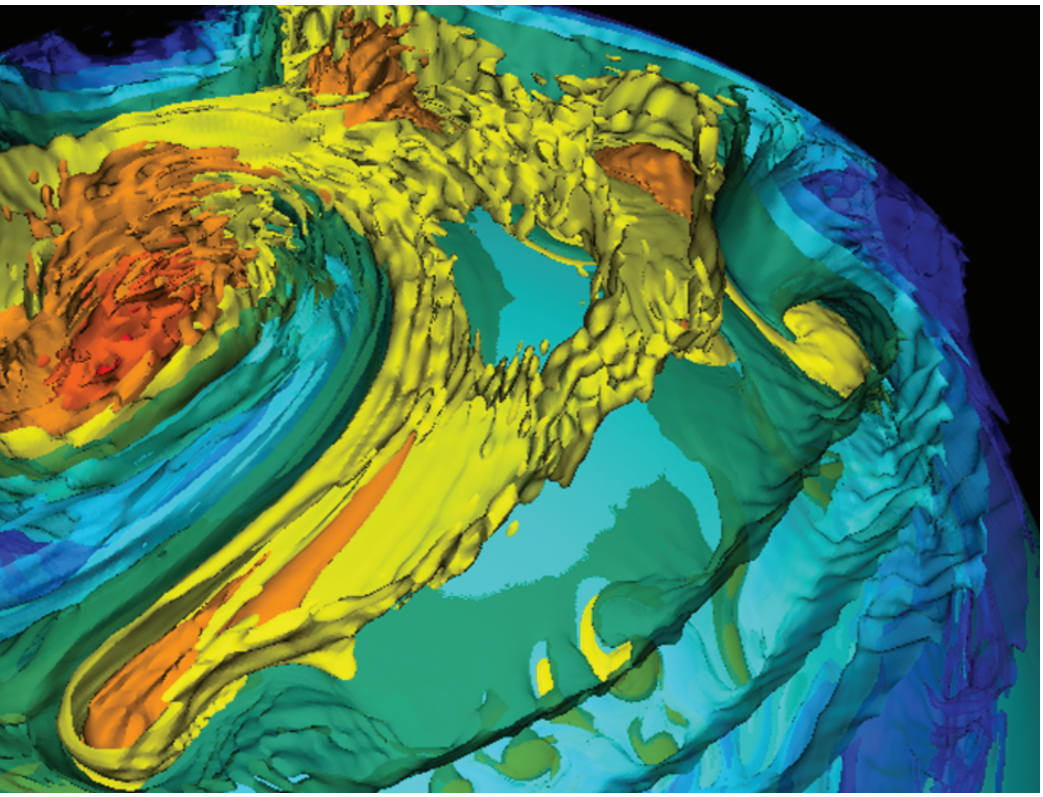
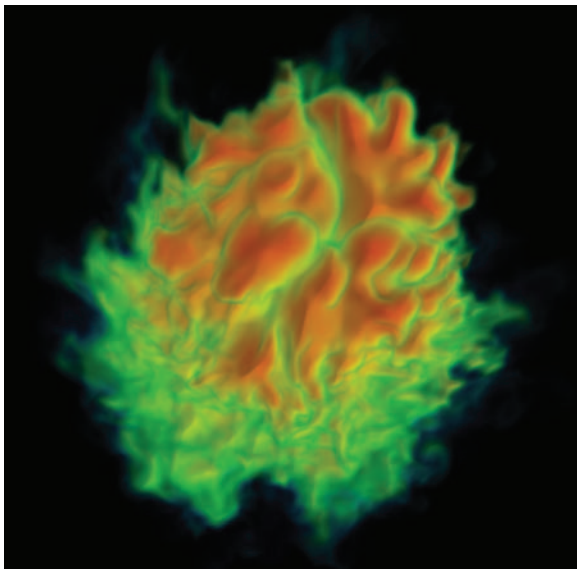
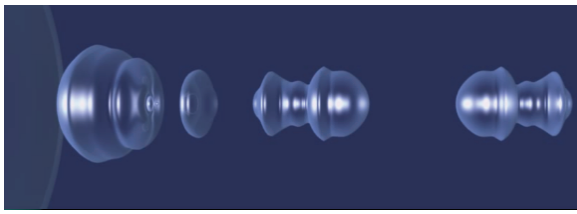
BERKELEY LAB







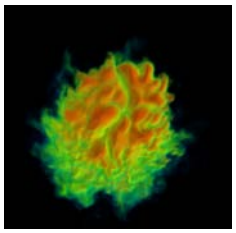
LAWRENCE BERKELEY NATIONAL LABORATORY

Computational Research Division

2009-2012

COVER IMAGES

Combustion Simulation

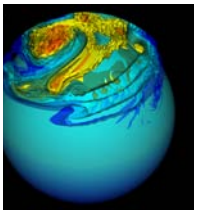


This 4608^2 image of a combustion simulation result was rendered by a hybrid-parallel (MPI+pthreads) raycasting volume rendering implementation running on 216,000 cores of the JaguarPF supercomputer. Combustion simulation data courtesy of J. Bell and M. Day (Center for Computional Sciences and Engineering, LBNL). For more information see "MPI-hybrid Parallelism for Volume Rendering on Large, Multi-core Systems" by M. Howison, E.W. Bethel, and H. Childs.

Fluid Pinch-off

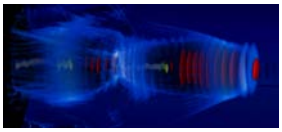
The CRD Math Group has built a new mathematical model applicable to the computations of fluid pinch-off. Their method correctly computes asymptotic constants of fluid self-similarity as breakup is approached, and the resulting rapid reconfiguration due to surface capillary waves and separation into microdroplets.

Global Cloud Resolving Model for Climate



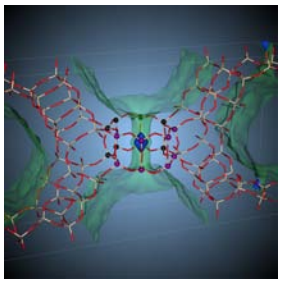
Accurate modeling of Earth's climate is one of the biggest challenges facing HPC today. One of the largest sources of error in existing climate models is the simulation of clouds. By increasing the grid resolution from 200 km to 2 km, clouds can be accurately resolved in simulations. Isocontours are plotted corresponding to different vorticity levels in the atmosphere. [Data provided by Dave Randall, Celal Konor and Ross Heikes (Colorado State Univ). Joint work with Karen Schuchardt, Bruce Palmer, Annette Koontz and Jeff Daily (PNNL). Visualization by the CRD Visualization Group's Prabhat.]

Laser Wakefield Particle Acceleration



Computational model of laser wakefield particle acceleration. Models of high-quality electron beams help scientists in the LOASIS program better understand beam acceleration dynamics. For more information see "Automatic Beam Path Analysis of Laser Wakefield Particle Acceleration Data" by Rübel et al. [Image by Oliver Rübel, CRD Visualization Group.]

Zeolite Structures



Visualization of zeolite material showing its bonding structure, and an isosurface visualization of the energy landscape for a diffusing carbon dioxide molecule (blue region indicates strong adsorption). By using Zeo++, they can identify carbon dioxide adsorbing materials much faster and by using far less computing power than through molecular simulation. This work is presented in more detail in a joint 2012 ChemPhysChem journal paper by Richard L. Martin et al. [Image generated by Richard L. Martin of the Scientific Computing Group using VisIt.]

LAWRENCE BERKELEY NATIONAL LABORATORY

Computational Research Division

2009-2012

CONTENTS

LBNL COMPUTATIONAL RESEARCH DIVISION

STRATEGIES, COMPETENCIES AND GOALS

David L. Brown
Division Director
January 2013

Summary

Lawrence Berkeley National Laboratory (LBNL) is a leader in mathematics and computer science with applications to research spanning many scientific disciplines, which are important to the U.S. Department of Energy (DOE) and the nation generally. LBNL's core strengths in computing sciences lie with its world-leading team of applied mathematicians, computer, data and computational scientists in the Computational Research Division (CRD), the research arm of LBNL's Computing Sciences Area. Spread across fourteen groups, CRD hosts approximately 150 researchers who develop mathematical models, algorithms, tools and software in support of computationally enabled scientific discovery. CRD supports the overall mission of the Computing Sciences Area, which is *to accelerate scientific discovery across a broad scientific community using advanced computing and mathematics*. This, in turn, supports the overall mission of LBNL, by addressing the most pressing scientific and technical challenges of our time, to transform the world energy economy and to provide a sustainable future for humankind.

CRD's research strengths can be described conveniently in terms of four core competencies, which are listed below:

» Applied Mathematics,

» Advanced Computer Science Technologies,

» Scientific Data Management and Analytics, and

» Software Technology Development.

The first three are research competencies, while the fourth is a supporting competency that cuts across the others. While many venues outside Berkeley Lab focus on the first three competencies as independent research focus areas, it is a hallmark of LBNL's computational research that a synergistic application of capabilities, from all of these competencies, is brought to bear on the science and engineering challenges being addressed. A major strength of LBNL computational research has been in the development of mathematics, algorithms and software supporting large-scale **modeling and simulation**. In the future, as large-scale experimental and observational science becomes increasingly dependent on mathematical and computational advances, computational solutions that enable scientific discovery will increasingly require a combination of modeling and simulation, advanced computer science technologies, scientific data management and analytics, integrated effectively into advanced **scientific data ecosystems**.

Our Science Engagement Strategy

The Computing Sciences Area's main strategy for accelerating scientific discovery across a broad scientific computing community is **science engagement**; the direct engagement of scientists in other domains to collaboratively advance scientific discovery by using advanced mathematical and computational technology. In many cases, our research staff members themselves have considerable credentials as scientists in science domains outside mathematics and computer science. In CRD, this science engagement strategy has driven our substantial track record in the development of mathematical models, algorithms, simulation and analysis tools, including exceptional software tools that have had a broad impact on DOE and industrial scientific and engineering applications, as well as on our own fields of applied mathematics and computer science.

As we look to the future, the rapidly growing data challenges at DOE's scientific user facilities, many of which are located at LBNL, drive a need for the development of advanced computational technology to enhance scientific discovery at those facilities. Over the past few years, we have begun to engage scientists in LBNL's other science divisions, many of which are responsible for DOE scientific user facilities, to understand and address their emerging computational challenges. We see significant research challenges that require the development of new analysis methods, new mathematical models and algorithms instantiated in highly effective simulation and analysis codes. The need for the development of scientific data ecosystems to address the management, transmission and analysis of data has become increasingly evident. As we have done for many science domains outside LBNL in the past, we will seek to provide LBNL science enterprises with advanced software technology solutions to enable and accelerate scientific discovery. Our vision is to significantly strengthen LBNL science through advanced mathematical and computer science research and advanced software technology.

CRD's Core Competencies and Future Goals

Applied Mathematics: CRD has one of the strongest, if not the strongest, applied mathematics competencies among the DOE National Laboratories. Most of the applied and computational mathematicians in CRD are found in the Center for Computational Sciences and Engineering, the Mathematics Group, the Applied Numerical Algorithms Group, the Complex Systems Group, and the Scientific Computing Group. Five of our applied mathematicians have been elected to the U.S. National Academies of Science and Engineering.[1] Berkeley Lab applied mathematicians are recognized internationally for their pioneering work in methods, algorithms and software for modeling and simulating complex scientific and engineering phenomena. Areas of research include high-resolution adaptive finite difference algorithms for multi-scale, multi-physics problems, techniques for computing complex flows with moving interfaces, numerical linear algebra, vortex and particle methods, and turbulence theory, among others.

Applied mathematics research is at the core of the U.S. Department of Energy's (DOE) Advanced Scientific Computing Research activities. CRD's goals for future developments in applied mathematics are to

» Continue our international leadership in the development of state-of-the-art mathematics, algorithms and computational techniques, impacting science and engineering challenges from DOE, other government agencies and industry;

» Develop simulation and analysis software, in collaboration with computer science teams, that provides exceptional performance on computing architectures from the desktop to the extreme-scale.

[1] National Academy of Sciences: John Bell, Alexandre Chorin, Philip Colella, James Demmel; National Academy of Engineering: James Demmel, James Sethian.

Advanced Computer Science Technologies: CRD's advanced CS effort is focused on computer science research that will continue to enable DOE scientists and engineers to use of the most advanced computational hardware for discovery science. Our research activities in this area are found predominantly in the Future Technologies Group. The current revolution in high performance computing hardware will result in a technology transition similar in impact to the transition from vector to MPP computing in the early 1990's. CRD's advanced CS research is focused on the hardware and software issues associated with this transition.

CRD's goals for future developments in advanced CS technologies are to

» Over the next decade, participate in computing architecture developments that can deliver one hundred times more usable operations per watt, thus enabling productive use of exascale technology for scientific computing within a realistic power budget;

» Translate these emerging architectural trends into technology requirements and software technology artifacts for operating systems, system software, languages, programming models and programming environments that will support productive exascale computing.

Scientific Data Management and Analytics: Scientific data management, storage, transport and analysis are significant core competencies within CRD, but continue to be exponentially growing challenges for DOE Office of Science (SC) Scientific User Facilities and large-scale modeling and simulation efforts. This is a technology-driven trend, largely due to the massive increase in both computational and observational capabilities. In addition, datasets become larger and prohibitively expensive to produce and/or reproduce. Therefore, it is increasingly important to store, curate and re-use these data for scientific discovery. CRD's R&D efforts supporting the development of the modern scientific data ecosystem will enable science to move forward with data as a first-class entity. Our goal for future development in this area is to:

» Expand our leadership in scientific data management and analysis by developing scientific data ecosystems that address the increasing challenges in this area faced by Scientific User Facilities and large-scale simulation codes.

Software Technology Development: CRD has a substantial track record in the production of advanced computational technology in the form of skillfully engineered software that leverages our mathematics and computer science research strengths, providing exceptional tools for scientific discovery. We develop and provide state-of-the-art software artifacts for broad use throughout the external community, and also implement the supporting computational technology required for basic and applied research activities within our organization supporting all three of the previous core competencies. Often these softwares are one in the same: much of the software infrastructure developed for internal purposes has been made available to external users as well. Software tools developed by CRD support scientific research in other science fields from high-energy physics, cosmology, climate, environmental science and biology to alternative energy, industrial processes and manufacturing.

A recent trend has been the increasing demand for scientific software solutions within Berkeley Lab. Advanced mathematics and computation have not traditionally been integrated into the largely experimental and observational science activities at LBNL, but rapidly increasing data collection capabilities at the LBNL scientific user facilities have created a need for advanced software solutions. CRD has begun to develop such solutions, again leveraging our mathematics and computer science expertise and skills, but there is no clear business model for supporting the need for a rapidly growing, agile and flexible computational technology workforce. To continue to be successful in bringing needed computational technology to LBNL science problems, it is important that we:

» Continue the production of advanced software technology solutions that enable and accelerate scientific discovery for the DOE, while developing a sustainable business model that will enable CRD to more broadly support LBNL with state-of-the-art software technology solutions.

The following three sections discuss our activities in the core competency areas of Applied Mathematics, Advanced Computer Science Technologies and Scientific Data Management and Analytics. Discussion of our work in software technology development cuts across these three competency areas and is included in all three sections. More complete information can be found by referring to CRD's website, located at http://crd.lbl.gov.

Applied Mathematics

Applied mathematics research is at the core of the U.S. DOE Advanced Scientific Computing Research (ASCR) activities, and CRD researchers have a long and distinguished track record in the development of advanced mathematical models, algorithms and simulation codes to enable and accelerate scientific discovery using computers of all scales from the desktop to the extreme scale. CRD applied mathematics researchers have had a broad impact worldwide, in many cases pioneering the development and application of new mathematical and algorithmic techniques for science fields ranging from combustion, materials science and astrophysics to biology, climate and the earth sciences. In addition, CRD research has resulted in the development of enhanced industrial processes and manufacturing techniques. The Center for Computational Sciences and Engineering (CCSE), the Applied Numerical Algorithms Group (ANAG), the Mathematics Group, the Complex Systems Group (CXG) and the Scientific Computing Group (SCG) focus on the development of new models, methods, algorithms and software for modeling and simulation. Researchers in SCG and the Center for Computational Cosmology (C3) use advanced simulations for their research in climate modeling, materials science and cosmology.

The **Center for Computational Sciences and Engineering** (CCSE) develops and applies advanced computational methodologies to solve large-scale scientific and engineering problems arising in DOE mission areas involving energy, environment, and industrial technology. CCSE focuses on the design of algorithms for multiscale, multiphysics problems described by nonlinear systems of partial differential equations, and in developing implementations of algorithms that target current and next-generation massively parallel computational architectures. CCSE researchers work collaboratively with application scientists to develop state-of-the art solution methodologies in the fields of combustion, porous media flow, fluctuating hydrodynamics, atmospheric modeling, cosmology and astrophysics. Leveraging advanced methods developed to model incompressible and low speed flows, CCSE has built a capability for numerically simulating laboratory-scale combustion devices and advanced ultra-low-emissions burners. Similar approaches have enabled the development of high-fidelity groundwater flow simulations, which have the potential for providing valuable insights into the long-term fate of groundwater contaminants. CCSE's low-Mach number astrophysics code has been used to perform full-star simulations of convection in a white dwarf, leading up to the ignition of a Type Ia supernova. Compressible astrophysics methodology based on similar technology has been applied to Type Ia and Type II supernovae as well as other astrophysical phenomena. Recent research in partnership with C3 has led to the development of the Nyx code to simulate gas and dark matter for cosmological studies. More recent research in CCSE has focused on the development of stochastic hybrid models and algorithms that include the effects of thermally induced fluctuations at molecular scales to provide a more realistic treatment of systems undergoing phase transitions, nucleation, barrier-crossing, Brownian motors, noise-driven instabilities and combustive ignition. CCSE is also a major participant in the ExaCT Co-Design center whose goal is to ensure that next generation extreme scale computer architectures will provide an effective computing environment for future combustion simulations.Recent research in partnership with C^3 has led to the development of the Nyx code to simulate gas and dark matter for cosmological studies. More recent research in CCSE has focused on the development of stochastic hybrid models and algorithms that include the effects of thermally induced fluctuations at molecular scales to provide a

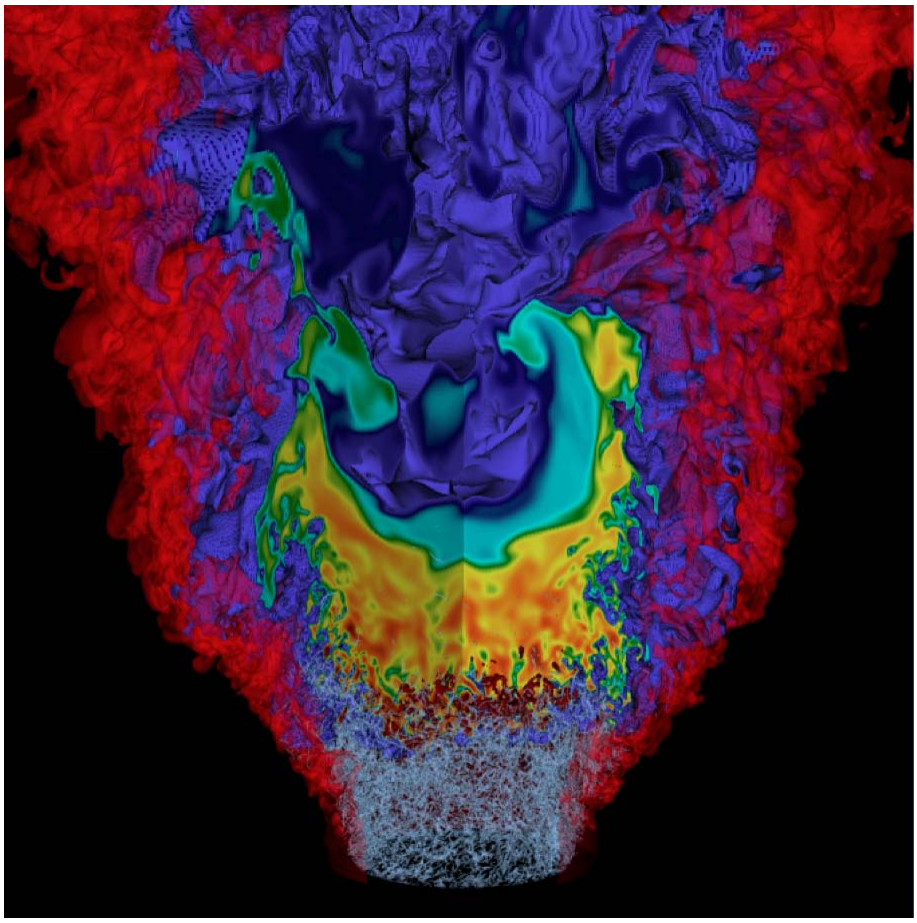


Figure 1: CCSE's high fidelity computational tools are leveraged to deliver an unprecedented ability to simulate high pressure flames, important for understanding next-generation stationary turbines based on low swirl injectors. This image shows the result of simulating NOx formation in a low swirl injector using a low-Mach number formulation combined with adaptive mesh refinement.

more realistic treatment of systems undergoing phase transitions, nucleation, barrier-crossing, Brownian motors, noise-driven instabilities and combustive ignition. CCSE is also a major participant in the ExaCT Co-Design Center whose goal is to ensure that next generation extreme-scale computer architectures will provide an effective computing environment for future combustion simulations.

The **Applied Numerical Algorithms Group** (ANAG) develops advanced numerical algorithms and software for partial differential equations integrated with the application of their software to problems of independent scientific and engineering interest. ANAG's particular focus is on high-resolution and adaptive finite difference methods for partial differential equations in complex geometries with applications to internal combustion engines, climate, and subsurface modeling. Recent work has focused on adaptive mesh, finite volume modeling of ice sheets, the development of an adaptive dynamical core for climate applications, and reactive transport for carbon sequestration. ANAG makes its computational technology available externally through the Chombo framework, which supports finite difference and finite volume discretizations of partial differential equations on block-structured adaptively-refined grids. Chombo includes embedded boundary and mapped multi-block methods for the representation of complex geometries, particle and particle-mesh algorithms within its Adaptive Mesh Refinement (AMR) framework, and higher-order methods in time and space for increased computational resolution. ANAG collaborates broadly across DOE through the SciDAC program, and participates in several collaborative

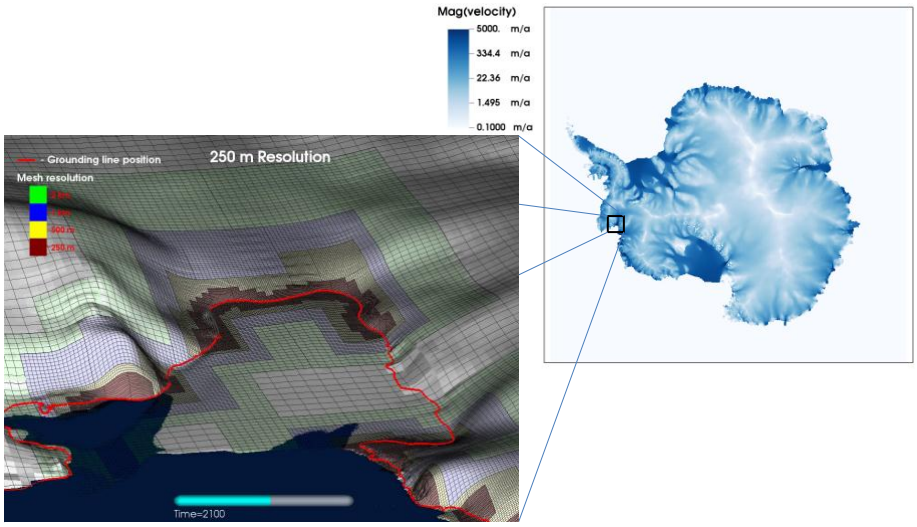


Figure 2: ANAG is developing scalable ARM ice sheet modeling capability in support of the Community Ice Sheet Model. This schematic shows computed ice velocity for Antarctica (right), and (left) meshing and grounding line location for the Pine Island Glacier.

efforts with computer scientists in CRD, including code optimization, compiler transformation techniques and overall numerical implementations for the CACHE project, abstract numerical algorithm specification and interaction with exascale co-design centers for the D-TEC project, and the demonstration of auto-tuning on scientific computing codes through the XTUNE X-stack project. In collaboration with LLNL, ANAG is also involved in the development of high-resolution methods for solving continuum kinetic systems in complex phase space geometries for plasma fusion modeling.

The **Mathematics Group** at LBNL develops new mathematical models, devises new algorithms, explores new applications, exports key technologies, and trains young scientists in support of DOE. Researchers use mathematical tools from a variety of areas in mathematics, physics, statistics, and computer science, including statistical physics, differential geometry, asymptotic analysis, graph theory, partial differential equations, discrete mathematics, and combinatorics. The Math Group attacks both technologically interesting and mathematically challenging problems, which has resulted in an impressive set of interrelated computing methodologies, and applications that have been applied to support the DOE energy mission and impact science and industrial challenges worldwide. Their recent research has developed methods for implicit sampling of many-dimensional non-Gaussian probability densities with a potential impact in fields as wide-ranging as economics and finance, weather forecasting, quantum field theory and statistics. The Math Group has also developed high-order methods for fluid-structure interaction with applications to vertical axis wind turbine simulations. Recent breakthrough work in mathematical and algorithmic methodologies for computing multiphase physics problems will have an impact on the modeling and understanding of industrial foams, materials, manufacturing and biological cell mechanics. The Math Group is also developing new computational tools to study dynamic problems with time-periodic forcing or symmetry. A key step in this research is to devise adjoint-based optimization techniques suitable for large-scale systems governed by nonlinear evolutionary partial differential equations.

The **Complex Systems Group** (CXG) was formed in 2010 to address the increasingly important role that mathematical research is playing in the analysis and optimization of complex systems, including application areas relevant to the DOE mission, such as networked interconnected systems, like computer

networks and electrical power grids. In addition to collaborative work with other groups in CRD, CXG works collaboratively with UC Davis, LLNL and Sandia Laboratories on a wide variety of topics in computer security. CXG also hosts the Center for Innovative Financial Technology (CIFT), which studies the behavior of the increasingly sophisticated, computationally driven aspects of the US and world financial markets.

The **Scientific Computing Group** (SCG) performs research to advance scientific computing by developing and enhancing applications in key scientific disciplines. It also develops tools and libraries for addressing general problems in computational science. Members of SCG work on diverse scientific applications in a wide variety of fields including atmospheric modeling, materials science and nano-science, chemical sciences, and biological sciences with relevance to LBNL and DOE. SCG also tackles generic problems in scientific computing, developing new algorithms and software for dense and sparse linear and eigen systems, mathematical libraries and templates, and software tools. SCG develops scalable sparse direct and hybrid linear solvers, and effective preconditioners, delivering much of this capability through SuperLU, a sparse solver package that is widely used internationally. SCG also develops scalable and robust algorithms and solvers for tackling large-scale linear and nonlinear eigenvalue problems using a wide variety of modern techniques. They also develop fast numerical methods for performing ground-state electronic structure analysis based on the Kohn-Sham density functional theory and numerical methods for excited state electronic structure analysis, important in a number of applications such as the design of photovoltaic devices and energy efficient batteries. In partnership with the LBNL Life Sciences, Chemical Sciences and Materials Science divisions, members of SCG are also performing research in a number of different areas in materials science, including methods for high-throughput discovery of improved scintillator materials, new methods for simulating f-electron soft X-ray spectroscopy, informatics tools for the analysis of void space in porous materials, similarity-driven discovery of porous materials for gas separations, and studying high surface area materials such as metal-organic frameworks. Collaborating with scientists at the Advanced Light Source (ALS), SCG has developed new theory, computational methods and HPC code for computing X-ray scattering patterns, with particular application to glancing-incidence small-angle X-ray scattering (GISAXS). Also relevant to current light sources, such as LCLS, the planned future light source, NGLS, and high-resolution electron microscopes is the development of efficient and robust algorithms for reconstructing 3D structures of macromolecules from 2D electron microscopy or X-ray images with unknown relative orientations. Recently, SCG has also begun to develop the theoretical tools and knowledge to understand pump-probe spectroscopy, a rapidly growing field capable of elucidating many aspects of complex materials physics. Finally, in collaboration with the LBNL Earth Sciences Division and other scientists in the climate community, SCG researchers are creating a large number of simulations of the CAM5.1 climate model, resolving features down to 25 km globally, to quantify how the chances of regional and local extreme events depend on past and future emissions.

CRD researchers play a significant role in the DOE SciDAC-3 **FASTMath** Institute, which develops and deploys scalable mathematical algorithms and software tools for reliable simulation of complex physical phenomena. CRD leverages its technical expertise to support a wide variety of applications in finite volume methods and particle/mesh methods in structured grid AMR frameworks, high order discretization methods, highly optimized linear and nonlinear solvers and eigensolvers. Software developed by FASTMath team

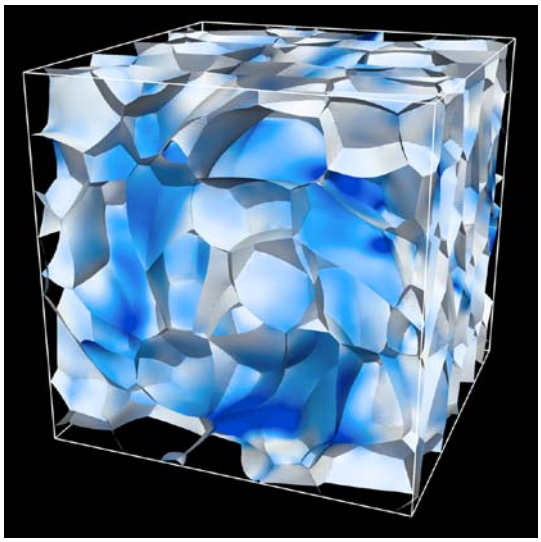


Figure 3: The Mathematics Group is developing mathematical and algorithmic methodologies for computing multi-phase physics with applications to industrial foams, materials, manufacturing, and biological cell mechanics. This figure shows an example computation for fluid flow and interface forces in agitators for fluid mixing in industrial devices.

members at LBNL comprises the framework and/or solver technology for application codes in accelerator modeling, astrophysics, climate, combustion, cosmology, materials science and porous media.

In addition, CRD researchers contribute to eleven **SciDAC-3 Application Partnerships**, bringing advanced mathematics and computer science to science applications that span the DOE Office of Science (SC). The SciDAC program is supported through partnerships between DOE ASCR and the other science programs in SC. The program brings together mathematicians, computer scientists and domains scientists from across the DOE Laboratory complex and U.S. academia to address difficult scientific challenges of interest to DOE through high performance computation. The partnerships supporting the climate program within the Biological and Environmental Research (BER) program will develop multi-scale atmospheric and oceanic parameterizations and improved ice sheet models for the Community Earth System Model (CESM) and the Community Ice Sheet Model (CISM). For the High Energy Physics (HEP) program, HPC tools and applications that support particle accelerator design will be developed and advanced simulation and analysis tools will be used to support cosmology experiments. For the Nuclear Physics (NP) program, large-scale computations will transform the fields of low-energy nuclear physics and astrophysics by addressing nuclear interactions and their uncertainties, *ab-initio* studies of light nuclei and their reactions, and studies of nucleonic matter and its astrophysical properties. Also for NP, CRD is partnered on the development of algorithms that support the linkage of a non-relativistic effective theory of nuclear physics to the exact theory of the strong interaction, quantum chromodynamics (QCD), by connecting the low-energy constants of the former to lattice QCD calculations of nucleon-nucleon scattering parameters.

CRD researchers also contribute to five of the SciDAC projects that partner ASCR with the Basic Energy Sciences (BES) program. These projects address simulations through first principles methods of the generation, evolution and fate of electronic excitations in molecular and nano-scale materials. They also address simulations through the development of advanced methods for excited state chemistry for the NWChem software suite, advanced modeling of ions in solutions, on surfaces and in biological environments, discontinuous methods for accurate, massively parallel quantum molecular dynamics, and scalable computational tools for the discovery and design of excited state phenomena in energy materials.

Advanced Computer Science Technologies

DOE Computational Science practitioners are facing a revolution in computing capability that promises to be as equally disruptive as the transition from vector to massively parallel computing in the 1990's. This revolution is driven by energy constraints—it is widely understood that an evolutionary development of high-end computers will not be possible within realistic DOE power budgets over the next decade. DOE and LBNL are addressing this challenge through a multi-pronged approach that includes the co-design of algorithms, software and hardware for targeted DOE applications. The computer science research efforts will require new approaches to programming, and the development of a software infrastructure that can support sophisticated users of emerging high-end architectures. This research will focus on the computer technology aspects of this revolution, but must mesh with concurrent developments and improvements in algorithm design for scientific software.

Current CRD research in advanced computer science technology occurs mostly in the Future Technologies Group (FTG) and focuses on understanding the performance and development of advanced software environments for today's systems.

Research supporting advanced software environments includes a substantial number of projects designed to enable high-performance parallel computing on today's systems. **GASNet** (Global-Address Space Networking) is a language-independent, low-level networking layer that provides network-independent high performance one-sided communication primitives. The library has been adopted by many vendors for their production software for one-sided messaging, and has been a critical for the

transport requirements for advanced research in parallel program languages including IBM's X10, Cray Chapel, and multiple implementations of Co-Array Fortran. **GASNet** is being extended through the DEGAS (Dynamic Exascale Global Address Space) project to support emerging programming environments that have asynchronous execution models to support exascale computing and advanced data analytics. **UPC** (Unified Parallel C) is an extension of the C programming language designed for high performance computing on large-scale parallel machines. The language provides a uniform programming model for both shared and distributed-memory hardware.

The **BLCR** (Berkeley Lab Checkpoint-Restart) library provides checkpoint-restart capability for parallel applications that communicate through MPI and is now included in every available implementation of the Linux operating system. Automatic Performance Tuning (or **Auto-tuning**) provides an effective means for providing performance portability across diverse hardware architectures (from GPU's to many-core CPUs). CRD's autotuning techniques have been developed and applied to multigrid, graph algorithms and finite difference stencil operations. CRD researchers also participate in UC Berkeley Par Lab's **SEJITS** (Selective Embedded Just-In-Time Specialization) project, which provides a methodology for developing Domain-Specific Embedded Languages (**DSEL**s). DSELs provide an interoperable approach to embedding a DSL (domain-specific language) into an existing language so that programmers can seamlessly use the DSL as well as features of the host language.

Looking towards the future, a number of advanced CS projects have initiated recently. The **DEGAS** Programming Environments project was funded as part of DOE's X-Stack2 program. DEGAS will develop the next generation of programming models, runtime systems and tools to meet the challenges of exascale systems. The objective is to develop a solution that will address the key challenges of scalability, programmability, performance portability, resilience, energy efficiency and interoperability posed by anticipated exascale systems. Also in the programming systems area, the **Tesselation** OS, developed in collaboration with UC Berkeley Par Lab, implements a modular OS structured around space-time partitioning and two-level scheduling between the global and partition runtimes. The partitioning, advanced resource management, and performance isolation features of **Tesselation** are viewed as essential for many-core client operating systems.

To help coordinate extreme-scale computing hardware architecture design across DOE, **CAL** was recently established as a joint SC/NNSA project in collaboration with Sandia National Laboratories. CAL will provide a common modeling and simulation infrastructure for the quantitative analysis of the trade-space for hardware and algorithmic software design. It will make use of advanced architectural modeling and simulation tools to explore novel hardware/architectural concepts in cooperation with industry partners and the DOE research community. This research intersects with DOE's **FastForward** effort, which partners with computer vendors to understand and develop CPUs, memory, storage and I/O in support of the creation of exascale computers over the next decade. Berkeley Lab also co-leads the **ExaCT** Combustion Co-design center for DOE, which works closely with the Computer Architecture Laboratory to better communicate the requirements of applications to industry hardware design teams, and also communicate information about future hardware constraints to advanced application development teams in DOE. DOE recognizes that a highly integrated hardware-software co-design process is essential to foster the rapid pace of innovation needed to create exascale computing systems that meet DOE mission needs. ExaCT brings together applied mathematicians, computer scientists, data scientists and combustion scientists to examine the entire computing workflow associated with the modeling and simulation of combustion processes relevant to the burning of alternative fuels.

CRD also contributes to **SUPER** (Institute for Sustained Performance, Energy, and Reslience), a broadly based SciDAC institute whose members have expertise in compilers and other system tools, performance engineering, energy management and resilience. SUPER ensures that DOE's computational scientists can successfully exploit the emerging generation of high performance computing (HPC) systems by providing application scientists with strategies and tools to productively maximize performance, conserve energy, and attain resilience.

From Scientific Data Management and Analytics to Scientific Data Ecosystems

In the past, the larger scientific data management and analysis challenges at DOE have been associated with the results from computational modeling and simulation. CRD's solutions for these challenges have focused on the management of datasets, data mining and visualization. Experimentally or observationally derived data has often been easier to deal with. These data have been easily manageable, with datasets easily stored and analyzed on desktop computers. With some exceptions, data is analyzed for scientific results and then discarded. This picture is changing as the technology for data collection improves rapidly. For example, CCD-based imaging technology is experiencing a Moore's-law-like exponential growth in the amount of data that can be collected.

Science domains, such as high energy physics, cosmology and climate, have been dealing with significant amounts of data for some time now, and custom computational tools to address the data workflows have been developed. However, as new instrumentation is deployed at a rapid pace, the need for these workflow solutions is no longer confined to a few isolated projects; at LBNL, we are seeing significant data challenges at many of the Advanced Light Source beamlines, the Joint Genome Institute, the National Center or Electron Microscopy, the Molecular Foundry, and similar facilities at other DOE Laboratories. We expect to see these challenges coming from emerging external facilities, such as the Square Kilometer Array and the LSST. It is no longer economically feasible to provide custom solutions for every data collection case. Individual PI's cannot write analysis software and perform the analysis for every experiment. As the data challenges become more complex, there is no equivalent set of data workflow libraries, as with MPI, LAPack, Chombo, PetSC for the modeling and simulation space.

Additionally, as datasets become larger, they become more difficult to store, analyze and reproduce. We expect to see a trend moving away from an individual PI's closely guarded dataset to produce a single research result. Instead, we expect to see the trend move toward the storage, curation, and availability of large datasets that can be accessed, combined, and used as a resource for general scientific discovery. An emerging concept that captures the essence of a new paradigm for data-driven scientific discovery is the *scientific data ecosystem*. This not only includes the software and hardware tools required to manage the entire data collection and analysis workflow, but also the scientists contributing and curating the data. Much like a physical ecosystem, the scientific data ecosystem evolves as data is added, and as data collection, management and analysis capabilities improve, supporting increasingly sophisticated scientific discovery from complex and often massive datasets.

The development of tools in this area is being supported by researchers in the Advanced Computing for Science (ACS) Department, the Biological Data Management and Technology Center (BDMTC) and in the Scientific Data Management (SDM), Computational Cosmology Center (C^3) and Visualization groups. The **Advanced Computing for Science Department** is developing software frameworks to enable the interoperability of software tools for DOE Fossil Energy's **Carbon Capture and Simulation Initiative**. They have developed an eco-informatics data management infrastructure that combines data processing, archiving, curation, user services and publication tracking to support a wide array of carbon data and subsurface flow projects, such as **ISCN, AmeriFlux, FLUXNET**, and digital hydrology projects. ACS was responsible for architecting the software infrastructure that supports the **ATLAS** experiment at CERN, recently recognized for its contribution to the discovery of the Higgs boson, Science Magazine's "Breakthrough of the Year" for 2012. In support of another one of the top 10 Science Breakthroughs of 2012, ACS was also the US lead for the software and computing support for the **Daya Bay Neutrino Experiment** in China. This software was essential in the successful measurement of the last unknown parameter θ_{13}, which describes how neutrinos transform at near-light-speed. Other data management projects, developed or under development by ACS, include the **Particle Data Group Workspace**, the **Advanced Light Source User Portal**, and the data management infrastructure for the **Materials Project**.

Recent research areas for the **Visualization Group** have included the development of topological data analysis methods, and the exploration of novel computer architectures for visualization, such as those using GPUs and SSDs. The group also works on the development of methods to scale visualization tools to tens of thousands of cores, and in some specialized cases, such as volume rendering, to hundreds of thousands of cores, opening the door for in situ visualization within very large simulation and analysis codes. The Visualization Group also supports the R&D 100 award-winning **VisIt** visualization system at NERSC, which helps users tailor the tool to specific user needs, sometimes substantially changing the infrastructure to support new types of data, such as adaptive block-structured meshes. The Visualization Group has developed visualization and analysis tools and methods that have supported DOE SciDAC scientists through the VACET Center and now the SDAV Institute. Members of the Visualization Group edited and published the book "High Performance Visualization: Enabling Extreme-Scale Scientific Insight", editors Wes Bethel and Hank Childs, released in 2012.

The LBNL-led **SDAV** Institute, announced in March 2012 as part of the President's Big Data Research and Development Initiative, also includes broadly applicable data management tools, some of which have been developed by the **SDM Group**. SDM is also responsible for the development of the **Advanced Performance Model** (APM), which improves the efficiency of resource utilization and the scheduling of scientific data transfers on high-speed networks. ADAPT, the **Adaptive Data Access and Policy-Driven Transfers** project, is a general-purpose data access framework for scientific collaborations that provides light-weight performance monitoring and estimation, fine-grained and adaptive data transfer management, and enforcement of site and VO policies for resource sharing. SDM is also well known for the R&D 100 award-winning **FastBit**, an efficient compressed bitmap index technology that has been widely used and adapted to search a variety of very large scientific datasets. Members of the SDM Group edited and published the book "Scientific Data Management: Challenges, Technology, and Deployment, editors Arie Shoshani and Doron Rotem, released in 2009.

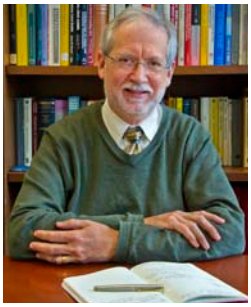
The SDM Group, the Visualization Group and the ACS Department have also combined their expertise to build data management and visualization capabilities for the Advanced Simulation Capabilities for Environmental Management (**ASCEM**) project.

The Biological Data Management and Technology Center has developed the **IMG** (Integrated Microbial Genomes) systems, which are a suite of microbial genome and metagenome data management and analysis systems that support the DOE Joint Genome Institute, as well as genomics researchers internationally. IMG supports operation involving genomes and/or metagenome datasets, genes, and functions that can be selected and explored individually or compared.

In recent years astrophysics has undergone a renaissance, transforming from a data-starved to a data-driven science. A new generation of experiments will gather datasets so massive that their analysis will require the use of leading-edge, high performance computing resources. Continuing a decade-long collaboration in this field, the Computational Research and Physics Divisions at LBNL have formed the **Computational Cosmology Center** (C^3). Recent successes in this group include the discovery of supernova PTF 111kly through the use of machine learning algorithms in the Real-time Transient Detection Pipeline running at NERSC. C^3 is also developing the computing tools needed to analyze data from the coming generation of cosmic microwave background polarization experiments, and applying them to the Planck satellite, Polarbear ground-based and EBEX balloon missions.

DIVISION DIRECTORS

DAVID L. BROWN



David Brown became Director of the LBNL Computational Research Division in August 2011. His 28-year career with the U. S. Department of Energy (DOE) National Laboratories includes fourteen years at Los Alamos National Laboratory (LANL) and thirteen years at Lawrence Livermore National Laboratory (LLNL), where most recently he served as the Deputy Associate Director for Science and Technology in the Computation Directorate. Brown's research expertise and interests lie in the development and analysis of algorithms for the solution of partial differential equations (PDEs). In particular, his research has focused on adaptive composite overlapping grid techniques for solving PDEs in complex moving geometries and in the analysis of difference approximations for PDEs. At LANL and LLNL, he led the highly successful Overture project, which in 2001 was named one of the 100 "most important discoveries in the past 25 years" by the DOE Office of Science.

In 2007, Brown convened an independent panel from the applied math research community to investigate how past, present and future math research, supported by DOE's Office of Advanced Scientific Computing Research (ASCR), could be applied to help tackle challenges being addressed by both the DOE Office of Science and the offices of Nuclear Energy, Fossil Energy, Environmental Management, Legacy Management, Energy Efficiency and Renewable Energy, Electricity Delivery and Energy Reliability, and Civilian Radioactive Waste Management, as well as the National Nuclear Security Administration.

Brown earned his Ph.D. in Applied Mathematics from the California Institute of Technology in 1982. He also holds a B.S. in Physics and an M.S. in Geophysics from Stanford University.

JONATHAN CARTER

Jonathan Carter is both the Computing Sciences Area Deputy Director and the Computational Research Division Deputy Director.

Before assuming the Deputy Associate Lab Director role for Computing Sciences, Carter was leader of the NERSC User Services Group (USG). He joined NERSC as a consultant in USG at the end of 1996, helping users learn to effectively use the computing systems. He became leader of USG at the end of 2005. Carter maintains an active interest in algorithms and architectures for high-end computing, and regularly participates in benchmarking and procurement activities to deploy new systems for NERSC. As part of collaborations with the Future Technologies Group in CRD, and the NERSC Advanced Technology Group, he has published several architecture evaluation studies, and looked at what it takes to move common simulation algorithms to exotic architectures. His application work on the Japanese Earth Simulator earned him a nomination as Gordon Bell Prize finalist in 2005.

Before coming to LBNL, Carter worked at the IBM Almaden Research Center as a developer of computational chemistry methods and software, and as a researcher of chemical problems of interest to IBM. Carter holds a Ph.D. and B.S. in chemistry from the University of Sheffield, UK, and performed his postdoctoral work at the University of British Columbia, Canada.

DEPARTMENT HEADS

DEB AGARWAL is the Advanced Computing for Science Department Head and the CyberInfrastructure Development Group Lead. Agarwal's research focuses on scientific tools that enable the sharing of scientific experiments, advanced networking infrastructure to support sharing of scientific data, scientific computation frameworks, data analysis support infrastructure for eco-science, and cybersecurity infrastructure to secure collaborative environments. Some of the projects Agarwal is working on include: AmeriFlux data processing and management, advanced computational subsurface modeling data management, and infrastructure for carbon capture simulations. As a member of the Berkeley Water Center collaboration between UC Berkeley and LBNL, Agarwal also leads a team that develops data server infrastructure to significantly enhance data browsing and analysis capabilities, and enables eco-science synthesis at the watershed-scale to understand hydrologic and conservation questions and at the global-scale to understand carbon flux. Agarwal received her Ph.D. in electrical and computer engineering from UC Santa Barbara and a B.S. in mechanical engineering from Purdue University.

ESMOND G. NG received his Ph.D. in computer science from the University of Waterloo, Canada in 1983. He is currently a senior scientist and the Department Head of the Applied Mathematics and Scientific Computing Department, as well as the acting Group Lead for the Scientific Computing Group. He has been at LBNL since 1999. Prior to joining LBNL, Ng was a senior research staff member and a group leader in the Mathematical Sciences Section of the Computer Science and Mathematics Division at Oak Ridge National Laboratory (ORNL). While he was at ORNL, he was also an adjunct professor in the Department of Computer Science at the University of Tennessee, Knoxville. Ng is well known for his work on sparse matrix computation. The algorithms and codes that he has developed are widely used. His research also includes numerical linear algebra, graph algorithms and complexity analysis, parallel computing, and mathematical software development and software engineering. In addition, he is interested in applying sparse matrix techniques to scientific and engineering applications. Ng has served on the editorial boards for the SIAM Journal on Matrix Analysis and Applications and the IEEE Transactions on Parallel and Distributed Systems. He is currently an associate editor of the SIAM Journal on Scientific Computing and also an associate editor of the Bulletin of Computational Mathematics.

VICTOR M. MARKOWITZ, is Senior Scientist and head of the Biological Data Management and Technology Center (BDMTC). Markowitz joined LBNL in 1987 where he led the development of data management and integration tools used for developing public and commercial genome databases, such as the Genome Database at the Johns Hopkins School of Medicine in Baltimore, and the German Genome Resource Center's Primary Database in Berlin. In 1997 he joined Gene Logic Inc. as CIO and Senior Vice President, Data Management Systems, responsible for the development and deployment of Gene Logic's data management products and software tools, including its flagship Genesis, one of the largest commercial gene expression data management deployed at tens of pharmaceutical and biotech companies worldwide. In 2004, Markowitz returned to LBNL's CRD to head BDMTC, where he oversaw the development of the DOE Joint Genome Institute's Integrated Microbial Genomes (IMG) family of systems used by a large community of scientists worldwide for conducting microbial genome and metagenome studies. Markowitz received his M.S. and D.S. degrees in computer science from Technion, the

Israel Institute of Technology. He has conducted research in various data management and database areas, and has authored over sixty articles and book chapters on various aspects of databases and scientific data management. He has served on review panels and program committees for database and bioinformatics programs and conferences.

JOHN SHALF is the Department Head of the Computer and Data Sciences Department and the Chief Technology Officer (CTO) of the National Energy Research Scientific Computing (NERSC) Division. Shalf has served on the DOE Exascale Steering Committee and is co-principal investigator of the Computer Architecture Laboratory (CAL), a joint activity of Sandia National Laboratories and Lawrence Berkeley National Laboratory for the DOE's ASCR. CAL is coordinating hardware architecture research and development activities across the DOE. Shalf also leads the CRD/NERSC Green Flash project, which is developing a novel HPC system design (hardware and software) for kilometer-scale global climate modeling that is hundreds of times more energy efficient than conventional approaches. Shalf is a coauthor of over 60 publications in the field of software frameworks and HPC technology. He was a member of the LBNL/NERSC team that won a 2002 R&D 100 Award for the RAGE robot. Shalf received a bachelor's degree in electrical engineering and a master's degree in computer engineering from Virginia Polytechnic Institute and State University. Before joining LBNL in 2000, he was a research programmer at the National Center for Supercomputing Applications at the University of Illinois and a visiting scientist at the Max-Planck-Institut für Gravitationphysick/Albert Einstein Institute in Potsdam, Germany, where he co-developed the Cactus code framework for computational astrophysics.

JOHN BELL is the Department Head of MACS (the Mathematics and Computational Sciences Department) and the group leader of CCSE. He received his B.S. from MIT and his Ph.D. from Cornell University, both in mathematics. Before coming to LBNL in 1996, Bell was the group leader of the Applied Math Group at Lawrence Livermore National Laboratory. His research interests include development and analysis of numerical methods for PDEs and stochastic systems. He contributed to research in finite volume methods, low-mach number methods, adaptive mesh refinement, numerical methods for stochastic PDEs, and parallel computing. He has worked on applications of these methods in a number of areas including astrophysics, combustion, subsurface flow, and micro-fluidics. He is also the deputy director of the ExaCT Co-Design Center.

AWARDS

2012

Kathy Yelick, ACM Fellow

Kathy Yelick, Associate Laboratory Director for Computing Sciences, was named a Fellow of the Association for Computing Machinery (ACM). Yelick was recognized for contributions to parallel languages that improve programmer productivity.

E. Wes Bethel, ACM Distinguished Scientist

Wes Bethel, leader of the Visualization Group, was named an ACM Distinguished Scientist by the Association for Computing Machinery.

Sean Peisert, IEEE Senior Member

Sean Peisert of CRD's Complex Systems Group and UC Davis was named an IEEE Senior Member, which requires at least 10 years of professional practice and significant performance over at least five of those years.

Hank Childs, DOE Early Career Award

Hank Childs of the Visualization Group was honored with a 2012 DOE's Early Career Award for "Data Exploration at the Exascale."

John Bell Elected to the National Academy of Sciences

John Bell, an applied mathematician and computational scientist who leads the Center for Computational Sciences and Engineering and the Mathematics and Computational Science Department, was elected to the National Academy of Sciences.

James Sethian, American Mathematical Society Fellow

In 2012, the Math Group's leader, James Sethian, was elected as an American Mathematical Society Fellow.

James Sethian, Einstein Fellowship Prize

James Sethian was honored as a Visiting Einstein Fellow at the Berlin Mathematical School in 2011 and 2012, Einstein Stiftung.

2011

James Sethian and Robert Saye, 2011 Cozzarelli Prize

James Sethian and Robert Saye, mathematicians who both hold joint appointments at LBNL and the UC Berkeley, won the 2011 Cozzarelli Prize for the best scientific paper in the category of Engineering and Applied Sciences for "The Voronoi Implicit Interface Method for computing multi-phase physics." The Cozzarelli Prize is sponsored by the Proceedings of the National Academy of Science.

Berkeley's Vern Paxson Honored for Lifetime Contributions to Internet Measurement, Security

Vern Paxson, who holds joint appointments at UC Berkeley, the International Computer Science Institute and LBNL's CRD, was named recipient of this year's ACM SIGCOMM Award "for his seminal contributions to the fields of Internet measurement and Internet security, and for distinguished leadership and service to the Internet community."

Advanced Trading Magazine's "Top 10 Innovators of the Decade"

David Leinweber, cofounder of the Center for Innovative Financial Technology (CIFT) in CRD, was named by Advanced Trading magazine as one of its "Top 10 Innovators of the Decade" for his work in developing a service that allows trading strategies to react to news the instant it breaks, managing what the magazine describes as "a fire hose of aggregated updates."

National Academy of Sciences

James W. Demmel, a professor at UC Berkeley, who has a joint appointment in LBNL's CRD, was elected to the National Academy of Sciences (NAS). The software and standards Demmel developed over the last two decades enable users to transition their computer programs to new high performace computers without having to re-implement the basic building blocks. The software is used by hundreds of sites worldwide, including all US DOE National Laboratories, NASA research laboratories, many universities, and companies in the aerospace, automotive, chemical, computer, environmental, medical, oil, and pharmaceutical industries.

INCITS Merit Award

Bruce Bargmeyer, leader of the Advanced Computing for Science Department's Metadata, Semantics, and Ecoinformatics Group, was recognized with the InterNational Committee for Information Technology Standards (INCITS) Merit Award for 2011. Bargmeyer was cited for "your more than 15 years of participation and contributions to INCITS/DM32.8 Technical Committee and ISO/IEC JTC 1/SC32. Your leadership and broad knowledge of important issues and dedicated efforts are what lead to the successful development and advancement of standards."

Sloan Research Fellows

Per-Olof Persson of CRD's Mathematics Group and **Koushik Sen** of CRD's Future Technologies Group, were awarded the prestigious Sloan Research Fellowship, given annually by the Alfred P. Sloan Foundation to scientists, mathematicians and economists who are at an early stage of their research careers. In 2010, Persson was only one of 38 researchers nationally who submitted a winning research proposal to the Air Force's Young Investigator Research Program. Sen who is also an assistant professor in the Department of Electrical Engineering and Computer Sciences at UC Berkeley, received a National Science Foundation Career Award in 2008 and the C. L. and Jane W.-S. Liu Award in 2004 for exceptional research promise.

James Sethian Named James H. Simon Chair in Mathematics

James Sethian was named joint Hewlett-Packard and James H. Simon Chair in Mathematics at UC Berkeley, 2011.

ICIAM Lagrange Prize

Alexandre Chorin won the 2011 Lagrange Prize from the International Council for Industrial and Applied Mathematics (ICIAM) for his ground breaking work in applied math. A member of the LBNL Mathematics Group and professor of Mathematics at UC Berkeley, Chorin was honored for "his fundamental and original contributions to applied mathematics, fluid mechanics, statistical mechanics, and turbulence modeling." The Lagrange Prize provides international recognition to mathematicians who have made an exceptional contribution to applied mathematics throughout their career.

ICIAM Pioneer Prize

James Sethian, leader of CRD's Mathematics Group, won the 2011 Pioneer Prize from the International Council for Industrial and Applied Mathematics (ICIAM). Sethian, who is also a professor of mathematics at UC Berkeley, was honored "for his fundamental methods and algorithms that have had a large impact in imaging and shape recovery in medicine, geophysics and tomography, and drop dynamics in inkjets." The Pioneer Prize recognizes pioneering work introducing applied mathematical methods and scientific computing techniques to an industrial problem area or a new scientific field.

Technion's Israel Pollak Prize

James Sethian, Group Lead for the Mathematics Group in LBNL's CRD, was awarded the Israel Pollak Distinguished Lectures Series 2011. His lectures included topics on "Advances in Advancing Interfaces."

2010

American Physical Society Fellow

Niels Gronbech-Jensen, a faculty scientist in CRD's Scientific Computing Group and a professor in the Applied Science Department at UC Davis, was named a fellow of the American Physical Society, nominated by the Computational Physics Division. He was cited "for his development and application of new computational algorithms and tools in Biological and Condensed Matter Physics, especially those involving massively parallel molecular dynamics, electrostatic interactions, ion implantation, and nonlinear physics."

ACM Distinguished Scientist

Kesheng John Wu of the Scientific Data Management Research Group was one of 41 ACM (the Association for Computing Machinery) members named as 2010 Distinguished Scientists. The Distinguished Scientist award recognizes ACM members who have at least 15 years of professional experience and who have achieved significant accomplishments or have made a significant impact on the computing field.

IEEE Sidney Fernbach Award

James Demmel, a member of the Future Technologies Group and a professor of Computer Science and Mathematics at UC Berkeley was the recipient of the 2010 IEEE Sidney Fernbach Award, which recognizes outstanding contributions in the application of high performance computing using innovative approaches. Demmel was cited "for computational science leadership for creating adaptive, innovative high performance linear algebra software."

US Air Force Young Investigator Research Program

Per-Olof Persson of CRD's Mathematics Group was one of 38 researchers who submitted winning research proposals through the Air Force's Young Investigator Research Program (YIP). Persson, who is also an assistant professor in mathematics at UC Berkeley, will work on efficient and robust high-order methods for fluid and solid mechanics.

NSF Faculty Early Career Development Award (CAREER)

Jon Wilkening of CRD's Mathematics Group received a five-year NSF Faculty Early Career Development Award (CAREER) to conduct research in optimization and continuation methods in fluid mechanics. Wilkening is also an assistant professor in mathematics at UC Berkeley.

Sir James Lighthill Distinguished Lectureship Award

The Sir James Lighthill Distinguished Lectureship Award, from the University of Florida, was given to the Mathematics Group's **Alexandre Chorin**, to honor his work in mathematical sciences and to influence the future generation of mathematicians.

NASA Group Public Service Achievement Award

Julian Borrill, Christopher Cantalupo and Theodore Kisner of the Computational Cosmology Center (C^3) were honored with a NASA Public Service Group Award for developing the supercomputing infrastructure for the US Planck Team's data and analysis operations at DOE's National Energy Research Scientific Computing Center (NERSC).

SIAM Junior Scientist Prize

Kamesh Madduri, an Alvarez Fellow working in the Scientific Data Management Research Group, was selected as the first winner of the Junior Scientist Prize established by the SIAM Activity Group on Supercomputing (SIAG/SC).

Australia-America Fellowship

Robert Saye, member of the Mathematics Group and post-graduate student of applied mathematics at UC Berkeley, was a recipient of the Australia-America Fellowship. His research focuses on developing novel mathematical and computational algorithms for the study of inter-connected interfaces.

DOE Computational Science Graduate Fellowship 2010

Jeff Donatelli was a recipient of DOE's Computational Science Graduate Fellowship to pursue his research in computational algorithms for imaging technology.

2009

Presidential Early Career Award for Scientists and Engineers (PECASE)

Cecilia Aragon, a staff scientist in CRD, was honored with a 2009 Presidential Early Career Award for Scientists and Engineers (PECASE), the US government's most prestigious award for early-career scientists.

AAAS Fellow

Juan C. Meza, Head of the High Performance Computing Research Department in the CRD, was elected a Fellow of the American Association for the Advancement of Science "for exemplary service to the federal energy laboratories and professional societies in enhancing research and research participation."

SIAM Fellows

John Bell, Alexandre Chorin, Phillip Colella, James Demmel and James Sethian of CRD were among the first group of Fellows announced May 1, 2009 by the Society for Industrial and Applied Mathematics (SIAM). The Fellows program was announced by SIAM in 2008 in part "to honor SIAM members who are recognized by their peers as distinguished for their contributions to the discipline."

ACM Gordon Bell Prize

A team of researchers from the IBM Almaden Research Center and **Horst Simon** of LBNL won the prestigious 2009 ACM Gordon Bell Prize in the special category for their development of innovative techniques that produce new levels of performance on a real application.

Hispanic Business Magazine's 2009 Women of Vision

Cecilia Aragon, a staff scientist in CRD, was honored by Hispanic Business magazine as one of 25 Women of Vision in 2009. Aragon is a founding member of Latinas in Computing and active in diversity and outreach programs at the Lab.

Hispanic Business Magazine's "100 Influentials"

Juan Meza, head of the High Performance Computing Research Department in LBNL's CRD, was named to Hispanic Business magazine's annual list of 100 influential Hispanics.

NSF Graduate Fellowship Recipient

Ben Preskill of the Mathematics Group received an NSF Graduate Fellowship in applied mathematics.

Honorary Fellow for the International Congress on Fracture

G.I. Barenblatt of the Mathematics Group and of the UC Berkeley Mathematics Department is now an Honorary Fellow of the International Congress on Fracture, which is concerned with the structural properties of materials and with advancing research for the development of new materials.

CONFERENCE HONORS AND RECOGNITION

2012

Invited Speaker

Ann Almgren, "Low Mach Number Models in Computational Astrophysics," SC12 Conference on High Performance Computing, Networking, Storage and Analysis, Salt Lake City, Utah, November 2012.

Keynote Speaker

Kathy Yelick, "Magellan: A Study in Cloud Computing for Science," 24th Systems and Software Technology Conference, Salt Lake City, Utah, April 2012.

Invited Speaker

Kathy Yelick, "More and Moore: Growing Computing Performance for Scientific Discovery," Symposium on 20th anniversary of the Federal Networking and Information Technology Research and Development (NITRD) Program, Washington, D.C., February 2012.

Best Paper in History of HPDC Conference

John Shalf (with Gabrielle Allen, Werner Benger, Tom Goodale, Hans-Christian Hege, Gerd Lanfermann, André Merzky, Thomas Radke and Edward Seidel), "The Cactus Code: A Problem Solving Environment for the Grid," selected as one of the top papers in the 20 years of publications from HPDC, the International ACM Symposium on High-Performance Parallel and Distributed Computing.

Plenary Invited Speaker

Maciej Haranczyk, "In Silico Design of Carbon Capture Materials," Commonwealth Scientific and Industrial Research Organization (CSIRO) Intelligent Processing Symposium and Workshop (CIPS 2012), Melbourne, Australia, May 2012.

Invited Plenary Addresses

James Sethian, "The Voronoi Implicit Interface Method", the Fourteenth International Conference on Hyperbolic Problems: Theory, Numerics, Applications, HYP2012.

Best Paper Award

P. Navratil, D. Fussell, C. Lin, and H. Childs, "Dynamic Scheduling for Large-Scale Distributed-Memory Ray Tracing," in Proceedings of EuroGraphics Symposium, on Parallel Graphics and Visualization, pages 6-70, May 2012.

Best Poster Award

Chris Rycroft, Robert Saye, Daniela Ushizima, James Sethian, "Mechanical simulation of cell and microenvironment interactions," Bay Area's Physical Sciences-Oncology Center Annual Meeting, 2012.

Invited Speaker

Gunther Weber, "Parallel Extraction of Crack-free Isosurfaces from Adaptive Mesh Refinement Data," ASTRONUM 2012: the 7th International Conference on Numerical Modeling of Space Plasma Flows, Big Island, HI, USA., June 2012.

2011

Best Paper Award

Devarshi Ghoshal, Indiana University; and Shane Canon and **Lavanya Ramakrishnan**, LBNL/NERSC, "I/O Performance of Virtualized Cloud Environments," The Second International Workshop on Data Intensive Computing in the Clouds (DataCloud-SC11), Seattle, November 2011.

Best Paper Award

Khaled Ibrahim, Steven Hofmeyr and Costin Iancu, "Characterizing the Performance of Parallel Applications on Multi-Socket Virtual Machines," CCGRID'11, the IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing 2011.

Best Poster Award

Chris Rycroft, Robert Saye, Kandice Tanner, Mina Bissell, Daniela Ushizima, James Sethian, "A Mechanics-Based Physical Computational Framework for Cell Modeling: Epithelial Cell Differentiation," Bay Area's Physical Sciences-Oncology Center Annual Meeting, 2011.

Invited Speaker

Gunther Weber, "Parallel Extraction of Crack-free Isosurfaces from Adaptive Mesh Refinement Data," Dagstuhl Scientific Visualization Seminar, Schloss Dagstuhl (Leibniz-Zentrum fuer Informatik), Wadern, Germany, June 2011.

SciDAC People's Choice Award

E. Cormier-Michel, D. L. Bruhwiler, M. Durant, D. Kindig, C. G. R. Geddes, M. Chen, O. Rübel, V. H. Ranjbar, B. Cowan, and J. R. Cary, won an award for "Colliding Laser Pulses Launch an Electron Beam into a Plasma Accelerator" at SciDAC's Visualization Night.

2010

Best Paper Award

Keith R. Jackson, Lavanya Ramakrishnan, Krishna Muriki, Shane Canon, Shreyas Cholia, **John Shalf**, Harvey J. Wasserman, and Nicholas J. Wright, "Performance Analysis of High Performance Computing Applications on the Amazon Web Services Cloud," IEEE International Conference on Cloud Computing Technology and Science (CloudCom 2010), Bloomington, Ind., Nov. 30-Dec.1, 2010.

Invited Speaker (Masterworks Series)

Phil Colella, "High-End Computing and Climate Modeling: Future Trends and Prospects," SC 10, the ACM-IEEE conference on high performance computing, networking, storage and analysis, New Orleans, La., November 2010.

Keynote Speaker

Katherine Yelick, "Science in the Clouds: The View from Berkeley," ISC Cloud '10 conference, Frankfurt, Germany, October 2010.

Keynote Speaker

Katherine Yelick, "Paving the Road to Exascale Computing," 39th International Conference on Parallel Processing (ICPP 2010), San Diego, Calif., September 2010.

Best Paper Award

Keith Jackson, Lavanya Ramakrishnan, Karl Runge (Physics) and **Rollin Thomas**, "Seeking Supernovae in the Clouds: A Performance Study," ScienceCloud 2010, the 1st Workshop on Scientific Cloud Computing, Chicago, Ill., June 2010.

Best Paper Award

Mark Howison, E. Wes Bethel and Hank Childs, "MPI-hybrid Parallelism for Volume Rendering on Large, Multi-core Systems," Eurographics Symposium on Parallel Graphics and Visualization (EGPGV'10), Norrköping, Sweden, May 2010.

Invited Speaker

Gunther Weber, "Topology-based Feature Definition and Analysis," ASTRONUM 2010: the 5th International Conference on Numerical Modeling of Space Plasma Flows, San Diego, California, June 2010.

Invited Speaker

James A. Sethian, "Advances in Advancing Interfaces," Richard von Mises Lecture, 200th Anniversary Celebration of Humboldt University, Berlin, 2010.

Best Paper Award

Mark Howison, E. Wes Bethel, and Hank Childs, "MPI-hybrid Parallelism for Volume Rendering on Large, Multi-core Systems," in Eurographics Symposium on Parallel Graphics and Visualization (EGPGV), Norrköping, Sweden, May 2010. LBNL-3297E.

SciDAC People's Choice Award 2010

Prabhat of the Visualization Group was winner of a People's Choice Award at the SciDAC Visualization Night, 2010.

SciDAC People's Choice Award 2010

The Visualization Group's **Hank Childs**, with other members of the group, won a People's Choice Award at the SciDAC Visualization Night for the visualization of "Type Ia Supernova: Turbulent Combustion on the Grandest Scale."

2009

SciDAC People's Choice Award 2009

The Visualization Group's **Hank Childs**, with other members of the group, won a People's Choice Award at the SciDAC Visualization Night for the visualization of "Simulation of the Turbulent Flow of Coolant in an Advanced Recycling Nuclear Reactor."

Best Paper Award

Shoaib Kamil, Cy Chan, Samuel Williams, Lenny Oliker, John Shalf, Mark Howison, E. Wes Bethel, Prabhat, "A Generalized Framework for Auto-tuning Stencil Computations," Cray User Group Conference, Atlanta, GA, May 2009.

Keynote Speaker

Leonid Oliker, "Green Flash: Designing an Energy Efficient Climate Supercomputer" IEEE International Parallel and Distributed Processing Symposium (IPDPS'09), Rome, Italy, May 2009.

Most Entertaining Talk

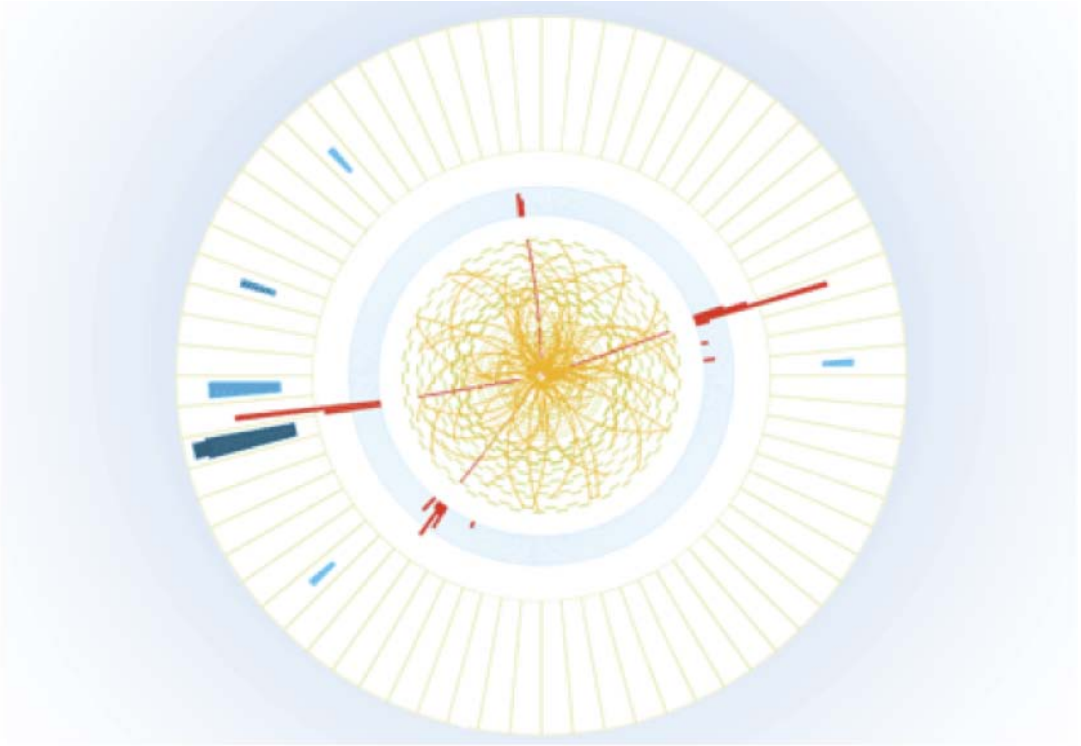
David Bailey and Jonathan Borwein, "High-precision, high-performance numerical integration," 3rd Workshop on High-Dimensional Approximation, Sydney, Australia, February 2009.

Invited Speaker

Gunther Weber, "The Contour Spectrum Revisited," Dagstuhl Scientific Visualization Seminar, Schloss Dagstuhl (Leibniz-Zentrum fuer Informatik), Wadern, Germany, June 2009.

Invited Speaker

Gunther Weber, "Recent Advances in VisIt: Streamlines and Query-Driven Visualization," ASTRONUM 2009: the 4th International Conference on Numerical Modeling of Space Plasma Flows, Chamonix, France, June/July 2009.



ADVANCED COMPUTING FOR SCIENCE DEPARTMENT

The Advanced Computing for Science (ACS) Department delivers innovative end-to-end computational and data analysis capabilities. ACS researchers are involved in basic research and applied research in three primary mission areas:

» Large and complex datasets: from basic science through application,

» Ground-breaking research in enabling data understanding,

» Custom interfaces enabling science.

ACS is broken up into three groups that work collaboratively, including: the CyberInfrastructure Development Group, the Sciences Software Group, and the Data Intensive Systems Group.

OVERVIEW OF HIGHLIGHTED PROJECTS

The Advanced Computing for Sciences Department works collaboratively with many groups and researchers both inside and outside of LBNL, and with DOE's supercomputing facility, the National Energy Research Scientific Computing Center (NERSC). Some of their collaborative work within CRD includes Advanced Simulation Capabilities for Environmental Management (ASCEM) with the Scientific Data Management Group, and the Materials Project with the Scientific Computing Group. They also participate in collaborative projects outside of CRD, like the DOE-funded Carbon Capture Simulation Initiative.

The highlighted projects for the ACS include: their development of an end-to-end solution for light source analysis, focusing on ALS users; Daya Bay offline computing; ATLAS core software; the Carbon Capture Simulation Initiative Toolkit; software technology in the service of science, and eco-science data management. For more collaborative work done by ACS, see the Materials Discovery project and the Advanced Simulation Capabilities for Environmental Management (ASCEM) project under "Collaborative Projects."

CYBERINFRASTRUCTURE GROUP RESEARCH STAFF

DEB AGARWAL is the ACS Department Head and the CyberInfrastructure Development Group Lead. Agarwal's research focuses on scientific tools that enable the sharing of scientific experiments, advanced networking infrastructure to support sharing of scientific data, scientific computation frameworks, data analysis support infrastructure for eco-science, and cybersecurity infrastructure to secure collaborative environments. Some of the projects Agarwal is working on include: AmeriFlux data processing and management, advanced computational subsurface modeling data management, and infrastructure for carbon capture simulations. As a member of the Berkeley Water Center collaboration between UC Berkeley and LBNL, Agarwal also leads a team that develops data server infrastructure to significantly enhance data browsing and analysis capabilities, and enables eco-science synthesis at the watershed-scale to understand hydrologic and conservation questions and at the global-scale to understand carbon flux. Agarwal received her Ph.D. in electrical and computer engineering from UC Santa Barbara and a B.S. in mechanical engineering from Purdue University.

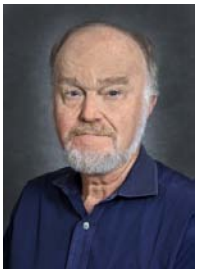
KEITH BEATTIE is a software engineering expert and data frameworks developer in the CyberInfrastructure Group. His interests are in distributed systems and effective software development practices and tools, specifically as applied in an interdisciplinary and multi-institutional scientific setting. Before beginning at LBNL two years ago, Beattie worked as a software engineer and release manager for a "dot-com." He currently co-leads the Software Development Support task element of the Carbon Capture Simulation Initiative (CCSI), and also deploys a 1 PB XRootD cluster on the Parallel Distributed Systems Facility (PDSF) at NERSC for the solenoidal tracker at Relativistic Heavy Ion Collider (STAR) experiment. Beattie was also a core member of the data acquisition development team for the IceCube neutrino detector installed at the South Pole and a co-lead developer on the NetLogger distributed systems monitoring pipeline. Beattie received an M.S. in computer science and a B.A. in mathematics from San Francisco State University.

VAL HENDRIX is a computational frameworks developer in the CyberInfrastructure Development Group. Hendrix has been working on **a t**oroidal **L**HC **a**pparatu**s** (ATLAS) experiment for the last year. She focuses on the deployment and configuration of ATLAS Tier-3 data analysis clusters (DAC) both on dedicated hardware and in the cloud. Out of the cloud, this work is an effort to reduce the maintenance burden of the scientist-turned system administrator by automating DAC configurations using the Puppet configuration management engine. In the cloud, Hendrix is working on an elastic DAC management solution that will allow a physicist to deploy a cluster in the cloud in minutes. Hendrix also has more than ten years of development experience with an interest in solving complex and large-scale computing problems.

KEITH R. JACKSON is a computer scientist in the CyberInfrastructure Development Group. His research focuses on developing tools and techniques to support distributed, large-scale, collaborative science. He is an expert in applying cloud computing to science. As a member of the Berkeley Water Center, he is developing tools to manage and transform satellite-based remote sensing data for eco-science. Some of his previous research includes the development of Python interfaces to the Globus Toolkit, distributed systems security, and work on scientific collaboration tools.

CRAIG LERES is a hardware, operating systems and networking expert. He is particularly familiar with FreeBSD internals including writing, modifying, and debugging device drivers. He is responsible for a number of critical systems at LBNL including DHCP, DNS and NTP. He is also a member of LBNL's Computer Protection Program and LBLnet, the network group.

SARAH POON is a user interface designer, web application developer, and Human Computer Interface (HCI) researcher in the CyberInfrastructure Development Group. As a practitioner, her focus is based on designing and developing web applications that take advantage of new technologies and practices of web interaction. Her research focus is in computer-supported cooperative work (CSCW). In particular, Poon is interested in understanding the challenges and the motivations for collaboration in distributed scientific teams and how technology can mediate collaboration.



CHUCK MCPARLAND is a computer scientist in the Advanced Computing for Science Department. He is responsible for the design and implementation of data acquisition systems for numerous accelerator-based HEP experiments at LBNL, CERN and the Relativistic Heavy Ion Collider (RHIC). Recently, he headed the data acquisition design team for the IceCube neutrino telescope laboratory at the South Pole Station, Antarctica. His contribution to the HEP field continues with his participation in the Particle Data Handbook upgrade project. He is also working with LBNL's Demand Response Research Center to standardize (as part of DOE's Smart Grid Initiative) the OpenADR utility protocol and to promote its use within California's power utility infrastructure. Lastly, he contributes to an LBNL-funded effort to define an energy efficient architecture suitable for a "purpose built" climate modeling supercomputer, called the "Green Flash" project.

DAVID ROBERTSON is a software engineer working on the ESnet OSCARS circuit reservation project. His primary interest is in web service development, including user interfaces. He has worked both at the Lawrence Livermore National Lab and in business positions. He received his M.S. in computer science from San Francisco State University in 1988.

CECILIA R. ARAGON has authored and co-authored over 30 peer reviewed publications in the areas of computer-supported cooperative work, human computer interaction, visualization, visual analytics, image processing, and machine learning. Aragon has an interdisciplinary background, including over 15 years of software development experience in industry and NASA, and a three year stint as the founder and CEO of a small company. She earned her Ph.D. in computer science from UC Berkeley in 2004, and she earned her B.S. in mathematics from the California Institute of Technology. She has received many awards for her research, including a Presidential Early Career Award for Scientists and Engineers (PECASE). She was also named one of the Top 25 Women of 2009 by Hispanic Business Magazine.

CRISTINE MORIN is a senior researcher in distributed computing. Morin is principal investigator for the EU Contrails cloud computing project and has been visiting and collaborating with ACS for the past year on sabbatical. Morin is the leader of the Myriads team at INRIA, Rennes. She is also an expert in cloud computing and other distributed computing topics.

CATHARINE VAN INGEN is an expert in data management and synthesis for eco-informatics. She has been a leader in the development of the FLUXNET Carbon Flux, the International Soil Carbon Network, and the California Watersheds data server architecture and data processing.

JOHN SHINN is an experienced senior engineer from the industrial energy sector serving as the Carbon Capture Simulation Initiative (CCSI) Industry Advisory Board (IAB) leader. Shinn organizes IAB meetings and provides guidance to the CCSI project team. Shinn recently retired from Chevron Corporation. He is also active in Engineers Without Borders and is a consultant for the World Bank.

ARTHUR WEIDMER is a graduate student at UC Berkeley. He is working on his Ph.D. with a focus on subsurface waste remediation. He is providing data expertise to the Advanced Simulation Capability for Environmental Management (ASCEM).

SCIENCE SOFTWARE SYSTEMS GROUP RESEARCH STAFF

CRAIG E. TULL is the Group Lead of the Science Software Systems Group. He has been developing scientific software and managing software projects for over 25 years. His interests are in component frameworks, generative programming, and using scripting languages to enhance the power and flexibility of scientific data exploration. Tull has worked on science frameworks in several experiments, including framework architect in the STAR Experiment, and as leader of the LBNL framework effort in ATLAS. He has worked on the PPDG (Particle Physics Data Grid) and the GUPFS (Global Unified Parallel File System) projects that aim to deliver innovative solutions to data-intensive computing in a distributed environment. He recently ended a three-year assignment in DOE headquarters as program manager for Computational High Energy Physics including HEP's SciDAC portfolio, and is currently the US manager of Software and Computing for the Daya Bay Neutrino Experiment in China. He has a Ph.D. in physics from UC Davis.

PAOLO CALAFIURA is an expert software engineering team lead and software developer. He works as a developer of large, high energy and nuclear physics (HENP) applications. Currently, he is chief software architect of the ATLAS Experiment at CERN. Calafiura coordinates the work of a "core software" team of physicists and engineers who have developed the libraries and tools that ATLAS physicists use to develop and run their physics applications, from trigger selection to event reconstruction, simulation, and analysis. The main components of the core software are the Athena framework and the data model foundation classes. Calafiura was invited to give a Plenary Session talk at the International Conference on Computing High Energy and Nuclear Physics 2009. Until 2003, Calafiura was a collaborator of CDF where he worked on code generation and handling, and on the event data model. Before moving to LBNL in 1998, he was a postdoc at CERN and at the Scuola Normale Superiore in Pisa where he also graduated in physics.

GARY KUSHNER is a computational pipeline and database expert. He is currently leading the development of the Advanced Light Source User Portal. He previously worked with the Joint Dark Energy Mission (JDEM) and Sloan Digital Sky Survey-III's (SDDS3) Baryon Oscillation Spectroscopic Survey (BOSS) astrophysics programs. On JDEM, Kushner was involved with the simulation, ground data systems and data reduction efforts. With BOSS, he was involved with the spectrographic analysis pipeline. In industry and academia, Kushner has been involved in many projects in the areas of physics, bioinformatics, databases, and wireless and hardware drivers. At various times, he has led both small and large engineering groups. Some of his industrial highlights include being a Senior System Architect at Oracle and Vice President of Engineering, CTO and acting CEO at Neomar. He has a B.A. in physics from UC Berkeley, and his current interests are in the areas of data reduction, databases and full-life cycle software development.

WIM LAVRIJSEN is a high-performance python development expert currently involved in the ATLAS experiment at CERN. He is the author of the PyROOT dynamic python bindings for the ROOT analysis software package, and he has developed several nimble performance monitoring tools that work under the heavy constraints of the ATLAS offline software. Lavrijsen's interests are in python parallelization for multi-core machines, optimization, and compilation.

IGOR GAPONENKO is an affiliate scientist in the Software Systems Group. His emphasis is in software and database solutions for large experiment data analysis systems.

CHARLES LEGGETT is a software framework developer with a background in physics. He is currently involved in developing the core software for the ATLAS analysis framework. Leggett is the Gaudi ATLAS coordinator managing, building and maintaining the Gaudi framework for ATLAS. He is also interested in studying the performance of the GNU C++ compiler, and he maintains a site tracking the results. Leggett received his Ph.D. in high energy physics from the University of Michigan, Ann Arbor in 1998, where he studied beta decays on the L3 experiment at CERN.

SIMON PATTON is a leader in the development of data transfer software and large-scale software frameworks. He is currently in charge of the core software infrastructure for the joint US and China Daya Bay Experiment. Over the last 20 years, he has had similar positions in the high energy physics and nuclear physics fields, ranging from database architect for Babar's object oriented database, to software coordinator for the CLEO experiment. Before joining the Daya Bay Experiment, Patton was the software architect for the IceCube Experiment based at the South Pole.

ROBERTO AGOSTINO VITILLO is a computational performance expert in the Software Systems Group. His research emphases are science cyberinfrastructure, performance monitoring and debugging tools.

DATA INTENSIVE SYSTEMS GROUP RESEARCH STAFF

DAN GUNTER is the Group Lead for the Data Intensive Systems Group. His research focus is on middleware, which is at the intersection of applications, databases, networks, and operating systems for distributed and collaborative computing. Thus, his research includes distributed workflows, network performance, visual analysis techniques, API and interface design, and complex data integration. He maintains the NetLogger application monitoring software. Gunter is currently a principal investigator on the Tigres workflow collaboratories project, the Materials Project Center for Functional Electronic Materials, and the Systems Biology Knowledgebase projects. Gunter received an M.S. in computer science from San Francisco State University.

JOSH BOVERHOF is a software frameworks developer with emphases on the development of science gateways, Python-based cyberinfrastructure, and science cyberinfrastructure. He received a B.S. in computer science from San Francisco State University. Boverhof has worked extensively on grid and web technologies. Currently, he leads the development of the Carbon Capture Simulation Initiative (CCSI) Turbine Science Gateway, which manages the execution of multiple simulations running on the cloud, Amazon EC2.

ABDELILAH ESSIARI is a software developer with an emphasis on science cyberinfrastructure development, workflow tools, and extensive expertise in the development of network security protocols. He received his M.S. in computer science from San Francisco State University. Essiari has previously committed to the Knowledge Discovery and Dissemination (KDD) BLACKBOOK project for semantic data-mining. He is currently performing software engineering and release engineering for the Carbon Capture Simulation Initiative (CCSI) Integration framework software and is working on the Advanced Light Source data analysis framework.

MONTE GOODE is a software engineer and database expert currently working on cyberinfrastructure for the sciences. Most recently, he has been focusing on data integration, storage and analysis for network monitoring and workflow systems, with a focus on current non-relational database tools. Previously, he participated in a collaboration with the Berkeley Water Center at UC Berkeley with an emphasis on environmental informatics and helped to build the Fluxdata.org International Soil Carbon Network datastore. Previous projects include contributions to the National Fusion Grid and the MyProxy Credential Management Service. He received an MLIS in library and information science from the University of Texas at Austin in 1996. Goode's fifteen plus year career includes work in both academic and commercial environments.

LAVANYA RAMAKRISHNAN is an expert in workflow tools. Her research interests include software tools for computational and data intensive science. Currently, she is actively engaged in projects that explore programming templates for data analysis workflow composition and data management techniques for elastic transient environments. She is also involved in projects evaluating cloud and big data technologies, such as Hadoop and NoSQL databases for scientific applications and performance, and energy and fault-tolerance trade-offs in infrastructure strategies. Ramakrishnan has worked with scientists from multiple domains including bioinformatics, biomedical science, storm-surge modeling, weather modeling, high energy physics and light source facilities. She previously worked as a research staff member at Renaissance Computing Institute and MCNC in North Carolina.

She has her master's and doctoral degrees from Indiana University and a bachelor's degree in computer engineering from VJTI, University of Mumbai. She joined LBNL as an Alvarez Postdoctoral Fellow in 2009.

TAGHRID SAMAK has an emphasis on "data science" at the intersection of statistical algorithms and computing, in the context of important scientific problems. She has developed workflow failure prediction algorithms, fingerprinting HPC codes, and two projects with the LBNL Joint Genome Institute: predicting protein function and genome assembly errors. Samak is investigating the applicability of machine learning algorithms in multiple science domains. She joined LBNL as a postdoc researcher in 2010. Her Ph.D. work was developing novel algorithms for distributed network policy configurations.

VERN PAXSON is a recognized leader in network intrusion detection. He is an associate professor on the faculty of the Electrical Engineering and Computer Sciences Department at UC Berkeley. He holds joint appointments at LBNL, where he has been a staff member since 1985, and a senior scientist since 1999 at the International Computer Science Institute in Berkeley. Paxson's research focuses on internet attacks and defenses, and network and defenses, which pursues a variety of research efforts in detecting and blocking network-borne attacks. He co-founded and served on the steering committees of the ACM Internet Measurement Conference, the ACM Workshop in Rapid Malcode, and the USENIX Workshop on Large-Scale Exploits and Emergent Threats. He has served as program chair or co-chair of ACM SIGCOMM, USENIX Security, IEEE Symposium on Security and Privacy, and ACM HotNets, and as an area director of the Internet Engineering Task Force. He is an ACM fellow and recipient of the 2008 ACM Grace Murray Hopper Award for his work on internet measurement. He has twice been a co-recipient of the IEEE Communications Society William R. Bennett Prize Paper award, and of the USENIX Lifetime Achievement award.

ROBIN SOMMER is an expert in network security monitoring. He is a staff researcher at the International Computer Science Institute in Berkeley, and he is also a member of ACS. His research focuses on network security monitoring in operational high performance settings, and he is one of the core developers of the open source Bro network intrusion detection system. Sommer received a Ph.D. from the Technical University of Munich in Germany, and postdoc fellowships from the German Academic Exchange Service, as well as from the Department of Homeland Security-funded Institute for Infrastructure Protection. He has participated in program committees for conferences and workshops, and he chaired the 2007 "Conference on the Detection of Intrusions & Malware, and Vulnerability Assessment."

CHRISTIAN KREIBICH received his diploma in computer science in 2002 from Technische Universitat Munchen, Germany, and he completed the Ph.D. program at the University of Cambridge Computer Laboratory. He began working for the Institute for Communication and Information Research (ICIR) in 2006, and is currently a staff research scientist at LBNL. His research focuses on topics in network architecture, distributed systems, and network security.

JOHN L. MCCARTHY graduated from Stanford University in 1964 and received his Ph.D. in history from Yale University in 1970, where he taught American history and quantitative methods for historical research for six years. In 1974, he moved to the Survey Research Center (SRC) at UC Berkeley, where he directed data analysis and educational services and helped develop UNIX-based tools for computer-assisted survey research, as well as a public opinion survey information retrieval system for the Congressional Research Service. Since retirement, McCarthy has also been working on voting integrity and election auditing issues as a volunteer with the Verified Voting Foundation and VerifiedVoting.org. McCarthy's major interests have included information systems design, metadata management, user interfaces, and collaborative software.

FRANK OLKEN is a researcher in databases and metadata registries. Prior to his recent retirement he served as a program director at National Science Foundation in the CISE Directorate. He was involved in standards development for the ISO/IEC 11179 metadata registry standard and development of the XMDR prototype metadata registry. His other standards work included the W3C RDF language standard and the W3C XML Schema Language standard. He has also worked on random sampling from databases. Olken received his Ph.D. from UC Berkeley in computer science.

TOWARDS AN END-TO-END SOLUTION FOR LIGHT SOURCE DATA

Research Highlights

Recent improvements in detector resolution and speed, and in source luminosity are yielding unprecedented data rates at the Basic Energy Sciences' (BES) national light source and neutron source facilities. In the past 2 years, Advanced Light Source (ALS) scientists have seen their data volumes grow from 65TB/year to 312TB/year, and in another 2 years the ALS will be generating 1.9PB/year. These data rates exceed the capabilities of data analysis approaches and computing resources utilized in the past, and will continue to outpace Moore's law scaling for the foreseeable future. The growing consensus within light source scientific communities is that scientific insight and discovery at BES facilities are now being limited by computational and computing capabilities much more than by detector or accelerator technology.

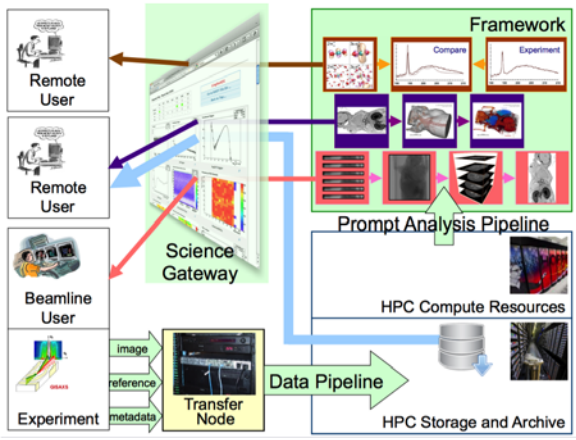


Figure 1: Advanced Light Source pipeline.

ACS researchers and their collaborators have proposed a systematic investigation and design of the light source analysis environment that can provide an end-to-end solution for data access, management, and analysis. This proposal will seamlessly be integrated with simulation codes and it will present easy-to-use web interfaces with single sign-ons tied to ALS users' unique identity. The result of the Laboratory Directed Research and Development (LDRD) will be a functional prototype end-to-end solution for current ALS data, and it will be designed to scale to the next-generation light source (NGLS) data rates. This prototype will enable research and development in the critical areas of data intensive computing that need to be addressed to enable large-scale photon science at the ALS, other BES national facilities, and eventually the NGLS.

ACS is also developing software to help scientists optimize their light source data. Many types of data cannot be directly "inverted" to yield information about the atomic arrangements or electronic states of the sample. Instead, users have had to use crude "physics-free" methods based on the experimental spectra of standards, or they have had to collaborate with theorists who perform *custom* simulations, which are matched to data. To optimize their beam time and to maximize their science output, these users need on-demand, near-real-time simulations, currently only available to a few. ACS personnel strive to deliver this service to all ALS users.

Scientific Impact

At the end of this LDRD, ACS and their collaborators will, through the prototype, have significantly improved the data analysis environment for the involved ALS beamlines. Photon scientists at these beamlines will be able to see, in near-real-time, the results of analysis on NERSC resources while running their experiments. The same users will be able to browse and download select data to their home institution, or run full analysis on NERSC resources afterwards. ACS will have identified what is needed to enable other ALS beamline analysis environments; making the prototype production-ready. Extending this effort to other ALS beamlines and to NGLS-scale analyses will require long-term funding from BES and ASCR DOE offices for the core components of the system.

CRD Contact

Craig Tull, Dan Gunter, Xiaoye Sherry Li, Simon Patton, Lavanya Ramakrishnan

ALS Collaborators

Elaine Chan, Alexander Hexemer, Stefano Marchesini, Matthew Marcus, Dilworth Parkinson

NERSC Collaborators

Richard Shane Canon, Jack Deslippe, Filipe Maia, David Skinner

ESnet Collaborators

Eli Dart, Inder Monga, Eric Pouyoul, Brian Tierney

Materials Sciences Division Collaborator

David Prendergast

DAYA BAY OFFLINE COMPUTING

Research Highlights

The Daya Bay Reactor Neutrino Experiment, a multi-national collaboration operating in the south of China, reported in March 2012 the first results of its search for the last, most elusive piece of a long-standing puzzle: how is it that neutrinos can appear to vanish as they travel? The answer requires a new understanding of fundamental physics and it may eventually solve the riddle of why the Universe contains more matter than antimatter.

For the first time, the neutrino experiment's copious data revealed a strong signal that the scientists were searching for, a so-called mixing angle, θ_{13} (pronounced theta one three), which the researchers measured with unmatched precision. θ_{13}, the last mixing angle to be precisely measured, expresses how electron neutrinos and their antineutrino counterparts mix and change into other flavors. The Daya Bay collaboration's first results indicate that $\sin^2(2\theta_{13})$ is equal to 0.092±0.017.

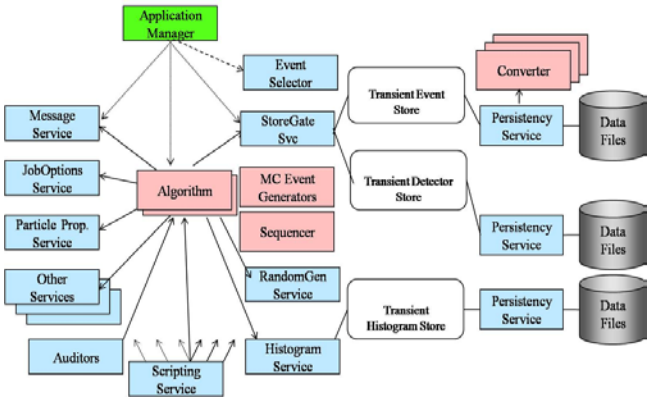


Figure 2: Daya Bay pipeline.

As US level 2 manager in charge of Daya Bay Offline Software and Computing, Craig Tull is responsible for the hardware, software, analysis, and simulation required to manage, archive, process, and analyze all Daya Bay data. Over the course of the experiment, Tull and his team have developed and refined simulation, data management, and analysis software to handle hundreds of terabytes of data in near-real-time. Their resultant software ecosystem includes:

» NuWa, which is a Gaudi-based simulation, processing, and analysis software framework used for all simulation and production processing for the experiment. NuWa extends the typical collider model for data analysis to accommodate the "window-based" analysis required by Daya Bay's detector system. NuWa uses ROOT for many analysis functions and GEANT4 for simulation.

» OfflineDB, which is a distributed, DBI-based (DataBase Interface) relational database system presenting all non-raw data information required to simulate and analyze experimental data. DBI permits the storage of parameters and conditions data in a time-varying schema that allows for the reproducible retrieval of any previous state of those parameters and data.

» Spade, which is an automated system for data migration, management, and archiving. Spade transfers data from the Daya Bay Nuclear Power Plant to Beijing and to LBNL for real-time analysis. Raw data are available at both LBNL and international high energy physics computing facilities within 30 minutes of a raw data file being closed.

» P-squared, which is an automated data pipeline for processing large numbers of data files integrated with Spade and Offline Data Monitor (ODM). P-squared is used for both the real-time and full production reprocessing of data and the running of simulations.

» ODM (Offline Data Monitor), which is used for real-time analysis and data QA monitoring for raw data. ODM is built on the NERSC Science Data Gateways and the NEWT toolkit. Spade-transfered raw data are automatically analyzed using NuWa, OfflineDB by P-squared, the results integrated with metadata from DAQ and DCS (Detector Control System) databases, and then presented in real-time (within two hours) to scientists at the experiment and around the globe.

The hardware provisioned for the experiment includes an on-site (at Daya Bay Nuclear Power Plant in China) computing facility, DayaNet (network connectivity to China's CSTNet), and PDSF compute and storage resources at NERSC. The constant 24/7 use of ESnet, CSTNet, and GLORIAD required extensive monitoring of the end-to-end system.

The multi-national, multi-institutional, multi-network nature of the transfers meant that the team had repeated outages and problems that needed to be diagnosed and solved. They had outages from typhoons and earthquakes, from bad hardware, firewalls, policy changes, and human error. In each instance, the team had the tools and information to communicate and resolve their problems.

Scientific Impact

After years of development, preparation, training, and research, the Daya Bay experiment has published the world's first unambiguous measurement of the neutrino oscillation mixing angle, θ_{13}. This scientific discovery was made possible by the data management and data analysis and simulation software and computing resources for which CRD is directly responsible. A central goal of US Offline Computing for the last 5 years has been "physics-ready software and computing on day 1." Daya Bay scientists were able to see anti-neutrinos with NuWa in the first filled detectors within 24 hours. Physicists, aided by the software, were able to see an anti-neutrino deficit in the far hall within days. And finally, they have been able to extract a high-quality θ_{13} value within only 75 days of the start of far-hall running. This is, of course, not solely due to Offline Computing efforts. However, Tull and his team's efforts were invaluable during data commissioning and physics analysis. They ensured that data were transferred to NERSC/PDSF, archived on NERSC/HPSS, and analyzed by ODM/Science-Gateway and were available to scientists within 2 hours. All the analysis was done using NuWa and the Science Software Systems Group's offline system. All data were transferred and analyzed using networking and computing systems designed and/or procured by US Offline management, and almost all US computing occurred on the NERSC/PDSF Tier-1 for Daya Bay.

CRD Contact

Craig Tull

CRD Collaborator

Simon Patton

ATLAS CORE SOFTWARE

Research Highlights

Athena is the object-oriented control framework used by **a** **t**oroidal **L**HC **a**pparatu**s** experiment, called ATLAS, at CERN, which is based on the Gaudi component architecture, a collaboration among several physics experiments. The framework provides physics application codes with an extensible series of core services including scheduling, configuration, data access, error handling, and message logging. The component architecture allows users to easily swap service implementations at runtime, an essential requirement for software that is run in thousands of different configurations.

The Athena framework belongs to the blackboard family. In this framework, data objects produced by knowledge modules are posted to StoreGate, an "in-memory database," from where other modules can access them and produce new data objects. This in-memory model greatly reduces the coupling between knowledge modules, which contain the algorithmic code for analysis and reconstruction, based on the fact that one knowledge module does not need to know which specific module can produce the information it needs nor does the module need to know which protocol it must use to obtain it.

This is also known as the interface explosion problem described in component software systems. Algorithmic code is known to be the least stable component of software systems and the blackboard approach has been very effective in reducing the impact of this instability. The trade-off of the data/knowledge object separation is the need for knowledge objects to identify data objects to be posted on or retrieved from the blackboard.

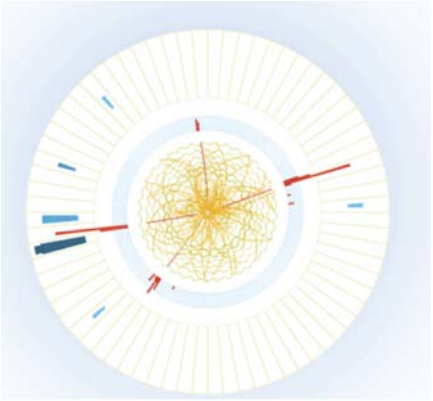


Figure 3: Particle collisions. It appears that the Higgs boson decays into two electrons and two positrons (shown in red). [Courtesy of L. Taylor and T. McCauley.]

Another key architectural choice in Athena is the transient/persistent data model separation: a data object transient representation, as a C++ object in memory, is kept independent from its persistent representation on disk (e.g., as a database entry, or a record in a binary file). StoreGate manages the conversion of a data object from/to its persistent form and provides an API to access data stored on persistent media. Keeping the transient data model independent from its representation on disk allows for the evolution of the two representations and their technologies independently, a key requirement for software that will run over several decades.

Scientific Impact

Athena is currently used by over 2,000 scientists worldwide to analyze multiple petabytes of collision data collected by the ATLAS experiment at the Large Hadron Collider. The framework manages about 4,000 C++ components totaling over 5 million LOC. Its flexibility allows it to be used in every aspect of ATLAS's complex workflow, from real-time data acquisition and selection, to Monte Carlo simulation and data analysis, such as the one leading to the discovery of Higgs boson-like particles. The CRD ATLAS team, which includes the ATLAS chief software architect, has lead the Athena development work for the last 10 years, and it is now extending its functionality to run efficiently time-critical ATLAS software on many-core architectures.

CRD Contact

Paolo Calafiura (ATLAS Chief Software Architect)

CRD Collaborators

Wim Lavrijsen, Charles Leggett, Roberto Vitillo

ATLAS Collaborators

Ian Hinchliffe (Physics), Vakhtang Tsulaia (Physics), Baptiste Wicht (HEFR), and the ATLAS Architecture Team

CARBON CAPTURE SIMULATION INITIATIVE

Research Highlights

The Carbon Capture Simulation Initiative (CCSI) develops and deploys state-of-the-art computational modeling and simulation tools to accelerate the commercialization of carbon capture technologies from discovery to development, demonstration, and ultimately the widespread deployment of these technologies to hundreds of power plants. The CCSI Toolset leadership and development involves personnel from across the Advanced Computing for Science (ACS) Department, which includes the CyberInfrastructure, Data Intensive Systems, and Science Software Systems groups. This toolset is a comprehensive, integrated suite of validated science-based computational models. CCSI provides these simulation tools to increase confidence in new carbon capture and storage designs, thereby reducing the risk associated with incorporating multiple innovative technologies into new carbon capture solutions.

In leading the Integration Framework and the Software Development Support, ACS researchers along with collaborators are developing the software framework needed to enable interoperability of the underlying carbon capture simulation tools. They are developing the Turbine Gateway, which enables and manages large numbers of simultaneous process simulations in support of uncertainty quantification, optimization, and reduced-order modeling efforts. In addition, ACS is responsible for the data management implementation that will link together all of the components and enable higher level services, such as risk assessment. The Energy Frontier Research Center is also helping to identify promising adsorber materials through advanced simulations.

The scientific underpinnings encoded into the suite of models will also ensure maximum learning for successive technology generations. These tasks implement the communication between the various tools, as well as data management, visualization, and software engineering. The CCSI Toolset is intended to accelerate the development and deployment cycle for bringing new carbon capture and storage (CCS) technologies to the market. (See Figure 4.)

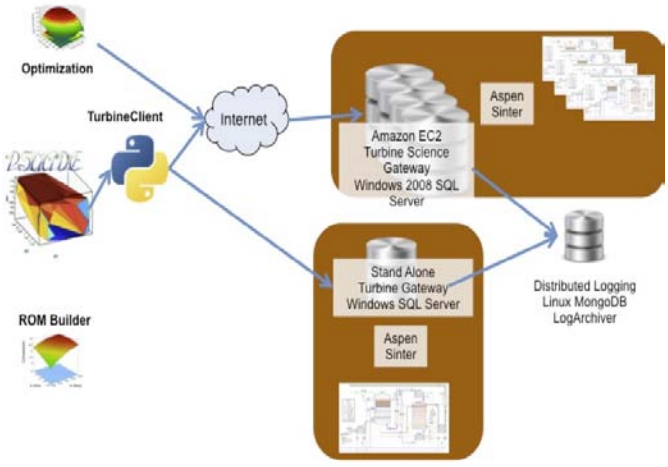


Figure 4: CCSI process optimization, Turbine Gateway and AspenSinter in operation.

Scientific Impact

Realizing the potential of CCSI depends on having an integrated carbon capture simulation environment that enables multi-scale simulations, uncertainty quantification, decision support, and optimization. ACS's Software Development Support in CCSI will provide comprehensive development support, and it will incorporate a comprehensive licensing, versioning, bug tracking, build, packaging, and test tools.

When completed, the CCSI Toolset will provide industry users with a comprehensive, integrated suite of scientifically validated models that enable the optimization of design, the assessment of uncertainty and the understanding of risks to support decision making capabilities. This toolset will also give users increased confidence in designs and reduce the risk associated with incorporating multiple innovative technologies into new carbon capture solutions.

Although carbon capture systems are not yet required at any of the 600 plus power plants across the US, CCSI works closely with industry to make carbon capture and storage tools easier to adapt to existing design technologies. CCSI's tools would make it easier for utility companies to meet such requirements if and when they are enacted, and could help companies doing business in countries where controls are already in place.

This past October, CCSI released their first set of tools and models at a meeting with representatives from the 19 companies that are currently participating on its industry advisory board.

CRD Contact

Deb Agarwal (CCSI Integration Framework Task Lead), Paolo Calafiura (CCSI Software Development Support Task Lead)

CRD Collaborators

Keith Beattie, Joshua Boverhof, Abdelilah Essiari, Daniel Gunter, Douglas Olson, Sarah Poon, Maciej Haranczyk, David Robertson

CCSI Institution and Industry Collaborators

National Energy Technology Laboratory (Overall Project Leadership), Lawrence Livermore National Laboratory, Los Alamos National Laboratory, Pacific Northwest National Laboratory, Carnegie Mellon University, and University of California (Berkeley). The CCSI's industrial partners provide representation from the power generation industry and the power equipment manufacturers. The initial industrial partners are ADA Environmental Solutions (ADA-ES), Alstom Power, Ameren, Babcock Power, Babcock & Wilcox, Chevron, the Electric Power Research Institute (EPRI), Eastman, Fluor, General Electric, Ramgen Power Systems, and Southern Company

ECO-SCIENCE DATA MANAGEMENT

Research Highlights

In eco-science, large-scale synthesis studies and simulations are becoming common. These studies and simulations are often coordinated by science teams, geographically distributed around the world, working on global-scale datasets. Over the last seven years, the Applied Computing for Science Department's groups evolved an eco-informatics data management infrastructure that combines data processing, archiving, curation, user services, and publication tracking. Variants of this data server infrastructure are now serving a wide array of carbon data and subsurface flow projects, such as ISCN, AmeriFlux and FLUXNET, and digital hydrology projects.

The principal goals of the International Soil Carbon Network (ISCN) are to produce databases, models, and maps that enable users to understand: a) how much carbon is stored in soils around the world, b) how long this carbon remains in the soil, and c) the factors that make this soil carbon vulnerable to being lost (i.e., emitted to the atmosphere). ISCN compiles soil sampling, and archives and analyzes protocols. To help ISCN, ACS made it possible for ISCN to share scientific, analytical and logistical infrastructure, synthesize products beneficial to stakeholders and scientists, and develop a community-accessible database.

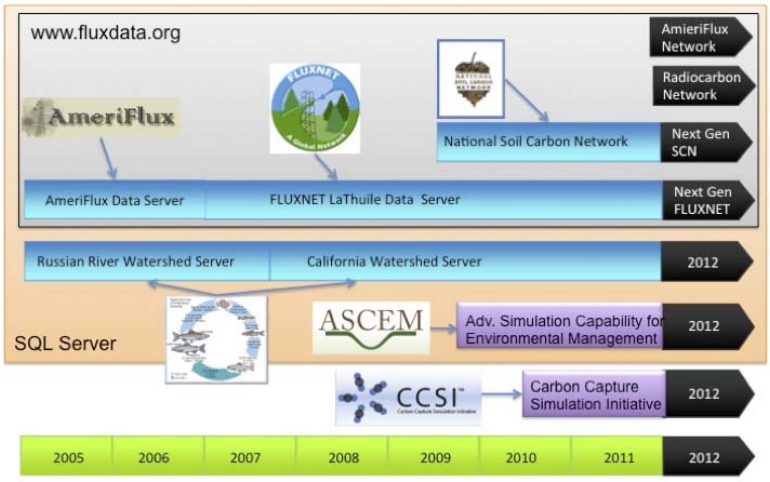


Figure 5: Evolution of the eco-science data server.

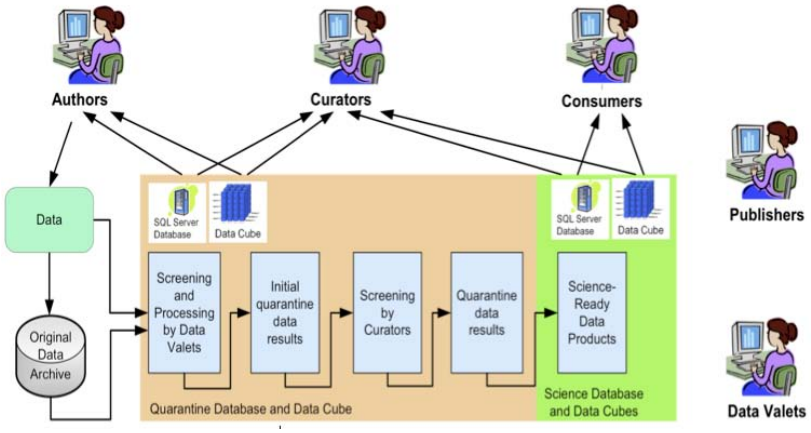


Figure 6: Data curation and publication support.

ACS researchers also developed a data server to support the collaborative analysis of carbon-climate data for FLUXNET. FLUXNET is a global network of over 400 carbon flux measurement sensor towers that provide a wealth of long-term carbon, water and energy flux data and metadata. The data from these towers are critical to understanding climate change through cross-site, regional, ecosystem, and global-scale analyses. The fluxdata.org data server now contains the FLUXNET data and is relied on by over 110 paper teams analyzing the data.

As part of the AmeriFlux Network Management Project, ACS helps leverage AmeriFlux's existing FLUXNET data management services and adds new capabilities to provide the network with data processing, archiving, user support, and publication tracking. AmeriFlux is a network of state-of-the-art flux sensors in a variety of ecosystems across North, Central and South America—from tundra to tropical forests, and everything in between—to study how carbon dioxide (CO_2), water vapor and energy cycles through the atmosphere, plants and soil. ACS's infrastructure helps Ameriflux researchers share data efficiently.

ACS also assists the scientists in hydrology, where the science has shifted to more sensor-based systems for stream flows and spatially distributed meteorological conditions obtained from ground and satellite systems. The sensor data are coupled with weather and climate models at finer and finer spatial resolutions. The net result is a vastly increased data collection rate and a generation of model simulations that require storage, retrieval, and archiving on a scale not previously contemplated. Because DOE recognizes that these globally-based datasets could benefit a variety of scientific communities beyond those interested in carbon flux, it funds the effort to make AmeriFlux data accessible to a wide-range of researchers.

Scientific Impact

For AmeriFlux and FLUXNET, ACS developed collaborative web portals where scientists can download data files, browse data summaries, update metadata and annotate the data through the portal. Synthesis teams can select sites and exchange e-mail with those sites through the portal. Users can also browse data directly from Excel spreadsheets or MatLab from the scientist's desktop; the scientist sees no difference between data "in the cloud" and on the desktop. The portal enables science researchers to concentrate on science rather than data management and the collaboration features enable continued growth and discovery for science.

Finally, the data server developed to assist hydrological studies is now used to support researchers working on a variety of questions including the impact of frost protection pumping, recovery of endangered fish populations, long term impact on a watershed of human activity, coastal lagoon dynamics, and modeling of annual watershed water balance.

CRD Contact

Deb Agarwal

CRD Collaborators

Arthur Wiedmer, Monte Goode, Valerie Hendrix, Keith Jackson, Lavanya Ramakrishnan

ESD Collaborators

Sebastien Biraud, Boris Faybishenko, Susan Hubbard, Marilyn Saarni, Margaret Torn

UCB Collaborators

Dennis Baldocci, James Hunt, Carolyn Remick, Youngryel Ryu, Rodrigo Vargas

Other Collaborators

Charlotte Ambrose (NMSF), Charleen Gavette (NMSF), William Winner (NMSF), Marty Humphrey (UVa), Norm Beekwilde (UVa), Catharine van Ingen (Microsoft Research), Dan Fay (Microsoft Research), Tony Hey (Microsoft Research), Chris Swanston (USFS), Kris Johnston (USFS), Luke Nave (USFS), Bev Law (OSU), Jen Harden (USGS), Marcus Reichstein (Max Planck Institute), Dario Papale (UTuscia, Italy)

SOFTWARE TECHNOLOGY DEVELOPMENT IN SERVICE OF SCIENCE

Research Highlights

ACS personnel are actively strengthening their software technology development partnerships with other LBNL divisions and with other offices in the DOE Office of Science. This effort complements ACS's already strong research partnering reputation and is designed to enhance confidence in CRD's commitment to the deployment of successful solutions and production infrastructures that advance the state of the art. Projects are currently underway in a wide array of science areas including: the Particle Data Group, High Energy Physics, Advanced Light Source, Energy Smart Grid, Energy Sciences Network, Materials Science, and Joint Genome Institute. A few example projects are described below.

Particle Data Group Workspace

The Particle Data Group (PDG) is an international collaboration charged with summarizing particle physics, as well as related areas of cosmology and astrophysics. The summaries are published in even-numbered years and is now a 1340-page book, *The Review of Particle Physics*. The review is also published and maintained as a fully hyper-linked website. The computing infrastructure supporting the PDG was conceived and built in the late '80s and, although it was modern for its time, it is no longer able to support the many participants in the review and the process of creating the review. ACS personnel work with PDG to design and build a new interactive workspace, a "User-Centered Design," that enables all the participants in the review process to collaborate and input data directly into the review infrastructure through the PDG Workspace. The initial components of this infrastructure went live in early 2012 and ACS continues to develop new features and infrastructure in collaboration with the PDG personnel.

Figure 7: PDG portal.

Smart Grid Protocol Development

Over the past several years, ACS personnel have worked closely with LBNL's Environmental Energy Technologies Division (EETD) on a series of research projects centered on the emerging electrical Smart Grid. The Smart Grid initiative intends to "re-engineer" major portions of the US energy generation and distribution system to achieve greater operational security and flexibility. Since most of the improvements mandated as part of this effort involve increased timely data communication between geographically dispersed grid components, ACS's experience in the areas of wide-area networking and real-time data acquisition and control has proven critical in understanding and evaluating the effect of proposed system upgrades. In particular, over the past two years, ACS has completed and distributed an open source implementation of EETD's automated demand response protocol (OpenADR) and is assisting in standardizing the OpenADR through the OASIS standards organization. ACS personnel are also working with major California utilities to evaluate and demonstrate secure Smart Meter to home communications of real-time energy usage through ACS's CEC-funded Open Smart Energy Gateway (OpenSEG) project. Finally, ACS personnel are currently completing a report for both DOE and CEC on the technical and operational status of the home area network deployments taking place in California, Texas, and Oklahoma.

Figure 8: The Advanced Light Source.

Advanced Light Source User Portal

The Advanced Light Source (ALS) welcomes researchers from universities, government labs, and industry who are interested in performing experiments at the general sciences and structural biology beamlines open to users. The ALS hosts over 2,000 users per year on-site. These users submit proposals to run experiments using one of the 39 beamlines at the ALS. The proposals are reviewed and if accepted, are scheduled for beam time. The user portal, in development by ACS personnel, provides the primary portal for users of the ALS with information tracking, proposal submission, proposal tracking, and visit related logistics, such as safety, scheduling, and badging. ACS works with ALS personnel to design and build this portal, which leverages existing systems at the lab and at other light sources.

Energy Sciences Network's Re-implementation of ESxSNMP

Reliable, real-time measurements of the core network are essential for the performance of next-generation networks. ESnet eXtensible SNMP System, or ESxSNMP, is a system developed at ESnet for collecting and storing large sets of SNMP timeseries measurements from network routers and devices. ESxSNMP makes performance data available via a remote REST-style interface, allowing for easy integration with other tools. The current ESxSNMP implementation uses a hybrid model for storing data, with a custom storage format (TSDB) for timeseries data and an SQL database for metadata. The current technology stack has proven too costly in developer time and attention to maintain. To address this problem, researchers in ACS are working with ESnet to re-implement the ESxSNMP back-end to use scalable NoSQL storage solutions. This solution will provide a solid low-maintenance foundation for the current workload and also provide scaling for faster networks, new data types, and new analysis approaches.

Joint Genome Institute: Sequence Proposal Tracking

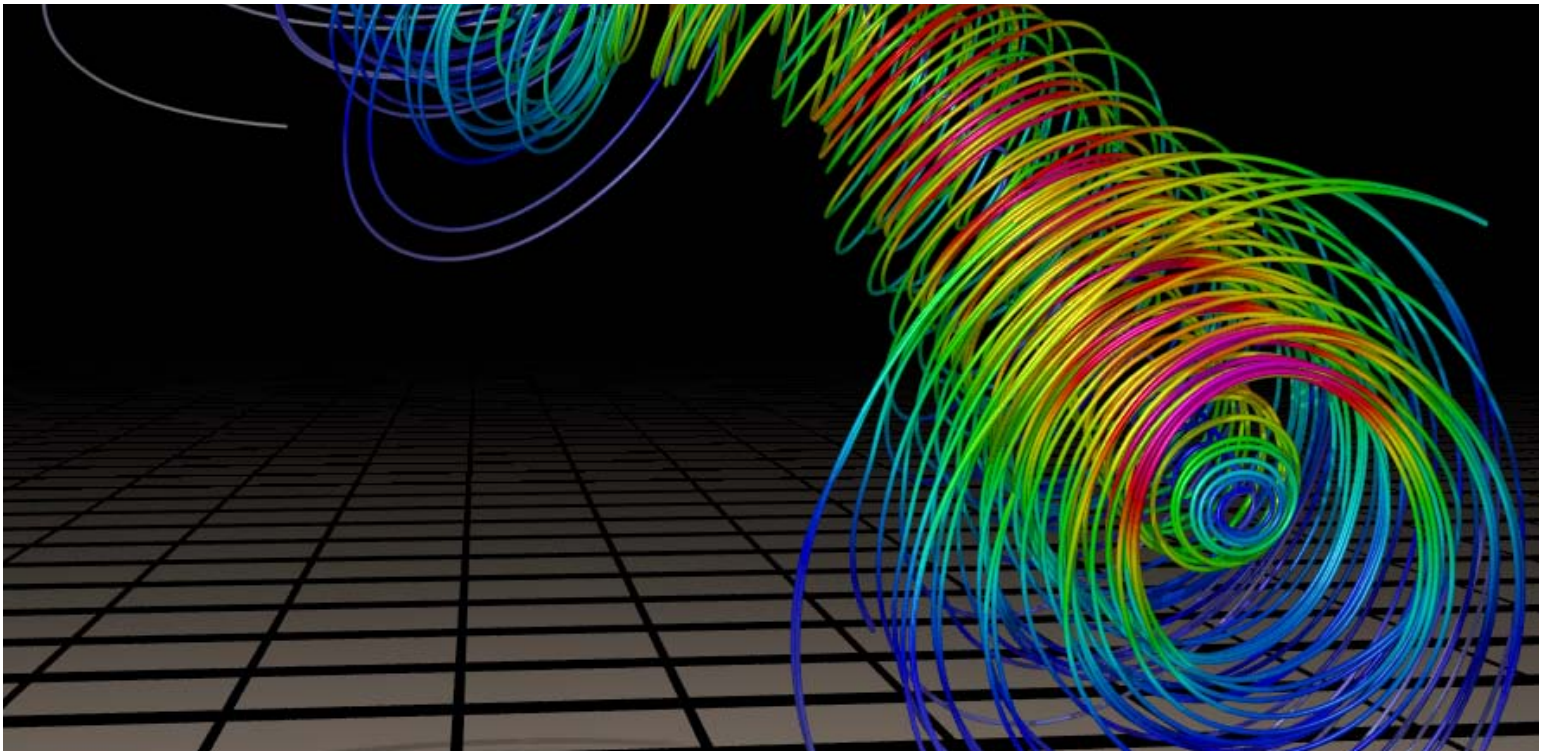
The central mission of the Genomic Technologies Department is to develop and effectively apply genomic technologies to accelerate DOE users' science. The Joint Genome Institute's (JGI) primary product continues to be a robust sample management and library construction coupled with state-of-the-art high-throughput DNA sequencing and analysis. ACS personnel helped on a short-term basis with the development of the sequence proposal tracking system. This work is concurrent with several research projects.

CRD Contact

Sarah Poon (Particle Data Group Workspace project), Charles McParland (Smart Grid Protocol Development), Craig Tull (Advanced Light Source User Portal project), Joshua Boverhof (Energy Sciences Network Reimplementation of ESxSNMP), Gary Kushner (Joint Genome Institute Sequence Proposal Tracking)

CRD Collaborators

Juerg Beringer (Particle Data Group), Mary Anne Piette (Environmental Energy Technology Division), Sue Bailey (Advanced Light Source User Services), James Floyd (Advanced Light Source User Services), Jon Dugan (Energy Sciences Networking Division)



APPLIED NUMERICAL ALGORITHMS GROUP

The mission of the Applied Numerical Algorithms Group (ANAG) is the development of advanced numerical algorithms and software for partial differential equations integrated with the application of the software to problems of independent scientific and engineering interest. The primary focus of ANAG's work is in the development of high-resolution and adaptive finite difference methods for partial differential equations in complex geometries, with applications to internal combustion engines and other industrial problems.

OVERVIEW OF HIGHLIGHTED PROJECTS

ANAG collaborates on multiple projects including the FASTMath project, funded by the SciDAC Institute, and the BISICLES project, a SciDAC program. More information about these projects can be found in "Collaborative Projects."

This section includes information about the ANAG's work on adaptive mesh refinement for atmospheric modeling, and nano-scale modeling of CO_2 sequestration.

RESEARCH STAFF

PHILLIP COLELLA received his A.B. in 1974, M.A. in 1976, and Ph.D. in 1979 from UC Berkeley, all in applied mathematics. He has been a staff scientist at LBNL and at the Lawrence Livermore National Laboratory, and from 1989 to 1995 he was a professor in the Mechanical Engineering Department at UC Berkeley.

Since 1996, Colella has been a Senior Staff Scientist and Group Lead for ANAG. He has also held a position as a Professor in Residence in the Electrical Engineering and Computer Science Department at UC Berkeley since 2010. His research has been in the area of numerical methods for partial differential equations, with contributions to high-resolution finite difference methods, adaptive mesh refinement, volume-of-fluid methods for irregular boundaries, and programming language and library design for parallel scientific computing.

He has also applied numerical methods in a variety of scientific and engineering fields, including shock dynamics, low-Mach number and incompressible flows, combustion, plasma physics, porous media flows, and astrophysical flows. Colella has led multiple multi-site/multi-disciplinary projects in scientific computing under the HPCC program in the 1990s and under the SciDAC program, including being the lead PI for the SciDAC Applied Partial Differential Equations Center (APDEC) from 2001 to 2011. His honors and awards include the IEEE Sidney Fernbach Award for high performance computing in 1998, the SIAM/ACM prize (with John Bell) for computational science and engineering in 2003, election to the US National Academy of Sciences in 2004, and election to the Inaugural class of SIAM Fellows in 2009.

PETER MCCORQUODALE received his B.S. in 1990 from Simon Fraser University, his M.S. in 1992 from Stanford University, and his Ph.D. in 1998 from MIT, all in mathematics. He was a postdoc fellow from 1998 to 2001 at the LBNL. In 2001 McCorquodale became a computational scientist in ANAG. His research has been in numerical methods for partial differential equations, particularly finite difference methods with adaptive mesh refinement. His work has been applied to problems in gas dynamics, accelerator physics, and climate modeling.

DANIEL GRAVES received his B.S. from the University of New Hampshire in 1989 and his Ph.D. from UC Berkeley in 1996, with both degrees in mechanical engineering. He has been a research scientist at LBNL since 1997 and he is currently a senior member in ANAG. His research is in the area of numerical methods for partial differential equations, with contributions in the areas of adaptive mesh refinement, Cartesian grid embedded boundary methods, massively parallel computation and library design for scientific computing. He has contributed significantly to algorithms for incompressible flow, shock physics, viscoelastic flows and elliptic solvers for magneto-hydrodynamics.

DANIEL MARTIN earned his B.S. in 1991 from the University of Florida and his Ph.D. in 1998 from UC Berkeley, both in mechanical engineering. He has been at the LBNL since 1998, first as a postdoc researcher, and now as a research scientist in ANAG. His research is in the development of numerical methods and software for solving partial differential equations, specializing in adaptive mesh refinement and finite volume methods. He has contributed to a wide range of scientific applications, including

incompressible flow, magneto-hydrodynamics, plasma physics, and land ice sheet models for climate research, while also contributing in various ways to the Chombo development and support efforts in ANAG. He is currently the lead developer for the SciDAC-funded Berkeley-ISICLES (BISICLES) land ice modeling effort. He is also the LBNL practicum coordinator for (and an alumnus of) the DOE Computational Graduate Fellowship (CSGF) Program.

PETER O. SCHWARTZ earned his B.S. in 1987 from UC Berkeley and his Ph.D. in 1994 from Ohio State University, both in mathematics. He has been at LBNL since 2001, first as a postdoc researcher, and now as a research scientist in ANAG. His research has been in the development of numerical methods and software for solving partial differential equations, specializing in environmental flows and systems biology. He has contributed to scientific applications modeling incompressible flow, reaction diffusion on complex geometric surfaces, shallow water discretizations, and land ice sheet models for climate research, while also contributing to Chombo infrastructure development. He is currently working on the SciDAC-funded Berkeley-ISICLES (BISICLES) land ice modeling effort.

TERRY LIGOCKI has been a core developer of the Chombo library for numerically solving partial differential equations using adaptive mesh refinement techniques since he joined ANAG in 2001. He focuses on the embedded boundary aspects of the Chombo library. Specifically, embedded boundary generation from implicitly defined volumes created using constructive solid geometry. His research interests include computational geometry, finite volume numerical algorithms, and transforming complex algorithms for large-scale computing. He was a member of APDEC for SciDAC and SciDAC-2, and is currently a member of FASTMath for SciDAC-3.

DAVID TREBOTICH is a computational scientist in ANAG. His research involves the development of high-resolution algorithms for complex flows in multi-scale systems using adaptive finite volume methods. Applications of interest are viscoelastic polymer flow, and subsurface flow and reactive transport at the microscopic pore scale. His current work in micro-scale fluid dynamics emphasizes multi-scale techniques to couple microscopic processes with macroscopic models. Prior to joining in LBNL in 2009, Trebotich held a staff scientist position at the Lawrence Livermore National Laboratory in the Center for Applied Scientific Computing.

GREG MILLER is a faculty scientist in ANAG and a professor of chemical engineering at UC Davis. His research involves the development of numerical methods for the solution of complex problems in engineering and science, mostly at the continuum scale. Specifically, Miller is interested in complex rheology as it pertains to multi-scale and continuum methods for incompressible viscoelastic flows, multi-phase flows for compressible and incompressible flows of fluids and elastic-plastic solids in time-dependent domains and interfacial processes, constitutive modeling (equations of state, thermodynamic modeling, symmetry invariants) and charged systems in plasma physics (boundary charge and bilayer problems).

BRIAN VAN STRAALEN received a B.S. in mechanical engineering in 1993 and a Master's of Mathematics in applied mathematics in 1995, both from University of Waterloo. His graduate research focused on the area of *a posteriori* error estimation for Navier–Stokes equations. He spent several work terms with Advanced Scientific Computing Ltd. working on computational fluid dynamics codes

written largely in Fortran 77 running on VAX and UNIX workstations 20 years ago. Straalen also worked as part of the Thermal Modeling Group with Bell Northern Research. He worked for Beam Technologies developing the partial differential equationsolve package: a combined symbolic manipulation package and finite element solver, running in parallel on some of the earliest NSF and DOE MPP parallel computers. He joined LBNL in 1998 to work with Phil Colella and to start up the Chombo Project. He is currently working on his Ph.D. in the Computer Science Department at UC Berkeley.

PAUL ULLRICH is an assistant professor of regional and global climate modeling in the Department of Land, Air and Water Resources at UC Davis. His undergraduate education is in applied mathematics and computer science from the University of Waterloo, where he also pursued a master's degree in applied mathematics and theoretical cosmology. He completed his doctorate at the University of Michigan, studying high-order finite volume methods for modeling atmospheric dynamics. His work led to the development of the MCore finite volume atmospheric dynamical core, a cutting-edge model for studying atmospheric dynamics. Ullrich is an associate editor of Monthly Weather Review and a lead organizer of the Dynamical Core Model Intercomparison Project (DCMIP). Further, he has over two years of industrial experience as a software engineer at companies such as Maplesoft and Sonic Foundry.

HANS JOHANSEN, received his Ph.D. in mechanical engineering in 1997 from UC Berkeley. He has been a staff researcher at LBNL since 2009, and his research is focused on high-performance, higher order finite volume discretizations for the embedded boundary and adaptive mesh refinement discretizations, with applications to heat and mass transfer and fluid dynamics. Prior to joining LBNL, Johansen served as executive director and chief infrastructure architect for JPMorganChase, where he was responsible for setting overall infrastructure technology direction for one of the world's largest banks. Prior to that, he was chief architect and first vice president at Washington Mutual. His projects and responsibilities spanned the implementation of high performance transaction systems, refactoring of major bank processing systems, and vendor evaluation and rationalization. Before 2002, Johansen was a senior consultant for capital markets with Random Walk Consulting (now Accenture), an early start-up architect, and a researcher in Accenture's Center for Strategic Technology. He brings broad technical expertise and 15 years of experience with financial services technology, high performance computing platforms and emerging technologies.

ADAPTIVE REFINEMENT FOR ATMOSPHERIC DYNAMICS

Research Highlights

Many features of atmospheric flows involve dynamics at multiple spatial and temporal scales. Current state large-scale simulations are limited in their ability to accurately and efficiently capture spatial features (such as clouds and orography on less than 1 km scales) and physical time scales (cloud dynamics and related weather phenomena) that interact with global dynamics. This leads to compromises between accuracy, resolution and complexity of the physics that can be included in these simulations. The resulting uncertainty in atmospheric dynamics leads to questions about the accuracy of weather prediction (such as tropical cyclones) and/or related effects in global climate change, like the frequency and severity of extreme events.

The ANAG team is implementing an atmospheric dynamic core (called a *dycore*) road map that uses the HPC software framework, Chombo, adapted to non-hydrostatic atmospheric flows (see Figure 1). The approach uses a new accurate higher order conservative finite volume approach for thin atmospheres. The cubed sphere mapping avoids polar singularities and large variations in grid cell aspect ratio; mapping discontinuities are handled carefully to be conservative to not introduce grid errors. In addition, a novel mathematical splitting and solver allows ANAG to avoid performance penalties due to extremely thin spherical layers, which are a common complication in climate modeling. The "thin cells" are typically dealt with by introducing physical simplifications, limiting total simulation time and/or creating parallel performance issues. ANAG's approach has excellent parallel efficiency and exhibits near-perfect scaling without any compromises to accuracy, resolution, or complexity.

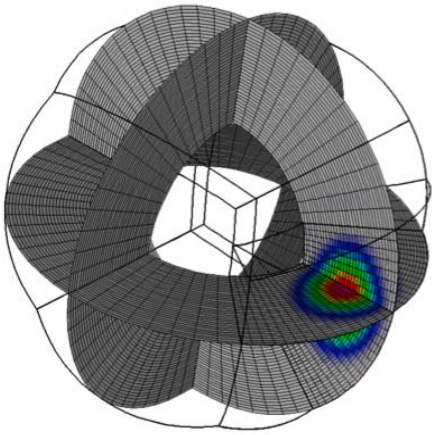


Figure 1: Cut-away of 3D cubed sphere domain for a simple scalar advection test case. Dark lines indicate block boundaries, as well as adaptive refinement boundaries around the patch.

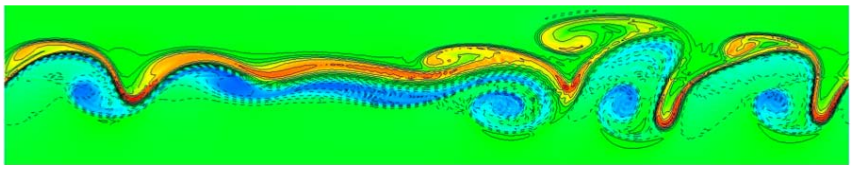


Figure 2: Barotropic instability using shallow water equations modeled on cubed sphere. This is a standard test case for validating atmospheric dycores.

Scientific Impact

When the project is completed, it will allow climate scientists to examine the effects of time accuracy and adaptive resolution on their simulation results. Currently, such studies are not within reach of even DOE leadership class machines at NERSC or ORNL. This will in turn help scientists validate other climate simulation codes and physical models, and improve the accuracy of both weather and climate predictions.

CRD Contact

H. Johansen, P. McCorquodale, P. Colella

Collaborators

P. Ullrich (UC Davis), William Collins (LBNL ESD)

ADAPTIVE REFINEMENT FOR ATMOSPHERIC DYNAMICS ADVANCED ALGORITHMS FOR HIGH PERFORMANCE COMPUTATIONS OF REACTIVE TRANSPORT PROCESSES ASSOCIATED WITH CARBON SEQUESTRATION

Research Highlights

The Chombo-Crunch project, funded by SciDAC-e, has been uniquely positioned from the project's beginning in 2009 to have a significant impact on the modeling and simulation effort of the Energy Frontier Research Center for Nanoscale Control of Geologic CO_2 (EFRC-NCGC) at LBNL. The EFRC is focusing on subsurface pore scales to explicitly resolve the key fluid–fluid (H_2O and CO_2) and fluid–solid interfaces, in order to arrive at a mechanistic understanding of how to control CO_2 injection in subsurfaces. By carefully understanding processes at the pore scale, the goal of this BES-funded research center is to bring such knowledge to bear on the macroscopic scale of a reservoir, which is the relevant scale for carbon sequestration and other important subsurface flow problems, such as the environmental remediation and storage of nuclear wastes.

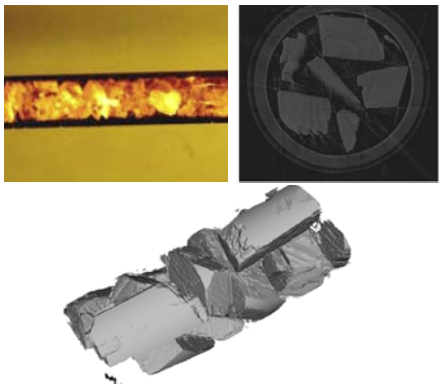


Figure 3: (Top left) EFRC reactive transport experiment in capillary tube (7 mm long) packed with crushed calcite. [Courtesy of L. Yang.] (Top right) Image slice of capillary tube obtained at ALS. [Courtesy of J. Ajo-Franklin.] (Bottom) Simulation grid for 1/6 section of crushed calcite in capillary tube.

The overall objective of the computational research is to develop a new high performance simulation capability based on advanced algorithms and software developed in CRD and, ultimately, to develop new approaches for representing chemical kinetics and interfacial dynamics in Darcy scale numerical models (upscaling). NCGC researchers hypothesize that a quantitative, mechanistic understanding of coupled physical and chemical processes at the pore-and-grain scale is possible, and that such an understanding can form a rational basis for upscaling geochemical kinetics to dynamic field scale systems.

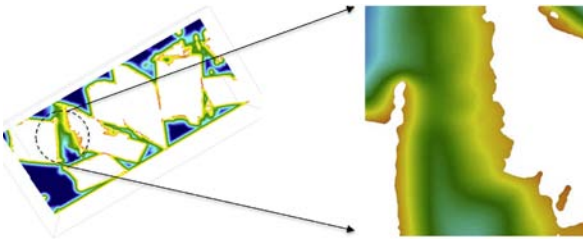


Figure 4: Simulated saturation index (slice plane) and image data resolution in simulation (1 µm).

ANAG's contribution is to use adaptive, embedded boundary methods developed in the Chombo framework to solve viscous, incompressible flow and conservative transport (advection-diffusion) in very complex geometries obtained from synthetic and realistic pore space (image data). Using operator splitting in a sequential non-iterative approach, ANAG researchers combine the geochemistry module of the software package CrunchFlow with Chombo to model multi-component geochemical reaction networks. Ultimately, the goal of this work is to model a multi-phase, reactive transport where individual phases are resolved and the pore space geometry is time-dependent in response to precipitation or dissolution.

Currently, ANAG has completed Chombo-Crunch simulation results for an EFRC validation experiment of reactive transport in a capillary tube packed with crushed calcite (see Figure 3). ANAG researchers achieved an image data resolution of the experiment at 1 micron such that the surface grains of the calcite are resolved (see Figure 4). They also developed proof-of-concept algorithms for time-dependent mineral boundaries and multi-phase flow.

Scientific Impact

Currently, the state of the art in multi-scale modeling is volume-averaging of pore scale results, which does not retain variability of the microscopic system nor heterogeneity of the pore space. The success of this project will result in an experimentally-validated simulation capability that can be used to model an ensemble of pore scale computational experiments. Such an ensemble can in turn be sampled to obtain a statistical representation of reactive transport parameters, such as permeability and reaction rates, providing more accurate parameterizations of the corresponding terms in the macroscopic Darcy equations, while retaining variability and heterogeneity. This work can also be applied to transport in other porous media, such as in energy storage materials.

CRD Contact

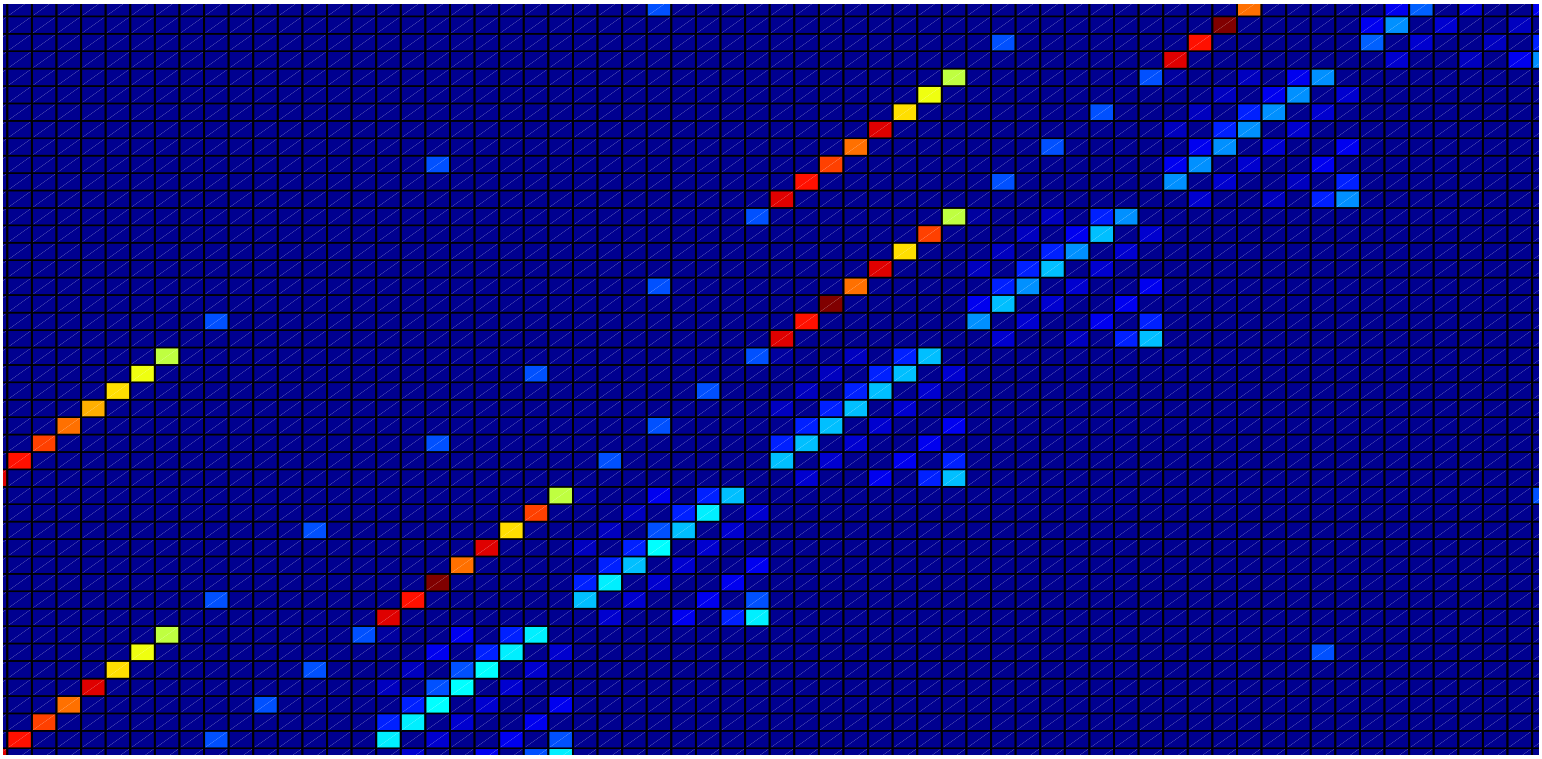
David Trebotich

CRD Collaborators

Dan Graves, Terry Ligocki, Brian Van Straalen

Other Collaborators

Greg Miller (CRD/UCD), Mark Adams (Columbia Univ.), Carl Steefel (LBNL/ESD), Sergi Molins (LBNL/ESD)



COMPLEX SYSTEMS GROUP

Mathematical research is playing an increasing role in the analysis and optimization of complex systems, including applications in network interconnected systems like computer networks and electrical power grids. LBNL's CRD has significant research activities in this area, and several projects in CRD develop the algorithms, tools and techniques to meet the analysis challenges posed by complex systems.

The Complex Systems Group (CXG) was started in 2010 to explore new opportunities in this field. CXG is currently engaged in research areas including the performance analysis of HPC systems, automatic performance tuning of scientific programs, mathematical analysis for cybersecurity, and financial system modeling.

OVERVIEW OF HIGHLIGHTED PROJECTS

Highlighted topics in this section include cybersecurity projects, intrusion detection for high performance computing, and an overview of the Center for Innovative Financial Technology (CIFT). Many of CXG's projects are a collaborative effort with researchers at UC Davis and other academic institutions.

RESEARCH STAFF

DAVID H. BAILEY, the Group Lead for CXG in CRD, is a leading figure in the high performance scientific computing and computational mathematics fields. He is the author of five books and more than 160 research papers. He has received many awards for his contributions to these fields, including the Sidney Fernbach Award from the IEEE Computer Society, the Gordon Bell Prize from the Association of Computing Machinery, and the Chauvenet Prize and the Merten Hasse Prize from the Mathematical Society of America.

AYDIN BULUÇ is a research scientist in CXG working on high performance graph analysis, libraries, and their applications in genomics and bioinformatics. His current research also includes parallel sparse matrix computations and communication-avoiding algorithms. Previously, he was a Luis W. Alvarez Fellow. He earned his doctorate in computer science from UC Santa Barbara in 2010.

ORIANNA DEMASI is a student and part-time researcher in CXG who is working on mathematical techniques to detect intrusions in high performance computer networks. Currently, she is a computer science Ph.D. student at UC Berkeley. In 2009, she received her B.S. in mathematics from McGill University in Montreal, Canada, where she received an undergraduate award to study differential equations with state-dependent delays. Some of her projects included a summer fellowship modeling tidal heights and velocities in the portion of the New England shelf, south of Martha's Vineyard.

DAVID LEINWEBER heads the Center for Innovative Financial Technology, created to help build a bridge between the computational science and financial markets communities. Leinweber was a Haas Fellow in finance at UC Berkeley from 2008 to 2010. His professional interests focus on how modern information technologies are best applied in trading and investing. As the founder of two financial technology companies and a quantitative investment manager, he is an active participant in today's transformation of markets. Prior to LBNL, Leinweber also served as Managing Director at First Quadrant, where long and market neutral strategies utilized a wide range of computerized techniques for stock selection and efficient trading. As a visiting faculty member at Caltech, Leinweber worked on practical applications of ideas at the juncture of technology and finance. Over the years, Leinweber has advanced the state of the art in the application of information technology in both the sell-side world of trading and the buy-side world of quantitative investment. He has published and spoken widely in both fields. He is an advisor to investment firms, stock exchanges, brokerages, and technology firms in areas related to financial markets. Leinweber graduated from MIT in physics and computer science and earned his Ph.D. in applied mathematics from Harvard University.

SEAN PEISERT has a joint appointment as a research scientist in CXG and he is an assistant adjunct professor and faculty member in the Computer Science Department at UC Davis. He performs research in computer security and is particularly interested in forensic analysis, intrusion detection, electronic voting, smart grid security, the insider threat, security policy modeling, and empirical studies of

security. Previously, he was an I3P Research Fellow at UC Davis and a computer security researcher at the San Diego Supercomputer Center (SDSC). He received his Ph.D., M.S., and B.S. in computer science from UC San Diego. Peisert is actively involved with the academic computer security community and is a steering committee member of the New Security Paradigms Workshop (NSPW) and the Workshop on Cyber Security Experimentation and Test (CSET). He is co-representative of LBNL to the Institute for Information Infrastructure Protection (I3P). Peisert serves as a reviewer, program committee member, and organizer for numerous journals and conferences, including the IEEE Symposium on Security and Privacy, the flagship conference for security research.

ALEXANDER SLEPOY is currently working on parallel scaling, cybersecurity and large data algorithms in CXG. He received his Ph.D. in physics from UC Davis in 2001, with a research focus in statistical mechanics. In subsequent work, mostly at DOE-funded national laboratories, he has worked in computational biology, whole cell modeling, protein folding and large-scale biochemical reaction systems. This involved extensive work in parallel algorithms, constant-scaling algorithms and large-scale parallel code design.

ANUBHAV JAIN is currently an Alvarez Postdoctoral Fellow in CXG. Jain received his B.S. at Cornell University in applied and engineering physics, and his Ph.D. at MIT in materials science and engineering. His background is in materials design using quantum chemistry codes. Jain currently works on the Materials Project, which is an effort to compute the properties of all known inorganic compounds and provide this data freely to materials designers around the world looking for technological breakthroughs. This effort involves a mix of theoretical physics, materials science, high performance computing, and cutting-edge web technologies.

INTRUSION DETECTION FOR HIGH PERFORMANCE COMPUTING

Research Highlights

The "Fingerprinting" high performance computing (HPC) programs project, now in its fourth year, addresses multiple elements of intrusion detection in HPC using a mathematical and data-driven approach. This project is a collaboration between LBNL, UC Davis, and the International Computer Science Institute (ICSI).

The overall goals of this project, also called the Cybersecurity project, are to develop mathematical and statistical methods to detect intrusions of HPC systems. CXG's mathematical analysis is predicated on the fact that large HPC systems represent unique environments, quite unlike unspecialized systems or general Internet traffic. User behavior on HPC systems tends to be much more constrained (often driven by research deadlines and limited computational resources), and is generally limited to certain paradigms of computation (the set of codes performing the bulk of execution provide a rich source of information). These special characteristics of an HPC environment permit researchers and collaborators to perform the equivalent of credit card fraud detection on the usage of these systems. In this research, CXG employed real system data, which they obtained in collaboration with staff in the NERSC Division.

There are many thrusts that this project explores. The most interesting thrust and most likely to have the greatest impact asks: what are people running on LBNL's HPC systems? Are they running what they usually run? Are they running what they requested cycle allocations to run? Are they running something illegal or something against policy?

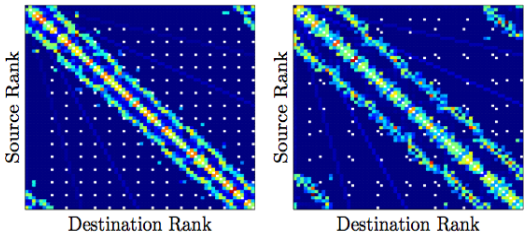


Figure 1: Adjacency matrices of HPC communications.

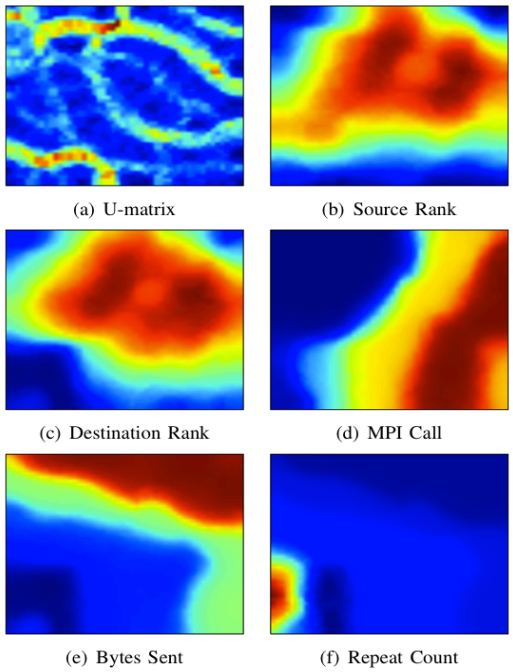


Figure 2: Self-organizing maps of IPM data.

Using both graph theoretic [25] and supervised machine learning [10, 13, 26-28] approaches, CXG developed techniques to classify HPC programs based on traces of communication patterns between processors (e.g., those collected using IPM[1] or Tau[2]). Specifically, CXG looks at the adjacency matrices (see Figure 1), performance counters, and other features that characterize the type of computational math being performed, with regard to a particular computational "dwarf" or "motif."

Understanding the important elements in intrusion detection is key in the success of this project. Ideally, CXG researchers identify the elements by automatically inferring the elements from the data [24]. To infer the elements, they employ two key techniques. First, they use self-organizing maps (see Figure 2), an unsupervised clustering

[1] Integrated Performance Monitoring (IPM): http://ipm-hpc.sourceforge.net.

[2] Tuning and Analysis Utilities (TAU): http://tau.uoregon.edu

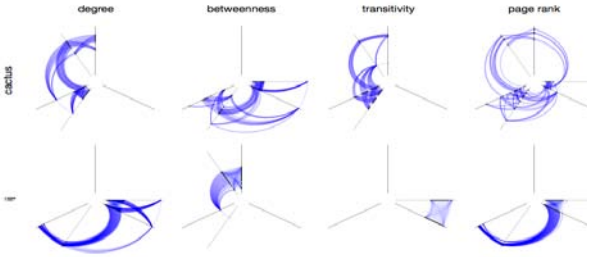


Figure 3: Hive plots of IPM data.

that uses a type of neural net. It is a nonlinear generalization of principle component analysis and is also like k-means clustering, but infers the number of clusters directly from the data.

Second, CXG researchers apply hive plots (see Figure 3) to analyze characteristics of message-passing networks formed by HPC codes. These plots suggest the features that contribute most to accurate classification. In particular, the plots visually reveal distinctive features of the communication patterns of these codes, as expected from different math algorithms embedded in the codes.

Related to the work on fingerprinting HPC programs, CXG researchers are also developing methods for data sanitization. Data sanitization is the process of transforming data (typically sensitive data) in such a way that the transformed version can be viewed by possible adversaries but such adversaries cannot determine the sensitive data. An example of this process is the data anonymization or "de-identifying" data (e.g., for releasing health records to medical researchers). Specifically, CXG research works to understand when data can be "sanitized" and when it cannot by taking into account both privacy requirements (what cannot be revealed) and analysis requirements (what must be revealed to enable appropriate use of data for either operational or research means). Further, the researchers assume an "open world" model in which adversaries have access to external knowledge that may help reveal even sanitized data elements. Thus, CXG seeks to determine what relationships will allow an adversary to undo sanitization [1].

In relation to this, CXG researchers are creating ways to develop audit trails that do not create better "insiders" by enabling mono- or bi-directional covert channels [19]. For example, in electronic voting, there needs to be a way to prevent voters from communicating with auditors, as this enables vote selling. Likewise, electronic voting methods need to prevent auditors from determining a voter's identity to prevent voter coercion. CXG researchers and collaborators are looking at codifying analyzability and anonymity relationships via ways of adding noise, enforcing regularity, and/or using procedural methods.

Access to substantial data and subject matter for experts is a key factor in the success of this research. CXG is currently using data from electronic health records supplied by the UC Davis Health System. They are also developing formal method techniques by constructing analyzable ontologies. Their ability to successfully apply the technique and successfully abstract the key elements should give researchers the ability to work with domain experts from other disciplines to construct ontologies and ultimately understand when data can be sanitized or not based on the requirements of the stakeholders. This technique should function regardless of the application to medical data or scientific/HPC data.

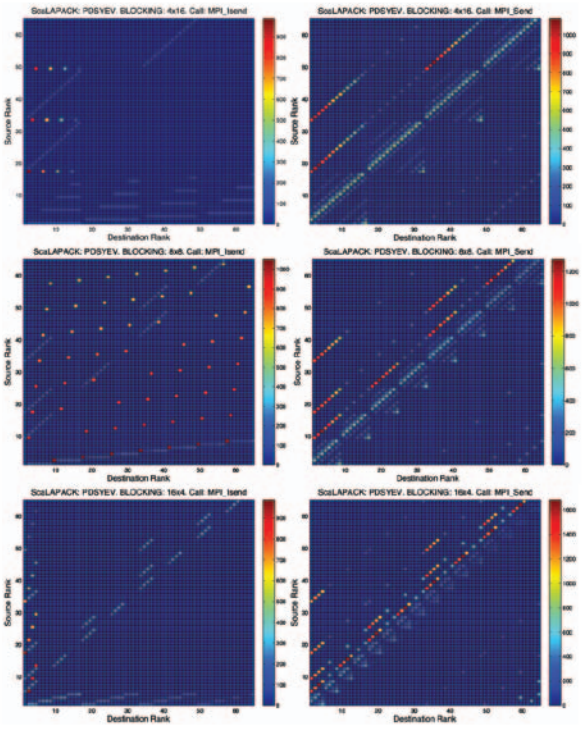


FIGURE 4: Communication pattern in ScaLAPACK: pdsyev (calculation of all eigenvalues) for MPI_Isend and MPI_Send with different processor blocking dimensions. Figures show that the blocking layout affects the adjacency matrix.

Scientific Impact

For fingerprinting HPC systems, CXG's results have been positive: they developed a Bayesian-based machine learning technique for the classification of dozens of scientific computations. They also developed a graph-theoretic approach using "approximate" graph matching techniques (e.g., graph edit distance); true positive rates of 85.5%, false positive rates of 0.6%. They employed a hybrid machine learning/graph theory approach that currently identifies test HPC codes with a 95–99% accuracy rate. And finally, they employed hive plots to give more intuition for the communication patterns than traditional CCDF plots.

Many studies have identified that there is a distinct behavior in a code's performance logs, but none of these studies address the uniqueness of this behavior, disparity from the behavior of other similar computations, or robustness of behavior to method parameters, problem size, and the number of nodes the code was run on. CXG began to address the question of how unique a code's behavior is by logging performance of the popular scientific codes in controlled environments [7]. The dataset they generated considers if a code's identity can be hidden under its scaling pattern. The goal of such a dataset is to generate a set of statistics that will uniquely identify codes regardless of the configuration that they were run in. CXG researchers then use their derived statistics with supervised machine learning methods to show that the statistics are flexible enough to be used in identifying a code.

An example of how runtime configurations affect the "appearance" of communication is shown in Figure 4. The figures show the same computation run under three different blocking configurations. Upon careful observation, a consistent pattern in computation is apparent, but such flexibility in the communication must be allowed for accurate identification of codes.

A significant contribution of CXG's work is using a modified rule ensemble method from machine learning on a large and diverse dataset [7]. They were able to correctly classify codes with high accuracy even though they had a large class imbalance and other features in the dataset that are representative of real HPC workloads.

To collect the dataset, CXG researchers gathered real runs of scientific codes at NERSC. This method allowed their data to have more depth and breadth than similar datasets that were used in previous studies. The observations collected are not just benchmark codes or multiple runs using a single functionality of a single code, but representative of the complex nature of computations that is routinely executed at supercomputing facilities. Because the dataset was collected from an uncontrolled environment, it represents real workloads well at supercomputing facilities. The datasets CXG worked with also had many more observations than previous sets, which allowed for better inference about behavior beyond the scope of their dataset.

CXG researchers modified a rule ensemble method for their application, and were able to correctly classify codes with up to 96% accuracy. Before applying the method to this dataset, they studied that extension of the method to multi-class problems and the use of different solvers to generate more compact models [8]. CXG also introduced the flexibility for the algorithm to classify codes as ambiguous and thus, introduced a way to draw further attention to particularly difficult observations. This flexibility allowed them to improve the classification even further as the majority of the observations that would have previously been misclassified are now flagged for further review.

CRD Contacts

David H. Bailey, Sean Peisert

LBNL Collaborators

Deborah Agarwal, Scott Campbell, Orianna DeMasi (also UCB), Sean Peisert (also UCD), Taghrid Samak, Alexander Slepoy

UCD Collaborators

Matt Bishop, Anhad Singh

*Researchers previously involved include Alex Kaiser (NYU), Juan Meza (UC Merced), Vern Paxson (LBNL, UC Berkeley/ICSI), Robin Sommer (ICSI/LBNL) and Sean Whalen (LBNL, UCD).

*Research collaborators from other institutions who have collaborated in an unpaid capacity include Prof. Sophie Engle (University of San Francisco) and Prof. Michael Hogarth (UC Davis School of Medicine).

Research Highlights

On May 6, 2010, around 2:45 PM, the Dow Jones Industrial Average plunged nearly 1000 points, and then recovered in a few minutes. Such severe fluctuations in financial markets have significantly undermined investor confidence.[1] Both investors and regulators are concerned that the dramatic rise of computer algorithm-based trading has introduced instabilities in financial markets.

The Center for Innovative Financial Technology (CIFT) was established in July 2010 in the Computational Research Division to build a bridge between high performance computing and financial markets, in an attempt to address the challenges posed by market instabilities.

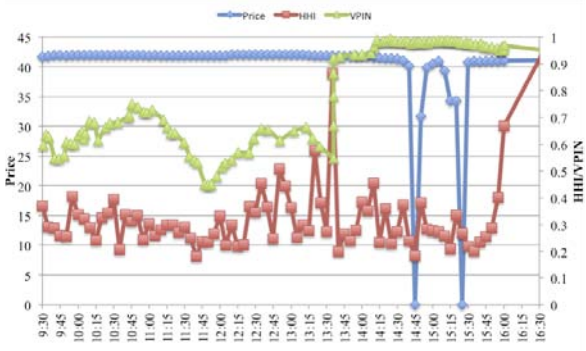


FIGURE 5: The price range (every five minutes) plotted against VPIN (red) and HHI (green) on May 6, 2010.

CIFT is a center for technology integration where technologies, developed in different parts of CRD, are applied to solve critical problems in financial markets. It is also a platform for an open exchange of ideas; where the market participants and academic research can bring their ideas together in an open science environment. This is reflected in the composition of CIFT, which includes computational experts, data management specialists, visual analytics gurus, academic statisticians, as well as a number of active traders and fund managers.

CIFT's hypothesis is that state-of-the-art large-scale data analysis and visualization technology can be used to better understand and regulate the behavior of modern electronic markets. In particular, "yellow light" indicators can be formulated and studied using CIFT's computer-based methods. This yellow light may be able to prevent catastrophic market swings more effectively and less intrusively than the "circuit breaker" approach, currently in use.

Based on professional trade and market regulator needs, CIFT has identified two indicators of market volatility: Volume Synchronized Probability of Informed Trading (VPIN) and the Herfindahl–Hirschman Index (HHI). VPIN is designed to measure imbalances between buying and selling in the market and can be applied to overall market indicators such as the Dow Jones Industrial Average, as well as individual financial instruments such as a stock or a commodity futures contract. HHI measures the market fragmentation, and the particular version of HHI used in CIFT's study is known as the volume HHI, which can be computed from the volumes of trades. Figure 5 shows VPIN and HHI on May 6, 2010. From this figure, both VPIN and HHI experienced a short jump around 13:30, about 75 minutes ahead of the first Flash Crash event, where the price of ACM (Accenture) dropped from roughly 10 to 0.01, only to mostly recover a few moments later. This suggests that VPIN and HHI could be suitable for detecting unusual events in stock trading.

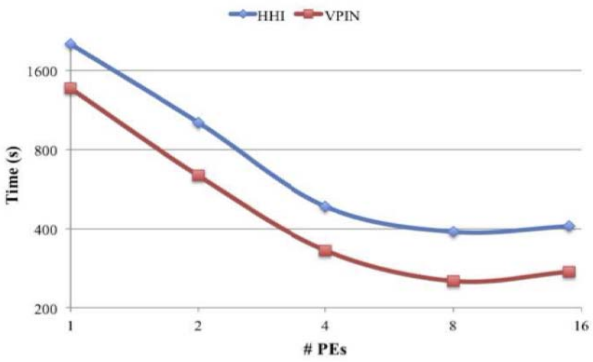


FIGURE 6: Time to compute VPIN and HHI for 25 Dow Jones Industrial Stocks can scale quite well with the number of processing elements (PE) used.

[1] For example, see http://www.nytimes.com/2012/08/02/business/unusual-volume-roils-early-trading-in-some-stocks.html.

Scientific Impact

Prior to CIFT's work, the computations on VPIN and HHI were only applied to the market indexes, such as the Dow Jones Industrial Average. CIFT, however, was the first to show that computing VPIN and HHI on individual stocks can actually detect unexpected events. CXG researchers at CIFT further parallelized the computation so that VPIN and HHI can be computed quickly. Figure 6 shows the time needed to compute VPIN and HHI scales well, as the number of processors increases.

CIFT's work demonstrates that VPIN and HHI are able to predict some unusual events in the market, and it is possible to reduce the computation time of these indicators to real-time. Following these initial steps, CIFT collaborators plan to further study the suitability of VPIN and HHI for detecting extreme volatility in the market. In future work, they also plan to work with other industrial partners to evaluate additional measures.

CRD Contacts

David H. Bailey, David Leinweber

CRD COLLABORATORS

E. Wes Bethel, K. John Wu, Oliver Rübel, Xin Guo

OUTSIDE COLLABORATORS

Ted Aronson (Principal, AJO Partners, former CFA Institute President), Aaron J. Lebovitz (Principal, Infinium Capital Management), Richard R. Lindsey (President and Chief Executive Officer, Callcott Group LLC), Marcos Lopez de Prado (Head of Quantitative Research, Tudor Investments), Ricardo Rojas (High-Frequency Trading, Deutsche Bank), Larry Tabb (CEO, Tabb Group)

COMPUTER SECURITY PROJECTS

Research Highlights

CXG is an active participant in a number of other projects in the arena of computer security, in addition to the "Mathematical and Data-Driven Approach to Intrusion Detection for High-Performance Computing" project. These projects span across UC Davis and LBNL and also include numerous other academic collaborators. Research sponsors currently include DOE's Advanced Scientific Computing Research program, DOE Office of Electricity Delivery and Energy Reliability's Cybersecurity for Energy Delivery Systems (CEDS) program, DOE's National Nuclear Security Administration (NNSA), and NSF's Security and Trustworthy Cyberspace (SaTC) program.

Application of Cybersecurity Techniques in the Protection of Efficient Cyber-Physical Energy Generation Systems

There are many physical systems today that impact or are impacted by networked computers. The physical systems themselves have always been designed with extremely high degrees of safety in mind. However, there are assets in physical plants, like the UC Davis Central Heating and Cooling Plant, that were never intended to be connected to any network. Furthermore, there is no mechanism in place, nor test or certification process, for the networked system, ensuring that machines collectively have no damaging behavior. As the targets of the *Stuxnet* worm now know, systems often *do* allow behavior that is damaging to individual devices since computer systems frequently can be misinformed or ignore tolerances of the physical systems they control.

In response to this problem, CXG researchers design and develop a security monitoring and analysis framework for control systems. The goal is to integrate the monitoring and analysis of network traffic and serial communication with an understanding of physical device constraints within a single intrusion detection system (IDS) to enhance the resilience of cyber-physical systems. CXG's technique monitors electrical loads, which has shown that this security problem is tractable [12].

This project is funded by DOE's CEDS program and is led by Chuck McParland and Sean Peisert, with Anna Scaglione (UC Davis) and graduate students. This project is also in collaboration with Robin Sommer (LBNL/ICSI) and researchers at the University of Twente.

Cyber Sciences Lab (CSL)

Eight DOE national laboratories, using seed funding from the NNSA Chief Information Officer, are working together on a vision for an enduring, national computer security research laboratory to address cybersecurity threats to the nuclear deterrent, energy infrastructure, and the nation's reliance on cyberinfrastructure at large. Research efforts that the laboratory would address vary from very short-range, tactical issues that leverage current capabilities, to very long-range research with results and solutions that may not be deployable for over 20 years, using techniques that likely do not currently exist. At LBNL, the CSL effort is led by Deb Agarwal and Sean Peisert.

Symbiosis in Byzantine Fault Tolerance and Intrusion Detection

Two principal components for providing protection in large-scale distributed systems are Byzantine fault tolerance (BFT) and intrusion detection systems (IDS). However, BFT traditionally suffers from high-latency and replication requirements. As these two components approach system security differently, though, intrusion detection has the potential to improve BFT. The integration of these two techniques, at both the fundamental and system levels, is the theme of this research effort.

As a first step in providing protection to distributed systems, CXG researchers and collaborators developed a new BFT algorithm called *h*BFT. *h*BFT is a leader-based protocol that uses speculation to reduce the cost of Byzantine agreement protocols with optimal resilience. *h*BFT satisfies three key goals. First, it tolerates Byzantine faulty clients. Second, it shows critical jobs can be moved to the clients without additional costs. And third, the notion of normal cases is defined, which means the primary is correct and the number of faulty backups does not exceed the threshold. From this protocol, CXG provided better performance for both fault-free cases and normal cases. They developed and validated this protocol by implementing it and testing it on the Emulab testbed (see Figure 7) [9].

CXG also developed a Byzantine fault-tolerant (BFT) state machine replication (SMR) protocol, *BChain* (see Figure 8), that detects and corrects errors while outperforming previous practical BFT protocols (e.g., PBFT, Zyzzyva, etc.), for both latency and throughput. They employ chain replication concepts to allow replicas to execute client requests in a pipeline manner. CXG also integrated this with a Byzantine fault diagnosis mechanism such that faulty replicas can eventually be removed from the system (e.g., if hardware has failed) or restored (e.g., if software is simply compromised). Thus, the mechanism performs better than previous BFT protocols while also adding proactive recovery, mitigating long-standing problems for long-lived systems.

This project is co-led by Karl Levitt and Sean Peisert at UC Davis and Keith Marzullo at UC San Diego/NSF, and is funded by the NSF's SaTC program. This project is also in collaboration with Hein Meling at University of Stavanger, Norway.

		PBFT	HQ	Fast	Zyzzyva	*h*BFT
Cost	Replicas	$\mathbf{3f+1}$	$\mathbf{3f+1}$	$5f+1$	$\mathbf{3f+1}$	$\mathbf{3f+1}$
Throughput (MAC ops/req)	Primary	$2+\frac{10f}{b}$	$4+4f^{\parallel}$	$2+\frac{9f}{b}$	$4+5f+\frac{3f}{b}^{\dagger}$	$\mathbf{2+\frac{3f}{b}}$
	Backup	$2+\frac{10f+1}{b}$	-$^{\parallel}$	$2+\frac{9f+1}{b}$	$3+5f+\frac{1}{b}^{*}$	$\mathbf{2+\frac{3f}{b}}$
	Client	$\mathbf{2+6f}^{**}$	$4+4f$	$2+9f$	$4+11f^{\ddagger}$	$\mathbf{2+6f}^{\P}$
1-way latency	Crit. pth.	$4/5^{\S}$	4	4	**3**	**3**
Handle concurrency well?		Yes	No	Yes	Yes	Yes
Handle faulty clients?		No	No	No	No	**Yes**

FIGURE 7: Comparison of BFT protocols to tolerate *f* faults with batch size *b*.

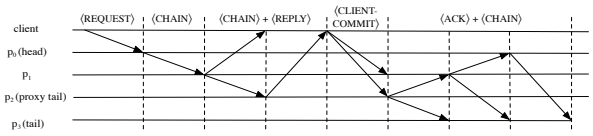


FIGURE 8: Communication pattern in chaining subprotocol.

Process-Based Technology to Support Comparison and Evaluation of the Security of Elections

Constructing process models of physical systems and human environments allow the analysis of those processes in the same way that source code can be analyzed using static analysis techniques. Specifically, process models are constructed by hand using expert, domain knowledge of how a system is supposed to function, containing descriptions of events, actors, and objects. Subsequently, models can be incrementally refined over time through expert analysis, as well as formal analysis. CXG researchers' and collaborators' work explores process composition tools as they apply to elections. This work includes the exploration of how to compose systems from pre-analyzed process components, how to analyze the vulnerability of these systems to attacks, and how to guarantee that important security properties are ensured for the resulting composed system.

The research team works closely with the Marin County Registrar of Voters' office and the Yolo County Clerk-Recorder's office. At present, they have implemented process models using the Little-JIL language in the context of elections (see Figure 9), and performed formal, static analyses on the fault tree to identify hazards in the election processes [22, 23]. They are continuing to expand this work by including enhanced provenance analysis by adding an artifact-centric analysis in addition to an actor-centric analysis. Finally, future work will allow this technique to be applied in additional domains to refine the processes of finite state verification and attack tree extraction from fault trees.

This project is funded by NSF's SaTC program and involves UC Davis researchers Matt Bishop, Bertram Ludäscher, and Sean Peisert, and UMass Amherst researchers Lee Osterweil, Lori Clarke, and George Avrunin.

Auditing Voting Systems While Preserving Secrecy and Anonymity

Election auditing verifies that the systems and procedures work as intended, and that the votes have been counted correctly. However, the audit trails, may contain information that either exposes the identity of the voter (enabling voter coercion); or contain information that communicates a message to a third party (enabling vote selling). The goal of this project is to determine the information needed to assess whether e-voting machines operate with the desired degree of assurance, especially with respect to anonymity and privacy [5, 6, 17-19]. The CXG researchers conducted a detailed analysis of the open-source Scantegrity electronic voting system and are broadening their study to include an analysis of penetration testing/red-teaming techniques.

This project is co-led by Matt Bishop and Sean Peisert and is funded by the NSF's SaTC program.

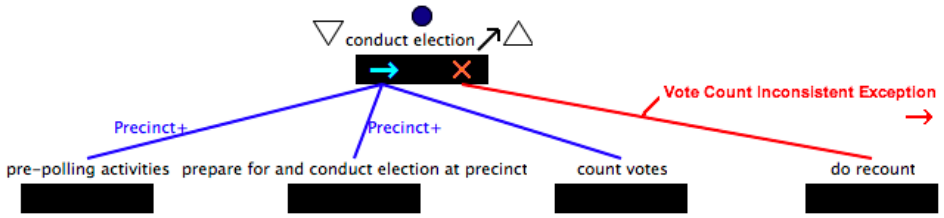


FIGURE 9: Highest level of election process.

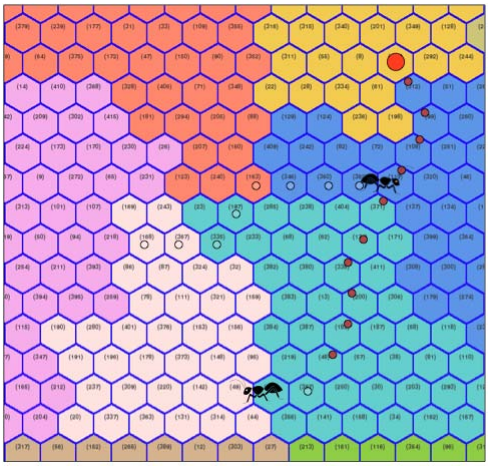


FIGURE 10: Digital ants.

The Hive Mind: Applying a Distributed Security Sensor Network to GENI

The Hive Mind project[1] defines and prototypes a security layer using a method of intrusion detection based on mobile agents and swarm intelligence (see Figure 10). The project's goal is to provide a lightweight, decentralized, intrusion detection method that is adaptable to changing threats while communicating suspicious activity across hierarchical layers to humans who can respond when needed. The goal is to augment, not replace, more traditional security mechanisms. For example, the Hive Mind should be effective where computing power is highly limited; where host-based IDSs would be impossible or in highly distributed systems without well-defined monitoring points making network-based detection infeasible. The Hive Mind can then be used in parallel with traditional IDSs.

The DETERlab and ProtoGENI served as primary implementation and system evaluation test beds for this project. This implementation focused on detecting a wide variety of attack patterns across an array of 400+ nodes, specifically virtual Linux hosts.

This project is led by Sean Peisert and is funded by NSF's CISE Directorate. It also involves collaborations with Matt Bishop at UC Davis, Carrie Gates at CA Labs, and previously involved Deborah Frincke at PNNL.

[1] The Hive Mind: http://hivemind.cs.ucdavis.edu/.

Other Past Recent Projects

Three recent, but not currently funded projects include research in computer forensics and the insider threat.

Forensics

CXG's work in computer forensics has sought to establish a rigorous scientific model of forensic logging and analysis that is both efficient and effective at establishing the data necessary to record in order to understand past events. While forensics traditionally looks at available data and attempts to draw conclusions from it, CXG, in contrast, seeks to understand the questions that want to be answered, and then derive what data is necessary to support answers to those questions [11, 15-18, 21].

Insider Threats

CXG's work on insider threats also takes a non-traditional approach. Whereas, security has traditionally been defined with respect to a perimeter, using static and binary access control decisions, CXG asserts that such a perimeter no longer exists and traditional access control techniques inhibit authorized users from performing their job. They define the *insider threat* as a combination of (a) access to a particular resource, (b) knowledge of a particular resource, and/or (c) trust of an individual by a particular organization. Moreover, the insider threat is clearly not binary, but a spectrum of *insiderness* based on the aforementioned qualities. They are developing access control solutions that integrate this understanding while also being informed by social science of how users may react optimally to system access control and countermeasures [2-4, 14].

Resilience and Clean Slate Solutions

CXG developed and evaluated a set of security requirements for critical systems, which included fundamental premises that those requirements would entail, and ideas for implementations that would instantiate those premises [20]. The overriding requirement guiding their paradigm was that "first principles" reflected the only real security strategy, where first principles are ideally provable, often measurable, and at least possible to order and bound. These principles allowed the

researchers to realize that many security policies may be in conflict, and as such, proofs, measures, and ordering gave an analyst (or even better, an automated system) the metrics needed in order to make informed decisions about how to resolve the conflicts. Several metrics that enable this, including state replication, data slicing, collusion, and information theory.

On the other hand, sometimes the use of Byzantine fault tolerance and formal methods as described in CXG's "first principles" solution comes only at great cost. Moreover, although such systems use *3f + 1* independent replicas and/or redundancy to protect against *f* failures, this model itself can be attacked by overwhelming the traditional *f* threshold. Methods exist to merge intrusion detection concepts with Byzantine fault tolerance, but solutions are still typically binary, and the resources to attack those solutions are often lopsidedly small. CXG researchers and collaborators focus on a non-conventional definition of *resilience* that incorporates more than availability and also addresses integrity and confidentiality [20]. They propose a dynamic solution that uses a risk-based measure of a network's current state of vulnerability [14] and an understanding of the complex failure modes of cyber-physical systems. Given the definition of *resilience*, we believe that some form of graceful transition and/or degradation to a pre-defined failure (but *survivable*) mode (e.g., fail-safe, fail-soft, fail-hard, fail-over) must be provided in the event that adversary resources exceed the *f* threshold.

This project is a collaboration between Matt Bishop, Chen-Nee Chuah, and Sean Peisert.

REFERENCES

[1] Matt Bishop, Justin Cummins, Sean Peisert, Anhad Singh, Deborah Agarwal, Deborah Frincke, and Michael Hogarth. "Relationships in Data Sanitization: A Study in Scarlet." In *Proceedings of the 2010 New Security Paradigms Workshop*, pages 151-164, Concord, MA, September 21-23 2010.

[2] Matt Bishop, Sophie Engle, Deborah A. Frincke, Carrie Gates, Frank L. Greitzer, Sean Peisert, and Sean Whalen. "A Risk Management Approach to the 'Insider Threat'." Christian W. Probst, Jerey Hunker, and Matt Bishop, editors. *Insider Threats in Cybersecurity, Advances in Information Security Series*, pages 115-138. Springer, Berlin, September 2010.

[3] Matt Bishop, Sophie Engle, Sean Peisert, Sean Whalen, and Carrie Gates. "We Have Met the Enemy and He is Us." In *Proceedings of the 2008 New Security Paradigms Workshop (NSPW)*, Lake Tahoe, CA, September 22-25, 2008.

[4] Matt Bishop, Sophie Engle, Sean Peisert, Sean Whalen, and Carrie Gates. "Case Studies of an Insider Framework." In *Proceedings of the 42nd Hawaii International Conference on System Sciences (HICSS), Cyber Security and Information Intelligence Research Minitrack*, Waikoloa, HI, Jan. 5-8, 2009.

[5] Matt Bishop, Mark Gra, Candice Hoke, David Jeerson, and Sean Peisert. "Resolving the Unexpected in Elections: Election Officials' Options." *Distributed by the Center For Election Excellence and the American Bar Association*, October 8, 2008.

[6] Matt Bishop, Sean Peisert, Candice Hoke, Mark Gra, and David Jeerson. E-Voting and Forensics: Prying Open the Black Box. In Proceedings of the 2009 Electronic Voting Technology Workshop/Workshop on Trustworthy Computing (EVT/WOTE '09), Montreal, Canada, August 10{11, 2009. USENIX Association.

[7] O.A. DeMasi and D.H. Bailey. On the uniqueness and identify-ability of high performance computing codes from their performance logs. In progress.

[8] O.A. DeMasi, J.C. Meza, and D.H. Bailey. *Dimension reduction using rule ensemble machine learning methods: A numerical study of three ensemble methods.* 2011. http://crd-legacy. lbl.gov/~dhbailey/dhbpapers/Ensemble_TechReport.pdf.

[9] Sisi Duan, Sean Peisert, and Karl Levitt. "hBFT: Fast Byzantine Fault Tolerance With Optimal Resilience." *IEEE Transactions on Dependable and Secure Computing (TDSC)*, (submitted) 2012.

[10] Sophie Engle and Sean Whalen. "Visualizing Distributed Memory Computations with Hive Plots." In *Proceedings of the 9th ACM International Symposium on Visualization for Cyber Security (VizSec)*, 2012.

[11] Karl Levitt, Lenore Zuck, and Sean Peisert. "Striking Back at the Privileged: Applicationbased Monitoring for Detection, Policy, Specification, Forensics, and Automated Response." In *Insider Threats: Strategies for Prevention, Mitigation, and Response,* no. 10341, Dagstuhl Seminar Proceedings. Dagstuhl, 2010.

[12] Xiao Li, Zhifang Wang, Vishak Muthukumar, Anna Scaglione, Chuck McParland, and Sean Peisert. "Networked Loads in the Distribution Grid." In *Proceedings of the 2012 APSIPA Annual Summit and Conference*, Hollywood, CA, Dec. 3{6 2012.

[13] Sean Peisert. *Fingerprinting Communication and Computation on HPC Machines.* Technical Report LBNL-3483E, Lawrence Berkeley National Laboratory, June 2010.

[14] Sean Peisert. "Optimistic Access Control and Anticipatory Forensic Analysis of Insider Threats." In *Insider Threats: Strategies for Prevention, Mitigation, and Response*, no. 10341 in Dagstuhl Seminar Proceedings. Dagstuhl, 2010.

[15] Sean Peisert, Matt Bishop, Sidney Karin, and Keith Marzullo. "Analysis of Computer Intrusions Using Sequences of Function Calls." *IEEE Transactions on Dependable and Secure Computing (TDSC)*, 4(2):137-150, April-June 2007.

[16] Sean Peisert, Matt Bishop, Sidney Karin, and Keith Marzullo. "Toward Models for Forensic Analysis." In *Proceedings of the Second International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE)*, pages 3-15, Seattle, WA, April 2007.

[17] Sean Peisert, Matt Bishop, and Keith Marzullo. "Computer Forensics In Forensis." *ACM Operating Systems Review (OSR) Special Issue on Computer Forensics*, 42(3):112-122, April 2008.

[18] Sean Peisert, Matt Bishop, and Keith Marzullo. "Computer Forensics In Forensis." In *Proceedings of the Third International IEEE Workshop on Systematic Approaches to Digital Forensic Engineering (IEEE-SADFE)*, pages 102-122, Oakland, CA, May 22, 2008.

[19] Sean Peisert, Matt Bishop, and Alec Yasinsac. "Vote Selling, Voter Anonymity, and Forensic Logging of Electronic Voting Machines." In *Proceedings of the 42nd Hawaii International Conference on System Sciences (HICSS), Digital Forensics, Pedagogy and Foundational Research Activity Minitrack*, Waikoloa, HI, Jan. 5-8, 2009.

[20] Sean Peisert, Ed Talbot, and Matt Bishop. "Turtles All The Way Down: A Clean-Slate, Ground-Up, First-Principles Approach to Secure Systems." In *Proceedings of the 2012 New Security Paradigms Workshop (NSPW)* (to appear), Bertinoro, Italy, September 19{21, 2012.

[21] Sean Philip Peisert. *A Model of Forensic Analysis Using Goal-Oriented Logging.* Ph.D. thesis, Department of Computer Science and Engineering, University of California, San Diego, March 2007.

[22] Huong Phan, George Avrunin, Matt Bishop, Lori A. Clarke, and Leon J. Osterweil. "A Systematic Process-Model-Based Approach for Synthesizing Attacks and Evaluating Them." In *Proceedings of the 2012 Electronic Voting Technology Workshop/Workshop on Trustworthy Computing* (EVT/WOTE'12), Redmond, WA, August 6-7, 2012. USENIX Association.

[23] Borislava I. Simidchieva, Sophie J. Engle, Michael Cliord, Alicia Clay Jones, Sean Peisert, Matt Bishop, Lori A. Clarke, and Leon J. Osterweil. "Modeling Faults to Improve Election Process Robustness." In *Proceedings of the 2010 Electronic Voting Technology Workshop/Workshop on Trustworthy Computing* (EVT/WOTE '10), Washington, D.C., August 11-13, 2010. USENIX Association.

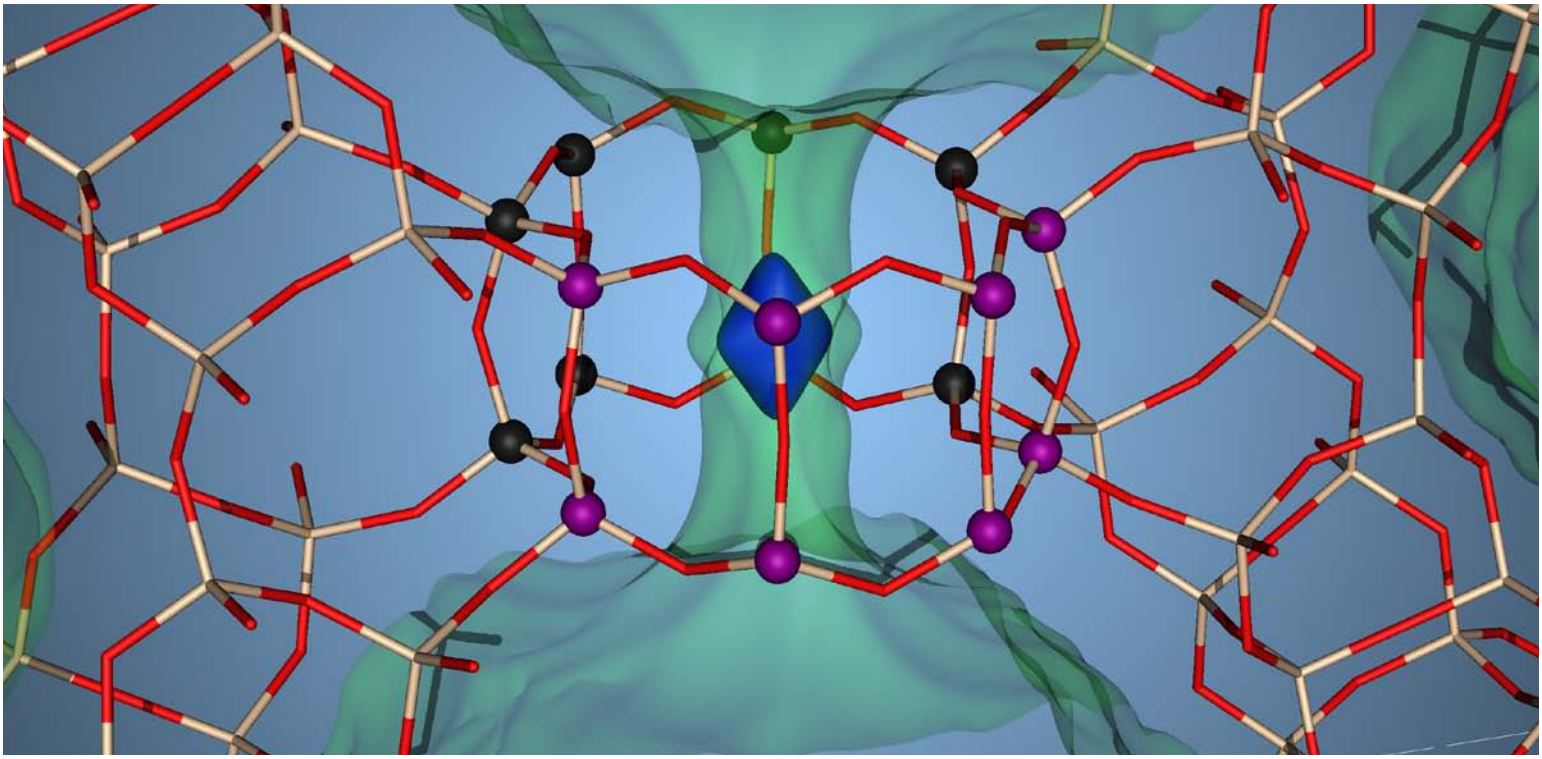
[24] Sean Whalen, Matt Bishop, and James Crutcheld. "Hidden Markov Models for Automated Protocol Learning." In *Proceedings of the 6th International ICST Conference on Security and Privacy in Communications Networks (SecureComm)*, Singapore, September 7-9, 2010.

[25] Sean Whalen, Sophie Engle, Sean Peisert, and Matt Bishop. "Network-Theoretic Classication of Parallel Computation Patterns." *International Journal of High Performance Computing Applications (IJHPCA)*, 26(2):159-169, May 2012.

[26] Sean Whalen, Sean Peisert, and Matt Bishop. "Network-Theoretic Classication of Parallel Computation Patterns." In *Proceedings of the First International Workshop on Characterizing Applications for Heterogeneous Exascale Systems (CACHES)*, Tucson, AZ, June 4, 2011.

[27] Sean Whalen, Sean Peisert, and Matt Bishop. "Multiclass Classification of Distributed Memory Parallel Computations." *Pattern Recognition Letters (PRL)*, (to appear) 2012.

[28] Sean Harrison Whalen. *Security Applications of the ε-Machine.* Ph.D. thesis, University of California, Davis, Dept. of Computer Science, September 2010.



SCIENTIFIC COMPUTING GROUP

The Scientific Computing Group's (SCG) ultimate goal is to advance scientific computing by developing and enhancing applications in key scientific disciplines, as well as developing tools and libraries for addressing general problems in computational science. Members of SCG work on diverse scientific applications in a wide variety of fields including atmospheric modeling, materials science and nanoscience, chemical science, and biological science. These scientific applications are relevant to the missions of LBNL and DOE.

OVERVIEW OF HIGHLIGHTED PROJECTS

The Scientific Computing Group's work is often in collaboration with other groups in CRD. This section includes information about their work with high performance linear solvers, and eigenvalue calculations in scientific applications. More information about the Scientific Computing Group's work can be found in "Collaborative Projects," under Berkeley **I**ce **S**heet **I**nitiative for **CL**imate **E**xtreme**S** (BISICLES) and Frameworks, Algorithms, and Scalable Technolgies for Mathematics (FASTMath).

RESEARCH STAFF

ESMOND G. NG received his Ph.D. in computer science from the University of Waterloo, Canada in 1983. He is currently a senior scientist and the Department Head of the Applied Mathematics and Scientific Computing Department and acting Group Lead for the Scientific Computing Group.

Prior to joining LBNL in 1999, Ng was a senior research staff member and a group leader in the Mathematical Sciences Section of the Computer Science and Mathematics Division at Oak Ridge National Laboratory (ORNL). While he was at ORNL, he was also an adjunct professor in the Department of Computer Science at the University of Tennessee, Knoxville. Ng is well known for his work on sparse matrix computation. The algorithms and codes that he has developed are widely used. His research also includes numerical linear algebra, graph algorithms and complexity analysis, parallel computing, and mathematical software development and software engineering. In addition, he is interested in applying sparse matrix techniques to scientific and engineering applications. Ng has served on the editorial boards for the SIAM Journal on Matrix Analysis and Applications and the IEEE Transactions on Parallel and Distributed Systems. He is currently an associate editor of the SIAM Journal on Scientific Computing and also an associate editor of the Bulletin of Computational Mathematics.

ZHAOJUN BAI is a professor in the Department of Computer Science and the Department of Mathematics at UC Davis. He obtained his Ph.D. from Fudan University, China in 1988, and joined UC Davis in 1999 after nine years at the University of Kentucky. He has held a faculty computer scientist position at LBNL since 2003. His main research interests are in the areas of scientific computing, numerical linear algebra design and analysis and mathematical software engineering. He participated in a number of large-scale synergistic computational science and engineering projects, such as LAPACK for solving the most common problems in numerical linear algebra and DOE's SciDAC projects PETAMAT, on next generation multi-scale quantum simulation software on multi-core systems with GPU acceleration. His recent research projects include the theoretical and computational study of minimization principles of the linear response theory for computing excitation energies of molecular systems. He currently serves on multiple editorial boards of journals in his research field, including Journal of Computational Mathematics and ACM Transactions on Mathematical Software. Previously, he has served as an associate editor of SIAM Journal on Matrix Analysis and Applications, vice chair of the 26th IEEE International Parallel and Distributed Processing Symposium and numerous other professional positions. His current research interests are in computer architecture and energy efficient parallel computing. More specifically, he is interested in optimizing data access performance for parallel computing and in utilizing heterogeneous computing power.

ANDREW CANNING is a staff scientist in SCG and an adjunct professor in the Chemical Engineering and Materials Science Department at UC Davis. Canning obtained his Ph.D. in computational physics from Edinburgh University, with a specialization in disordered magnetic systems. After, he moved to Switzerland as a postdoc to work at the EPFL, where he focused on first-principles electronic structure methods based on the Car–Parrinello approach. In 1998 Canning joined LBNL after working in Roberto Car's group for three years. His research has focused on the development of new computational methods for first-principles calculations, particularly parallel algorithms for large-scale parallel computers. He has also worked in many application areas ranging from nano-structures to actinide materials and gamma ray detectors. He won the Gordon Bell prize in 1998 and was a Computerworld Smithsonian Laureate in 2001.

JAMES DEMMEL is a distinguished professor of computer science and mathematics at UC Berkeley. His personal research interests are in numerical linear algebra, high performance computing, and communication avoiding algorithms. He is known for his work on the LAPACK and ScaLAPACK linear algebra libraries. He is also a member of the National Academy of Sciences, National Academy of Engineering, a fellow of the ACM, IEEE and SIAM, and winner of the IEEE Computer Society Sidney Fernbach Award, the SIAM J. H. Wilkinson Prize in Numerical Analysis and Scientific Computing, and numerous best paper prizes. He was an invited speaker at the 2002 International Congress of Mathematicians and the 2003 International Congress on Industrial and Applied Mathematics.

MACIEJ HARANCZYK is a research scientist in SCG. Haranczyk received a Ph.D. and M.S. in chemistry from University of Gdansk, Poland. He spent his postdoc appointment as a 2008 Glenn T. Seaborg Fellow at LBNL. His research interests include the development of methods, tools and approaches to enable efficient molecular and materials discovery.

LEX KEMPER is a 2012 Luis W. Alvarez Fellow in Computing Sciences. He studied physics at University of Florida, Gainesville and received his Ph.D. from Stanford University where he worked on applying computational approaches to studying high-temperature superconductivity. Before joining LBNL, Kemper continued work at Stanford as a postdoc researcher. His main research interest is in pump-probe spectroscopy, which is a rapidly growing field with the capability of elucidating many aspects of complex materials physics. His project focuses on developing the theoretical tools and knowledge necessary to both understand the current experiments and to provide vision for future directions.

JOHN GILBERT was a primary architect and developer of Matlab's sparse matrix capability and of the SuperLU solver library. He received his Ph.D. in computer science from Stanford in 1981 and was a teacher, researcher, and manager at Cornell University and Xerox PARC. He has done fundamental work in sparse matrix algorithms and software. Gilbert now directs the Combinatorial Scientific Computing Laboratory at UC Santa Barbara, where his current research applies linear algebra and sparse matrix methods to discrete problems in the analysis of large graphs and networks. Gilbert has served in a number of advisory and consulting positions for industry and for government laboratories, and in various editorial and leadership capacities in the profession of applied mathematics and computational science. He is also a fellow of the Society for Industrial and Applied Mathematics.

NIELS GRONBECH-JENSEN has been a faculty scientist in SCG since 1999. He received his M.S. in applied mathematical physics in 1989 and his Ph.D. in physics in 1991, both from the Technical University of Denmark. While working at LBNL, he has also been a professor and researcher at Stanford University, Los Alamos National Laboratory, UCLA, UC Davis, the Niels Bohr Institute, University of Copenhagen, and finally the chair of the applied science program at UC Davis. Jensen's research interests are theoretical condensed matter and statistical physics, dynamical systems, molecular dynamics for hard and soft matter, self-assembly, electrostatic effects, ion irradiation and radiation damage, superconducting device physics, macroscopic quantum systems, and numerical analysis and algorithms. He is a fellow of the American Physical Society, and among other awards, Jensen received an R&D 100 award and was honorable mention for the Gordon Bell Prize in 1993 for the IEEE Computer Society.

XIAOYE SHERRY LI works on fast and robust sparse linear solvers and next-generation computing for X-ray science. She develops scalable sparse direct and hybrid linear solvers, and effective preconditioners, which can solve the most challenging linear systems that are often too difficult for iterative methods. She also developed new theory for computational methods, HPC code and user interfaces for computing X-ray scattering patterns with complex nano-structures for the ALS and NGLS.

RICHARD LUIS MARTIN is a postdoc research fellow in SCG. He received his Ph.D. in cheminformatics and his B.S. in computer science and mathematics from the University of Sheffield. His research interests include the development of computational methods for the representation, searching and screening of chemical and material data.

DÁITHÍ STONE is a research scientist in SCG. His research concerns the detection and attribution of climate change and its effects on natural and human systems. One particular focus is on extreme events, assessing the degree to which the chance of these events may have been altered because of anthropogenic emissions. His roles include leading the international C20C Detection and Attribution project and the Weather Risk Attribution Forecast exercise, as well as serving as lead or contributing author on parts of the Fourth and Fifth Assessment Reports of the Intergovernmental Panel on Climate Change. Stone studied in Canada and France, and has worked in the U.K. and South Africa before joining LBNL.

MICHAEL F. WEHNER is a staff scientist in SCG. His current research concerns the behavior of extreme weather events in a changing climate, especially heat waves, intense precipitation, drought and tropical cyclones. Before joining LBNL in 2002, Wehner was an analyst at the Lawrence Livermore National Laboratory in the Program for Climate Modeling Diagnosis and Intercomparison. He is the author or co-author of over 90 scientific papers. He was also a member of the lead author team for the 2009 White House report, "Global Climate Change Impacts in the United States" and is currently a lead author for both the Fifth Assessment Report of the Intergovernmental Panel on Climate Change and the upcoming Third US National Assessment on Climate Change. Wehner earned his master's degree and Ph.D. in nuclear engineering from the University of Wisconsin–Madison, and his bachelor's degree in physics from the University of Delaware.

EIGENVALUE CALCULATIONS IN SCIENTIFIC APPLICATIONS

Research Highlights

Eigenvalue calculations arise in a number of DOE applications, such as accelerator modeling, materials sciences and chemistry, and nuclear physics. In accelerator modeling, eigenvalue calculations are used to study cavity resonance so that cavity design can be optimized.

In materials sciences, chemistry and nuclear physics, eigenvalue calculations are used to approximate the solution of the many-body Schrödinger equation, which can in turn be used to study the electronic, chemical and optical properties of materials. Eigenvalue calculations also play an essential role in data analysis and classification.

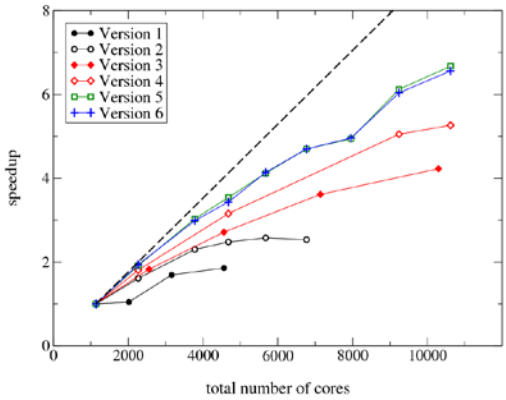


Figure 1: Speedup achieved by using a topology awared eigensolver that makes use of hybrid OpenMP/MPI on an nuclear structure eigenvalue problem (^{10}B, N_{max}= 8, M_j=1). Version 1 is the previous implementation. Version 6 contains all improvements made in 2011.

Developing efficient and scalable algorithms and implementations for eigenvalue calculations is a major component of the SciDAC project. SCG's eigenvalue calculation efforts date back to the TOPS project in SciDAC-1 and 2, and is currently funded within the FASTMath institute in SciDAC-3.

In addition to core algorithmic development and scalable implementation, SCG has also been working with application scientists on integrating eigensolvers with application codes. For example, they have integrated the Parallel ARnoldi PACKage (PARPACK) with Omega3P, a finite element frequency domain Maxwell equation solver for particle accelerators. SCG also integrated PARPACK with MFDn, a configuration interaction code for computing both the ground and excited states of light nuclei. They plan to integrate their research in SciDAC-3 with quantum chemistry codes such as Qchem, NWChem and materials science codes such as BerkeleyGW and CP2K. In almost all applications, the eigenvalue problems SCG solves are large and sparse or structured.

There are two main components in SCG's latest work in eigenvalue calculations. First, they are improving the performance of Lanczos-based eigensolvers for computing a few eigenpairs of symmetric sparse matrices on a many/multi-core system [1]. Their goal is to reduce the communication overhead as much as possible. SCG researchers are using two main techniques to achieve this. Their first technique is to reorganize the computational tasks and mapping them to processing units in a way that exploits the network topology. The second technique uses hybrid and overlapped OpenMP/MPI implementation to hide communication overhead. Figure 1 shows that by using these two techniques, SCG was able to speed up the previous implementation of the Lanczos method substantially for an application problem arising from nuclear structure. Second, SCG researchers are developing algorithms that can be used to compute a relatively large number of eigenpairs of symmetric sparse matrices (e.g., one percent of the eigenvalues of a million-dimensional matrix) using two approaches: (i) spectrum slicing, where the basic idea is to divide the desired part of the spectrum into several subintervals and compute eigenvalues within each interval in parallel; and (ii) penalized trace minimization. This last approach relies on BLAS3 operations and reduces the number of Rayleigh–Ritz calculations which often do not scale well on large-scale parallel computers.

In SciDAC-3, SCG along with other collaborators are involved in a number of parternship projects that require eigenvalue computation. One type of problem is the density functional theory problem that requires solving nonlinear eigenvalue problems in which the matrix depends nonlinearly on a subset of its eigenvectors. SCG developed

efficient preconditioning technique to accelerate a class of quasi-Newton methods for solving these type of problems. Another type of problem arises from using the configuration interacting technique to solve quantum many-body problems in computational chemistry. SCG researchers and collaborators are in the process of developing fast iterative methods and preconditioning techniques to tackle large problems for chemistry.

Scientific Impact

The development of a highly efficient eigensolver enables scientists at SLAC to simulate the electromagnetic field and resonance properties of new cavity designs for the International Linear Collider (ILC) in much greater detail.

Improving the performance of the eigensolver in the MFDn program enables nuclear physics collaborators to predict properties of several light nuclei. For example, they were recently able to predict properties of fluorine-14, an exotic short-lived nucleus for the first time. These properties were later confirmed by experiments performed at Texas A&M University's Cyclotron Institute.

The development of highly efficient nonlinear eigensolvers allows materials scientists to examine a number of microscopic properties of complex materials quickly, and enable them to design and screen new materials that are suitable for new energy applications.

CRD Contact

Chao Yang

SCG Collaborators

Hasan Metin Aktulga, Sherry Li, Lin Lin, Esmond Ng

LBNL Collaborators

Steven Louie (Materials Sciences Division/ Molecular Foundry), Jeff Neaton (Materials Sciences Division/Molecular Foundry), Martin Head-Gordon (Chemical Sciences Division/UCB), Wick Haxton (Nuclear Physics Division)

Other Collaborators

James Vary (Iowa State University), Pieter Maris (Iowa State University), Jim Cheliokowsky (University of Texas, Austin), Yousef Saad (University of Minnesota), John Pask (LLNL), Roberto Car (Princeton University), Weinan E (Princeton University), Christopher Cramer (University of Minnesota), Laura Gagliardi (University of Minnesota), Bert de Jong (PNNL)

References

[1] H. M. Aktulga, C. Yang, E. G. Ng, P. Maris, and J. P. Vary. "Topology-aware mappings for large-scale eigenvalue problems." C. Kaklamanis, T. Papatheodorou, and P. G. Spirakis, editors, *Proceedings of the Europar2012 Conference, Lecture Notes in Computer Science*, Berlin, Heidelberg, Springer 2012, pages 830-842.

HIGH PERFORMANCE LINEAR SOLVERS

Research Highlights

For high performance linear solvers, SCG's objective is to develop scalable sparse direct linear solvers and effective preconditioners that are critical in the simulations of numerically challenging problems. Their main goal is the development of three types of linear solvers: direct solvers, the Schur complement based hybrid direct and iterative solver, and near-optimal preconditioners using hierarchical matrix techniques.

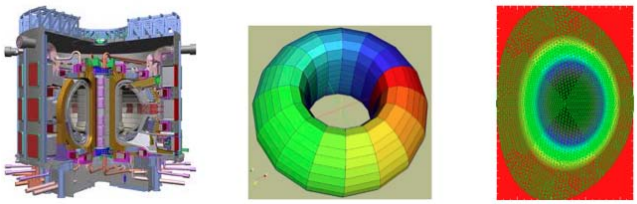


Figure 2: ITER Tokamak (left). Illustration of simulation code M3D-C1 high-order finite element discretization of the tokamak structure. Top view: 16–32 toroidal prisms (center). Slice view: ~104 nodes per plane (right).

Direct solvers are based on LU decomposition and a triangular solution. The main code base is SuperLU, which is widely adopted in academia and industry. SCG's new development focuses on enabling the effective use of many-core systems with heterogeneous node architectures, such as NUMA multi-core nodes. The significant accomplishments are summarized as follows. In recent years, the SuperLU solver suite has been downloaded on the order of 10,000 times annually. In particular, it received over 24,000 downloads in fiscal year 2012. SuperLU is used in numerous large-scale simulation codes in DOE and in many academic institutions. SuperLU is also used in a variety of commercial applications, including Walt Disney Feature Animation, airplane design (e.g., Boeing), oil industry (e.g., Chevron), and semiconductor industry (e.g., AMD). It has been adopted by many commercial mathematical libraries and software, including Cray's LibSci, FEMLAB, HP's MathLib, IMSL, NAG, OptimaNumerics, and Python (SciPy).

The second is the Schur complement based hybrid direct and iterative solver. SCG uses the non-overlapping domain decomposition framework, which blends direct solvers for the solution of the subdomain variables with preconditioned iterative solvers for the solution of the interface variables (i.e., Schur complement system). The main code base is called the Parallel Domain decomposition Schur complement based linear solver (PDSLin). A salient feature of PDSLin is to employ hierarchical parallelism to avoid increasingly ill-conditioning of the Schur complement system when increasing the number of processing cores. A significant accomplishment SCG achieved was to solve the 3D accelerator RF cavity design problem, involving highly indefinite linear systems of 52 million unknowns using 8,192 processing cores.

The third type of linear solver is near-optimal preconditioners using the hierarchical matrix techniques. SCG employs the truncated singular value decomposition (SVD) type of compact representation for the numerical low-rank submatrices arising from many discretized PDE problems. In particular, they have been using hierarchically semi-separable (HSS) matrices to represent such low-rank structures with hierarchical bases, which lead to near-linear time and low-memory factorization algorithms.

One significant accomplishment achieved by using this linear solver was the solution to the industrial-strength problem: 3D anisotropic Helmholtz

Figure 3: The accelerator hardware (left), a routinely solved RF cavity (center), and a 3-cryomodule RF Unit of the International Linear Collider Main Linac (right). The nonlinear eigenvalue problems arise when designing the cavities with waveguide coupling for multiple waveguide modes.

equations for seismic imaging with 6.4 billion unknowns using 4,096 processing cores. No other solvers were able to solve this type of problem with the scale and resolution.

Some important applications of SCG's linear solvers are in fusion physics and accelerator physics. The study of fusion physics applications is important in discovering how to harness the ultimate clean energy, fusion energy. Fusion energy does not contribute to global warming, the fuel is inexhaustible, there are no safety or weapons spin-off concerns, and there is no intrinsic long-term radioactive byproducts. Scientists at Princeton Plasma Physics Laboratory and LBNL have used computer modeling for a prototype fusion reactor, the International Thermonuclear Experimental Reactor (ITER), which is now being constructed in France. The US is one of seven nations participating in its funding and operation. ITER is based on the tokamak concept, which is a doughnut shaped vessel filled with a very high temperature gas (plasma) and extremely strong magnetic fields. In order to optimize the operation and output of ITER, scientists need to solve the equations that describe the magnetized plasma, just as aircraft designers solve the equations of fluid flow to optimize the design of an aircraft. These equations are very complex, and require specialized software. Several of the leading codes that are being used for predicting and optimizing the performance of ITER and other smaller experiments have used SuperLU, which is by far the most efficient software package for solving these equations. SuperLU makes calculations of unprecedented accuracy and efficiency that are of critical importance if scientists are to get the most out of the $10 billion ITER experiment. This also gives the US scientists somewhat of an edge over our international partners, and may also give us an advantage in designing and building the follow-on "DEMO" commercial fusion power plant.

Another example is in particle accelerator physics. Fourteen of the twenty-eight facilities in the DOE twenty-year outlook on Facilities for the Future of Science are particle accelerators. Scientists at SLAC National Accelerator Laboratory have developed an electromagnetics modeling code, Omega3P, to optimize cavity shapes for increased accelerating gradient and beam current. Here, they perform the frequency domain analysis using finite element discretization, which can provide high fidelity modeling for complex geometries of cavities. This requires finding the eigenmodes of electromagnetic cavities modeled by Maxwell's equations. The computational challenges come from the nonlinearity of the eigenvalue problems and the need for finding the interior eigenpairs. The most reliable approach to solving such eigenvalue problem is shift-invert Lanczos in which the demanding task is to solve the indefinite and ill-conditioned linear systems. SCG's linear solvers, SuperLU and PDSLin, have been used to solve these challenging equations reliably and efficiently. A significant accomplishment is the successful modeling of the 3D radio frequency (RF) cavity shape, which involves the solution of the highly indefinite linear systems of 52 million unknowns using 8,000 or more processing cores.

Scientific Impact

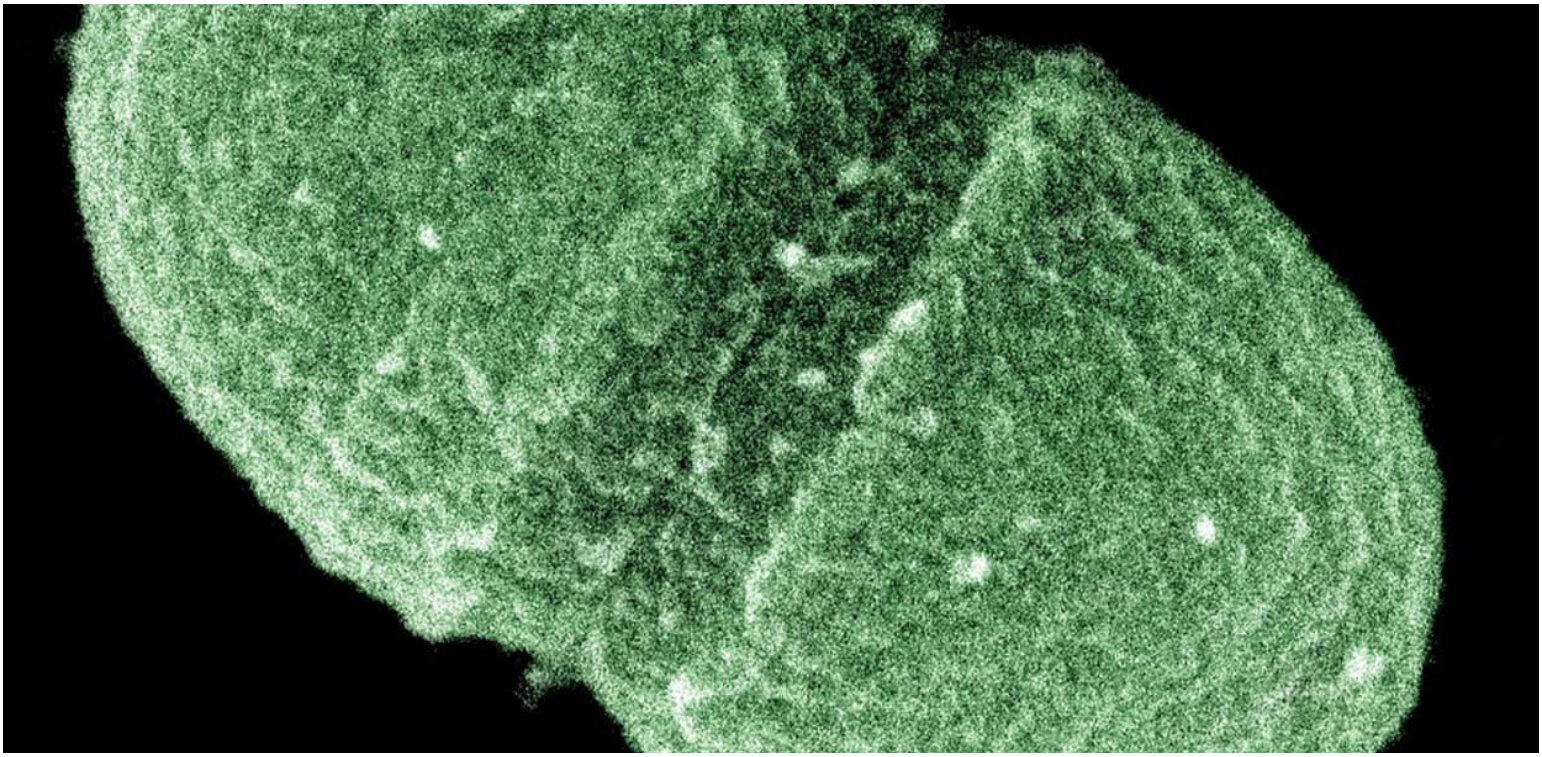
SCG's sparse solvers have been used in many scientific simulation codes worldwide. They provide fast and reliable solutions for very challenging linear systems arising from many scientific disciplines, such as accelerator, fusion, materials, quantum chemistry, and fluid mechanics.

CRD Contact

X. Sherry Li

Collaborators

James Demmel, John Gilbert, Laura Grigori, Meiyue Shao, Ichitaro Yamazaki



BIOLOGICAL DATA MANAGEMENT AND TECHNOLOGY CENTER

The Biological Data Management and Technology Center (BDMTC) is responsible for the development and maintenance of the Integrated Microbial Genomes (IMG) systems. The IMG systems enable scientists to study microbial genomes, single cell genomes, and microbial community metagenomes in the integrated context of an expanding universe of genome and metagenome datasets generated by sequencing centers worldwide. The IMG systems consist of a data warehouse, genome and metagenome annotation (data interpretation) pipelines, data integration utilities, and application specific analysis toolkits.

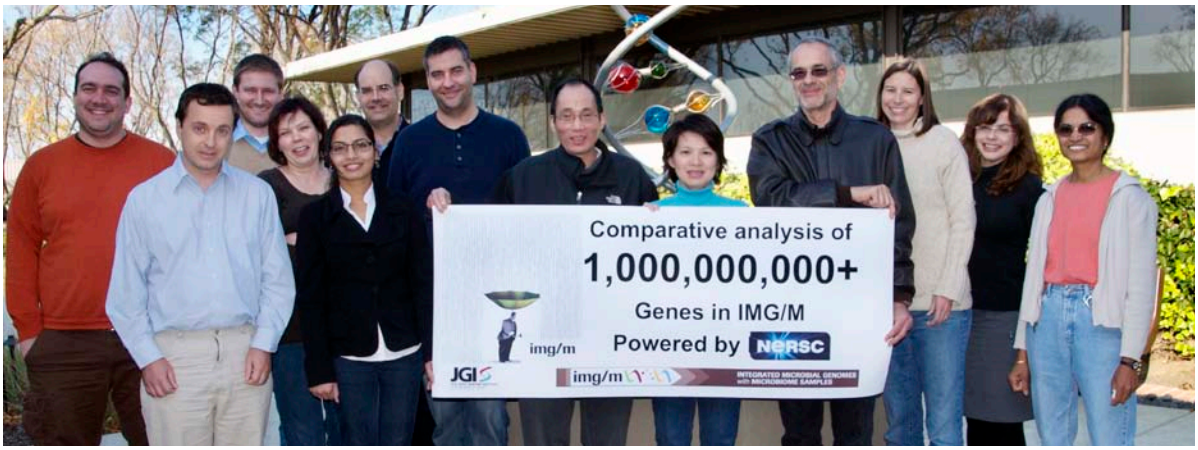
The Database Systems Group develops and maintains the data management infrastructure for the IMG systems; the tracking system for the microbial and metagenome annotation and integration; the functional annotation and data integration pipelines for isolate microbial genomes; and the data integration pipeline for metagenomes. In addition, this group is responsible for creating and maintaining the content for individual genome and metagenome portals for data distribution.

The Data Analysis Group develops, deploys, and maintains data exploration and comparative analysis tools and viewers that are part of the IMG user interfaces.

OVERVIEW OF HIGHLIGHTED PROJECTS

The Biological Data Management and Technology Center highlighted projects include information about the Integrated Microbial Genomes systems, focusing on data management and data analysis.

The IMG system has been developed in partnership with scientists of the Microbial Genome and Metagenome Program (MGMP) at the DOE Joint Genome Institute (JGI), in particular Nikos C. Kyrpides, head of MGMP, Natalia N. Ivanova, and Kostas Mavrommatis. Shane Cannon (NERSC) has contributed to extensions of the IMG data management infrastructure and data processing pipelines that addressed data content scalability challenges.

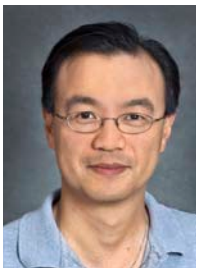


BDMTC DEPARTMENT HEAD

VICTOR M. MARKOWITZ is a Senior Scientist and the Department Head of the Biological Data Management and Technology Center (BDMTC). Markowitz joined LBNL in 1987 where he led the development of data management and integration tools used for developing public and commercial genome databases, such as the Genome Database at the Johns Hopkins School of Medicine in Baltimore, and the German Genome Resource Center's Primary Database in Berlin.

In 1997 he joined Gene Logic Inc. as CIO and Senior Vice President, Data Management Systems, responsible for the development and deployment of Gene Logic's data management products and software tools, including its flagship Genesis, one of the largest commercial gene expression data management deployed at tens of pharmaceutical and biotech companies worldwide. In 2004, Markowitz returned to LBNL's Computational Research Division to head BDMTC, where he oversaw the development of the DOE Joint Genome Institute's Integrated Microbial Genomes (IMG) family of systems used by a large community of scientists worldwide for conducting microbial genome and metagenome studies. Markowitz received his M.S. and D.S. degrees in computer science from Technion, the Israel Institute of Technology. He has conducted research in various data management and database areas, and has authored over sixty articles and book chapters on various aspects of databases and scientific data management. He has served on review panels and program committees for database and bioinformatics programs and conferences.

DATA ANALYSIS GROUP RESEARCH STAFF



KEN CHU is the Group Lead for the Data Analysis Group in BDMTC. He is responsible for the design, development, and deployment of the analysis tools and user interfaces for the IMG family of systems. He received a B.S. in computer science and electrical engineering from University of Toronto. Before joining BDMTC in 2006, Chu was a software engineer at Gene Logic where he developed key components of Gene Logic's flagship Genesis and Ascenta data management platforms for gene expression data, one of the most successful data management systems of their kind, deployed to tens of pharmaceutical and biotech companies.

JENNY HUANG is currently working on the analysis and visualization of large genomic and metagenomic datasets in the IMG family of systems. She received her Ph.D. in molecular and cellular biology, her M.S. in civil and environmental engineering from the University of Delaware, and her B.S. in biology from University of Science & Technology, China. Before joining BDMTC in 2010, Huang worked at Gene Logic on user interface components of Gene Logic's flagship Genesis data management platforms for gene expression data.

YINGJIE LIN is a software engineer in BDMTC who develops user interface components for the IMG family of systems. She received her Ph.D. in biomedical sciences and bioinformatics from Mount Sinai School of Medicine at New York University and a B.S. in biological sciences from Peking University. Before joining BDMTC in 2012, Lin worked on developing structural biology web applications.

BIJU JACOB joined BDMTC in 2009 as a software engineer responsible for the development of user interface components for the IMG family of systems. He received his M.S. in computer science from the University of Illinois, and another M.S. in mathematics from the University of Madras, India. He also received a B.S. in physics, chemistry, and mathematics from Bangalore University, India. Before joining BDMTC, Jacob worked as a software engineer at Elsevier, which publishes science and health information serving over 30 million scientists worldwide.

ANNA RATNER develops visualization tools for microbial and metagenome data in the IMG family of systems. She received her Ph.D. in immunology from UCLA. Before joining BDMTC in 2008, Ratner worked for seven years as a software engineer at Gene Logic where she designed and developed various components of Gene Logic's flagship Genesis and Ascenta data management platforms for gene expression data, one of the most successful data management systems of their kind, deployed to tens of pharmaceutical and biotech companies.

DATABASE SYSTEMS GROUP RESEARCH STAFF

I-MIN A. CHEN is the Group Lead of Database Systems Group in BDMTC. She received her Ph.D. in computer science from the University of Southern California. Her research focuses are on semantic and object-oriented data modeling, and biological data management. Chen currently works on improving the consistency and completeness of genomic data annotations in the IMG database.

KRISHNA PALANIAPPAN, as a member of the Database Systems Group, is responsible for the design, integration and maintenance of the Oracle data warehouse of the IMG system, which involves integration of genomic and metagenomic data, functional annotation data, and transcriptomics and proteomics data. She earned a Ph.D. in computer science from Nova Southeastern University, an M.S in computer science from Northern Illinois University, and a B.S. in biology from University of Madras, India. Her research focuses are on applying machine learning methods for predicting genomic features, genomic data modeling, and large-scale genomic data integration for comparative genome analysis. Palaniappan is also working on improving the current IMG data loading pipeline steps, Oracle query performance and tuning, as well as improving the consistency and completeness of genomic data annotations in the IMG database.

YURI GRECHKIN is responsible for enzyme and metabolic data loading in IMG. He received his M.S. in computer science from Moscow Technical University, and he joined LBNL in 2005. Grechkin helps with genome review, and genome and gene replacement in the IMG database and, in addition, he manages the data and metadata synchronization between the IMG database and Genomes Online Database (GOLD).

ERNEST SZETO is a bioinformatics software engineer in the Database Systems Group. He received his M.S. in bioinformatics from Polytechnic University of NYU, and his B.S. in civil engineering from Stanford University. Szeto joined LBNL in 1981, and has implemented many application tools for data management, functional analysis and prediction, and data loading. He currently focuses on issues of scaling up computation and data management for very large amounts of genomic and metagenomic data consisting of billions of genes.

JON BERTSCH is a bioinformatics engineer in the Database Systems Group. He received a Ph.D. in plant molecular and physiologic biology from UC Berkeley and a B.S. in biochemistry from Cambridge University, England. He is currently working in the IMG group on projects related to the IMG database pipeline for data loading and maintenance of the IMG biological databases. Before he joined BDMTC, Bertsch worked in biology research and web applications.

MANOJ PILLAY received his M.S. in computer and information sciences from the University of Delaware. Before joining BDMTC in 2010, Pillay worked at the Delaware Biotechnology Institute, University of Delaware, where he was a member of the production bioinformatics team responsible for acquiring, processing and integrating public genomic data with expression data. He is currently responsible for integrating metagenome datasets into the Integrated Microbial Genomes (IMG) data warehouse and for the distribution of individual genome and metagenome datasets to IMG users.

MICROBIAL GENOME AND METAGENOME DATA MANAGEMENT WITH THE IMG SYSTEM

System Highlights

CRD's Biological Data Management and Technology Center (BDMTC) developed and currently maintains the Integrated Microbial Genomes (IMG) system.[1] The IMG system provides the integrated data management environment required for studying microbial genomes and microbial community metagenomes in the context of an expanding universe of genome and metagenome datasets generated by sequencing centers worldwide.

At the core of the IMG system is a *data warehouse* that contains genome and metagenome datasets provided by scientific users, as well as public bacterial, archaeal, eukaryotic, and viral genomes from the US National Center for Biotechnology Information genomic archive and a rich set of engineered, environmental and host associated metagenomes. Genomes and metagenome samples are processed using IMG's microbial genome and metagenome sequence data processing pipelines. The genomes and metagenomes are then integrated into the data warehouse using IMG's data integration pipelines. IMG data integration involves metadata characterization, as illustrated in Figure 1.

Microbial genome and metagenome analysis is conducted in terms of genes, genomes and metagenome samples. From its first release in

[1] http://img.jgi.doe.gov/

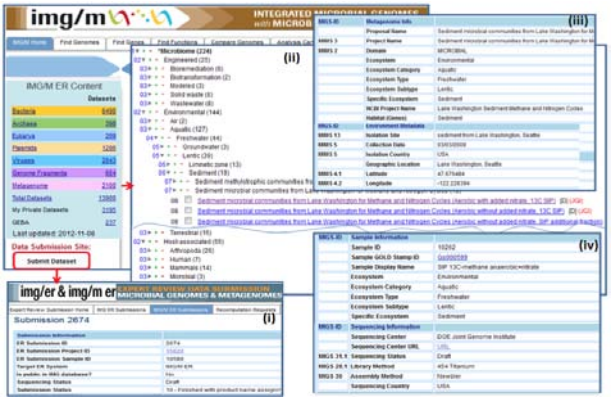


Figure 1: Figure 1: IMG data submission and metadata characterization. (i) Genome and metagenome datasets are provided by scientists for processing and inclusion into IMG via IMG's data submission site. (ii) Datasets in IMG/M are organized using a hierarchical classification that is critical for conductive comparative analysis. (iii) Datasets are associated with metadata, such as environmental data, and (iv) sample and sequencing information.

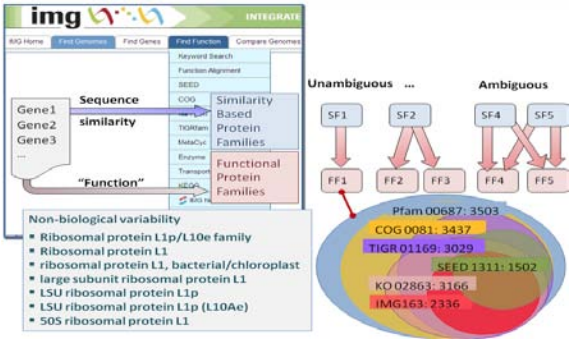


Figure 2: Reviewing functional annotation consistency in IMG. Functional annotations from different sources in IMG provide support for identifying annotation discrepancies. This figure shows the relationship between gene sets annotated with the same functional role by different protein classifications and annotation resources, namely COG, Pfam, TIGRfam, SEED, and KEGG Orthology.

2005, IMG has grown from an initial content of about 300 genomes, with a total of 2 million genes, to 10,137 bacterial, archaeal, eukaryotic and viral genomes, and 2,100 metagenome samples, with about 8 billion genes (as of November 10, 2012). 2,431 genomes and 829 metagenome samples in IMG are "private," that is, they are owned by scientists who have password protected access to study them before their public release. About 30% of the genomes and 75% of the metagenome samples in IMG have been sequenced at DOE's Joint Genome Institute, with the rest sequenced at other centers worldwide.

The effective comparative analysis of microbial genomes and metagenomes requires a consistent and complete view of biological data. The consistency of the biological data regards the biological coherence of annotations, while the completeness of the data regards the extent and coverage of functional characterization for genomes and metagenomes. Microbial genomes and metagenomes are characterized in IMG using different functional annotation resources, thus providing a comprehensive framework for identifying and resolving annotation discrepancies, as illustrated in Figure 2. IMG also contains organism metadata that include experimentally observed phenotypes ranging from relatively simple features, such as the ability to produce certain enzymes, to more complex characteristics,

such as growth requirements or energy metabolism type. A rule-based system for predicting phenotypes in IMG provides a powerful mechanism for validating functional annotations, whereby the phenotypic traits of an organism are inferred based on the presence of certain metabolic reactions and pathways and compared to experimentally observed phenotypes.

IMG's maintenance activities are mainly of a production nature, including regular (bi-weekly) genome and metagenome content updates and tuning the underlying data management infrastructure consisting of the Oracle database management system and the open source Berkeley DB (database) system. IMG's infrastructure is continuously revised in order to cope with the rapid increase in the number and the size of the genome and metagenome datasets. The infrastructure also strives to improve query performance and accommodate new data types, such as new functional annotations or experimental results.

IMG's data processing and integration processes involve massively parallel data intensive computations and, therefore, require high performance computing capabilities. The iterative nature of these processes requires a high-throughput computing environment. LBNL's National Energy Research Scientific Computing Center (NERSC) provides the computing infrastructure employed by IMG's data processing and integration pipelines.

Scientific Impact

Figure 3: Distribution by country of 2,585 IMG registered users worldwide.

IMG provides support to scientists conducting microbial genome and metagenome studies[1] and, as of November 15th, 2012, had 2,585 registered users worldwide (Figure 3).

The development of IMG's large-scale data processing, integration and analysis capabilities has required computer science, software engineering, and bioinformatics expertise. Running IMG's data processing and integration pipelines on a regular basis requires a high performance and high-throughput computing infrastructure. Therefore, IMG has become a unique resource for scientists conducting microbial genome and metagenome analysis worldwide.

[1] A list of publications enabled by IMG is available at: https://img.jgi.doe.gov/publication.

CRD Contacts

Victor M. Markowitz, I-Min A. Chen, Ken Chu

Collaborators

Nikos C. Kyrpides (MGMP/JGI), Natalia N. Ivanova (JGI), Kostas Mavrommatis (JGI), Shane Cannon (NERSC)

MICROBIAL GENOME AND METAGENOME DATA ANALYSIS WITH THE IMG SYSTEM

System Highlights

CRD's Biological Data Management and Technology Center (BDMTC) has developed and maintains the Integrated Microbial Genomes (IMG) system. IMG's data warehouse provides the integrated environment required for studying microbial genomes and microbial community metagenomes. Microbial genome and metagenome specific user interfaces provide access to different subsets of IMG's data and analysis toolkits: IMG[1] and IMG/M,[2] which support the analysis of publicly available isolate genomes (IMG/W) and metagenomes (IMG/M) in the context of all public IMG datasets; IMG ER[3] and IMG/M-ER[4] which provide registered IMG users with tools for expert review analysis and the curation of their password protected genomes (IMG ER) and metagenomes (IMG/M ER) in the context of all public IMG datasets.

IMG genome and metagenome data analysis consists of operations involving genomes and/or metagenome datasets, genes, and functions, which can be selected, explored individually, as illustrated in Figure 4, and compared, as illustrated in Figure 5.

[1] https://img.jgi.doe.gov/w
[2] https://img.jgi.doe.gov/m
[3] https://img.jgi.doe.gov/er
[4] https://img.jgi.doe.gov/mer

The examples in Figures 4 and 5 illustrate several key characteristics of conducting analyses with IMG. First, genome/metagenome analysis is an iterative process that involves a sequence (composition) of data exploration and comparative analysis operations. The composition of operations is facilitated by *carts* used for transferring results consisting of sets of objects (genomes, scaffolds, genes, functions) of interest between individual operations.

Second, analysis results are presented in multiple formats, including domain specific viewers, in order to facilitate the exploration of large results. Some domain specific viewers are available as open source tools, but need to be adapted to and integrated seamlessly into the IMG data environment.

Third, metagenome dataset comparisons need to take into account the stochastic nature of such datasets and test whether differences can be ascribed to chance variation or not, with results including an assessment of statistical significance, as illustrated in Figure 5.

Finally, the iterative nature of the IMG analysis process entails rapid response time (high performance) for individual operations. As the

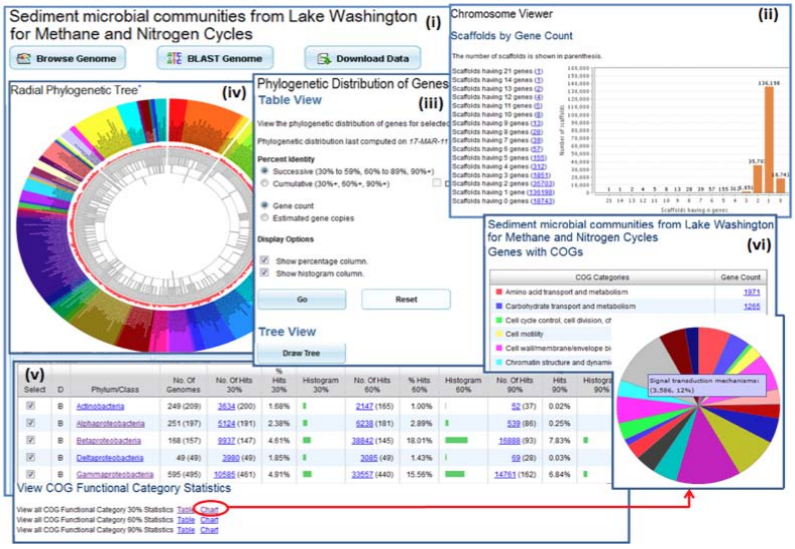


Figure 4: IMG data exploration. (i) Datasets can be examined using browsing, searching, or data sharing tools. (ii) "Scaffold Cart" allows the user to select scaffolds based on properties such as gene content. (iii) "Phylogenetic Distribution of Genes" provides an estimate of the phylogenetic composition of a genome or metagenome sample, with the result displayed using (iv) the "Radial Phylogenetic Tree" viewer, or (v) a tabular format consisting of a histogram. The histogram shows the quantity of protein-coding genes in the sample. It species which genes have the best similarity matches to proteins of isolate genomes in each phylum or class with different percent identities. (vi) The organization of genes by their assignment to COGs can be displayed in a pie chart format.

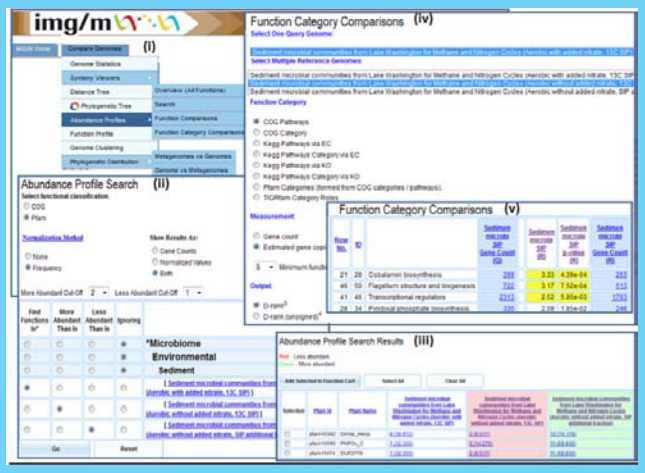


Figure 5: IMG comparative analysis. "Abundance Profile Search" allows users to find protein families (e.g., Pfams) in metagenomes and genomes based on their relative abundance, such as (ii) finding all protein families (e.g., Pfams) in a sample that are at least twice as abundant as in a second sample and are at least twice less abundant than in a third sample. (iii) The "Abundance Profile Search Results" consist of a list of protein families that satisfy the search criteria together with the metagenomes or genomes involved in the comparison and their associated raw or normalized gene counts. (iv) The "Function Category Comparison" tool allows users to compare a metagenome dataset with other metagenome or reference genome datasets in terms of the relative abundance of functional categories (e.g., COG Pathway). (v) The result of this comparison lists for each function category, F, the number of genes and estimated gene copies in the target (query) metagenome associated with F, and for each reference genome/metagenome the number of genes or estimated gene copies associated with F, as well as an assessment of statistical significance in terms of associated p-value and d-ran.

number and size of new genome and metagenome datasets included into IMG increases rapidly, the organization of these datasets needs to be continuously revised in order to maintain the performance of analysis operations.

IMG analysis tools are effective in revealing gaps in microbial genome and metagenome annotations, such as genes with missing or inconsistent protein product names, missing enzymes in the context of biological pathways, and genes missed by automated gene prediction. The Expert Review (ER) versions of IMG user interfaces provide the curation tools needed to address such annotation gaps, whereby these tools are coupled seamlessly with the analysis tools used for reviewing annotations.

Scientific Impact

The development of IMG's data analysis tools is driven by and evaluated within the context of microbial genome and metagenome studies, whereby these tools substantially help reduce the time required for typical analyses. IMG provides support to a large community of scientists worldwide[1] conducting microbial genome and metagenome studies.[2]

The IMG analysis tools are continuously improved in terms of their analytical capabilities, usability and performance. Since 2007, hands-on IMG training workshops have been attended by 615 scientists from 53 countries and have served as a source of feedback on the effectiveness of IMG's tools.

An education-specific version of IMG, IMG EDU, has provided support for teaching over 210 microbial genome and metagenome analysis courses to over 4,500 students at tens of colleges across the US.

CRD Contacts

Victor M. Markowitz, I-Min A. Chen, Ken Chu

Collaborators

Nikos C. Kyrpides (Head of MGMP, JGI), Natalia N. Ivanova (JGI), Kostas Mavrommatis (JGI)

[1] As of November 15th, 2012, IMG had 2,585 registered users in 63 countries across 6 continents.

[2] A list of publications enabled by IMG is available at: https://img.jgi.doe.gov/publication.

FUTURE TECHNOLOGIES GROUP

The Future Technologies Group (FTG) is a research group committed to the design and development of hardware and software systems that allow scientists to more effectively use extreme-scale computers. The members of this group work to create and improve computer architectures (both software and hardware), languages, operating systems, and conduct performance evaluations for all aspects of computers and applications. The group seeks to push computer science to new limits in order to allow scientists to analyze and solve the most challenging problems.

OVERVIEW OF HIGHLIGHTED PROJECTS

Members of FTG work closely with different application scientists (e.g., scientists in climate modeling, astrophysics, fusion simulation, life sciences, and nanoscience), along with faculty and students from the Electrical Engineering and Computer Sciences Department at UC Berkeley. They also work with the NERSC production computing facility. They have access to leading-edge production computing platforms, as well as hardware prototypes of experimental systems. FTG members publish many of their findings in top journals and conferences, and they have developed software systems that are widely used outside of the group.

The Future Technologies Group's research includes work on resiliency for large-scale computers, auto-tuning X-Stack, GASNet: Software Abstraction of Network Interconnect for PGAS Languages and Libraries, and Berkeley UPC (BUPC): Open-source Portable High-performance UPC. Their work is often in collaboration with multiple institutions, including MIT and the Princeton Plasma Physics Laboratory.

RESEARCH STAFF

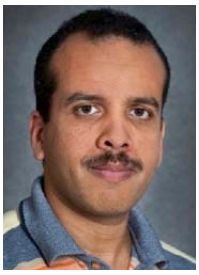
ERICH STROHMAIER is a Senior Scientist and Group Lead of FTG. His current research focuses on performance characterization, evaluation, modeling, and performance engineering. He is the Principal Investigator of the CACHE Institute, a joint mathematics and computer science institute focused on communication avoiding and communication hiding at extreme scales. He is also involved in a leading role in the DEGAS (Dynamic Exascale Global Address Space) programming environment project. Strohmaier is one of the founders of the TOP500 project.

DANIEL BURKE serves as project manager for the new Computer Architecture Lab to explore low-energy approaches for the DOE's Exascale Computing Initiative. Since 2006, he has collaborated on joint projects while at UC Berkeley, including RAMP and GreenFlash, which evolved into the co-design thrust of the Exascale Computing Initiative. Burke is primarily engaged in system design and project management of next generation reconfigurable platform systems in support of processor, system, and ASIC (application-specific integrated circuit) research. Burke is currently working on projects like investigations of high core-count, many-core systems and associated optimized software, and ultra high efficiency systems and implementation (algorithm design, operating system interaction, RTL emulation). He received his B.S. in physics and math from Texas A&M University, and his M.S. in computer engineering from the University of Illinois, with research design experience at Caltech, UIUC, and NCSA serving in roles ranging from design to project management.

COSTIN IANCU is currently working on programming models and code optimization for large-scale parallel systems, while focusing on simple and practical designs. Iancu graduated in 2001 with a Ph.D. in computer science from UC Santa Barbara. Over the years, he has been involved in many projects connected to LBNL, including the Berkeley UPC, FastOS, DEGAS, Corvette, and THOR.

AYDIN BULUÇ is a research scientist in FTG working on high performance graph analysis, libraries, and their applications in genomics and bioinformatics. His current research also includes parallel sparse matrix computations and communication-avoiding algorithms. Previously, he was a Luis W. Alvarez Postdoctoral Fellow. He received his Ph.D. in computer science from UC Santa Barbara in 2010 and his B.S. in computer science and engineering from Sabanci University, Turkey in 2005. As a graduate student, he spent a semester at the Math Department of MIT, and a summer at Sandia National Labs' CSRI institute, New Mexico. BuluÇ has architected and contributed to parallel graph libraries like the Combinatorial BLAS and the Knowledge Discovery Toolbox (KDT).

CY CHAN is a computational science postdoc research fellow in FTG, working on developing new techniques for software optimization and novel programming models for HPC systems. His current research interests include the analysis and optimization of multi-physics fluid simulation codes for hardware/software co-design and the development of new programming methodologies to automate this process on exascale systems. His recent publications have introduced and studied auto-tuners for sparse linear systems including multigrid solvers. In addition to these pursuits, Chan is also interested in developing collaborations in other areas of renewable energy applications research.



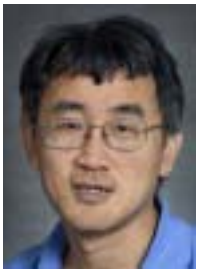
KHALED IBRAHIM is a research scientist in FTG. He joined LBNL in 2009, coming from INRIA, Bregtane, France. He obtained his Ph.D. in computer engineering from North Carolina State University in 2003. Some of his research interests include code optimization for high performance computing and the use of accelerators for cost/power effective supercomputing; virtualization and cloud computing environments for high performance computing; speculation in shared-memory multiprocessor system, architectural support for secure computing on multiprocessor systems; and finally, power-aware scheduling of embedded applications on multiprocessor system-on-chip.

PAUL HARGROVE has been a full-time staff member at LBNL since 2000. He received his Ph.D. from Stanford University in 2004. His current activities include the Berkeley UPC Compiler and Runtime, the GASNet communications library, and the Berkeley Lab Checkpoint Restart (BLCR). Hargrove's primary research interests are the systems software and middleware that enable current and next-generation high performance computing. Some of Hargrove's previous work at LBNL include two multi-institution efforts in the HPC middleware/systems software areas: Coordinated Infrastructure for Fault Tolerant Systems (CIFTS) and the SciDAC Scalable Systems Software (SSS) ISIC; and the Modular Virtual Interface Architecture (M-VIA) software.

LEONID OLIKER is a senior computer scientist in FTG. He received multiple bachelor's degrees in computer engineering and finance from the University of Pennsylvania, and performed both his Ph.D. and postdoc work at NASA Ames Research Center. Oliker has co-authored over 90 technical articles, five of which received best paper awards. His research interests include HPC evaluation, and multi-core auto-tuning.

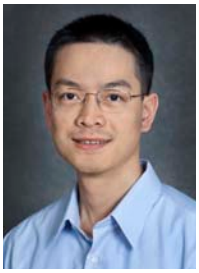
STEVEN HOFMEYR is a Computer Systems Engineer in FTG. He received a Ph.D. in computer science from the University of New Mexico in 1999. Previously, he was founder and Chief Scientist at a cybersecurity company, Sana Security, which was acquired by AVG in 2009. His research interests include information security modeling of complex systems, such as the Internet and financial markets; understanding the impact of policies and regulations for controlling cyberthreats on a large scale; future operating systems for many-core architectures; and scheduling and load balancing for parallel programming. He was named one of Infoworld's Top 12 Innovators of the Year in 2004, and he received the MIT Technology Review TR100 Innovators of the Year award in 2003 for advances in network security.

ERIC ROMAN is a Computer Systems Engineer in FTG. He received his Ph.D. in physics from UC Berkeley in 2010. His main research interests are operating systems for parallel computers, with a focus on system resiliency in exascale platforms. He collaborates with the Par Lab at UC Berkeley on the Tessellation Operating System, and he is also one of the developers for the Berkeley Lab's Checkpoint Restart (BLCR).



HONGZHANG SHAN received his Ph.D. in computer science from Princeton University in 2001. After working a year for Storage Networks, he joined FTG. His research focuses on parallel programming models, performance modeling, benchmarking, and optimization. He has published over 30 technical papers on international journals and conferences. One of his papers received the Gordon Bell Prize at SC08.

DIDEM UNAT is currently designing and evaluating programming models for future exascale architectures, as part of the Hardware Software co-design project. She is the 2012 recipient of the Luis W. Alvarez Fellowship. Her research interest is primarily in parallel programming models, compiler analysis and performance modeling.



YILI ZHENG has been a computer research scientist at LBNL since 2008. He works on multiple areas in parallel computing, from system software to applications. Zheng's most recent research projects include parallel programming system design and implementation. In the recently completed DARPA ECHELON project, he worked with collaborators at NVIDIA to design a new programming system, Phalanx, for hierarchical computer systems with heterogeneous processors. With colleagues at LBNL and UC Berkeley, he has been researching and developing compiler-free PGAS programming systems, such as the PGAS C++ template library and PyGAS, a Python extension for productive parallel programming. He is one of the key members in the Berkeley UPC Team. Zheng received his Ph.D. in electrical and computing engineering from Purdue University.

SAMUEL WILLIAMS is a research scientist in FTG. He received multiple bachelor's degrees in electrical engineering, applied mathematics, and physics from the Southern Methodist University. He received both his M.S. and Ph.D. from UC Berkeley. His research interests focus on performance modeling and performance optimization (nominally via auto-tuning) of HPC applications running on many-core architectures. He has co-authored 45 technical reports, book chapters, and papers, two of which he received best paper awards.

RESILIENCY FOR LARGE-SCALE COMPUTERS

Research Highlights

FTG's current resilience efforts address operating systems and resilience problems that are relevant to the DOE community. The resilience efforts in the DEGAS (Dynamic Exascale Global Address Space) project expand beyond system-level checkpoint/restart by integrating a new programming model (Containment Domains) for application-level resiliency, operating system (BLCR) and runtime (GASnet) support for fault tolerance.

The current resiliency research:

1. Supports coordinated checkpoints for GASNet and UPC.
2. Develops hybrid recovery schemes using Containment Domains and checkpointing (BLCR or application-based).
3. Reduces checkpoint time by reducing the I/O bottleneck when creating checkpoint files: improve process migration, node local or in-memory checkpoints, deduplication, checkpoint replication, and writeback.

Algorithm-based fault tolerance with memory exclusion reduces checkpoint overhead for further scalability.

BCLR

Berkeley Lab's Checkpoint/Restart for Linux (BLCR) is a Linux kernel module that provides checkpoint/restart for Linux processes. BLCR supports a wide variety of existing applications. Over 20 languages are tested regularly. In most cases applications can be checkpointed without requiring modification to the sources. FTG has designed BLCR in such a way that the changes required for checkpoint/restart support are limited to the parallel runtime. Few, if any, changes to application code are required to take advantage of BLCR, but BLCR does make checkpoint/restart functions available to applications that need them.

The Containment Domain (CD) model is a promising new approach to achieve resilience at extreme scales. A CD is a hierarchical (nesting) recovery container which "contains" the dependency of a calculation on other CDs. CDs isolate errors, in that an error in a CD requires recovery (rollback) of only that specific CD. In contrast, global checkpoint/restart approaches must always rollback an entire application to a known consistent state. Successful recovery of a domain is local to that domain, otherwise recovery escalates to the parent domain.

Scientific Impact

The BLCR effort began in 2001 as part of SciDAC Scalable Systems Software ISIC, and continued as part of the Coordinated and Improved Fault Tolerance for High Performance Computing Systems (CIFTS) project in 2006.

Early releases of BLCR began in 2002. Since 2002, BLCR has seen a wide adoption worldwide. BLCR is now part of the Fedora Project repository and Debian repository. BLCR support has been added to TORQUE and SLURM batch queuing systems by external collaborators.

In addition to support by MPICH2, MVAPICH, and Open MPI, BLCR support has been added by vendors, such as CRAY and SGI, to their MPI implementations and job launching runtime systems. BLCR runs on x86 and x86–64 systems (with or without the Xen hypervisor), PPC 32-bit and 64-bit systems (for future deployment on IBM Blue Gene systems) and ARM (for low-power and embedded environments). BLCR is used by the research community; Google Scholar reports 90 citations of the 2003 paper describing BLCR. FTG's collaboration with North Carolina State University has resulted in three new capabilities for BLCR: in-place rollback, incremental checkpoints, and live migration.

CRD Contacts

Eric Roman, Paul Hargrove

Collaborators

Mattan Erez (UT), Frank Mueller (NCSU)

BERKELEY UNIFIED PARALLEL C (UPC) LANGUAGE AND GLOBAL ADDRESS SPACE NETWORKING (GASNET)

Research Highlights

The goal of the Berkeley Unified Parallel C (BUPC) team is to further the adoption of the Partitioned Global Address Space (PGAS) programming model, and the team members perform multi-disciplinary research in application development, compiler and runtime technologies. The BUPC project has implemented a portable, high-performance, production-quality, open source compiler and runtime for the Unified Parallel C (UPC) language. The related Global Address Space Networking (GASNet) project provides the network communications runtime for BUPC, and for numerous other PGAS languages and libraries, as shown in Figure 2.

PGAS programming models provide a global data view for increased productivity, and one-sided communication semantics matching modern RDMA-capable hardware for increased performance. Throughout the years, the team has demonstrated the performance and productivity advantages of UPC over the dominant message-passing model of MPI, especially for applications with irregular communication patterns. They have demonstrated that communication overlap is a feasible technique for latency hiding; and they have addressed bisection bandwidth limitations on large-scale systems. The BUPC team has also shown that the performance portability of codes across systems is achievable using a combination of compile time and runtime dynamic analysis and code generation guided by performance models (Figure 1).

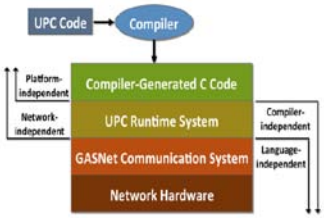


Figure 1: Structure of the Berkeley UPC compiler, including the GASNet communications runtime library.

The Berkeley UPC and GASNet projects are joint efforts between LBNL and UC Berkeley. The team members have contributed significantly to the UPC language specification, and to the general research in the field of PGAS programming models. These contributions would not have been possible without BUPC and GASNet as vehicles for prototyping and testing new ideas.

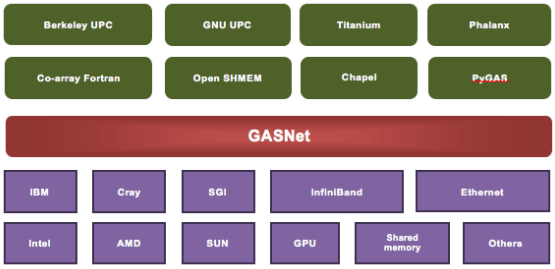


Figure 2: GASNet stands for "Global Address Space Networking," a library that enables one-sided communication for many parallel programming systems on diverse computer architectures.

Scientific Impact

Berkeley UPC was the first widely available UPC compiler, and it is one of the most widely deployed compilers to date, due to its portability. Berkeley UPC is installed and maintained on all of NERSC's production HPC systems, and on other DOE leadership class systems. This wide availability of the UPC language is one key to its adoption both by application authors and by educators of future computational scientists. GASNet is the first open-source portable and high-performance runtime library for Active Messages and one-sided communication, and it is the *de facto* standard in its field. As shown in the upper half of Figure 2, GASNet is used in many PGAS programming languages and libraries in production systems and in research projects. At LBNL, this includes the BUPC compiler plus a number of research projects, such as ECHELON and PyGAS. GASNet has enabled or accelerated deployment of many PGAS programming systems by significantly reducing their software development time, while delivering excellent performance across a wide range of systems from laptops to supercomputers.

CRD Contacts

Paul Hargrove, Costin Iancu, Katherine Yelick, Yili Zheng

Collaborators

Numerous users of BUPC and GASNet contributed to its development and testing.

X-STACK: AUTO-TUNING FOR PERFORMANCE AND PORTABILITY

Research Highlights

In order for scientists to meet exascale computing's energy efficiency goals, scientists will inevitably have to focus on the explicit management of hardware functionality through software. The energy efficiency goals of exascale computing will inevitably necessitate less virtualization of functionality in hardware and more explicit management through software.

When coupled with the increasing diversity of hardware (commodity CPUs, GPUs, many-core coprocessors, etc.), automated performance tuning of applications (auto-tuning) becomes essential. FTG's X-Stack efforts are designed to attack this problem for different application arenas using dramatically different approaches to auto-tuning. The first, which is now in its third and final year, is focused on the data locality (scatter/gather) and data synchronization challenges endemic to particle-in-cell (PIC) codes, sparse linear algebra, and graph analytics via auto-tuned libraries, runtimes, and DSLs.

The second project started in October and it focuses on improving performance through the application of data movement minimizing techniques to Adaptive Mesh Refinement Multigrid (AMR MG) applications and tensor contractions via source-to-source compilation techniques.

In the first X-Stack project, FTG members and collaborators focused on three major application drivers: particle-in-cell (PIC) codes, graph analytics, and sparse linear algebra. PIC codes represent an algorithmically superior O(N) approach to particle simulation. They are most often found in fusion simulation codes like the Gyrokinetic Toroidal Code (GTC), which is used to simulate the plasmas in Tokamak fusion reactors. Unfortunately, the code's performance often suffers on modern CPU and GPU architectures due to profound data locality and data synchronization challenges arising in their charge deposition (scatter-add) and push (gather) phases. In collaboration with Princeton Plasma Physics Laboratory (PPPL) and Penn State, FTG developed a number of techniques including static and dynamic buffering/bucketing, and highly-optimized floating-point atomics to improve GTC's performance by as much 2X on CPUs [4,5]. They also delivered the first highly optimized implementation of GTC running on NVIDIA GPUs. As current system power is often loosely tied to performance, there were proportionate improvements in energy efficiency (Figure 1).

Prior work showed that graph algorithms, like breadth first search, could be expressed in the language of linear algebra using an alternate semi-ring. Thus, FTG's work efforts are separated

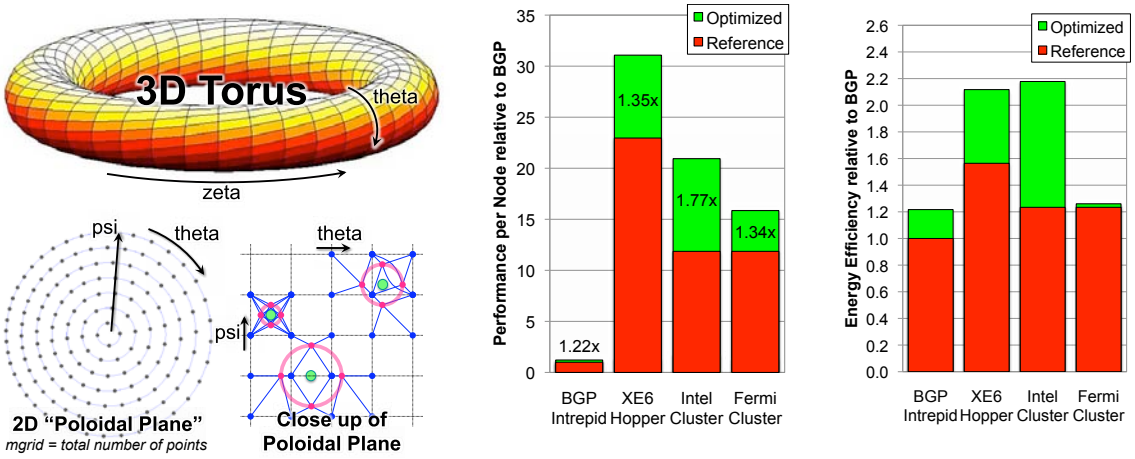


Figure 1: Conceptualization of the charge deposition phase within GTC; and performance and energy efficiency (normalized BlueGene/P = 1.0) before and after optimization. Baseline for the GPU is fastest Fortran version on the Intel Cluster. Speedups through optimization are labeled.

into two arenas: optimization of sparse linear algebra primitives within the Combinatorial BLAS (CombBLAS) [2,3]; and encapsulation of data structure and code optimization behind a python interface—the Knowledge Discovery Toolbox (KDT) [1]. More recently, FTG's efforts have focused on improving the performance and efficiency of filtered graph queries on large semantic graphs. As designed, such filters (i.e., all tweets sent on Monday) nominally require a call to the Python interpreter and yield poor performance. In collaboration with UCB, UCSB and FTG have leveraged the capability of Selective Embedded Just-In-Time Specialization (SEJITS) to JIT the filters on-the-fly. As such, the filters may now be performed at native hardware speeds and provide an order-of-magnitude performance boost to some graph algorithms.

Scientific Impact

FTG's technology for improving the performance of scatter-add operations has been demonstrated in a GTC benchmark on a variety of platforms. It is now being integrated into production versions of GTC and GTC-P (which incorporate more physics and a 2D domain decomposition) so that scientists may exploit the latest generation of supercomputers including both the BlueGene/Q "Mira" at ANL and the GPU-accelerated XK7 "Titan" at ORNL.

Using KDT and CombBLAS, application and big data scientists can exploit the parallelism of a large distributed-memory supercomputer for graph analytics using a straightforward sequential interface. Moreover, using the recently developed and integrated SEJITS aspects, domain scientists can guarantee very high performance and efficiency.

CRD Contacts
Samuel Williams, Aydín Buluç, Leonid Oliker, Khaled Ibrahim

Collaborators
John Gilbert (UCSB), Stephane Ethier (PPPL), Bei Wang (PPPL), Kamesh Madduri (Penn State), Armando Fox (UCB), Shoaib Kamil (MIT)

REFERENCES

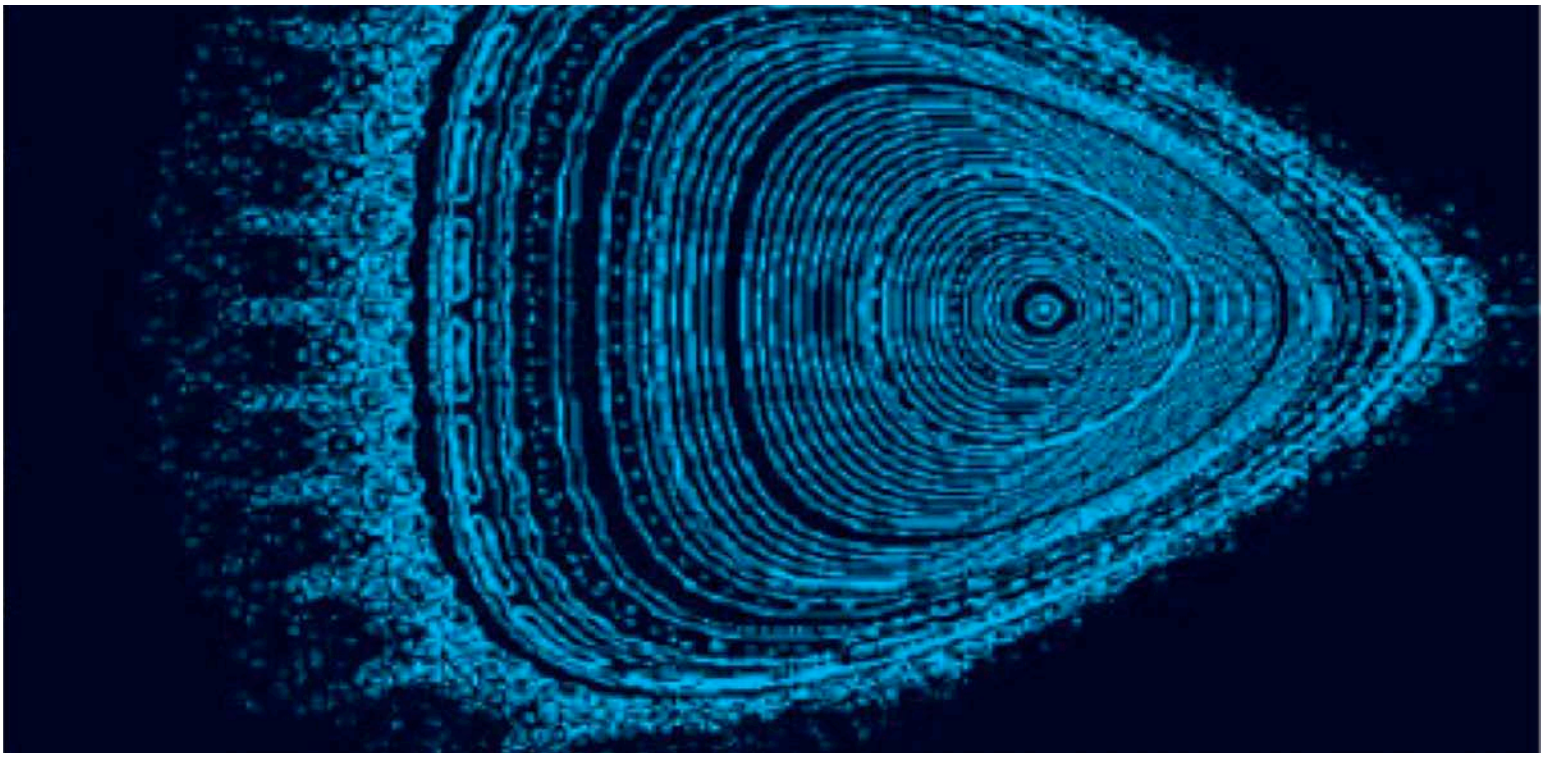
[1] Knowledge Discovery Toolbox Release v0.2, March 2012.

[2] A. Buluc and J. Gilbert. "The Combinatorial BLAS: Design, implementation, and applications." In *Intl. J. High Performance Computing Applications*, 2011.

[3] Combinatorial BLAS. release v1.2.1. April 2012. http://gauss.cs.ucsb.edu/~aydin/CombBLAS/html.

[4] K. Madduri, E. J. Im, K. Ibrahim, S. Williams, S. Ethier, and Leonid Oliker. "Gyrokinetic particle-in-cell optimization on emerging multi- and manycore platforms." *Parallel Computing*, 2011. http://dx.doi.org/10.1016/j.parco.2011.02.001.

[5] Kamesh Madduri, Khalid Ibrahim, Samuel Williams, Eun-Jin Im, Stephane Ethier, John Shalf, and Leonid Oliker. "Gyrokinetic toroidal simulations on leading multi- and manycore hpc systems." In Supercomputing (SC), 2011.



SCIENTIFIC DATA MANAGEMENT GROUP

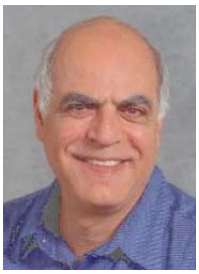
The mission of the Scientific Data Management (SDM) Group is to conduct research and development of data management technology for scientific applications, and to provide deployment and support of products developed by the group. These two types of activities complement each other. Research and development activities produce new concepts, algorithms, simulations, and prototype implementations, which are documented in published papers and other documents. Successful prototypes and technologies are then deployed in various application domains. Deployment and support includes the scaling of software products, making them robust, and applying them to emerging hardware architectures. In turn, when products are deployed, new challenges are identified for further research. Another aspect of the group's mission is to lead and support scientific data management and related activities for DOE. Such activities include leading the Scalable Data Management, Analysis, and Visualization (SDAV) Institute, and contributing to various workshops conducted by DOE for setting the agenda for future research.

OVERVIEW OF HIGHLIGHTED PROJECTS

The SDM Group works collaboratively with other groups in CRD on projects that require joint expertise. In particular, SDM has collaborated on several projects with the Visualization Group, like the extreme-climate data project and the trillon-particle simulation, and with other groups in the Advanced Computing for Science Department.

This section includes information about their work on applying bitmap indexing to genomic data search and advanced performance modeling with combined passive and active monitoring. In the "Collaborative Projects" section, they are collaborators on the ASCEM project and the extreme-climate data project.

RESEARCH STAFF

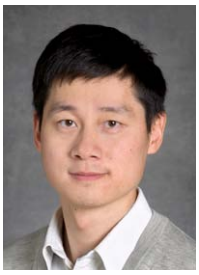


ARIE SHOSHANI is a Senior Staff Scientist in SDM. He joined LBNL in 1976, after he received his Ph.D. from Princeton University in 1969. He is the Group Lead of SDM and his current areas of work include data models, query languages, temporal data, statistical and scientific database management, storage management on tertiary storage, and grid storage middleware. Shoshani is also the director of the SDAV Institute under DOE's SciDAC program. He has published over 150 technical papers in refereed journals and conferences, and chaired several conferences in database management.

SURENDRA BYNA is a research scientist in SDM. Before joining LBNL, he was a researcher at NEC Labs America where he was a part of the Computer Systems Architecture Department and was involved in the Heterogeneous Cluster Computing project. He was also a research assistant professor in the Department of Computer Science at Illinois Institute of Technology (IIT), a guest researcher at the Math and Computer Science Division of the Argonne National Laboratory, and a faculty member of the Scalable Computing Software Laboratory at IIT. His current research interests are in computer architecture and energy efficient parallel computing. And more specifically, he is interested in optimizing data access performance for parallel computing and in utilizing heterogeneous computing power.

DORON ROTEM is currently a Senior Staff Scientist and member of SDM. Before joining LBNL, Rotem was a tenured associate professor in the Department of Computer Science, University of Waterloo, Canada. He also co-founded and served as CTO of a start-up company, called CommerceRoute, that made software products in the area of workflow and data integration. In CommerceRoute, he led a development team that built software to integrate heterogeneous databases and ERP systems, such as SAP and PeopleSoft. His research interests include energy efficient storage systems, scientific data management and parallel and distributed computing and algorithms. He has published over 80 papers in these areas.

ALEX SIM currently works on DOE projects that focus on network performance prediction and optimization, and applying SDM's tools on Storage Resource Management (BeStMan) to the Earth System Grid and Open Science Grid. He has also worked on the Particle Physics Data Grid, two Next Generation Internet projects and the HENP Data Grand Challenge project. Sim previously worked for an AI consulting firm where he developed expert systems involving object-oriented programming and artificial intelligence techniques for the Navy and other industries. He research interests are in resource management, data modeling, knowledge engineering, computer vision, and artificial intelligence.



JOHN WU is currently working on indexing technology for searching large datasets. He primarily focuses on improving bitmap index technology with compression, encoding and binning. He is the key developer of the FastBit bitmap indexing software, which has been used in a number of applications including high energy physics, combustion, network security, and query-driven visualization. Wu has also been working on a number of scientific computing projects including the development of the Thick-Restart Lanczos (TRLan) algorithm for solving eigenvalue problems and devising statistical tests for deterministic effects in broad-band timeseries. Wu received a Ph.D. in computer science from the University of Minnesota, an M.S. in physics from the University of Wisconsin-Milwaukee, and a B.S. in physics from Nanjing University, China.

JAESIK CHOI s is a postdoc fellow in SDM. He received his Ph.D. in computer science from University of Illinois at Urbana-Champaign in 2012, and his B.S. in computer engineering *magna cum laude* from Seoul National University in 2004. His research interests are in inference and learning in large-scale graphical models with an emphasis on intelligent hybrid (cyber-physical) systems. His technical work includes: scaling up probabilistic reasoning and learning in hybrid system models, learning large-scale dynamic systems, and a noble combination of motion planning and AI planning for robotic manipulations. He has published 20 papers in referred journals and conferences.

JUNMIN GU is a member of SDM. Currently, she is a key member and developer for the project on data transfers that utilizes policies for resource sharing. Gu has already developed several versions of the Storage Resource Management software for accessing and manipulating scientific data across the grid, which is currently deployed by many sites in the Open Science Grid and the Earth System Grid. Previously, she worked on projects for federating OLAP and Object Databases. Prior to LBNL, Gu worked for a consulting firm that develops AI solutions for government and industries. Gu received her master's degrees in mathematics and computer science from the University of Wisconsin–Madison. Her interests include data management, resource management, and distributed computing.

ALEX ROMOSAN has been a member of SDM since 2001. He received a B.S. in applied science from the University of Toronto and a Ph.D. in high energy physics from Columbia University. From 1997 to 2001, he was a postdoc research associate at LBNL working on the BaBar experiment at SLAC. Romosan has designed and implemented systems to address numerous aspects of data management requirements to support the analysis of large volumes of scientific data. He made contributions in the area of algorithm development, where he improved the scheduling of compute jobs on a cluster of servers, and he has been studying the effects of various caching and acceptance policies on the behavior of the Storage Resource Manager. He is currently working on the ASCEM project (data management of groundwater contamination wells database system), and the efficient genome properties search project.

ADVANCED PERFORMANCE MODELING WITH COMBINED PASSIVE AND ACTIVE MONITORING

Research Highlights

Network technologies have evolved rapidly in recent years. Advanced science networks, such as those managed by ESnet and Internet2 in the US, operate at speeds up to 100GB/s. Despite improvements in network technologies though, optimal selection of shared network resources and efficient scheduling of data transfers in a distributed and collaborative environment are challenging tasks in achieving superior performance in accessing data. In this project, SDM researchers study the network data traffic patterns, and build models to estimate the bandwidth utilization performance.

This data traffic modeling effort is in many ways similar to the modeling of highway vehicle traffic patterns and the estimation of travel time for travel planning. For example, it takes a vehicle roughly 1.5 hours to travel from Berkeley to the San Francisco Airport at 8:00 AM, or about 40 minutes around 1:00 PM. The network bandwidth utilization in scientific environments, such as high-speed networks, is more organized than web traffic, however, and cannot be directly modeled, estimated or predicted with existing network models based on active measurements. To understand the network traffic patterns, SDM developed a statistical performance estimation model based on historical measurements. This tool allows for the identification of recurrent patterns, and for the extraction of variations associated with normal and repeating events. The insight gained from such traffic patterns will be the basis for a predictive model of the network performance. The statistical information SDM is gathering can be used to estimate throughput performance of large data transfers for a given time window. It will even be possible to provide a confidence interval for the estimates. Such predictions will improve network usage and enable predictable data movement on high-bandwidth networks. The predictive estimation model and the framework will be helpful in optimizing the performance and utilization of terabit and future high-speed networks, as well as improving the utilization of shared resources with predictable performance for scientific collaborations, especially in data intensive applications.

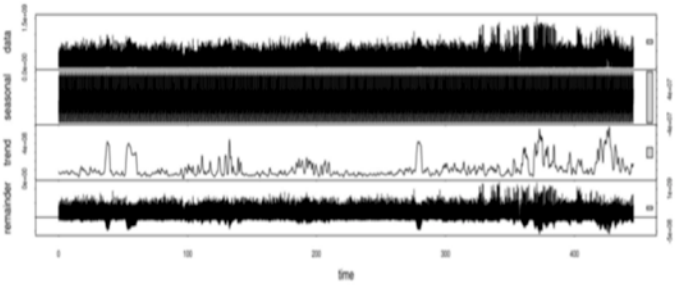


Figure 1: Statistical timeseries analysis from SNMP network utilization data at the NERSC gateway router shows regular patterns when examined over long periods of time (from 5/2011 to 5/2012). Timeseries seasonal adjustment plots are shown based on the periodicity of a day, consisting of four subplots of original data, seasonal component, trend component and irregular component from top to bottom.

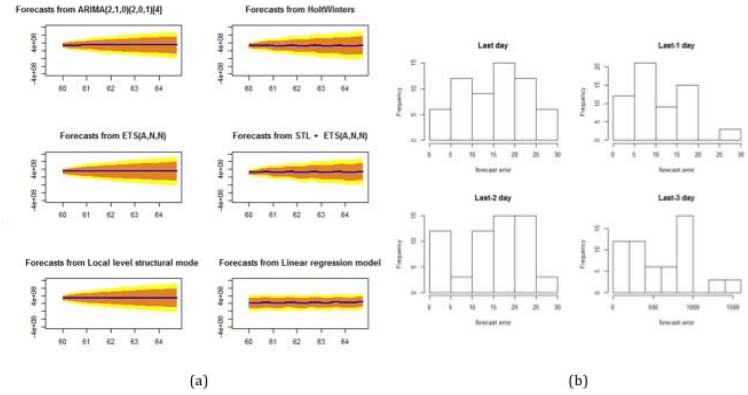


Figure 2: (a) Prediction comparison in different models and two confidence intervals (orange shades for 90% confidence interval, yellow shades for 95% confidence interval). (b) Prediction error based on the current statistical model with 24 hour periodicity. The forecast errors are less than 20% in most cases, and around 10% in some cases.

Scientific Impact

Scientists will benefit from the work on network traffic prediction because they can now utilize the resources offered by high-bandwidth network infrastructures more effectively. The network performance prediction model enables more efficient resource scheduling with accurate time estimates for data placement and access, as well as long-term capacity and traffic engineering planning of network infrastructures for large-scale data intensive applications.

CRD Contact

Alex Sim

Collaborators

Constantine Dovrolis (Georgia Institute of Technology), Greg Bell, Brian Tierney (ESnet)

FASTBIT SUPPORT FOR GENE CONTEXT ANALYSIS

Research Highlights

Genes located near each other, or in a chromosomal neighborhood, are often functionally related. These neighborhoods provide a gene context for the function of genes. Protein Cluster Context Analysis is a technique that allows for the examination of gene correlations based on their collocation on a chromosome and their functional annotations (such as associations with Clusters of Orthologous Groups, called COGs, or protein families, called Pfams). This technique first identifies "cassettes," which are defined as a tuple of adjacent genes that are separated by less than 300 base-pairs, as shown in Figure 1. Thus, every genome has multiple cassettes and each cassette has a set of associated functional annotations (or properties). Currently, the number of such functional properties is about 20,000.

A typical search for functionally related cassettes requires millions of comparisons—a complex and inefficient task with current data management tools. The Integrated Microbial Genome (IMG) system provides analysis tools that support such searches. In order to support interactive genome analysis, all possible matches are pre-computed and stored in IMG's data warehouse. As the number of genomes and genes in IMG grows, the time required for pre-computations (over a week), the space required for storing their results in the data warehouse (hundreds of millions of rows), and the time required for executing queries underlying gene context analysis (tens of minutes) is becoming prohibitively large.

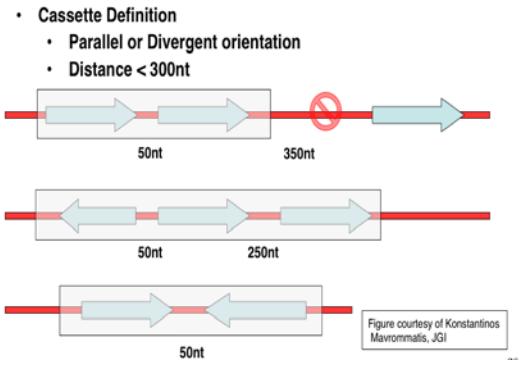


Figure 1: Visual definition of a cassette.

In order to overcome this problem, SDM members along with their collaborators have employed a bitmap representation of the functional properties associated with cassettes. By doing so, SDM is able to represent the data in a compact form, as shown in Figure 3. This representation of cassettes amounts to the *and* of bitmaps that represent the functional properties. However, the bitmaps can be very long and the computation inefficient. Thus, SDM members have used their FastBit indexing technique to compress the bitmaps. FastBit is a compressed indexing method that executes logical operations (such as *and* and *or*) without decompressing the bitmaps, and is therefore very efficient. However, they had to extend its functionality to allow for multiple columns to be *and*-ed (that is to support "multi-valued" attributes). Consequently, now such searches can be performed in seconds, with the underlying

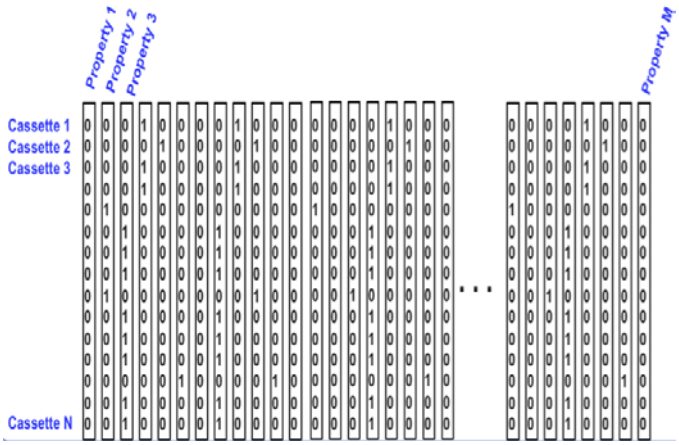


Figure 2: Bitmap representation of cassettes.

data representation of about a gigabyte of space residing in main memory. This solution eliminates the need for pre-computations since all searches can now be done in-place.

Searching for cassettes that share properties with other cassettes is only one type of query. Another, more challenging query, which is gene context analysis, involves a reference genome (that typically has multiple cassettes), and a set of selected genomes (each with multiple cassettes). This query finds cassettes in each of the genomes that have two or more properties in common. Using the bitmap representation in Figure 4, this requires 2^n searches over all the cassettes involved, which is not scalable. The solution was to invert the indexes so that they are represented as bitmap rows rather than columns, and then perform row comparison to find two or more matches. They also developed an algorithm that speeds up the search by removing duplicate patterns in intermediate steps. Consequently, this query can now be performed in seconds (linearly with the number of organisms) compared to many minutes required previously with pre-computed data. This represents an important breakthrough allowing the scaling of gene context analysis in the IMG system.

Scientific Impact

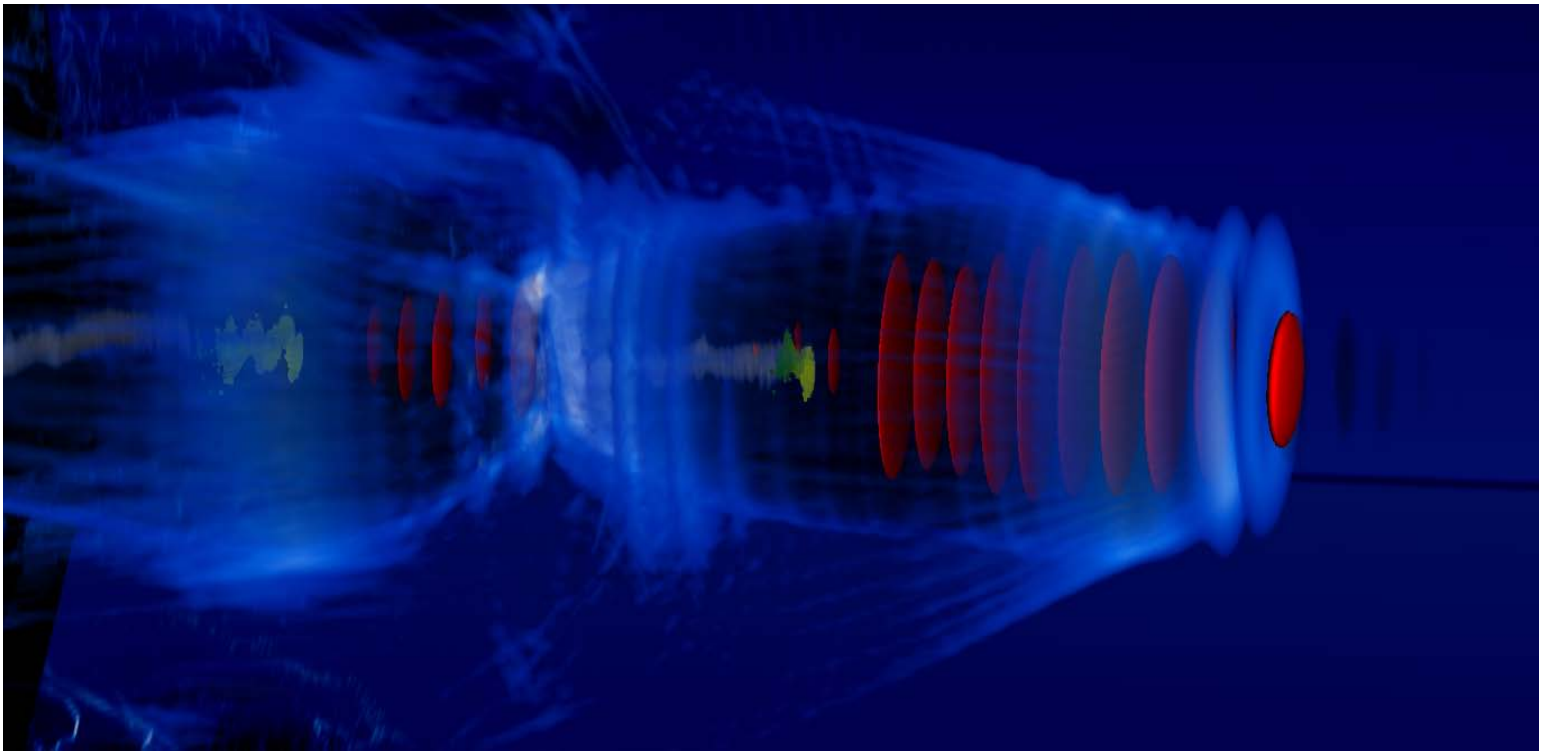
SDM's award-winning FastBit indexing (a 2008 R&D 100 award) has been used in multiple projects to accelerate scientific understanding and the discovery of patterns in real and simulation data. In this case, FastBit indexing allows biologists to analyze data in real-time in order to discover similarity patterns in genomic structures.

CRD Contact

Alex Romosan, John Wu, Arie Shoshani

Collaborators

Kostas Mavrommatis (Joint Genome Institute), Ken Chu, Victor Markowitz (BDMTC/CRD)



VISUALIZATION GROUP

The Visualization Group's mission is to assist researchers in achieving their scientific goals more quickly through visualization and analysis, while simultaneously advancing the state of the art through their own research. The group's objective is to develop new capabilities in high performance visualization and analysis. The development of these capabilities is driven by the needs of contemporary computational and experimental science projects central to the mission of DOE's Office of Science.

OVERVIEW OF HIGHLIGHTED PROJECTS

The Visualization Group collaborators are diverse, ranging from theoretical astrophysicists to computational biologists. The collaborators all have a common theme: the need to understand complex systems and phenomena through visual data exploration and analysis of scientific data.

The Visualization Group has many collaborations within CRD, as well as within LBNL. For example, they worked with the Scientific Data Management Group on the trillion particle simulation (included in this section), and also with the Scientific Computing Group on an extreme-scale climate application. Outside of CRD, the Visualization Group works with other divisions, like the Earth Sciences Division on a CO_2 sequestration project.

Team members of the Visualization Group are also part of the Advanced Simulation Capabilities for Environmental Management project, funded by DOE (see "Collaborative Projects").

RESEARCH STAFF

WES BETHEL , a senior computer scientist, is the Group Lead for the Visualization Group. He earned a B.S. and an M.S. in computer science from the University of Tulsa in 1984 and 1986, and a Ph.D. in computer science from UC Davis in 2010. Bethel and his team conduct applied research and development on techniques and tools for enabling scientific knowledge discovery in some of the world's largest collections of scientific data generated by computational models and simulations. This work lies at the nexus of data management, data analysis, scientific visualization, and high performance computing. He was the Coordinating Principal Investigator for the SciDAC Visualization and Analytics Center for Enabling Technology (VACET) which made production-quality, petascale-capable visualization a reality at DOE supercomputing facilities, and has produced software tools in use by a worldwide scientific community in disciplines ranging from high energy physics to climate modeling. His research interests include software architecture, high performance computing, scientific and information visualization, and computer graphics. He is an ACM Distinguished Scientist.

DAVID CAMP is a Computer Systems Engineer in the Visualization Group. He received his B.S., M.S., and Ph.D. in computer science from UC Davis in 1999, 2000, and 2012. Camp joined LBNL in 2010. His work focuses on increasing scientific productivity and understanding through increased parallel algorithm performance and data analysis using graphics techniques. Camp's research interests include computer graphics algorithms, visualization and analytics software architecture, distributed visualization algorithms, hybrid-parallelism designs, parallel algorithm techniques, and high performance computing.

HANK CHILDS has been a Computer Systems Engineer in the Visualization Group since June of 2009. Previously, he was at Lawrence Livermore National Laboratory from 1999 to 2009. He received a B.S. in 1999 and a Ph.D. in 2006, both in computer science and from UC Davis. Childs is the architect of the VisIt project, an end user visualization tool for very large data that has users throughout the world. It has been downloaded over two hundred thousand times, and won an R&D 100 award. Childs has also served as the Chief Software Architect of three different multi-institution, multi-million dollar efforts, including SciDAC VACET. Childs is currently a member of the executive council for the SciDAC Institute on Scalable Data Management, Analysis, and Visualization (SDAV). Finally, he received the DOE Early Career award in July of 2012 to research "Exploration at the Exascale."

MARK HOWISON holds a B.S. in mathematics from Brown University (2006) and an M.S. in computer science from UC Berkeley (2009). In his previous research, he has fabricated mathematical artwork, designed visualization tools, and investigated how people use technology to learn about mathematics and science. His high performance computing research has included profiling and optimizing parallel I/O, tuning image processing filters for multi-core processors and GPUs, and studying parallelization strategies for visualization algorithms. Howsion holds joint appointments with the Brown University Center for Computation and Visualization and CRD's Visualization Group.

HARINARAYAN KRISHNAN is a Computer Systems Engineer in the Visualization Group. Krishnan recently joined LBNL and has participated as a technical contributor in seven lab related projects. He is currently the Chief Software Architect for one of these projects. With a Ph.D. in computer science, Krishnan's primary focus is on scientific visualization, high performance computing, and software development.

BURLEN LORING is a Computer Systems Engineer in the Visualization Group. He received an M.S. in applied mathematics. Loring worked for more than eight years in professional software engineering, high performance computing, and scientific data visualization before joining LBNL. His areas of specialty include parallel computing, vector field analysis and visualization, and astrophysics and plasma physics visualization and analysis. He has extensive experience with key visualization software libraries and visualization tools, such as VTK and ParaView. His research interests include parallel scientific visualization algorithms, large data visualization, and the application of many-core technologies to scientific data visualization.

JÖRG MEYER is a Computer Systems Engineer in Visualization Group, where his research focuses on large-scale, parallel scientific data visualization and high performance computing support for visualization applications. Meyer is also a member of the NERSC Analytics Team, as well as a member of the Remote Data Analysis and Visualization (RDAV) support staff of the National Institute for Computational Sciences (NICS) at Oak Ridge National Laboratory. Previously, Meyer held faculty positions at UC Irvine, where he conducted his research on high performance computing and visualization techniques at the California Institute for Telecommunications and Information Technology (Calit2). He received his Ph.D. from the University of Kaiserslautern, Germany, in 1999, and he held an appointment as a postdoc researcher and lecturer in the Computer Science Department at UC Davis, from 1999 to 2000. He maintains an adjunct assistant professorship in the Computer Science and Engineering Department at Mississippi State University, where he was also affiliated with an NSF Engineering Research Center from 2000 to 2002. Meyer has led and served on various conference and program committees for multiple professional organizations, including IEEE, ACM SIGGRAPH and EuroVis. He has published over 152 journal articles, book chapters, conference papers, abstracts and posters in his research field.

DMITRIY MOROZOV is a postdoc fellow, and he has been with the Visualization Group since January 2011. Morozov received a Ph.D. in computer science from Duke University in 2008. Before joining LBNL, he spent two years as a postdoc scholar in computer science and mathematics at Stanford University. Morozov's research is in the field of computational geometry and topology. At LBNL, he has worked on parallel computation of topological descriptors as well as on extracting geometric descriptors of porous materials.

PARDEEP PALL recently joined the Visualization Group as a Computer Systems Engineer. He received an undergraduate degree in physics from University College London, and a Ph.D. in atmospheric physics from Oxford University. His research focuses on determining whether the risks of damaging weather related events—in particular floods—are affected by anthropogenic climate change. This involves generating large datasets of weather using climate models, and analyzing this using visualization software developed at LBNL. He has previous experience with the climate*prediction*.net public distributed computing project, and specializes in data mining and extreme value statistics.

PRABHAT is a research scientist in the Visualization Group. He has been at LBNL for the past five years, and holds a B. Tech in computer science and engineering from IIT Delhi, and an M.S. in computer science from Brown University. He is currently pursuing a Ph.D. in earth and planetary sciences at UC Berkeley. Prabhat's current research interests span scientific data management, parallel I/O, high performance computing, applied statistics, machine learning and big data analytics.

OLIVER RÜBEL is a Computer Systems Engineer in the Visualization Group, as well as a member of the NERSC Analytics Team. He was as a student assistant at LBNL from 2006 to 2010 and has been a staff Computer Systems Engineer since 2011. He received his M.S. and Ph.D. in computer science in 2006 and 2009 from the University of Kaiserslautern, Germany. From 2010 to 2011, Rübel worked as a postdoc researcher in the Data Analysis Group of the Center for Applied Scientific Computing at Lawrence Livermore National Lab. During his career, he has worked with scientists in different disciplines, including life sciences, high energy physics, climate and computational finance. Rübel has contributed substantially to research in the areas of query-driven visualization and linked multiple views. He is one of the lead developers for PointCloudXplore, the standard software tool of the Berkeley Drosophila Transcription Network Project for the visualization and analysis of 3D gene expression data. In close collaboration with accelerator scientists, Rübel has also developed a large range of methods for visualization and analysis of large particle datasets and automatic feature detection.

DANIELA USHIZIMA is a research scientist in the Visualization and Math groups in CRD, as well as a consultant for NERSC. Her work focuses on image analysis and pattern recognition as it applies in diverse scientific domains, ranging from biomedical pictures to porous material imaging with applications to carbon sequestration. She has been a principal investigator and co-principle investigator on multiple projects, including Quantitative Image Analysis for Computational Modeling, and Visualization and Analysis for Nanoscale Control of Geologic Carbon Dioxide. In 2012, she also lead a Statistical and Applied Mathematical Sciences (SAMSI) Working Group on Imaging, as part of the SAMSI Program on Massive Datasets. Before joining LBNL in 2007, she was a professor of computer science at the Catholic University of Santos, Sao Paulo, Brazil. Ushizima received her Ph.D. and M.S. from the University of Sao Paulo in computational physics, and she earned her undergraduate degree in computer science in 1997 at the Federal University of Sao Carlos. Her research interests include computer vision, machine learning, signal processing, quantitative microscopy, and high performance computing.

GUNTHER H. WEBER has been a research scientist in the Visualization Group since January 2007 and an adjunct assistant professor of computer science at UC Davis since July 2008. He earned his Ph.D. and M.S. in computer science in 2003 and 1999, and his B.S. in computer science with a minor in physics in 1995, all from the University of Kaiserslautern, Germany. Prior to joining the LBNL Visualization Group and the NERSC Analytics Team, Weber was first a postdoc scholar and later a project scientist at the Institute for Data Analysis and Visualization at UC Davis, working on the visualization of 3D gene expression data (with researchers of LBNL's Genomics and Life Sciences divisions), topology-based exploration of scalar data, and visualization of brain imaging data and experimental earthquake data. His research interests include computer graphics, scientific visualization and data analysis with a focus on topological data analysis methods, parallelization of visualization algorithms, and hierarchical data representation methods.

CO_2 SEQUESTRATION AND STORAGE: FROM RAW MICRO-CT TO QUALITY MEASUREMENTS

Research Highlights

Since 1751, nearly 337 billion tons of CO_2 were emitted into the atmosphere as the result of combustion of fossil fuels and cement production. If CO_2 is disruptive to the earth's climate, then reduction in atmospheric CO_2 is necessary to avoid an environmental catastrophe. One solution to controlling the CO_2 concentration in the atmosphere is to store it in deep subsurface rock formations using a safe and effective technology.

LBNL researchers are developing new experiments and simulations to increase the understanding of processes in fluid-rock systems, which relate to the geologic sequestration of CO_2. For example, they have designed materials and collected subsurface samples of porous materials, which are exposed to high energy X-rays using synchrotron radiation, while a detector captures information that can be reconstructed as image slices. These images are used to construct computational geometries for fluid dynamics simulations.

The research team working on the CO_2 sequestration research has developed tools to recover material structures from micro tomographic images. These tools perform an essential pre-processing step for subsequent analysis, such as the extraction of pore networks from porous materials, porosity and permeability estimation, and quantification of CO_2 during flow. The image analysis workflow, developed by Ushizima et al. [5], automates segmentation, and is deployed as part of the ImageJ plug-in known as Quant-CT, a new threaded, shared-memory parallel package that combines 3D nonlinear smoothing with the 3D region merging for material segmentation. This plug-in leverages standard algorithms, such as the bilateral filtering and the statistical region merging (SRM), and adapts them to deal with artifacts found in micro-CT images. The plug-in's substantial contributions come from a new scheme to estimate the photometric parameters of the bilateral filtering using the coefficient of variance, extracted from subimages (or patches). Another advance, in collaboration with Andrea Bianchi [6], was the algorithm that controls the coarseness of the segmentation and over-segmentation, called material assignment based on similarity histograms (MASH). Even in cases of sharp brightness variations across the image slices, MASH tackles over-segmentation that cannot be handled by SRM alone.

After splitting the micro-CT image stacks into volumes corresponding to dense material and empty spaces, feature extraction takes place in order to determine properties of the porous material. Morozov and Weber [4] implemented a set of algorithms using topological analysis to quantify maximum flow through porous networks (see Figure 2). They introduce a new approach to characterize porous materials using the extraction of Reeb graphs to represent the structure of the interstitial volume. In addition, they calculate flow graphs which approximate the network branch capacity for carrying flow. Sophisticated visualization tools emphasize the most prominent pore bodies of a porous material, which corresponds to the loci where liquid can accumulate.

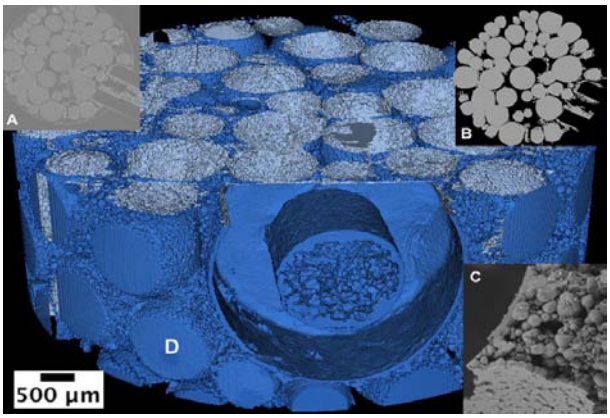
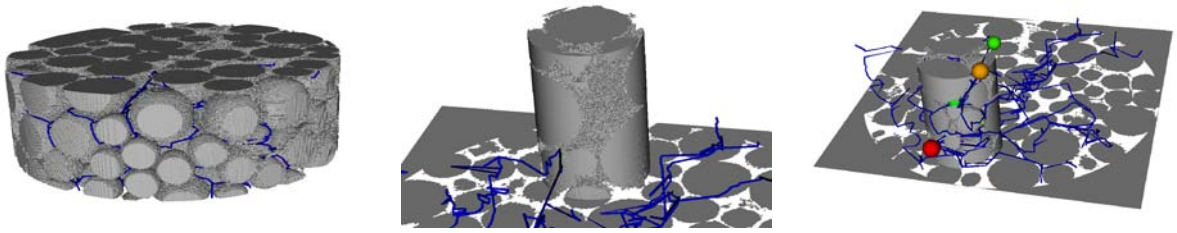


Figure 1: Micro-CT of porous media: (A) cross section of glass bead column, inoculated with ***S. pausterii*** that promote calcite precipitation; cross section is input to our software Quant-CT, which outputs segmented slices as in (B); rendering of the segmentation result for the whole stack in (D) using VisIt; SEM image in (C) emphasizes the result of biomineralization, which clogs the void space, cementing the pore channels.



(a) (b) (c)

Figure 2: (a) Segmented sample and the flow graph following the voids in the sample. (b) Cylindrical cut through a region of high flow, pocket spheres, and flow graph. (c) Focusing on a cylindrical region with little flow shows that pores are "blocked'" by calcite precipitation produced by microbes ***(S. pasteurii)*** in the experiments.

Scientific Impact

The developed tools and algorithms help domain scientists quantify material properties that are required for developing technologies to store CO_2 safely in deep surface rock formations. Currently, the analysis and feature extraction pipeline enables the detection of the solid phase from the micro-CT and the quantification of material's porosity automatically. The Reeb graph computation allows for an efficient extraction of pockets and pore network information. The pipeline also enables the comparison between different materials, and between experimental results and simulation outputs. Finally, part of the pipeline shows the increase of the image throughput and the decrease of the delay between data collection and characterization.

CRD Contacts

Wes Bethel, Andrea Bianchi, Dmitriy Morozov, Daniela Ushizima, Gunther Weber

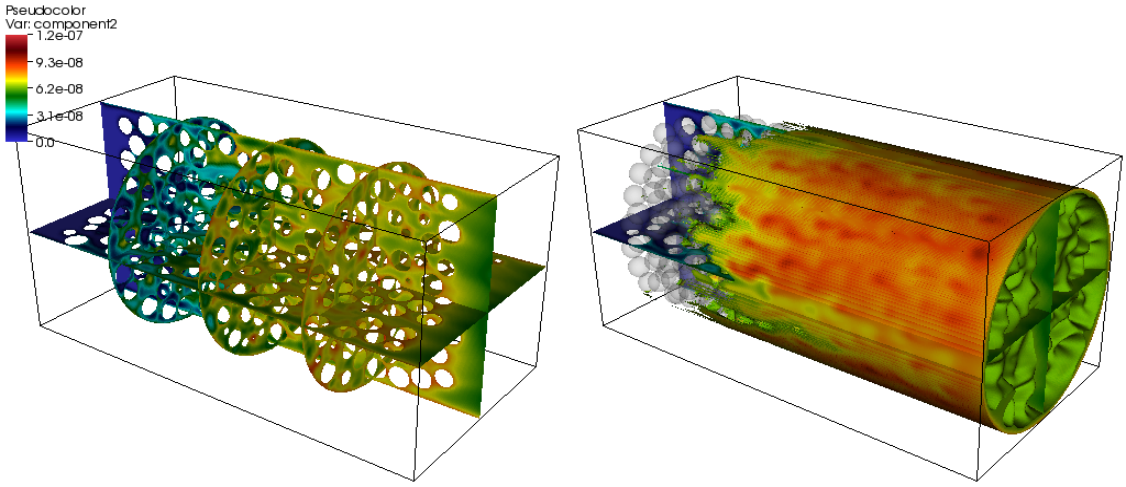


Figure 3: Visualization of fluid flow simulation of a glass bead pack.

PARALLEL I/O, ANALYSIS, AND VISUALIZATION OF A TRILLION PARTICLE SIMULATION

Research Highlights

The study of magnetic reconnection benefits from petascale computing, which enables the study of highly resolved, 3D problem configurations that are more physically realistic than previous 2D configurations. Recent studies conducted by CRD researchers and their collaborators use VPIC, a particle-in-cell plasma physics code, run at 120K-way concurrency on NERSC's Hopper with one trillion particles. This problem scale provides the fidelity required to investigate specific science questions, such as the nature of energetic particle behavior in the region of reconnection "hot spots." However, problems of this scale also expose significant data-related challenges that impact storage, analysis, and visualization of simulation results.

One such challenge is in the extreme-scale parallel I/O. The simulation produces approximately 32TB of particle data per timestep each of which contains one trillion particles. They incorporated H5Part [3], a veneer API that encapsulates the complexity of parallel HDF5 developed jointly by CRD researchers and collaborators from the high energy physics community, into the VPIC codebase for writing the data to a single shared HDF5 file, an approach that avoids the pitfalls associated with the traditional *file-per-process* I/O model. Working with NERSC center staff, they manually tuned the HDF5, Lustre and MPI-IO parameters to achieve an approximate 27GB/s average write performance, which is a significant fraction of system peak I/O rate (approximately 35GB/s).

Another challenge is implementing high performance index/query for analysis and visualization. The core of visual data analysis is being able to quickly search through scientific data for subsets that meet specific criteria that correspond to features or phenomena of interest. Building upon a substantial body of work within CRD in the area of query-driven visualization and analysis, CRD researchers developed a hybrid parallel version of FastQuery [2], a distributed-memory parallel implementation of bitmap indexing. With this new hybrid parallel implementation on Hopper, the team was able to perform a trillion particle dataset in about 10

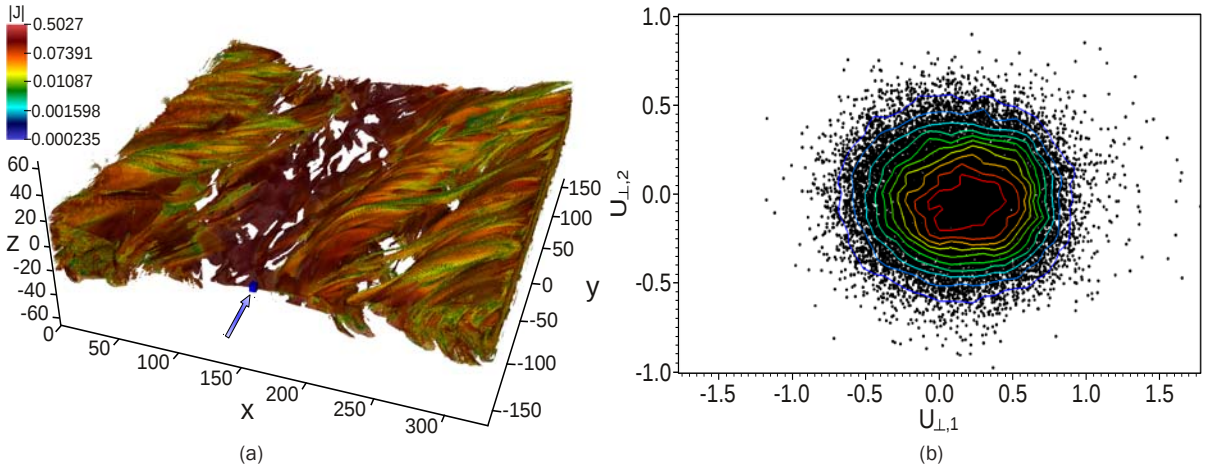


Figure 3: (a) Particle scatter plot (black) of $U_{perpedicular,\,1}$ vs. $U_{perpendicular,\,2}$ of all energetic particles (with Energy > 1.3) contained in the box in the x-line region indicated in Figure 4b. Additional isocontours indicate the associated particle density (blue = low density and red = high density). The elliptical shape of the particle distribution is indicative of agyrotropy in the x-line region. (b) Isosurface plot of the positron particle density n_p with color indicating the magnitude of the total current density |J|. The blue box (indicated by the arrow) is located in the x-line region of the simulation and illustrates the query (157.654 < x < 1652.441) && (-165 < y <-160.025) && (-2.5607 < z < 2.5607), which is used in Figure 4a to study agyrotropy.

minutes, and execute queries in less than three seconds.

These new capabilities—high performance I/O and index/query—enabled the team to study specific science questions about magnetic reconnection that were previously out of reach. One discovery was that energetic particles are confined by the magnetic flux ropes (see Figure 4a), another was the agyrotropic distribution of energetic particles in the region around a magnetic reconnection "hot spot" (see Figure 4b).

Scientific Impact

This work, which is documented in an SC12 technical program publication [1], demonstrates several "first-ever" accomplishments. First, this work is the first time anyone ever ran a realistic trillion particle simulation on 120K Hopper cores. A plasma physics simulation of this scale produces phenomena never seen before. Second, this is the largest-ever write to a single shared 32TB HDF5 file with collective I/O. Third, during the simulation run and subsequent parallel write operations, the team sustained approximately 27GB/s in I/O performance, with occasional reaches to the full approximate 35GB/s theoretical peak performance. This level of I/O performance has never before been achieved. Fourth, the combination of FastBit/FastQuery/H5part has never before operated at this scale. Fifth, the science collaborators have never before looked at the one trillion particle data, which in turn lead to new science discoveries, namely the confinement of energetic particles to magnetic flux ropes and the agyrotropic behavior of energetic particles around the magnetic reconnection hot spots. Sixth, scientists examined the energy spectrum of the trillion particles and discovered a power-law distribution. Power laws are routinely found in natural phenomena, and have been conjectured about in plasma physics simulations. This is the first such discovery in a 3D kinetic plasma simulation. Finally, the technological capabilities are deployed in software that is accessible by the worldwide software community (VisIt, FastQuery). These accomplishments reflect a successful collaboration between science researchers and several different programs at LBNL (visualization, scientific data management, NERSC).

CRD Contacts

Prabhat, Suren Byna, Oliver Rübel, E. Wes Bethel, Mark Howison, Arie Shoshani, Kesheng Wu, Andrew Uselton

Collaborators

Homa Karimabadi (UCSD), Vadim Roytershteyn (UCSD), Bill Daughton (LANL)

REFERENCES

[1] Surendra Byna, Jerry Chou, Oliver Rübel, Prabhat, Homa Karimabadi, William S. Daughton, Vadim Roytershteyn, E. Wes Bethel, Mark Howison, Ke-Jou Hsu, Kuan-Wu Lin, Arie Shoshani, Andrew Uselton, and Kesheng Wu. "Parallel I/O, Analysis, and Visualization of a Trillion Particle Simulation." In *Proceedings of SuperComputing 2012*, Nov. 2012. LBNL-5832E.

[2] Jerry Chou, Kesheng Wu, Oliver Rübel, Mark Howison, Ji Qiang, Prabhat, Brian Austin, E. Wes Bethel, Rob D. Ryne, and Arie Shoshani. "Parallel Index and Query for Large Scale Data Analysis." In *Proceedings of Supercomputing 2011*, Seattle, WA, USA, November 2011. LBNL-5317E.

[3] Mark Howison, Andreas Adelmann, E. Wes Bethel, Achim Gsell, Benedikt Oswald, and Prabhat. "H5hut: A High-Performance I/O Library for Particle-Based Simulations." In *Proceedings of 2010 Workshop on Interfaces and Abstractions for Scientific Data Storage* (IASDS10), Heraklion, Crete, Greece, September 2010. LBNL-4021E.

[4] Daniela Ushizima, Dmitriy Morozov, Gunther H. Weber, Andrea Bianchi, and E. Wes Bethel. "Augmented topological descriptors of pore network." In *VisWeek,* Seattle, WA, USA, October 2012. To appear. LBNL report number pending.

[5] Daniela Ushizima, Dilworth Parkinson, Peter Nico, Jonathan Ajo-Franklin, Alastair MacDowell, Benjamin Kocar, Wes Bethel, and James Sethian. "Statistical segmentation and porosity quantification of 3D x-ray microtomography." *SPIE*, 2011.

[6] Daniela M. Ushizima, Andrea Bianchi, Christina deBianchi, and E. Wes Bethel. "Material Science Image Analysis using Quant-CT in ImageJ." In *Proceedings of ImageJ User and Development Conference*, October 2012. To appear. LBNL report number pending.

COMPUTATIONAL COSMOLOGY CENTER

In recent years astrophysics has undergone a renaissance, transforming from a data-starved to a data-driven science. A new generation of experiments will gather datasets so massive that their analysis will require the use of leading-edge, high performance computing resources. Continuing a decade-long collaboration in this field, the Computational Research and Physics Divisions at LBNL have formed the Computational Cosmology Center (C^3).

C^3 is a focused collaboration of astrophysicists and computational scientists whose goals are to develop the tools, techniques and technologies to meet the analysis challenges posed by present and future cosmological datasets.

OVERVIEW OF HIGHLIGHTED PROJECTS

Members of C^3 conduct research in a number of areas where high performance computing is needed to support theoretical and observational cosmology, or where massively parallel cosmology codes can help to drive computational science research and development. This section focuses on their work on the astrophysics imaging pipeline, simulations and analyses for the polarization of the CMB, and CMB data analysis from the Planck Satellite Mission.

RESEARCH STAFF

PETER NUGENT is a senior staff scientist and Group Lead for C^3 and an adjunct professor in the Astronomy Department at UC Berkeley. Nugent attended Bowdoin College and received his M.S. and Ph.D. in physics with a concentration in astronomy from the University of Oklahoma. He joined LBNL in 1996 as a postdoc fellow working with Saul Perlmutter on the measurement of the accelerating universe with Type Ia supernova, for which Perlmutter received the Nobel Prize in physics in 2011. Nugent has presented his work as a participant on *PBS News Hour*, NASA's *Space Science Update* program, CNN, NOVA, NPR, and BBC. His work has also been featured in *Time*, *Newsweek*, *Science* and *Nature*.

JULIAN BORRILL is a senior staff scientist and co-founder of C^3. He leads the cosmic microwave background group's work on massively parallel analysis algorithms and implementations, and their application to data from the Planck satellite, EBEX balloon, and PolarBear ground-based experiments. He is also the Computational Systems Architect for the US Planck collaboration. He holds multiple master's degrees in mathematics and political science from the University of Cambridge, astrophysics and information technology from the University of London, and a Ph.D. in theoretical physics from the University of Sussex.

REIJO KESKITALO recently started in C^3 to work on scalable mapmaking for the next generation of astrophysical experiments. He received his Ph.D. in theoretical physics from University of Helsinki in 2009. He has also been a member of the Planck collaboration since 2005. Previously, Keskitalo was a postdoc at the California Institute of Technology where he worked on cosmic microwave background mapmaking and residual noise characterization. He continues to work on Planck through both low and high frequency instrument consortia.

TED KISNER is a Computer Systems Engineer working in C^3. Kisner received his M.S. and Ph.D. in physics from UC Santa Barbara in 2002 and 2008, respectively. His work focuses on algorithm development and using modern software engineering practices to improve the quality and scalability of software for cosmological data analysis. Kisner's current projects involve architecting a large software toolkit for cosmic microwave background experiments and prototyping techniques for maximum likelihood extraction of spectra from the proposed BigBOSS project. Kisner also works on packaging and managing production cosmology software across the machines at NERSC, the National Energy Research Scientific Computing Center.

AARON COLLIER is a Computer Systems Engineer working in C^3. He holds degrees in computer science and mathematics from Widener University. His current work is primarily focused on application performance analysis and code optimization for massively parallel platforms. He has worked on a variety of codes including nonlinear algebraic equation solvers and nonlinear tokamak microturbulence codes. Collier previously worked at Lawrence Livermore National Laboratory and General Atomics.

ZARIJA LUKIC has been a postdoc in C^3 since 2011. He received his Ph.D. in 2008 at the University of Illinois at Urbana–Champaign, and was a postdoc research associate in the Theoretical Division at Los Alamos National Laboratory from 2008 to 2011. The main focus of his research is in the large-scale structure of the Universe, and accurate determination of cosmological parameters with particular emphasis on theoretical predictions for the Lyman-alpha forest using the Nyx code, which is relevant to current and future DOE-HEP experiments. Lukic's research also includes aspects of nuclear physics, and he has published research on the practical usage of cosmic rays to obtain material identification of scanned cargos, as well as to remotely diagnose the state of damaged cores at the Fukushima reactors.

CASEY STARK is a third year astrophysics graduate student at UC Berkeley, working with Peter Nugent in C^3. He received his B.S. in physics and math at the University of Southern California. His work focuses on cosmological simulations and building open source, easy-to-use simulation tools. Currently, Stark is working on simulating the Lyman-alpha forest to aid observational analysis and to learn about the high-redshift IGM.

ROLLIN THOMAS has been a staff scientist in C^3 since 2012. He received his B.S. in physics from Purdue University and his Ph.D. from the University of Oklahoma in 2003. Thomas is a computational astrophysicist focused on bringing high performance and parallel computing techniques to bear on problems related to dark energy. In particular, he works on experimental and theoretical efforts to better understand Type Ia supernovae to measure the accelerating expansion of the Universe. On the experimental side, he is interested in the discovery, follow-up, classification, and analysis of transient astrophysics data in evermore efficient and automated ways. His interests in supernova theory are related to numerical radiative transfer and involve creating new, data-driven codes to interpret supernova observations.

CHELSEA HARRIS is a graduate student at UC Berkeley. Her undergraduate studies were at UC Santa Barbara in the College of Creative Studies Physics Department. She received a B.S. in physics with Highest Academic Honors and Distinction in the major. In 2011, Harris participated in an internship focusing on creating synthetic radio line profiles for simulated molecular cloud cores. Her interests include large-scale cosmology, supernovae, binary evolution, binary formation and accretion disk instabilities.

ASTROPHYSICS IMAGING PIPELINES

Research Highlights

The Palomar Transient Factory (PTF), the La Silla Supernova Search (LSSN) and the Dark Energy Survey Supernova Search (DESSN) are experiments designed to systematically explore the optical transient and variable sky. The main goal of these projects is the discovery and follow-up of Type Ia supernovae—cosmic probes of distance—whose observation led to the discovery of the accelerating universe and today remains of the best ways to study dark energy.

LBNL's Physics Division is a major participant in these programs, and they along with Fermi National Accelerator Laboratory, Argonne National Laboratory and SLAC Laboratory are leading the effort on DESSN. In addition to this science, these surveys will fill the gaps in our present-day knowledge of the optical transient phase space and will provide deep co-additions of imaging data for the target selection of faint galaxies and quasars for future surveys like BigBOSS.

For one example of C³'s projects, PTF is currently the largest optical astronomical survey in the world and in four years of operations, researchers at Caltech, LBNL, IPAC, Berkeley, LCOGT, Oxford, Columbia and the Weizmann Institute have acquired over 2.5 million digital images spanning most of the extragalactic sky visible from the Northern Hemisphere. During each clear night, they acquire nearly 100GB of digital images generating over 60 thousand transients detections (the needles), hidden within 2 million false positives due to image artifacts, mis-subtractions, cosmic rays, etc. (the haystack). Of these 60 thousand transients, only a handful are truly new and exciting astrophysical events (supernovae, gamma-ray bursts, etc.), and the majority are well-known and repeating events, such as variable stars, asteroids and active galaxies.

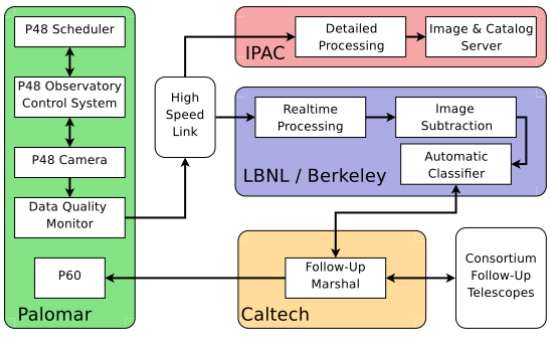


Figure 1: A schematic of the Palomar Transient Factory pipeline. Over 100GB of data come from the telescope each night and are processed in near real-time at LBNL/NERSC. Machine learning classifiers look at well over one million potential candidates every night to automatically and a handful of new and exciting transients like the Type Ia supernova SN 2011fe.

Figure 2: A false-color image of SN 2011fe in the Pinwheel galaxy. It is the closest Type Ia supernova in the last 25 years and the brightest in the last 40 ears. Its discovery was made only 11 hours after explosion. The early discovery and announcement to the worldwide community has made it the best studied supernova of its type to date.

The data is processed at NERSC and made available over the web and through databases on the Science Gateway nodes housed at the supercomputing facility. Images are taken every minute during the night and processed in near real-time generating over a thousand potential candidates per minute. These candidates need to be compared in real-time to past events, as well as to known nearby astrophysical objects, and then analyzed with a suite of machine learning algorithms to assess their quality.

Scientific Impact

The rapid turnaround is necessary in order to detect the supernovae as fast as possible with additional instrumentation to understand the physics, which drives these events. The pipelines used to process the data, sift through the candidates, and output the results to their respective collaborations were developed in the Computational Cosmology Center.

While DESSN and LSSN are just beginning, PTF has been very successful. With nearly 2,000 spectroscopically confirmed supernovae, PTF has been the most successful survey of its generation and a forerunner of surveys like the DOE-NSF Large Synoptic Survey Telescope. PTF has discovered several interesting new astrophysical phenomena and has by far expanded our knowledge of the early phases of supernovae more than any previous survey. With three papers published in Nature, two in Science and over fifty-one papers total just in the past four years, the survey has been extremely productive scientifically. The pipelines created at NERSC for these surveys have been great test beds for other experiments. They fully explore the transfer of data, processing, analysis and distribution of science to large international collaborations with real-time turnarounds.

CRD Contact

Peter Nugent

Collaborators

Shri Kulkarni (PTF/Caltech), Charles Baltay (LSSN/Yale), Masao Sako (DESSN/UPenn)

SIMULATIONS AND ANALYSIS OF CMB POLARIZATION DATA AT THE PETASCALE AND BEYOND

Research Highlights

To achieve the necessary signal-to-noise to detect Cosmic Microwave Background (CMB) polarization, the size of the datasets being gathered will grow thousand-fold over the next 15 years (just as datasets have increased over the last 15 years). This will require simulation and analysis tools to scale through the next 10 iterations of Moore's law, into the realm of exascale computing.

For the last four years, C^3 has been working to address the challenges of scaling their codes to next-generation HPC systems. This project, developed from their work for the Planck satellite mission, will directly support C^3's collaborations on the PolarBear and EBEX experiments and will be made available to all next-generation CMB experiments (a large fraction of which use NERSC's HPC resources). Using the full capability of next-generation HPC systems requires two broad areas of work, though: first, addressing C^3's code I/O and communication bottlenecks, and second, enabling the codes to use new types of compute hardware (e.g., GPUs, Intel MIC).

Given the hardware available at the projects outset (primarily NERSC's Hopper system), C^3 staff first addressed the I/O and communication bottlenecks. As shown in Figure 3, C^3 staff improved their simulation and mapping (SimMap) performance by a factor of 250 over the last six years, with half the increased performance time coming from Moore's Law and half from algorithm and implementation improvements. This work also demonstrated that at the highest concurrencies communication-bound codes may benefit from running with one process and 24 threads per Hopper node, despite the inevitable NUMA overhead.

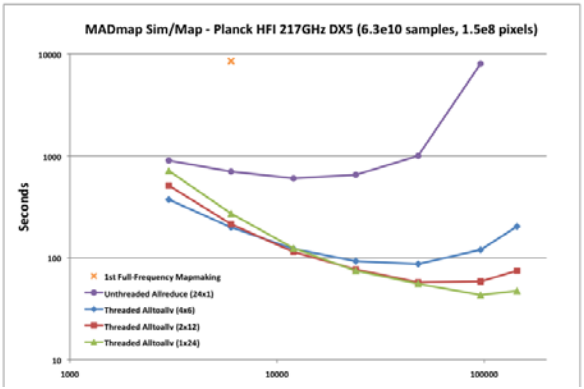


Figure 3: Performance of C^3's core simulation and map-making software at NERSC, from the rest run on 6000 cores of Seaborg to the most recent runs on 144,000 cores of Hopper, showing the impact of hybridization and MPI-optimization.

Scientific Impact

The primary goal of post-Planck CMB experiments is to detect the cosmological B-mode signal that is expected to be generated by gravitational radiation from the epoch of inflation. Such a signal would both confirm and constrain the inflationary paradigm. The faintness of the signal both requires enormous data volumes to overcome the instrument noise and exquisitely precise handling of systematic effects. Both of these requirements need large numbers of detailed simulations that this code development is targeting. Already, C^3 has been able to achieve Planck's simulation goals, and are pushing towards PolarBear-scale runs (10x Planck's data volume). In the coming year, C^3 will also transition to Leadership Class Facilities and NERSC's new Edison system, enabling researchers the ability to both extend their scaling to even higher concurrencies and begin the work of extending the code to take advantage of GPU and MIC architectures.

CRD Contacts

Julian Borrill, Aaron Collier, Josquin Errard, Reijo Keskitalo, Theodore Kisner

Collaborators

Carlo Baccigalupi (SISSA, Trieste), Andrew Jae (IC, London), Radek Stompor (APC, Paris)

COSMIC MICROWAVE BACKGROUND DATA ANALYSIS FOR THE PLANCK SATELLITE MISSION

Research Highlights

The Planck satellite mission will gather the largest, most detailed observations of the Cosmic Microwave Background (CMB) to date. However, the analysis of certain datasets, like the CMB datasets and the Planck dataset, given its precision and volume in particular, is a significant computational challenge.

For the last 10 years NASA has funded the CMB data analysis group in CRD to develop the high performance computing tools needed to perform the most computationally challenging elements of the Planck data analysis on NERSC supercomputers. For the last five years, this program has been formalized by a unique memorandum of understanding between DOE and NASA, under which DOE guarantees Planck an annual allocation of NERSC resources, and NASA guarantees the Berkeley group 2.5 full-time employees, both for the lifetime of the mission.

The Planck program builds upon C^3's previous work on the Maxima and BOOMERanG balloon experiments, and the program itself provides the starting point for the development of the HPC tools needed for even larger datasets to be gathered by post-Planck suborbital experiments, like EBEX and PolarBear.

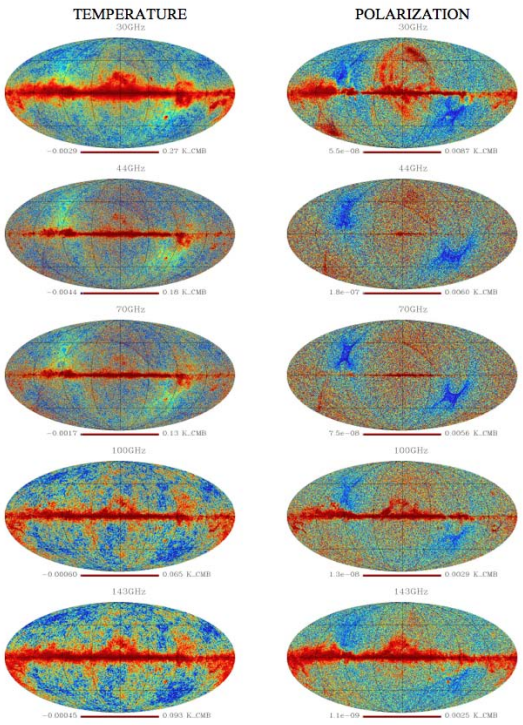


FIGURE 4: FFP6 temperature and polarization maps corresponding to one realization of the Planck data at all of its nine frequencies (seven polarized, two unpolarized). These data include both the CMB and astrophysical foregrounds, and instrument noise, band-pass and beam effects.

The CMB provides a unique image of the Universe, as it was only 400,000 years after the Big Bang, and it provides a powerful probe of both cosmology and ultra-high energy physics. Previous CMB data, in tandem with observations of Type Ia supernovae, overturned the standard model of cosmology and replaced it with a Universe dominated by a mysterious dark energy; all of the experiments now being fielded that try to understand this depend upon the Planck results to break degeneracies in their parameter space.

Controlling statistical and systematic uncertainties is critical to CMB data analyses. These analyses rely on huge ensembles of simulations. C^3's goal is to develop the tools to generate O(10,000) synthetic Planck observation time-streams and apply the same map-making to each as was used for the real data, and to do so quick and cheap enough to meet the mission's schedule and resource constraints.

A series of evermore complex and complete simulations of the Planck full focal plane (FFP) has been generated over the past five years. C^3 has recently completed FFP6, generating around 50TB of simulated data and over 100,000 maps, using approximately 5 million CPU-hours on up to 100,000 Hopper cores over a period of two months. This will be crucial in validating the tools and quantifying the results to be published in the first Planck cosmology release in March 2013.

Scientific Impact

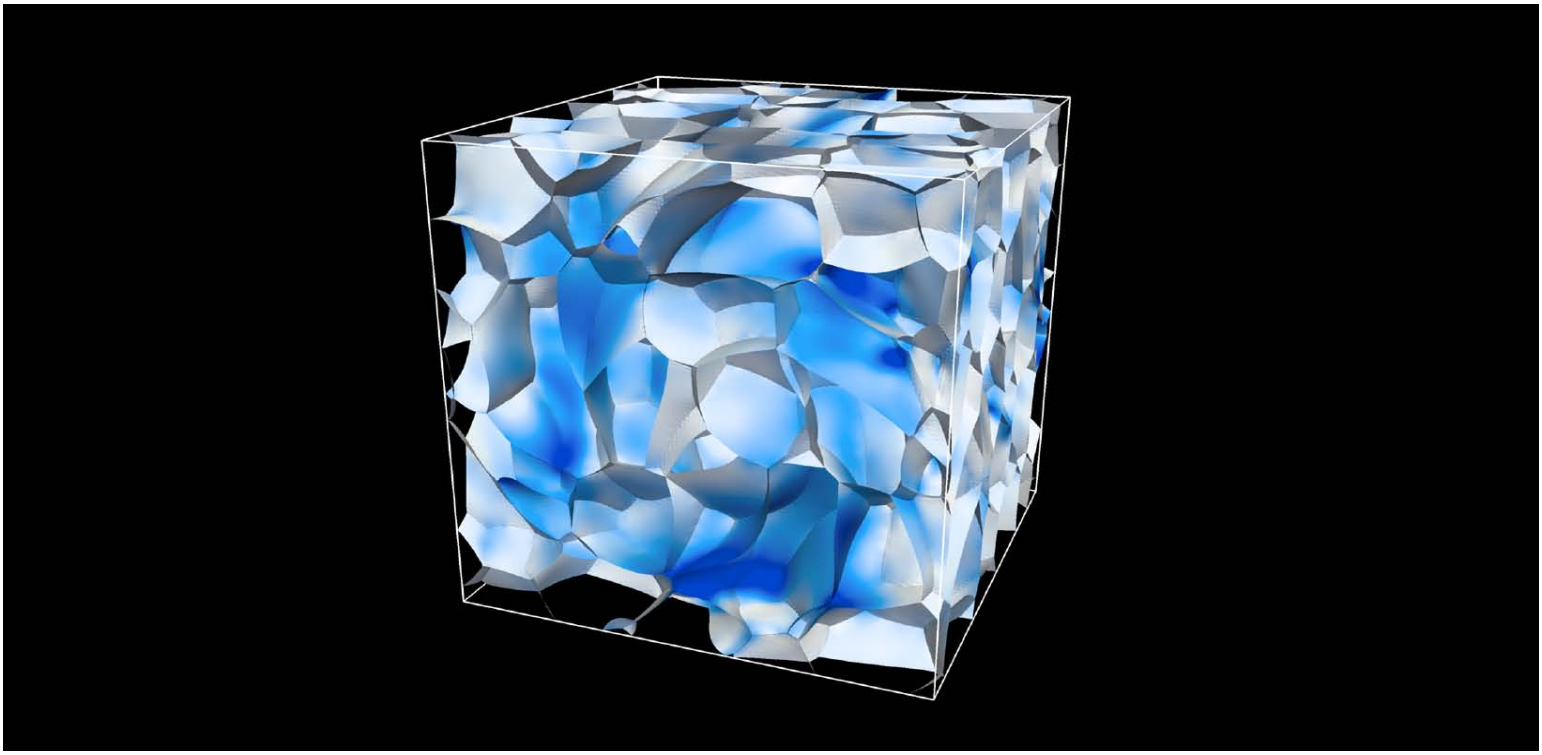
All current and forthcoming cosmology experiments will depend upon Planck results to complement their observations. For at least the next decade Planck will provide the definitive temperature and the only high-resolution, all-sky, polarization dataset. However, the fluctuations in the CMB that encode the fundamental science are so tiny that without very precise control, they would be overwhelmed by statistical and systematic uncertainties. This work is therefore absolutely essential for the comprehensive scientific exploitation of the dataset being gathered by Planck. Over the last five years, C^3 has increased the scale of their simulation suites approximately 250-fold, with half of that coming from increases in NERSC's computing capability and half from improvements to C^3's algorithms.

CRD Contacts

Julian Borrill, Aaron Collier, Reijo Keskitalo, Theodore Kisner

COLLABORATORS

The Planck Collaboration



MATHEMATICS GROUP

The Mathematics Group's research centers on the development of numerical and analytical methods and their application to challenging, difficult problems in physics and engineering. Current areas of specialized interest are vortex and particle methods, solid mechanics and fracture, interface techniques, turbulence theory, dynamics of polymeric systems, parallel implementation of codes for large-scale scientific computing, and fast algorithms.

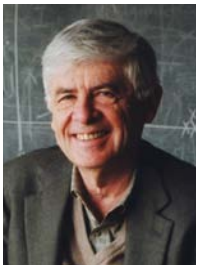
OVERVIEW OF HIGHLIGHTED PROJECTS

The Mathematics Group is in collaboration with the UC Berkeley Mathematics Department. All of the staff hold joint appoints at LBNL and UC Berkeley and therefore, their research reflects a team effort among senior staff scientists, professors, postdocs, and graduate students. Their highlighted projects here are only a few of many, which include implicit sampling, mathematical and algorithmic methodologies for computing multi-phase, multi-physics problems, math applications to vertical axis wind turbine simulations, and long-time dynamics and optimization of nonlinear partial differential equations.

RESEARCH STAFF

JAMES A. SETHIAN is a professor of mathematics at UC Berkeley, as well as a Senior Faculty Scientist and Group Lead of the Mathematics Group. He received his Ph.D. in applied mathematics from UC Berkeley in 1982. Sethian continued his research with an NSF postdoc fellowship at the Courant Institute of Mathematics, and then returned to Berkeley as an assistant professor in 1985. He now holds the James H. Simons Chair in Mathematics at UC Berkeley. Sethian is the author of many scientific articles and books, and serves as an associate editor on several journals. He is a member of the National Academy of Engineering, and the recipient of numerous prizes and awards. His awards include the Norbert Wiener Prize in Applied Mathematics, which is awarded jointly by the American Mathematical Society (AMS) and the Society for Industrial and Applied Mathematics (SIAM), for his representations of the motion of curves, surfaces, interfaces, and wave fronts, and for his applications of mathematical and computational ideas to scientific problems. He was also awarded the 2011 Pioneer Prize, which is awarded by the International Council for Industrial and Applied Mathematics (ICIAM) for his contributions to applications in imaging and shape recovery in medicine, geophysics, tomography, and drop dynamics in ink jets. Sethian works the theory, algorithms, and applications of moving interfaces as they are applied to problems in fluid mechanics, materials science, and industrial processes, such as ink jet printing, semiconductor fabrication, biological and medical imaging, and geophysics.

GRIGORY BARENBLATT is an Emeritus Professor of Mathematics at UC Berkeley, and he is an Emeritus Senior Faculty Scientist at LBNL. Before joining LBNL in 1997, Barenblatt graduated in 1950 and completed his Ph.D. in 1953 at Moscow University's Department of Mechanics and Mathematics. He received his D.Sc. in 1957 and obtained his professorship in 1967 from Moscow University as well. From 1992 to 1994, Barenblatt was the G.I. Taylor Professor of Fluid Mechanics at Cambridge University. His work includes research on fracture mechanics, the theory of fluid and gas flows in porous media, the mechanics of non-classical deformable solids, turbulence, and self-similarities, nonlinear waves and intermediate asymptotics. His studies led to non-classical problems in mathematical physics with wide practical applications. Barenblatt has had many awards and honors for his work in mathematics. His honors include the Timoshenko Medal from the American Society of Mechanical Engineers, and Maxwell Prize from the International Council for Industrial and Applied Mathematics, and the G.I. Taylor Medal from the US Society of Engineering Sciences, and foreign membership in the US National Academy of Sciences.



ALEXANDRE CHORIN is a UC Berkeley professor of mathematics and a Senior Faculty Scientist in the Mathematics Group. He received his Ph.D. from the Courant Institute of Mathematics at New York University in 1966, becoming an associate professor in 1971 before joining the Berkeley faculty in 1972. He was head of the LBNL Mathematics Department from 1986-1985. Chorin's awards include the National Academy of Sciences award in applied mathematics and numerical analysis, the Wiener prize of the American Mathematical Society, and the Lagrange prize of the International Council on Applied Mathematics. Chorin is known for his contributions to computational fluid mechanics, turbulence theory, and computational statistics, including the invention of the ubiquitous projection method for modeling incompressible fluids and the random vortex method for computing turbulent flow.

RESEARCH STAFF

OLE HALD is a professor of mathematics at UC Berkeley. He received his Ph.D from the Courant Institute of Mathematics at New York University in 1972. He works in numerical analysis and inverse problems with a wide collection of applications to geophysical simulations, inverse sampling, stochastic particle methods, and fluid mechanics.

JON WILKENING is an associate professor in the Mathematics Department at UC Berkeley and a member of the Mathematics Group. Wilkening graduated *summa cum laude* with honors with a B.S. in engineering physics from the University of Arizona in 1996. He received a Ph.D. in mathematics from UC Berkeley in 2002, where he studied microchip failure due to stress-driven grain boundary diffusion and electromigration. He was a DOE Computational Science Graduate Student Fellow from 1997 to 2001 and was chosen as a Frederick A. Howes Scholar in Computational Science in 2003. Wilkening was a Courant Instructor at New York University from 2002 to 2005. He joined the faculty at UC Berkeley in 2005, and won an NSF CAREER award in 2010. His research interests are in numerical analysis, computational physics, partial differential equations and high performance computing. He has studied a wide variety of physical systems, including the failure of microchips, optimal transportation, gas dynamics, shape-optimization of robotic crawlers, lubrication theory, mode-locked lasers, water waves, and rolling tires.

PER-OLOF PERSSON has been an assistant professor of mathematics at UC Berkeley since 2008, and he is also an affiliate in the Mathematics Group at LBNL, and the Department of Aeronautics and Astronautics at MIT. Between 2005 and 2008, Persson was an instructor of applied mathematics at MIT. His current research interests are in high-order methods for fluid and solid mechanics, and in particular techniques based on the discontinuous Galerkin method. Persson developed the 3DG software package based on new discretization techniques and efficient solvers, which can be applied to problems in aerodynamics and solid dynamics, such as aero-acoustics, biologically inspired flapping flight, wind turbines, and micro-mechanical devices.

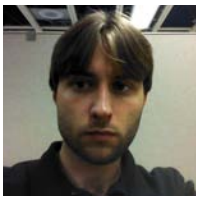
F. ALBERTO GRÜNBAUM is a Senior Faculty Scientist in the Mathematics Group at LBNL. He also holds a full-time position in the Mathematics Department at UC Berkeley. Before joining LBNL in 1976, Grünbaum taught at the Courant Institute at NYU, was a research scientist at the IBM Research Center in Yorktown Heights, NY and then joined the Applied Mathematics Department at Caltech before moving to Berkeley in 1974. He has served as director of the Center for Pure and Applied Mathematics and then Chairman of the UC Berkeley Mathematics Department from 1989–1992. He has been published extensively in the area of imaging, and his recent research interests include the study of quantum walks and its applications to different areas of science and technology.

BEN STAMM was a Morrey Visiting Assistant Professor in the Department of Mathematics at UC Berkeley and was affiliated with the Mathematics Group at LBNL from 2010 to 2012. He is currently an assistant professor with chair (Maître de conférence avec chaire) at the Laboratoire Jacques–Louis Lions at the University of Pierre et Marie Curie (UPMC-Paris VI). He received his Ph.D. in 2008 and his M.S. in 2005 both in mathematics from the École Polytechnique Fédérale de Lausanne (EPFL). His research interests include numerical analysis, partial differential equations, discontinuous Galerkin methods, stabilization strategies, reduced basis method, scattering problems, electrical field integral equation, and *a posteriori* estimates.

CHRIS RYCROFT is a Morrey Assistant Professor in UC Berkeley's Mathematics Department and the Mathematics Group at LBNL. He works on developing parallel and high-performance simulation methods to address problems in solid and fluid mechanics. Rycroft has also worked on simulations of dense granular materials where he wrote a free software library, Voro++, for carrying out 3D cell-based computations of the Voronoi tessellation. Currently, he is working with the Bay Area Physical Sciences-Oncology Center on computational cell modeling.

AUGUST JOHANSSON, a native of Sweden, received his M.S. in engineering physics from Chalmers University of Technology and his Ph.D. in computational science and engineering from Umea University. He is currently a Miller Fellow at the Department of Mathematics, UC Berkeley. Johansson is interested in developing numerical methods for partial diferential equations, primarily using finite elements. He is particularly interested in problems involving phenomena that are described by several partial diferential equations coupled to each other, which may involve different temporal and spatial scales. He is also developing fixed grids methods where the finite element mesh does not need to fit the computational domain.

MATTHIAS MORZFELD is a mathematician postdoc fellow at LBNL. He received his Ph.D. in mechanical engineering from UC Berkeley in 2011 and his M.S. in 2009 also in mechanical engineering from UC Berkeley. He graduated from the Technical University Darmstadt, Department of Mechanical Engineering in 2007 with distinction. His research interests include Monte Carlo methods, data assimilation, vibration and controls, and model order reduction. His goal is to create methods and algorithms that are useful in applications in oceanography, geomagnetism, robotics, structural dynamics and earthquake engineering.



JEFF DONATELLI is a sixth year math graduate student at UC Berkeley. He received his B.S. in mathematics with high honors from the University of Maryland, College Park in 2007. He is a DOE Computational Science Graduate Fellow. His current research includes algorithm design for next generation X-ray imaging, computational harmonic analysis, phase retrieval, and orientation determination.

BRADLEY FROEHLE is a graduate student and research assistant at UC Berkeley and in the Mathematics Group where he works on high-order fluid and fluid-structure interaction methods. He received his B.S. in math, *summa cum laude* and with high distinction, and in physics, with high distinction, from University of Minnesota, Twin Cities in 2003. Froehle's research interests include numerical analysis, fluid and solid mechanics, micromagnetics, computational physics, partial differential equations, mesh-based partial differential equation methods, deformable domains, discontinuous Galerkin finite elements, high performance computing, multigrid, numerical optimization, time periodicity, and GPU programming (CUDA/OpenCL).

MICHAEEL KAZI is a Ph.D. student at UC Berkeley in the Mathematics Department. He received his B.S. in computer science and applied mathematics from the Florida Institute of Technology in 2007. Kazi's research works on the reconstruction of surfaces with level set methods, given volume-of-material data.

TREVOR POTTER received his B.A. in mathematics from Gustavus Adolphus College in 2006. His current research is on time-periodic solutions of viscoelastic treaded rollers and the effects of viscoelasticity on steady state rollers. Some of Potter's past research has been on numerical studies of effective dynamics equations for solutions to the nonlinear Schrödinger equation.

BEN PRESKILL received his B.S. in mathematics from Harvey Mudd College in Claremont, California. He is currently a fourth year graduate student in the Mathematics Department at UC Berkeley and an NSF Graduate Fellow. His research is working on elastic interface evolution in biological systems.

ROBERT SAYE studied at the Australian National University and received a Bachelor of Philosophy with First Class Honours (2007), specializing in applied mathematics. He expects to graduate from UC Berkeley with a Ph.D. in applied mathematics in 2013. His research interests include problems involving multiple moving interfaces, development of advanced numerical methods, multi-scale physics, fluid flow and fluid-structure interaction, high performance computing methods and scientific visualization.

LUMING WANG received a B.S. in mathematics from Zhejiang University, China, and he is currently a third-year graduate student in applied mathematics at UC Berkeley. He has also studied at UCLA and Uppsala University, Sweden during his undergraduate studies. His research interests include high-order accurate numerical method, unstructured mesh generation and physics-based simulations.

JUE CHEN is a second year math graduate student at UC Berkeley. She received her B.S. in mathematics and applied mathematics from Zhejiang University, Chu Kochen Honors College in 2010, with an emphasis in pure mathematics. Her current research interests include the compact finite difference method on linear elasticity problems, and the boundary integral method on wave equations.

MARIA GARZON is a full-time professor at Oviedo University, Oviedo, Spain, and she is also a visiting scholar at LBNL. She graduated from the Superior Mining School, Oviedo, Spain in mining engineering, and she received her Ph.D. in mining engineering from the University of Oviedo as well. Her research is in embedded potential flow models and level set methods, and numerical methods for free boundary problems. Garzon is currently developing embedded potential methods for electrical charged drops in a 3D formulation to simulate industrial electro-spray devices.

JAKUB KOMINIARCZUK studied at the Massachusetts Institute of Technology and received two degrees: a B.S. in mathematics and a B.S. in physics, both in 2007. He expects to graduate from UC Berkeley with a Ph.D. in mathematics in 2013, specializing in applied mathematics. His research interests include problems involving high-dimensional probability distributions which appear frequently in problems arising in condensed matter and quantum physics, biology and social sciences.

FORMER STUDENTS AND FELLOWS

D. Adalsteinsson (Chapel Hill)
U. Adiga (GE)
J. Akao (Mathworks)
C. Anderson (UCLA)
S. Baden (UCSD)
J. Barber (Los Alamos)
P. Bernard (Maryland)
C. Borgers (Tufts)
Y. Burlakov (City Bank)
T. Buttke (Wall Street)
M. Callahan (Harvard)
C. Cameron (Wall Street)
M. Cameron (Courant)
C.C. Chang (Taiwan Nat. Univ)
A. Cheer (UC Davis)
A. Chertock (NC State)
A. Chester
D. Chopp (Northwestern)
P. Colella (LBNL)
R. Cortez (Tulane)
T. Deschamps (GE)
A. Edelman (MIT)
C. Fenimore (NIST)
R. Fattal (Tel Aviv)
D. Fishelov (Tel Aviv)
A. Fogelson (Utah)
S. Fok (Applied Networks)
S. Fomel (Univ. Texas)
A.F. Ghoniem (MIT)
D. Givon (Princeton)
A. Gottlieb (SDSU)
H. Glaz (Maryland)
C. Greengard (IBM)
G. Groh (ETH Zurich)
M. Gu (Berkeley)
J. Huh (Oak Ridge)
E. Ingerman (GE)
K. Kamrin (MIT)
I. Karassalo (Stockholm)
A. Kast (NetWork Physics)
R. Kimmel (Technion)
I. Kliakhandler (Illinois)
M. Kobayashi (IBM-Japan)
D. Krapp (Mitsubishi)
R. Krasny (Michigan)
R. Krause (Arizona)
R. Kupferman (Jerusalem)
J. Kurylo (McGill)
D. Levy (Maryland)
K. Lin (Arizona)
S.S. Lin (Taiwan)
T. Ligocki (NERSC-LBNL)
J. Lou (DOD)
R. Malladi (GE)
M. McCoy (LLNL)
R. Mendez (Japan Institute)
R. Miller (Oregon)
B. Milne (Xilinx, Inc)
M. Minion (North Carolina)
C. Nitsch (Naples)
D. Nolan (Univ. Miami)
P. Okunev (Finance)
R. Pember (LLNL)
M. Perlman (Israel)
W. Proskurowski (USC)
D. Pryor (Ikos)
E.G. Puckett (UC Davis)
J. Qi (Wall Street)
I. Ravve (Israel)
C. Rhee (Korea Automotive)
S. Roberts (Aus. Nat. Univ.)
A. Sarti (Bologna)
J.A. Sethian (UC Berkeley)
Y. Shan (D.E. Shaw)
J. Shearer (Michigan)
A. Shestakov (LLNL)
H. Shvets (UC Berkeley)
G. Sod (Tulane)
P. Stinis (Stanford)
J. Strain (UC Berkeley)
M. Tokman (UC Merced)
E. Tau (Duke Medical School)
X. Tu (Kansas)
A. Vladimirsky (Cornell)
H. Wang (UC Santa Cruz)
J. Weare (Courant)
N. Whittaker (U. Mass)
J. Wilkening (Courant)
A. Wiegmann (Kaiserslautern)
C. Willis (Wall Street)
J. Wright(USD)
J. Yu (Wall Street)
H. Zhou (North Carolina)

IMPLICIT SAMPLING

Research Highlights

The difficulty in sampling many-dimensional probability densities on a computer is a major obstacle in many fields of science, for example, in economics and finance, weather forecasting, quantum field theory, and statistics. The problem is that there are typically too many states to sample, most of which have a very small probability, and the location of the important states is not known in advance.

The Mathematics Group at LBNL has developed a methodology, implicit sampling, which finds unbiased high-probability samples efficiently through a variationally enhanced process, which zeroes in on the high-probability regions of the sample space.

One major application of their methodology so far, has been to the problem of data assimilation, where one is trying to make predictions on the basis of an uncertain model and a stream of noisy data (as one does, for example, in meteorology and in economic forecasting). As an example, suppose the coast guard is trying to locate a dinghy carrying survivors from a shipwreck (see Figure 1). The previous method of calculating the dinghy's trajectory can guess where the dinghy would be on the basis of imprecise information about wind directions and currents (this is the noisy model). Let's say a ham radio operator has heard from someone on the dinghy, but this information is also imprecise (this is a noisy datum). How does one combine the two to find an optimal guess, without first tracing out a lot of possible paths and then finding out that most of the traced paths are, in fact, made unlikely by the data? Implicit sampling makes it possible to do this quickly and efficiently.

The real-life applications carried out so far include applications to oceanography, meteorology, geomagnetism, biology, and engineering. Figure 2 shows numerical results from an application to combustion, where one is trying to estimate the state of a flame from a mathematical model supplemented by noisy data. The graph shows the mean and variance of the error as a function of the number of "particles" (i.e., samples), with two estimation methods: one in wide use, and one based on implicit sampling. The latter has much smaller errors at much lower costs.

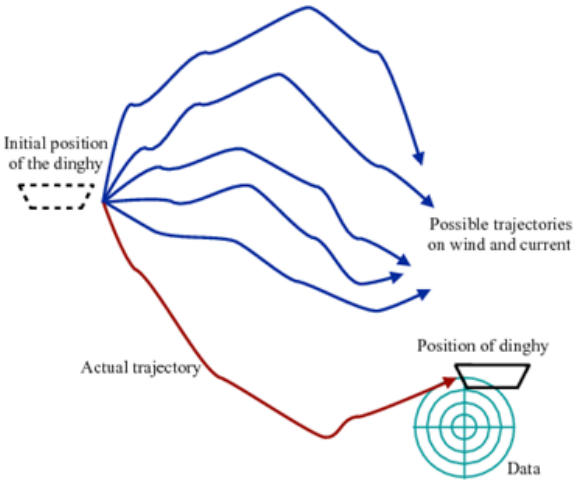


Figure 1: Example of how implicit sampling can accurately estimate the position of a dinghy (shown by the path in red) without having to trace improbable paths (shown in blue).

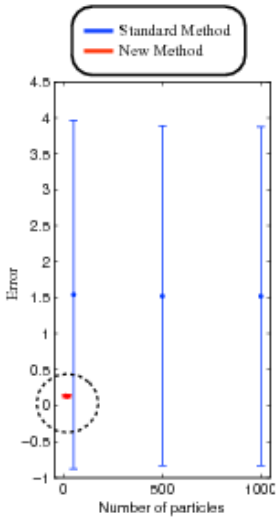


Figure 2: Mean error and standard deviation of the standard method (blue) and the implicit sampling method (red). The standard method requires thousands of particles for the estimation, and the implicit sampling method requires only a few particles.

Scientific Impact

The efficient sampling of many-dimensional probability densities will have a large impact throughout science. At LBNL, mathematics researchers are currently considering applications to quantum mechanics, statistical physics, meteorology and climate prediction, and underground hydrodynamics.

CRD Contact

A. Chorin (UCB/LBNL), M. Morzfeld, J. Kominiarczuk

MATHEMATICAL AND ALGORITHMIC METHODOLOGIES FOR COMPUTING MULTI-PHASE MULTI-PHYSICS: INDUSTRIAL FOAMS, MATERIALS, MANUFACTURING AND CELL MECHANICS

Research Highlights

Many problems involve the physics of multiply-connected moving interfaces. Some examples of these problems include liquid foams (e.g., soap bubbles, polyurethane and colloidal mixtures), and solid foams, such as wood and bone. Manufactured solid foams lead to lightweight cellular engineering materials, including crash absorbent aluminum foams; and controlling foams are critical in chemical processing.

These problems have multiple domains which share common walls meeting at multiple junctions. The domain boundaries move under forces which depend on both local and global geometric properties, such as surface tension and volume constraints, as well as long-range physical forces, including incompressible flow, membrane permeability, and elastic forces. For example, the interaction of multiple phases in fluid simulations requires a robust treatment of the interplay between membrane dynamics, fluid mechanics, surface tension, and intricate jump conditions [3], [4].

Producing good mathematical models and numerical algorithms that capture the motion of these interfaces is challenging, especially at junctions where multiple interfaces meet, and when topological connections change. Several methods have been proposed to track interfaces, including front tracking, volume of fluid, variational, and level set methods. It has remained a challenge to robustly and accurately handle the wide range of possible motions for an evolving, highly complex, multiply-connected interface that separates a large number of phases under time-resolved physics.

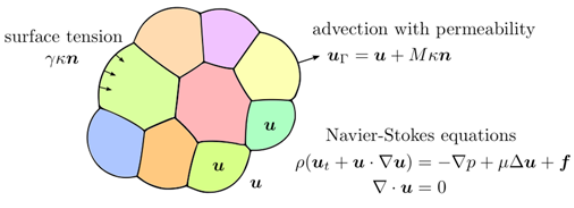


Figure 3: In a typical multi-phase fluid flow problem, the Navier–Stokes equations, determined by the velocity u of the multi-phase fluid, are coupled to the interface motion, which in turn couples to the fluid via surface tension.

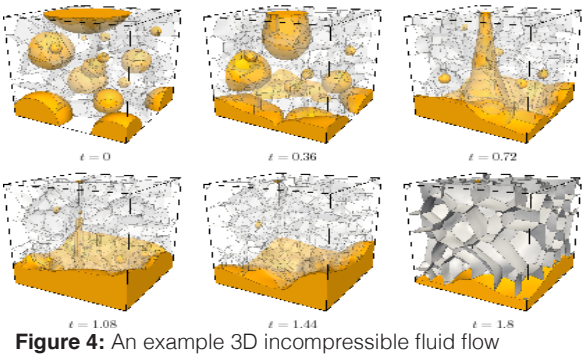


Figure 4: An example 3D incompressible fluid flow simulation, mixing phases at various densities under gravity and surface tension, where t denotes time.

CRD Math researchers have recently built a computational methodology, known as the Voronoi Implicit Interface Method, to track multiple coupled interfaces moving under complex physics constraints. Mathematically, the researchers provide a consistent formulation of 3D multi-phase physics as a single time-dependent partial differential equation, which is solved using intertwined steps of initial value upwind solvers and fast Eikonal solvers for Voronoi reconstructions. The results are algorithms which efficiently and accurately compute topological

change while coupling to fluid and solid mechanics, surface tension, and geometrical constraints. The methods are highly efficient and parallelizable, and they are now being used to compute industrial mixing and materials manufacturing.

Scientific Impact

The Math Group's algorithms allow for a more accurate understanding of a vast array of science problems and applications. The algorithms are applicable to many computing problems in both industry and research, including fluid mixing of multiple species in combustion and reactor designs, grain metal and materials growth in industrial manufacturing, semiconductor failure analysis of metal interconnect lines, liquid foam drainage in chemical and biological devices, and cell mechanics and structural stability analyses in biofuels. The original paper for this research was awarded the 2011 Proceedings of the National Academy Sciences Cozzarelli Prize for Best Paper in Applied Sciences and Engineering [2].

CRD Contact

R. Saye, B. Preskill, D. Maddix, J. A. Sethian

HIGH-ORDER METHODS FOR FLUID-STRUCTURE INTERACTION WITH APPLICATIONS TO VERTICAL AXIS WIND TURBINE SIMULATIONS

Research Highlights

The Mathematics Group has developed new numerical schemes for high-order accurate simulations of fluids and solids [1]. This has led to efficient solvers that scale well on the new generation of multi-core computer architectures. The research focuses on applying these methods to relevant applications, such as wind turbine simulations. The Math Group's studies have been used to determine optimal operating conditions and to maximize the power outputs of vertical-axis wind turbines, or VAWTs.

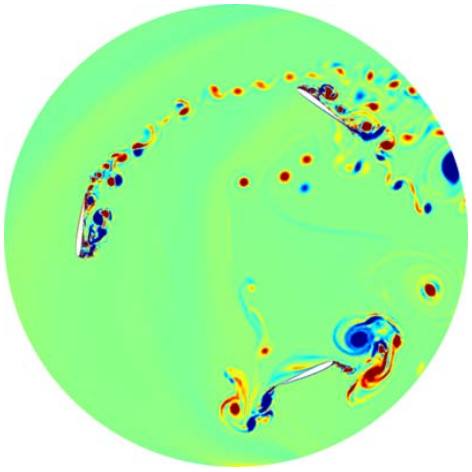


Figure 5: VAWT Simulation, 3kW unit, showing the vorticity plot of a three-blade 3kW unit.

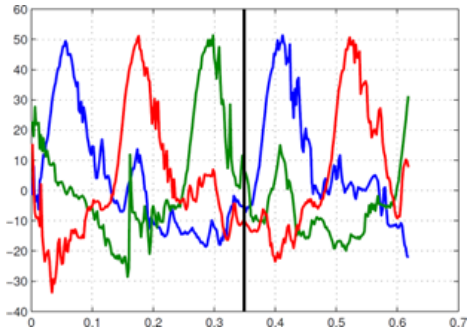


Figure 6: Power output for each of the blades vs. time.

In more detail, the Math Group has developed new numerical methods for solving the equations that arise from these physical application problems. One recent example is the Line-DG method, which is a class of high-order schemes that are 1–2 orders of magnitude more efficient than current approaches. They also developed implicit-explicit timestepping schemes for an accurate fluid-structure simulation without coupled Jacobian matrices. And finally, they improved the parallel block-ILU(0) and multigrid preconditioners that are used in the solvers, which scale well up to thousands of CPU cores.

Scientific Impact

Math researchers have implemented the methods in the 3DG software package, which is used by a number of research groups at UC Berkeley, LBNL, and MIT, for simulations in areas such as wind turbines, flapping flight, and aeroacoustics.

Their techniques replace the traditional low-order techniques for a range of simulation tools. They provide reliable levels of accuracy for challenging problems, involving wave propagation, multiple scales, multi-physics, and nonlinear interactions.

CRD Contact

P.-O. Persson

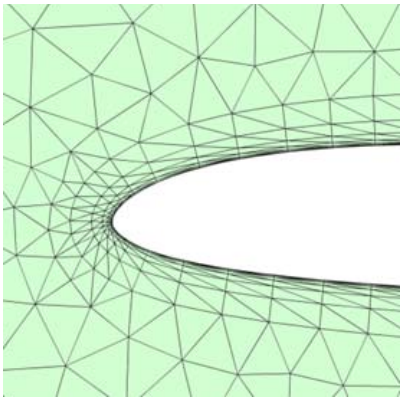


Figure 7: Zoom-in of a blade-tip of the computational high-order mesh, showing the stretched elements required to resolve the turbulent boundary layer.

LONG-TIME DYNAMICS AND OPTIMIZATION OF NONLINEAR PARTIAL DIFFERENTIAL EQUATIONS

Research Highlights

The Math Group is developing new computational tools to study dynamic problems with time-periodic forcing or symmetry. A key step in this research is to devise adjoint-based optimization techniques suitable for large-scale systems governed by nonlinear evolutionary partial differential equations. Because the solutions of the partial differential euqations are time-periodic, Floquet theory can be used to characterize their stability. The framework developed to compute time-periodic solutions is easily adapted to compute Floquet multipliers in parallel on a supercluster. These algorithms have been demonstrated in a wide array of application areas, including water waves, viscoelastic fluids, gas dynamics, nonlinear optics, and solid mechanics.

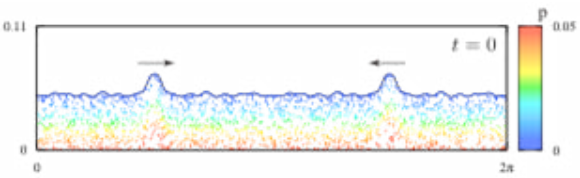


Figure 9: Stability of rogue waves.

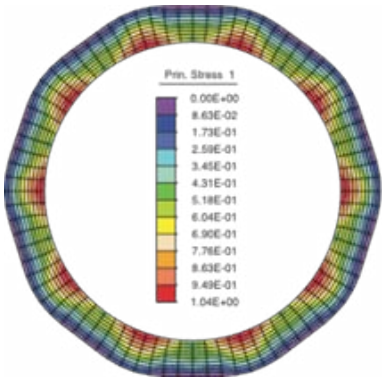


Figure 8: Dynamics of rolling tires. One of the applications in which time-periodic dynamics are important.

Figure 10: Parallel geometric multigrid.

Scientific Impact

Using these methods, the Math Group has solved a 60 year old open question on the limiting form of extreme standing water waves due to Penney and Price (1952) and G.I. Taylor (1953). Their question was whether the largest-amplitude standing wave would form a 90 degree corner angle, providing an example of finite-time singularity formation in Euler's equations in which the solution regains smoothness after the event. The new result or answer is that the wave crests of extreme standing waves do not sharpen to a corner in a self-similar fashion as predicted, but instead develop an oscillatory structure on small scales near the crest.

The Math Group is currently using these methods to study stress states in rolling vehicle tires with treads, and to study the low-dimensional nature of standing water waves by combining high-resolution numerical simulations with reduced-order modeling techniques. They also use these methods to gain insight about the onset of elastic turbulence in viscoelastic mixing problems, and to optimize over-a-cycle in complex fluid-structure problems, such as flapping flight. These techniques provide vast improvements in efficiency and robustness through parallelism and reformulation as minimization problems.

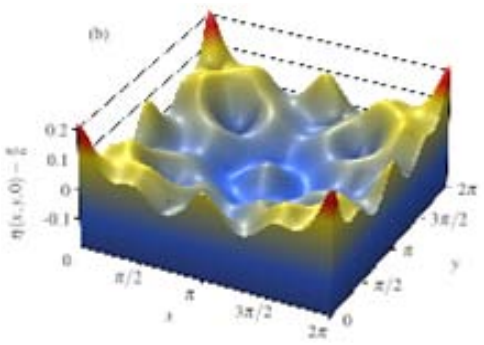


Figure 11: Standing water waves and pattern formation. Another application example for long-time dynamics and optimization.

CRD Contact

J. Wilkening

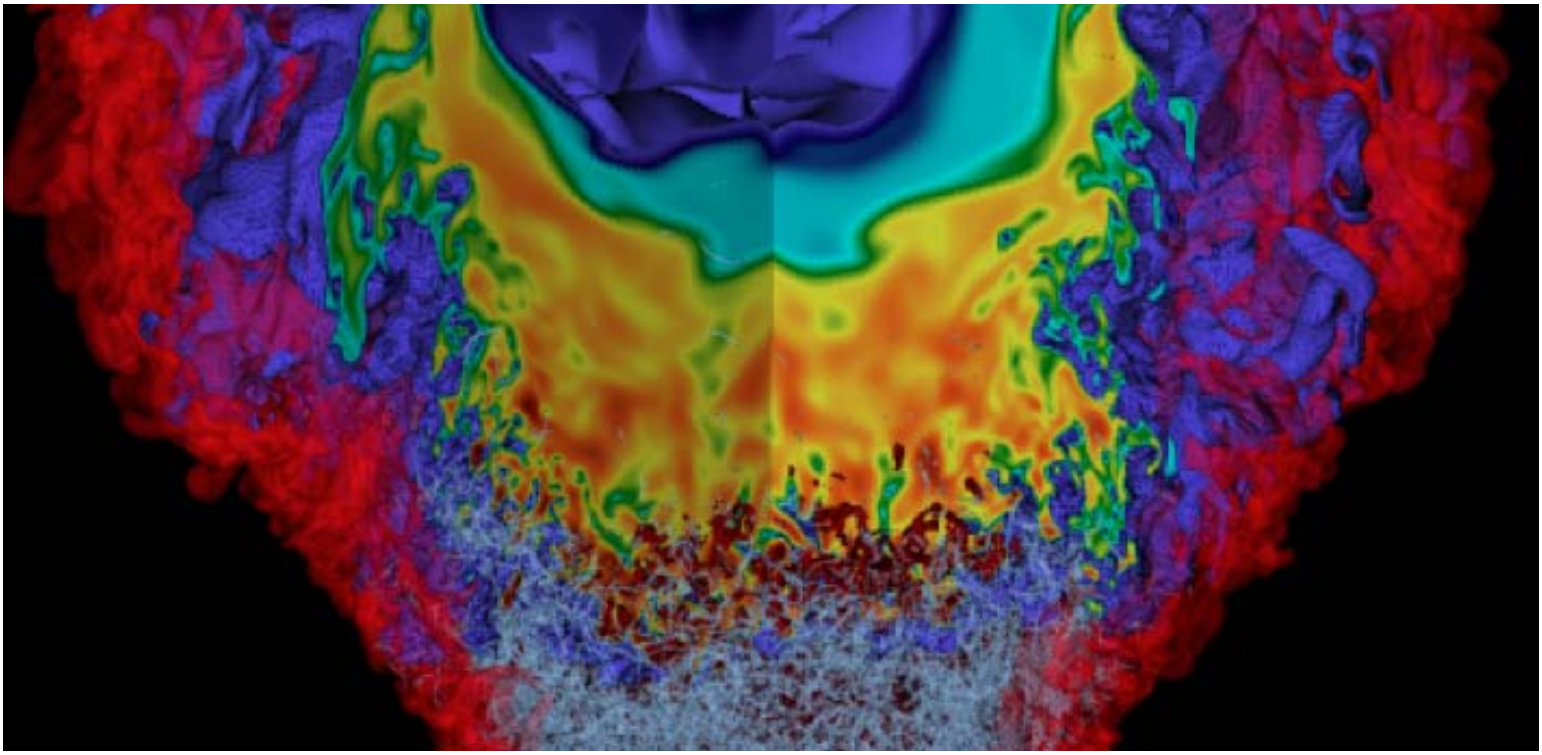
REFERENCES

[1] Per-Olof Persson. "A Sparse and High-Order Accurate Line-Based Discontinuous Galerkin Method for Unstructured Meshes." Journal of Computational Physics, 233:414-429, 2013.

[2] R. I. Saye and J. A. Sethian. "The Voronoi Implicit Interface Method for computing multi-phase physics." Proceedings of the National Academy of Science USA, 108(49):19498-19503, 2011.

[3] R. I. Saye and J. A. Sethian. "Analysis and applications of the Voronoi Implicit Interface Method." Journal of Computational Physics, 231(18):6051-6085, 2012.

[4] Robert I. Saye and James A. Sethian. "The Voronoi Implicit Interface Method and Computational Challenges in Multi-phase Physics." Milan Journal of Mathematics, 80:369-379, 2012.



CENTER FOR COMPUTATIONAL SCIENCES AND ENGINEERING

The Center for Computational Sciences and Engineering (CCSE) develops and applies advanced computational methodologies to solve large-scale scientific and engineering problems arising in the Department of Energy (DOE) mission areas involving energy, environment, and industrial technology. The primary focus of CCSE researchers is on designing algorithms for multi-scale, multi-physics problems described by nonlinear systems of partial differential equations, and in developing implementations of algorithms that target current and next-generation massively parallel computational architectures. Sample application areas include combustion, porous media flow, fluctuating hydrodynamics, atmospheric modeling, cosmology and astrophysics. CCSE researchers work collaboratively with application scientists to develop state-of-the-art solution methodologies in these fields.

CCSE is also the home of BoxLib, a software framework for massively parallel block-structured adaptive mesh refinement (AMR) codes. BoxLib is used for the development of new algorithms and is the basis for many mature AMR applications.

OVERVIEW OF HIGHLIGHTED PROJECTS

CCSE's highlighted projects include an overview of the Center for Exascale Simulation of Combustion in Turbulence (ExaCT) and a discussion of their work on high-order multi-physics coupling strategies.

RESEARCH STAFF

JOHN BELL is the Department Head of MACS (the Mathematics and Computational Sciences Department) and the group leader of CCSE. He received his B.S. from MIT and his Ph.D. from Cornell University, both in mathematics. Before coming to LBNL in 1996, Bell was the group leader of the Applied Math Group at Lawrence Livermore National Laboratory. His research interests include development and analysis of numerical methods for partial differential equations and stochastic systems. He contributed to research in finite volume methods, low Mach number methods, adaptive mesh refinement, numerical methods for stochastic partial differential equations, and parallel computing. He has worked on applications of these methods in a number of areas including astrophysics, combustion, subsurface flow, and micro-fluidics. He is also the deputy director of the ExaCT Co-Design Center.

ANN ALMGREN is a staff scientist in CCSE. She received her B.A. in physics from Harvard University and her Ph.D. in mechanical engineering from UC Berkeley. She visited the Institute for Advanced Study in Princeton, NJ, prior to joining Lawrence Livermore National Laboratory, then moved to LBNL in 1996. Her primary research interest is in computational algorithms for solving numerical partial differential equations for fluid dynamics in a variety of application areas. Her current projects include the development and implementation of new multi-physics algorithms in high-resolution adaptive mesh codes that are designed for the latest multi-core architectures. She is particularly interested in low-Mach number methods for atmospheric and astrophysical applications.

KAUSHIK BALAKRISHNAN is a postdoc in CCSE. He received his B.Tech in aerospace engineering from the Indian Institute of Technology, Madras, India, and his M.S. and Ph.D. in aerospace engineering from the Georgia Institute of Technology, Atlanta. He joined CCSE in 2010 and works on numerical methods for simulation of explosions, reactive flows, multi-phase flows, and fluctuating hydrodynamics.

ANUJ CHAUDHRI is a postdoc in CCSE. He received his B.E. in mechanical engineering from Osmania University in India, his M.S. from the University of Tennessee, Knoxville, and his Ph.D. in applied mechanics and mechanical engineering from the University of Pennsylvania. Before joining LBNL in 2012, he was a postdoc in theoretical biophysics at the University of Chicago and a visiting scientist at Genentech. His research interests include statistical physics, biophysics, multi-scale modeling, stochastic processes and developing numerical schemes for stochastic partial differential equations.

MARC DAY is a staff scientist in CCSE. He received his B.S. in nuclear engineering from UC Berkeley and his Ph.D. in plasma physics and nuclear engineering from UCLA. Day joined CCSE in 1995, first through a joint postdoc appointment with the NERSC computing center, and then full-time with CCSE in 1996. His primary focus has been developing and implementing block-structured adaptive numerical methods for a variety of problems in reacting fluids, with applications in astrophysical flows, terrestrial combustion, and multi-phase sub-surface flows.

MIKE LIJEWSKI is one of the principal architects of the BoxLib framework, and is an expert in many aspects of high performance computing. He regularly designs and implements new features in the BoxLib code while maintaining and improving the efficiency and performance of existing features across a range of platforms. He received a B.A. in math from Illinois Wesleyan University and an M.A. in applied mathematics from University of Maryland, College Park. He started with CCSE after working at NSWC (Naval Surface Warfare Center), Cornell Theory Center, NASA Goddard, The Royal Hong Kong Jockey Club, and Rogue Wave Software.

ANDY NONAKA is a research scientist in CCSE. He received his B.S. in electrical engineering from University of the Pacific and his Ph.D. in engineering applied sciences from UC Davis. Nonaka joined CCSE as a postdoc in 2007 and became a research scientist in 2011. His interests include the development of high-order spatial and temporal discretizations to solve problems described by compressible, incompressible, and low-Mach number models in combustion, astrophysics, and fluctuating hydrodynamics problems.

WEIQUN ZHANG is a project scientist in CCSE. He received his B.S. in physics from the University of Science and Technology of China and his Ph.D. in astronomy and astrophysics from UC Santa Cruz. He was a postdoc at Stanford University and New York University before he joined CCSE in 2010. He has worked on numerical methods for relativistic magnetohydrodynamics and radiation hydrodynamics, and applications in astrophysics. His recent projects include the development of high-order narrow stencils for second-order derivatives and the development of a new combustion code that solves compressible Navier–Stokes equations for viscous multi-component reacting flows.

VINCE BECKNER is a Computer Systems Engineer and programmer in CCSE. He received his B.S. in mechanical engineering from the University of Florida and his M.S. in computer science from the California Polytechnic State University, San Luis Obispo. He started with the group in 1992 when it was at the Lawrence Livermore National Laboratory, and moved with the group to LBNL in 1996 when it became CCSE. His interests include large-scale parallel programming, performance tuning, I/O optimization, data analytics, and visualization.

MATTHEW EMMETT is a postdoc in CCSE. He received his B.S. in mathematical physics from Simon Fraser University, his M.S. in applied mathematics from the University of Calgary, and his Ph.D. in applied mathematics from the University of Alberta. He was a postdoc at the University of North Carolina at Chapel Hill before joining CCSE in 2012. His research interests currently focus on numerical schemes for evolving time-dependent partial differential equations, including iterative, multi-level timestepping schemes, and methods for their parallelization across the time domain.

CENTER FOR EXASCALE SIMULATION OF COMBUSTION IN TURBULENCE

Research Highlights

Extrapolating current trends in computer architecture leads to machines that have unrealistic power requirements, are intractable to build and maintain, and are too expensive. Reducing power and cost leads to many-core architectures with reduced memory per core and high costs for data movement. Effective utilization of these types of architectures requires that scientists fundamentally rethink their simulation methodology, while simultaneously considering the algorithms used to model the relevant physics. The scientists have to consider the programming models used to implement those algorithms and how different architectural features affect the performance of the simulation methodology. These issues need to be addressed as part of an iterative co-design loop that includes mathematicians, computer science researchers and vendors to ensure that scientists can effectively utilize future exascale architectures.

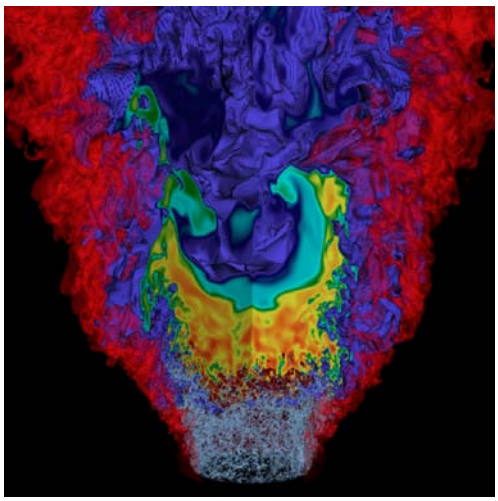


Figure 1: Simulation of emissions from a low swirl injector fueled by hydrogen at laboratory conditions. Exascale computing will enable these types of simulations to be performed at conditions found in a gas turbine used for power generation.

The goal of the Center for Exascale Simulation of Combustion in Turbulence, or ExaCT, is to address these co-design issues in the context of combustion simulation, one of the critical DOE applications that would benefit significantly from a substantial increase in simulation capability. ExaCT is a DOE-funded collaboration across multiple national laboratories and universities. The US is at a critical crossroad where issues of energy security, climate change, and economic competitiveness are driving a need for new high-efficiency, low-emissions combustion systems for transportation, power generation and industrial processes that are able to burn an evolving set of new fuels. Simulations are urgently needed to facilitate the development of these new systems and fuels. Meeting this need requires simulations with sufficient chemical fidelity to distinguish the combustion properties of alternative fuels in the combustion regimes that are relevant in novel combustion systems.

During the first year of the project, CCSE, as part of the ExaCT team, focused on three basic tasks. The first task was a detailed characterization of the behavior of current combustion simulation methodology. This study, performed at the request of the vendor community, focused on measuring the instruction mix, memory requirements, byte-to-flop ratios and communications of current codes. These measurements provided an initial baseline for defining hardware requirements and helped to obtain an initial characterization of how algorithms are likely to need change to effectively utilize potential exascale architectures.

The second task the ExaCT team pursued was the development of proxy applications that provide simplified codes representative of the characteristics of combustion simulation. These proxy applications are designed to capture key elements of combustion algorithms while retaining sufficient simplicity. By retaining simplicity, proxy applications can be easily implemented in a new programming model or tested on prototype hardware. During the first year, CCSE researchers constructed a compressible combustion code, SMC, that is a fully functional reacting flow code based on the compressible Navier–Stokes equations. They also created a simplified version of the SMC code, CNS, that solves the compressible Navier–Stokes equations without species transport or kinetics and assumes constant properties.

And finally, they developed a multigrid proxy that solves variable coefficient elliptic partial differential equations using various forms of multigrid relaxation. The first two proxy applications are representative of the types of stencil operations and single-point physics computations used in reacting flow simulations. The multigrid application is representative of the linear systems that need to be solved as part of a low Mach number formulation of the reacting flow equations.

The third task the CCSE team, in collaboration with members of FTG, pursued was the development of tools needed to model the performance of algorithms and make estimates of the impact of different hardware design choices on performance. The starting point for this analysis is a methodology for compactly representing the behavior of a particular algorithm. The initial focus was on the representation of stencil-like operations. Researchers developed an approach that, for each loop in an algorithm, captures the floating point operations, data dependencies, memory traffic and communication associated with that loop. This representation can be generated from compiler analysis or it can be generated from a mathematical description of the algorithm. This resulting description can then be used as part of an analytic performance model that captures floating-point behavior, cache and memory behavior, and communication characteristics of the algorithm. The performance model also provides an analysis of data-dependencies between loops. Using a number of parameters that characterize a particular architecture, ExaCT researchers can then investigate sensitivities of the performance to various machine characteristics. ExaCT has also been able to use this methodology to explore the effect of various code transformations on performance. The data from the code representation can also be used to drive a network simulator, such as the SST Macro, to model the impact of realistic network effects, such as contention, on performance. Thus far, this methodology has been applied to both the CNS and SMC codes and results have been disseminated to the vendor community as well as the broader DOE exascale research community.

Scientific Impact

The work being done in this project will ensure that future exascale machines will be able to perform large-scale reacting flow simulations effectively. Simulations of this type will enable combustion scientists to perform first principles direct numerical simulations with detailed chemical kinetics in relevant thermochemical and flow regimes to glean fundamental insights into turbulence-chemistry interactions at high pressure. The results will form the basis of science-based predictive models, which will be used to design advanced combustors that meet pollutant and greenhouse gas emissions targets, reduce dependence on petroleum and promote economic competitiveness.

More broadly, the methodology being developed as part of ExaCT represents core algorithm technology used across a wide range of DOE applications including astrophysics, climate, fusion, materials science and subsurface flow, as well as a number of national security and defense applications. Consequently, results from this project will also have a direct impact in these other areas.

CRD Contacts

John Bell (ExaCT Deputy Director), Jon Shalf

COLLABORATORS

Jacqueline Chen (ExaCT Director, SNL), Scott Klasky (ORNL), Pat McCormick (LANL), Bob Moser (UT Austin), Dan Quinlan (LLNL)

HIGH-ORDER MULTI-PHYSICS COUPLING STRATEGIES

Research Highlights

Researchers in CCSE are developing new numerical coupling strategies to solve complex multi-physics problems that are described by systems of time-dependent partial differential equations containing vastly different length and time scales. Current application areas include combustion, atmospheric modeling, astrophysics, and fluctuating hydrodynamics. The mathematical and algorithmic developments expand upon the current state of spectral deferred correction (SDC) theory. SDC methods were originally developed to solve time-dependent ordinary differential equations with an arbitrarily high order of accuracy. More recent developments demonstrate that for multi-physics problems, different processes can be integrated at different rates while iteratively retaining full coupling between processes. This allows for each component to be integrated in time using a numerical scheme well-suited for the mathematical character of that component. SDC-based approaches can also leverage the spatial adaptivity provided by adaptive mesh refinement (AMR), and can be used to develop new efficient time-parallel techniques, allowing CCSE researchers to create novel high-order AMR algorithms that exhibit strong coupling between mesh hierarchies and efficient parallelism across the time domain.

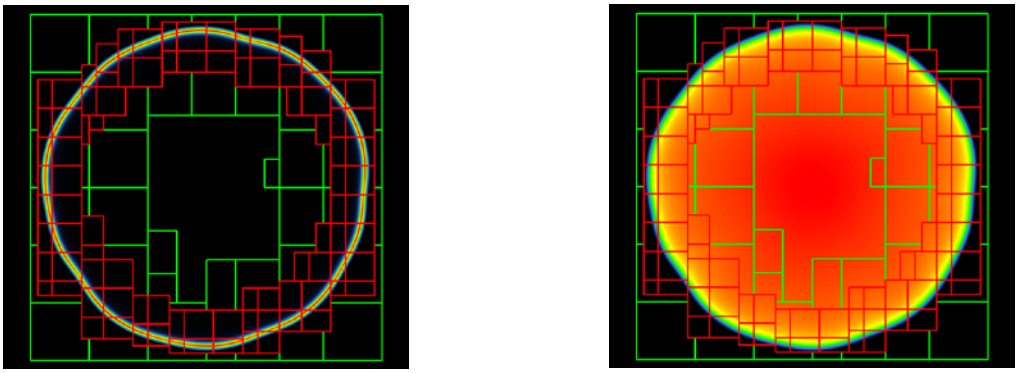


Figure 2: Hydrogen flame simulation using AMR with 9 chemical species and 21 reactions. In the CCSE researchers' approach, they treat advection explicitly with a second-order Godunov method, diffusion semi-implicitly with a Crank–Nicolson discretization; and reactions are integrated using a variable-order backward differentiation formula temporal integrator publicly available in the VODE package. The green and red boxes represent grids containing cells of increasingly higher resolution. (Left)Contour of hydrogen flame front. (Right) Temperature profile: red = 2000K, black = 198K.

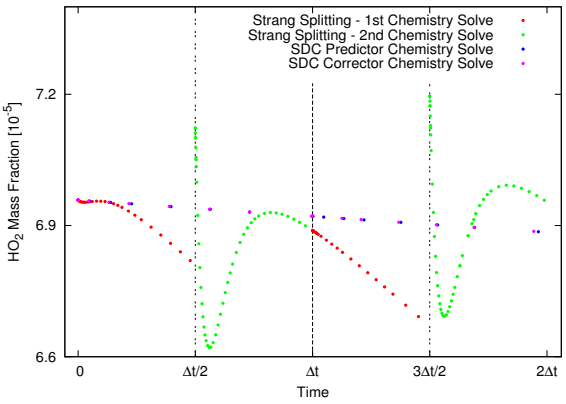


Figure 3: H evolution of the mass fraction of HO_2 in one particular cell over two timesteps for a Strang splitting and an SDC-type algorithm. Each dot represents a point in time where VODE evaluates the chemical burning rates. The red and green trajectories represent two separate calls to VODE required by the Strang splitting algorithm; and the discontinuity between the trajectories represents the solution update in the advection-diffusion step of the Strang splitting algorithm. The blue and pink trajectories represent the predictor and corrector calls to advance the thermodynamic variables in the SDC algorithm. Note that the SDC scheme is much more efficient in terms of the number of VODE evaluations.

Scientific Impact

The work being done in this project will ensure that researchers will be able to solve complex multi-physics systems efficiently and accurately on current and future massively parallel architectures. CCSE researchers are creating an open SDC framework that will allow others to incorporate SDC concepts into a wide variety of settings. For example, SDC methods can be incorporated into existing multi-physics and AMR codes that are used throughout many scientific domains—such as tsunami modelling, climate modelling, turbulence, etc.—to create more accurate and efficient codes with increased parallelism. These approaches will shorten the time required for simulations to complete and also utilize massively parallel machines to their fullest potential.

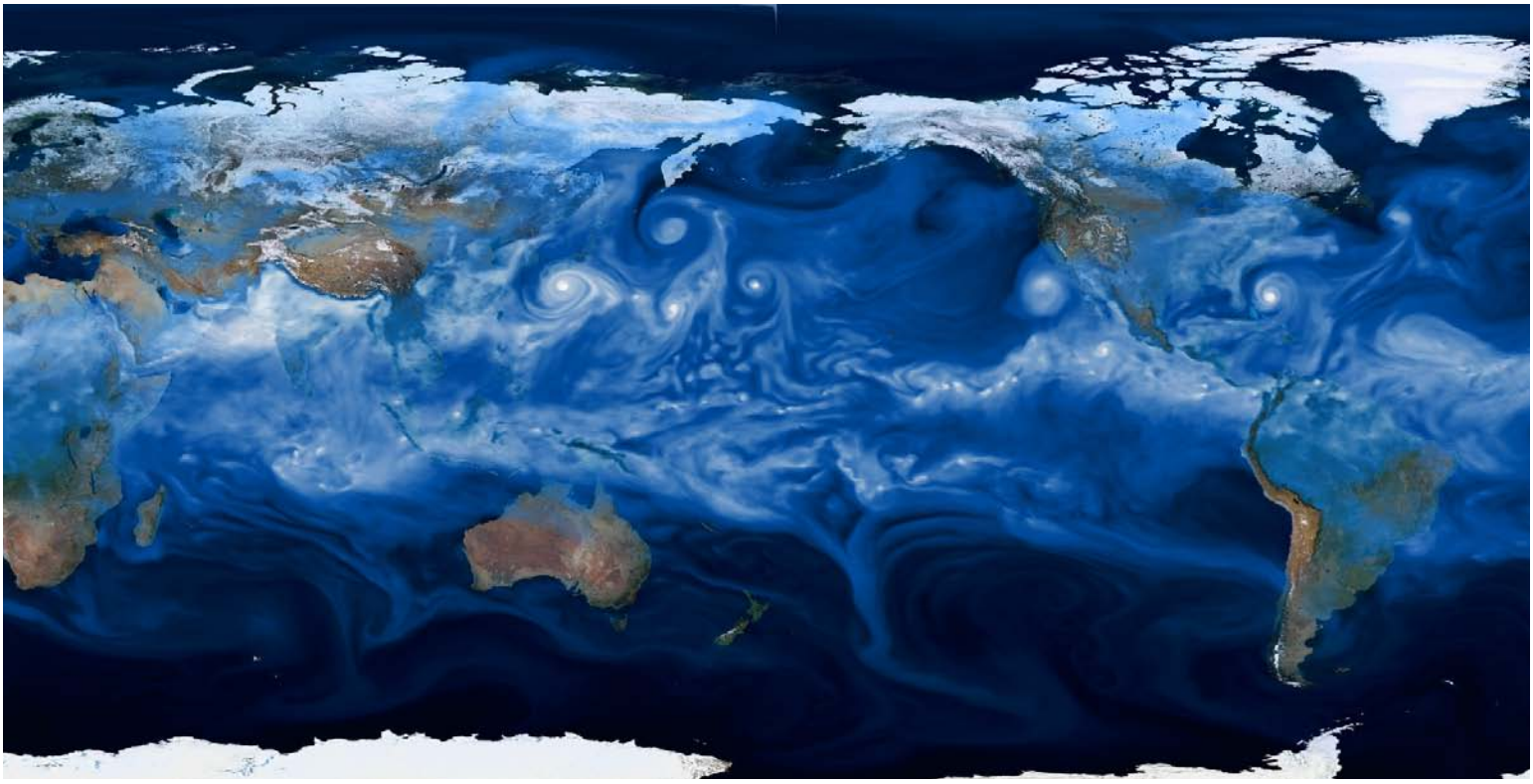
CRD Contacts

John Bell, Andy Nonaka, Matt Emmett, Marc Day, Ann Almgren, Weiqun Zhang

COLLABORATORS

Michael Minion (Stanford University), Michael Zingale (Stony Brook University)

Building a Bigger Picture



COLLABORATIVE PROJECTS ACROSS CRD

BERKELEY ICE SHEET INITIATIVE FOR CLIMATE EXTREMES (BISICLES)

Research Highlights

Recent observations show that the Greenland and Antarctic ice sheets are losing mass at increasing rates. If recent trends continue, ice sheets will make a dominant contribution to the 21st century sea-level rise (SLR), far exceeding the projections of the IPCC's Fourth Assessment Report, AR4. Growing ice mass losses not only could raise sea level, but the mass loss could also affect other parts of the climate system, such as the Atlantic Meridional Overturning Circulation and its associated poleward heat transport, through increased freshwater discharge to high-latitude oceans. In the IPCC's AR4 report, the computational state-of-the-art modeling capability for ice sheets was deemed inadequate due to the inability to make realistic projections on SLR. In response, the DOE-funded **I**ce **S**heet **I**nitiative for **CL**imate **E**xtreme**S**, or ISICLES, program aimed to bring ASCR-developed computational expertise to bear on the ice sheet modeling problem.

The dynamics of ice sheets span a wide range of scales. Correctly resolving the dynamics of localized regions, such as grounding lines and ice stream shear margins, requires extremely fine resolutions (finer than 1 km in places). In particular, resolving the dynamics of the grounding line (where the land-based ice sheet meets the ocean and begins to float) is critical to understanding the dynamics of the ice sheet. Modeling an entire continental-scale ice sheet at such resolutions is impractical or impossible with current computational resources. At the same time, such fine resolution is unnecessary over large dynamically quiescent regions, which makes ice sheet modeling an ideal candidate for adaptive mesh refinement (AMR).

Since 2009, ANAG has developed a scalable AMR ice sheet modeling code built on the Chombo framework called Berkeley-ISICLES (BISICLES), which has become a part of the Community Ice Sheet Model (CISM). With a dynamical core based on the vertically integrated model of Schoof and Hindmarsh (2010), BISICLES can resolve dynamically important regions at the sub-kilometer scale while using a much coarser resolution

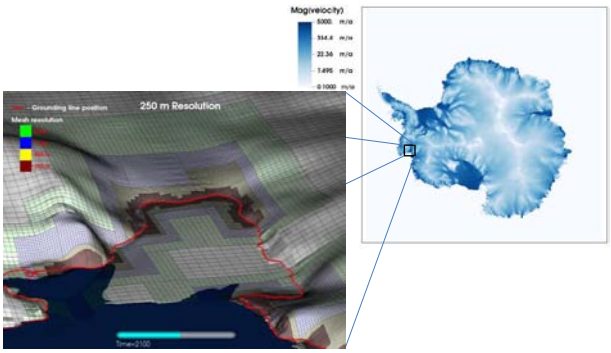


Figure 1: Schematic showing computed ice velocity for Antarctica (right), and (left) meshing and grounding line location for the Pine Island Glacier.

where appropriate [4]. As an example, Figure 1 shows a computed velocity field for Antarctica. The inset shows the adaptive meshing and grounding line location for the Pine Island Glacier in the Amundsen Sea Sector.

As a part of the BISICLES project, an effort was made by FTG to improve the performance of the modeling code using auto-tuning techniques. Auto-tuning is a feedback-driven, semi-automated, optimization technique that attempts to find the best implementation for a given platform/code combination by automating the code transformation and evaluation process to find the best implementation and parameters. FTG collaborated with a visiting scientist, Matthias Christen, from the University of Basel, in order to quantify the benefits of auto-tuning in the multigrid (MG) algorithm in the context of AMR applications, like BISICLES.

FTG has developed a number of semi-automatic techniques for optimizing stencils, including 27-point stencils [5]. Although dense stencils, like the 27-point, are not currently a part of BISICLES, they may become a part of ice sheet modeling codes as the codes transition to 3D. FTG collaborated with Christen on the development of code transformation and auto-parallelization technology, and evaluated it on a number of

platforms [3]. They obtained the following insights. The tiny grids in the AMR codes (32^2 or smaller in the depth-integrated BISICLES code) easily fit in modern caches and thus, see no benefit for parallelization or cache blocking. Moreover, AMR MG codes typically apply one operation at a time to the set of boxes within a level. There is no temporal locality to exploit, and therefore no need for temporal blocking.

Finally, analysis showed that it was theoretically possible to reorder and fuse the application of operators to improve, reuse and cache locality. These ideas did not get incorporated into the tools mentioned above. To that end, FTG set about constructing a simple test harness that would allow them to evaluate the potential benefit of this style of optimization, as well as quantifying the benefits of communication-avoiding (time skewing/temporal blocking) for Gauss–Seidel, Red-Black relaxations and advanced many-core architectures in the MG V-cycle [17].

Results showed that the reordering and fusion of operations could yield nearly a factor of two in the benchmark while communication-avoiding and many-core architectures could improve the V-cycle performance by as much as an additional 3.5X. As this was an *ad-hoc*, manual, proof-of-concept, FTG proposed an X-Stack2 project, called XTune, to automate this process. XTune has subsequently been funded and will address some of the implementation performance challenges of AMR MG on advanced architectures in a programmer-friendly, productive manner.

Scientific Impact

BISICLES participated in the MISMIP3D grounding line dynamics model inter-comparison [14] in which the accuracy and effectiveness of ANAG's AMR model was demonstrated, as was the need for very fine spatial resolution to resolve the dynamics of ice sheets near grounding lines.

In a study by Payne et al. [15], the BISICLES code was employed to explore the effects of uncertain climate forcing on projections of the West Antarctic ice sheet over the 21st and 22nd centuries. The use of the BISICLES model ensured that the grounding line dynamics were sufficiently resolved to accurately model the glacial retreat scenarios being explored. As an example, Figure 2 shows a plot of the SLR contribution from the Amundsen Sea sector in Antarctica for a typical climate scenario examined by Payne et al. for different fine mesh resolutions. Almost no contribution is present with a uniform 4 km mesh, while at the 500 m resolution, one can see the asymptotic regime with a much more credible prediction. The results in this work were used to help inform the IPCC AR5 report and its discussion of sea-level rise.

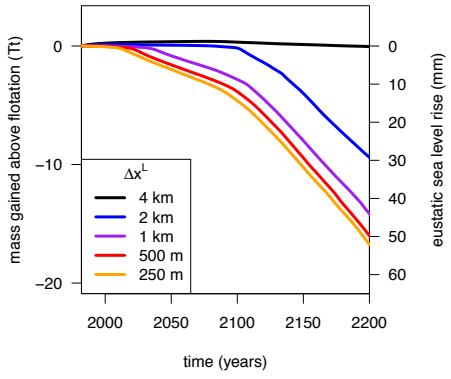


Figure 2: Contribution to sea-level rise from the Amundsen Sea region in Antarctica for one possible climate scenario at different mesh resolutions.

CRD Contacts

Esmond Ng (LBNL PI), Dan Martin (BISICLES technical lead), Sam Williams (auto-tuning technical lead)

LAS ALAMOS NATIONAL LABORATORY COLLABORATORS

William Lipscomb, Stephen Price, Doug Ranken

UNIVERSITY OF BRISTOL (U.K.) COLLABORATORS

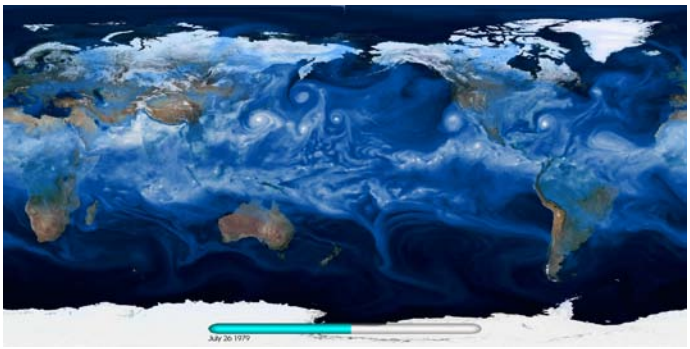
Stephen Cornford, Antony J. Payne

EXTREME-SCALE ANALYSIS OF CLIMATE DATA

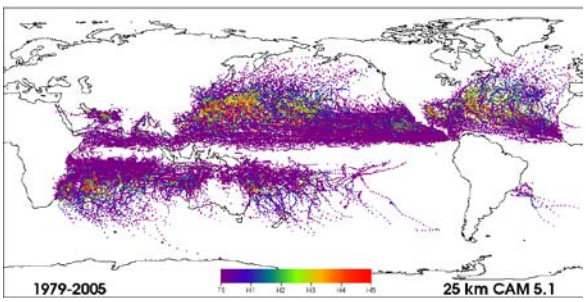
Research Highlights

Modern climate simulations produce massive amounts of data. Commonly used models such as the Community Earth System Model (CESM) and Weather Research and Forecasting model (WRF) routinely produce tens of gigabytes of data, and the next generation of global cloud resolving simulations will produce terabytes of data on a per-timestep basis. Coordinated multi-model ensembles of climate model simulations, such as the Coupled Model Intercomparison Project (CMIP3), produce hundreds of gigabytes today, while the CMIP5 code presently produces around 100TB of data for a single 25-year simulated period, a volume of data that will explode when the code is run in ensemble. These ensembles are a crucial tool for international and national organizations, such as the Intergovernmental Panel on Climate Change (IPCC), tasked to assess the human role in climate change and to assess potential adaptation and mitigation strategies for reducing our role in climate change. Legacy software tools, such as those developed in the 1980s and used for the past three decades, have served well, but fall short of what is needed to meet the challenge of growing data size, evolving computational platforms, and ever more sophisticated lines of climate science inquiry.

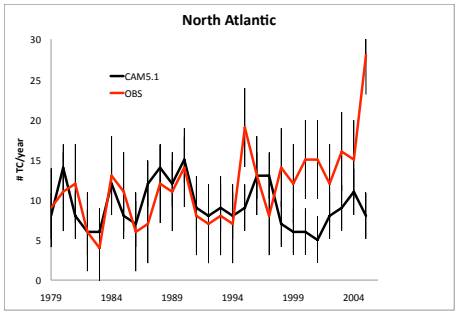
Funded through DOE's Office of Biological and Environmental Research (BER), this CRD research team includes staff from several different groups (Visualization, Scientific Data Management, Scientific Computing). CRD researchers are



(a) Visualization of water vapor in 0.25° CAM5 output.



(b) Cyclone tracks computed from 100TB of CAM5 output, colored by storm categories on the Sapphir-Simpson scale.



(c) Comparing the annual count of observed vs. modeled cyclones in the Atlantic basic.

Figure 3: In this example, a high resolution atmospheric code produces massive amounts of data and the science objective is to study the number of cyclones that form over time. One timestep of model output is visualized (a). The TECA code is run in parallel to identify cyclones and their tracks over time (b). These results are compared to the counts of cyclones observed over the same time period (c). The data source for this particular problem was about 100TB of CAM5 output processed on 80,000 cores in about two hours, compared to an estimated serial processing time of about 10 years.

engaged in research to produce next-generation visualization and analysis software infrastructure that runs on large-scale parallel platforms, so as to meet the challenges of exploding data size, and to provide the ability to perform advanced analysis of the sort needed to better understand a changing climate. This CRD team consists of climate, computer, and computational scientists working together so that climate science research questions are answered and consequently this research is driving the evolution of software and algorithms, which are proven on tough, large-scale climate problems, then released as software products to the climate science community.

One of the software products developed by this team is a new analysis framework named Toolkit for Extreme Climate Analysis (TECA). TECA is sufficiently flexible because it performs feature identification, tracking, and analysis of several different types of extreme weather events, such as Tropical Storms (see Figure 3) and Atmospheric Rivers (see Figure 4), and is capable of running in parallel on tens of thousands of cores to reduce runtime from months or years to minutes or hours [16, 1]. The basic idea is that structured data is partitioned in space and or time, the infrastructure applies a user-specified analysis kernel to detect a feature, the results (features) are then gathered for aggregation and analysis. One particular design strength of this approach is the adaptability to different feature detection scenarios that result from the user-definable detection kernel.

A second software thrust from this team is infrastructure that can support the parallel execution of R-based codes. The long-term objective is to have the infrastructure manage the execution of user-supplied R code, including the marshaling of data, gathering of results, and connections with the surrounding software environment. This approach is useful in applying new statistical techniques based on Extreme Value Theory to study precipitation and temperature extremes in North America [8]; reflecting new climate science results that stem from advances in parallel analysis technology. The new infrastructure is deployed in VisIt so that R-based tools can take advantage of VisIt's scalability and its existing infrastructure for loading, processing, analyzing, and visualizing data. Recent work has added the ability to support parallel operations that involve collective communication, so as to support a wide variety of R-based analysis as well as those that are embarassingly parallel, along with the ability to take advantage of multi-core processors.

Scientific Impact

Climate scientists are now using this capability to perform climate data analysis work on large computational platforms that was previously impractical or impossible due to sheer data size. The accomplishments range from drastic reductions in processing time to new climate science results. One performance example is the application of TECA to detect and analyze Tropical Cyclones in 26 years of 0.25° CAM5 output in about an hour on 80,000 cores compared to an estimated 10 years of processing time if run in serial.

In terms of climate science discoveries, TECA is presently in use by LBNL climate scientists. These scientists conduct several different types of climate model intercomparison projects that focus on comparing extreme weather events in the historical record with different models run at varying resolutions and climate scenarios. Early results, some of which are under review for publication at this time, show that increasing model resolution produces extreme weather events (cyclones) that more closely match the historical record. Other results, also not yet published, show how some models have biases that result in a better match with the historical record by region, such as one model matching the onset and count of storms in the Atlantic Basin better than those in the Pacific Basin.

As a starting point, TECA is an infrastructure upon which new capabilities may be added. TECA has been adapted for use in detecting many different types of extreme weather (tropical cyclones, extra-tropical cyclones, atmospheric rivers) simply by changing a small detection kernel. In addition, diagnostic capabilities have been added, such as

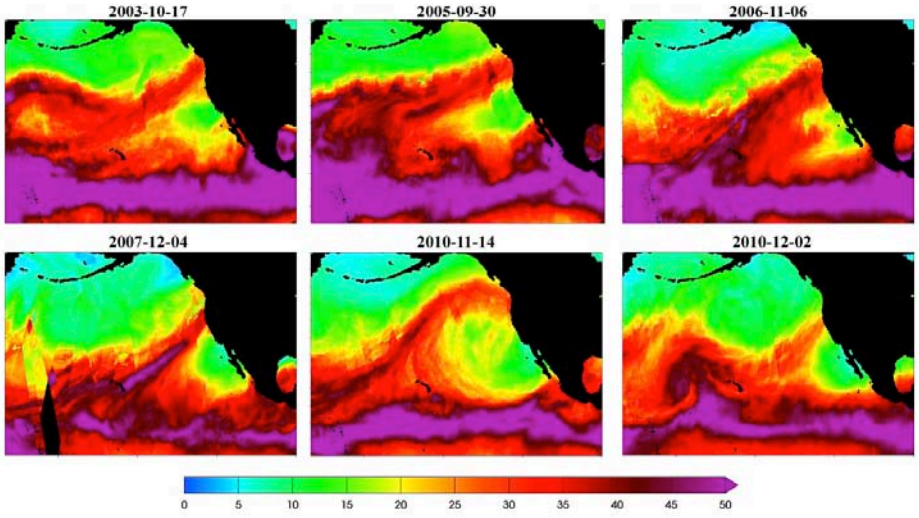


Figure 4: Sample CAM5 atmospheric river (AR) detections from TECA's parallel code. Note that this method is able to detect AR events of different shapes and sizes. A selection of events in the SSM/I satellite record detected by TECA's pattern recognition technique shows atmospheric rivers. Note that they share a similar horseshoe-like pattern but vary greatly in their spatial extent. Events that pass over Hawaii, such as occurred on November 6, 2006, are often referred to as "The Pineapple Express."

recording maximum wind, wind radius, precipitation, and so forth, that help climate scientists better understand the formation and evolution of extreme weather events in different model scenarios. The impact of this type of flexible infrastructure has just begun to emerge in the form of many climate science publications in gestation at this time, as well as new project ideas that result from having a new technology capable of processing massive climate datasets quickly on large parallal platforms. For example, Michael Wehner (LBNL) recently received a 2013 INCITE Award for 150 million hours on the ANL BG/Q for a project entitled "Attributing Changes in the Risk of Extreme Weather and Climate," a project that will make extensive use of TECA to study extreme event attribution.

The team developing and applying TECA to climate science problems has received numerous requests from the broader climate science community for this software, which in its initial form was implemented as a standalone capability, but that is now being ported into VisIt to simplify its use and widen its dissemination, including for distribution as part of the UV-CDAT package. The parallel R-based infrastructure is also presently deployed in VisIt, and the combined capability will be included in a future VisIt release. Industry has expressed interest as well, as one Swiss firm has used our team's tropical cyclone analysis of CAM5 data to perform storm damage assessments for North American locations under different future climate scenarios.

CRD Contact

Prabhat, Suren Byna, Michael Wehner, Oliver Rübel, Kesheng Wu, E. Wes Bethel

FRAMEWORKS, ALGORITHMS, AND SCALABLE TECHNOLGIES FOR MATHEMATICS (FASTMATH)

Research Highlights

The Frameworks, Algorithms, and Scalable Technolgies for Mathematics (FASTMath) SciDAC Institute develops and deploys scalable mathematical algorithms and software tools for the reliable simulation of complex physical phenomena on current high performance computing architectures, as well as those that will be deployed in the next few years at DOE high performance computing facilities. Members of FASTMath collaborate with DOE domain scientists to ensure the usefulness and applicability of FASTMath technologies.

The FASTMath Institute is one of four applied mathematics/computer science institutes, funded under the current SciDAC-3 program. It was started in the summer of 2011 and will span five years. LBNL is one of the institutional members of the FASTMath Institute. The FASTMath technical expertise in the CRD at LBNL spans many areas in support of a wide variety of applications, and leverages expertise that have been developed over the years and funded by the Applied Mathematics, SciDAC-1, and SciDAC-2 programs at DOE. Current areas of expertise include finite volume methods and particle/mesh methods in structured grid AMR frameworks, high-order discretization methods, highly optimized linear and nonlinear solvers, and eigensolvers. The following will describe CRD's contributions in these areas.

Region-Based Adaptive Mesh Refinement (RAMR) and Optimal Subcycling

Collaborators in CRD have developed a new paradigm for adaptive mesh refinement (AMR) in BoxLib. In the newly-named region-based AMR (RAMR), different regions at the same level of spatial refinement may have a different temporal refinement. When combined with optimal subcycling, which selects a timestep for each region based on maximizing the efficiency of the overall algorithm, the use of RAMR can result in significant computational savings over traditional AMR approaches. During this past year, CRD's FASTMath team has implemented a branch of the BoxLib source distribution that supports RAMR as opposed to traditional AMR, and they have also developed a prototype RAMR application as a branch of the Nyx cosmology code (based on BoxLib) that is used in SciDAC-3. The initial version of RAMR requires that regions at the same spatial resolution be separated by at least one ghost cell at that level.

High-Order Finite-Volume Methods

CRD's FASTMath team has completed an initial implementation of support for high-order methods on mapped-multiblock grids. Software components include extensions of the fourth-order methods developed for Cartesian grids for interpolation. The software components maintain conservation at refinement boundaries for AMR, including ones that intersect block boundaries. The software also uses specialized versions of interblock interpolation that improve performance in situations in which various anisotropies arise, such as problems for which the coordinate system is a tensor product of a mapped-multiblock grid and a Cartesian grid, or in which AMR is applied in a fixed subset of the coordinate directions. CRD researchers have also extended the AMR grid generation, i.e., the covering of a collection of tagged points by a union of rectangles satisfying suitable proper nesting conditions, to the case of a multiblock grid hierarchy. And finally, they provided generic versions of an explicit and a semi-implicit fourth-order Runge–Kutta method, using template classes, required to obtain temporal accuracy that matches the spatial accuracy of the methods. All of the methods described are implemented in parallel, and represent the components required to implement simulation capabilities for strategic Office of Science applications in fusion (gyrokinetic edge plasma modeling) and climate (AMR for atmospheric modeling).

Embedded Boundary Methods

CRD's work with embedded boundary methods focuses on three topics. The first is grid generation for geometries specified from a variety of sources, including high-resolution image data and stereolithographic (STL) representations of geometries. The former is part of a larger effort to implement a general capability for generating cut-cell geometric representations to any order, and for any number of dimensions using implicit function representations. Such techniques are essential for developing higher order EB (embedded boundary) discretizations; for moving boundaries (which can be viewed as cut cells in 4D space-time), and for free boundaries (stationary or moving). The FASTMath team uses similar techniques to generate cut-cell geometric information from STL, which is the standard engineering representation of surface geometries. The second activity has been the parallel optimization of the EB incompressible flow capability so that it weakly scales to 100,000 processors. This requires the introduction of a variety of low-level optimizations, as well as coupling to the PETSc library. The third activity has been a collaborative effort with the SDAV SciDAC Center on visualization techniques of the results of EB simulations at such large scales. The improved scaling, grid generation from image data, and visualization tools have been used in simulations for an Energy Frontier Research Center, the Center for Nanoscale Control of Geologic CO_2.

Sparse Matrix Solvers

CRD's FASTMath team is developing scalable sparse direct linear solvers and effective preconditioners for the most challenging linear systems, which are often too difficult for iterative methods. Their focal effort is the development of three types of linear solvers. The first, encapsulated in the SuperLU_DIST code base, is a pure direct solver using LU factorization followed by triangular solutions. The second, encapsulated in the software PSDLin, is a Schur complement-based hybrid solver. The team uses a non-overlapping domain decomposition framework, which blends direct solvers for the solution of the sub-domain variables with preconditioned iterative solvers for the solution of the interface variables (i.e., Schur complement system). The third type is the nearly-optimal preconditioners using low-rank approximate factorization of the dense sub-matrices. For many discretized partial differential equation problems, the numerical ranks of the off-diagonal blocks are often very small, thus rendering the truncated single value decomposition (SVD) type of compact representation. They have been using the Hierarchically Semi-separable (HSS) matrices to represent such low-rank structures efficiently, which lead to nearly linear time approximate factorization algorithms. The FASTMath team is developing parallel codes for this type of factorization to be used as direct solvers or as preconditioners.

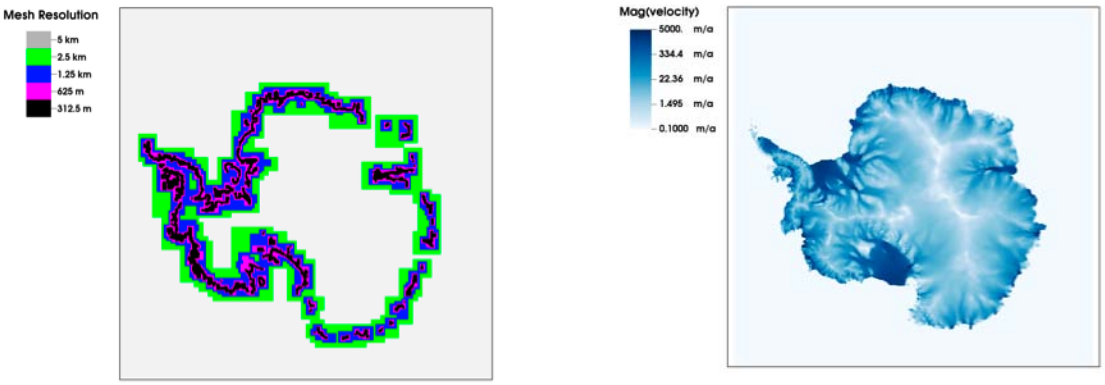


Figure 5: (Left) Vertically integrated Antarctic ice speeds computed by the Chombo-based BISICLES code. (Right) Mesh resolution: 5 km (white); 2.5 km (green); 1.25 km (blue), 625 m (purple); and 312.5 m (black) mesh. Slightly less than 8% of the domain is refined to the finest level.

Eigenvalue Calculations

The goal of the eigensolvers work in FASTMath is to develop scalable and robust algorithms and solvers for tackling large-scale linear and nonlinear eigenvalue problems. Techniques to be considered include, implicitly restarted Lanczos and Arnoldi iterations for linear eigenvalue problems; nonlinear Arnoldi, Jacobi–Davidson, rational Krylov, second-order Arnoldi methods for one-parameter nonlinear eigenvalue problems; and optimization approaches for solving eigenvalue problems with nonlinearity in the eigenvectors. In the first year of FASTMath, CRD collaborators developed effective preconditioning techniques for solving nonlinear eigenvalue problems in which the matrix depends nonlinearly on the eigenvectors. They also examined ways to improve the performance of Lanczos-based eigensolvers for computing a few eigenpairs of symmetric sparse matrices on a many/multi-core system. CRD researchers developed techniques to reduce the communication overhead as much as possible. One of the techniques was designed to exploit the network topology. Another technique was to employ hybrid and overlapped OpenMP/MPI implementation to hide communication overhead. And finally, they designed algorithms to compute a relatively large number of eigenvalues for a million-dimensional matrix.

Software

Software developed by FASTMath team members at LBNL comprises the framework and/or solver technology for application codes in accelerator modeling, astrophysics, combustion, cosmology, materials science and porous media. This is evident by the fact the FASTMath team members at LBNL are involved in 10 out of the 18 SciDAC-3 Scientific Computation Application Partnerships: 5 in Basic Energy Sciences, 1 in Biological and Environmental Research, 2 in High Energy Physics, and 2 in Nuclear Physics.

Software developed by LBNL FASTMath team members is publicly available by download through the individual web sites. Their collection includes

- Block-structured adaptive mesh refinement: BoxLib, Chombo,
- Particle and Particle/Mesh Methods: BoxLib, Chombo,
- Embedded boundary methods: Chombo,
- Direct solution of linear systems: SuperLU,
- Sparse eigensolvers: PARPACK.

These software packages are widely used in scientific applications at DOE and elsewhere.

Scientific Impact

The RAMR paradigm is being developed in the context of the Nyx cosmology code, a SciDAC-3 applications code based on the FASTMath-supported BoxLib framework. Nyx contains block-structured AMR hyperbolic and elliptic solvers for the gas evolution and self-gravity, respectively, as well as particles representing dark matter. As such, Nyx serves as an ideal computational test bed for the new ideas, and in return, Nyx will become the first large-scale massively parallel cosmology code to reap the benefits of RAMR and optimal subcycling, both of which potentially result in substantial computational savings. Increased efficiency in cosmological simulations results in larger domains and higher resolutions, and therefore, a higher fidelity of the physical processes, resulting in more reliable simulations for comparison with observation. Ultimately these simulations are used to guide our understanding of both the past and present history of the Universe.

Chombo is currently used in a variety of applications within DOE. CRD is using Chombo in many of their projects, which include applications to AMR simulations of land ice dynamics and to pore-scale modeling of carbon sequestration, as well as a new effort to develop an AMR capability for atmospheric modeling. Other DOE applications include the use of the elliptic solver and particle capabilities for beam dynamics in accelerators and gyrokinetic models of edge plasmas in tokamaks. Chombo is also being used to replace the current AMR infrastructure in the FLASH code, a simulation capability being funded by NNSA for high energy

density physics (HEDP) applications. In all of these cases, Chombo is viewed by the applications scientists as a stable, highly capable platform to carry out high performance simulations that would otherwise require a much larger investment to develop stand-alone capabilities from scratch and to support those capabilities.

In the area of eigenvalue calculations, CRD researchers have been working closely with computational physicists to improve a symmetric eigensolver in many fermion dynamics (MFDn), which is a nuclear structure calculation code. For over a five-year period under SciDAC-2, the collaboration has resulted in an improvement of the performance of MFDn by a factor of almost four on a standard test case. The improvements over time have enabled MFDn to make scientific discoveries. One version of the MFDn code has been used in predictive discoveries of the properties of fluorine-14, which is an exotic nucleus with extreme proton-to-neutron ratio. In this particular case, the calculations were completed before the isotope was observed in experiments. The order of the matrix was close to 2 billion. Another version of MFDn has been used successfully to understand why the carbon-14 isotope has such a long half-life, which is over 5,700 years. (The half-lives of other unstable carbon isotopes are extremely small—less than minutes.) Without the collaboration with computational physicists, these calculations would not have been possible, as the calculations would just take too long to run.

CRD Contacts

Ann Almgren, Phil Colella, Esmond Ng

CRD COLLABORATORS

John Bell, Daniel Graves, Xiaoye Sherry Li, Terry Ligocki, Michael Lijewski, Peter McCorquodale, Daniel Martin, Brian Van Straalen, Chao Yang

OTHER LBNL COLLABORATORS

Wick Haxton (Lead PI of an NP SciDAC SAP), Martin Head-Gordon (PI of a BES SciDAC SAP)

COLLABORATORS AT OTHER INSTITUTIONS

Roberto Car (Princeton University, Lead PI of a BES SciDAC SAP), Joseph Carlson (LANL, Lead PI of an NP SciDAC SAP), James Chelikowsky (University of Texas, Austin, Lead PI of a BES SciDAC SAP), Christopher Cramer (University of Minnesota, Lead PI of a BES SciDAC SAP), Salman Habib (ANL, Lead PI of a HEP SciDAC SAP), William Lipscomb (LANL, Lead PI of a BER SciDAC SAP), Panagiotis Spentzouris (FNAL, Lead PI of a HEP SciDAC SAP)

MATERIALS DISCOVERY

Research Highlights

In order to accelerate the discovery of new materials for important energy applications, CRD researchers are developing and applying novel algorithms and cutting-edge computer science solutions to energy applications like gas separations, energy storage and catalysis. CRD's resulting tools allow scientists to handle the complexity of materials search space represented by large databases. The tools let users collect, characterize and screen large molecular and material databases, discover materials with outstanding properties, perform data-mining for new materials knowledge, as well as disseminate results via interactive visualization and web solutions. The developed tools and databases are community accessible and developed in collaboration with domain scientists of the Berkeley-based Energy Frontier Research Centers for Gas Separations and the recently funded DOE Basic Energy Sciences Center for Functional Electronics.

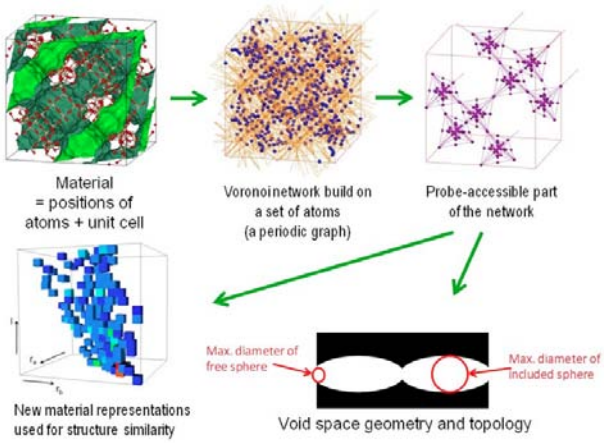


Figure 6: Zeo++: a tool for high-throughput analysis of voids in crystal structures based on Voronoi decomposition, which for a set of atoms in a periodic material's structure, provides the Voronoi network map of the voids therein. The resulting Voronoi network is analyzed to provide information on the geometry and topology of the void space, as well as used to derive materials' similarity.

The research in materials discovery can be grouped into four main themes. The first theme is the development of high-throughput analysis and data-mining tools. While very large databases of material structures are being developed, these tools are needed to analyze and screen structure libraries, which can contain millions of entries. These tools and approaches are critical to the success of this project's endeavors, which enable automatic, unsupervised analysis of these materials, as well as development of suitable descriptors, which encode chemical information, and which can be used for screening and data-mining. CRD researchers have developed a suite of algorithms that involve various applied mathematics methodologies, i.e., Voronoi decomposition, partial differential equations, the Fast Marching Method, or graph theory. One example of such a methodology is Zeo++, a Voronoi decomposition-based package initially developed to characterize the void space of porous materials in a high-throughput manner, which is now being adopted to investigate ion diffusion paths in battery materials (see Figures 6 and 7).

Efficient materials discovery approaches are the second main theme of this project. The search space of possible materials is often so large that it cannot be efficiently enumerated and screened by a brute-force approach, in which each structure is characterized using an expensive molecular simulation methodology. CRD researchers have developed efficient discovery approaches, where only important and statistically relevant materials are characterized, without using any computationally expensive methodologies. These approaches rely on either the identification of structural similarities that lead to the clustering of materials exhibiting similar properties, or an optimization procedure of abstracted "alchemical" material structures. The first approach has been recently demonstrated in the screening of zeolite materials for the CO_2 capture and sequestration application, where the majority of the best performing materials was discovered at a fraction of the cost of brute-force screening. The latter approach was in turn used to discover new high-surface area materials, outperforming the current record-holding materials by about 80%.

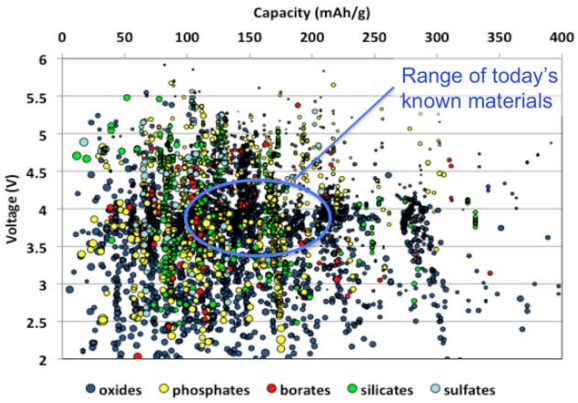


Figure 7: Potential battery materials screened by the Materials Project as a function of predicted voltage and capacity. Note the comparatively narrow range of properties exhibited by known materials.

A third theme is the development of high-throughput workflows. Dramatic increases in parallel computing power have made the computational determination of many fundamental properties of matter possible on an unprecedented scale. The need to run and collect data from complex calculations, such as Vienna the *Ab initio* Simulation Package (VASP), at the scale of thousands materials per month, on systems such as NERSC's Hopper, led CRD researchers to develop a novel intertwined workflow and database infrastructure, which they have used to create an unprecedented database of materials properties.

And finally, a community-accessible datastore and analysis tools are the fourth theme in the materials discovery project. Given the intellectual effort required to compute materials properties at a large scale, the resulting datastore is a precious resource. To open this datastore to the broadest set of experimentalists and theorists, both power (search and analytics) and usability are important. CRD researchers, in collaboration with MIT, have developed a flexible and scalable system for storing the data and disseminating it from a web-based gateway, called the Materials Project.

The Materials Project [10] is pioneering the "fourth paradigm" of data-intensive science in the materials science community. This data-centered framework allows scientists to collaborate and derive scientific insight in materials science in a way that is not possible with current scientific tools. For example, users can store material searches as URL permalinks that make the sharing of results natural and easy. In addition, LBNL researchers and their collaborators' Python Materials Genomics (*pymatgen*) library [13], an open-source Python library, provides a collaborative platform for sophisticated analyses of materials that leverages the Materials Project database via RESTful interfaces. Programmatic access (a representational state transfer API, or REST API) and well-defined data formats allow external researchers to analyze and use the data locally.

The Materials Project infrastructure leadership and development involves personnel from Data Intensive Systems Group (DIS) and Complex Systems Group (CXG) in the Center for Computational Sciences and Engineering. DIS personnel have partnered in the development of a database and data integration infrastructure to enable new calculations and integrate those calculations with an envisioned online "materials design environment." CXG personnel have led the work on a high-throughput workflow engine that meets the needs of materials computations on HPC cluster resources. These efforts are integrated through a common data infrastructure, allowing scientific reproducibility through integrated provenance. Staff at NERSC have been essential partners in the design, development, and deployment the Materials Project website on NERSC resources.

The Center for Functional Electronic Materials, funded by the DOE Basic Energy Sciences's Predictive Theory and Modeling for Materials

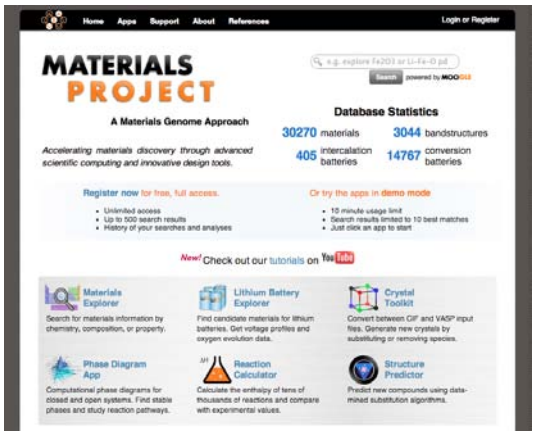


Figure 8: Screenshot of the Materials Project homepage showing statistics on compounds in the database, as well as the menu of analysis applications for searching materials, drawing phase diagrams, etc.

and Chemical Sciences program, will extend the Materials Project with high-throughput calculations, and electronic structure methods, as well as novel data-mining algorithms for surface, defect, electronic and finite temperature property predictions.

The Materials Project is one of the leading projects in a new era of computer-aided materials development, the keystone of the Materials Genome Initiative, that has been recognized by the US Government as "crucial to achieving global competitiveness in the 21st century" [9]. As of the beginning of 2013, the Materials Project website allows anyone in the world the ability to explore the results of detailed and difficult electronic structure calculations on 30,498 materials, which includes 3,044 band structures, 400 intercalation batteries, and 14,721 conversion batteries. Both the quantity and quality of data being made available to the scientific community are virtually unmatched.

The current Materials Project implementation runs as a 24/7 production system with, as of January 2013, over 2,500 registered users. The framework has performed computations to screen over 80,000 inorganic compounds for a variety of applications, including lithium-ion and sodium-ion batteries [2, 6, 7, 11, 12]. The Materials Project infrastructure is being leveraged by newly-funded Materials Science projects at DOE and NSF. As new collaborators and capabilities come online, it is not unreasonable to expect the popularity of the site to continue to grow, along with the corresponding impact on the community both in terms of best-practiced informatics and, primarily, as a data and analysis resource.

Scientific Impact

Scientists, research groups and industry benefit from the work done by CRD researchers by being able to access these databases and use the many developed tools. The flexible workflow and database infrastructure that can be configured to add new and to modify existing systems to generate new systems (such as defects, surface calculations, etc.), thus, accelerating the rate of generating new databases for screening. The Materials Project website, shown in Figure 8, is a gateway to the materials database and has over 3,000 registered users. Several publications have already been produced by users who have mined the database, in addition to a number of papers produced by the collaboration. Similarly, the Carbon Capture Materials Database serves as a resource of physicochemical basic data related to CO_2 capture materials. The database has over 100 users from academia and energy industry. It is also connected to the tools developed within the Carbon Capture Simulation Initiative (CCSI) community.

The CRD-developed materials discovery software tools, such as Zeo++ or (with MIT) pymatgen, are provided to the community on an open source basis. The application of these tools, both internally and by other researchers around the globe, has led to the discovery of many materials with outstanding properties.

CRD Contacts

Maciej Haranczyk (SCG), Dan Gunter (DIS), Chris Rycroft (Math), Kristin Persson (EETD, project co-lead), Wei Chen (EETD), Michael Kocher (EETD)

LBNL COLLABORATORS

Richard Martin (SCG), Anubhav Jain (CXG), Lavanya Ramakrishnan (DIS), David Skinner (NERSC), Shreyas Cholia (NERSC), Annette Greiner (NERSC)

OTHER LBNL COLLABORATORS

Berend Smit (UCB), Mark Asta (UCB), Gerbrand Ceder (MIT), Shyue Ping Ong (MIT), William Davidson Richards (MIT), Evgueni Chtykov (MIT), Geoffroy Hautier (Université Catholique de Louvain), Alan Dozier (University of Kentucky), Matthew J. Beck (University of Kentucky), Stefan Adams (National University of Singapore)

ADVANCED SIMULATION CAPABILITY FOR ENVIRONMENTAL MANAGEMENT (ASCEM)

Research Highlights

In an effort to provide computational infrastructure for environmental management solutions, DOE created the Advanced Simulation Capability for Environmental Management (ASCEM) initiative—a multi-year program that involves geologists, hydrologists, physicists and computer scientists from various national laboratories, including Pacific Northwest National Laboratory (PNNL), Savannah River National Laboratory (SRNL), Los Alamos National Laboratory (LANL) and Lawrence Berkeley National Laboratory (LBNL). ASCEM is envisioned to be a state-of-the-art approach for integrating data, software and scientific understanding to improve subsurface contaminant fate and transport simulations used to support environmental management decisions.

The goal of this program is to develop models, simulation software and visual analysis tools for the evaluation of measured and simulated data with respect to nuclear waste, waste products, and other contaminants in soil and ground water. Historical data from existing contamination sites is taken into account to calibrate soil and water flow models and to verify simulation results. The simulation code is then used to extrapolate this data into the future, incorporating different treatment options. The ultimate goal of the simulation and analysis of the resulting data is to arrive at a decision for a potential future closure of the site. This is usually equivalent to ending environmental treatment measures.

Due to the complexity of the underlying data models developed by teams of geologists and physicists, it was necessary to employ HPC methods both for the simulation and for the subsequent visual analysis of the data. This work is a collaborative effort among a team of CRD researchers in the Scientific Data Management (SDM) Group, the Advanced Computing Sciences (ACS) Department, the Visualization Group, and the Center for Computing Sciences and Engineering.

Management and Visual Analysis of Heterogeneous Environmental Site Characterization Data

Scientists, site managers and regulators usually need to access, create, and manage large amounts of heterogeneous data, including disparate site characterization data. The data are necessary to enable the development of conceptual and numerical models, parameter estimation, and multiple simulations. Using the DOE Savannah River Site F-Area and Hanford specific databases (well layout, concentration, groundwater level, lithology, meteorological records, etc.) as case studies, the CRD team designed a methodology to effectively manage the staging and views of environmental data to ensure that site characterization and monitoring data are readily available as inputs for parameter estimation, numerical predictions, uncertainty quantification, and risk analysis. The CRD team developed methods for organizing different databases and data types into a common framework that allows the user to browse the data as a single coherent dataset, and then display the data in its geospatial context.

The Observational Data Management System (ODMS) was designed with two goals: (1) provide a way to ingest data from heterogeneous data sources into a single database schema, and (2) generate common data views for driving a generic interface for data browsing, searching, display, and output. These goals avoid the need for special purpose database schemas and user interfaces for

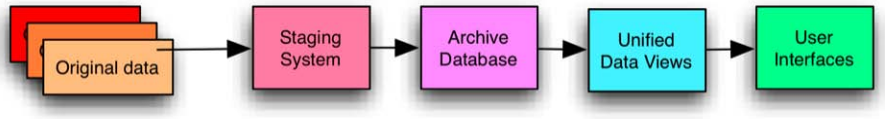


Figure 1: ASCEM data management system architecture view.

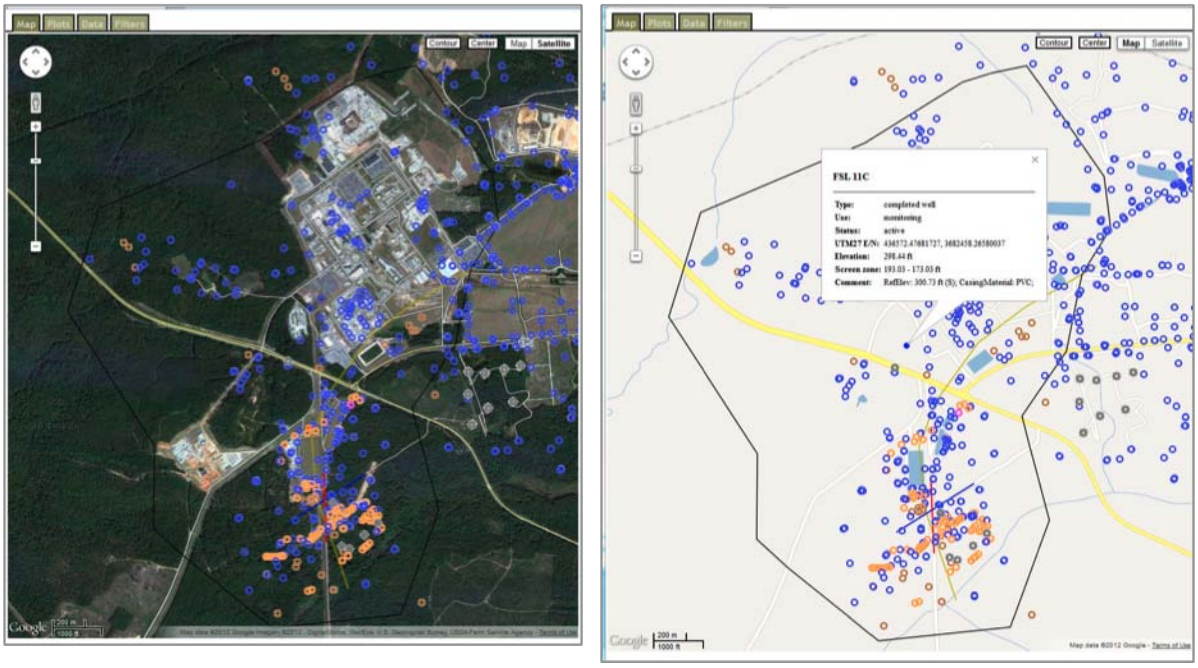
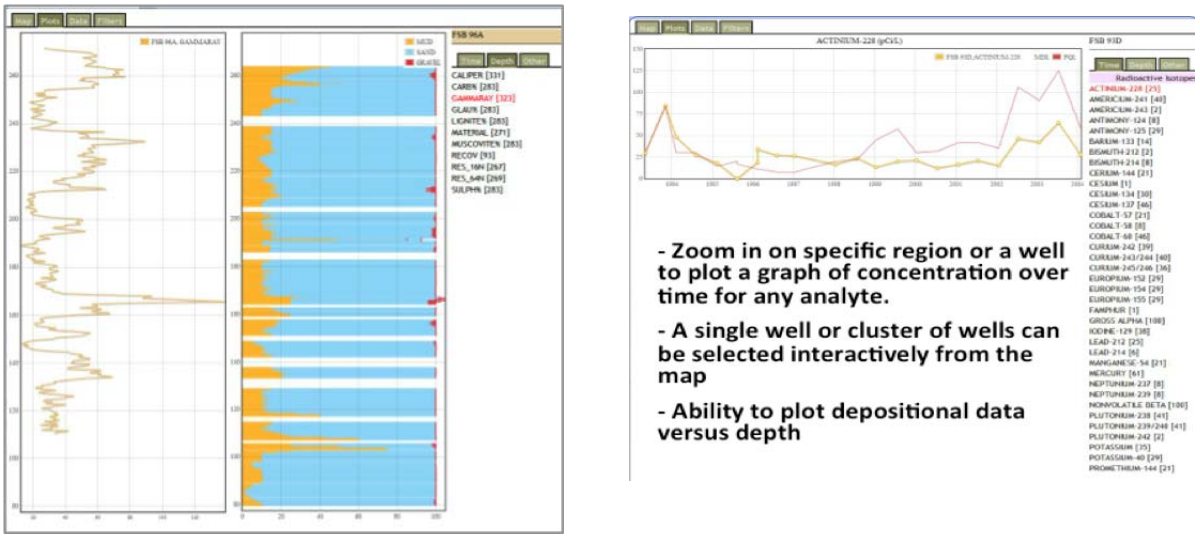


Figure 2: Displaying wells positions on real map.

each new database, and the goals ensure a unified management and view of heterogeneous datasets. The data management modular architecture components are shown in Figure 1. The system design begins with the development of the staging system. Staging includes ingesting the data into the database, cleaning the data, standardization of variable names, and any unit conversions required. Once the data have been staged, they are moved to the archive database. The archive database stores the data in a uniform format, allowing new data types to be incrementally added to the database without requiring a major redesign of the database schema. CRD's team used a canonical schema, where common concepts, such as station/well and types of data (e.g., time series curves for displaying analyte concentration over time, and depth-based curves for depth data, such as provided by cone penetrometer), are normalized. The unified data views are instances of the data organized in ways

Figure 3: Browsing of filtered data by time and depth.



that the user would like to access the data. For example, a data view may reconstruct the time history of contaminants in the wells. The user interfaces utilize the unified data views to present data processing, access, and browsing interfaces.

The main accomplishments of this project are the development of a unified database system, and an interface that is intuitive for the user to use. The team used PostgreSQL to manage and query common environmental site characterization, monitoring, and modeling data. They used Google Maps to display the wells in the site (Figure 2). Also, the CRD team developed browsing and search interfaces and added tools (MAP, PLOT, FILTER) to display interactive maps and plots (Figure 3).

Users can plot a graph of concentration over time for any listed analyte for the selected wells. Additionally, for a given aquifer and analyte users, can overlay a contour plot of the analyte concentration as measured by all the wells screened in that particular aquifer for any calendar year in the measured range.

To generate the contour plots, CRD's team used the graphical capabilities of the VisIt server. Upon selection of the aquifer, analyte and year, all concentration data for every well satisfying the selection criteria are retrieved from the database and passed via an AJAX call to the VisIt server along with the geographic coordinates of the wells (Figure 4).

In addition to 2D Google Maps plot overlays, the team also generated 3D plots of the terrain, the wells and selected stratigraphic layers. As this work focuses on the visual display of historical data, the starting point centers around loading such data into VisIt. First, raw field data was obtained from scientists working at the site. The data formats included plain text (comma-delimited), GDAL GeoTIFF images, and ESRI Shapefiles. Most data types were loaded into VisIt with the included readers. However, the team had to construct a custom plug-in to read in comma-delimited plain text files with a specific layout describing the depositional environment, as recorded in wellbores at the site. Other plain text files record concentration values for contaminants measured at specific depths within each monitoring well or describe the subsurface coordinates of impermeable lithologic surfaces that form aquifer boundaries. Geo-referenced true-color aerial photographs and Digital Elevation Model (DEM) terrain data were loaded as GDAL GeoTIFF images. ESRI Shapefiles loaded into VisIt contain vector data defining roads and structures. Also, in order to apply elevations from the DEM topography to the aerial photographs, the DEM had to be resampled to match the pixel resolution of the photographs. The CRD team managed to subset all datasets to be within the same spatial extent.

The Visualization team used a Pseudocolor plot to visualize the concentration values, varying the color from white to red by using a range of concentration values from 0 to 400 pCi/L. Prior to drawing the plot, the team subset the concentration values using the threshold operator to confine the measurements within a particular aquifer defined by the depth at which the measurements were recorded. Subsequently, they

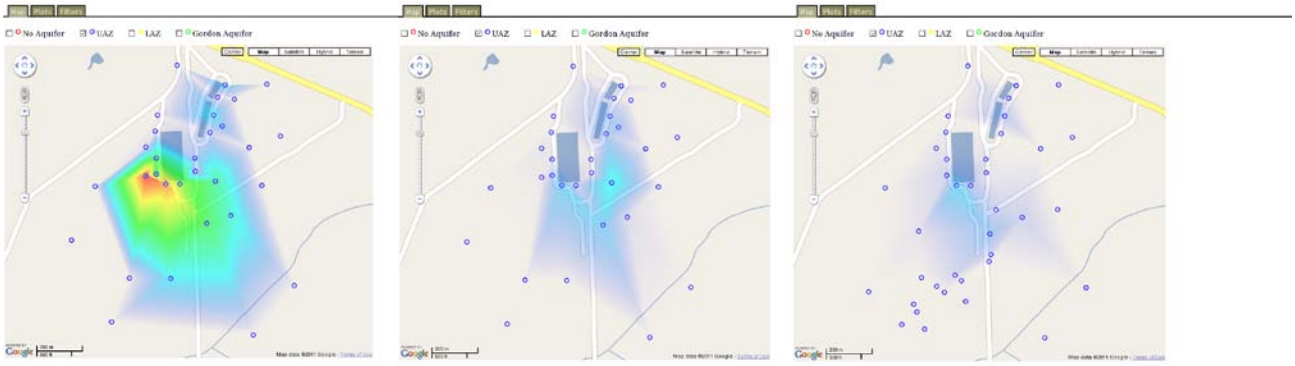


Figure 4: Plan view map of the oil and groundwater U-238 contamination at the Savannah River Site (SRS F-area, upper aquifer zone) in (a) 1994, (b) 2001 and (c) 2008, displayed in a web browser as a Google Maps bitmap overlay created in VisIt.

created a surface from those values by applying a Delaunay triangulation. A Delaunay triangulation was used, because it avoids errors typical for other types of interpolation, such as Shepard's, Sibson's or Hardy's method. A Delaunay triangulation was also applied to the lithologic information that defines a particular aquifer boundary and gave the resulting surface the color green.

VisIt has a feature that allows the recording of all the commands used to open databases, create and modify plots, modify legend attributes, and add text annotation (Figure 5).

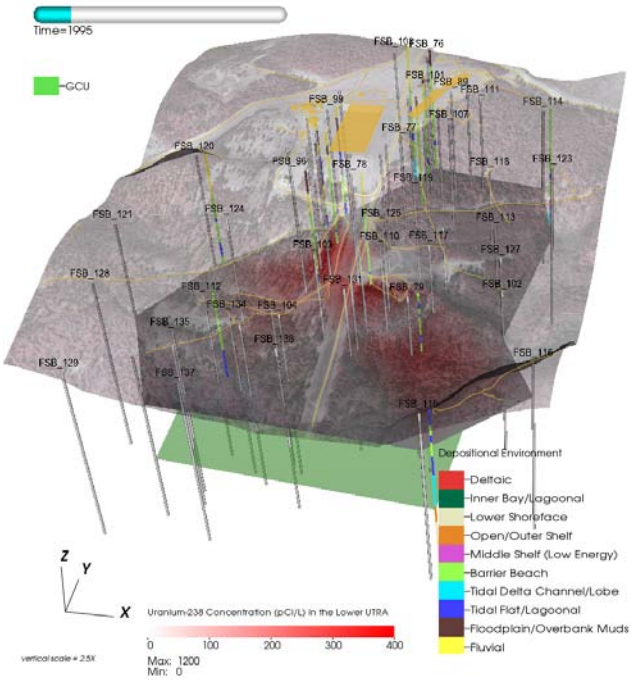


Figure 5: 3D visualization (computed in VisIt) of Uranium-238 plume migrating through the subsurface of the F-area. This still image was taken from an animation, which includes an aerial image with elevations applied from a digital elevation map, wells that have information about the depositional environment (lithology with respect to depth), buildings, structures, and roads (in yellow/orange color), and the lower boundary for the GCU aquifer (green). It also shows Uranium-238 concentration as a pseudocolor plot (white to red).

Scientific Impact

The team's implementation of the database management system provided the scientists with an easy access to the diverse experimental data collected at remediation sites. This allowed the data to be directly used from the database in developing parameter estimations, site characterization, calibration, and model validation activities. This ensures that the latest and best data is used. Having ready access to data search and visualization capabilities also allows researcher to explore the contamination patterns over time, as well as the effects of remediation activities over time, providing a better understanding of site dynamics.

CRD Contacts

Deb Agarwal, Arthur Wiedmer, Gary Kushner, Alex Romosan, Arie Shoshani (Data Management) Joerg Meyer (Visualization), Harinarayan Krishnan (Visualization), Jennifer Horsman (Visualization)

Collaborators

Boris Faybishenko (Earth Science Division), Tad Whiteside (Savannah River National Laboratory), James Hunt (Department of Civil and Environmental Engineering, UC Berkeley)

REFERENCES

[1] Surendra Byna, Prabhat, Michael F. Wehner, and Kesheng Wu. "Detecting Atmospheric Rivers in Large Climate Datasets." In Proceedings of the 2nd International Workshop on Petascale Data Analytics: Challenges, and Opportunities (PDAC-11/Supercomputing 2011), Seattle,WA, USA, November 2011. ACM/IEEE.

[2] G. Ceder, G. Hautier, A. Jain, and S.P. Ong. "Recharging lithium battery research with rstprinciples methods." *MRS Bulletin*, 36(03):185-191, 2011.

[3] M. Christen, N. Keen, T. Ligocki, L. Oliker, J. Shalf, B. van Straalen, and S. Williams. "Automatic thread-level parallelization in the Chombo AMR library." Technical report, Lawrence Berkeley National Laboratory, 2011.

[4] S. L. Cornford, D. F. Martin, D. T. Graves, D. F. Ranken, A. M. Le Brocq, R. M. Gladstone, A. J. Payne, E. G. Ng, and W. H. Lipscomb. "Adaptive mesh, nite volume modeling of marine ice sheets." *Journal of Computational Physics*, 232(1):529-549, 2013.

[5] K. Datta, S. Williams, V. Volkov, J. Carter, L. Oliker, J. Shalf, and K. Yelick. "Auto-tuning the 27-point stencil for multicore." In *International Workshop on Automatic Performance Tuning (iWAPT)*, 2009.

[6] Georoy Hautier, Anubhav Jain, Hailong Chen, Charles Moore, Shyue Ping Ong, and Gerbrand Ceder. "Novel mixed polyanions lithium-ion battery cathode materials predicted by highthroughput ab initio computations." *J. Mater. Chem.*, 21:17147-17153, 2011.

[7] Georoy Hautier, Anubhav Jain, Shyue Ping Ong, Byoungwoo Kang, Charles Moore, Robert Doe, and Gerbrand Ceder. "Phosphates as lithium-ion battery cathodes: An evaluation based on high-throughput ab initio calculations." *Chemistry of Materials*, 23(15):3495-3508, 2011.

[8] K. E. Kunkel, T. R. Karl, H. Brooks, J. Kossin, J. H. Lawrimore, D. Arndt, L. Bosart, D. Changnon, S. L. Cutter, N. Doesken, K. Emanuel, P. Y. Groisman, R. W. Katz, T. Knutson, J. O'Brien, C. J. Paciorek, T. C. Peterson, K. Redmon, D. Robinson J. Trapp, R. Vose, S. Weaver, M. Wehner, K. Wolter, and D. Wuebbles. "Monitoring and Understanding Trends in Extreme Storms: State of Knowledge." *Bulletin of the American Meteorological Society*, 2012.

[9] Materials Genome Initiative for Global Competitiveness. http://www.whitehouse.gov/mgi.

[10] Materials Project. http://www.materialsgenome.org/.

[11] Yifei Mo, Shyue Ping Ong, and Gerbrand Ceder. "First principles study of the li10gep2s12 lithium super ionic conductor material." *Chemistry of Materials*, 24(1):15-17, 2012.

[12] Shyue Ping Ong, Vincent L. Chevrier, Georoy Hautier, Anubhav Jain, Charles Moore, Sangtae Kim, Xiaohua Ma, and Gerbrand Ceder. "Voltage, stability and diusion barrier dierences between sodium-ion and lithium-ion intercalation materials." *Energy Environ. Sci.*, 4:3680-3688, 2011.

[13] Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Georoy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L. Chevrier, Kristin A. Persson, and Gerbrand Ceder. "Python materials genomics (pymatgen): A robust, open-source python library for materials analysis." *Computational Materials Science*, 68(0):314-319, 2013.

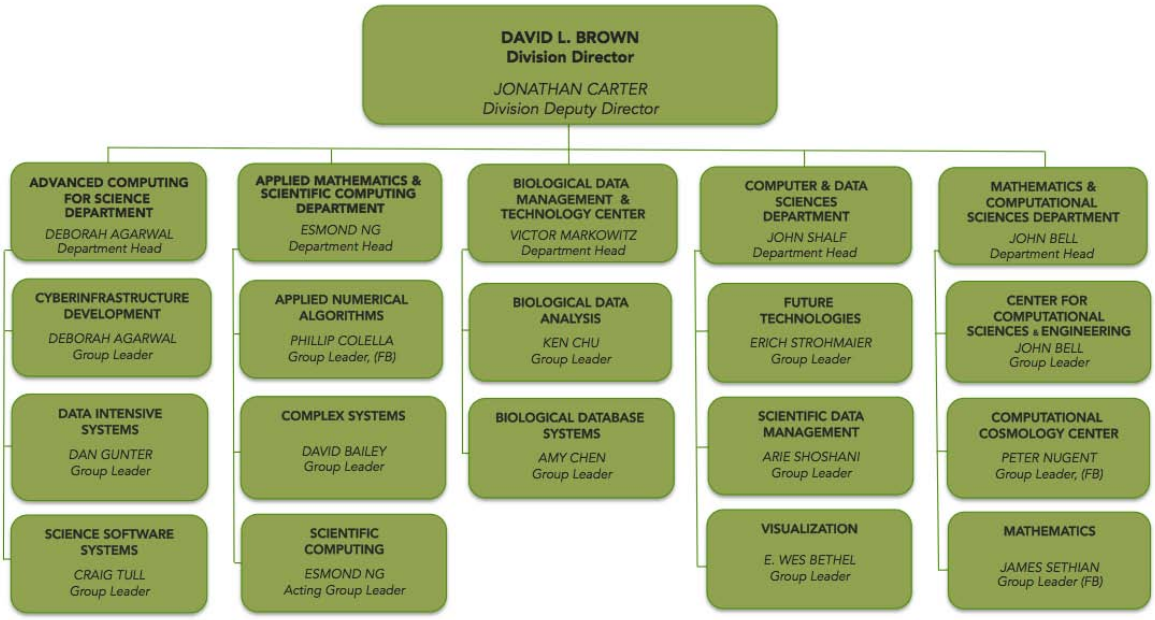
[14] Frank Pattyn, Laura Perichon, Gael Durand, Lionel Favier, Olivier Gagliardini, Richard C. A. Hindmarsh, Thomas Zwinger, Torsten Albrecht, Stephen Cornford, David Docquier, Johannes J. Furst, Daniel Goldberg, G. Hilmar Gudmundsson, Angelika Humbert, Moritz Hutten, Philippe Huybrechts, Guillaume Jouvet, Thomas Kleiner, Eric Larour, Daniel Martin, Mathieu Morlighem, Anthony J. Payne, David Pollard, Martin Ruckamp, Oleg Rybak, Helene Seroussi, Malte Thoma, and Nina Wilkens. "Grounding-line migration in plan-view marine icesheet models: results of the Ice2Sea MISMIP3d intercomparison." *Journal of Glaciology,* 2012. Submitted.

[15] A. J. Payne, S. L. Cornford, D. F. Martin, C. Agosta, M. R. van den Broeke, T. L. Edwards, R. M. Gladstone, H. H. Hellmer, G. Krinner, A. M. Le Brocq, S. R. M. Ligtenberg, W. H. Lipscomb, S. R. Shannon E. G. Ng, R. Timmerman, and D. G. Vaughan. "Impact of uncertain climate forcing on projections of the West Antarctic ice sheet over the 21st and 22nd centuries." *Earth and Planetary Science Letters*, 2012. Submitted.

[16] Prabhat, Oliver Rübel, Surendra Byna, Kesheng Wu, Fuyu Li, Michael Wehner, and E. Wes Bethel. "TECA: A Parallel Toolkit for Extreme Climate Analysis." In *Third Worskhop on Data Mining in Earth System Science (DMESS 2012) at the International Conference on Computational Science (ICCS 2012)*, Omaha, Nebraska, June 2012. LBNL-5352E.

[17] S. Williams, D. Kalamkar, A. Singh, A. Deshpande, B. Van Straalen, M. Smelyanskiy, A. Almgren, P. Dubey, J. Shalf, and L. Oliker. "Optimization of geometric multigrid for emerging multiand manycore processors." In *Supercomputing (SC)*, November 2012.

COMPUTATIONAL RESEARCH DIVISION ORGANIZATION CHART



PUBLICATIONS 2009–2012

ADVANCED COMPUTING FOR SCIENCE DEPARTMENT

2012

Zacharia Fadika, Madhusudhan Govindaraju, Shane Canon, and Lavanya Ramakrishnan, "Evaluting hadoop for data-intensive scientific operations," *IEEE Cloud Computing*, 2012.

Deb Agarwal, Arthur Wiedmer, Boris Faybishenko, Tad Whiteside, James Hunt, Gary Kushner, Alex Romosan, and Arie Shoshani Arie, "A methodology for management of heterogeneous site characterization and modeling data," to appear in *XIX International Conference on Water Resources (CMWR)*, Urbana-Champaign, IL, June 2012.

Elif Dede, Zacharia Fadika, Jessica Hartog, Modhusudhan Govindaraju, Lavanya Ramakrishnan, Daniel Gunter, and Richard Shane Canon, "MARISSA: MApReduce Implementation for Streaming Science Applications," In *eScience*, 2012.

Adel El-Atawy and Taghrid Samak, "End-to-end verification of qos policies," In 2012 IEEE Network Operations and Management Symposium, Maui, HI, USA, April 16-20, 2012, pp. 426–434, 2012.

Ezra Kissel, Ahemd El-Hassany, Guilherme Fernandes, Martin Swany, Dan Gunter, Taghrid Samak, and Jennifer Schopf, "Scalable integrated performance analysis of multi-gigabit networks," In *5th Workshop on Distributed Autonomous Network Management System (DANMS)*, co-located with NOMS 2012, 2012.

Dario Papale, Deborah Agarwal, Dennis Baldocchi, Robert Cook, Joshua B, Fisher, and Catharine van_Ingen, "Database maintenance, data sharing policy, collaboration," Marc Aubinet, Timo Vesala, and Dario Papale, editors, In *Eddy Covariance*, pp. 399–424, Springer, 2012.

Lavanya Ramakrishnan, Richard Shane Canon, Krishna Muriki, Iwona Sakrejda, and Nicholas J. Wright, "Evaluating interconnect and virtualization performance for high performance computing," In *Special Issue of ACM Performance Evaluation Review*, 40(2), 2012.

Taghrid Samak, Dan Gunter, Monte Goode, Ewa Deelman, Gideon Juve, Fabio Silva, and Karan Vahi, "Failure analysis of distributed scientific workflows executing in the cloud," In *8th International Conference on Network and Service Management, CNSM 2012*, Las Vegas, USA, October 22-26, 2012.

Taghrid Samak, Dan Gunter, and Valerie Hendrix, "Scalable analysis of network measurements with hadoop and pig," In *5th Workshop on Distributed Autonomous Network Management System (DANMS),* co-located with NOMS 2012, 2012.

Taghrid Samak, Dan Gunter, and Zhong Wang, "Prediction of Protein Solubility in E, Coli," In *eScience*, 2012.

Karen Schuchardt, Deb Agarwal, Stefan Finsterle, Carl Gable, Ian Gorto, Luke Gosink, et al., "Akuna - integrated toolsets supporting advanced subsurface flow and transport simulations for environmental management," to appear in *XIX International Conference on Water Resources (CMWR)*, Urbana-Champaign, IL, June 2012.

Karan Vahi, Ian Harvey, Taghrid Samak, Dan Gunter, Kieran Evans, David Rogers, Ian Taylor, Monte Goode, Fabio Silva, Eddie Al-Shakarchi, Gaurang Mehta, Andrew Jones, and Ewa Deelman, "A General Approach to Real-time Workflow Monitoring," In *The Seventh Workshop on Workflows in Support of Large-Scale Science (WORKS12)*, in conjunction with SC 2012, Salt Lake City, November 10-16 2012, Salt Lake City, USA, November 2012.

2011

Deb Agarwal, You-Wei Cheah, Dan Fay, Jonathan Fay, Dean Guo, Tony Hey, Marty Humphrey, Keith Jackson, Jie Li, Christophe Poulain, Youngryel Ryu, and Catharine van Ingen, "Data-intensive science: The terapixel and modisazure projects," *International Journal of High Performance Computing Applications*, 25(3):304–316, 2011.

Elif Dede, Lavanya Ramakrishnan, Dan Gunter, and Madhusudhan Govindaraju, "Riding the Elephant: Managing Ensembles with Hadoop," In *The 4th Workshop on Many-Task Computing on Grids and Supercomputers (MTAGS)*, IEEE, November 2011.

Zacharia Fadika, Elif Dede, Madhusudhan Govindaraju, and Lavanya Ramakrishnan, "Benchmarking mapreduce implementations for application usage scenarios," *Grid 2011: 12th IEEE/ACM International Conference on Grid Computing*, 0:1–8, 2011.

Zacharia Fadika, Elif Dede, Madhusudhan Govindaraju, and Lavanya Ramakrishnan, Mariane: "Mapreduce implementation adapted for HPC environments," *Grid 2011: 12th IEEE/ACM International Conference on Grid Computing,* 0:1–8, 2011.

Devarshi Ghoshal, Richard Shane Canon, and Lavanya Ramakrishnan, "I/O performance of virtualized cloud environments," In *The Second International Workshop on Data Intensive Computing in the Clouds (DataCloud-SC11)*, 2011.

Dan Gunter, Taghrid Samak, Ewa Deelman, Christopher H. Brooks, Monte Goode, Gideon Juve, Gaurang Mehta, Priscilla Moraes, Fabio Silva, Martin Swany, and Karan Vahi, "Online Workflow Management and Performance Analysis with STAMPEDE," *7th International Conference on Network and Service Management (CNSM 2011)*, 2011.

Keith R, Jackson, Krishna Muriki, Lavanya Ramakrishnan, Karl J, Runge, and Rollin C, Thomas, "Performance and Cost Analysis of the Supernova Factory on the Amazon AWS Cloud," *Sci, Program*, 19:107–119, April 2011.

Ezra Kissel, Dan Gunter, Taghrid Samak, Ahmed El-Hassany, Guilherme Fernandes, and Martin Swany, "An Instrumentation and Measurement Framework for End-to-End Performance Analysis," Technical Report 2011/04, University of Delaware, 2011.

David C. Miller, Deborah A. Agarwal, Xin Sun, and Charles Tong, "iCCSI and the role of advanced computing in accelerating the commercial deployment of carbon capture systems," In *Proceedings of the SciDAC 2011 Conference*, Denver, CO, July 2011.

L. Ramakrishnan, P. T. Zbiegel, S. Campbell, R. Bradshaw, R.S. Canon, S. Coghlan, I. Sakrejda, N. Desai, T. Declerck, and A. Liu, "Magellan: Experiences from a science cloud," In *Proceedings of the 2nd international workshop on Scientific cloud computing, ScienceCloud '11*, pages 49–58, San Jose, CA, June 2011, ACM.

Lavanya Ramakrishnan, Richard Shane Canon, Krishna Muriki, Iwona Sakrejda, and Nicholas J.Wright, "Evaluating interconnect and virtualization performance for high performance computing," In *Proceedings of 2nd International Workshop on Performance Modeling, Benchmarking and Simulation of High Performance Computing Systems (PMBS11)*, 2011.

Lavanya Ramakrishnan, Jeffrey Chase, Dennis Gannon, Daniel Nurmi, and Rich Wolski, "Deadline-sensitive workflow orchestration without explicit resource control," *J. Parallel Distributed Computing*, 71(3):343–353, March 2011.

Taghrid Samak, Dan Gunter, Ewa Deelman, Gideon Juve, Gaurang Mehta, Fabio Silva, and Karan Vahi," Online Fault and Anomaly Detection for Large-Scale Scientific Workflows," In *13th IEEE International Conference on High Performance Computing and Communications (HPCC-2011)*, Banff, Alberta, Canada, September 2011, IEEE Computer Society.

Taghrid Samak, Dan Gunter, Monte Goode, Ewa Deelman, Gaurang Mehta, Fabio Silva, and Karan Vahi, "Failure Prediction and Localization in Large Scientific Workflows," In *The Sixth Workshop on Workflows in Support of Large-Scale Science (WORKS11)*, Seattle, WA, USA, November 2011.

Taghrid Samak, Daniel Gunter, Monte Goode, Ewa Deelman, Gideon Juve, and Fabio Silva, "Using Machine Learning Techniques for Online Failure Prediction in Large Scientific Workflows," 2011.

2010

D. Agarwal, M. Humphrey, N. Beekwilder, K. Jackson, M. Goode, and C. van Ingen, "A data-centered collaboration portal to support global carbon-flux analysis," *Concurrency and Computation: Practice and Experience - Successes in Furthering Scientific Discovery*, December 2010, LBNL-1827E.

M. Altunay, D. Olson, and M. Helm, "OSG-ESnet Joint Identity Management Workshop Report," *OSG Doc 931*, January 2010.

S. J. Bailey, G. Aldering, P. Antilogus, C. Aragon, C. Baltay, S. Bongard, C. Buton, M. Chidress, N. Chotard, Y. Copin, H. Fakhouri, E. Gangler, S. Loken, P. Nugent, R. Pain, E. Pecontal, R. Pereira, S. Perlmutter, D. Rabinowitz, G. Rigaudier, K. Runge, R. Scalzo, G. Smadja, C. Tao, R. C. Thomas, and C. Wu, Using spectra to make cosmology's best standard candles even better," *Bulletin of the American Astronomical Society*, 42:313, 2010.

S. Binet, P. Calafiura, S. S nyder, W. Wiedenmann, and F. Winklmeier, "Harnessing multicores: strategies and implementations in ATLAS," *J. Phys. Conf.* (Ser,219:042002,2010), 2010, 17th International Conference on Computing in High Energy and Nuclear Phsics, Prague, Czech Republic, March 2009.

M. Bishop, J. Cummins, S. Peisert, A. Singh, B. Bhumiratana, D. Agarwal, D. Frincke, and M. Hogarth, Relationships and data sanitization: a study in scarlet, In *Proceedings of the 2010 workshop on New security paradigms*, pages 151–164, ACM, 2010.

Scott Callaghan, Ewa Deelman, Dan Gunter, Gideon Juve, Philip Maechling, Christopher Brooks, Karan Vahi, Kevin Milner, Robert Graves, Edward Field, David Okaya, and Thomas Jordan, "Scaling up Workflow-based Applications," *Journal of Computer and System Sciences*, 76(6):428–446, September 2010.

L. Hajdu, L. Dideno, J. Lauret, and D. Olson, "Comparison of cloud approaches: Overview and perspectives from STAR," In *International Symposium on Grid Computing*, Taipei, March 2010.

K. Jackson, L. Ramakrishnan, R. Thomas, and K.j, Runge, "Seeking supernovae in the clouds: A performance study," In *High Performance Distributed Computing*, Chicago, IL, June 2010.

Raj Kettimuthu, Alex Sim, Dan Gunter, Bill Allcock, PeerT, Bremer, John Bresnahan, Andrew Cherry, Lisa Childers, Eli Dart, Ian Foster, Kevin Harms, Jason Hick, Jason Lee, Michael Link, Jeff Long, Keith Miller, Vijaya Natarajan, Valerio Pascucci, Ken Raffenetti, David Ressman, Dean Williams, Loren Wilson, and Linda Winkler, "Lessons learned from moving Earth System Grid data sets over a 20 Gbps wide-area network," In *International Symposium on High Performance Distributed computing: Advanced Computing for Machinery*, June 2010.

O. Komogortsev, S. I. Jayarathna, C. Aragon, and M. Mahmoud, "Biometric identification via an oculomotor plant mathematical model," In *Eye Tracking Research and Applications, ETRA*, Austin, TX, 2010.

J. Li, D. Agarwal, M. Humphrey, C. van Ingen, K. Jackson, and Y. Ryu, "Escience in the cloud: A modis satellite data reprojection and reduction pipeline in the Windows Azure platform," In *24th IEEE International Parallel and Distributed Precessing Symposium, IPDPS*, Atlanta, Georgia, April 2010.

J. Li, M. Humphrey, Y,W, Cheah, Y. Ryu, D. Agarwal, K. Jackson, and C. van Ingen, "Fault Tolerance and Scaling in e-Science Cloud Applications: Observations from the Continuing Development of MODISAzure," In *2010 IEEE Sixth International Conference on eScience*, pp. 246–253, IEEE, 2010.

L. Ramakrishnan, K. Jackson, S. Canon, S. Cholia, and J. Shalf, "Defining future platform requirements for e-science cloud" (position paper), In *ACM Symposium on Cloud Computing*, Indianapolis, Indiana, June 2010.

A. Meier, C. Aragon, B. Hurwitz, D. Mujumdar, T. Peffer, D. Perry, and M. Pritoni, "How people actually use thermostats," In *The American Council for an Energy-Efficient Economy proceedings*, ACEEE 2010, 2010.

L. Ramakrishnan and B. Plale, "A multidimensional classification model for scientific workflow characteristics," In *1st Int'l Workshop on Workflow Approaches for New Data-Centric Science*, Indianapolis, Indiana, June 2010.

L. Ramakrishnan, D. Gannon, and B. Plale, "Workem: Representing and emulating disturbed scientific workflow execution state," In *10th IEEE/ACM International Symposium on Cluster, Cloud and Grip Computing, CCGid*, Melbourne, Australia, 2010.

L. Ramakrishnan, C. Guok, K. Jackson, E. Kissel, D. M, Swany, and D. Agarwal, "On-demand overlay networks for large scientific data transfer," In *10th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, CCGrid*, Melbourne, Australia, 2010.

R. Roberts, T. Trucano, P. Pope, C. Aragon, M. Jiang, T. Wei, L. Chilton, and Alan Bakel, "On the verification and validation of geospatial image analysis algorithms," In *IEEE International Geoscience and Remote Sensing Symposium, IGARSS*, 2010.

R. Scalzo, G. Aldering, P. Antilogus, C. Aragon, C. Baltay, S. Bailey, S. Bongard, C. Buton, M. Childress, N. Chotard, Y. Copin, E. Gangler, S. Loken, P. Nugent, R. Pain, E. Pecontal, R. Pereira, S. Perlmutter, D. Rabinowitz, G. Rigaudier, K. Runge, G. Smadja, H. Swift, C. Tao, R.C. Rabinowitz, G. Rigaudier, K. Runge, G. Smadja, H. Swift, C. Tao, R.C. Thomas, and C. Wu, "Nearby Supernova Factory observations of Sn 2007if: First Total Mass Measurement of a Super-Chandrasekhar-mass Progenitor," *Astrophysical Journal*, 713:1073–1094, March 2010.

R. A. Scalzo, C. Akerlof, G. Aldering, P. Antilogus, C. Aragon, S. Bailey, C. Baltay, A. Bauer, D. Birchall, S. Bongard, C. Buton, M. Childress, Y. Copin, A. Gal-Yam, E. Gangler, J. Holtzman, Sl Hoyer, M. Kasliwal, S. Loken, J. Miller, P. Nugent, R. Pain, E. Pecontal, R. Pereira, S. Perlmutter, C. Peters, G. Smadja, C. Tao, R.C. Thomas, J. Thorstensen, B. Weaver, C. Wu, and F. Yuan, "Nearby Supernova Factory Observations of Sn 2007if: Evidence for Shell Structure in the Explosion Of A Super-Chandrasekhar-Mass White Dwarf," AAS Meeting #215, #343,06: *Bulletin of the American Astronomical Society*, 42:450, 2010.

A. Sim, D. Gunter, V. Natarajan, A. Shoshani, D. Williams, J. Long, J. Hick, J. Lee, and E. Dart, "Efficient Bulk Data Replication for the Earth System Grid," *Data Driven E-science: Use Cases and Successful Applications of Distributed Computing Infrastructures (Isgc 2010)*, pp. 435, 2010.

A. Sim, D. Gunter, V. Natarajan, A. Shoshani, D. Williams, J. Long, J. Hick, J. Lee, and E. Dart, "Efficient bulk data replication for the Earth Systems Grid," In *International Symposium on Grid Computing, National Science Council*, March 2010.

Y. Simmhan and L. Ranakrishnan, "Comparison of resource platform selection approaches for scientific workflows," In *1st Workshop on Scientific Cloud Computing, Colocated with ACM and High Performance Distributed Computing*, Chicago, Illinois, 2010.

Y. Simmhan, E. Soroush, C. van Ingen, D. Agarwal, and L. Ramakrishnan, "Brew: Blackbox resource selection for e-science workflows," In *Fifth Workshop on Workflows in Support of Large-SCale Science(WORKS10)*, pp. 1–10, New Orleans, LA, November 2010, IEEE.

2009

Y. Yao, P. Calafirua, J. Poffet, A. Cavalli, C., Leggett, and B. Frederic, "Virtual machine logbook: Enabling virtualization for ATLAS," *J. Phys. Conference Series*, (219:032031), 2010, 17th International Conference on Computing in High Energy and Nuclear Physics 2009.

Cecilia R, Aragon, "Collaborative analytics for astrophysics explorations," Shimon Y. Nof, editor, In *Springer Handbook of Automation*, Springer-Verlag, 2009.

C. Aragon, S. Poon, A. Monroy-Hernandez, and D. Aragon, "A tale of two online communities: Fostering collaboration and creativity in scientists and children," In *ACM Conference on Creativity and Cognition*, Berkeley, CA, 2009.

C. R. Aragon, S. Poon, G. Aldering, R. C. Thomas, and R. Quimby, "Using visual analytics to develop situation awareness in astrophysics," *Journal of Information Visualization*, 2009.

S. Bailey, G. Aldering, P. Antilogus, C. Aragon, C. Baltay, S. Bongard, C. Buton, M. Childress, N. Chotard, E. Gangler Copin, S. Loken, P. Nugent, R. Pain, E. Pecontal, R. Pereira, S. Perlmutter, D. Rabinowitz, G. Rigaudier, K. Runge, R. Scalzo, G. Smadja, H. Swift, C,, Tao, R. C, Thomas, and C. Wu, "Using spectral flux ratios to standardize sn ia luminosities," *Astronomy & Astrophysics Letters*, 2009.

S. Bailey, G. Aldering, P. Antilogus, C. Aragon, C. Baltay, S. Bongard, C. Buton, M. Childress, N. Chotard, Y. Copin, E. Gangler, S. Loken, P. Nugent, R. Pain, E. Pecontal, R. Pereira, S. Perlmutter, D. Rabinowitz, G. Rigaudier, P. Ripoche, K. Runge, R. Scalzo, G. Smadja, C. Tao, R. Thomas, and C. Wu, "Generalized Spectral Correlations of Type Ia Supernovae From The Nearby Supernova Factory," *Bulletin of the American Astronomical Society*, 41, 2009.

S. Bailey, G. Aldering, P. Antilogus, C. Aragon, C. Baltay, S. Bongard, C. Buton, M. Childress, N. Chotard, Y. Copin, E. Gangler, S. Loken, P. Nugent, R. Pain, E Pecontal, R. Pereira, S. Perlmutter, D. Rabinowitz, G. Rigaudier, K. Runge, R. Scalzo, G. Smadja, H. Swift, C. Tao, R. C, Thomas, and C. Wu, "Using Spectral Flux Ratios To Standardize Sn Ia Luminosities," *Astronomy & Astrophysics Letters*, 2009.

B. Bargmeyer, "Metadata registries: Research, standards development and practice," *International Journal of Metadata, Semantics and Ontologies, Special Edition of Semantics and Ontologits*, 4(4), November 2009.

Bruce Bargmeyer, "Emerging semantic technology standards - making life easier in the information world," In *Proceedings of the 12th Open Forum on Metadata Registries*, Seoul, Korea, June 2009.

Bruce Bargmeyer (editor), "Metadata registries: Research, standards development and practice," *International Journal of Metadata, Special Edition of Semantics and Ontologies (IJMSO)*, 2009.

Bruce Bargmeyer and Sam Chance, "Environmental data: Edge issues and the path forward," Palle Haastrup and Jorgen Wuertz, editors, In *Environmental Data Exchange Network For Inland Water*, Elsevier, 2009, ISBN-13: 978-0-444-52973-2.

B. Bargmeyer, "Metadata registries and cloud computing-standards challenges," In *13th Open Forum on Metadata Registries*, Kunming, China, May 2009.

S. Callaghan, E. Deelman, and others (in submission), "Scaling up workflow-based applications," Journal of Computer and System Sciences, 2009, (Special Issue on Workflows).

K. Chard, J. Boverhof, W. Tan, R. Madduri, and I. Foster, "Wrap scientific applications as wsrf grid services using gravi," In *The International Conference on Web Services*, Los Angeles, CA, July 2009.

Y. Copin, G. Aldering, P. Antilogus, C. Aragon, S. Bailey, C. Baltay, S. Bongard, C. Buton, M. Childress, E. Gangler, J. Jerke, S. Loken, P. Nugent, R. Pain, E. Pecontal, R. Pereira, S. Perlmutter, D. Rabinovitz, G. Rigaudier, P. Ripoche, K. Runge, R. Scalzo, G. Smadja, H. Swift, C. Tao, R. C. Thomas, and C. Wu, "The Nearby Supernova Factory: First results," *EAS Publications Series*, 36:11–15, 2009.

C. R. Aragon and K. J, Runge, "Workflow management for high volume supernova search," In *ACM Symposium on Applied Computing SAC '09*, 2009.

Fred Gey, "Dynamic change in classification systems: A preliminary model for instance-based matching," Technical report, Lawrence Berkeley National Laboratory, 2009.

M. Humphrey, D. Agarwal, and C. van Ingen, "Fluxdata,org: Publication and Curation of Shared Scientific Climate and Earth Sciences Data," In *2009 Fifth IEEE International Conference on e-Science*, pages 118–125, IEEE, 2009.

Kevin Keck, Bruce Bargmeyer, et al., "Committee draft 1 of ISO/IEC 11179 information technology – metadata registries (mdr) part 3 edition 3," 2009, Standard - published for international ballot.

O. Komogortsev, S. Jayarathna, C. Aragon, and M. Mahmoud, "Biometric identificaton via an oculomotor plant mathematical model," Technical report, Texas State University, 2009,

J. Li, Y. Ryu, K. Jackson, D. Agarwal, M. Humphrey, and C. van Ingen, "Process-based estimates of terrestrial C02 assimilation and eveovs," In *Trans. American Geophysical Union*, San Francisco, CA, December 2009, Fall Meet, Suppl., Abstract B51B-0308.

J. Li, Y,Ryu, K. Jackson, D. Agarwal, M. Humphrey, and C. van Ingen, A cloud computing service for remote sensing data reprojection and science reduction" In *EOS Trans. American Geophysical Union*, San Francisco, CA, December 2009, Fall Meet Suppl., Abstract H51H-0857.

L,Ramkrishnan, D. Nurmi, A. Mandal, C. Koelbel, D. Gannon, T,M, Huang, Y. S, Kee, G. Oberteli, K. Thyagaraja, R. Wolski, A. Yarhan, and D. Zagorodnov, "Vgrads: Enabling escience workflows on grids and clouds with fault tolerance," In *ACM/IEEE SC2009 Conference on High Performance Computing, Networking, Storage and Analysis*, Portland, Oregon, November 2009.

C. McParland, "The evolution of the internet community and the 'yet-to-evolve' smart grid community: parallels and lessons-to-be-learned," In *Grid-Interop Forum*, Denver,CO, November 2009.

M. A. Piette, G. Ghatikar, S. Kiliccote, E. Koch, D. Hennage, P. Palensky, and C. McParland, "Open automated demand response communications specification (version 1,0)," Technical Report LBNL-1779E, Lawrence Berkeley National Laboratory, 2009, CEC OpenADR - Version 1,0 Report, California Energy Commission, PIER Program, CEC-500-2009-063.

Sarah S. Poon, Rollin C. Thomas, Cecilia R. Aragon, and Brian Lee, "Context-linked intelligent user interfaces," In *Workshop on Human Interaction with Intelligent and Networked Systems at the IUI International Conference on Intelligent User Interfaces*, Sanibel Island, FL, 2009.

Lavanya Ramakrishnan and Daniel A. Reed, "predictable quality of service atop degradable distributed systems," *Cluster Computing*, 2009.

L. Ramakrishnan, D. Nurmi, A. Mandal, C. Koelbel, D. Gannon, T. M, Huang, Y,-S, Kee, G. Oberteli, K. Thyagaraja, R. Wolski, A. Yarkhan, and D. Zagorodnov, "Automated analysis for detecting beams in laser wakefield simulations," In *ACM/IEEE SC* in proceedings, 2009.

Bronis R. Supinski, Sadaf Alam, David H. Bailey, Laura Carrington, Chris Daley, Anshu Dubey, Todd Gamblin, Dan Gunter, Paul D. Hovland, Heike Jagode, Karen Karavanic, Gabriel Marin, John Mellor-Crummey, Shirley Moore, Boyana Norris, Leonid Oliker, Catherine Olschanowsky, Philip C. Roth, Martin Schulz, Sameer Shende, Allan Snavely, Wyatt Spear, Mustafa Tikir, Jeff Vetter, Pat Worley, and Nicholas Wright, "Modeling the Office of Science ten year facilities plan: The PERI Architecture Tiger Team," *Journal of Physics: Conference Series*, 180(1):012039+, July 2009.

APPLIED MATHEMATICS AND SCIENTIFIC COMPUTING DEPARTMENT

APPLIED NUMERICAL ALGORITHMS GROUP

2012

B. Kallemov, G. H. Miller, S. Mitran and D. Trebotich, "Calculation of Viscoelastic Bead-Rod Flow Mediated by a Homogenized Kinetic Scale with Holonomic Constraints," *Molecular Simulation*, 2012.

G. H. Miller and D. Trebotich, "An Embedded Boundary Method for the Navier-Stokes Equations on a Time-Dependent Domain," *Communications in Applied Mathematics and Computational Science*, 7(1):1-31, 2012.

S. Molins, D. Trebotich, C. I. Steefel and C. Shen, "An Investigation of the Effect of Pore Scale Flow on Average Geochemical Reaction Rates Using Direct Numerical Simulation," *Water Resour. Res.*, 48(3), 2012.

S. Guzik, P. McCorquodale, P. Colella, "A Freestream-Preserving High-Order Finite-Volume Method for Mapped Grids with Adaptive-Mesh Refinement," *50th AIAA Aerospace Sciences Meeting* Nashville, TN, 2012.

P.S. Li, D.F. Martin, R.I. Klein, and C.F. McKee, "A Stable, Accurate Methodology for High Mach Number, Strong Magnetic Field MHD Turbulence with Adaptive Mesh Refinement: Resolution and Refinement Studies," *The Astrophysical Journal Supplement Series*, 2012.

2011

B. Wang, G.H. Miller, and P. Colella, "A Particle-in-Cell Method with Adaptive Phase-Space Remapping for Kinetic Plasmas," *SIAM J. Sci. Comput.*, 2011.

Q. Zhang, "High-Order, Multidimensional, and Conservative Coarse-Fine Interpolation for Adaptive Mesh Refinement," *Comput. Methods Appl. Mech. Engrg.*, 200(45-46):3159-3168 2011, 2011.

C. Shen, D. Trebotich, S. Molins, D. T. Graves, B. Van Straalen, et al., "High Performance Computations of Subsurface Reactive Transport Processes at the Pore Scale," *Journal of Physics: Conference Series, Proceedings of SciDAC* 2011, 2011.

F. Miniati and D.F. Martin, "Constrained-transport Magnetohydrodynamics with Adaptive Mesh Refinement in CHARM," *The Astrophysical Journal Supplement Series*, 195(1):5, 2011.

B. Wang, "A Particle-in-Cell Method with Adaptive Phase-space Remapping for Kinetic Plasmas," June 2011.

B. Van Straalen, P. Colella, D. T. Graves, N. Keen, "Petascale Block-Structured AMR Applications Without Distributed Metadata," *Euro-Par 2011 Parallel Processing-17th International Conference*, Euro-Par 2011, August 29 - September 2, 2011, Proceedings, Part II. Lecture Notes in Computer Science 6853 Springer 2011, Bordeaux, France, 2011.

B. Kallemov, G. H. Miller, S. Mitran and D. Trebotich, "Multiscale Rheology: New Results for the Kinetic Scale," *NSTI-Nanotech 2011*, Vol. 2, pp. 575-578 2011, 2011.

B. Kallemov, G. H. Miller and D. Trebotich, "A Higher-Order Accurate Fluid-Particle Algorithm for Polymer Flows," *Molecular Simulation*, 37(8):738-745 2011, 2011.

P. McCorquodale and P. Colella, "A High-Order Finite-Volume Method for Conservation Laws on Locally Refined Grids," *Communications in Applied Mathematics and Computational Science*, Vol. 6 2011, No. 1, 1-25, 2011.

P. Colella, M.R. Dorr, J.A.F. Hittinger, and D.F. Martin, "High-Order Finite-Volume Methods in Mapped Coordinates," Journal of Computational Physics, 230(8):2952-2976 2011, 2011.

2010

R.K. Crockett, P. Colella, and D.T. Graves, "A Cartesian Grid Embedded Boundary Method for Solving the Poisson and Heat Equations with Discontinuous Coefficients in Three Dimensions," *Journal of Computational Physics*, 230(7):2451-2469, 2010.

R.H. Cohen, J. Compton, M. Dorr, J. Hittinger, W.M Nevins, T.D. Rognlien, et al., "Testing and Plans for the COGENT edge kinetic code," (abstract) submitted to Sherwood 2010, 2010.

B. Kallemov, G. H. Miller and D. Trebotich, "Numerical Simulation of Polymer Flow in Microfluidic Devices," *009 Proceedings of the Fourth SIAM Conference on Mathematics for Industry (MI09)* pp. 93-98, 2010.

E. Ateljevich, P. Colella, D.T. Graves, T.J. Ligocki, J. Percelay, P.O. Schwartz, Q. Shu, "CFD Modeling in the San Francisco Bay and Delta," *2009 Proceedings of the Fourth SIAM Conference on Mathematics for Industry (MI09)*, pp. 99-107, 2010.

Caroline Gatti-Bono, Phillip Colella and David Trebotich, "A Second-Order Accurate Conservative Front-Tracking Method in One Dimension," *SIAM J. Sci. Comput.*, 31(6):4795-4813, 2010.

Schwartz, P., and Colella P., "A Second-Order Accurate Method for Solving the Signed Distance Function Equation," *Communications in Applied Mathematics and Computational Science*, Vol. 5 2010, No. 1, 81-97, 2010.

Prateek Sharma, Phillip Colella, and Daniel F. Martin, "Numerical Implementation of Streaming Down the Gradient: Application to Fluid Modeling of Cosmic Rays," *SIAM Journal on Scientific Computing*, Vol. 32(6), 3564-3583, 2010.

Q. Zhang, H. Johansen and P. Colella, "A Fourth-Order Accurate Finite-Volume Method with Structured Adaptive Mesh Refinement for Solving the Advection-Diffusion Equation," *SIAM Journal on Scientific Computing*, Vol. 34, No. 2. (2012), B179, 2010.

Q. Zhang and P. L.-F. Liu, "Handling Solid-Fluid Interfaces for Viscous Flows: Explicit Jump Approximation vs. Ghost Cell Approaches," *Journal of Computational Physics*, 229(11):4225-4246, 2010.

2009

B. Kallemov, G. H. Miller and D. Trebotich, "A Duhamel Approach for the Langevin Equations with Holonomic Constraints," *Molecular Simulation*, 35(6):440-447, 2009.

P. Colella, M. Dorr, J. Hittinger, D.F. Martin, and P. McCorquodale, "High-Order Finite-Volume Adaptive Methods on Locally Rectangular Grids," *2009 J. Phys.: Conf. Ser.* 180 012010, 2009.

A. Nonaka, D. Trebotich, G. H. Miller, D. T. Graves, and P. Colella, "A Higher-Order Upwind Method for Viscoelastic Flow," *Comm. App. Math. and Comp. Sci.*, 4(1):57-83, 2009.

P. Colella, M. Dorr, J. Hittinger, P.W. McCorquodale, and D.F. Martin, "High-Order Finite-Volume Methods on Locally-Structured Grids," *Numerical Modeling of Space Plasma Flows: Astronum 2008–Proceedings of the 3rd International Conference*, June 8-13, 2008, St John, U.S. Virgin Islands, 2009, pp. 207-216, 2009.

B.V. Straalen, J. Shalf, T. Ligocki, N. Keen, and W. Yang, "Scalability Challenges for Massively Parallel AMR Application," *23rd IEEE International Symposium on Parallel and Distributed Processing*, 2009.

COMPLEX SYSTEMS GROUP

2013

David H. Bailey, "A New Kind of Science: Ten years later," *Irreducibility and Computational Equivalence, edited by Hector Zenil*, Springer, 2013.

2012

David H. Bailey, Jonathan M. Borwein, "Experimental Applied Mathematics," *Princeton Companion for Applied mathematics*, (July 22, 2012).

David H. Bailey, Jonathan M. Borwein, "Normal numbers and pseudorandom generators," *Workshop on Computational and Analytical Mathematics* in Honour of Jonathan Borwein's 60th Birthday, Springer, 2012.

David H. Bailey, Jonathan M. Borwein, Christian S. Calude, et al., "An empirical approach to the normality of pi," *Experimental Mathematics*, 2012, 21:375-384.

David H. Bailey, Jonathan M. Borwein, "Ancient Indian square roots: An exercise in forensic paleo-mathematics," *American Mathematical Monthly*, October 2012, 119:646-657.

David H. Bailey and Marcos M. Lopez de Prado, "Balanced baskets: A new approach to trading and hedging risks," *Journal of Investment Strategies*, August 6, 2012, to appear.

David H. Bailey, Jonathan M. Borwein and Richard E. Crandall, "Computation and theory of extended Mordell-Tornheim-Witten sums," *Mathematics of Computation*, July 31, 2012, to appear.

David H. Bailey, Jonathan M. Borwein, "On Eulerian log-gamma integrals and Tornheim-Witten zeta functions," *Ramanujan Journal*, July 28, 2012.

Francisco J. Aragon Artacho, David H. Bailey, Jonathan M. Borwein, Peter B. Borwein, "Walking on real numbers," *Mathematical Intelligencer*, July 23, 2012, to appea.

David H. Bailey, Marcos M. Lopez de Prado, "The Sharpe ratio efficient frontier," *Journal of Risk*, July 20, 2012.

David H. Bailey and Jonathan M. Borwein, "Nonnormality of Stoneham constants," *Ramanujan Journal*, May 30, 2012, to appear.

David H. Bailey and Marcos M. Lopez de Prado, "The strategy approval decision: A Sharpe ratio indifference curve approach," *Social Science Research Network*, March 27, 2012.

David H. Bailey, Jonathan M. Borwein, "Exploration, Experimentation and Computation," *International Advances in Mathematics*, 2012, 31:1-14.

David H. Bailey and Jonathan M. Borwein, "Hand-to-hand combat with thousand-digit integrals," *Journal of Computational Science*, 2012, 3:77-86.

David H. Bailey, Roberto Barrio, Jonathan M. Borwein, "High precision computation: Mathematical physics and dynamics," *Applied Mathematics and Computation*, 2012, 218:10106-1012.

David H. Bailey, Jonathan M. Borwein, Richard E. Crandall, John Zucker, "Lattice sums arising from the Poisson equation," November 2, 2012.

David H. Bailey, Jonathan M. Borwein, Alexander D. Kaiser, "Automated simplification of large symbolic expressions," August 22, 2012.

David H. Bailey, Jonathan M. Borwein, Richard E. Crandall, Michael G. Rose, "Expectations on fractal sets," July 22, 2012.

David H. Bailey, Orianna DeMasi, Juan Meza, "Feature selection and multi-class classification using a rule ensemble method," May 25, 2012.

David H. Bailey, Jonathan M. Borwein, Cristian S. Calude, et al., "Normality and the digits of Pi," February 3, 2012.

A. Buluç, K. Madduri, "Graph partitioning for scalable distributed graph computations," *AMS Contemporary Mathematics, Proc. of the 10th DIMACS Implementation Challenge Workshop*, 2012.

A. Buluç, J. Gilbert, "Parallel sparse matrix-matrix multiplication and indexing: Implementation and experiments," *SIAM Journal on Scientific Computing (SISC)*, 2012.

K. Kandalla, A. Buluç, H. Subramoni, K. Tomko, J. Vienne, L. Oliker, D. K. Panda, "Can network-offload based non-blocking neighborhood MPI collectives improve communication overheads of irregular graph algorithms?" *International Workshop on Parallel Algorithms and Parallel Software (IWPAPS 2012)*, 2012.

A. Lugowski, D. Alber, A. Buluç, J. Gilbert, S. Reinhardt, Y. Teng, A. Waranis, "A flexible open-source toolbox for scalable complex graph analysis," *SIAM Conference on Data Mining (SDM)*, 2012.

A. Lugowski, A. Buluç, J. Gilbert, S. Reinhardt, "Scalable complex graph analysis with the knowledge discovery toolbox," *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, March 2012.

Matt Bishop, Sean Peisert, "Security and Elections," *IEEE Security & Privacy*, September 2012, 10(5):64-67.

Sean Whalen, Sophie Engle, Sean Peisert, Matt Bishop, "Network-Theoretic Classification of Parallel Computation Patterns," *International Journal of High Performance Computing Applications (IJHPCA)*, May 2012, 26(2):159-169.

Xiao Li, Zhifang Wang, Vishak Muthukumar, Anna Scaglione, Chuck McParland, Sean Peisert, "Networked Loads in the Distribution Grid," *Proceedings of the 2012 APSIPA Annual Summit and Conference*, Hollywood, CA, December 3, 2012.

Sean Peisert, Ed Talbot, Matt Bishop, "Turtles All the Way Down: A Clean-Slate, Ground-Up, First-Principles Approach to Secure Systems," *Proceedings of the 2012 New Security Paradigms Workshop (NSPW)*, ACM, September 2012.

2011

A. Buluç, J. R. Gilbert, V. B. Shah, "Implementing Sparse Matrices for Graph Algorithms," *Graph Algorithms in the Language of Linear Algebra*. SIAM Press, 2011.

A. Buluç, J. R. Gilbert, "New Ideas in Sparse Matrix-Matrix Multiplication," *Graph Algorithms in the Language of Linear Algebra*. SIAM Press, 2011.

Samuel W. Williams, David H. Bailey, "Parallel Computer Architecture," *Performance Tuning of Scientific Applications*, edited by David H. Bailey, Robert F. Lucas, Samuel W. Williams, CRC Press, 2011, Pages: 11-33.

David H. Bailey, Lin-Wang Wang, Hongzhang Shan, Zhengji Zhao, Juan Meza, et al., "Tuning an electronic structure code," *Performance Tuning of Scientific Applications*, edited by David H. Bailey, Robert F. Lucas, Samuel W. Williams, (CRC Press: 2011) Pages: 339-354.

David H. Bailey, Allan Snavely, Laura Carrington, "Performance modeling: The convolution approach," *Performance Tuning of Scientific Applications*, edited by David H. Bailey, Robert F. Lucas, Samuel W. Williams, (CRC Press: 2011) Pages: 151-164.

David H. Bailey, Robert F. Lucas, Samuel W. Williams, "Introduction," *Performance Tuning of Scientific Applications*, edited by David H. Bailey, Robert F. Lucas and Samuel W. Williams, (CRC Press: 2011) Pages: 1-10.

David H. Bailey, Robert F. Lucas, Samuel W. Williams, ed., *Performance Tuning of Scientific Applications*, CRC Press: 2011.

Orianna DeMasi, Juan Meza and David H. Bailey, "Dimension reduction using rule ensemble machine learning methods: A numerical study of three ensemble methods," August 30, 2011.

A. Buluç, J. Gilbert, "The Combinatorial BLAS: Design, implementation, and applications," *International Journal of High-Performance Computing Applications (IJHPCA)*, 2011.

A. Buluç, K. Madduri, "Parallel breadth-first search on distributed memory systems," *Supercomputing (SC)*, November 2011.

A. Buluç, S. Williams, L. Oliker, J. Demmel, "Reduced-Bandwidth Multithreaded Algorithms for Sparse Matrix-Vector Multiplication," *International Parallel Distributed Processing Symposium (IPDPS)*, May 2011.

David H. Bailey and Jonathan M. Borwein, "The greatest mathematical discovery?" May 1, 2011.

David H. Bailey, "A compendium of BBP-type formulas for mathematical constants," January 14, 2011.

A. Kaiser, S. Williams, K. Madduri, K. Ibrahim, D. Bailey, J. Demmel, E. Strohmaier, "TORCH Computational Reference Kernels: A Testbed for Computer Science Research," LBNL Technical Report, 2011.

David H. Bailey, Jonathan M. Borwein, "Exploratory experimentation and computation," *Notices of the American Mathematical Society*, November 2011, 58:1410-1419.

David H. Bailey, Jonathan M. Borwein, Andrew Mattingly, Glenn Wightwick, "The computation of previously inaccessible digits of Pi^2 and Catalan's constant," *Notices of the American Mathematical Society*, July 5, 2011.

David H. Bailey, Jonathan M. Borwein, "High-precision numerical integration: Progress and challenges," *Journal of Symbolic Computation*, 2011, 46:741-754.

Matt Bishop, Justin Cummins, Sean Peisert, Bhume Bhumitarana, Anhad Singh, et al., "Relationships and Data Sanitization: A Study in Scarlet," *Proceedings of the 2010 New Security Paradigms Workshop (NSPW)*, Concord, MA, ACM, September 2011, 151-164.

Sean Whalen, Sean Peisert, Matt Bishop, "Network-Theoretic Classification of Parallel Computation Patterns," *Proceedings of the First International Workshop on Characterizing Applications for Heterogeneous Exascale Systems (CACHES)*, Tucson, AZ, IEEE Computer Society, June 4, 2011.

2010

David H. Bailey, Jonathan M. Borwein, David Broadhurst and Wadim Zudlin, "Experimental mathematics and mathematical physics," *Gems in Experimental Mathematics*, edited by Tewodros Amdeberhan, Luis A. Medina, Victor H. Moll, (American Mathematical Society: January 1, 2010) Pages: 41-58.

David H. Bailey, Jonathan M. Borwein, "Experimental computation with oscillatory integrals," *Gems in Experimental Mathematics*, edited by Tewodros Amdeberhan, Luis A. Medina, Victor H. Moll, (American Mathematical Society: January 1, 2010) Pages: 25-40.

E. Strohmaier, S. Williams, A. Kaiser, K. Madduri, K. Ibrahim, D. Bailey, J. Demmel,, "A Kernel Testbed for Parallel Architecture, Language, and Performance Research," *International Conference of Numerical Analysis and Applied Mathematics (ICNAAM)*, June 1, 2010.

A. Kaiser, S. Williams, K. Madduri, K. Ibrahim, D. Bailey, J. Demmel, E. Strohmaier, "A Principled Kernel Testbed for Hardware/Software Co-Design Research," *Proceedings of the 2nd USENIX Workshop on Hot Topics in Parallelism (HotPar)*, 2010.

A. Buluç, J. R. Gilbert, C. Budak, "Solving path problems on the GPU," *Parallel Computing*, 36(5-6):241 - 253., 2010.

David H. Bailey, Jonathan M. Borwein, Richard E. Crandall, "Advances in the theory of box integrals," *Mathematics of Computation*, July 2010, 79:1839-1966.

Peter G. Neumann, Matt Bishop, Sean Peisert, Marv Schaefer, "Reflections on the 30th Anniversary of the IEEE Symposium on Security and Privacy," *Proceedings of the 31st IEEE Symposium on Security and Privacy*, Oakland/Berkeley, CA, IEEE Computer Society, May 2010, 3-13.

2009

David H. Bailey, "The NAS Parallel Benchmarks," *Encyclopedia of Parallel Computing*, (Springer: November 1, 2009).

David H. Bailey, Jonathan M. Borwein, "Experimental mathematics and computational statistics," *Wiley Interdisciplinary Reviews: Computational Statistics*, edited by Edward Wegman, Yasmin H. Said, David W. Scott, (April 2009).

A. Buluç, J. T. Fineman, M. Frigo, J. R. Gilbert, C. E. Leiserson, "Parallel sparse matrix-vector and matrix-transpose-vector multiplication using compressed sparse blocks," *SPAA '09 Proceedings of the twenty-first annual symposium on Parallelism in algorithms and architectures*, 2009.

Zhengji Zhao, Juan Meza, Byounghak Lee, Hongzhang Shan, Eric Strohmaier, et al., "The linearly scaling 3D fragment method for large scale electronic structure calculations," *Journal of Physics: Conference Series*, July 1, 2009.

David H. Bailey, "Misleading performance claims in scientific computation," *Proceedings of DAC2009*, May 2009.

Juan Meza, Scott Campbell, David H. Bailey, "Mathematical and statistical opportunities in cyber security," March 1, 2009.

David H. Bailey, Jonathan M. Borwein, "PSLQ: An algorithm to discover integer relations," *Computeralgebra-Rundbrief*, October 2009, 45:8-11.

David H. Bailey, "Can an evolutionary process generate English text?" *Biological Theory*, March 1, 2009, 4:125-131.

Bronis R. Supinski, Sadaf Alam, David H. Bailey, Laura Carrington, et al., "Modeling the Office of Science ten year facilities plan: The PERI Architecture Tiger Team," *Journal of Physics: Conference Series*, 2009.

David H. Bailey, Jonathan M. Borwein, Richard E. Crandall, "Resolution of the Quinn-Rand-Strogatz constant of nonlinear physics," *Experimental Mathematics*, 2009, 18:107-116.

Sean Peisert, Matt Bishop, Laura Corriss, Steven J. Greenwald, "Quis Custodiet ipsos Custodes? A New Paradigm for Analyzing Security Paradigms," *Proceedings of the 2009 New Security Paradigms Workshop (NSPW)*, The Queen's College, Oxford, United Kingdom, ACM, September 2009, 133-144.

Matt Bishop, Sean Peisert, Candice Hoke, Mark Graff, David Jefferson, "E-Voting and Forensics: Prying Open the Black Box," *Proceedings of the 2009 Electronic Voting Technology Workshop/Workshop on Trustworthy Elections (EVT/WOTE)*, Montreal, Canada, USENIX, August 2009.

SCIENTIFIC COMPUTING GROUP

2012

Abhinav Sarje, Xiaoye S. Li, Slim Chourou, Elaine R. Chan, Alexander Hexemer, "Massively Parallel X-ray Scattering Simulations," *Supercomputing*, November 2012.

S. Chourou, A. Sarje, X. Li, E. Chan, A. Hexemer, "High-Performance GISAXS Code for Polymer Science," *Synchrotron Radiation in Polymer Science*, April 2012.

Slim T. Chourou, Abhinav Sarje, Xiaoye Li, Elaine Chan and Alexander Hexemer, "HipGISAXS: A High-Performance Software for Simulating GISAXS Data," *Macromolecules*, 2012.

A. Canning, J. Shalf, N.J. Wright, S. Anderson and M. Gajbe, "A Hybrid MPI/OpenMP 3d FFT for Plane Wave First-principles Materials Science Codes," *Proceedings of CSC12 Conference* (2012).

S. Derenzo, G. Zhang, A. Canning, E. Bourret-Courchesne, "Experimental and Theoretical Studies of Donor-Acceptor Scintillation from PbI2," to appear in *Journal of Luminescence* (2012).

Gaigong Zhang, Andrew Canning, Niels Gronbech-Jensen, S. Derenzo, L-W Wang, "Shallow impurity level calculations in semiconductors using *ab initio* methods," Submitted to *Physical Review Letters* (2012)

Richard L. Martin, Thomas F. Willems, Li-Chiang Lin, Jihan Kim, et al., "Similarity-Driven Discovery of Zeolite Materials for Adsorption-Based Separations," *ChemPhysChem*, August 22, 2012.

Jihan Kim, Li-Chiang Lin, Richard L. Martin, Joseph A. Swisher, Maciej Haranczyk, et al., "Large-Scale Computational Screening of Zeolites for Ethane/Ethene Separation," Langmuir, July 11, 2012.

Li-Chiang Lin, Adam H. Berger, Richard L. Martin, Jihan Kim, et al., "In Silico Screening of Carbon Capture Materials," *Nature Materials*, May 27, 2012, 11:633–641.

Jihan Kim, Richard L. Martin, Oliver Rübel, Maciej Haranczyk and Berend Smit, "High-throughput Characterization of Porous Materials Using Graphics Processing Units," *Journal of Chemical Theory and Computation*, March 16, 2012, 8:1684–1693, LBNL 5409E.

Richard L. Martin, Prabhat, David D. Donofrio, James A. Sethian and Maciej Haranczyk, "Accelerating Analysis of void spaces in porous materials on multicore and GPU platforms," *International Journal of High Performance Computing Applications*, February 5, 2012, 26:347-357.

J. Gonzalez-Domınguez, O. Marques, M. Martın, G. Taboada, J. Tourino, "Design and Performance Issues of Cholesky and LU Solvers using UPCBLAS," *10th IEEE International Symposium on Parallel and Distributed Processing with Applications*, Madrid, 2012.

H. Hu, C. Yang, K. Zhao, "Absorption correction A* for cylindrical and spherical crystals with extended range and high accuracy calculated by Thorkildsen & Larsen analytical method," in press Acta Crystallographica, A, 2012.

H. M. Aktulga, C. Yang, E. G. Ng, P. Maris, J. P. Vary, "Improving the Scalability of a Symmetric Iterative Eigensolver for Multi-core Platforms," submitted to *Concurrency and Computation: Practice & Experience*, November 16, 2012.

H. M. Aktulga, L. Lin, C. Haine, E. G. Ng, C. Yang, "Parallel Eigenvalue Calculation based on Multiple Shift-invert Lanczos and Contour Integral based Spectral Projection Method," submitted to *Parallel Computing*, November 13, 2012.

Zaiwen Wen, Chao Yang, Xin Liu, Stefano Marchesini, "Alternating direction methods for classical and ptychographic phase retrieval," Inverse Problems, November 2012, 28.
L. Lin, C. Yang, "Elliptic preconditioner for accelerating the self-consistent field iteration in Kohn-Sham Density Functional Theory," submitted to *SIAM J. Sci. Comp*., 2012.

D. Y. Parkinson, C. Yang, C. Knoechel, C. A. Larabell, M. Le Gros, "Automatic alignment and reconstruction of images for soft X-ray tomography," *J Struct Biol*, February 2012, 177:259—266.

L. Lin, M. Chen, C. Yang, L. He, "Accelerating Atomic Orbital-based Electronic Structure Calculation via Pole Expansion and Selected Inversion," submitted to *J. Comp. Phys*., 2012.

P. Maris, H. M. Aktulga, S. Binder, A. Calci, U. V. Catalyurek, et al., "No Core CI calculations for light nuclei with chiral 2- and 3-body forces," *J. Phys. Conf. Ser*., in press, December 21, 2012.

P. Maris, H. M. Aktulga, M. A. Caprio, U. V. Catalyurek, E. G. Ng, et al., "Large-scale Ab-initio Configuration Interaction Calculations for Light Nuclei," *J. Phys. Conf. Ser*., December 18, 2012, 403 0120.

Z. Zhou, E. Saule, H. M. Aktulga, C. Yang, E. G. Ng, P. Maris, J. P. Vary, U. V. Catalyurek, "An Out-of-core Eigensolver on SSD-equipped Clusters," *2012 IEEE International Conference on Cluster Computing (CLUSTER)*, Beijing, China, September 26, 2012, 248 – 256.

Z. Zhou, E. Saule, H. M. Aktulga, C. Yang, E. G. Ng, P. Maris, J. P. Vary, U. V. Catalyurek, "An Out-Of-Core Dataflow Middleware to Reduce the Cost of Large Scale Iterative Solvers," *2012 41st International Conference on Parallel Processing Workshops (ICPPW)*, Pittsburgh, PA, September 10, 2012, 71 – 80.

H. M. Aktulga, C. Yang, P. Maris, J. P. Vary, E. G. Ng, "Topology-Aware Mappings for Large-Scale Eigenvalue Problems," *Euro-Par 2012 Parallel Processing Conference*, Rhode Island, Greece, August 31, 2012, LNCS 748:830-842.

H. M. Aktulga, L. Lin, C. Haine, E. G. Ng, C. Yang, "Parallel Eigenvalue Calculation based on Multiple Shift-invert Lanczos and Contour Integral based Spectral Projection Method," submitted to *Parallel Computing*, November 13, 2012.

H. M. Aktulga, C. Yang, E. G. Ng, P. Maris, J. P. Vary, "Improving the Scalability of a Symmetric Iterative Eigensolver for Multi-core Platforms," submitted to *Concurrency and Computation: Practice & Experience*, November 16, 2012.

H. M. Aktulga, J. C. Fogarty, S. A. Pandit, A. Y. Grama, "Parallel reactive molecular dynamics: Numerical methods and algorithmic techniques," *Parallel Computing*, 38 (4-5), 245-259, April 2, 2012

H. M. Aktulga, S. A. Pandit, A. C. T. van Duin, A. Y. Grama, "Reactive Molecular Dynamics: Numerical Methods and Algorithmic Tehniques," *SIAM J. Sci. Comput.*, 34(1), C1–C23, January 31, 2012

C. Mendl, L. Lin, "Towards the Kantorovich dual solution for strictly correlated electrons in atoms and molecules," submitted to *Phys. Rev. B*, 2012.

L. Lin, S. Shao, W.E, "Efficient iterative method for solving the Dirac-Kohn-Sham density functional theory," submitted to *J. Comput. Phys*., 2012.

L. Lin, L. Ying, "Element orbitals for Kohn-Sham density functional theory," *Phys. Rev. B*, 2012, 85:235144.

L. Lin, C. Yang, "Elliptic preconditioner for accelerating the self-consistent field iteration in Kohn-Sham Density Functional Theory," submitted to *SIAM J. Sci. Comp*., 2012.

L. Lin, J. Lu, L. Ying and W. E, "Optimized local basis set for Kohn-Sham density functional theory," *J. Comput. Phys*., 2012, 231:4515.

L. Lin, J. Lu, L. Ying and W. E, "Adaptive local basis set for Kohn-Sham density functional theory in a discontinuous Galerkin framework I: Total energy calculation," *J. Comput. Phys*., 2012, 231:2140.

D. Flammini, A. Pietropaolo, R. Senesi, C. Andreani, F. McBride, et al., "Spherical momentum distribution of the protons in hexagonal ice from modeling of inelastic neutron scattering data," *J. Chem. Phys*., 2012, 136:024504.

L. Lin, M. Chen, C. Yang, L. He, "Accelerating Atomic Orbital-based Electronic Structure Calculation via Pole Expansion and Selected Inversion," submitted to *J. Comp. Phys*., 2012.

H.-M. Eiter, M. Lavagnini, R. Hackl, E.A. Nowadnick, A.F. Kemper, T.P et al., "Alternative route to charge density wave formation in multiband systems," *Proceedings of the National Academy of Sciences*, 2012.

A.F. Kemper, M. Sentef, B. Moritz, C.C. Kao, Z.X. Shen, J.K. Freericks, T.P. Devereaux, "Mapping of the unoccupied states and relevant bosonic modes via the time dependent momentum distribution," *arXiv* 2012.

B. Moritz, A. F. Kemper, M. Sentef, T. P. Devereaux, J. K. Freericks, "Correlation tuned cross-over between thermal and nonthermal states following ultrafast transient pumping," *arXiv*, 2012.

J. K. Freericks, A. Y. Liu, A. F. Kemper, T. P. Devereaux, "Pulsed high harmonic generation of light due to pumped Bloch oscillations in noninteracting metals," *Physica Scripta*, 2012,

P. Maris, H. M. Aktulga, S. Binder, A. Calci, U. V. Catalyurek, et al., "No Core CI calculations for light nuclei with chiral 2- and 3-body forces," *J. Phys. Conf. Ser.*, in press, December 21, 2012.

P. Maris, H. M. Aktulga, M. A. Caprio, U. V. Catalyurek, E. G. Ng, et al., "Large-scale Ab-initio Configuration Interaction Calculations for Light Nuclei," *J. Phys. Conf. Ser.*, December 18, 2012, 403 0120.

Z. Zhou, E. Saule, H. M. Aktulga, C. Yang, E. G. Ng, P. Maris, J. P. Vary, U. V. Catalyurek, "An Out-of-core Eigensolver on SSD-equipped Clusters," *2012 IEEE International Conference on Cluster Computing (CLUSTER)*, Beijing, China, September 26, 2012, 248 – 256.

Z. Zhou, E. Saule, H. M. Aktulga, C. Yang, E. G. Ng, P. Maris, J. P. Vary, U. V. Catalyurek, "An Out-Of-Core Dataflow Middleware to Reduce the Cost of Large Scale Iterative Solvers," *2012 41st International Conference on Parallel Processing Workshops (ICPPW)*, Pittsburgh, PA, September 10, 2012, 71 – 80.

H. M. Aktulga, C. Yang, P. Maris, J. P. Vary, E. G. Ng, "Topology-Aware Mappings for Large-Scale Eigenvalue Problems," *Euro-Par 2012 Parallel Processing Conference*, Rhode Island, Greece, August 31, 2012.

Fuyu Li, William D. Collins, Michael F. Wehner, Ruby L. Leung, "Hurricanes in an Aquaplanet World: Implications of the Impacts of External Forcing and Model Horizontal Resolution," to appear in the *Journal of Modeling the Earth System*, 2012.

Seung-Ki Min, Xuebin Zhang, Francis Zwiers, Hideo Shiogama, Yu-Shiang Tung, and Michael Wehner, "Multi-Model Detection and Attribution of Extreme Temperature Changes," submitted to *Journal of Climate*, 2012.

Donald Wuebbles, Gerald Meehl, Katharine Hayhoe, Thomas R. Karl, Kenneth Kunkel, Benjamin Santer, Michael Wehner, Brian Colle, Erich M. Fischer, Rong Fu, Alex Goodman, Emily Janssen, Huikyo Lee, Wenhong Li, Lindsey N. Long, Seth Olsen, Anji Seth, Justin Sheffield, Liqiang Sun "CMIP5 Climate Model Analyses: Climate Extremes in the United States," submitted to the *Bulletin of the American Meteorological Society*, 2012.

V. V. Kharin, F. W. Zwiers, X. Zhang, M. Wehner "Changes in temperature and precipitation extremes in the CMIP5 ensemble," submitted to *Climatic Change*, 2012.

Russell S. Vose, Scott Applequist, Mark A. Bourassa, Sara C. Pryor, Rebecca J. Barthelmie, Brian Blanton, Peter D. Bromirski, Harold E. Brooks, Arthur T. DeGaetano, Randall M. Dole, David R. Easterling, Robert E. Jensen, Thomas R. Karl, Katherine Klink, Richard W. Katz, Michael C. Kruk, Kenneth E. Kunkel, Michael C. MacCracken, Thomas C. Peterson, Bridget R. Thomas, Xiaolan L. Wang, John E. Walsh, Michael F. Wehner, Donald J. Wuebbles, and Robert S. Young "Monitoring and Understanding Changes in Extremes: Extratropical Storms, Winds, and Waves," submitted to the *Bulletin of the American Meteorological Society*, 2012.

John C. H. Chiang, C. Y. Chang and M.F. Wehner (2012) Long-term trends of the Atlantic Interhemispheric SST Gradient in the CMIP5 Historical Simulations, Submitted to J. Climate.

Santer, B.D., J. Painter, C. Mears, C. Doutriaux, P. Caldwell, J.M. Arblaster, P. Cameron-Smith, N.P. Gillett, P.J. Gleckler, J.R. Lanzante, J. Perlwitz, S. Solomon, P.A. Stott, K.E. Taylor, L. Terray, P.W. Thorne, M.F. Wehner, F.J. Wentz, T.M.L. Wigley, L. Wilcox and C.-Z. Zou "Identifying Human Influences on Atmospheric Temperature: Are Results Robust to Uncertainties?" To appear in *Proceedings of the National Academy of Sciences*, 2012.

Prabhat, Oliver Ruebel, Surendra Byna, Kesheng Wu, Fuyu Li, Michael Wehner and Wes Bethel "TECA: A Parallel Toolkit for Extreme Climate Analysis," *International Conference on Computational Science, ICCS 2012, Workshop on Data Mining in Earth System Science, Procedia Computer Science*, 2012.

Michael F. Wehner (2012) Methods of Projecting Future Changes in Extremes. Chapter 8 in *Extremes in a Changing Climate: Detection, Analysis and Uncertainty*, A. AghaKouchak et al. (eds.), Water Science and Technology Library 65, DOI 10.1007/978-94-007-4479-0 8, Springer.

Fuyu Li, Daniele Rosa, William D. Collins, and Michael F. Wehner "'Super-Parameterization'–a Better Way to Simulate Regional Extreme Precipitation?" To appear in *Journal of Advances in Modeling Earth Systems*, 2012.

T.C. Peterson, R. Heim, R. Hirsch, D. Kaiser, H. Brooks, N.S. Diffenbaugh, R. Dole, J. Giovannettone, K. Guiguis, T.R. Karl, R.W. Katz, K. Kunkel, D. Lettenmaier, G. J. McCabe, C.J. Paciorek, K.Ryberg, S.Schubert, V.B.S. Silva, B. Stewart, A.V. Vecchia, G. Villarini, R.S. Vose, J. Walsh, M.Wehner, D. Wolock, K. Wolter, C.A. Woodhouse and D. Wuebbles "Monitoring and Understanding Changes in Heatwaves, Coldwaves, Floods and Droughts in the United States: State of Knowledge," to appear in the *Bulletin of the American Meteorological Society*, 2012.

K.E. Kunkel, T.R. Karl, H. Brooks, J. Kossin, J. Lawrimore, D. Arndt, L. Bosart, D. Changnon, S. Cutter, N. Doesken, K. Emanuel, P.Ya. Groisman, R.W. Katz, T. Knutson, J. O'Brien, C. Paciorek, T. Peterson, K. Redmond, D. Robinson, J. Trapp, R. Vose, S. Weaver, M. Wehner, K. Wolter, D. Wuebbles "Monitoring and Understanding Trends in Extreme Storms: State of Knowledge," To appear in the *Bulletin of the American Meteorological Society*, 2012. 10.1175/BAMS-D-11-00262.1

2011

Michael F. Wehner, Leonid Oliker, John Shalf, David Donofrio, Leroy A. Drummond, Ross Heikes, Shoaib Kamil, Celal Konor, Norman Miller, Huiro Miura, Marghoob Mohiyuddin, David Randall, Woo-Sun Yang, "Hardware/Software Co-design of Global Cloud System Resolving Models," *Journal of Advances in Modeling Earth Systems 3*, M10003, 2011.

Surendra Byna, Prabhat, Michael F. Wehner, Kesheng Wu, "Detecting Atmospheric Rivers in Large Climate Datasets," *Proceedings of the 2nd International Workshop on Petascale Data Analytics: Challenges, and Opportunities* (PDAC-11/ Supercomputing11/ ACM/IEEE) November 14, 2011. Seattle, WA, USA.

B.D. Santer, Carl Mears , Charles Doutriaux , Peter Gleckler , Tom Wigley , Nathan Gillett , Detelina Ivanova , Thomas Karl , John Lanzante , Gerald Meehl , Peter Stott , Karl Taylor , Peter Thorne , Michael Wehner , Frank Wentz, "Separating Signal and Noise in Atmospheric Temperature Changes: The Importance of Timescale," *Journal of Geophysical Research-Atmospheres* 116, D22105, 2011.

John Shalf, David Donofrio, Chris Rowen, Leonid Oliker, Michael Wehner, "Green Flash: Climate Machine," to appear in *The Encyclopedia of Parallel Computing*, 2011.

Michael Wehner, David R. Easterling, Jay H. Lawrimore, Richard R. Heim Jr., Russell S. Vose, Benjamin Santer, "Projections of Future Drought in the Continental United States and Mexico," *Journal of Hydrometerology* 12, 1359–1377, 2011.

Fuyu Li, William Collins, Michael Wehner, David Williamson, Jerry Olson "Response of precipitation extremes to idealized global warming in an aqua-planet climate model: towards a robust projection across different horizontal resolutions," *Tellus* 63, 876-883, 2011.

Fuyu Li, William Collins, Michael Wehner, David Williamson, Jerry Olson, Christopher Algieri, "Impact of horizontal resolution on simulation of precipitation extremes in an aqua-planet version of Community Atmospheric Model (CAM3)," *Tellus* 63, 884-823, 2011.

Chang, C.-Y., J. C. H. Chiang, M. F. Wehner, A. R. Friedman, R. Ruedy, "Sulfate Aerosol Control of Tropical Atlantic Climate over the Twentieth Century," *Journal of Climate*, 24, 2540–2555, 2011.

A. Chaudhry, A, Canning, R. Boutchko, M. Weber, N. Gronbech-Jensen and S. Derenzo, "First-principles studies of Ce-doped RE2M2O7 (RE=Y, La; M = Ti, Zr, Hf): A class of non-scintillators," *Journal of Applied Physics*, Vol. 109, pp 083708, 2011.

A. Canning, A. Chaudhry, R. Boutchko and N. Gronbech-Jensen, "First principles studies of luminescence in Ce doped inorganic scintillators," *Phys. Rev. B* 83, 125115 2011.

Richard L. Martin, Berend Smit and Maciej Haranczyk, "Addressing Challenges of Identifying Geometrically Diverse Sets of Crystalline Porous Materials," *Journal of Chemical Information and Modeling*, November 18, 2011, 52:308–318.

Beketayev, K., Weber, G.H., Haranczyk, M., Bremer, P.-T., Hlawitschka, M., and Hamann, B., "Topology-based Visualization of Transformation Pathways in Complex Chemical Systems," *Computer Graphics Forum* (Special Issue, Proc. Eurographics / IEEE Symposium on Visualization), June 2011, 663-672, LBNL 5242E.

Maciej Haranczyk, Richard L. Martin, Prabhat, James A. Sethian and E. Wes Bethel, "Computational Approaches for the High-Throughput Analysis of Porous materials for Energy related applications," *Scientific Discovery through Advanced Computing*, 2011.

Jihan Kim, Alice Koniges, Richard L. Martin, Maciej Haranczyk, et al., "GPU Computational Screening of Carbon Capture Materials," *Scientific Discovery through Advanced Computing 2011*, 2011.

Filipe R.N.C. Maia, Chao Yang, Stefano Marchesini, "Compressive auto-indexing in femtosecond nanocrystallography," *Ultramicroscopy*, 2011, 111:807--811, LBNL 4598E.

L. Lin, C. Yang, J. Lu, L. Ying, W. E, "A fast parallel algorithm for selected inversion of structured sparse matrices with application to 2D electronic structure calculations," *SIAM J. Sci. Comput.*, 2011, 33:1329.

Lin, C. Yang, J. Meza, J. Lu, L. Ying, W. E, "SelInv -- An algorithm for selected inversion of a sparse symmetric matrix," *ACM Trans. Math. Software*, 2011, 37:40.

H. M. Aktulga, C. Yang, U. V. Catalyurek, P. Maris, J. P. Vary, E. G. Ng, "On Reducing I/O Overheads in Large-Scale Invariant Subspace Projections," *Euro-Par 2011: Parallel Processing Workshops*, Bordeaux, France, August 29, 2011, LNCS 715:305-314.

E. G. Ng, J. Sarich, S. M.Wild, T. Munson, H. M. Aktulga, C. Yang, et al., "Advancing Nuclear Physics Through TOPS Solvers and Tools," *SciDAC 2011 Conference*, Denver, CO, July 10, 2011, arXiv:1110.1708.

H. M. Aktulga, C. Yang, P. Maris, J. P. Vary, E. G. Ng, "Large-scale Parallel Null Space Calculation for Nuclear Configuration Interaction," *2011 International Conference on High Performance Computing and Simulation (HPCS)*, Istanbul, Turkey, July 8, 2011, 176 – 185.

L. Lin, J.A. Morrone and R. Car, "Correlated tunneling in hydrogen bonds," *J. Stat. Phys*., 2011, 145:365.

L. Lin, J.A. Morrone, R. Car, M. Parrinello, "Momentum distribution, vibrational dynamics and the potential of the mean force in ice," *Phys. Rev. B* (Rapid Communication), 2011, 83:220302.

L. Lin, J. Lu and L. Ying, "Fast construction of hierarchical matrix representation from matrix-vector multiplication," *J. Comput. Phys*., 2011, 230:4071.

H. M. Aktulga, C. Yang, U. V. Catalyurek, P. Maris, J. P. Vary, E. G. Ng, "On Reducing I/O Overheads in Large-Scale Invariant Subspace Projections," *Euro-Par 2011: Parallel Processing Workshops*, Bordeaux, France, August 29, 2011, LNCS 715:305-314.

E. G. Ng, J. Sarich, S. M.Wild, T. Munson, H. M. Aktulga, C. Yang, et al., "Advancing Nuclear Physics Through TOPS Solvers and Tools," *SciDAC 2011 Conference*, Denver, CO, July 10, 2011, arXiv:1110.1708.

2010

A. Canning, J. Shalf, H Wassermann and M. Gajbe, "A Comparison of Different Communication Structures for Scalable Three Dimensional FFTs in First Principles Codes," *Parallel Computing: From Multicores and GPUs to Petascale*, IOS Press, B. Chapman (Eds.) 2010.

Rostyslav Boutchko, Andrew Canning, Anurag Chaudhry, Stephen Derenzo, "A theoretical study of the relative importance of chemical and geometric effects for Ce-based scintillation in La and Y aluminum perovskites," *Nuclear Science Symposium Proceedings* (NSS/MIC), 2010 IEEE p275 2010

G. Gundiah, A. Canning et. al., "Scintillation Properties of Eu2+ Activated Barium Fluoroiodide," *IEEE transactions on nuclear science*. 57. No. 3. 2010.

L. Lin, J.A. Morrone, R. Car and M. Parrinello, "Displaced path integral formulation for the momentum distribution of quantum particles," *Phys. Rev. Lett*., 2010, 105:110602.
J. C. Fogarty, H. M. Aktulga, A. van Duin, A. Y. Grama, S. A. Pandit, "A Reactive Simulation of the Silica-Water Interface," *J. Chem. Phys*. 132, 174704, May 4, 2010,

H. M. Aktulga, A. Y. Grama, S. Plimpton, A. Thompson, "A Fast ILU Preconditioning-based Solver for the Charge Equilibration Problem," *CSRI Summer Proceedings 2009*, January 11, 2010.

D. Hasenkamp, A. Sim, M. Wehner and K. Wu, "Finding Tropical Cyclones on a Cloud Computing Cluster: Using Parallel Virtualization for Large-Scale Climate Simulation Analysis," *Proceedings of the 2nd IEEE International Conference on Cloud Computing Technology and Science*, Indianapolis November 30-December 3, 2010.

Michael F. Wehner, G. Bala, Phillip Duffy, Arthur A. Mirin, and Raquel Romano "Towards Direct Simulation of Future Tropical Cyclone Statistics in a High-Resolution Global Atmospheric Model," *Advances in Meteorology*, Vol. 2010, Article ID 915303, 13 pages, 2010.

Wehner, M.F., "Sources of uncertainty in the extreme value statistics of climate data," *Extremes* 13, 205-217, 2010.

Wehner, M.F. ,R. Smith, P. Duffy, G. Bala, "The effect of horizontal resolution on simulation of very extreme US precipitation events in a global atmosphere model," *Climate Dynamics* 32, 241-247, 2010.

2009

A. Canning, J. Dongarra, J. Langou, O. Marques, S. Tomov, C. Voemel, and L.-W. Wang, "Interior State Computation of Nanostructures," *Proceedings of Para08*, Norway, 2009.

A. Canning, A. Chaudhry R. Boutchko and S. E. Derenzo, "First-Principles Studies and Predictions for New Ce doped Gamma Detector Materials," *IEEE transaction on nuclear science*, Vol. 53, 3 p944, 2009

M. Gajbe, A. Canning, J. Shalf, L-W Wang, "Tuning of 3 Dimensional Fast Fourier Transforms on the Cray XT4," *Proceedings of Cray User Group Conference*, Atlanta, March 2009.

A. Chaudhry, A. Canning, R. Boutchko, S. E. Derenzo and N. Gronbech-Jensen, "First principles calculations for Ce-doped rare-earth oxyhalides REOX (RE = Y , La; X = F, Cl, Br, I) for Gamma Detector Materials," to appear, *IEEE transaction on nuclear science*, Vol. 53, 3, p949, 2009.

R. Boutchko, A. Chaudhry, A. Canning et al., "Cerium Activated Scintillation in Yttrium Halides: First Principles Theory and Prediction," to appear, *IEEE transaction on nuclear science*, Vol. 53, 3, p977, 2009.

E. D. Bourret-Courchesne, G. Bizarri, R. Borade, Z. Yan, S. M. Hanrahan, G. Gundiah, A. Chaudhry, A. Canning, and S. E. Derenzo, "Eu2+-doped Ba2CsI5, a new high-performance scintillator," *Nucl. Instrum. Methods Phys. Res. A* 612, 138, 2009.

L. Lin, J. Lu, L. Ying, W. E, "Pole-based approximation of the Fermi-Dirac function," *Chin. Ann. Math.,* 2009, 30B:729.
L. Lin, J. Lu, L. Ying, R. Car, W. E, "Fast algorithm for extracting the diagonal of the inverse matrix with application to the electronic structure analysis of metallic systems," *Commun. Math. Sci.,* 2009, 7:755.

J.A. Morrone, L. Lin, R. Car, "Tunneling and delocalization effects in hydrogen bonded systems: A study in position and momentum space," *J. Chem. Phys*., 2009, 130:204511.

L. Lin, J. Lu, R. Car, W. E, "Multipole representation of the Fermi operator with application to the electronic structure analysis of metallic systems," *Phys. Rev. B*, 2009, 79:115133.

Y. Park, H. M. Aktulga, A. Y. Grama, A. Strachan, "Strain relaxation in Si/Ge/Si nanoscale bars from molecular dynamics simulations," *J. Appl. Phys.* 106, 034304 2009, August 6, 2009

Easterling, D. R., and M. F. Wehner, "Is the climate warming or cooling?" *Geophys. Res. Lett.*, 36, L08706, 2009.
B.D. Santer, K.E. Taylor, P.J. Gleckler, C. Bonfils, T.P. Barnett, D.W. Pierce, T.M.L. Wigley, C. Mears, F.J. Wentz, W. Brueggemann, N.P. Gillett, S.A. Klein, S. Solomon, P.A. Stott, and M.F. Wehner, "Incorporating Model Quality Information in Climate Change Detection and Attribution Studies," *Proceeding of the National Academy of Sciences*, 2009.

Michael Wehner, Lenny Oliker, John Shalf, "Low Power Supercomputers," *IEEE Spectrum*, October 2009.

David Donofrio, Leonid Oliker, John Shalf, Michael Wehner, Chris Rowen, Jens Krueger, Shoaib Kamil and Marghoob Mohiyuddin, "Energy-Efficient Computing for Extreme Scale Science," *IEEE Computer*, 42(11). October 2009, pp 62-71.

BIOLOGICAL DATA MANAGEMENT AND TECHNOLOGY CENTER

2012

V. M. Markowitz, I.-M. A. Chen, K. Palaniappan, K. Chu, E. Szeto, et al. "IMG: the integrated microbial genomes database and comparative analysis system," *Nucleic Acids Res.*, 40: D115-D122, 2012.

Markowitz VM, Chen IMA, Chu K, Szeto E, Palaniappan K, et al., "IMG/M: the integrated metagenome data management and comparative analysis system," *Nucleic Acid Res.*, 40: D123-129, 2012.

Markowitz VM, Chen IMA, Chu K, Szeto E, Palaniappan K, et al., "IMG/M-HMP: A metagenome comparative analysis system for the Human Microbiome Project," *PLoS ONE* 7(7): e40151, 2012.

The Human Microbiome Project Consortium, "Structure, function and diversity of the healthy human microbiome," *Nature* 486: 207-214, 2012.

The Human Microbiome Project Consortium, "A framework for human microbiome research," *Nature* 486: 215-221, 2012.

V. M. Markowitz, I.-M. A. Chen, K. Palaniappan, K. Chu, E. Szeto, et al. "IMG: the integrated microbial genomes database and comparative analysis system," *Nucleic Acids Res.*, 40: D115-D122, 2012.

Markowitz VM, Chen IMA, Chu K, Szeto E, Palaniappan K, et al., "IMG/M: the integrated metagenome data management and comparative analysis system," *Nucleic Acid Res.*, 40: D123-129, 2012.

Markowitz VM, Chen IMA, Chu K, Szeto E, Palaniappan K, et al., "IMG/M-HMP: A metagenome comparative analysis system for the Human Microbiome Project," *PLoS ONE* 7(7): e40151, 2012.

The Human Microbiome Project Consortium, "Structure, function and diversity of the healthy human microbiome," *Nature* 486: 207-214, 2012.

The Human Microbiome Project Consortium, "A framework for human microbiome research," *Nature* 486: 215-221, 2012.

Pagani, I, Liolios, K, Jansson, J, Chen, IMA, Smironova, T, Nosrat, B, Markowitz, VM, and Kyrpides, NC., "The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata," *Nucleic Acids Res.*, 40: D571-D579, 2012.

2010

Markowitz, V.M. et al. "The integrated microbial genomes system: an expanding comparative analysis resource," *Nucleic Acids Research 2010*, Vol. 38, Database issue D382-390K.

Liolios, K. et al., "The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata," *Nucleic Acids Research 2010*, Vol. 38, Database issue D346-354.

2009

Sharon, I., Pati, A., Markowitz, V.M., Pinter R.Y. "A Statistical Framework for the Functional Analysis of Metagenomes," *Lecture Notes in Computer Science* 5541, RECOMB 2009, pp. 570-585, 2009.

Markowitz, V.M., Mavromatis, K., Ivanova, N. et al., "IMG ER: A System for Microbial Genome Annotation Expert Review and Curation," *Bioinformatics,* 2009.

Mavromatis, K., Chu, K., Ivanova, N.N., Hooper, S., Markowitz, V.M., Kyrpides, N.C. Mavromatis, K. et al. "Gene Context Analysis in the Integrated Microbial Genomes (IMG) Data Management System," *PLoS ONE* 4(11): e7979, 2009.

COMPUTER AND DATA SCIENCES DEPARTMENT

FUTURE TECHNOLOGIES GROUP

2012

Samuel Williams, "Optimization of Geometric Multigrid for Emerging Multi- and Manycore Processors," *Supercomputing (SC),* November 2012.

Michael Garland, Manjunath Kudlur, Yili Zheng, "Designing a Unified Programming Model for Heterogeneous Machines," *Supercomputing (SC)*, November 2012.

Evangelos Georganas, Jorge González-Domínguez, Edgar Solomonik, Yili Zheng, et al., "Communication Avoiding and Overlapping for Numerical Linear Algebra," *Supercomputing (SC)*, November 2012.

S. Williams, D. Kalamkar, A. Singh, A. Deshpande, B. Van Straalen, et al., "Optimization of Geometric Multigrid for Emerging Multi- and Manycore Processors," *Supercomputing (SC)*, November 2012.

B. Wang, S. Ethier, W. Tang, K. Ibrahim, K. Madduri, S. Williams, "Advances in gyrokinetic particle in cell simulation for fusion plasmas to Extreme scale," *Supercomputing (SC)*, 2012.

A. Buluç, K. Madduri, "Graph partitioning for scalable distributed graph computations," *AMS Contemporary Mathematics, Proc. of the 10th DIMACS Implementation Challenge Workshop*, 2012.

K. Madduri, J. Su, S. Williams, L. Oliker, S. Ethier, K. Yelick, "Optimization of Parallel Particle-to-Grid Interpolation on Leading Multicore Platforms," *Transactions on Parallel and Distributed Systems (TPDS)*, October 1, 2012.

K. Kandalla, A. Buluç, H. Subramoni, K. Tomko, J. Vienne, L. Oliker, D. K. Panda, "Can network-offload based non-blocking neighborhood MPI collectives improve communication overheads of irregular graph algorithms?" *International Workshop on Parallel Algorithms and Parallel Software (IWPAPS 2012),* 2012.

B. Edwards, T. Moore, G. Stelle, S. Hofmeyr, S. Forrest, "Beyond the blacklist: Modeling malware spread and the effect of interventions," *New Security Paradigms Workshop*, September 19, 2012.

Jun Zhou, Didem Unat, Dong Ju Choi, Clark C. Guest, Yifeng Cui, "Hands-on Performance Tuning of 3D Finite Difference Earthquake Simulation on GPU Fermi Chipset," *Procedia CS*, 2012, Vol 9:976-985.

J. Krueger, P. Micikevicius, S. Williams, "Optimization of Forward Wave Modeling on Contemporary HPC Architectures," LBNL Technical Report, 2012.

A. Buluç, J. Gilbert, "Parallel sparse matrix-matrix multiplication and indexing: Implementation and experiments," *SIAM Journal on Scientific Computing (SISC)*, 2012.

Mads Kristensen, Yili Zheng, Brian Vinter, "PGAS for Distributed Numerical Python Targeting Multi-core Clusters," *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 2012.

Didem Unat, Jun Zhou, Yifeng Cui, Scott B. Baden, Xing Cai, "Accelerating a 3D Finite Difference Earthquake Simulation with a C-to-CUDA Translator," *Computing in Science and Engineering*, May 2012, Vol 14:48-59.

A. Lugowski, D. Alber, A. Buluç, J. Gilbert, S. Reinhardt, Y. Teng, A. Waranis, "A flexible open-source toolbox for scalable complex graph analysis," *SIAM Conference on Data Mining (SDM)*, 2012.

A. Lugowski, A. Buluç, J. Gilbert, S. Reinhardt, "Scalable complex graph analysis with the knowledge discovery toolbox," *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, March 2012.

B. Edwards, S. Hofmeyr, G. Stelle, S. Forrest, "Internet Topology over Time," *Arxiv.org*, February 17, 2012.

Mitesh R. Meswani, Laura Carrington, Didem Unat, Allan Snavely, Scott B. Baden, Stephen Poole, "Modeling and Predicting Performance of High Performance Computing Applications on Hardware Accelerators," *IPDPS Workshops*, IEEE Computer Society, 2012.

Han Suk Kim, Didem Unat, Scott B. Baden, Jürgen P. Schulze, "Interactive Data-centric Viewpoint Selection," *Visualization and Data Analysis, Proc. SPIE* 8294, January 2012.

A. Napov, "Conditioning Analysis of Incomplete Cholesky Factorizations with Orthogonal Dropping," 2012, LBNL 5353E.

A. Napov and Y. Notay, "An Algebraic Multigrid Method with Guaranteed Convergence Rate," *SIAM J. Sci. Comput*., Vol.43, pp. A1079-A1109, 2012.

2011

Mitesh R. Meswani, Laura Carrington, Didem Unat, Joshua Peraza, et al., "Modeling and Predicting Application Performance on Hardware Accelerators," *International Symposium on Workload Characterization (IISWC)*, IEEE, November 2011.

J. Krueger, D. Donofrio, J. Shalf, M. Mohiyuddin, S. Williams, L. Oliker, F. Pfreundt, "Hardware/Software Co-design for Energy-Efficient Seismic Modeling," *Supercomputing (SC)*, November 2011.

A. Buluç, K. Madduri, "Parallel breadth-first search on distributed memory systems," *Supercomputing (SC)*, November 2011.

Matt Bishop, Justin Cummins, Sean Peisert, Bhume Bhumitarana, Anhad Singh, et al., "Relationships and Data Sanitization: A Study in Scarlet," *Proceedings of the 2010 New Security Paradigms Workshop (NSPW)*, Concord, MA, ACM, September 2011, 151-164.

S. Williams, et al., "Stencil Computations on CPUs, Stanford Earth Sciences Algorithms and Architectures Initiative" *(SESAAI)*, 2011.

S. Williams, et al., "Performance Optimization of HPC Applications on Multi- and Manycore Processors," *Workshop on Hybrid Technologies for NASA Applications, 4th International Conference on Space Mission Challenges for Information Technology*, 2011.

J. Demmel, K. Yelick, M. Anderson, G. Ballard, E. Carson, I. Dumitriu, et al., "Rethinking Algorithms for Future Architectures: Communication-Avoiding Algorithms," *Hot Chips* 23, 2011.

S. Williams, et al, "Stencil Computations on CPUs," *Society of Exploration Geophysicists High-Performance Computing Workshop (SEG)*, July 2011.

S. Hofmeyr, T. Moore, S. Forrest, B. Edwards, G. Stelle, "Modeling Internet scale policies for cleaning up malware," *Workshop on the Economics of Information Security (WEIS 2011)*, June 14, 2011.

Chang-Seo Park, Koushik Sen, Paul Hargrove, Costin Iancu, "Efficient data race detection for distributed memory parallel programs," *Supercomputing (SC)*, 2011.

Didem Unat, Xing Cai, Scott B. Baden, "Mint: realizing CUDA performance in 3D stencil methods with annotated C," *ICS '11 Proceedings of the international Conference on Supercomputing*, ACM, June 2011, 214-224.

Sean Whalen, Sean Peisert, Matt Bishop, "Network-Theoretic Classification of Parallel Computation Patterns," *Proceedings of the First International Workshop on Characterizing Applications for Heterogeneous Exascale Systems (CACHES)*, Tucson, AZ, IEEE Computer Society, June 4, 2011.

Didem Unat, Han Suk Kim, Jurgen Schulze, Scott Baden, "Auto-optimization of a feature selection algorithm," *Emerging Applications and Many-Core Architecture*, June 2011.

A. Buluç, J. Gilbert, "The Combinatorial BLAS: Design, implementation, and applications," *International Journal of High-Performance Computing Applications (IJHPCA)*, 2011.

A. Buluç, S. Williams, L. Oliker, J. Demmel, "Reduced-Bandwidth Multithreaded Algorithms for Sparse Matrix-Vector Multiplication," *International Parallel Distributed Processing Symposium (IPDPS)*, May 2011.

P. Narayanan, A. Koniges, L. Oliker, R. Preissl, S. Williams, N. Wright, et al., "Performance Characterization for Fusion Co-design Applications," *Cray Users Group (CUG)*, May 2011.

A. Kaiser, S. Williams, K. Madduri, K. Ibrahim, D. Bailey, J. Demmel, E. Strohmaier, "TORCH Computational Reference Kernels: A Testbed for Computer Science Research," LBNL Technical Report, 2011.

David H. Bailey, Robert F. Lucas, Samuel W. Williams, ed., *Performance Tuning of Scientific Applications*, CRC Press, 2011.

M. Wehner, L. Oliker, J. Shalf, D. Donofrio, L. Drummond, et al., "Hardware/Software Co-design of Global Cloud System Resolving Models," *Journal of Advances in Modeling Earth Systems (JAMES)*, 2011.

Kamesh Madduri, Khaled Z. Ibrahim, Samuel Williams, Stephane Ethier, et al., "Gyrokinetic Toroidal Simulations on Leading Multi- and Manycore HPC Systems," *The International Conference for High Performance Computing, Networking, Storage, and Analysis*, January 2011.

Samuel Williams, Leonid Oliker, Jonathan Carter, John Shalf, "Extracting Ultra-scale Lattice Boltzmann Performance via Hierarchical and Distributed Auto-tuning," *SC 11*, New York, NY, USA, ACM, January 2011.

Kamesh Madduri, Eun-Jin Im, Khaled Z. Ibrahim, Samuel Williams, Stephane Ethier, Leonid Oliker, "Gyrokinetic Particle-in-cell Optimization on Emerging Multi- and Manycore Platforms," *Parallel Computing*, January 2011, 37:501–520.

R. Sudarsan, J. Borrill, C. Cantalupo, T. Kisner, K. Madduri, L. Oliker, H. Simon, Y. Zheng, "Cosmic Microwave Background Map-Making at the Petascale and Beyond," *International Conference of Supercomputing (ICS)*, 2011.

A. Napov and Y. Notay, "Smoothing Factor, Order of Prolongation and Actual Multigrid Convergence," *Numerische Mathematik*, Vol. 118, pp. 457-483, 2011.

A. Napov and Y. Notay, "Algebraic Analysis of Aggregation-Based Multigrid," *Numer. Lin. Alg. Appl.,* Vol.18, pp. 539-564, 2011.

M. Christen, N. Keen, T. Ligocki, L. Oliker, J. Shalf, B. van Straalen, S. Williams, "Automatic Thread-Level Parallelization in the Chombo AMR Library," LBNL Technical Report, 2011.

Rajesh Nishtala, Yili Zheng, Paul Hargrove, Katherine A. Yelick, "Tuning collective communication for Partitioned Global Address Space programming models," *Parallel Computing*, 2011, 37(9):576-591.

Samuel W. Williams, David H. Bailey, "Parallel Computer Architecture," *Performance Tuning of Scientific Applications*, edited by David H. Bailey, Robert F. Lucas, Samuel W. Williams, (CRC Press: 2011) pp. 11-33.

2010

S. Williams, N. Bell, J. W. Choi, M. Garland, L. Oliker, R. Vuduc, "Sparse Matrix-Vector Multiplication on Multicore and Accelerators," chapter in *Scientific Computing with Multicore and Accelerators*, edited by Jack Dongarra, David A. Bader, Jakub Kurzak, 2010.

K. Datta, S. Williams, V. Volkov, J. Carter, L. Oliker, J. Shalf, K. Yelick, "Auto-tuning Stencil Computations on Multicore and Accelerators," chapter in *Scientific Computing with Multicore and Accelerators*, edited by Jack Dongarra, David A. Bader, (2010).

S. Williams, "The Roofline Model," chapter in *Performance Tuning of Scientific Applications*, (2010).

L. Oliker, J. Carter, V. Beckner, J. Bell, H. Wasserman, M. Adams, S. Ethier, E. Schnetter, *Large-Scale Numerical Simulations on High-End Computational Platforms*, Chapman & Hall/CRC Computational Science, edited by D. H. Bailey, R. F. Lucas, S. W. Williams, (CRC Press: 2010).

K. Datta, S. Williams, V. Volkov, J. Carter, L. Oliker, J. Shalf, K. Yelick, *Auto-Tuning Memory-Intensive Kernels for Multicore*, Chapman & Hall/CRC Computational Science, edited by D. H. Bailey, R. F. Lucas, S. W. Williams, CRC Press: 2010. Pages: 219

Khaled Z. Ibrahim, Erich Strohmaier, "Characterizing the Relation Between Apex-Map Synthetic Probes and Reuse Distance Distributions," *The 39th International Conference on Parallel Processing (ICPP),* 2010, 353 -362.

G. Hendry, J, Chan, S, Kamil, L. Oliker , J. Shalf, L. Carloni , K. Bergman, "Silicon Nanophotonic Network-On-Chip using TDM Arbitration," *Hot Interconnects*, August 2010.

A. Kaiser, S. Williams, K. Madduri, K. Ibrahim, D. Bailey, J. Demmel, E. Strohmaier, "A Principled Kernel Testbed for Hardware/ Software Co-Design Research," Proceedings of the 2nd USENIX Workshop on Hot Topics in Parallelism," *Proceedings of the 2nd USENIX Workshop on Hot Topics in Parallelism (HotPar)*, 2010.

J. A. Colmenares, S. Bird, H. Cook, P. Pearce, D. Zhu, J. Shalf, et al., "Resource Management in the Tessellation Manycore OS," *2nd Usenix Workshop on Hot Topics in Parallelism (HotPar)*, June 15, 2010.

E. Strohmaier, S. Williams, A. Kaiser, K. Madduri, K. Ibrahim, D. Bailey, J. Demmel,, "A Kernel Testbed for Parallel Architecture, Language, and Performance Research," *International Conference of Numerical Analysis and Applied Mathematics (ICNAAM)*, June 1, 2010.

Sean Peisert, "Fingerprinting Communication and Computation on HPC Machines," Lawrence Berkeley National Laboratory, Technical Report, June 2010, LBNL LBNL-3483E.

S. Ethier, M. Adams, J. Carter, L. Oliker, "Petascale Parallelization of the Gyrokinetic Toroidal Code," *VECPAR: High Performance Computing for Computational Science*, June 2010.

A. Kaiser, S. Williams, K. Madduri, K. Ibrahim, D. Bailey, J. Demmel, E. Strohmaier, "A Principled Kernel Testbed for Hardware/ Software Co-Design Research," *Proceedings of the 2nd USENIX Workshop on Hot Topics in Parallelism (HotPar)*, 2010.

Shoaib Kamil, Cy Chan, Leonid Oliker, John Shalf, Samuel Williams, "An Auto-tuning Framework for Parallel Multicore Stencil Computations," *International Parallel and Distributed Processing Symposium (IPDPS)*, April 20, 2010.

Yili Zheng, Filip Blagojevic, Dan Bonachea, Paul H. Hargrove, et al., "Getting Multicore Performance with UPC," *SIAM Conference on Parallel Processing for Scientific Computing*, February 2010.

S. Williams, et al., "Memory-Efficient Optimization of Gyrokinetic Particle-to-Grid Interpolation for Multicore Processors," *SIAM Conference on Parallel Processing for Scientific Computing (SIAM PP10)*, 2010.

A. Napov, "Algebraic Analysis of V-Cycle Multigrid and Aggregation-Based Two-Grid Methods," 2010.

A. Chandramowlishwaran, S. Williams, L. Oliker, I. Lashuk, G. Biros, R. Vuduc, "Optimizing and Tuning the Fast Multipole Method for State-of-the-Art Multicore Architectures," *International Parallel & Distributed Processing Symposium (IPDPS)*, 2010.

A. Uselton, M. Howison, N. Wright, D. Skinner, N. Keen, J. Shalf, K. Karavanic, L. Oliker, "Parallel I/O Performance: From Events to Ensembles," *International Parallel & Distributed Processing Symposium (IPDPS)*, 2010.

Khaled Z. Ibrahim, "Bridging the Gap Between Complex Software Paradigms and Power-efficient Parallel Architectures," *International Conference on Green Computing*, 2010, 417-424.

A. Napov and Y. Notay, "When Does Two-Grid Optimality Carry Over to the V-Cycle?" *Numer. Lin. Alg. Appl.*, Vol.17, pp. 273-290, 2010.

A. Napov and Y. Notay, "Comparison of Bounds for V-Cycle Multigrid," *Appl. Numer. Math.*, vol.60, pp.176-192, 2010.

2009

J. Demmel, J. Dongarra, A. Fox, S. Williams, V. Volkov, K. Yelick, "Accelerating Time-to-Solution for Computational Science and Engineering," *SciDAC Review*, Number 15, December 2009.

D. Donofrio, L. Oliker. J. Shalf, M. Wehner, C. Rowen, J. Krueger, S. Kamil, M. Mohiyuddin, "Energy-Efficient Computing for Extreme-Scale Science," *IEEE Computer*, November 2009, 42, No. :62-71.

Kamesh Madduri, Samuel Williams, Stephane Ethier, Leonid Oliker, et al., "Memory-Efficient Optimization of Gyrokinetic Particle-to-Grid Interpolation for Multicore Processors," *Supercomputing (SC)*, November 2009.

J. Shalf, M. Wehner, L. Oliker, "The Challenge of Energy-Efficient HPC," *SciDAC Review*, Fall, 2009.

M. Mohiyuddin, M. Murphy, L. Oliker, J. Shalf, J. Wawrzynek, S. Williams, "A Design Methodology for Domain-Optimized Power-Efficient Supercomputing," *Supercomputing (SC)*, 2009.

M. Wehner, L. Oliker., and J. Shalf, "Low Power Supercomputers," *IEEE Spectrum*, October 2009.

S. Zhou, D. Duffy, T. Clune, M. Suarez, S. Williams, M. Halem, "The Impact of IBM Cell Technology on the Programming Paradigm in the Context of Computer Systems for Climate and Weather Models," *Concurrency and Computation: Practice and Experience (CCPE)*, August 2009.

R. Liu, K. Klues, S. Bird, S. Hofmeyr, K. Asanovic, J. D. Kubiatowicz, "Tessellation: Space-Time Partitioning in a Manycore Client OS," *First USENIX Workshop on Hot Topics in Parallelism*, June 15, 2009.

Shoaib Kamil, Cy Chan, Samuel Williams, Leonid Oliker, John Shalf, et al., "A Generalized Framework for Auto-tuning Stencil Computations," *Cray User Group Conference (CUG)* (Best Paper Award), Atlanta, GA, May 4, 2009, LBNL 2078E.

Paul Hargrove, Jason Duell, Eric Roman, "Berkeley Lab Checkpoint/Restart (BLCR): Status and Future Plans," *Dagstuhl Seminar: Fault Tolerance in High-Performance Computing and Grids*, May 2009.

S. Williams, J. Carter, L. Oliker, J. Shalf, K. Yelick, "Resource-Efficient, Hierarchical Auto-Tuning of a Hybrid Lattice Boltzmann Computation on the Cray XT4," *Proceedings of the Cray User Group (CUG)*, Atlanta, GA, 2009.

S. Williams, A. Waterman, D. Patterson, "Roofline: An Insightful Visual Performance Model for Floating-Point Programs and Multicore Architectures," *Communications of the ACM (CACM)*, April 2009.

Paul Hargrove, Jason Duell, Eric Roman, "System-level Checkpoint/Restart with BLCR," *TeraGrid 2009 Fault Tolerance Workshop*, March 19, 2009.

S. Williams, L. Oliker, R. Vuduc, J. Shalf, K. Yelick, J. Demmel, "Optimization of Sparse Matrix-Vector Multiplication on Emerging Multicore Platforms," *Parallel Computing Journal*, 2009.

J. Gebis, L. Oliker, J. Shalf, S. Williams, K. Yelick, "Improving Memory Subsystem Performance using ViVA: Virtual Vector Architecture," *Architecture of Computing Systems (ARCS), Delft, Netherlands*, 2009.

K. Datta, S. Kamil, S. Williams, L. Oliker, J. Shalf, K. Yelick, "Optimization and Performance Modeling of Stencil Computations on Modern Microprocessors," *SIAM Review (SIREV)*, cover article, 2009.

John Shalf, David Donofrio, "Green Flash: Extreme Scale Computing on a Petascale Budget," 2009.

Didem Unat, Theodore Hromadka III, Scott B. Baden, "An Adaptive Sub-sampling Method for In-memory Compression of Scientific Data," *DCC*, IEEE Computer Society, 2009.

K. Datta, S. Williams, V. Volkov, J. Carter, L. Oliker, J. Shalf, K. Yelick, "Auto-Tuning the 27-point Stencil for Multicore," *Proceedings of Fourth International Workshop on Automatic Performance Tuning (iWAPT2009)*, January 2009.

Khaled Z. Ibrahim, J. Jaeger, Z. Liu, L.N. Pouchet, P. Lesnicki, et al., "Simulation of the Lattice QCD and Technological Trends in Computation," *The 14th International Workshop on Compilers for Parallel Computers (CPC 09)*, 2009.

G. Hendry, S.A. Kamil, A. Biberman, J. Chan, B.G. Lee, M Mohiyuddin, et al., "Analysis of Photonic Networks for Chip Multiprocessor Using Scientific Applications," *NOCS2009,* 2009.

J. Borrill, L. Oliker, J. Shalf, H. Shan, A. Uselton, "HPC Global File System Performance Analysis Using A Scientific-Application Derived Benchmark," *Parallel Computing Journal*, 2009.

S. Kamil, L. Oliker, A. Pinar, J. Shalf, "Communication Requirements and Interconnect Optimization for High-End Scientific Applications," *IEEE Transactions on Parallel and Distributed Systems*, IEEE, 2009.

S. Kamil, L. Oliker, A. Pinar, J. Shalf, "Communication Requirements and Interconnect Optimization for High-End Scientific Applications," *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, 2009.

John Shalf and Jason Hick (Arie Shoshani and Doron Rotem), "Storage Technology Fundamentals," *Scientific Data Management: Challenges, Technology, and Deployment, Volume*. Chapman & Hall/CRC, 2009.

G. Hendry, S. Kamil, A. Biberman, J. Chan, B. Lee, M. Mohiyuddin, A. Jain, et al., "Analysis of Photonic Networks for a Chip Multiprocessor Using Scientific Applications," *International Symposium on Networks-on-Chip (NOCS)*, 2009.

S. Amarasinghe, D. Campbell, W. Carlson, A. Chien, W. Dally, E. Elnohazy, et al., "ExaScale Software Study: Software Challenges in Extreme Scale Systems," 2009.

John Shalf, Thomas Sterling, "Operating Systems For Exascale Computing," 2009.

R. Biswas, J. Vetter, L. Oliker, "Revolutionary Technologies for Acceleration of Emerging Petascale Applications," Guest Editors, *Parallel Computing Journal*, 2009.

John Shalf, Erik Schnetter, Gabrielle Allen, Edward Seidel, "Cactus and the Role of Frameworks in Complex Multiphysics HPC Applications," 2009.

Samuel Williams, Jonathan Carter, Leonid Oliker, John Shalf, Katherine Yelick, "Optimization of a Lattice Boltzmann Computation on State-of-the-art Multicore Platforms," *J. Parallel Distrib. Comput.*, 2009, 69:762—777.

SCIENTIFIC DATA MANAGEMENT GROUP

2012

Surendra Byna, Jerry Chou, Oliver Rubel, Prabhat, Homa Karimabadi, William S. Daughton, Vadim Roytershteyn, E. Wes Bethel, Mark Howison, Ke-Jou Hsu, Kuan-Wu Lin, Arie Shoshani, Andrew Uselton, and Kesheng Wu, "Parallel I/O, Analysis, and Visualization of a Trillion Particle Simulation," *SuperComputing conference*, SC'12, November 2012.

Joerg Meyer, Harinarayan Krishnan, Jennifer Horsman, Alexandru Romosan, E. Wes Bethel, "Visual Data Analysis as an Integral Part of Environmental Management," to be published in *Transactions on Visualization and Computer Graphics,* IEEE TVCG, October 2012.

Benson Ma, Arie Shoshani, Alex Sim, Kesheng Wu, Yong-Ik Byun, Jaegyoon Hahm and Min-Su Shin, "Efficient Attribute-based Data Access in Astronomy Analysis," *International workshop on Network-aware Data Management*, November 11, 2012.

Prabhat, Oliver Rubel, Surendra Byna, Kesheng Wu, Fuyu Li, Michael Wehner and Wes Bethel, "TECA: A Parallel Toolkit for Extreme Climate Analysis," *International Conference on Computational Science*, ICCS June, 2012.

Elaheh Pourabbas, Arie Shoshani, and Kesheng Wu, "Minimizing Index Size by Reordering Rows and Columns," *International Conference on Scientific and Statistical Database Management (SSDBM)*, June 2012.

Y. Yin, S. Byna, H. Song, X.-H. Sun, and R. Thakur, "Boosting Application-Specific Parallel I/O Optimization Using IOSIG," accepted for publication at *IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)* 2012.

2011

E. Wes Bethel, David Leinweber, Oliver Ruebel, Kesheng Wu, "Federal Market Information Technology in the Post Flash Crash Era: Roles for Supercomputing," In *Proceedings of the fourth workshop on High performance computational finance (WHPCF '11)*. ACM, New York, NY, USA.

Surendra Byna, Prabhat, Michael F. Wehner, and Kesheng John Wu, "Detecting atmospheric rivers in large climate datasets," In *Proceedings of the 2nd international workshop on Petascal data analytics: challenges and opportunities (PDAC '11)*. ACM, New York, NY, USA, 2011.

Cummings, Klasky, Podhorszki, Barreto, Lofstead, Schwan, Docan, Parashar, Sim, Shoshani, "EFFIS: and End-to-end Framework for Fusion Integrated Simulation," *PDP* 2011

Jerry Chou, Jinoh Kim and Doron Rotem, "Energy-Aware Scheduling in Disk Storage Systems," in *ICDCS* 2011

Jerry Chou, Jinoh Kim and Doron Rotem, "Energy Saving Techniques for Disk Storage Systems," Book chapter in *Handbook of Energy-Aware and Green Computing* edited by Ishfaq Ahmad and Sanjay Ranka ,2011.

J. Chou, Kesheng Wu; Prabhat, "FastQuery: A Parallel Indexing System for Scientific Data," *Cluster Computing (CLUSTER), 2011 IEEE International Conference*, vol., no., pp.455-464, 26-30 Sept. 2011.

J. Chou, K. Wu, O. Rubel, M. Howison, J. Qiang, Prabhat, B. Austin, E. W. Bethel, R. D. Ryne, and A. Shoshani, "Parallel Index and Query for Large Scale Data Analysis," *SuperComputing Conference*, 2011.

Junmin Gu, Dimitrios Katramatos, Xin Liu, Vijaya Natarajan, Arie Shoshani, Alex Sim, Dantong Yu, Scott Bradley, Shawn McKee, "StorNet: Co-Scheduling of End-to-End Bandwidth Reservation on Storage and Network Systems for High Performance Data Transfers," *Proceedings of IEEE INFOCOM HSN 2011*, Shanghai China, 2011.

Jinoh Kim, Hasan Abbasi, Luis Chacon, Ciprian Docan, Scott Klasky, Qing Liu, Norbert Podhorszki, Arie Shoshani, Kesheng Wu, "Parallel In Situ Indexing for Data-intensive Computing," *IEEE Symposium on Large Data Analysis and Visualization*, October 23 - 24, 2011.

Jinoh Kim, Jerry Chou and Doron Rotem, "Energy Proportionality and Performance in Data Parallel Computing Clusters," *SSDBM* 2011.

Jinoh Kim, Doron Rotem, "Energy proportionality for disk storage using replication," *EDBT* 2011: 81-92.

Jinoh Kim, Doron Rotem "Using Replication for Energy Conservation in RAID Systems," *PDPTA* 2010: 703-709

K. Madduri and K. Wu, "Massive-scale RDF Processing Using Compressed Bitmap Indexes," SSDBM 2011. *Lecture Notes in Computer Science*, 2011, Vol. 6809/2011, 470-479.

A. Shoshani, et al., "The Scientific Data Management Center: Available Technologies and Highlights," *SciDAC Conference*, 2011.

D. Williams, A. Shoshani, et al., "Earth System Grid Center for Enabling Technologies (ESG-CET): A Data Infrastructure for Data-Intensive Climate Research," *SciDAC Conference* 2011.

Kesheng Wu, Surendra Byna, Doron Rotem, and Arie Shoshani, "Scientific data services: a high-performance I/O system with array semantics," In *Proceedings of the first annual workshop on High performance computing meets databases (HPCDB '11)*. ACM, New York, NY, USA, 9-12, 2011.

K. Wu, R. R Sinha, C. Jones, S. Ethier, S. Klasky, K.-L. Ma, A. Shoshani and M. Winslett, "Finding regions of interest on toroidal meshes," *Computational Science & Discovery*, Vol. 4, page 015003. 2011.

2010

M. Balman, E. Chaniotakis, A. Shoshani, A. Sim, "An Efficient Reservation Algorithm for Advanced Network Provisioning," *ACM/IEEE Supercomputing Conference* 2010 (SC10).

Julian Cummings, Jay Lofstead, Karsten Schwan, Alexander Sim, Arie Shoshani, Ciprian Docan, Manish Parashar, Scott Klasky, Norbert Podhorszki, Roselyne Barreto, "EFFIS: An End-to-end Framework for Fusion Integrated Simulation," *Euromicro Conference on Parallel, Distributed, and Network-Based Processing*, pp. 428-434, 2010, 18th Euromicro Conference on Parallel, Distributed and Network-based Processing, 2010.

G. Garzoglio, J. Bester, K. Chadwick, D. Dykstra, D. Groep, J. Gu, T. Hesselroth, O. Koeroo, T. Levshina, S. Martin, M. Salle, N. Sharma, F. Siebenlist, A. Sim, A. Verstegen, "Adoption of a SAML-XACML Profile for Authorization Interoperability across Grid Middleware in OSG and EGEE," *Proceedings of the 18th International Conference on Computing in High Energy and Nuclear Physics* (CHEP 2010), 2010.

Junmin Gu, Dimitrios Katramatos, Xin Liu, Vijaya Natarajan, Arie Shoshani, Alex Sim, Dantong Yu, Scott Bradley, Shawn McKee, "StorNet: Co-Scheduling of End-to-End Bandwidth Reservation

on Storage and Network Systems for High Performance Data Transfers," *Proceedings of IEEE INFOCOM HSN 2011*, Shanghai China, 2011.

Junmin Gu, Dimitrios Katramatos, Xin Liu, Vijaya Natarajan, Arie Shoshani, Alex Sim, Dantong Yu, Scott Bradley, Shawn McKee, "StorNet: Integrated Dynamic Storage and Network Resource Provisioning and Management for Automated Data Transfers," *Proceedings of the 18th International Conference on Computing in High Energy and Nuclear Physics (CHEP 2010)*, Taipei Taiwan, 2010

D. Hasenkamp, A. Sim, M. Wehner, K. Wu, "Finding Tropical Cyclones on a Cloud Computing Cluster: Using Parallel Virtualization for Large-Scale Climate Simulation Analysis," *CloudCom2010*. 2010.

Jinoh Kim, Doron Rotem "Using Replication for Energy Conservation in RAID Systems," *PDPTA* 2010: 703-709

Ekow J. Otoo, Doron Rotem, Shih-Chiang Tsao: "Dynamic Data Reorganization for Energy Savings in Disk Storage Systems." *SSDBM* 2010: 322-341

E. Pourabbas and A. Shoshani, "Improving Estimation Accuracy of Aggregate Queries on Data Cubes," *Data & Knowledge Engineering* 69 (2010) 50–72.

A. Sim, M. Balman, D. Williams, A. Shoshani, and V. Natarajan, "Adaptive Transfer Adjustment in Efficient Bulk Data Transfer Management for Climate Datasets," *Parallel and Distributed Computing and Systems (PDCS2010)*, Nov. 2010.

A. Sim, D. Gunter, V. Natarajan, A. Shoshani, D. Williams, J. Long, J. Hick, J. Lee, E. Dart, "Efficient Bulk Data Replication for the Earth System Grid," LBNL-3821E, 2010.

A. Sim, D. Gunter, V. Natarajan, A. Shoshani, D. Williams, J. Long, J. Hick, J. Lee, E. Dart, "Efficient Bulk Data Replication for the Earth System Grid," *Proceedings of International Symposium on Grid Computing (ISGC)*, 2010.

Arie Shoshani, Scott Klasky, Rob Ross, "Scientific Data Management: Challenges and Approaches in the Extreme Scale Era," *SciDAC Conference*, 2010.

K. Wu, K. Madduri, S. Canon, "Multi-Level Bitmap Indexes for Flash Memory Storage," *IDEAS'10*. 2010.

K. Wu, A. Shoshani, and K. Stockinger, "Analyses of Multi-Level and Multi-Component Compressed Bitmap Indexes," *ACM TODS,* Vol. 35, Article 2, 2010.

I. Yamazaki, Z. Bai, H. Simon, L.-W. Wang, and K. Wu, "Adaptive Projection Subspace Dimension for the Thick-Restart Lanczos Method," *ACM Trans. Math. Softw*. 37, 3, Article 27 (September 2010).

Oliver Rübel, Sean Ahern, E. Wes Bethela, Mark D. Biggin, Hank Childs, Estelle Cormier-Michel, Angela DePace, Michael B. Eisen, Charless C. Fowlkes, Cameron G.R. Geddes, Hans Hagen, Bernd Hamann, Min-Yu Huang, Soile V.E. Keränen, David W. Knowles, Cris L. Luengo Hendriks, Jitendra Malik, Jeremy Meredith, Peter Messmer, Prabhat,Daniela Ushizima,Gunther H. Weber, Kesheng Wu, "Coupling visualization and data analysis for knowledge discovery from multi-dimensional scientific data," *International Conference on Computational Science*, ICCS 2010. Pages 1757–1764

2009

Scientific Data Management: Challenges, Technology, and Deployment, Editors Arie Shoshani and Doron Rotem, Chapman & Hall/CRC Computational Science, December 2009.

Ekow J. Otoo and Kesheng Wu, "Accelerating queries on very large datasets," Book Chapter in forthcoming book entitled *Scientific Data Management: Challenges, Technology, and Deployment*, Editors Arie Shoshani and Doron Rotem, Chapman & Hall/CRC Computational Science, 2009.

Per Svensson, Peter Boncz, Milena Ivanova, Martin Kersten, Niels Nes, Doron Rotem, "Emerging vertical database systems in support of scientific data," Book Chapter in forthcoming book entitled *Scientific Data Management: Challenges, Technology, and Deployment*, Editors Arie Shoshani and Doron Rotem, Chapman & Hall/CRC Computational Science, 2009.

Lifeng He, Yuyan Chao, Kenji Suzuki, Kesheng Wu, "Fast connected-component labeling," *Pattern Recognition* 42(9): 1977-1987 (2009).

K. Madduri and D.A. Bader, "Compact Graph Representations and Parallel Connectivity Algorithms for Massive Dynamic Network Analysis," *The 23rd IEEE International Parallel and Distributed Processing Symposium (IPDPS 2009)*, Rome, Italy, May 25-29, 2009.

M. Nagappan, Kesheng Wu; M. A.Vouk, "Efficiently Extracting Operational Profiles from Execution Logs Using Suffix Arrays," *Software Reliability Engineering,* 2009. ISSRE '09. pp.41-50, 16-19 Nov. 2009.

Ekow Otoo, Doron Rotem and Shih-Chiang Tsao, "Analysis of Trade-Off Between Power Saving and Response Time in Disk Storage Systems," *The Fifth Workshop on High-Performance, Power-Aware Computing*, May 2, 2009, Rome, Italy.

Ekow J. Otoo, Doron Rotem, and Shih-Chiang Tsao, "Energy smart management of scientific data," *21st Int'l. Conf. on Sc. and Stat. Database Management (SSDBM'2009)*, New Orleans, Louisiana, USA, Jun. 2009.LBNL Tech. Report No LBNL-2185E.

Ekow J. Otoo, Doron Rotem, and Shih-Chiang Tsao, "Workload-adaptive management of energy-smart disk storage systems," *Cluster Computing and Workshops*. IEEE International Conference on CLUSTER '09. 2009

Elaheh Pourabbas and Arie Shoshani, "Improving estimation accuracy of aggregate queries on data cubes," *Data & Knowledge Engineering*, 2009.

Torben Bach Pedersen, Junmin Gu, Arie Shoshani, Christian S. Jensen, "Object-extended OLAP querying," *Data & Knowledge Engineering* 68 (2009) 453–480.

Morris Riedel, Erwin Laure, Th. Soddemann, Alex Sim, Vijaya Natarajan, Arie Shoshani, Junmin Gu, et al., "Interoperation of world-wide production e-Science infrastructures," *Concurrency and Computation: Practice and Experience* 21(8): 961-990 (2009)

Oliver Rübel, Prabhat, Kesheng Wu, Hank Childs, Jeremy Meredith, Cameron G. R. Geddes, Estelle Cormier-Michel, Sean Ahern, Gunther H. Weber, Peter Messmer, Hans Hagen, Bernd Hamann, E. Wes Bethel, "High performance multivariate visual data exploration for extremely large data." *SC 2008*: 51. LBNL-716E.

Oliver Rubel, Prabhat, Kesheng Wu, Hank Childs, Jeremy Meredith, Cameron G.R. Geddes, Estelle Cormier-Michel, Sean Ahern, Gunther H. Weber, Peter Messmer, Hans Hagen, Bernd Hamann and E. Wes Bethel, "Application of High-performance Visual Analysis Methods to Laser Wakefield Particle Acceleration Data," *IEEE Visualization* 2008. LBNL-952E.

Kurt Stockinger, John Cieslewicz, Kesheng Wu, Doron Rotem and Arie Shoshani, "Using Bitmap Index for Joint Queries on Structured and Text Data," *New Trends in Data Warehousing and Data Analysis Annals of Information Systems*, 2009, Vol. 3, 1-23.

Arie Shoshani, Flavia Donno, Junmin Gu, Jason Hick, Maarten Litmaath, Alex Sim, "Dynamic Storage Management," Book Chapter in forthcoming book entitled *Scientific Data Management: Challenges, Technology, and Deployment*, Editors Arie Shoshani and Doron Rotem, Chapman & Hall/CRC Computational Science, 2009.

Arie Shoshani and Doron Rotem, editors, *Scientific Data Management: Challenges, Technology, and Deployment*, Chapman & Hall/CRC Computational Science., Dec. 2009.

Rishi Rakesh Sinha, Marianne Winslett, Kesheng Wu. "Finding Regions of Interest in Large Scientific Datasets." *SSDBM* 2009: 130-147.

D. N. Williams, R. Ananthakrishnan, D. E. Bernholdt, A. Shoshani, A. Sim, et al., "The Earth System Grid: Enabling Access to Multimodel Climate Simulation Data," *American Meteorological Society*, Vol. 90, No. 2, 195-205, 2009.

Kesheng Wu, Sean Ahern, E. Wes Bethel, Junmin Gu, Ekow Otoo, Arie Shoshani, Alexander Sim, E. Otoo, et al., "FastBit: Interactively Searching Massive Data." *SciDAC 2009*. LBNL-2164E.

Kesheng Wu, Sean Ahern, E. Wes Bethel, Jacqueline Chen, Hank Childs, Estelle Cormier-Michel, Cameron Geddes, Junmin Gu, Hans Hagen, Bernd Hamann, Wendy Koegler, Jerome Lauret, Jeremy Meredith, Peter Messmer, Ekow Otoo, Victor Perevoztchikov, Arthur Poskanzer, Prabhat, Oliver Rubel, Arie Shoshani, Alexander Sim, Kurt Stockinger, Gunther Weber, and Wei-Ming Zhang. FastBit: Interactively Searching Massive Data. In *Proc. SciDAC,* 2009.

Kesheng Wu, Sean Ahern, E. Wes Bethel, Junmin Gu, Ekow Otoo, Arie Shoshani, Alexander Sim, E. Otoo, et al., "FastBit: Interactively Searching Massive Data," *SciDAC* 2009. LBNL-2164E.
D. N. Williams, R. Ananthakrishnan, D. E. Bernholdt, A. Shoshani, A. Sim, et al., "The Earth System Grid: Enabling Access to Multimodel Climate Simulation Data," *American Meteorological Society*, Vol. 90, No. 2, 195-205, 2009

Kesheng Wu, Kurt Stockinger, and Arie Shoshani, "Breaking the Curse of Cardinality on Bitmap Indexes," *SSDBM* 2008. Tech Report LBNL-173E. Submitted for publication.

Kesheng Wu, Kurt Stockinger and Arie Shoshani, "Analyses of Multi-Level and Multi-Component Compressed Bitmap Indexes," LBNL Tech Report LBNL-60891. 2009. in review.

Kesheng Wu, Ekow J. Otoo, and Kenji Suzuki. "Optimizing two-pass connected component labeling algorithms." *Pattern Analysis and Applications*. Vol. 12, No. 2 (2009), 117-135.

Ekow Otoo and Kesheng Wu, "Accelerating Queries on Very Large Datasets," In *Scientific Data Management: Challenges, Technology, and Deployment*, Chapman & Hall/CRC Computational Science. 2009.

VISUALIZATION GROUP

2012

Daniela M. Ushizima, Dmitriy Morozov, Gunther H. Weber, Andrea G.C. Bianchi, James A. Sethian, and E. Wes Bethel, "Augmented topological descriptors of pore networks for material science," *IEEE Transactions on Visualization and Computer Graphics*, Proc. IEEE Vis 2012, 18(12):2041-2050, 2012.

Gunther H. Weber, Hank Childs, and Jeremy S. Meredith, "Efficient Parallel Extraction of Crack- free Isosurfaces from Adaptive Mesh Renement (AMR) Data," In *Proceedings of IEEE Symposium on Large Data Analysis and Visualization (LDAV)*, pages 31-38, October 2012. LBNL-5799E.

E. Wes Bethel, Hank Childs, and Charles Hansen, editors. *High Performance Visualization—Enabling Extreme-Scale Scientific Insight*. Chapman & Hall, CRC Computational Science. CRC Press/Francis–Taylor Group, Boca Raton, FL, USA, November 2012.

K.E. Kunkel, T.R. Karl, H. Brooks, J. Kossin, J.H. Lawrimore, D. Arndt, L. Bosart, D. Changnon, S.L. Cutter, N. Doesken, K. Emanuel, P.Y. Groisman, R.W. Katz, T. Knutson, J. O'Brien, C.J. Paciorek, T.C. Peterson, K. Redmon, D. Robinson, J. Trapp, R. Vose, S. Weaver, M. Wehner, K. Wolter, and D. Wuebbles, "Monitoring and Understanding Trends in Extreme Storms: State of Knowledge," In *Bulletin of the American Meteorological Society*, 2012.

Patrick Oesterling, Christian Heine, Gunther H. Weber, and Gerik Scheuermann, "Visualizing global structure of nd point clouds as topological 1d height fields to enable supervised local data analysis," In *IEEE Transactions on Visualization and Computer Graphics*, 2012. Accepted for publication.

E. Wes Bethel, Surendra Byna, Jerry Chou, Estelle Cormier-Michel, Cameron G. R. Geddes, Mark Howison, Fuyu Li, Prabhat, Ji Qiang, Oliver Rübel, Robert D. Ryne, Michael Wehner, and Kesheng Wu, "Big Data Analysis and Visualization: What Do LINACS

and Tropical Storms Have In Common?" *In 11th International Computational Accelerator Physics Conference, ICAP 2012*, Rostock-Warnemünde, Germany, August 2012.

Surendra Byna, Jerry Chou, Oliver Rübel, Prabhat, Homa Karimabadi, William S. Daughton, Vadim Roytershteyn, E. Wes Bethel, Mark Howison, Ke-Jou Hsu, Kuan-Wu Lin, Arie Shoshani, Andrew Uselton, and Kesheng Wu, "Parallel I.O, Analysis, and Visualization of a Trillion Particle Simulation," *Supercomputing 2012 (SC12)*, Salt Lake City, Utah, Nov. 10-16, 2012.

E. Wes Bethel, David Camp, Hank Childs, Mark Howison, Hari Krishnan, Burlen Loring, Joerg Meyer, Prabhat, Oliver Ruebel, Daniela Ushizima, Gunther Weber, "Towards Exascale: High Performance Visualization and Analytics–Project Status Report," Technical Report LBNL-5767E. Lawrence Berkeley National Laboratory, Berkeley CA, USA, 94720. June, 2012. In DOE Exascale Research Conference, April 16–18, 2012, Portland OR, USA.

Gunther H. Weber, Dmitriy Morozov, Kenes Beketayev, John Bell, Peer-Timo Bremer, Marc Day, Bernd Hamann, Christian Heine, Maciej Haranczyk, Mario Hlawitschka, Valerio Pascucci, Patrick Oesterling, and Gerik Scheuermann, "Topology-based Visualization and Analysis of High-dimensional Data and Time-varying Data at the Extreme Scale," In *DOE Exascale Research Conference,* Portland, OR, April 2012.

Gunther H. Weber and Peer-Timo Bremer, "In-situ Analysis: Challenges and Opportunities," Technical Report LBNL-5692E. Lawrence Berkeley National Laboratory, Berkeley, CA, USA, 94720, April 2012. Position paper for DOE Exascale Research Conference in Portland, OR.

Gunther H. Weber, Kenes Beketayev, Peer-Timo Bremer, Bernd Hamann, Maciej Haranczyk, Mario Hlawitschka, and Valerio Pascucci, "Comprehensible Presentation of Topological Information," Technical Report LBNL-5693E. Lawrence Berkeley National Laboratory, Berkeley, CA, USA, 94720, April 2012. Status report for DOE Exascale Research Conference in Portland, OR.

Mehmet Balman, Eric Pouyoul, Yushu Yao, E. Wes Bethel, Burlen Loring, Prabhat, John Shalf, Alex Sim, and Brian L. Tierney, "Experiences with 100G Network Applications," In *Proceedings of the 5th International Workshop on Data Intensive and Distributed Computing (DIDC 2012)*. Delft, Netherlands. June 2012.

Karen L. Schuchardt, Deborah A. Agarwal, Stefan A. Finsterle, Carl W. Gable, Ian Gorton, Luke J. Gosink, Elizabeth H. Keating, Carina S. Lansing, Joerg Meyer, William A.M. Moeglein, George S.H. Pau, Ellen A. Porter, Sumit Purohit, Mark L. Rockhold, Arie Shoshani, and Chandrika Sivaramakrishnan, "Akuna - Integrated Toolsets Supporting Advanced Subsurface Flow and Transport Simulations for Environmental Management," *XIX International Conference on Computational Methods in Water Resources (CMWR 2012)*, University of Illinois at Urbana-Champaign, June 17-22, 2012.

E. Wes Bethel, "Exploration of Optimization Options for Increasing Performance of a GPU Implementation of a Three-dimensional Bilateral Filter," Technical Report, LBNL-5406E, Lawrence Berkeley National Laboratory, Berkeley CA, USA, 94720. 2012.

E. Wes Bethel and Mark Howison, "Multi-core and Many-core Shared-memory Parallel Raycasting Volume Rendering Optimization and Tuning," *International Journal of High Performance Computing Applications*, November 2012. LBNL-5362E.

Samuel Gerber, Oliver Rübel, Peer-Timo Bremer, Valerio Pascucci and Ross T. Whitaker, "Morse-Smale Regression," *Journal of Computational and Graphical Statistics* (in press) (published online Jan. 2012). LBNL-5682E.

Allen R. Sanderson, Brad Whitlock, Oliver Rübel, Hank Childs, Gunther H. Weber, Prabhat and Kesheng Wu, "A System for Query Based Analysis and Visualization," *Third International Eurovis Workshop on Visual Analytics EuroVA 2012*, K. Matkovic and G. Santucci (Editors), Vienna, Austria, June, 2012, LBNL-5507E.

E. Wes Bethel, David Leinweber, Oliver Rübel, Kesheng Wu (authors in alphabetical order), "Federal Market Information Technology in the Post Flash Crash Era: Roles of Supercomputing," *The Journal of Trading*, Spring 2012, Vol. 7, No. 2, pp. 9-24.

P. Navratil, D. Fussell, C. Lin, and H. Childs, "Dynamic Scheduling for Large-Scale Distributed-Memory Ray Tracing," *Proceedings of EuroGraphics Symposium on Parallel Graphics and Visualization (EGPGV)*, pages 61-70, May 2012, LBNL-5930E. Best paper winner.

K. P. Gaither, H. Childs, K. Schulz, C. Harrison, B. Barth, D. Donzis, and P. Yeung, "Using Visualization and Data Analysis to Understand Critical Structures in Massive Time Varying Turbulent Flow Simulations," *IEEE Computer Graphics and Applications*, 32(4):34-45, 2012.

Jihan Kim, Richard Martin, Oliver Rübel, Maciej; Haranczyk and Berend Smit, "High-throughput Characterization of Porous Materials Using Graphics Processing Units," *Journal of Chemical Theory and Computation*, 2012, 8 (5), pp 1684–1693, DOI: 10.1021/ct200787v, March, 2012.

Prabhat, Oliver Rübel, Surendra Byna, Kesheng Wu, Fuyu Li, Michael Wehner and E. Wes Bethel, "TECA: A Parallel Toolkit for Extreme Climate Analysis," *Third Workshop on Data Mining in Earth System Science (DMESS 2012)* at the International Conference on Computational Science (ICCS 2012), Omaha, Nebraska, June, 2012.

O. Rübel, S.V.E. Keränen, M.D. Biggin, D.W. Knowles, G.H. Weber, H. Hagen, B. Hamann, and E.W. Bethel, "Linking Advanced Visualization and MATLAB for the Analysis of 3D Gene Expression Data," Series: *Mathematics and Visualization, Visualization in Medicine and Life Sciences II, Progress and New Challenges*, Springer Verlag, Heidelberg, Germany, 2012, pp 267-285.

Mark Howison, E. Wes Bethel, Hank Childs, "Hybrid Parallelism for Volume Rendering on Large, Multi- and Many-core Systems," *IEEE Transactions on Visualization and Computer Graphics*, 18(1):17-29, January 2012. LBNL-4370E.

Ushizima, D.M., Weber, G., Morozov, D., Bethel, W., Sethian, J.A., "Algorithms for Microstructure Description applied to Microtomography," *Carbon Cycle 2.0 Symposium*, LBNL, Feb. 10. 2012.

David Camp, Hank Childs, Christoph Garth, Dave Pugmire, Kenneth Joy, "Parallel Stream Surface Computation for Large Data Sets," In *Proceedings IEEE Symposium on Large-Scale Data Analysis and Visualization* 2012. (LBNL-5776E)

2011

MacCarthy, B., Carr, H., Weber, G.H., "Topological Galleries: A High Level User Interface for Topology Controlled Volume Rendering," Technical Report LBNL-5019E Lawrence Berkeley National Laboratory, Berkeley CA, USA, 94720. 2012.

Huang, M.-Y., Mackey, L., Keraenen, S.V.E., Weber, G.H., Jordan, M.I., Knowles, D.W., Biggin, M.D. and Hamann, B. (2011), "Visually relating gene expression and in vivo DNA binding data," in: Wu, F.-X., Zaki, M., Morishita, S., Pan, Y., Wong, S., Christianson, A. and Hu, X., eds., *Proceedings of IEEE International Conference on Bioinformatics and Biomedicine 2011 (IEEE BIBM 2011),* IEEE Computer Society Press, Los Alamitos, California, pp. 586-589.

E. Wes Bethel, David Leinweber, Oliver Rübel, Kesheng Wu (authors in alphabetical order), "Federal Market Information Technology in the Post Flash Crash Era: Roles of Supercomputing," *Workshop on High Performance Computational Finance at Supercomputing 2011* (SC11).

Michael Philpott, Prabhat and Yoshiyuki Kawazoe, "Magnetism and Bonding in Graphene Nanodots with H modified Interior, Edge and Apex." *J. Chem Physics*, 135-8, August 2011.

Richard Martin, Prabhat, David Donofrio and Maciek Haranczyk, "Accelerating Analysis of void spaces in porous materials on multicore and GPU platforms." *International Journal of High Performance Computing Applications.* February 2011.

Suren Byna, Prabhat, Michael Wehner and John Wu, "Detecting Atmospheric Rivers in Large Climate Datasets." *Supercomputing 2011 Workshop on Petascale Data Analytics: Challenges and Opportunities*. 2011.

Tina Zhuo, Prabhat, Chris Paciorek and Cari Kaufman, "Distributed Kriging Analysis for Large Spatial Data." *IEEE ICDM Workshop on Knowledge Discovery from Climate Data*, Vancuover, December, 2011.

Weber, G.H., Bremer, P.-T. and Pascucci, V., "Topological Cacti: Visualizing Contour-based Statistics," in *Topological Methods in Data Analysis and Visualization II*, pp. 63-76, Springer-Verlag, Heidelberg, Germany, 2011.

Weber, G.H., Bremer, P.T., Gyulassy A., and Pascucci, V., "Topology-based Feature Definition and Analysis," in *Numerical Modeling of Space Plasma Flows: Astronum-2010* (Astronomical Society of the Pacific Conference Series), Vol. 444, pp. 292-297, 2011.

Deines E., Weber, G.H., Garth, C., Van Straalen, B. Borovikov, S., Martin, D.F., and Joy, K.I., "On the computation of integral curves in adaptive mesh refinement vector fields," Hagen, H., ed., *Scientific Visualization, Schloss Dagstuhl - Leibnitz Center for Informatics*, Vol. 2, pp. 73-91 2011.

Ushizima, D.M., Parkinson, D., Nico, P., Ajo-Franklin, J., Macdowell, A., Kocar, B., Bethel E.W, Sethian J.A., "Statistical segmentation and porosity quantification of 3D x-ray microtomography," in *XXXIV Applications of Digital Image Processing: Proceeding of SPIE* 2011, August 2011, San Diego, CA, USA.

Prabhat, Quincey Koziol, Karen Schuchardt, E. Wes Bethel, Jerry Chuo, Mark Howison, Mike McGreevy, Bruce Palmer, Oliver Ruebel and John Wu, "ExaHDF5: An I/O Platform for Exascale Data Models, Analysis and Performance," *Scientific Discovery Through Advanced Computing* 2011, Invited Paper.

Ushizima, D.M., Weber, G.H., Ajo-Franklin, J., Kim, Y., Macdowell, A., Morozov, D., Nico, P., Parkinson, D., Trebotich, D., Wan, J., and Bethel E.W., "Analysis and visualization for multiscale control of geologic CO2," in J*ournal of Physics: Conference Series, Proceedings of SciDAC 2011*, July 2011, Denver, CO, USA. LBNL 4948E.

Jerry Chou, Kesheng Wu, Oliver Rübel, Mark Howison, Ji Qiang, Prabhat, Brian Austin, E. Wes Bethel, Rob D. Ryne, and Arie Shoshani. Parallel Index and Query for Large Scale Data Analysis. *In Proceedings of Supercomputing 2011*, Seattle, WA, USA, 2011.

Bremer, P.-T., Weber, G.H., Tierny, J., Pascucci, V., Day, M.S., and Bell, J.B., "Interactive Exploration and Analysis of Large Scale Turbulent Combustion Using Topology-based Data Segmentation" in *IEEE Transactions on Visualization and Computer Graphics*. 2011.

Luke J. Gosink, Christoph Garth, John C. Anderson, E. Wes Bethel, and Kenneth I. Joy, "An Application of Multivariate Statistical Analysis for Query-Driven Visualization," *IEEE Transactions on Visualization and Computer Graphics*, 17(3):264–275, 2011.

Beketayev, K., Weber, G.H., Haranczyk, M., Bremer, P.-T., Hlawitschka, M., and Hamann, B., "Topology-based Visualization of Transformation Pathways in Complex Chemical Systems," in *Computer Graphics Forum* (Special Issue, Proc. Eurographics/ IEEE Symposium on Visualization), 30(3), pp. 663-672, European Association for Computer Graphics, June 2011.

E. Wes Bethel, John van Rosendale, Dale Southard, Kelly Gaither, Hank Childs, Eric Brugger, Sean Ahern, "Visualization at Supercomputing Centers: The Tale of Little Big Iron and the Three Skinny Guys," *IEEE Computer Graphics and Applications*, 31(1):90–95, Jan/Feb 2011.

G.H. Weber, P.-T. Bremer, M.S. Day, J.B. Bell, and V. Pascucci, "Feature tracking using reeb graphs," In: V. Pascucci, X. Tricoche, H. Hagen, and J. Tierny, eds., *Topological Methods in Data Analysis and Visualization: Theory, Algorithms, and Applications*, pp. 241-253. Springer Verlag, 2011.

David Camp, Hank Childs, Amit Chourasia, Christoph Garth, Kenneth Joy, "Evaluating the Benefits of An Extended Memory Hierarchy for Parallel Streamline Algorithms". In *Proceedings IEEE Symposium on Large-Scale Data Analysis and Visualization*, 2011, pp 57-64.

H. Childs, S. Ahern, J. Meredith, M. Miller, and K. I. Joy. Comparative Visualization Using Cross-Mesh Field Evaluations and Derived Quantities. In *Scientific Visualization: Interactions, Features, Metaphors*, Vol. 2 of Dagstuhl Follow-Ups, pages 59-72. Dagstuhl, Germany, 2011.

H. Childs, E. Brugger, B. Whitlock, J. Meredith, S. Ahern, K. Bonnell, M. Miller, G. H. Weber, C. Harrison, D. Pugmire, T. Fogal, C. Garth, A. Sanderson, E. W. Bethel, M. Durant, D. Camp, J. M. Favre, O. Ruebel, P. Navratil, M. Wheeler, P. Selby, and F. Vivodtzev, "VisIt: An End-User Tool For Visualizing and Analyzing Very Large Data," In *Proceedings of SciDAC*, 2011 (LBNL-5945E).

C. Harrison, H. Childs, and K. P. Gaither. "Data-Parallel Mesh Connected Components Labeling and Analysis." In *Proceedings of EuroGraphics Symposium on Parallel Graphics and Visualization*, pages 131-140, April 2011.

Jerry Chuo, John Wu, Prabhat, "FastQuery: A Parallel Indexing System for Scientific Data," *Workshop on Interfaces and Abstractions for Scientific Data Storage*, IEEE Cluster 2011.

Prabhat, Dmitry Zubarev and W. A. Lester, "Statistical Exploration of Electronic Structure of Molecules from Quantum Monte Carlo Simulations." LBNL Technical Report.

Jerry Chuo, John Wu, Prabhat, "Design of FastQuery: how to generalize Indexing and Querying systems for scientific data." LBNL Technical Report.

2010

Min-Yu Huang, Gunther H. Weber, Xiao-Yong Li, Mark D. Biggin, and Bernd Hamann, "Quantitative Visualization of ChIP-chip Data by Using Linked Views." In *Proceedings IEEE International Conference on Bioinformatics and Biomedicine 2010 (IEEE BIBM 2010) Workshops, Workshop on Integrative Data Analysis in Systems Biology (IDASB)*, pp. 195-200.

Patrick Oesterling, Gerik Scheuermann, Sven Teresniak, Gerhard Heyer, Steffen Koch, Thomas Ertl, Gunther H. Weber, "Two-stage Framework for a Topology-Based Projection and Visualization of Classified Document Collections," In *Proceedings IEEE Symposium on Visual Analytics Science and Technology (IEEE VAST)*, pp. 91-98. Salt Lake City, Utah, USA, October

Mark Howison, Quincey Koziol, David Knaak, John Mainzer, John Shalf, "Tuning HDF5 for Lustre File Systems," In *IASDS10–Workshop on Interfaces and Abstractions for Scientific Data Storage*, Heraklion, Crete, Greece. September 2010.

Mark Howison, Andreas Adelmann, E. Wes Bethel, Achim Gsell, Benedikt Oswald, Prabhat, "H5hut: A High-Performance I/O Library for Particle-based Simulations," In *IASDS10–Workshop on Interfaces and Abstractions for Scientific Data Storage*, Heraklion, Crete, Greece. September 2010.

E.A. Carvalho, D.M. Ushizima, F.N.S. Medeiros, C.I.O. Martins, R.C.P. Marques, I.N.S. Oliveira, "SAR imagery segmentation by statistical region growing and hierarchical merging," *Digital Signal Processing*, Elsevier, Vol. 20, Issue 5, Pages 1365-1378, Sept 2010. 20 in Top 25 Hottest Articles in Digital Signal Processing.

D. M. Ushizima, F. N. S. Medeiros, J. Cuadros, C. I. O. Martins, "Vessel Network Detection Using Contour Evolution and Color Components," *Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Buenos Aires, Argentina. Sept 2010.

C. Rycroft, D. M. Ushizima, R. Saye, C. M. Ghajar, J. A. Sethian, "Building a physical cell simulation and comparing with confocal microscopy," *Bay Area Physical Sciences-Oncology Center (NCI) Meeting 2010*, UCSF Medical Sciences, Sept 2010.

M. Howison, E. W. Bethel, and H. Childs, "Hybrid Parallelism for Volume Rendering on Large, Multi-core Systems," In *Journal of Physics: Conference Series, Proceedings of SciDAC 2010*, July 2010, Chattanooga TN, USA.

M. Howison, E. W. Bethel, and H. Childs, "Hybrid Parallelism for Volume Rendering on Large, Multi-core Systems," In *Astronum 2010, 5th International Conference on Numerical Modeling of Space Plasma Flows*, June 2010, San Diego, CA, USA.

Oliver Rübel, Sean Ahern, E. Wes Bethel, Mark. D Biggin, Hank Childs, Estelle Cormier-Michel, Angela DePace, Michael B. Eisen, Charless C. Fowlkes, Cameron G. R. Geddes, Hans Hagen, Bernd Hamann, Min-Yu Huang, Soile V. E. Keränen, David W. Knowles, Cris L. Luengo Hendriks, Jitendra Malik, Jeremy Meredith, Peter Messmer, Prabhat, Daniela Ushizima, Gunther H. Weber, and Kesheng Wu, "Coupling Visualization and Data Analysis for Knowledge Discovery from Multi-dimensional Scientific Data," In *Procedia Computer Science, Proceedings of International Conference on Computational Science*, ICCS 2010, May 2010.

M. Howison, "Comparing GPU Implementations of Bilateral and Anisotropic Diffusion Filters for 3D Biomedical Datasets," Technical Report LBNL-3425E.

T.Fogal, H. Childs, S. Shankar, J. Kruger, R.D. Bergeron, and P. Hatcher, "Large Data Visualization on Distributed Memory Multi-GPU Clusters," *High Performance Graphics 2010*, Saarbruecken, Germany, June 2010.

J. Meredith and H. Childs, "Visualization and Analysis-Oriented Reconstruction of Material Interfaces," *Eurographics/IEEE Symposium on Visualization (EuroVis)*, Bordeaux, France, June 2010.

A.Uselton, K.Antypas, D.M.Ushizima, J.Sukharev, "File System Monitoring as a Window Into User I/O Requirements," *CUG-2010*, Edinburgh, UK, May 24-27th, 2010.

Daniela Ushizima, Cameron Geddes, Estelle Cormier-Michel, E. Wes Bethel, Janet Jacobsen, Prabhat, Oliver Rbel, Gunther Weber, Bernard Hamann, Peter Messmer, Hans Hagen, "Automated Detection and Analysis of Particle Beams in Laser-plasma Accelerator Simulations," *In-Tech*, 367-389, 2010.

D. M. Ushizima, F. Medeiros, "Retinopathy diagnosis from ocular fundus image analysis," *Modeling and Analysis of Biomedical Image, SIAM Conference on Imaging Science (IS10)*, Chicago, Il, April 12-14th, 2010.

D. M. Ushizima and J. Cuadros, "Ocular fundus for retinopathy characterization," *Bay Area Vision Meeting*, Feb 5th, Berkeley, CA 2010.

Hank Childs, David Pugmire, Sean Ahern, Brad Whitlock, Mark Howison, Prabhat, Gunther Weber, E. Wes Bethel, "Extreme Scaling of Production Visualization Software on Diverse Architectures," *IEEE Computer Graphics and Applications*, 30(3):22-31, May/June 2010.

Martin Isenburg, Peter Lindstrom, and Hank Childs, "Parallel and Streaming Generation of Ghost Data for Structured Grids," *IEEE Computer Graphics and Applications*, 30(3):32-44, May/June 2010.

Mark Howison, E. Wes Bethel, and Hank Childs, "MPI-hybrid Parallelism for Volume Rendering on Large, Multi-core Systems," In *Proceedings of the Eurographics Symposium on Parallel Graphics and Visualization*, May 2–3, 2010, Norköping, Sweden. LBNL-3297E. Best Paper Award.

E. Wes Bethel, "Using wesBench to Study the Rendering Performance of Graphics Processing Units," Technical Report LBNL-3025E, Lawrence Berkeley National Laboratory, Berkeley, CA, USA, 94720, 2010.

P.-T. Bremer, G.H. Weber, V. Pascucci, M.S. Day and John B. Bell, "Analyzing and Tracking Burning Structures in Lean Premixed Hydrogen Flames," In *IEEE Transactions on Visualization and Computer Graphics,* 16(2): 248-260, 2010. LBNL-2276E.

O. Rübel, G. H. Weber, M-Y Huang, E. W. Bethel, M. D. Biggin, C. C. Fowlkes, C. Luengo Hendriks, S. V. E. Keränen, M. Eisen, D. Knowles, J. Malik, H. Hagen and B. Hamann, "Integrating Data Clustering and Visualization for the Analysis of 3D Gene Expression Data," *IEEE Transactions on Computational Biology and Bioinformatics*, Vol. 7, No.1, pages 64-79, Jan/Mar 2010. LBNL-382E.

G. H. Weber, S. Ahern, E.W. Bethel, S. Borovikov, H.R. Childs, E. Deines, C. Garth, H. Hagen, B. Hamann, K.I. Joy, D. Martin, J. Meredith, Prabhat, D. Pugmire, O. Rübel, B. Van Straalen and K. Wu, "Recent Advances in VisIt: AMR Streamlines and Query-Driven Visualization," In *Numerical Modeling of Space Plasma Flows: Astronum-2009* (Astronomical Society of the Pacific Conference Series). LBNL-3185E. Vol. 429, pp. 329-334, 2010.

O. Rübel, "Linking Automated Data Analysis and Visualization with Applications in Developmental Biology and High-energy Physics," *Schriftenreihe Informatik, Band 28, Der Dekan (hrsg), Fachbereich Informatik*, Technische Universität Kaiserslautern, December, 2009.

David Camp, Christoph Garth, Hank Childs, David Pugmire, Kenneth Joy, "Streamline Integration using MPI-Hybrid Parallelism on Large Multi-Core Architecture," *IEEE Transactions on Visualization and Computer Graphics*, 17(11):1702–1713, November 2011, (LBNL-4563E).

2009

Mark Howison and Carlo Sequin, "CAD Tools for Creating Space-filling 3D Escher Tiles," In *Computer-Aided Design and Applications*, (6):737-748, Reno, NV, USA, 2009. LBNL-2467E.

C. Garth, E. Deines, K. Joy, E. W. Bethel, H. Childs, G. Weber, S. Ahern, D. Pugmire, A. Sanderson, C. Johnson, "Twists and Turns: Vector Field Visual Data Analysis for Petascale Computational Science," *SciDAC Review*, Number 15, Winter 2009. (LBNL-2983E).

E. Wes Bethel, Hank Childs, Ajith Mascarenhas, Valerio Pascucci, and Prabhat, "Scientific Data Management Challenges in High Performance Visual Data Analysis," In Arie Shoshani and Doron Rotem, editors, *Scientific Data Management: Challenges, Existing Technology, and Deployment*. Chapman & Hall/CRC Press, 2009.

P.-T. Bremer, G.H. Weber, J. Tierny, V. Pascucci, M.S. Day and J.B. Bell, "A Topological Framework for the Interactive Exploration of Large Scale Turbulent Combustion," In *Proceedings of the 5th IEEE International Conference on e-Science*, pp. 247-254, 10.1109/e-Science.2009.42, Oxford, UK, December 2009.

D. Pugmire, H. Childs, C. Garth, S. Ahern, G.H. Weber, "Scalable computation of streamlines on very large datasets." In *Proc. Supercomputing SC09*. Portland, OR, USA. November 2009.

O. Rübel, C.G.R. Geddes, E. Cormier-Michel, K. Wu, Prabhat, G.H. Weber, D.M. Ushizima, P. Messmer, H. Hagen, B. Hamann, and E.W. Bethel, "Automatic Beam Path Analysis of laser Wakefield Particle Acceleration Data," *IOP Computational Science & Discovery*, 2:015005, pp 38, Nov, 2009.

E. W. Bethel, C. Johnson, S. Ahern, J. Bell, P.-T. Bremer, H. Childs, E. Cormier-Michel, M. Day, E. Deines, T. Fogal, C. Garth, C. G. R. Geddes, H. Hagen, B. Hamann, C. Hansen, J. Jacobsen, K. Joy, J. Krüger, J. Meredith, P. Messmer, G. Ostrouchov, V. Pascucci, K. Potter, Prabhat, D. Pugmire, O. Rübel, A. Sanderson, C. Silva, D. Ushizima, G. Weber, B. Whitlock, K. Wu, "Occam's Razor and Petascale Visual Data Analysis," In *Journal of Physics Conference Series, Proceedings of SciDAC*, 2009.

K. Wu, S. Ahern, E. W. Bethel, J. Chen, H. Childs, E. Cormier-Michel, C. Geddes, J. Gu, H. Hagen, B. Hamann, W. Koegler, J. Lauret, J. Meredith, P. Messmer, E. Otoo, V. Perevoztchikov, A. Poskanzer, Prabhat, O. Rübel, A. Shoshani, A. Sim, K. Stockinger, G. Weber, W.-M. Zhang, "FastBit: Interactively Searching Massive Data," *In Journal of Physics Conference Series, Proceedings of SciDAC,* 2009.

N. Shah, N., S. E. Dillard, G.H. Weber, B. Hamann, "Volume visualization of multiple alignment of large genomic DNA," In *Mathematical Foundations of Scientific Visualization, Computer Graphics, and Massive Data Exploration*, Springer-Verlag, Heidelberg, Germany, pp. 325-342, July, 2009.

C. G. R. Geddes, E Cormier-Michel, E. H. Esarey, C. B. Schroeder, J.-L. Vay, W. P. Leemans, D. L.. Bruhwiler, J. R. Cary, B. Cowan, M. Durant, P. Hamill, P. Messmer, P. Mullowney, C. Nieter, K. Paul, S. Shasharina, S. Veitzer, G. Weber, O. Rübel, D. Ushizima, Prabhat, E. W.Bethel, K. Wu, "Large Fields for Smaller Facility Sources," *SciDAC Review*, No. 13, Summer 2009.

Luke Gosink, Kesheng Wu, E. Wes Bethel, John D. Owens, Kenneth I. Joy, "Data Parallel Bin-based Indexing for Answering Queries on Multi-core Architectures," In *Proceedings of the 21st International Conference on Scientific and Statistical Database Management (SSDBM)*, pp 110-129. New Orleans, LA, June 2009.

Shoaib Kamil, Cy Chan, Sam Williams, Leonid Oliker, John Shalf, Mark Howison, E. Wes Bethel, Prabhat, "A Generalized Framework for Auto-tuning Stencil Computations," In *CUG 2009*, May 4-7, 2009, Atlanta GA, USA. Best Paper Award winner.

G. H. Weber, O. Rübel, M.-Y. Huang, A. H. DePace, C. C. Fowlkes, S. V. E. Keränen, C. L. Luengo Hendriks, H. Hagen, D. W. Knowles, J. Malik, M. D. Biggin and B. Hamann, "Visual exploration of three-dimensional gene expression using physical views and linked abstract views," In *IEEE Transactions on Computational Biology and Bioinformatics*, 6(2), April-June, pp. 296-309, 2009.

S. E. Dillard, V. Natarjan, G. H. Weber, V. Pascucci and B. Hamann, "Topology-guided Tessellation of Quadratic Elements," In *International Journal of Computational Geometry and Applications (IJCGA),* 19(2), April, special issue, pp. 196-211, 2009.

M. Hlawitschka, G.H. Weber, A. Anwander, O.T. Carmichael, B. Hamann and G. Scheuermann, "Interactive Volume Rendering of Diffusion Tensor Data." In *Visualization and Processing of Tensor Fields: Advances and Perspectives,* Springer-Verlag, Heidelberg, Germany, pp. 161-176, April, 2009.

E. Wes Bethel, "High Performance, Three-dimensional Bilateral Filtering," Lawrence Berkeley National Laboratory, Tech Report LBNL-1601E, 2009.

Yunfeng Yang, Daniel P Harris, Feng Luo, Wenlu Xiong, Marchin Joachimiak, Liyou Wu, Paramvir Dehal, Janet Jacobsen, Anthony V Palumbo, Adam P. Arkin, and Jizhong Zhou, "Snapshot of iron response in Shewanella oneidensis by gene network reconstruction," *BMC Genomics 2009*, 10:131.

Sharon Borglin, Dominique Joyner, Janet Jacobsen, Aindrila Mukhopadhyay, Terry C. Hazen, "Overcoming the anaerobic hurdle in phenotype microarrays: Generation and visualization of growth curve data for Desulfovibrio vulgaris Hildenborough," *J. Microbiological Methods*, 76 (2), 159-168. February 2009.

E. Wes Bethel, Oliver Rübel, Prabhat, Kesheng Wu, Gunther H. Weber, Valerio Pascucci, Hank Childs, Ajith Mascarenhas, Jeremy Meredith, and Sean Ahern, "Modern Scientific Visualization is More than Just Pretty Pictures," In *Numerical Modeling of Space Plasma Flows: Astronum-2008* (Astronomical Society of the Pacific Conference Series).

C.I.O.Martins, D.M.Ushizima, F.N.S.Medeiros, F.N.Bezerra, R.C.P.Marques, N.D.A.Mascarenhas, "Iterative Self-dual Reconstruction on Radar Image Recovery," *Proc. of IEEE Workshop on Applications of Computer Vision*, Snowbird, Utah, 37-42, 2009. LBNL-3846E.

MATHEMATICS AND COMPUTATIONAL SCIENCE DEPARTMENT

CENTER FOR COMPUTATIONAL SCIENCES AND ENGINEERING

2012

A. Donev, A. Nonaka, Y. Sun, T. Fai, A. Garcia and J. Bell, "Low Mach Number Fluctuating Hydrodynamics of Diffuively Mixing Fluids," submitted for publication.

A. Nonaka, J. B. Bell, M. S. Day, C. Gilet, A. S. Almgren, and M. L. Minion, "A Deferred Correction Coupling Strategy for Low Mach Number Flow with Complex Chemistry," *Combustion Theory and Modelling*, Vol. 16, No. 6, pp. 1053-1088, 2012.

R. Sisneros, C. Malone, A. Nonaka, and S. Woosley, "Investigation of Turbulence in the Early Stages of a High Resolution Supernova Simulation," *Proceedings of the Supercomputing 2012 Conference*, 2012.

W. Schmidt, J. Schulz, L. Iapichino, A. S. Almgren, "Shear-induced mass stripping in a minor merger scenario: Influence of adaptive mesh refinement and the hydro solver," submitted for publication.

R. Tiron, A.S. Almgren, R. Camassa, "Shear instability of internal solitary waves in Euler fluids with thin pycnoclines," *Journal of Fluid Mechanics*, available on CJO2012 doi:10.1017/jfm.2012.366.

S. Williams, D. Kalamkar, A. Singh, A. Deshpande, B. Van Straalen, M. Smelyanskiy, A. Almgren, P. Dubey, J. Shalf, L. Oliker, "Optimization of Geometric Multigrid for Emerging Multi- and Manycore Processors," *Proceedings of the Supercomputing 2012 Conference*, to appear, 2012.

A. Nonaka, A. J. Aspden, M. Zingale, A. S. Almgren, J. B. Bell, and S. E. Woosley, "High-Resolution Simulations of Convection Preceding Ignition in Type Ia Supernovae Using Adaptive Mesh Refinement," *Astrophysical Journal*, 745, 73, 2012.

J. Nordhaus, T.D. Brandt, A. Burrows, A. Almgren, "The Hydrodynamic Origin of Neutron Star Kicks," *Monthly Notices of the Royal Astronomical Society*, 423, Issue 2, June 2012.

J. B. Bell, M. S. Day and M. J. Lijewski, "Simulation of Nitrogen Emissions in a Premixed Hydrogen Flame Stabilized on a Low Swirl Burner," *Proceedings of the Combustion Institute*, 2012.

K. Balakrishnan, A. L. Kuhl, J. B. Bell and V. E. Beckner, "An Empirical Model for the Ignition of Explosively Dispersed Aluminum Particle Clouds," to appear in *Shock Waves*, 2012.

F. Balboa Usabiaga, J. Bell, R. Delgado-Buscalioni, A. Donev, T. Fai, B. Griffith, C. Peskin, "Staggered Schemes for Fluctuating Hydodynamics," to appear in *Multiscale Modeling and Simulation*.

K. Balakrishnan, J. B. Bell and A. L. Kuhl, "A Riemann Solver for a Heterogeneous Two-Phase Flow Model," submitted for publication, 2011.

A.S. Almgren, A.J. Aspden, J. B. Bell, and M. L. Minion, "On the Use of Higher-Order Projection Methods for Incompressible Turbulent Flow," *SIAM J. Sci. Comput.*, to appear.

G. S. H. Pau, J. B. Bell, A. S. Almgren, K. M. Fagnan and M. J. Lijewski, "An Adaptive Mesh Refinement Algorithm for Compressible Two-Phase Flow In Porous Media," *Computational Geosciences*, 16, 577-592, 2012.

2011

K.M. Fagnan, M.J. Lijewski, G.Pau, N.J. Wright, "Acceleration of porous media simulations on the Cray XE6 platform," *Proceedings of CUG2011*, Fairbanks, Alaska.

Emmanouela Rantsiou, Adam Burrows, Jason Nordhaus, Ann Almgren, "Induced Rotation in Three-Dimensional Simulations of Core-Collapse Supernovae: Implications for Pulsar Spins," Astrophysical Journal, 732, 57, 2011.

K. Balakrishnan, A. L. Kuhl, J. B. Bell and V. E. Beckner, "An Empirical Model for the Ignition of Aluminum Particle Clouds Behind Blast Waves," *23rd International Colloquium on the Dynamics of Explosions and Reactive Systems*, Irvine, CA,July 24-29, 2011.

K. Balakrishnan, A. L. Kuhl, J. B. Bell and V. E. Beckner, "Ignition of Aluminum Particle Clouds Behind Reflected Shock Waves," *23rd International Colloquium on the Dynamics of Explosions and Reactive Systems*, Irvine, CA,July 24-29, 2011.

A. L. Kuhl, J. B. Bell, V. E. Beckner and K. Balakrishnan, "Spherical Combustion Clouds in Explosions," *23rd International Colloquium on the Dynamics of Explosions and Reactive Systems*, Irvine, CA,July 24-29, 2011.

A. Nonaka, A. S. Almgren, J. B. Bell, H. Ma, S. E. Woosley, and M. Zingale, "From Convection to Explosion: End-to-End Simulation of Type Ia Supernovae," *Proceedings of SciDAC 2011*, Denver, Colorado, July 2011.

K. Balakrishnan, J. B. Bell, A. L. Kuhl and W. M. Howard, "Riemann Solver for the Nigmatulin Model of Two-Phase Flow," *17th Biennial International Conference of the APS Topical Group on Shock Compression of Condensed Matter*, Chicago, IL, June 26-July 1, 2011.

W. Zhang, L. Howell, A. Almgren, A. Burrows, and J. Bell "CASTRO: A New Compressible Astrophysical Solver. II. Gray Radiation Hydrodynamics," *Astrophysical Journal Supplement Series*, 196, 20, (2011)

A. Donev, A. de la Fuente, J. B. Bell, and A. L. Garcia, "Enhancement of Diffusive Transport by Nonequilibrium Thermal Fluctuations," *JSTAT*, Vol. 2011, P06014, (2011).

A. Donev, A. de la Fuente, J. B. Bell, and A. L. Garcia, "Diffusive Transport by Thermal Velocity Fluctuations," *Phys. Rev. Lett.*, Vol. 106, No. 20, page 204501, 2011.

M. Day, S. Tachibana, J. Bell, M. Lijewski, V. Beckner and R. Cheng, "A combined computational and experimental characterization of lean premixed turbulent low swirl laboratory flames. I. Methane flames," *Combustion and Flame*, 159(1) 275-290 (2012).

A. J. Aspden, J. B. Bell, S. Dong, and S. E. Woosley, "Burning Thermals in Type Ia Supernovae," *Astrophysical Journal*, 738, 94-107, (2011).

M. Zingale, A. Nonaka, A. S. Almgren, J. B. Bell, C. M. Malone, and S. E. Woosley, "The Convective Phase Preceding Type Ia Supernovae," *Astrophysical Journal*, 740, 8, (2011).

A. Nonaka, S. May, A. S. Almgren, and J. B. Bell, "A Three-Dimensional, Unsplit Godunov Method for Scalar Conservation Laws," *SIAM J. Sci. Comput.*, Vol. 33, No. 4, 2011.

S. May, A. Nonaka, A. S. Almgren, and J. B. Bell, "An Unsplit, Higher-Order Godunov Method Using Quadratic Reconstruction for Advection in Two Dimensions," *Comm. App. Math. and Comp. Sci.*, Vol. 6, No. 1, 2011.

A. J. Aspden, M. S. Day, and J. B. Bell, "Turbulence-flame interactions in lean premixed hydrogen: transition to the distributed burning regime," *Journal of Fluid Mechanics*, 680, 287-320, (2011). doi:10.1017/jfm.2011.164.

J. Bell and M. Day, "Adaptive Methods for Simulation of Turbulent Combustion" in *Turbulent Combustion Modeling, Advances, New Trends and Perspectives Series: Fluid Mechanics and Its Applications*, Echekki and Epaminondas (Eds.), Springer, Vol. 95 (2011).

A. J. Aspden, J. B. Bell, and S. E. Woosley, "Turbulent Oxygen Flames in Type Ia Supernovae," *Astrophysical Journal*, 730, 144-151, (2011).

S. E. Woosley, A. R. Kerstein, and A. J. Aspden, "Flames in Type Ia Supernova: Deflagration-Detonation Transition in the Oxygen Burning Flame," *Astrophysical Journal*, 734, 37-41, (2011).

A. J. Aspden, M. S. Day, and J. B. Bell, "Characterization of Low Lewis Number Flames," *Proc. Combust. Inst.*, 33 (2011), 1463-1471.

A. J. Aspden, M. S. Day, and J. B. Bell, "Lewis Number Effects in Distributed Flames," *Proc. Combust. Inst.,* 33 (2011), 1473-1480.

M. S. Day, J. B. Bell, X. Gao and P. Glarborg "Numerical Simulation of Nitrogen Oxide Formation in Lean Premixed Turbulent Flames," *Proc. Combust. Inst.*, 33 (2011), 1591-1599.

M. S. Day, X. Gao, and J. B. Bell, "Properties of Lean Turbulent Methane-Air Flames with Significant Hydrogen Addition," *Proc. Combust. Inst*., 33 (2011), 1601-1608.

C. M. Malone, A. Nonaka, A. S. Almgren, J. B. Bell, and M. Zingale "Multidimensional Modeling of Type I X-ray Bursts. I. Two-Dimensional Convection Prior to the Outburst of a pure 4He Accretor," *Astrophysical Journal*, 728, 118, Feb. 2011.

Ke-Jung Chen, Alexander Heger, and Ann Almgren, "Multi-Dimensional Simulations of Pair-Instability Supernovae," *Computer Physics Communications*, 182:1, 254-256, January 2011.

2010

A. Almgren, J. Bell, D. Kasen, M. Lijewski, A. Nonaka, P. Nugent, C. Rendleman, R. Thomas, and M. Zingale, "MAESTRO, CASTRO and SEDONA -- Petascale Codes for Astrophysical Applications," *Proceedings of SciDAC 2010*, Chattanooga, Tennessee, July 2010.

H. Ma, M. Zingale, S. E. Woosley, A. J. Aspden, J. B. Bell, A. S. Almgren, A. Nonaka, and S. Dong, "Type Ia Supernovae: Advances in Large Scale Simulation," *Proceedings of SciDAC 2010*, Chattanooga, Tennessee, July 2010.

J. Bell, M. S. Day, X. Gao, M. J. Lijewski, "Simulation of Nitrogen Emissions in a Low Swirl Burner," Proceedings of SciDAC 2010, Chattanooga, Tennessee, July 2010.
J. Nordhaus, A. Burrows, A. Almgren, and J. Bell, "Dimension as a Key to the Neutrino Mechanism of Core-Collapse Supernova Explosions," *Astrophysical Journal*, 720, 694, Sept. 2010.

C. C. Joggerst, A. Almgren, and S. E. Woosley, "Three-dimensional simulations of Rayleigh-Taylor mixing in core collapse supernovae with CASTRO," *Astrophysical Journal*, 723, 353, October 2010.

A. Nonaka, A.S. Almgren, J. B. Bell, M. J. Lijewski, C. M. Malone, and M. Zingale, "MAESTRO: An Adaptive Low Mach Number Hydrodynamics Algorithm for Stellar Flows," *Astrophysical Journal Supplement Series*, 188, 358-383, June 2010.

A. S. Almgren, V.E. Beckner, J.B. Bell, M.S. Day, L.H. Howell, C.C. Joggerst, M.J. Lijewski, A. Nonaka, M. Singer, M. Zingale, "CASTRO: A New Compressible Astrophysical Solver. I. Hydrodynamics and Self-Gravity," *Astrophysical Journal*, 715, 1221-1238, June 2010.

Ke-Jung Chen, Alexander Heger, and Ann Almgren, "Two Dimensional Simulations of Pair-Instability Supernovae," *Conference on The First Stars and Galaxies: Challenges for the Next Decade*, Austin, Texas, March 8-11, 2010, Conference Proceedings, published by the American Institute of Physics.

G. S. H. Pau, J. B. Bell, K. Pruess, A. S. Almgren, M. J. Lijewski, K. Zhang, "High resolution simulation and characterization of density-driven flow in CO2 storage in saline aquifers," *Advances in Water Resources*, 33(4):443-455, 2010.

A. Donev, J. B. Bell, A. L. Garcia, and B. J. Alder, "A hybrid particle-continuum method for hydrodynamics of complex fluids," *SIAM J. Multiscale Modeling and Simulation*, 8(3):871-911, 2010.

A. J. Aspden, J. B. Bell, and S. E. Woosley, "Distributed Flames in Type Ia Supernovae," *Astrophysical Journal*, 710, 1654-1663, February 2010.

C. C. Joggerst, A. Almgren, J. Bell, Alexander Heger, Daniel Whalen, and S. E. Woosley, "Primordial Core-Collapse Supernovae and the Chemical Abundances of Metal-Poor Stars,'' *Astrophysical Journal*, 709, 11-26, January 2010.

A. Donev, E. Vanden-Eijnden, A. Garcia, and J. Bell, "On the Accuracy of Explicit Finite-Volume Schemes for Fluctuating Hydrodynamics,'' *Communications in Applied Mathematics and Computational Science*, 5(2):149-197, 2010.

N. Vasudeo, T. Echekki, M. Day, and J. Bell, "The regime diagram for premixed flame kernel-vortex interactions - revisited," *Phys. Fluids* (22) 043602.

2009

A. Donev, A. L. Garcia, and B. J. Alder, " A Thermodynamically-Consistent Non-Ideal Stochastic Hard Sphere Fluid," *J. Stat. Mech*., P11008, 2009.

S. E. Woosley, A. R. Kerstein, V. Sankaran, A. J. Aspden and F. Ropke "Type Ia Supernovae: Calculations of Turbulent Flames Using the Linear Eddy Model," *Astrophysical Journal*, 704, pp.255-273, October 10, 2009.

P.-T. Bremer, G. Weber, J. Tierny, V. Pascucci, M. Day and J. Bell, "A Topological Framework for the Interactive Exploration of Large Scale Turbulent Combustion," *Proceedings of the 5th IEEE International Conference on e-Science* pp. 247-254 (2009)

M. Zingale, A. S. Almgren, J. B. Bell, A. Nonaka, and S. E. Woosley, "Low Mach Number Modeling of Type Ia Supernovae. IV. White Dwarf Convection," *Astrophysical Journal*, 704, 196-210, 2009.

M. M. Scase, A. J. Aspden, and C. P. Caulfield, "The effect of sudden source buoyancy flux increases on turbulent plumes," *Journal of Fluid Mechanics*, 635 pp. 137-169, Sept 2009.

George S. H. Pau, John B. Bell, Karsten Pruess, Ann S. Almgren, Michael J. Lijewski, Keni Zhang, "Numerical studies of density-driven flow in CO2 storage in saline aquifers," *Proceedings of TOUGH Symposium*, September 14-16 2009, Berkeley, California, USA.

George S. H. Pau, Ann S. Almgren, John B. Bell, Michael J. Lijewski, Eric Sonnenthal, Nic Spycher, Tianfu Xu, Guoxiang Zhang, "A Parallel Second-Order Adaptive Mesh Algorithm for Reactive Flow in Geochemical Systems," *Proceedings of TOUGH Symposium*, September 14-16 2009, Berkeley, California, USA.

M.S. Day, J.B. Bell, R.K. Cheng, S. Tachibana, V.E. Beckner and M.J. Lijewski, "Cellular burning in lean premixed turbulent hydrogen-air flames: coupling experimental and computational analysis at the laboratory scale," *SciDAC 2009, J. of Physics: Conference Series*, San Diego, California, July 2009.

E. W. Bethel, C. Johnson, S. Ahern, J. Bell, P.-T. Bremer, H. Childs, E. Cormier-Michel, M. Day, E. Deines, T. Fogal, C. Garth, C. G. R. Geddes, H. Hagen, B. Hamann, C. Hansen, J. Jacobsen, K. Joy, J. Kruger, J. Meredith, P. Messmer, G. Ostrouchov, V. Pascucci, K. Potter, Prabhat, D. Pugmire, O. Rubel, A. Sanderson, C. Silva, D. Ushizima, G. Weber, B. Whitlock, K. Wu "Occam's razor and petascale visual data analysis," *SciDAC 2009, J. of Physics: Conference Series*, San Diego, California, July 2009.

G. S. H. Pau, A. S. Almgren, J. B. Bell, and M. J. Lijewski, "A Parallel Second-Order Adaptive Mesh Algorithm for Incompressible Flow in Porous Media," *Phil. Trans. R. Soc.* A 367, 4633-4654, 2009. LBNL Report LBNL-176E.

S.E. Woosley, A.S. Almgren, A.J. Aspden, J.B. Bell, D. Kasen, A.R. Kerstein, H. Ma, A. Nonaka and M. Zingale, "Type Ia Supernovae: Advances in Large Scale Simulation ," *SciDAC 2009, J. of Physics: Conference Series*, 180, July 2009.

A. S. Almgren, J. B. Bell, A. Nonaka, M. Zingale, "A New Low Mach Number Approach in Astrophysics," *Computing in Science and Engineering*, Vol. 11, No. 2, pp. 24-33, March/April 2009.

Y. Maday, N. C. Nguyen, A. T. Patera, and G. S. H. Pau, "A General Multipurpose Interpolation Procedure: The Magic Points," *Communications on Pure and Applied Analysis*, 8(1), 383–404, 2009.

COMPUTATIONAL COSMOLOGY CENTER

2012

K. Maguire, M. Sullivan, R. S. Ellis, et al., "Hubble Space Telescope studies of low-redshift Type Ia supernovae: evolution with redshift and ultraviolet spectral trends," *Monthly Notices of the Royal Astronomical Society*, 426, 2012.

N. M. Law, A. L. Kraus, R. Street, et al., "Three New Eclipsing White-dwarf-M-dwarf Binaries Discovered in a Search for Transiting Planets around M-dwarfs," *The Astrophysical Journal*, 757, 2012.

R. L. Barone-Nugent, C. Lidman, J. S. B. Wyithe, et al., "Near-infrared observations of Type Ia supernovae: the best known standard candle for cosmology," *Monthly Notices of the Royal Astronomical Society*, 425, 2012.

R. Scalzo, G. Aldering, P. Antilogus, et al., "A Search for New Candidate Super-Chandrasekhar-mass Type Ia Supernovae in the Nearby Supernova Factory Data Set," *The Astrophysical Journal*, 757, 2012.

B. Dilday, D. A. Howell, S. B. Cenko, et al., "PTF 11kx: A Type Ia Supernova with a Symbiotic Nova Progenitor," *Science*, 337, 2012.

E. O. Ofek, R. Laher, J. Surace, et al., "The Palomar Transient Factory photometric catalog 1.0," *Publications of the Astronomical Society of the Pacific*, 124, 2012.

M. M. Kasliwal, S. R. Kulkarni, A. Gal-Yam, et al., "Calcium-rich Gap Transients in the Remote Outskirts of Galaxies," *The Astrophysical Journal*, 755, 2012.

J. C. van Eyken, D. R. Ciardi, K. von Braun, et al., "The PTF Orion Project: A Possible Planet Transiting a T-Tauri Star," *The Astrophysical Journal*, 755, 2012.

J. P. Bernstein, R. Kessler, S. Kuhlmann, et al., "Supernova Simulations and Strategies for the Dark Energy Survey," *The Astrophysical Journal*, 753, 2012.

J. T. Parrent, D. A. Howell, B. Friesen, et al., "Analysis of the Early-time Optical Spectra of SN 2011fe in M101," *The Astrophysical Journal*, 752, 2012.

Y. Cao, M. M. Kasliwal, J. D. Neill, et al., "Classical Novae in Andromeda: Light Curves from the Palomar Transient Factory and GALEX," *The Astrophysical Journal*, 752, 2012.

F. K. Roepke, M. Kromer, I. R. Seitenzahl, et al., "Constraining Type Ia Supernova Models: SN 2011fe as a Test Case," *The Astrophysical Journal*, 750, 2012.

D. Polishook, E. O. Ofek, A. Waszczak, et al., "Asteroid rotation periods from the Palomar Transient Factory survey," *Monthly Notices of the Royal Astronomical Society*, 421, 2012.

X. Wang, L. Wang, A. V. Filippenko, et al., "Evidence for Type Ia Supernova Diversity from Ultraviolet Observations with the Hubble Space Telescope," *The Astrophysical Journal*, 749, 2012.

S. B. Cenko, J. S. Bloom, S. R. Kulkarni, et al., "PTF10iya: a short-lived, luminous flare from the nuclear region of a star-forming galaxy," *Monthly Notices of the Royal Astronomical Society*, 420, 2012.

A. Corsi, E. O. Ofek, A. Gal-Yam, et al., "Evidence for a Compact Wolf-Rayet Progenitor for the Type Ic Supernova PTF 10vgv," *The Astrophysical Journal*, 747, 2012.

N. Smith, S. B. Cenko, N. Butler, et al., "SN 2010jp (PTF10aaxi): a jet in a Type II supernova," *Monthly Notices of the Royal Astronomical Society*, 420, 2012.

A. Horesh, S. R. Kulkarni, D. B. Fox, et al., "Early Radio and X-Ray Observations of the Youngest nearby Type Ia Supernova PTF 11kly (SN 2011fe)," *The Astrophysical Journal*, 746, 2012.

E. O. Ofek, R. Laher, N. Law, et al., "The Palomar Transient Factory Photometric Calibration," *Publications of the Astronomical Society of the Pacific*, 124, 2012.

J. S. Bloom, D. Kasen, K. J. Shen, et al., "A Compact Degenerate Primary-star Progenitor of SN 2011fe," *The Astrophysical Journal*, 744, 2012.

Planck Collaboration, N. Aghanim, M. Arnaud, et al., "Planck intermediate results. I. Further validation of new Planck clusters with XMM-Newton," *Astronomy and Astrophysics*, 543, 2012.

2011

W. Li, J. S. Bloom, P. Podsiadlowski, et al., "Exclusion of a luminous red giant as a companion star to the progenitor of supernova SN 2011fe," *Nature*, 480, 2011.

P. E. Nugent, M. Sullivan, S. B. Cenko, et al., "Supernova SN 2011fe from an exploding carbon-oxygen white dwarf star," *Nature*, 480, 2011.

K. Maguire, M. Sullivan, R. C. Thomas, et al., "PTF10ops - a subluminous, normal-width light curve Type Ia supernova in the middle of nowhere," *Monthly Notices of the Royal Astronomical Society*, 418, 2011.

R. C. Thomas, G. Aldering, P. Antilogus, et al., "Type Ia Supernova Carbon Footprints," *The Astrophysical Journal*, 743, 2011.

I. Arcavi, A. Gal-Yam, O. Yaron, et al., "SN 2011dh: Discovery of a Type IIb Supernova from a Compact Progenitor in the Nearby Galaxy M51," *The Astrophysical Journal*, 742, 2011.

A. Corsi, E. O. Ofek, D. A. Frail, et al., "PTF 10bzf (SN 2010ah): A Broad-line Ic Supernova Discovered by the Palomar Transient Factory," *The Astrophysical Journal*, 741, 2011.

M. A. Agüeros, K. R. Covey, J. J. Lemonias, et al., "The Factory and the Beehive. I. Rotation Periods for Low-mass Stars in Praesepe," *The Astrophysical Journal*, 740, 2011.

Z. Cano, D. Bersier, C. Guidorzi, et al., "XRF 100316D/SN 2010bh and the Nature of Gamma-Ray Burst Supernovae," *The Astrophysical Journal*, 740, 2011.

D. Levitan, B. J. Fulton, P. J. Groot, et al., "PTF1 J071912.13+485834.0: An Outbursting AM CVn System Discovered by a Synoptic Survey," *The Astrophysical Journal*, 739, 2011.

K. Krisciunas, W. Li, T. Matheson, et al., "The Most Slowly Declining Type Ia Supernova 2001ay," *The Astronomical Journal*, 142, 2011.

A. Gal-Yam, M. M. Kasliwal, I. Arcavi, et al., "Real-time Detection and Rapid Multiwavelength Follow-up Observations of a Highly Subluminous Type II-P Supernova from the Palomar Transient Factory Survey," The Astrophysical Journal, 736, 2011.

J. C. van Eyken, D. R. Ciardi, L. M. Rebull, et al., "The Palomar Transient Factory Orion Project: Eclipsing Binaries and Young Stellar Objects," *The Astronomical Journal*, 142, 2011.

A. J. Levan, N. R. Tanvir, S. B. Cenko, et al., "An Extremely Luminous Panchromatic Outburst from the Nucleus of a Distant Galaxy," *Science*, 333, 2011.

I. K. W. Kleiser, D. Poznanski, D. Kasen, et al., "Peculiar Type II supernovae from blue supergiants," *Monthly Notices of the Royal Astronomical Society*, 415, 2011.

R. M. Quimby, S. R. Kulkarni, M. M. Kasliwal, et al., "Hydrogen-poor superluminous stellar explosions," *Nature*, 474, 2011.

Z. Cano, D. Bersier, C. Guidorzi, et al., "A tale of two GRB-SNe at a common redshift of z=0.54," *Monthly Notices of the Royal Astronomical Society*, 413, 2011.

M. Childress, G. Aldering, C. Aragon, et al., "Keck Observations of the Young Metal-poor Host Galaxy of the Super-Chandrasekhar-mass Type Ia Supernova SN 2007if," *The Astrophysical Journal*, 733, 2011.

A. J. Barth, M. L. Nguyen, M. A. Malkan, et al., "Broad-line Reverberation in the Kepler-field Seyfert Galaxy Zw 229-015," *The Astrophysical Journal*, 732, 2011.

M. Sullivan, M. M. Kasliwal, P. E. Nugent, et al., "The Subluminous and Peculiar Type Ia Supernova PTF 09dav," *The Astrophysical Journal*, 732, 2011.

N. Chotard, E. Gangler, G. Aldering, et al., "The reddening law of type Ia supernovae: separating intrinsic variability from dust using equivalent widths," *Astronomy and Astrophysics*, 529, 2011.

M. Kerschhaggl, G. Aldering, P. Antilogus, et al., "Cosmology with the Nearby Supernova Factory," *Progress in Particle and Nuclear Physics*, 66, 2011.

A. M. Smith, S. Lynn, M. Sullivan, et al., "Galaxy Zoo Supernovae," *Monthly Notices of the Royal Astronomical Society*, 412, 2011.

M. M. Kasliwal, S. R. Kulkarni, I. Arcavi, et al., "PTF 10fqs: A Luminous Red Nova in the Spiral Galaxy Messier 99," *The Astrophysical Journal*, 730, 2011.

A. A. Miller, L. A. Hillenbrand, K. R. Covey, et al., "Evidence for an FU Orionis-like Outburst from a Classical T Tauri Star," *The Astrophysical Journal*, 730, 2011.

R. L. C. Starling, K. Wiersema, A. J. Levan, et al., "Discovery of the nearby long, soft GRB 100316D with an associated supernova," *Monthly Notices of the Royal Astronomical Society*, 411, 2011.

R. C. Thomas, P. E. Nugent, and J. C. Meza, "SYNAPPS: Data-Driven Analysis for Supernova Spectroscopy," *Publications of the Astronomical Society of the Pacific*, 123, 2011.

J. Cooke, R. S. Ellis, M. Sullivan, et al., "Hubble Space Telescope Studies of Nearby Type Ia Supernovae: The Mean Maximum Light Ultraviolet Spectrum and its Dispersion," *The Astrophysical Journal*, 727, 2011.

Covey, K. R., L. A. Hillenbrand, A. A. Miller, et al., "PTF10nvg: An Outbursting Class I Protostar in the Pelican/North American Nebula," *The Astronomical Journal*, 141, 2011.

J. D. Neill, M. Sullivan, A. Gal-Yam, et al., "The Extreme Hosts of Extreme Supernovae," *The Astrophysical Journal*, 727, 2011.

Planck Collaboration, N. Aghanim, M. Arnaud, et al., "Planck early results. XXVI. Detection with Planck and confirmation by XMM-Newton of PLCK G266.6-27.3, an exceptionally X-ray luminous and massive galaxy cluster at z ~ 1," *Astronomy and Astrophysics*, 536, 2011.

Planck Collaboration, A. Abergel, P. A. R. Ade, et al., "Planck early results. XXV. Thermal dust in nearby molecular clouds," *Astronomy and Astrophysics*, 536, 2011.

Planck Collaboration, A. Abergel, P. A. R. Ade, et al., "Planck early results. XXIV. Dust in the diffuse interstellar medium and the Galactic halo," *Astronomy and Astrophysics*, 536, 2011.

Planck Collaboration, P. A. R. Ade, N. Aghanim, et al., "Planck early results. XXIII. The first all-sky survey of Galactic cold clumps," *Astronomy and Astrophysics*, 536, 2011.

Planck Collaboration, P. A. R. Ade, N. Aghanim, et al., "Planck early results. XXII. The submillimetre properties of a sample of Galactic cold clumps," *Astronomy and Astrophysics*, 536, 2011.

Planck Collaboration, A. Abergel, P. A. R. Ade, et al., "Planck early results. XXI. Properties of the interstellar medium in the Galactic plane," *Astronomy and Astrophysics*, 536, 2011.

Planck Collaboration, P. A. R. Ade, N. Aghanim, et al., "Planck early results. XX. New light on anomalous microwave emission from spinning dust grains," *Astronomy and Astrophysics*, 536, 2011.

Planck Collaboration, P. A. R. Ade, N. Aghanim, et al., "Planck early results. XIX. All-sky temperature and dust optical depth from Planck and IRAS. Constraints on the 'dark gas' in our Galaxy," *Astronomy and Astrophysics*, 536, 2011.

Planck Collaboration, P. A. R. Ade, N. Aghanim, et al., "Planck early results. XVIII. The power spectrum of cosmic infrared background anisotropies," *Astronomy and Astrophysics*, 536, 2011.

Planck Collaboration, P. A. R. Ade, N. Aghanim, et al., "Planck early results. XVII. Origin of the submillimetre excess dust emission in the Magellanic Clouds," *Astronomy and Astrophysics*, 536, 2011.

Planck Collaboration, P. A. R. Ade, N. Aghanim, et al., "Planck early results. XVI. The Planck view of nearby galaxies," *Astronomy and Astrophysics*, 536, 2011.

Planck Collaboration, J. Aatrokoski, P. A. R. Ade, et al., "Planck early results. XV. Spectral energy distributions and radio continuum spectra of northern extragalactic radio sources," *Astronomy and Astrophysics*, 536, 2011.

Planck Collaboration, P. A. R. Ade, N. Aghanim, et al., "Planck early results. XIV. ERCSC validation and extreme radio sources," *Astronomy and Astrophysics*, 536, 2011.

Planck Collaboration, P. A. R. Ade, N. Aghanim, et al., "Planck early results. XIII. Statistical properties of extragalactic radio sources in the Planck Early Release Compact Source Catalogue," *Astronomy and Astrophysics*, 536, 2011.

Planck Collaboration, N. Aghanim, M. Arnaud, et al., "Planck early results. XII. Cluster Sunyaev-Zeldovich optical scaling relations," *Astronomy and Astrophysics*, 536, 2011.

Planck Collaboration, P. A. R. Ade, N. Aghanim, et al., "Planck early results. XI. Calibration of the local galaxy cluster Sunyaev-Zeldovich scaling relations," *Astronomy and Astrophysics*, 536, 2011.

Planck Collaboration, N. Aghanim, M. Arnaud, et al., "Planck early results. X. Statistical analysis of Sunyaev-Zeldovich scaling relations for X-ray galaxy clusters," *Astronomy and Astrophysics*, 536, 2011.

Planck Collaboration, N. Aghanim, M. Arnaud, et al., "Planck early results. IX. XMM-Newton follow-up for validation of Planck cluster candidates," *Astronomy and Astrophysics*, 536, 2011.

Planck Collaboration, P. A. R. Ade, N. Aghanim, et al., "Planck early results. VIII. The all-sky early Sunyaev-Zeldovich cluster sample," *Astronomy and Astrophysics*, 536, 2011.

Planck Collaboration, P. A. R. Ade, N. Aghanim, et al., "Planck early results. VII. The Early Release Compact Source Catalogue," *Astronomy and Astrophysics*, 536, 2011.

Planck HFI Core Team, P. A. R. Ade, N. Aghanim, et al., "Planck early results. VI. The High Frequency Instrument data processing," *Astronomy and Astrophysics*, 536, 2011.

A. Zacchei, D. Maino, C. Baccigalupi, et al., "Planck early results. V. The Low Frequency Instrument data processing," *Astronomy and Astrophysics*, 536, 2011.

Planck HFI Core Team, P. A. R. Ade, N. Aghanim, et al., "Planck early results. IV. First assessment of the High Frequency Instrument in-flight performance," *Astronomy and Astrophysics*, 536, 2011.

A. Mennella, R. C. Butler, A. Curto, et al., "Planck early results. III. First assessment of the Low Frequency Instrument in-flight performance," *Astronomy and Astrophysics*, 536, 2011.

Planck Collaboration, P. A. R. Ade, N. Aghanim, et al., "Planck early results. II. The thermal performance of Planck," *Astronomy and Astrophysics*, 536, 2011.

Planck Collaboration, P. A. R. Ade, N. Aghanim, et al., "Planck early results. I. The Planck mission," *Astronomy and Astrophysics*, 536, 2011.

2010

S. Darbha, B. D. Metzger, E. Quataert, et al., "Nickel-rich outflows produced by the accretion-induced collapse of white dwarfs: light curves and spectra," Monthly Notices of the Royal Astronomical Society, 409, 2010.

E. O. Ofek, I. Rabinak, J. D. Neill, et al., "Supernova PTF 09UJ: A Possible Shock Breakout from a Dense Circumstellar Wind," *The Astrophysical Journal*, 724, 2010.

M. M. Kasliwal, S. R. Kulkarni, A. Gal-Yam, et al., "Rapidly Decaying Supernova 2010X: A Candidate ".Ia" Explosion," *The Astrophysical Journal*, 723, 2010.

S. Kozlowski, C. S. Kochanek, D. Stern, et al., "SDWFS-MT-1: A Self-obscured Luminous Supernova at z ~= 0.2," *The Astrophysical Journal*, 722, 2010.

D. Poznanski, P. E. Nugent, & A. V. Filippenko, "Type II-P Supernovae as Standard Candles: The SDSS-II Sample Revisited," *The Astrophysical Journal*, 721, 2010.

I. Arcavi, A. Gal-Yam, M. M. Kasliwal, et al., "Core-collapse Supernovae from the Palomar Transient Factory: Indications for a Different Population in Dwarf Galaxies," *The Astrophysical Journal*, 721, 2010.

B. D. Metzger, G. Martinez-Pinedo, S. Darbha, et al., "Electromagnetic counterparts of compact object mergers powered by the radioactive decay of r-process nuclei," *Monthly Notices of the Royal Astronomical Society*, 406, 2010.

S. B. Cenko, N. R. Butler, E. O. Ofek, et al., "Unveiling the Origin of Grb 090709A: Lack of Periodicity in a Reddened Cosmological Long-Duration Gamma-Ray Burst," *The Astronomical Journal*, 140, 2010.

R. Amanullah, C. Lidman, D. Rubin, et al., "Spectra and Hubble Space Telescope Light Curves of Six Type Ia Supernovae at 0.511 z; 1.12 and the Union2 Compilation," *The Astrophysical Journal*, 716, 2010.

A. A. Miller, J. M. Silverman, N. R. Butler, et al., "SN 2008iy: an unusual Type IIn Supernova with an enduring 400-d rise time," *Monthly Notices of the Royal Astronomical Society*, 404, 2010.

R. A. Scalzo, G. Aldering, P. Antilogus, et al., "Nearby Supernova Factory Observations of SN 2007if: First Total Mass Measurement of a Super-Chandrasekhar-Mass Progenitor," *The Astrophysical Journal*, 713, 2010.

N. Smith, A. Miller, W. Li, et al., "Discovery of Precursor Luminous Blue Variable Outbursts in Two Recent Optical Transients: The Fitfully Variable Missing Links UGC 2773-OT and SN 2009ip," *The Astronomical Journal*, 139, 2010.

D. Poznanski, R. Chornock, P. E. Nugent, et al., "An Unusually Fast-Evolving Supernova," *Science*, 327, 2010.

P. Natoli, G. de Troia, C. Hikage, et al., "BOOMERanG constraints on primordial non-Gaussianity from analytical Minkowski functionals," *Monthly Notices of the Royal Astronomical Society*, 408, 2010.

R. Keskitalo, M. A. J. Ashdown, P. Cabella, et al., "Residual noise covariance for Planck low-resolution data analysis," *Astronomy and Astrophysics*, 522, 2010.

Leahy, J. P., M. Bersanelli, O. D'Arcangelo, et al., "Planck prelaunch status: Expected LFI polarisation capability," Astronomy and Astrophysics, 520, 2010.

N. Mandolesi, M. Bersanelli, R. C. Butler, et al., "Planck prelaunch status: The Planck-LFI programme," *Astronomy and Astrophysics*, 520, 2010.

J. A. Tauber, N. Mandolesi, J.-L. Puget, et al., "Planck pre-launch status: The Planck mission," *Astronomy and Astrophysics*, 520, 2010.

C. M. Cantalupo, J. D. Borrill, A. H. Jaffe, T. S. Kisner, & R. Stompor, "MADmap: A Massively Parallel Maximum Likelihood Cosmic Microwave Background Map-maker," *The Astrophysical Journal Supplement Series*, 187, 2010.

2009

N. M. Law, S. R. Kulkarni, R. G. Dekany, et al., "The Palomar Transient Factory: System Overview, Performance, and First Results," *Publications of the Astronomical Society of the Pacific*, 121, 2009.

A. Rau, S. R. Kulkarni, N. M. Law, et al., "Exploring the Optical Transient Sky with the Palomar Transient Factory," *Publications of the Astronomical Society of the Pacific*, 121, 2009.

A. Gal-Yam, P. Mazzali, E. O. Ofek, et al., "Supernova 2007bi as a pair-instability explosion," *Nature*, 462, 2009.

J. D. Simon, A. Gal-Yam, O. Gnat, et al., "Variable Sodium Absorption in a Low-extinction Type Ia Supernova," *The Astrophysical Journal*, 702, 2009.

S. Nobili, V. Fadeyev, G. Aldering, et al., "Constraining Dust and Color Variations of High-z SNe Using NICMOS on the Hubble Space Telescope," *The Astrophysical Journal*, 700, 2009.

S. Bailey, G. Aldering, P. Antilogus, et al., "Using spectral flux ratios to standardize SN Ia luminosities," *Astronomy and Astrophysics*, 500, 2009.

D. Rubin, E. V. Linder, M. Kowalski, et al., "Looking Beyond Lambda with the Union Supernova Compilation," *The Astrophysical Journal*, 695, 2009.

D. Poznanski, N. Butler, A. V. Filippenko, et al., "Improved Standardization of Type II-P Supernovae: Application to an Expanded Sample," *The Astrophysical Journal*, 694, 2009.

M. Sullivan, R. S. Ellis, D. A. Howell, et al., "The Mean Type Ia Supernova Spectrum Over the Past Nine Gigayears," *The Astrophysical Journal*, 693, 2009.

K. Barbary, K. S. Dawson, K. Tokita, et al., "Discovery of an Unusual Optical Transient with the Hubble Space Telescope," *The Astrophysical Journal*, 690, 2009.

Veneziani, M., A. Amblard, A. Cooray, et al., "Subdegree Sunyaev-Zel'dovich Signal from Multifrequency BOOMERANG Observations," *The Astrophysical Journal*, 702, 2009.

M. A. J. Ashdown, C. Baccigalupi, J. G. Bartlett, et al., "Making maps from Planck LFI 30 GHz data with asymmetric beams and cooler noise," *Astronomy and Astrophysics*, 493, 2009.

J. Borrill, "Cosmic Microwave Background Data Analysis: From Time-Ordered Data to Angular Power Spectra," *Data Analysis in Cosmology*, 665, 2009.

MATHEMATICS GROUP

2012

M.J. Cantero, F.A. Grünbaum, L. Moral, and L. Velazquez, "One-dimensional quantum walks with one defect," *Rev. Math Phys.* 24 (2) (2012), 1250002, 52 pages.

A.J. Chorin and X. Tu, "An iterative implementation of the implicit nonlinear filter," *Math. Model. Numer. Anal.* 46 (2012), pp. 535-543.

A.J. Chorin, M. Morzfeld, and X. Tu, "A survey of implicit particle filters for data assimilation," in *State Space Models with Applications in Economics and Finance*, ed. S. Wu, 2012.

M. Ganesh, J. Hesthaven, and B. Stamm, "A reduced basis method for multiple electromagnetic scattering in three dimensions," submitted to *J. Comput. Phys.*, 2012.

M. Garzon, L. Gray, and J.A. Sethian, "Axisymmetric boundary integral formulation for a two fluid system," *Int. J. Numer. Methods. Engrg.*, 2012.

M. Garzon, L. Gray, and J.A. Sethian, "Droplet and bubble pinch-off computations using level sets," *J. Comput. Appl. Math*. 236 (12) (2012), pp. 3034-3041.

S. Govindjee and P.-O. Persson, "A time-domain discontinuous Galerkin method for mechanical resonator quality factor computations," submitted to *J. Comput. Phys.*, 2012.

M. Haranczyk, C. H. Rycroft, and T. Willems, "Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials," *Microporous Mesoporous Mater.* 149 (2012), pp. 134--141.

J. Hesthaven, B. Stamm, and S. Zhang, "Certified reduced basis method for the electric field integral equation," *SIAM J. Sci. Comput.*, 2012.

J. Hesthaven, B. Stamm, and S. Zhang, "New greedy algorithms for the empirical interpolation and reduced basis methods: with applications to high dimensional parameter spaces," submitted to *ESAIM-Math. Model. Num*, 2012.

R. Martin, Prabhat, D. Donofrio, J.A. Sethian, and M. Haranczk, "Accelerating analysis of void space in porous materials on multicore and GPU platforms," I*nt. J. High Perform. Comput.* DOI: 10.1177/1094342011431591, (2012).

M. Morzfeld and A.J. Chorin, "Implicit particle filters for models with partial noise, with an application to geomagnetic data assimilation," submitted to *Nonlinear Proc. Geophys,* 2012.

M. Morzfeld, E. Atkins, X. Tu, and A.J. Chorin, "A random map implementation of implicit filters," *J. Comput. Phys.* 231(4) (2012), pp. 2049-2066.

R. H. Nochetto and B. Stamm, "*A posteriori* error estimates for the Electric Field Integral Equation on polyhedra," arXiv:1204.3930v1 [math.NA], 2012.

P.-O. Persson, "A sparse and high-order accurate line-based discontinuous Galerkin method for unstructured meshes," arXiv:1204.1533 [math.NA], submitted to *J. Comput. Phys*, 2012.

P.-O. Persson, "High-order Navier-Stokes simulations using a sparse line-based discontinuous Galerkin method," Paper AIAA-2012-436, *Proc. 50th AIAA Aerospace Sciences Meeting*, Nashville, TN, 2012.

P.-O. Persson, D. Willis, and J. Peraire, "Numerical simulation of flapping wings using a panel method and a high-order Navier-Stokes solver," *Internat. J. Numer. Methods Engrg.*, 2012.

C.H. Rycroft and F. Gibou, "Simulations of a stretching bar using a plasticity model from the shear transformation zone theory," *J. Comput. Phys.* 231 (2012), pp. 2155--2179.

C.H. Rycroft, T. Lind, S. Güntay, and A. Dehbi, "Granular flow in pebble bed reactors: dust generation and scaling," *Proc. ICAPP 2012*, Paper 12328, Chicago, 2012.

R. Saye and J.A. Sethian, "Analysis and applications of the Voronoi implicit interface method," submitted to *J. Comput. Phys*, 2012.

R. Saye and J.A. Sethian, "The Voronoi implicit interface method and computational challenges in multiphase physics," *Milan J. Math*. 80 (1) (2012).

B. Stamm and T.P. Wihler, "A total variation discontinuous Galerkin approach for image restoration," submitted to *SIAM J. Imaging Sci*, 2012.

2011

A. Babataheri, M. Roper, M. Fermingier, and O. du Rour "Tethered fleximags as artificial cilia," *J. Fluid Mech.* 678 (2011), pp. 5-13.

G.I. Barenblatt, *Flow, Deformation and Fracture*, Cambridge Univ. Press, 2011.

D. Donofrio, M. Haranzcyk, R. Martin, Prabhat, and J.A. Sethian, "Accelerating analysis of void space in porous materials on multicore and GPU platforms," *Proc. Internat. Conf. Chemical Structures*, March, 2011.

J.L. Eftang and B. Stamm, "Parameter multi-domain hp empirical interpolation," submitted to *Int. J. Numer. Methods Engrg*, 2011.

M. Garzon, L. Gray, and J.A. Sethian, "Simulation of the droplet-to-bubble transition in a two-fluid system," *Phys. Rev. E* 83(4) (2011), pp. 046318-1-14.

F.A. Grünbaum, "The Darboux process and a noncommutative bispectral problem: Some explorations and challenges," in *Geometric Aspects of Analysis and Mechanics, Progr. Math.* Vol. 292, eds. E.P. van den Ban and J.A.C. Kolk, Birkhauser, 2011, pp. 161-177.

F.A. Grünbaum and M. Rahman, "A system of multivariable Krawtchouk polynomials and a probabilistic application," *SIGMA* 7 (2011), 17 pages.

F.A. Grünbaum, I. Pacharoni, and J.A. Tirao, "Two stochastic models of a random walk in the U(n)-spherical duals of U(n+1)," *Ann. Mat. Pura Appl.*, DOI: 10.1007/s10231-011-0232-z, (2011).

F.A. Grünbaum, M. de la Iglesia, and A. Martinez-Finkelshtein, "Properties of matrix valued orthogonal polynomials via their Riemann-Hilbert characterization," *SIGMA 7* (2011) 098, 31 pages.

F.A. Grunbaum and M. Rahman, "On a family of 2-variable orthogonal Krawtchouk polynomials," *SIGMA, 6* (2011) 090,12 pages.

J. Hesthaven, Y. Maday, and B. Stamm, "Reduced basis method for the parametrized electric field integral equation," *J. Comput Phys*., 2011.

Jon Wilkening, "Breakdown of self-similarity at the crests of large-amplitude standing water waves," *Phys. Rev. Lett.*, 107 (2011), pp. 184501-1-5.

D.T. Kawano, M. Morzfeld, F. Ma, "The decoupling of defective linear dynamical systems in free motion," *J. Sound Vib.* 330 (21) (2011), pp. 5165-5183.

L.C. Lee, S. J. S. Morris and J. Wilkening, "Stress concentrations, diffusionally-accommodated grain boundary sliding and the viscoelasticity of polycrystals," *Proc. Royal Soc. A*, 467 (2011), pp. 1624-1644.

F. Ma and M. Morzfeld, "A general methodology for decoupling linear dynamical systems," *Proc. 12th East Asia-Pacific Conference on Structural Engineering and Construction* (EASEC-12), Hong Kong, 2011.

F. Ma, M. Morzfeld, and D. T. Kawano, "A method and algorithm for transforming damped linear systems into independent equations," *Proc. 2011 ASME Computers and Information in Engineering Conference*, Washington, DC, DETC2011-47474, 4 pp., 2011.

M. Morzfeld and F. Ma, "The decoupling of damped linear systems in configuration and state spaces," *J. Sound Vib.* 330 (2) (2011), pp. 155-161.

M. Morzfeld and F. Ma, "The decoupling of linear dynamical systems," *Proc. COMPDYN, III ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering*, Corfu, Greece, 2011.

M. Morzfeld, F. Ma, B. N. Parlett, "The transformation of second-order linear systems into independent equations," *SIAM J. Appl. Math*. 71 (4) (2011), pp. 1026-1043.

M. Morzfeld, X. Tu, E. Atkins, and A.J. Chorin, "Random map implementation of implicit filters," submitted to *J. Comput. Phys*, 2011.

J. Peraire and P.-O. Persson, "High-Order Discontinuous Galerkin Methods for CFD," in *Adaptive High-Order Methods in Computational Fluid Dynamics, World Scientific Series Adv. Comput. Fluid Dynamics*, Vol. 2, ed. Z. J. Wang, 2011.

P.-O. Persson, "High-order LES simulations using implicit-explicit Runge-Kutta schemes," Paper AIAA-2011-684, *49th AIAA Aerospace Sciences Meeting and Exhibit*, Orlando, FL, January 2011.

P.-O. Persson and D. J. Willis, "High fidelity simulations of flapping wings designed for energetically optimal flight," Paper AIAA-2011-568, *49th AIAA Aerospace Sciences Meeting and Exhibit*, Orlando, FL, January 2011.

R. Saye and J.A. Sethian, "The Voronoi implicit interface method for computing multiphase physics," *Proc. Nat. Acad. Sci*. USA, 108 (49) (2011), pp. 19498-19503.

B. Stamm, "*A posteriori* estimates for the bubble stabilized discontinuous Galerkin method," *J. Comput. Appl. Math.*, 2011.

M. Theillard, C. H. Rycroft, and F. Gibou, "A multigrid method on non-graded adaptive octree/quadtree cartesian grids," submitted to *J. Sci. Comput*, 2011.

A. Uranga, P.-O. Persson, M. Drela, and J. Peraire, "Implicit large eddy simulation of transition to turbulence at low reynolds numbers using a discontinuous Galerkin method," *Internat. J. Numer. Methods Engrg.*, 87 (1-5) (2011), pp. 232-261.

A. Uranga, P.-O. Persson, M. Drela, and J. Peraire, "Preliminary investigation into the effects of cross-flow on low Reynolds number transition," Paper AIAA 2011-3558, *Proc. 20th AIAA CFD Conference*, Honolulu, HI, 2011.

J. Wilkening, "Overdetermined shooting methods for computing time-periodic water waves with spectral accuracy," submitted to *Comput. Sci. Disc*, 2011.

J. Wilkening and J. Yu, "A local construction of the Smith normal form of a matrix polynomial," *J. Symbolic Comput.*, 46 (2011), pp. 1-22.

T.F. Willems, C.H. Rycroft, M. Kazi, J. C. Meza, and M. Haranczyk "High-throughput analysis of the void space in crystalline porous materials using the Voronoi decomposition," submitted to *Microporous Mesoporous Mater*, 2011.

M.O. Williams, J. Wilkening, E. Shlizerman and J. N. Kutz, "Continuation of periodic solutions in the waveguide array mode-locked laser," *Physica D*, 240/22 (2011), pp. 1791-1804.

M.O. Williams, E. Shlizerman, J. Wilkening and J. N. Kutz, "The low dimensionality of time-periodic standing waves in water of finite and infinite depth," submitted to *SIAM J. Appl. Dyn. Syst*, 2011.

2010

D.M. Ambrose and J. Wilkening, "Computation of symmetric, time-periodic solutions of the vortex sheet with surface tension," *Proc. Natl. Acad. Sci*. USA, 107/8 (2010), pp. 3361-3366.

D.M. Ambrose and J. Wilkening, "Computation of time-periodic solutions of the Benjamin-Ono equation," *J. Nonlinear Sci.*, 20/3 (2010), pp. 277-308.

D.M. Ambrose and J. Wilkening, "Computing time-periodic solutions of nonlinear systems of partial differential equations," *Proc. Hyperbolic Problems: Theory, Numerics, and Applications*, Beijing, 2010.

G.I. Barenblatt and P.J.M. Monteiro, "Scaling laws in nano-mechanics," *Phys. Mesomechan.*, 13 (2010), pp. 245-248.

G.I. Barenblatt, M. Bertsch, and L. Giacomelli, "Steady and quasi-steady thin viscous flows near the edge of a solid surface," *Eur. J. Appl. Math.*, 21 (2010), pp. 253-270.

V. Carey, D. Estep, A. Johansson, M. Larson, and S. Tavener, "Blockwise adaptivity for time dependent problems based on coarse scale adjoint solutions," *SIAM J. Sci. Comput*., 32 (2010), pp. 2121-2145.

E.A. Carvalho, D.M. Ushizima, F.N.S. Medeiros, C.I.O. Martins, R.C.P. Marques, and I.N.S. Oliveira, "SAR imagery segmentation by statistical region growing and hierarchical merging," *Digital Signal Processing*, Elsevier, 2010.

A.J. Chorin and X. Tu, "Iterative implementation of implicit filters," submitted to *Math. Mod. Num. Anal*, 2010.

A.J. Chorin, and X. Tu, "Interpolation and iteration for nonlinear filters," *Math. Mod. Num. Anal*., 2010.

A.J. Chorin, M. Morzfeld, and X. Tu, "Implicit filters for data assimilation," *Commun. Appl. Math. Comp. Sci.*, 5 (2010), pp. 221-240.

M. Haranczyk and J.A. Sethian, "Automatic structure analysis in high-throughput characterization of porous materials," *J. Chem. Theory Comput*., 6(11) (2010), pp. 3472-3480.

L.C. Lee, S.J.S. Morris, J. Wilkening and T.I. Zohdi, "Effects of stress concentrations on the attenuation by diffusionally assisted grain boundary sliding," *Proc. 16th US National Congress of Theoretical and Applied Mechanics*, State College, PA, 2010.

G.C. Leite, D.M. Ushizima, G.G. de Lima, and F.N.S. Medeiros, "Wavelet analysis for wind fields estimation," *Sensors*, 2010.

T. Lind, S. Güntay, A. Dehbi, Y. Liao, and C. H. Rycroft, "PSI project on HTR dust generation and transport," *Proc. HTR* 2010.

F. Ma, M. Morzfeld, and A. Imam, "The decoupling of damped linear systems in free and forced vibration," *J. Sound Vib.* 329 (15) (2010), pp. 3182-3202.

V.-T. Nguyen, J. Peraire, K.B. Cheong, and P.-O. Persson, "A discontinuous Galerkin front tracking method for two-phase flows with surface tension," *Comput. & Fluids*, Vol. 39, No. 1, pp. 1-14, 2010.

P.-O. Persson, D. J. Willis, and J. Peraire, "The numerical simulation of flapping wings at low Reynolds numbers," Paper AIAA-2010-724, *48th AIAA Aerospace Sciences Meeting and Exhibit*, Orlando, FL, January 2010.

T. Potter, "Effective dynamics for N-solitons of the Gross-Pitaevskii equation," 2010, arXiv:1009.4910.

O. Ruebel, S. Ahernc, E.W. Bethel, M. D. Biggin, H. Childs, E. Cormier-Michele, A. DePace, M.B. Eiseng, C.C. Fowlkes, C.G.R. Geddes, H. Hagen, B. Hamann, M. Huang, S.V.E. Keranend, D.W. Knowlesl, C L. Luengo-Hendriks, J. Malik, J. Meredith, P. Messmer, Prabhat, D. Ushizima, G H. Weber, and K. Wua, "Coupling Visualization and Data Analysis for Knowledge Discovery from Multi-dimensional Scientific Data," *International Conference on Computational Science*, ICCS 2010.

C.H. Rycroft, M. Koeppe, and F. Lam, "Recombination lower bounds for missing and circular data: a branch and bound approach," 2010.

C. H. Rycroft, Y. Wong, and M. Z. Bazant, "Fast spot-based multiscale simulations of granular flow," Powder Technol., 200 (2010), pp. 1–11.

D. Ushizima, F. Medeiros, "Retinopathy diagnosis from ocular fundus image analysis," *Modeling and Analysis of Biomedical Image, Imaging Science, SIAM* (accepted) 2010.

D.M. Ushizima, C. G. Geddes, E. Cormier-Michel, E. W. Bethel, J. Jacobsen, Prabhat, O. Ruebel, G. Weber, P. Messmer, B. Hamann, H. Haggen, "Automated detection and analysis of particle beams in laser-plasma accelerator simulations," *Machine Learning*, In-Tech (book chapter), 2010.

M. Westdickenberg and J. Wilkening, "Variational particle scheme for the porous medium equation and for the system of isentropic Euler equations," *ESAIM-Math. Model. Num.*, 44 (2010), pp. 133-166.

J. Wilkening, "Non-existence of a limiting standing wave of extreme form on deep water," (2010), arXiv:1011.2476.

J. Yu, "A local construction of the Smith normal form of a matrix polynomial, and time-periodic gravity-driven water waves," Ph. D. Thesis, UC Berkeley, May, 2010.

2009

J.A. Acebron, S. Lozano, and A. Arenas, "Enhancement of signal response in complex networks induced by topology and noise," *Applications of Nonlinear Dynamics Model and Design of Complex Systems*, Series: Understanding Complex Systems, Springer, 2009.

D. M. Ambrose and J. Wilkening, "Global paths of time-periodic solutions of the Benjamin-Ono equation connecting pairs of traveling waves," *Commun. Appl. Math. Comput. Sci.* 4 (2009), pp. 177-215.

A. Arenas, A. Fernandez, and S. Gomez, "An optimization approach to the structure of the neuronal layout of C. elegans," *Handbook on Biological Networks*, 2009.

A. Arenas, A. Fernandez, and S. Gomez, "Analysis of the structure of complex networks at different resolution levels," *New. J. Phys*., (2009).

A. Arenas, J. Duch, A. Fernandez, and S. Gomez, "Size reduction of complex networks preserving modularity," *New J. Phys.* (2009).

A. Arenas, A. Cabrales, L. Danon, A. Diaz-Guilera, R. Guimera, and F. Vega-Redondo, "Optimal information transmission in organizations: search and congestion," *Rev. Econom. Design*, 2009.

E.W. Bethel, C. Johnson, S. Ahern, J. Bell, P.-T. Bremer, H. Childs, E. Cormier-Michel, M. Day, E. Deines, T. Fogal, C. Garth, C.G.R. Geddes, H. Hagen, B. Hamann, C. Hansen, J. Jacobsen, K. Joy, J. Kruger, J. Meredith, P. Messmer, G. Ostrouchov, V. Pascucci, K. Potter, Prabhat, D. Pugmire, O. Rubel, A. Sanderson, C. Silva, D. Ushizima, G. Weber, B. Whitlock, and K. Wu., "Occam's razor and petascale visual data analysis," *J. Phys. Conference Series, Proc. SciDAC* 2009.

M.K. Cameron, S.B. Fomel, and J.A. Sethian, "Study of a Cauchy problem for a nonlinear elliptic PDE for seismic velocity estimation from time migration," *J. Comput. Phys.*, 228, 7388-7411, 2009.

A.J. Chorin and X. Tu, "Implicit sampling for particle filters," *Proc. Nat. Acad. Sci.* USA, 106 (2009), pp. 17249-17254.

A.J. Chorin and O.H. Hald, *Stochastic Tools for Mathematics and Science*, 2nd edition, Springer, 2009.

Y. Frenkel and V. Roytburd, "Traveling solitary waves for Maxwell-Duffing media: Computation via simulated annealing," *Appl. Math. Letters,* 2009.

Y. Frenkel, I. Gabitov, A. Maimistov, and V. Roytburd, "Propagation of extremely short electromagnetic pulses in doubly-resonant media," *Phys. Rev. A*, 2009.

M. Garzon, J.A. Sethian, L. Gray, "Potential fluid flow computations involving free boundaries with topological changes," *Proc. International Conference on Computational and Mathematical Methods on Science and Engineering*, CMMSE 2009. ISBN 978-84-612 9727-6.

M. Garzon, L. Gray, and J.A. Sethian, "Numerical simulation of non-viscous liquid pinch off using a coupled level set boundary integral method," *Journal Computational Physics*, 228, 17, pp. 6079-6106, 2009

C.G.R. Geddes, E. Cormier-Michel, E.H. Esarey, C.B. Schroeder, J.-L. Vay, W.P. Leemans, D.L. Bruhwiler, J.R. Cary, B. Cowan, M. Durant, P. Hamill, P. Messmer, P. Mullowney, C. Nieter, K. Paul, S. Shasharina, S. Veitzer, G. Weber, O. Rubel, D. Ushizima, Prabhat, E.W. Bethel, and K. Wu, "Large fields for smaller facility sources," *SciDAC Review*, Number 13, Summer 2009.

D. Givon, P. Stinis, and J. Weare, "Dimensional reduction for particle filters of systems with time-scale separation," *IEEE Transactions on Signal Processing* Vol.57 No.2 (2009), pp. 424-435.

M. Haranczyk and J.A. Sethian, "Navigating molecular worms inside chemical labyrinths," *PNAS*, 106, pp. 21472-21477, 2009.

H. Kim and X. Tu, "A three-level BDDC algorithm for Mortar discretization," *SIAM J. Numer. Anal.*, 2009.

J. M. Kinser, *Python for Bioinformatics*, Jones-Bartlett, 2009.

J. Li and X. Tu, "Convergence analysis of a balancing domain decomposition method for solving interior Helmholtz equations," *Numer. Linear Algebra Appl.*, 2009.

R.C.P. Marques, F.N.S. Medeiros, and D.M. Ushizima, "Target detection on SAR images using level set methods," *IEEE Transactions on Systems, Man and Cybernetics*, C(39):2:214-222, 2009.

C.I.O. Martins, D.M. Ushizima, F.N.S. Medeiros, F.N. Bezerra, R.C.P. Marques, and N.D.A. Mascarenhas, "Iterative self-dual reconstruction on radar image recovery," *Proc. of IEEE Workshop on Applications of Computer Vision*, Snowbird, Utah, 37-42, 2009.

P.-O. Persson, "Scalable parallel Newton-Krylov solvers for discontinuous Galerkin discretizations," *Proc 47th AIAA Aerospace Sciences Meeting*, Paper AIAA-2009-606, 2009.

P.-O. Persson, J. Peraire, "Curved mesh generation and mesh refinement using Lagrangian solid mechanics," *Proc. 47th AIAA Aerospace Sciences Meeting*, Paper AIAA-2009-949, 2009.

P.-O. Persson, J. Bonet, J. Peraire, "Discontinuous Galerkin Solution of the Navier-Stokes Equations on Deformable Domains," *Comput. Meth. in Appl. Mech. Engr*., Vol. 198, No. 17-20, pp. 1585-1595, 2009.

M. Roper and M.P. Brenner, "A non-perturbative approximation for the moderate Reynolds number Navier-Stokes equations," *Proc. Nat. Acad. Sci.* USA, March 3, 2009 vol. 106 no. 9 2977-2982.

M. Roper, R. Pepper, M.P. Brenner, and A. Pringle, "Explosively launched spores of ascomycete fungi have drag-minimizing shapes," *Nat. Acad. Sci*. USA, (2009).

A. Roytburd and V. Roytburd, "Domain evolution and polarization of continuously graded ferroelectric films," LBNL-1429E, 2009.

O. Ruebel, C G. Geddes, E. Cormier-Michel, K. Wu, J. Jacobsen, Prabhat, G. Weber, D.M. Ushizima, P. Messmer, H. Haggen, B. Hamann, E.W. Bethel, "Automatic Beam Path Analysis of LaserWakefield Particle Acceleration Data," *IOP Computational Science & Discovery,* 2009.

C.H. Rycroft, "Voro++: A three-dimensional Voronoi cell library in C++," LBNL-1432E, 2009, Chaos 19, 041111 (2009).

C.H. Rycroft, K. Kamrin, and M.Z. Bazant, "Assessing continuum postulates In simulations of granular flow," *J. Mech. Phys. Solids*, 57, 828-839 (2009).

C.H. Rycroft, A.V. Orpe, and A. Kudrolli, "Physical test of a particle simulation model in a sheared granular system," LBNL-1430E, 2009.

C.H. Rycroft, A.V. Orpe, and A. Kudrolli, "Physical test of a particle simulation model in a sheared granular system," *Phys. Rev. E*. 80, 031305, 2009.

J. Suckale and J.-C. Nave, "An alternative numerical model of buoyancy driven models," *Journal of Geophysical Research*, 2009.

J. Suckale, J.-C. Nave, and B. H "Simulations of buoyancy-driven flow in the presence of large viscosity contrasts: I. Numerical approach," submitted, *Geophysical Research*, 2009.

J. Suckale, B. Hager, J.-C. Nave, and L.T. Elkins-Tanton, "Simulations of buoyancy-driven flow in the presence of large viscosity contrasts: II. Ramifications for explosive basaltic," submitted, *Geophysical Research*, 2009.

A. Uranga, P.-O. Persson, M. Drela, and J. Peraire, "Implicit large eddy simulation of transitional flows over airfoils and wings," *Proc. 19th AIAA CFD Conference*, 2009.

J. Weare, "Particle filtering with path sampling, with an application to a bimodal ocean current problem," *J. Comp. Phys.* 228 (2009), pp. 4321-4331.

J. Wilkening, "Practical error estimates for Reynolds' lubrication approximation and its higher order corrections," *SIAM J. Math. Anal.*, 41(2):588-630, 2009.