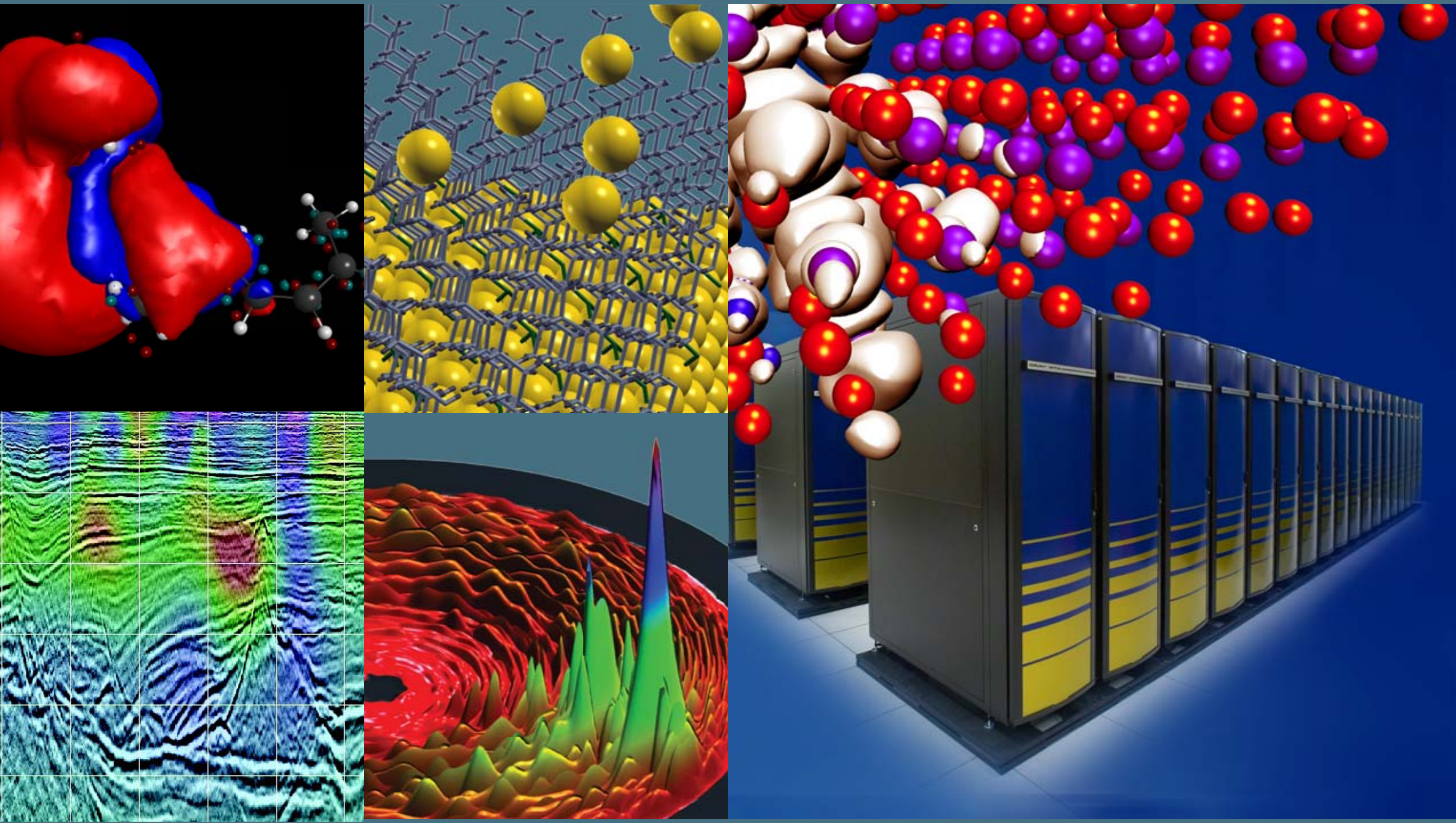
LARGE SCALE COMPUTING AND STORAGE REQUIREMENTS



Basic Energy Sciences

Report of the NERSC / BES / ASCR
Requirements Workshop
February 9 and 10, 2010

DISCLAIMER

This is LBNL report LBNL-XXX, published April 2011

Large Scale Computing and Storage Requirements for Basic Energy Sciences Research

Workshop Report

Conducted February 9-10, 2010

DOE Office of Science
Office of Basic Energy Sciences (BES)
Office of Advanced Scientific Computing Research (ASCR)
National Energy Research Scientific Computing Center (NERSC)

Editors

Richard A. Gerber, NERSC

Harvey J. Wasserman, NERSC

Lawrence Berkeley National Laboratory
One Cyclotron Road
Berkeley, CA 94720

NERSC is funded by the United States Department of Energy, Office of Science, Advanced Scientific Computing Research (ASCR) program. Yukiko Sekine is the NERSC Program Manager and Mark Pederson, Nick Woodward, and Jim Davenport serve as the BES allocation managers for NERSC.

NERSC is located at the Lawrence Berkeley National Laboratory, which is operated by the University of California for the U.S. Department of Energy.

This work was supported by the Directors of the Office of Science, Office of Basic Energy Sciences, and the Office of Advanced Scientific Computing Research, Facilities Division, of the U.S. Department of Energy under contract DE-AC02-05CH11231.

Table of Contents

1 Executive Summary

The National Energy Research Scientific Computing Center (NERSC) is the leading scientific computing facility supporting research within the Department of Energy's Office of Science. NERSC provides high-performance computing (HPC) resources to approximately 4,000 researchers working on about 400 projects. In addition to hosting large-scale computing facilities, NERSC provides the support and expertise scientists need to effectively and efficiently use HPC systems.

In February 2010, NERSC, DOE's Office of Advanced Scientific Computing Research (ASCR) and DOE's Office of Basic Energy Sciences (BES) held a workshop to characterize HPC requirements for BES research through 2013. The workshop was part of NERSC's legacy of anticipating users' future needs and deploying the necessary resources to meet these demands.

Workshop participants reached a consensus on several key findings, in addition to achieving the workshop's goal of collecting and characterizing computing requirements. The key requirements for scientists conducting research in BES are:

1. Larger allocations of computational resources;
2. Continued support for standard application software packages;
3. Adequate job turnaround time and throughput;
4. Guidance and support for using future computer architectures.

This report expands upon these key points and presents others. Several "case studies" are included as significant representative samples of the needs of science teams within BES. Research teams' scientific goals, computational methods of solution, current and 2013 computing requirements, and special software and support needs are summarized in these case studies. Also included are researchers' strategies for computing in the highly parallel, "multi-core" environment that is expected to dominate HPC architectures over the next few years.

NERSC has strategic plans and initiatives already underway that address key workshop findings. This report includes a brief summary of those relevant to issues raised by researchers at the workshop.

2 Office of Basic Energy Sciences Mission

The Office of Basic Energy Sciences (BES) supports world-class research in areas relevant to energy resources, production, conversion, and efficiency and to the mitigation of the adverse impacts of energy production and use, i.e., in areas of materials sciences and engineering, chemical sciences, geosciences, and biosciences.

The Materials Sciences and Engineering Division supports research that explores the origin of macroscopic material behaviors and their fundamental connections to atomic, molecular, and electronic structures. A major goal of the division is to enable a paradigm shift for the deterministic design and discovery of new materials with novel structures, functions, and properties. This Division also seeks to conceptualize, calculate, and predict processes underlying physical transformations, tackling challenging real-world systems—for example, materials with many atomic constituents, with complex architectures, or that contain defects; systems that exhibit correlated emergent behavior; and systems that are far from equilibrium. Basic research in a broad range of energy technologies leads to the development of materials that improve the efficiency, economy, environmental acceptability, and safety of energy generation, conversion, transmission, storage, and use.

The Chemical Sciences, Geosciences, and Energy Biosciences Division supports research that explores fundamental aspects of chemical reactivity and energy transduction over an enormous range of scale and complexity. Phenomena are studied over spatial scales from the sub-nanometer, as defined by the structure of atoms and molecules, to kilometers, appropriate to the behavior of subsurface geological structures, and over time scales defined by the motions of electrons in atoms, attoseconds (10^{-18} seconds), to millennia over which geological change must be understood. At the heart of this research lies the quest to understand and control chemical reactions and the transformation of energy at the molecular scale in systems ranging from simple atoms and molecules, to active catalysts, to complex biochemical or geochemical moieties. This knowledge serves as a basis for the development of new processes for the generation, storage, and use of energy and for mitigation of the environmental impacts of energy use.

BES supports the Office of Science (SC) mission to foster, formulate, and support forefront basic and applied research programs that advance the science and technology foundations necessary to accomplish Department of Energy (DOE) missions.

3 About NERSC

The National Energy Research Scientific Computing (NERSC) Center, which is supported by the U.S. Department of Energy's Office of Advanced Scientific Computing Research (ASCR), serves about 4,000 scientists working on some 400 projects of national importance. Operated by Lawrence Berkeley National Laboratory (LBNL), NERSC is the primary high-performance computing facility for scientists in all research programs supported by the Department of Energy's Office of Science. These scientists, working remotely from DOE national laboratories, universities, other federal agencies, and industry, use NERSC resources and services to further the research mission of the Office of Science (SC). Computational science conducted at NERSC spans a range of scientific disciplines, including physics, materials science, energy research, climate change, and the life sciences. This large and diverse user community runs hundreds of different application codes. NERSC users generate about 1,500 peer-reviewed scientific papers per year. NERSC activities and scientific results are also described in the center's annual reports, newsletter articles, technical reports, and extensive online documentation. In addition to providing computational support for projects allocated by the Office of Science program offices (ASCR, BER, BES, FES, HEP and NP), NERSC directly supports scientific partnerships *across* offices through support of the Scientific Discovery through Advanced Computing (SciDAC[1]) and ASCR Leadership Computing Challenge[2] Programs, as well as several international collaborations in which DOE is engaged. In short, NERSC supports the computational needs of the entire spectrum of DOE open science research.

The DOE Office of Science supports three major High Performance Computing Centers: NERSC and the Leadership Computing Facilities at Oak Ridge and Argonne National Laboratories. NERSC has the unique role of being solely responsible for providing HPC resources to all open scientific research areas sponsored by the Office of Science. The leadership computing facilities support a limited number of select projects, whose research areas may not span all Office of Science objectives and are not restricted to mission-relevant investigations but instead can come from other national or industrial priorities.

This report illustrates NERSC's alignment with, and responsiveness to, DOE program office needs, in this case the needs of the Office of Basic Energy Sciences. The large number of projects supported by NERSC, the diversity of application codes, and its role as an incubator for scalable application codes present unique challenges to the center. As demonstrated by the overall scientific productivity by NERSC users, however, the combination of effectively managed resources and excellent user support services, the NERSC Center continues its 35-year history as a world leader in advancing computational science across a wide range of disciplines.

---

[1] http://www.scidac.gov

[2] http://www.sc.doe.gov/ascr/incite/AllocationProcess.pdf

For more information about NERSC visit the web site at http://www.nersc.gov.

4 Workshop Background and Structure

In support of its mission and to maintain its reputation as one of the most productive scientific computing facilities in the world, NERSC regularly collects requirements from a variety of sources. The NERSC Energy Research Computing Allocations Process (ERCAP) involves questions for applicants about their applications, libraries, and other software needs. Selected detailed studies of the NERSC workload are also carried out focusing on trends in particular areas of computational science. There are also frequent communications with DOE program managers and scientists who use the facility.

In February 2010, the DOE Office of Advanced Scientific Computing Research (ASCR, which manages NERSC), the DOE Office of Basic Energy Sciences (BES), and NERSC held a workshop to gather HPC requirements for current and future science programs supported by BES. This report presents the findings of that workshop.

This document presents several consensus findings. In support of these, "case study" summary reports are included as specific representative samples of the research conducted within BES. The case studies were chosen by the DOE Program Office Managers and NERSC personnel to provide broad coverage in geologic science, chemistry, and materials science. However, BES funds many research endeavors in these fields and the case studies presented here do not necessarily represent the entirety of BES research. Each case study describes its scientific goals today and through 2013, its computational method of solution, and its current computing needs and expected future needs.

Since supercomputer architectures are trending toward systems with chip multiprocessors containing hundreds or thousands of cores per socket and perhaps millions of cores per system, participants were asked to describe their strategy for computing in such a highly parallel, "multi-core" environment.

Requirements presented in this document will serve as input to the NERSC planning process for systems and services, and will help ensure that NERSC continues to provide world-class resources for scientific discovery to scientists and their collaborators in support of the DOE Office of Science, Office of Basic Energy Sciences.

Specific findings from the workshop follow**.**

5 Workshop Demographics

5.1 Participants

Name	Affiliation	NERSC Repo
Brian Austin	University of California, Berkeley	mp208
Hai-Ping Cheng	University of Florida	m526
Peter Cummings	Vanderbilt University	mp138, m526
Randall Cygan	Sandia National Laboratories	
James Davenport	Office of Basic Energy Sciences	
Thomas Devereaux	SLAC National Accelerator Laboratory	m772
Andrew Felmy	Pacific Northwest National Laboratory	mp119
Richard Gerber	NERSC User Services Group	
Mark Jarrell	University of Cincinnati	
Anthony Ladd	University of Florida	
Thomas Miller	California Institute of Technology	m822, mp54
Burkhard Militzer	University of California, Berkeley	m744, m1036
Normand Modine	Sandia National Laboratories	
James Muckerman	Brookhaven National Laboratory	m783
Habib Najm	Sandia National Laboratory	m914, m401
Jeffrey Neaton	Lawrence Berkeley National Laboratory	m387, mp149, mp173
Greg Newman	Lawrence Berkeley National Laboratory	m372
Mark Pederson	Office of Basic Energy Sciences	
Yukiko Sekine	Office of Advanced Scientific Computing Research	
David Skinner	NERSC Outreach	
G. Malcolm Stocks	Oak Ridge National Laboratory	m641
Lin-Wang Wang	Lawrence Berkeley National Laboratory	mp304
Harvey Wasserman	NERSC User Services Group	
Nick Woodward	Office of Basic Energy Sciences	
Margie Wylie	NERSC Communications	
Katherine Yelick	NERSC Director	

5.2 NERSC Projects Represented by Case Studies

BES has more projects at NERSC (166 total; 69 in Chemical Sciences; 88 in Material Sciences; and 9 in Geosciences) than any other office in the DOE Office of Science. Workshop attendees represented projects that used 26 percent of the computer time used by all BES researchers in 2009. The following table lists the projects represented at the workshop. Hours used at NERSC are for 2009.

Project ID (Repo)	**NERSC Computational Project Title**	**Principal Investigator**	**Hours Used at NERSC**
	Materials Sciences		
mp304	*Large Scale Nanostructure Electronic Structure Calculations*	*Lin-Wang Wang*	1.6 M
mp261	*Multiscale Simulations of Particle-, Molecule-Surface Interactions, Simulations of nanowires: Structure, Dynamics, and Quantum Transport, and Structure and Electronic Structure of High T_c Materials*	*Hai-Ping Chen*	1.0 M
m526	*Computational Resources for the Nanomaterials Theory Institute at the Center for Nanophase Materials Sciences*	*Paul Kent, ORNL*	5.3 M
m772	*Simulation of Photon Spectroscopies for Correlated Electron Systems*	*Thomas Devereaux*	0.7 M
m387	*Theory of Nanostructured Materials*	*Jeffrey Neaton*	2.6 M
Total of Materials Sciences projects represented by case studies (23% of Material Sciences total)			**11.2 M**
NERSC 2009 Materials Sciences Total			**49.4 M**
	Chemical Sciences		
mp138	*Molecular-Based Simulation of Complex and Nanostructured Fluids*	*Peter Cummings*	1.0 M
m822	*Sampling diffusive dynamics on long timescales, and simulating the coupled dynamics of electrons and nuclei*	*Thomas Miller*	0.8 M
mp208	*QMC for the Electronic Structure of Molecules*	*William Lester*	4.5 M
m744	*First-principles simulation of dense water, oxygen, and hydrogen at high pressure*	*Burkhard Militzer*	1.2 M
m914; m401	*Computations of Reacting Flow with Detailed Kinetics; A Computational Facility for Reacting Flow Science*	*Habib Najm*	0.05 M
m783	*Computational Studies at BNL of the Chemistry of Energy Production and Use*	*James Muckerman*	0.3 M
Total of Chemical Sciences projects represented by case studies (22% of Chemical Sciences total)			**7.9 M**
NERSC Chemical Sciences Total			**35.7 M**
	Geosciences		
m372	*Large Scale 3D Geophysical Inversion & Imaging*	*Gregory Newman*	3.6 M
mp119	*Computational Studies in Molecular Geochemistry*	*Andrew Felm*	0.8 M
*	*Fracture Dissolution and the Evolution of Permeability: From 2-D to 3-D Simulations*	*Tony Ladd*	0.2 M
Total of Geosciences projects represented by case studies (84% of Geosciences total)			**4.6 M**
NERSC Geosciences Total			**5.5 M**
		Total Represented by Case Studies (26% of BES total)	**23.8 M**
		All BES at NERSC	**91.4 M**

* Ladd was not a NERSC user in 2009. The "Hours Used" value is his estimate of computing required for a fracture dissolution project at the time of the workshop.

6 Findings

6.1 Summary of Requirements

The following is a summary of consensus requirements derived from the case studies.

6.1.1 Scientists in BES need larger allocations of computational resources to meet their research goals.

a) Researchers in BES anticipate needing 1.5 billion hours of computing time at NERSC to support their research in 2013, 16 times more than they used in 2009.

b) Scientists can readily use "all the time that will be realistically available" in the foreseeable future and some important computational techniques (e.g., Landau-Wang) require resources beyond those available at NERSC today.

c) More accurate simulations (e.g., *ab initio* instead of classical methods), which need additional resources, are required to make scientific progress in some fields. More resources are needed to model additional physical interactions and processes (e.g., to study electron-phonon interactions) and to add needed complexity to existing ones (e.g., treating solvent molecules quantum mechanically). More resources are needed to model larger, more realistic systems of interest and to enable a well-demonstrated need to move from 2-D to 3-D simulations for geological phenomena that are inherently asymmetric.

d) Significant allocations – on the order of millions of hours – are needed to support code development, testing, and scaling studies.

6.1.2 The chemistry and material science communities need NERSC to continue to supply and support application software that is standard in their fields.

a) Researchers rely on the increased productivity that NERSC-built application software affords by eliminating the need to spend research time building, debugging, and tuning codes.

b) NERSC should ensure that these applications continue to be optimized for its future systems.

c) NERSC should share its procedures for building open-source software so users can build customized versions for specific science needs.

d) Some researchers would like NERSC to organize a repository containing application software contributed by, and available to, NERSC users.

e) BES researchers need NERSC to continue furnishing and supporting the key math software packages on which these standard application codes strongly depend.

6.1.3 BES researchers need adequate job turnaround and throughput to effectively support their research.

a) Many chemistry and material science studies require a quick job turnaround time because modeling a single physical system requires a *suite* of techniques choosing the methods as the study progresses is an interactive process.

b) Some simulations, such as those involving catalytic redox systems or path integral methods, require fast turnaround for large numbers of simultaneous small- and medium-sized jobs.

c) Improved throughput for larger-concurrency jobs is needed to allow timely modeling of larger physical systems.

d) Some electronic structure calculations, including some using proprietary third-party applications, cannot be checkpointed and require system queue policies that permit very long-running (~several weeks) jobs.

6.1.4 Science teams need guidance implementing, and support for running, their codes on future architectures.

a) NERSC should provide training and technical support for running on multicore and manycore architectures.

b) It is essential that NERSC platforms in 2013 support codes based on MPI and Fortran 90.

c) If NERSC installs a manycore-based system that requires new programming models (e.g., CUDA), these methods need to be supported within existing programming environments (e.g. F90 & MPI), rather than replacing the current programming model completely.

d) NERSC should install available versions of standard community codes on NERSC-hosted architectures, including testbeds like GPU clusters.

e) BES and NERSC should support code development projects to implement (on future systems) codes that have widespread use within BES.

6.2 Other Significant Observations

a) There is key synergy between simulation carried out at NERSC and experimental science, especially at important facilities such as the Molecular Foundry at Berkeley Lab. Successful interplay between theory and simulation depends on the ability to produce simulation results rapidly since the simulations often guide subsequent experimental work.

b) Moving codes to computational systems considerably larger than those available at NERSC today may involve substantial development, debugging, load-balancing analysis, and scaling studies.

c) NERSC has the opportunity to provide facilities and support for the archiving, processing, and sharing necessary for huge volumes of data from ramp-up of the SLAC LINAC Coherent Light Source (LCLS).

d) Architectures with large single-system images and large globally-addressable memory are needed to support some standard quantum chemistry applications (e.g., Gaussian) that do not scale well, yet are crucial to many material science and chemistry researchers.

e) Three key types of BES simulations (joint imaging/inversion of electromagnetic and seismic data, Quantum Monte Carlo, and molecular dynamics) appear to be amenable to GPU acceleration, at least for some types of computation. At the time of this workshop several key issues for these areas remain, however, such as the need for double-precision floating point and hardware square root, and the question of the generality of GPU-accelerated molecular dynamics.

6.3 Computing Requirements

The following table lists the projected computational hours required by research projects represented by case studies in this report and a multiplier these projected hours represent relative to 2009 use. All but one of the projects had an allocation at NERSC in 2009. “Total Scaled Requirement” in the table represents the hours needed by all NERSC BES projects in 2013 if the total is scaled by the same factor as that needed by the projects represented by case studies in this report.

Computing Hours Required for BES Case Study Projects

NERSC Computational Project Title	**Principal Investigator**	**Hours Needed in 2013**	**Increase Over 2009 NERSC Use**
Materials Sciences			
Large Scale Nanostructure Electronic Structure Calculations	*Lin-Wang Wang*	10 M	6
Multiscale Simulations of Particle-, Molecule-Surface Interactions, simulations of nanowires: Structure, dynamics, and quantum transport, and structure and electronic structure of high T_c materials	*Hai-Ping Cheng*	10 M	10
Computational Resources for the Nanomaterials Theory Institute at the Center for Nanophase Materials Sciences	*Paul Kent*	50 M	10
Simulation of Photon Spectroscopies for Correlated Electron Systems	*Thomas Devereaux*	8 M	11
Theory of Nanostructured Materials	*Jeffrey Neaton*	25 M	10
Total of Material Sciences projects represented by case studies		**103 M**	**9.1**
NERSC Materials Sciences Total Scaled Requirement		**450 M**	**9.1**
Chemical Sciences			
Molecular-Based Simulation of Complex and Nanostructured Fluids	*Peter Cummings*	50 M	50
Sampling Diffusive Dynamics on Long Time Scales, and Simulating the Coupled Dynamics of Electrons and Nuclei	*Thomas Miller*	30 M	40
Quantum Monte Carlo for the Electronic Structure of Molecules	*William Lester*	100 M	22
First-Principles Simulation of Dense Water, Oxygen, and Hydrogen at High Pressure	*Burkhart Militzer*	20 M	16
Computations of Reacting Flow with Detailed Kinetics; A Computational Facility for Reacting Flow Science	*Habib Najm*	35 M	650
Computational Studies at BNL of the Chemistry of Energy Production at Use	*James Muckerman*	2.4 M	8
Total of Chemical Sciences projects represented by case studies		**237 M**	**30**
NERSC Chemical Sciences Total Scaled Requirement		**1,000 M**	**30**
Geosciences			
Large Scale 3D Geophysical Inversion and Imaging	*Gregory Newman*	20 M	5.6
Computational Studies in Molecular Geochemistry	*Andrew Felmy*	20 M	25
Fracture Dissolution and the Evolution of Permeability: From 2-D to 3-D Simulations	*Tony Ladd*	2 M	10
Total of Geosciences projects represented by case studies		**42 M**	**9.1**
NERSC Geosciences Total Scaled Requirement		**50 M**	**9.1**
	Total Represented by Case Studies	**382 M**	**16**
	NERSC BES Total Scaled Requirement	**1,500 M**	**16**

7 NERSC Initiatives and Plans

NERSC has initiatives already underway and long-term strategic plans that address some requirements presented in this report. A summary of these initiatives and plans is presented in this section.

7.1 Compute Resources

NERSC plans to increase its computational resources with the Hopper system (the NERSC-6 Project) in 2011 and NERSC-7 approximately three years later. Hopper represents a 4.5-fold increase in aggregate application performance over the Franklin quad-core system that went into production in mid-2009. Technology trends suggest that NERSC-7 will continue this trend of ~2X per year with nearly constant space and hardware costs, but growing electrical costs will require a larger budget even to meet this overall increase of approximately 16X over 2009 capacity by 2013. These increases are projected for application performance, translated to a normalized "core hour" on the Franklin XT4 system. Manycore processing nodes and other architectural changes may yield different per-core performance, or make the notion of a core obsolete, so these plans should be viewed as increases in application throughput of the machine.

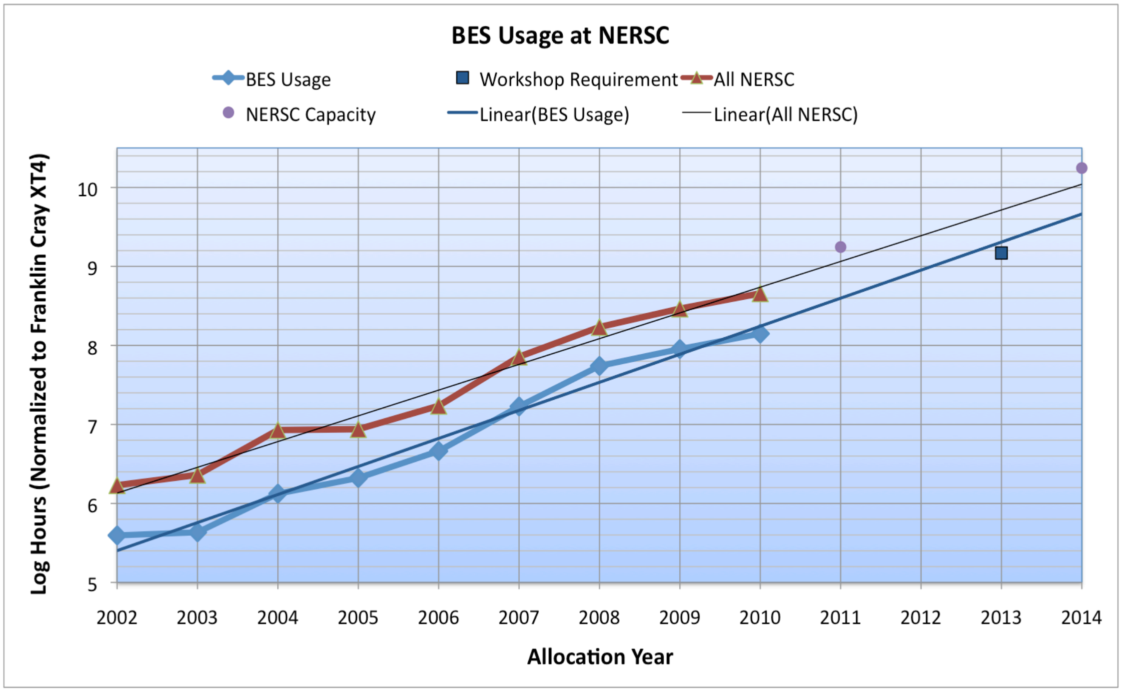


Figure: NERSC delivers an increasing computational capacity at a rate of about 2X per year.

The requirements for computing hours set forth in this report are expected to be accommodated by NERSC's existing plans, which are contingent on the funding required to procure, run, and support these future systems.

7.2 Software Support

NERSC recognizes that BES scientists rely on standard community software applications for much of their research. NERSC is committed to maintaining a rich repository of application software and programming libraries for the BES community. NERSC acquires, builds, installs and supports optimized versions of about 15 of the leading chemistry and materials sciences codes, including VASP and NAMD. The NERSC consultants have extensive experience with these applications and how to optimize their use for particular science problems. NERSC installs and supports as many community codes as feasible given funding levels for software acquisition and consulting support.

While some software is adapted for leading-edge HPC systems quickly, some popular packages are not, and NERSC may require additional resources (either directly to NERSC or indirectly to third-party software developers) to support these packages on the newest systems.

NERSC will investigate the feasibility of creating a repository of software contributed by its user community. As part of the SciDAC Outreach Center NERSC provides a source code software repository that is available to all DOE researchers but NERSC does not have a repository specifically tailored to installations for NERSC machines.

7.3 Job Turnaround and Throughput

NERSC monitors job queues regularly and tries to optimize them to maximize throughput and minimize wait times for all users, although in general, longer wait times are a side effect of running systems at high utilization. NERSC offers premium queues with higher scheduling priority (and double the charge factor) so researchers can get fast turnaround on special occasions. NERSC is exploring ways to achieve better predictability via advanced scheduling algorithms and the Magellan cloud computing research project is investigating other service models that would allow different levels of service with respect to turnaround.

To address testing at scale NERSC allows users to schedule dedicated time on all or part of the Cray XT4 Franklin system. Some users have taken advantage of this reservation system to run and debug at scale interactively. NERSC is investigating the ramifications of offering such a service on a regularly scheduled basis to projects that require data processing on a predictable schedule.

Running systems with high availability facilitates optimal throughput and turnaround. NERSC has a long history of working closely with vendors to reduce system software and hardware failures. Vendors are subject to strict contractual uptime and availability metrics on the major NERSC systems. NERSC plans to have major systems overlap in lifetimes, typically supporting two major systems at a time. This strategy mitigates the impact of downtimes during a new system's "breaking in" period by having the stability provided by a mature system.

NERSC recognizes that some community applications do not checkpoint easily and queues are set up to try to accommodate those codes. NERSC implements long-running queues subject to constraints imposed by each system's mean time between failure and maintenance schedule.

7.4 User Support on Future Architectures

NERSC will continue its leading role in identifying and resolving performance, programmability, and other issues raised by new architectures. However, radically different architectures — such as manycore and GPU systems — will require more staff to help users adapt their codes and workflows to run efficiently (or perhaps run at all) on these systems.

NERSC is fielding early architecture testbeds, including a 42-node GPU system and the Magellan cloud system, as requested in this series of workshops. NERSC has the expertise to play a leadership role in the upcoming architecture and software transformation; its impact limited only by staffing levels and funds to acquire testbed machines. Because of the critical nature of upcoming architecture innovations, which may dramatically change the kinds of machines NERSC can deploy, NERSC is working with the Berkeley Lab Computational Research Division (CRD) and the rest of the DOE community to help influence vendor designs, evaluate emerging architectures, and influence the programming models, algorithms and software that will be needed.

NERSC recognizes the investment its users have made in MPI-based codes, and we expect support for MPI to continue through NERSC-7. However, MPI-only codes may not perform optimally on future systems and NERSC has already started holding training classes and workshops on mixed-mode MPI/OpenMP programming. Additionally, NERSC has an aggressive set of activities addressing emerging programming models, such as UPC and Co-Array Fortran, through support for these models in machine procurements, close interaction with CRD, and in collaboration with Cray in our "Programming Models Center of Excellence." The SciDAC-e Postdoc program at NERSC also has postdocs working on CAF, UPC, CUDA, OpenMP, OpenCL, and other novel programming systems as researchers embedded within NERSC user teams.

8 Geosciences

8.1 BES Geoscience Overview

Andrew Felmy, PNNL

The BES Geosciences research program supports research aimed at developing an understanding of fundamental Earth processes that can be used as a foundation for efficient, effective, and environmentally sound use of energy resources, and provides an improved scientific basis for advanced energy and environmental technology. The program's primary research focuses on examining the geochemistry of mineral-fluid interactions, geophysical interrogation of the Earth’s crust, and the basic properties of rocks, fluids, and minerals.

Research challenges in the geosciences cover a broad range of time and length scales from the molecular (from Angstroms to microns) where the focus is on *mineral-water complexity and dynamics* and *nanoparticle and colloid chemistry and physics*; to the mesoscale (microns to millimeters) where the focus is more on *dynamic imaging of flow and transport*; to the field scale (meters to kilometers) where the focus centers around *integrated characterization, modeling, and monitoring of geologic systems* and *transport properties and in situ characterization of fluid trapping, isolation and immobilization* (Figure 8-1). From a computational standpoint such a range of scientific challenges is difficult to address since at its most basic form one is often solving different fundamental sets of equations (e.g. at the molecular scale the Schrödinger equation $H\psi=E\psi$ or at the mesoscale or field scale the convective-dispersion continuum equations). Each research challenge has its own issues in terms of spatial or temporal discretization, basis function representation, and resulting mathematical algorithms.

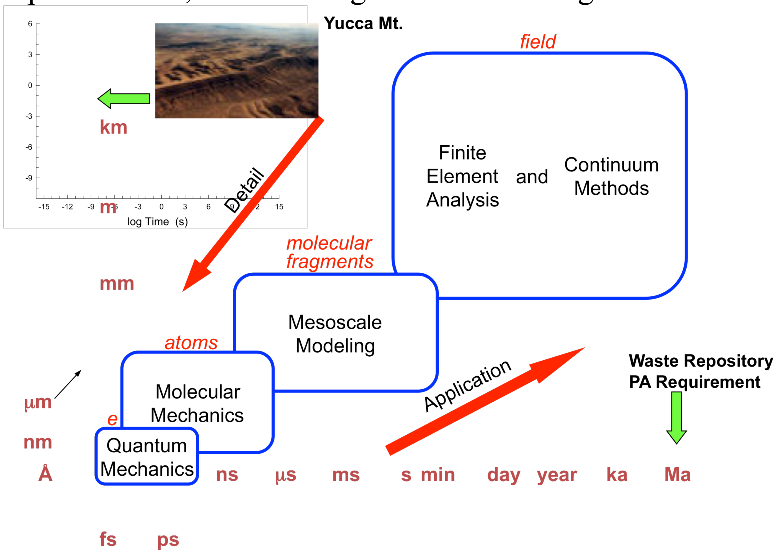


Figure 8-1. Range of time and length scales of importance in Geosciences research exhibiting the different computational approaches at each scale.

The area of molecular simulation is one that significantly affects the Geosciences research program primarily in the field of geochemistry as described above. However, the computational resources needed to support this research also support the broader mission of the Chemical Sciences, Geosciences, and Biosciences Division within BES and are described in detail in other sections of this workshop report. The remainder of this chapter consists of two case studies emphasizing computation needs for problems involving greater length and time scales. The first highlights the need to move from 2-D to 3-D for simulations of flow and reactivity. This case study relates directly to the priority research direction *transport properties and in-situ characterization of fluid trapping, isolation and immobilization* identified in the "Basic Research Needs for Geosciences: Facilitating 21st Century Energy Systems" report from 2007, and specifically to the disposal of CO_2 in deep geologic formations. The second case study focuses on geophysical imaging of subsurface geological structure and fluids, which relates to the grand challenge of *integrated characterization, modeling, and monitoring of geologic systems* identified in the same 2007 report.

An active area of geoscience research at NERSC for which no case study is presented is computational studies in molecular geochemistry. This effort consists of molecular level simulations, including: *ab initio* molecular modeling of interactions between Fe(III) reducing bacteria and iron-oxides, which has implications for biogeochemical activity in subsurface environments; modeling aimed at providing a molecular scale understanding of surface complexation reactions at oxide, oxyhydroxide, and silicate minerals; first-principles molecular dynamics simulations to improve understanding of natural processes that lead to the concentration of metal species in natural waters and deposition of ore rich formations; and using first-principles simulation to model and characterize the mechanism of dissociative reduction of halogenated hydrocarbons by iron-oxide and nanoparticulate iron systems. A summary of computational requirements for a project covering these areas is presented in the following table (PI, Andrew Felmy, PNNL) and included in the summary tables in Chapters 5 and 6, above.

	Current (2009)	**In 2013**
Computational Hours	800 K	20 M
Parallel Concurrency	100	1 K – 20 K
Wall Hours per Run	100	100
Aggregate Memory	100 Gb	20 GB – 25 TB
Memory per Core	1 GB	2 GB

8.2 BES Geoscience Case Studies

8.2.1 Fracture Dissolution and the Evolution of Permeability: From 2-D to 3-D Simulations

Principal Investigator: Tony Ladd, University of Florida

8.2.1.1 Summary and Scientific Objectives

Storing excess CO_2 in geologic formations of carbonate phases is a possible means of mitigating the contribution of fossil fuel emissions to global warming. Modeling the sequestration of CO_2 in these formations must take into account the evolution of permeability through dissolution and precipitation in the network of fractures that run throughout the formations. Theories of fracture evolution typically assume that dissolution is uniform in the direction normal to the flow, and that the front is therefore planar. Under these circumstances, a rapidly eroding front can penetrate only a short distance into the fracture. However, numerical simulations have shown that, in some cases, dissolution is highly non-uniform, leading to a much more rapid increase in permeability. Recently, we showed by a linear stability analysis that non-uniform dissolution is the expected behavior on geophysical scales rather than the exception. A typical example is shown in Figure 8-2, below. Here a dissolving fluid is injected into a smooth fracture, with small variations (less than 1 %) in aperture. Rather quickly an unstable situation develops where the flow becomes concentrated into a few channels. These channels compete with each other and eventually flow in the shorter channels ceases and they become retarded. Eventually all the flow is being drained through a single channel.

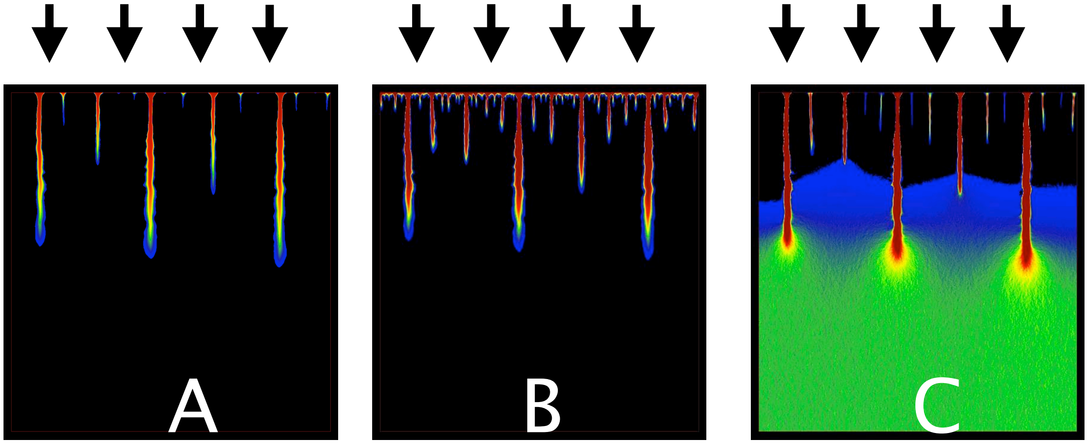


Figure 8-2. Concentration (A), aperture (B) and flow fields (C) in a narrow fracture.

We concluded that the evolution of fracture permeability is better approximated by the growth of localized regions of high permeability than by conventional models based on uniform aperture growth. In particular, estimates of breakthrough times can differ by orders of magnitude in scenarios of relevance to sequestration. Our results suggest that, after sufficient dissolution, the characteristic length scale of the variation in fracture aperture will be determined by the reaction kinetics and flow rate, and not by the initial heterogeneity of the fracture.

Our work relies on numerical simulations to provide crucial insight and input into our developing theoretical and modeling efforts. We find that we can represent the pressure field about long channels quite accurately and very efficiently using a conformal mapping technique. The pressure fields in Figure 8-3 are determined by direct numerical simulation (upper) and conformal mapping (lower). If we can understand and predict how the individual channels grow then we can model the later development of the fracture aperture with a simple and fast conformal solver. At present the challenges are to 1) understand the transition from the linear to non-linear regimes and 2) to understand the growth of individual channels in the pressure field of the neighboring channels.

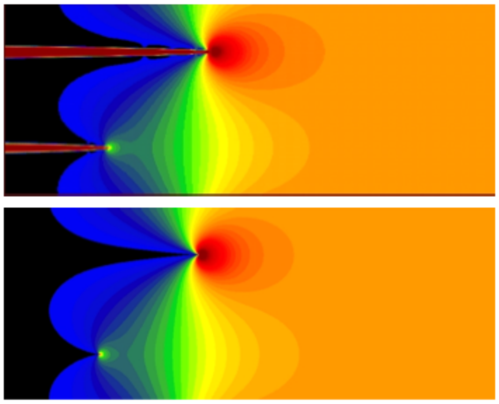


Figure 8-3. Pressure field around two channels in a narrow fracture. The upper panel is from a 2-D simulation and the lower panel is from a conformal mapping.

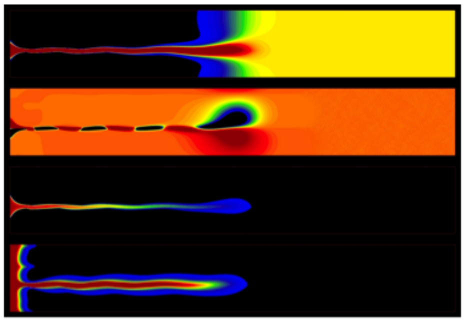


Figure 8-4 An individual channel. From top to bottom: axial velocity, transverse velocity, aperture, and concentration fields.

Much of our work uses 2-D simulations, which are simple, robust and fast. Figure 6-4 shows the results of a typical simulation of a single channel growing in a narrow sample. Although the channel is nearly straight (3rd panel down), the transverse velocity field (2nd panel down) shows a very periodic sequence of oscillations, which we do not yet fully understand. This suggests that to really comprehend the details of the growth of individual channels three-dimensional simulations are necessary.

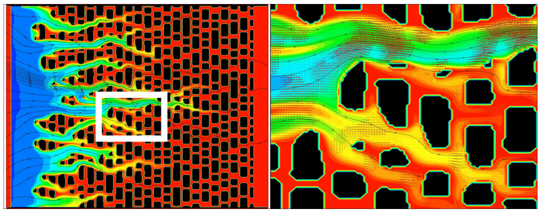


Figure 8-5 Concentration (color contours), velocity field (arrows) and streamlines in a very rough fracture.

Figure 8-5 illustrates the level of detail available from a 3-D simulation, showing a slice through a portion of the fracture. The patches of high and low velocity exist because the velocity is a fully 3-D field. This information is missing from 2-D simulations. Three-dimensional simulations are important but computationally very time consuming. Below we sketch out requirements for such simulations.

8.2.1.2 Methods of Solution

We use a hybrid lattice-Boltzmann/finite-difference scheme for dissolution in fractured media. Bézier polynomials are used to construct a piecewise continuous surface, which is robust in complex topographies. A standard MRT lattice-Boltzmann method with interpolated boundary conditions is used to solve for the flow field at each iteration. However, in the absence of an upwind differencing scheme, LB methods are limited in the range of grid Péclet numbers that can be accessed, and we therefore use a finite-difference method for the concentration field. For more details see Yu and Ladd, *J. Comp. Phys*. **229**, 6450-6465 (2010).

8.2.1.3 HPC Requirements

A reasonable fracture area to allow study of the initiation of a substantial number of channels on scales of geophysical significance would be 1 x 1 m. With a mean fracture aperture in the fully dissolved system of 1 mm that means a volume of 10^{-3} m^3. The scale of the fracture aperture in the undissolved state would be 0.1 mm, so with ten points across the channel the resolution of the finest grid would be 10^{-5} m. Thus, we would need on the order of 10^{12} grid points for the concentration, but considerably less for the fluid solver – say 10^{10}. The total memory requirement would be about 10 TB for a single copy of the data or 20 TB if update in place cannot be used. In addition, information to store the geometry of the surface is required – of the order of 1TB. The result would be about 10,000-20,000 processes with about 1 GB of memory per process.

There are two stages to the computation – first the fluid equations are solved, followed by the transport equations. Past experience suggests of the order of 1,000 iterations for the flow solver and 100 iterations of the concentration solver for each cycle of erosion. Usually we use of the order of 1000 cycles to model the dissolution process. For the fluid equations, a single iteration of the flow solver would take about 2,000 CPU seconds for 10^{10} grid points. The total for the fluid solver would then be about 2 x 10^9 CPU seconds or about 3 days of computing with 10^4 processes. The concentration solver would take about 20,000 CPU seconds for 10^{12} grid points. So the computation time for the flow and concentration solvers is similar, as we have found in smaller scale runs. We would therefore need about 150-200 hours per simulation on 10,000 processors. Probably on the order of 10 runs would be needed to make substantial advances, so a total CPU time of about 2,000 hours on 10,000 processors would be enough to make a real difference.

Prior to this there should be substantial development and testing, to make sure this large resource is well utilized. A reasonable path forward would begin with simulations on the order of 100 processes, similar to what we have already done and scale to 1,000 processes over time. Here we would need substantial access – on the order of 1,000 hours on 1,000 processors to develop and test. Finally we could then proceed to the large-scale calculations. The computational needs are summarized in the following table.

8.2.1.4 Computational and Storage Requirements Summary

	Current (2009)	**In 2013**
Computational Hours	200 K	2 M
Parallel Concurrency	100	10 K
Wall Hours per Run	100	100
Aggregate Memory	100 GB	10 TB
Memory per Core	1 GB	1 GB
Archival Storage	200 K	2 M

8.2.1.5 Support Services and Software

We made no use of DOE support services or software.

8.2.1.6 Emerging HPC Architectures and Programming Models

All our codes are currently within the MPI framework – we do not have a multithreaded code at the present time.

8.2.2 Geophysical Subsurface Imaging: Jointly Image/Invert Electromagnetic and Seismic Data

Principal Investigator: Gregory Newman, LBNL
NERSC Repo: m372

8.2.2.1 Summary and Scientific Objectives

Three-dimensional geophysical imaging is now receiving considerable attention for mapping geophysical subsurface attributes (electrical conductivity and seismic velocity). Uses of this technology include mapping subsurface fluids in potential oil and gas and geothermal reservoirs and monitoring sequestered carbon dioxide, which is of significant relevance to the Department of Energy missions in energy security and environmental stewardship of its facilities.

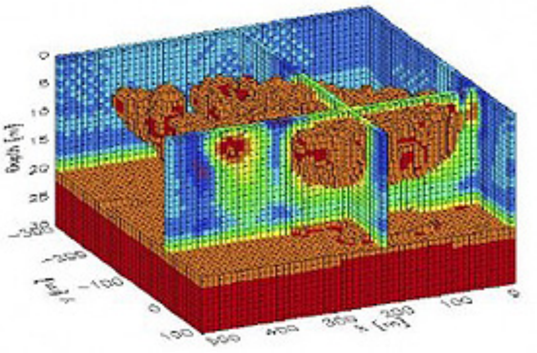


Figure 8-6 EMGEO ElectroMagnetic Geological Mapper

An important objective in the next three to five years is to *jointly* image/invert geophysical data obtained from electromagnetic (EM) and seismic surveys to obtain images of how the geophysical attributes are distributed in the Earth in 3D. Up to this point we have only been able to image EM or seismic data separately. By completing the imaging experiment using joint data sets we expect a significant improvement in the resolution at which the subsurface can be imaged. To ensure that both data sets have similar spatial resolution, seismic data will be Laplace-Fourier transformed before inversion with the EM data. Results from this endeavor will then be useful in constructing initial velocity models critical for successfully reverse time migration and full wave form seismic imaging schemes for high resolution imaging of subsurface reflectivity and subtle velocity variations. We envisage imaging data sets involving 1,000s of shots (geophysical sources) where each shot involves the solution of wave equations in acoustic and electromagnetic wave propagation in three dimensions.

In order to achieve our goal we need to reduce the time to solution of the imaging problem by exploiting graphic processing units (GPUs) and field programming gate arrays (FPGAs). GPUs and FPGAs offer a potential ten-fold speed up over existing and established CPU technologies. Because the GPU and FPGA technology is rapidly evolving, we propose to investigate both types of hardware accelerators.

8.2.2.2 Methods of Solution

The 3-D imaging problem has large computational demands, due to the computationally expensive solution of wave propagation problems for EM and acoustic fields on a 3-D finite-difference grid. These wave propagation problems give rise to large and sparse linear systems of equations that are complex-symmetric and are solved using iterative Krylov methods. The imaging component of the algorithm is based on non-linear conjugate gradient and a steepest decent optimization scheme. A line search is also involved over the parameter space to find the optimal model step size to reduce the data misfit. Multiple levels of parallelization are exploited, both over the data and model space (domain decomposition) using MPI. Excellent scaling of the algorithm has been observed up to 32,000 cores.

The computational workhorses in the algorithm are the forward/adjoint solve of the wave equations needed to compute predicted data, the gradient of the objective function that is being minimized, and the line search, which insures an acceptable model step in the update.

Each solve is expected to be on the order of 25 million field unknowns, each requiring close to 0.5 GB of memory. To be successful, the algorithm should scale up to tens of thousands of cores, with the solver for each shot/source distributed across subsets of these cores (100s of cores). However, with GPU/FPGA hardware, the domain decomposition of the solver may be no longer necessary, resulting in significant reduction in the computational resources needed for the imaging problem, by an order of magnitude.

The computational bottleneck is in the Krylov solver, specifically a matrix-vector multiply. Time to access the cache memory to complete the multiply is limiting. New technologies are needed to achieve tenfold speedup, where clustered GPUs or FPGAs offer significant potential. Note: the algorithmic approach adopted using standard CPU technology is not IO constrained because of the modest data volumes; it is only CPU constrained.

8.2.2.3 HPC Requirements

Our algorithms exploit a hierarchical parallel framework, using several levels of parallelism where both model space and data space are distributed over an arbitrarily large number of cores. They have been shown to scale to tens of thousands of processors and use the MPI interface extensively. In one imaging experiment, 32,768 tasks/processors on the IBM Watson Research Blue Gene/L supercomputer were successfully utilized. Over a 24-hour period we were able to image a large-scale field data set that previously required over four months of processing time on 1,024 cores of an Intel or AMD / InfiniBand cluster.

In the next three to five years a tenfold increase in HPC resources would provide the means to attack the 3-D joint geophysical inverse problem – imaging problems involving different geophysical attributes and data types, where there is no established rock physics model to couple the different attributes and data; i.e. EM, gravity and seismic. It would enable a critical step forward in understanding subsurface geological systems and processes in a self-consistent manner. It would further improve the likelihood of success of full waveform imaging and migration of seismic data at its greatest resolution and detail and would allow for the possibility of imaging across multiple scale lengths, incorporating different types of geophysical data and attributes in the process.

It is essential that this hierarchical framework with the MPI interface be preserved; otherwise, migration to new hardware will be too costly. However, that being said, one should still be able to port to different computing environments without too much difficulty, and this specifically includes GPU/FPGA clusters when such platforms come into production in the next three to five years. Thus, it is critical that new computing hardware preserve old computing paradigms such as MPI and Fortran90, coupled with new programming languages (CUDA) to allow for easier migration of software to new hardware, and still achieve the anticipated speedup in computation.

8.2.2.4 Computational and Storage Requirements Summary

	Current (2009)	**In 2013**
Main Science Driver	**Separate Imaging**	**Joint Imaging**
Computational Hours	3.6 M	20 M
Parallel Concurrency	5 K	25 K
Wall Hours per Run	76 Hours	760 Hours
Aggregate Memory	2.5 TB	10 TB
Memory per Core	400 MB	4 GB
I/O per Run	100 GB	500 GB
On-Line Storage Needed	0.1 GB	100 GB
Data Transfer	10 GB/Month	100 GB/Month
Archival Storage	0	1 GB

8.2.2.5 Support Services and Software

Support for the following languages, MPI and Fortran 90, CUDA or OpenCL if viable, needs to be a priority. Technical support on effectively running many MPI threads across multiple GPUs and or FPGAs is also important.

8.2.2.6 Emerging HPC Architectures and Programming Models

GPUs and FPGAs are emerging in the HPC world as viable technologies to accelerate scientific computation by at least tenfold. However achieving the anticipated accelerations is a different matter. Programming on GPUs and FPGAs is still in its infancy and has at times been described as jungle programming. Programming FPGAs has different issues, specifically knowledge of assembly/machine languages and detailed hardware design that the research scientists will have limited time and or motivation to master. Either type of accelerator is also known not to handle large amounts of I/O too efficiently or effectively. It is still unclear which technology will emerge the winner; currently GPUs appear to be most popular. CUDA, the programming language for NVIDIA® GPUs, is still evolving. As an example, CUDA does not support complex square root. Moreover, until now, C and C++ developers targeting GPU accelerators have had to rely on language extensions to their programs. Use of GPUs from Fortran applications has been extremely limited. Programmers of x64+GPU systems have been required to program at a detailed level including a need to understand and specify data usage information and manually construct sequences of calls to manage all movement of data between the x64 host and GPU. However, NVIDA has been working with The Portland Group to develop a CUDA compiler that provides Fortran language support for NVIDIA's CUDA-enabled GPUs. It is anticipated that Fortran developers and Fortran legacy codes with data parallel problems will be able to use this compiler to harness the massive parallel computing capability of NVIDIA GPUs to create high performance applications for scientific computing. Other issues with GPUs concern double precision arithmetic, and memory checking to avoid dropping bits in a scientific computation. It is anticipated that fast double precision GPUs will be on the market in the next several months. With existing GPU technology double precision computation is about eight times slower than single precision. The imaging problems discussed in this case study require double precision computation, without dropping bits in computation, for effective iterative Krylov solver methods.

9 Material Science and Engineering

9.1 BES Material Science and Engineering Overview

Jim Davenport, Program Manager
Theoretical Condensed Matter Physics
Materials Sciences and Engineering Division
Office of Basic Energy Sciences

Materials Science is a large area of research devoted to the study of the properties of metals and alloys, magnets, superconductors, semi-conductors, biomaterials, and other aggregates of matter. Research in materials science is carried out in BES in the Materials Sciences and Engineering Division and also in the Scientific User Facilities Division. Examples of work supported by each Division are given below in the case studies.

The Materials Sciences and Engineering Division supports research that explores the origin of macroscopic material behaviors and their fundamental connections to atomic, molecular, and electronic structures. The portfolio stresses the need to probe, understand, and control the interactions of phonons, photons, electrons, and ions with matter to direct and control energy flow in materials systems over multiple time and length scales. Such understanding and control are critical to science-guided design of highly efficient energy conversion processes, such as new electromagnetic pathways for enhanced light emission in solid-state lighting and multi-functional nanoporous structures for optimum charge transport in batteries and fuel cells.

This Division also seeks to conceptualize, calculate, and predict processes underlying physical transformations, tackling challenging real-world systems - for example, materials with many atomic constituents, with complex architectures, or that contain defects; systems that exhibit correlated emergent behavior; and systems that are far from equilibrium. Such understanding will be critical to developing predictive capability for complex systems behavior, such as in superconductivity and magnetism. The Division also plays a major role in enabling the nanoscale revolution. The development of new nanoscale materials, as well as the methods to characterize, manipulate, and assemble them, create an entirely new paradigm for developing new and revolutionary energy technologies.

The Scientific User Facilities Division supports the operation of a nationwide suite of major facilities that provide open access to sophisticated instrumentation needed to probe and create materials for scientists of many disciplines from academia, national laboratories, and industry. These large-scale user facilities consist of a complementary set of intense x-ray sources, neutron scattering centers, electron beam characterization capabilities, and research centers for nanoscale science.

The BES-supported suite of facilities and research centers provides a unique set of analytical tools for studying the atomic structure and functions of complex materials. These facilities provide key capabilities to correlate the microscopic structure of materials with their macroscopic properties. The synchrotron light sources, producing photons largely over a very wide range of photon energies (from the infrared to hard x-rays), shed light on fundamental aspects of the physical world, investigating energy, momentum, and position using the techniques of spectroscopy, scattering, and imaging applied over various time scales. Neutron sources take advantage of the electrical neutrality and special magnetic properties of the neutron to probe atoms and molecules and their assembly into materials. Electron beam instruments provide the spatial resolution needed to observe individual nanostructures and even single atoms by exploiting the strong interactions of electrons with matter and the ability to readily focus beams of charged particles. The Nanoscale Science Research Centers provide the ability to fabricate complex nanostructures using chemical, biological, and other synthesis techniques, and to characterize, assemble, and integrate them into devices.

9.2 Material Science and Engineering Case Studies

9.2.1 Density Functional Theory: Carrier Dynamics in Nano Solar Cells

Principal Investigators: Lin-Wang Wang, Lawrence Berkeley National Laboratory
NERSC Repo: mp304
Hai-Ping Cheng, Department of Physics and QTP, University of Florida
NERSC Repos: mp261 and m526

9.2.1.1 Summary and Scientific Objectives

The goal of this work is the study of carrier dynamics and electron transport in nanoparticle-based solar cells. Solar cells made of inorganic nanocrystals, or mixtures of nanocrystals and organic polymers and molecules have the potential to provide cheaper photovoltaic power than traditional thin film solar cells. But there are unique challenges in using nanostructures for solar cell applications. These include the large number of surface states, the strong exciton binding energies, the nano-interfaces, the lack of doping, and the possibility of unintended internal electric fields. To understand the working mechanism of a nano solar cell, one has to understand the carrier dynamics in a nanostructure. The carrier dynamics is first determined by the electronic structures of the nanosystem, then by the electron-hole interaction, internal electric field, and by electron-phonon interactions. Through this project, different nanosystems are investigated. A typical example can be a core shell CdSe/CdS seeded nanorod. At one side, a gold particle is attached to the nanorod for electron collection, while the other surface of the CdS is attached to a random poly(3-hexylthiophene) (P3HT) polymer for hole collection.

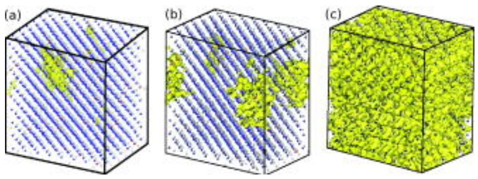


Figure 9-1 Isosurface plot of $ZnTe_{0.9685}O_{0.0315}$ alloy wave functions.

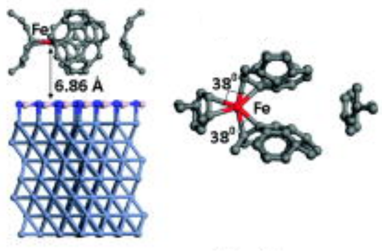


Figure 9-2. Most stable structure of Fe-doped monolayer C60 on a Nickel surface from a spin-dependent generalized gradient approximation calculation.

Simulating the carrier dynamics of such a nanosystem requires a *suite* of techniques and computer codes, and this is a key distinguishing feature of this work. This is because of the complex nature of the problem and the different levels of approximation in different electronic structure methods and codes. A single code cannot capture both the diverse physical phenomena involved and the computational techniques related to the carrier dynamics. This is often the case for material science simulations. Also, a fixed, single large-scale simulation cannot produce the results. Instead, different simulations need to be carried out, and new simulations need to be pursued based on the results of previous simulations. It is thus an interactive process. The computer is used as a tool to aid our scientific inquiry. Like any other inquiries, human intelligence is at the center of the process, and different tools need to be used intermittently in a nimble fashion. Even with modern large-scale computers, this feature of material science simulation remains the same. This requires that each simulation be finished in a timely fashion, and the computer (the tool) be available around the clock. As a mid-term to long-term effort, we will integrate codes into multi-scale and multi-task computing frameworks and automate the procedure of end-to-end simulations.

9.2.1.2 Methods of Solution

In order to study the electron transport in a nanocrystal, we first need to know the electronic eigen energies and wave functions in such a system. In our example, this would be the electronic states for the CdSe/CdS seeded core/shell nanorod, and the electronic states for the random P3HT polymers. There are two different ways to approach this. The first is to use the non-self-consistent charge patching method (CPM) where the charge density of the nanocrystal is constructed from the atomic charge motifs. After the charge density of the system is determined, and hence the single particle Hamiltonian, the single particle Schrodinger's equation can is solved using the folded spectrum method implemented in the PEscan code. The second way is to use the linear scaling three-dimensional fragment (LS3DF) method to self-consistently solve the charge density of a ten- to hundred-thousand atom system under the density functional theory (DFT). This self-consistent calculation is necessary when there are internal electric fields in the system. However, computationally, the LS3DF method is much more expensive than the CPM. Nevertheless, the LS3DF can be thousand times faster than the direct DFT calculations as implemented in codes such as VASP and PEtot. The reason is that the direct DFT method scales as the third power of the number of atoms, while LS3DF scales linearly with the number of atoms.

There are many problems where the atomic structures of the system need to be determined computationally. This can be the case of a semiconductor-metal nanocontact (e.g., the Au/CdS nanocontact), the polymer-semiconductor surface attachment (the P3HT/CdS attachment), or the surface passivation. There are two different ways to study such problems. The first is to propose different atomic structures (e.g., for the cation exchange formed at the Au/CdS interface), and then use the DFT method (e.g., VASP code or PWSCF) to relax the atomic structure and yield the total energies of different proposed structures. The lowest energy case will likely to be the correct atomic structure. However, there are situations where the possible configurations are too numerous. In that case, a DFT based molecular dynamics (MD) simulation can be performed, e.g., to study the amorphous Pt/CdS interface. Such simulation is very expensive, and often can only be carried out for a few picoseconds (where at least 10-100 times that would be desired). For some organic systems (e.g., the P3HT polymer), especially when there is no chemical reaction happening, classical force field (CFF) based MD can be used. Those simulations (e.g., using the LAMMPS and/or DL_Poly codes) can be carried out for millions of atoms and carried out for hundreds of nanoseconds. The VASP, PWSCF, and PEtot codes are plane wave pseudopotential codes. They can be scaled to a thousand processors or so to calculate systems of up to a thousand atoms.

One important issue in nanosystem carrier dynamics is in the exciton dissociation. This requires the calculation of excitons in nanowires or at nanointerfaces. Due to the lack of doping in nanosystems, hence the lack of a conventional *p-n* junction, exciton dissociation often happens at the interface with a stacked (type-II) band alignment. In our example CdSe/CdS core/shell system, the electron will stay in CdS while the hole will stay in CdSe, thus forming an interface exciton. The exciton calculation can be done following the standard GW plus Bethe-Salpeter equation (BSE) approach. However, it is impossible to evaluate the GW+BSE approach directly for a thousand-atom nanosystem and approximations must be made to simplify the problem. This will involve use of the PEscan code to calculate hundreds of band edge states, constructing the electron-hole pair configurations, calculating the Coulomb and exchange integrals between these configurations, and diagonalizing the resulting BSE Hamiltonian. Various code choices exist to calculate the Coulomb and exchange integrals among the electron-hole configurations, and to diagonalize the resulting matrix.

Carrier dynamics are often strongly affected by electron-phonon interactions. This is the case for initial carrier cooling after carrier generation by the absorption of a photon. This is also true by carrier hopping transport in organic systems (e.g., the disordered P3HT), and perhaps from one nanocrystal to a nearby nanocrystal, or across the boundary between the semiconductor and the polymer (e.g., from CdSe/CdS to P3HT). The carrier cooling or possible trapping events are strongly related to possible surface states. The surface states are determined by the surface passivation and atomic structure, which can be investigated by DFT total energy calculations or MD simulations as discussed above. The phonon modes can be obtained directly by evaluation of the dynamic Hessian matrix. This requires numerical evaluation of the atomic forces caused by displacement of each atom in the system; for a thousand-atom system, this would be a few thousand self-consistent calculations. Another possible way is to approximate the Hessian matrix, e.g., by a patching scheme, or by a classical force field. After the phonon modes are calculated, the electron-phonon interaction can be calculated by evaluating the electron-electron wave function integral over the potential change of the Hamiltonian due to a particle phonon mode. This can be done using a direct DFT calculation, or by approximated schemes like CPM. After all the relevant electron-phonon coupling constants have been calculated, the phonon assisted hopping transport, or cooling process can be simulated by a master equation. For organic systems (e.g., the disordered P3HT), to yield a converged carrier mobility, a multiscale model has to be used.

Finally, some coherent or tunneling transport (e.g., for the hole in CdSe to cross the CdS barrier to reach P3HT polymer) can be calculated using the quantum transport method. We have been using our own planewave transport code PEtot_trans and the non-equilibrium Green's function code (SIESTA_trans) to do such calculations. While SIESTA_trans is based on atomic basis functions, the PEtot_trans code deploys a special algorithm that uses auxiliary periodic boundary conditions to study the transport problem. As a result, its calculation is similar to a conventional ground state supercell calculation.

9.2.1.3 HPC Requirements

Due to the existence of different methods and codes at different levels of approximations, it is difficult to say what is the exact requirement for the computing resources, although more resources will allow us to do more accurate simulations (e.g., change the simulation from classical to *ab initio* methods), and to do new type of calculations (e.g., to study electron-phonon interactions). In the foreseeable future, we should have no difficulty to use up all the computer time realistically available to us. Up to now, the problems we decided to do and the approaches we chose to use were determined by the computer time we were allocated, instead of the other way around: first decide a computational approach, then ask for the computer time. For a given scientific problem, one can easily come up with a high level computational approach to use up all the computer time available. Historically, the limited computing resources have been the motivation for the developments of many innovative algorithms and approximation methods.

As discussed in the summary, material science simulation is often an interactive process involving frequent decision points based on simulation output. As such, one of the requirements is fast turn-around time for the submitted jobs, since that will save human time, which often is more valuable. Long waiting time is equivalent to the unavailability of the tool, and it significantly reduces the usefulness of the resource. This is the single most important requirement for the current HPC systems. It will be ideal if the waiting time is similar to the program execution time. This can be achieved by changing the policies. First, instead of requiring the resources to run at 98% of its capacity, it might be better to require its use at, say, 80% capacity. Considering the waste of human time (and the amount of money to support the scientist users versus the money to support the machine), jamming the machine up to 98+% of its capacity is a bad idea. The reduction of the usage can be achieved by allocating only ~80% of the total computer time available. Currently, the computer time is probably over allocated. Another change can be the formula used to determine the queue position of a submitted job. To guarantee that the waiting time of a job is similar to the requested execution time, the formula should advance a job's position in the queue as the waiting time of this job approaches its requested execution time. The current formula prioritizes the jobs regardless of their requested execution times; either it is 5 minutes, or 12 hours.

Another requirement is to have some long running jobs and memory. We have encountered this problem for some of the GW calculations using VASP. The memory is a problem (e.g., each core requires more than 16 GB), so is the run time limit (more than 72 hours). Many of the DFT codes do not have check-pointed capability, and the runs cannot be separated into a few shorter jobs. This is particularly a problem when we use third party or commercial codes (e.g., VASP) for which we cannot modify the source and the way they are run. Usually such jobs don't need large core counts (the codes typically don't scale so well). Thus, some special unit or part of a machine with long queues will be very useful. The inability to scale to large number of processors for such runs doesn't mean it does not belong to HPC, thus, is not the responsibility of NERSC. Because these codes are part of our toolset, they are often used together with some of our other tools (codes), and it is thus highly desirable to run them on the NERSC machines, particularly so that data files can be shared amongst codes.

Today, our calculations are mostly run on the Franklin Cray XT4 machine. For the LS3DF runs, we can use up to 20,000 cores for a few hours. This is used to carry out one self-consistent calculation for a nanocrystal (e.g., a 5,000 atom CdSe/CdS core/shell structure). For a typical PEscan run, it uses 128 to 512 cores for a few hours to calculate a few states near the band edge, either for the nanocrystal or the polymers. But we might repeat such runs for many times (e.g., 10 times) for a given nanosystem or a variation of a system (e.g., changing the size or the shape). The CPM codes do not take time at all. We have used the VASP to carry out *ab initio* MD for a 700-atom interface problem. This used 1,000 cores, and many consecutive runs have been carried out with accumulated 100 hours of wall clock time for each MD simulation. Several such simulations have been performed. The PEtot code has been used to obtain a self-consistent solution of thousand-atom systems. Each of such run is performed on 1,000 cores for five or ten hours. The classical MD simulations are run using LAMMPS on a few hundred processors for a few hours.

In the next 3 years, we plan to perform more *ab initio* simulations to study the surface atomic structure and passivation of the nanocrystal, and the surface electronic states. This will require more *ab initio* MD simulations. Such simulations are quite expensive and we expect our need for computer time will increase by a factor of five at least. We also plan to study electron-phonon interactions and the related carrier dynamics in nanocrystal and mixed nanocrystal/organic systems. The computational time requirement will depend on whether the CPM is good enough to describe the change of the single-particle Hamiltonian under a phonon mode. If CPM can be used, then the computational time requirement is minimal; otherwise, thousands of DFT self-consistent calculations for thousand-atom systems are needed. The required computational time is similar, or somewhat greater, than the DFT MD simulation.

Access to 50x resources would allow us to perform time domain simulations of carrier dynamics, which will be critical for fully understanding photovoltaics and solar cells. This involves evolving the electron wave functions following their time dependent Schrodinger's equation while progressing the nuclei position using molecular dynamics. Due to the small time step (10^{-3} fs) needed for the electron, and the long simulation time (many ps) for the incoherent state collapse to happen, there is a need for large computer resources and massive parallelization. We plan to develop an approach to parallelize that simulation to speed up the time evolution. With the 50x computer resources, we would also change our *ab initio* search for the surface atomic structure. Some systematic approach would need to be developed to search for the minimum energy atomic configurations.

In general, with the 50x increase in resources, the best road forward is not just to apply our current computational approaches and codes to some larger systems or higher planewave energy cutoff, and real space grids. For example, with the conventional $O(N^3)$ scaling DFT code, that will only increase our system size by a factor of four, which certainly is not very impressive. Rather, new approaches (e.g., linear scaling methods) need to be developed that are more suitable to the new level of resources available (certain computational schemes will only make sense at certain level of computing powers), and new scientific problems could be simulated with the increased computer power. All these need investment in algorithm and code development.

9.2.1.4 Computational and Storage Requirements Summary

PI: Lin-Wang Wang

	Current (2009)	**In 2013**
Computational Hours	1.6 M	~10 M
Parallel Concurrency	Up to 20 K	16 to 100 K
Wall Hours per Run	20 min to 100 hours	20 min to 100 hours
Aggregate Memory	20 TB	50 TB
Memory per Core	2 GB (could be 8GB)	2 GB (could be > 32 GB)
I/O per Run	20 GB	20 GB
On-Line Storage Needed	1 TB	4 TB
Data Transfer	40 GB	40 GB
Archival Storage	30 GB	500 GB

The small jobs are as important as the large jobs.

PI: Hai-Ping Cheng

	Current (2009)	**In 2013**
Computational Hours	1 M	~10 M
Parallel Concurrency	64-512	5 to 20 K
Wall Hours per Run		5 K
Aggregate Memory	500 GB	1 TB
Memory per Core	2 GB	4 GB
I/O per Run	100 GB	1 TB
On-Line Storage Needed	1 GB	3 GB
Data Transfer	2 TB / month	3 TB / month
Archival Storage	30 GB	500 GB

9.2.1.5 Support Services and Software

We usually do our own visualization off the main computer. We do only minimal graphics (e.g., gnuplot) online to analyze our data. But it will be very valuable if we can do more visualization on the computer on which we do our simulations, so everything is done in the same environment and no large potentially time-consuming data transfer is required.

The "scratch" directory space is heavily used, and most recent data are stored there. We found that the NERSC policy of no frequent data cleansing on the scratch directory is a very valuable one. That provides a large working space (desk) which is very useful. We made a heavy use of the HPSS archive system.

We certainly require the commonly used math libraries, including LAPACK, ScaLAPACK, arpack, FFT, etc. We also take the advantage of the compiled third party codes in the NERSC platform, e.g., the VASP code and the LAMMPS code. It will be advantageous to us if NERSC can organize a software repository contributed from NERSC users on a volunteer basis, allowing the whole NERSC community to share codes and capabilities. A forum can be developed related to such a repository.

9.2.1.6 Emerging HPC Architectures and Programming Models

We are currently using fortran90 and MPI as our programming and parallelization model. We do plan to experiment with OpenMP and the multicore environment. We are very interested in testing DFT codes (e.g. PEtot) on the GPU, but have not done that due to the lack of human resources and funding. We advocate enthusiastically, that BES and NERSC support jointly new major code development projects. Such effort in Switzerland, for example, is already underway to revolutionize CMPD and BigDFT (both DFT based electronic structure with MD) that combined OpenMP and MPI, and CUDA such that these code can take advantage of GPU and run parallel jobs on ~10^6 processors.

9.2.2 Computational Resources for the Nanomaterials Theory Institute at the Center for Nanophase Materials Sciences

Principal Investigator: Paul Kent, Oak Ridge National Laboratory
NERSC Repo m526

9.2.2.1 Project Summary & Scientific Objectives for the Next 5 Years

This project performs breakthrough calculations into the behavior and properties of nanoscale systems, ranging from new energy efficient nanoscale catalysts to simulations of DNA used for molecular electronics. The systems under study are increasingly realistic, in most cases incorporating large length scales to properly simulate experimental conditions. A variety of methods and implementations are applied to a variety of systems of different sizes etc.: there is no single "target problem" for which we need better accuracy/increased length/timescale. Instead we need progress in all areas. The calculations are therefore challenging, requiring extensive use of high performance computing facilities at NERSC.

Investigations supported by this project include:

- The Reactivity and Stability of Nanoparticle Alloys for the Electrocatalytic Reduction of Oxygen
- Catalytic reforming and combustion of hydrocarbons on nanoparticles
- Quantum transport through oxide junctions
- Computational determination of grain boundary resistance
- Van der Waals Interactions
- DNA derivatives
- Oxide surfaces for fuel cell and catalytic applications
- Structure-property relationships for the Mo-V-Te-Nb/Ta-O mixed metal oxides catalysts

9.2.2.2 Current HPC Usage and Methods

The methods used are primarily atomistic simulation techniques: classical molecular dynamics (e.g. the LAMMPS implementation), density functional theory (e.g. the VASP implementation of the projector-augmented wave plane-wave pseudopotential method), and quantum chemistry (e.g. Gaussian, NWChem).

The concurrency requirements are broad. For example, in nanoscience it is often better to investigate a variety of systems than focus on a single system. Often, the throughput for all the systems considered is good, time to solution is not a problem, and scaling requirements are not strict. However, in the case of molecular dynamics (classical or *ab initio*), the time to solution is absolutely critical, and parallel efficiency is routinely sacrificed to obtain a solution more quickly (e.g., in days, not weeks, etc.). The "one minute per ionic step" typically achievable for mid-sized systems using VASP places a limit of a few 10ps on achievable timescale. Only new architectures or dramatically better parallelization techniques are likely to significantly increase this.

It is important to note that since most of the computation is performed in support of user projects at the Center for Nanophase Materials Sciences, we are able to choose the appropriate implementation for a given calculation. Hence we broadly monitor implementations of the various methods. For some problems, NAMD might be more appropriate than LAMMPS, and VASP is not necessarily the best choice for periodic DFT calculations.

9.2.2.3 HPC Requirements

Many of the calculations require days or weeks of computation. This is achieved via checkpointing. Unfortunately, not all methods of all codes implement checkpointing well enough. For example, for one project we improved the checkpointing of frozen phonon calculations in VASP so that large supercells could be easily studied.

For long-lived calculations, particularly molecular dynamics, sustained high throughput is important. (We have found that users generally do not like restarting calculations, using checkpoints etc., and often cite the long queue limits on some NSF resources as particularly attractive compared to DOE sites. Better automation and education is required.)

Nearly all of our calculations are severely compute hour constrained. For example, it has been clear for several years that most of our DFT calculations should be done using so-called hybrid functionals. Extensive testing and benchmarking (vs experiment and quantum chemical approaches) shows these functionals to be more accurate. Unfortunately, hybrid functionals are formally N4 scaling compared to the conventional N3 scaling of local density approaches. In practice these calculations are typically 10-100 times more expensive than our current methods. The errors are physically significant, particularly for delicate systems such as catalysts.

9.2.2.4 Computational and Storage Requirements Summary

	Current (2009)	**In 2013**
Computational Hours	5.3 M	50 M
Parallel Concurrency	16 – 512	128 – 32K
Wall Hours per Run	As long as possible	As long as possible
Aggregate Memory		
Memory per Core	2 GB	All available (planning for 512 – 1024 MB)
I/O per Run	100 GB	1 TB
On-Line Storage Needed	500 GB	1 TB
Data Transfer	100 GB/user/year	100 GB/user/year
Archival Storage	35 TB	50 TB

9.2.2.5 Support Services and Software

Methods and implementations are continuing to evolve. Memory per compute unit is going down. Out of node bandwidth per compute unit is going down. Broadly speaking, we need software that can run reliably on current and future systems delivering good time to solution. We are likely to need particular help with threading applications (OpenMP etc.) and in profiling them.

9.2.2.6 Emerging HPC Architectures and Programming Models

Our strategy involves:

(i) developing new implementations and new algorithms on new architectures for methods and codes that do not have critical mass, e.g., for novel classical molecular dynamics potentials and new quantum Monte Carlo methods;

(ii) watching and learning from development in Europe, which has very good people and is ahead of the U.S. in terms of practical ability for researchers to develop electronic structure codes;

(iii) sharing our experiences with other users, with HP centers, and with code authors if we are not the main developer; and

(iv) most importantly, developing informal performance models for codes and methods as applied to our actual research problems, as opposed to other people's research problems. This allows us to assess suitability of new architectures and programming models for actual research tasks.

9.2.3 Photon Spectroscopy and Light-Matter Interactions

Principal Investigators: Thomas Devereaux, SLAC National Accelerator Laboratory and Jeffrey Neaton, Lawrence Berkeley National Laboratory
NERSC Repos: m974, m772, and m387

9.2.3.1 Project Summary & Scientific Objectives for the Next 5 Years

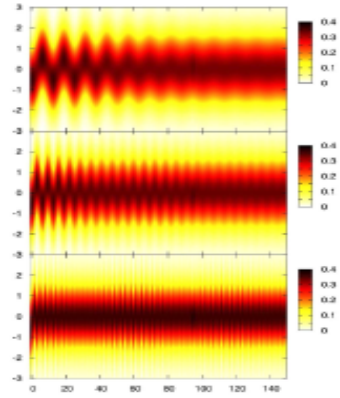


Figure 9-3. Computed photoemission spectra.

Light-matter interactions are central to a variety of materials phenomena relevant to energy conversion. Photon spectroscopies are essential probes of our understanding of these interactions. They provide much-needed characterization of novel materials, including nanostructures and soft matter, and comprise some of the most important tools available for observation and discovery. Advances in third-generation light sources and detectors have provided detailed information pertaining to the BESAC five grand challenge questions. A large volume of data will continue to accrue at ever accelerating rates with the upgrades at current facilities, such as SPEAR III at Stanford Synchrotron Light Source, the MERLIN beamline at Lawrence Berkeley National Labs (LBNL), and the NSLS II at Brookhaven, as well as with the commissioning of new time-resolved spectroscopies at large-scale facilities, such as the Linac Coherent Light Source (LCLS) at SLAC National Accelerator Laboratory. In concert with the rapid advances in experimental capabilities, continued development and application of theory and simulation for treating light-matter interactions are crucial to provide meaning to and understanding of experimental results, to predict and provide insight into new physics and phenomena, and to guide the development of new principles and to chart new directions.

At LCLS, the short duration of hard x-ray pulses has opened many new possibilities for probing matter at extremely small length and time scales. Some of the scientific motivators are the ability to drive chemical transformations by controlled optical or infrared pulses and derive an understanding the atomic and electronic transformation with x-rays. This entails capturing, with snapshots on the femtosecond timescale, the making and breaking of chemical bonds and the crucial transition-state intermediates in chemical reactions. Also, time-domain imaging and spectroscopy will help derive an understanding of the origins of nanoscale charge and spin order and their dynamics in correlated materials through high-resolution energy- and time-dependent x-ray spectroscopies.

During the past few years, the Photon Spectroscopies and Light-Matter Interactions (PSLMI) team has begun establishing activities in the Bay Area National Lab settings, building up teams of investigators conducting state-of-the-art simulations of photon-related spectroscopies. The codes used by these investigators are based on an extremely diverse but complimentary set of theoretical methods, including determinant quantum

Monte Carlo (DQMC), exact diagonalizations (ED), pump-probe dynamical mean field theory (DMFT), and density functional theory (DFT) approaches both for ground-state properties, and, using many-body perturbation theory (within the GW approximation and using a Bethe-Salpeter equation approach) or time-dependent DFT, for excited-state properties and charge dynamics.

One set of our scientific objectives includes developing codes and algorithms that focus on ultrafast nonequilibrium phenomena in strongly correlated materials. We want to understand how electron-electron correlations and large driving fields can change the way materials behave and how this behavior can provide insight into the microscopic theories for such complex systems and their emergent strongly correlated behavior. Another important class of systems is nanostructures, which are often distinguished by their large surface-to-volume ratios that, upon integration into devices, can lead to a high density of nanoscale interfaces. However, the relationship between structure and electronic properties of interfaces, and device function, is not yet well understood. This is particularly the case in the context of solar energy applications, where interfaces are central to device function and knowledge of electronic excited states – inaccessible with ground-state DFT – and their time evolution are central to evaluating and understanding the efficacy of newly-synthesized nanomaterials. These systems are believed to have the potential for wide applications within energy science due to their high tunability and the variation in properties as a response to external perturbations or driving forces. Beyond improving our understanding of strongly correlated electrons, excited-states, and nonequilibrium charge dynamics in new materials, the results obtained in this research program have direct consequences on our global understanding of the behavior of any dynamical system with strong interactions, and is thus applicable to all of the grand challenge questions.

9.2.3.2 Current HPC Usage and Methods

For DQMC, the code architecture makes it almost perfectly parallel under weak scaling. The majority of our production runs at NERSC concentrate on providing adequate data for post-processing analytic continuation. This involves spawning a large number of independent Markov chains of the same length. However, this code typically does not run efficiently on more than 128 or 256 processors due to "warm-up" overhead. We use two variations of the same exact diagonalization (ED) code: one simply calling the PARPACK sparse matrix libraries, based on parallel Arnoldi iterative methods, and a second that also includes calls to ParMETIS (Parallel Graph Partitioning and Fill-reducing Matrix Ordering), an MPI-based parallel library that implements a variety of algorithms for partitioning and repartitioning unstructured graphs. The number of rows or columns of the sparse matrix that can fit in memory limit the number of cores that the code can use. This is dependent on the sparsity of the matrix for any single problem. The vectors storing the eigenfunctions of the first N eigenstates of the sparse matrix, each one the length of the Hilbert space dimension, must also fit in the memory allocated to a single core.

Using both PARPACK and ParMETIS, the code demonstrates superlinear behavior to tens of thousands of processors. The matrix-vector product used in the iterative Arnoldi process is a key component that has an important impact on overall performance of the program. The Hamiltonian matrix for this problem is very large (on the order of 108x108) and also sparse (approximately 500 nonzero elements per column or row); therefore, it is more efficient to calculate the matrix-vector product using graph partitioning provided by the ParMETIS libraries.

The basic algorithm for our pump-probe DMFT-based codes is an iterative solution of a set of coupled nonlinear equations. Starting with an initial guess for the self-energy, one calculates the local Green's function on the lattice, then extracts the effective medium by using the local Dyson equation, then solves the impurity problem in that effective medium for the impurity Green's function, and finally extracts the impurity self-energy via Dyson's equation. This loop is iterated until it converges. Because of its iterative nature, we cannot tell *a priori* how many iterations will be needed to achieve convergence (although experience tells us it is usually between 10 and 200 iterations with more iterations needed for larger interaction strengths or smaller fields). Within this basic structure, there are two critical elements. The first is the determination of the local Green's function. For nonequilibrium problems, this is achieved via a two-dimensional matrix-valued quadrature routine where each quadrature point requires one matrix inversion and two matrix multiplications to determine the value of the integrand. Standard LAPACK and BLAS routines are used because the matrices are general complex dense matrices and the codes are well optimized.

For studies of optical and photoemission spectroscopies in the visible-to-ultraviolet, we use a DFT-based excited-state formalism, many-body perturbation theory (MBPT) within the GW approximation, treating electron-hole interactions within the Bethe-Salpeter equation (BSE) approach. These methods are heavily reliant upon linear algebraic operations, such as matrix multiplies, diagonalizations, and inversions, as well as fast Fourier transforms (FFT), and thus our parallel implementations require heavy communication between nodes. The starting point for these excited-state methods is a large set of Kohn-Sham wavefunctions, computed within the pseudopotential approximation and expanded in a plane-wave basis set. Many plane-waves are often needed for useful comparison with experiment. A significant bottleneck for our plane-wave pseudopotential excited-state calculations is a required sum over the large number of unoccupied states. The number of occupied states grows linearly with the volume of the system supercell, which can be large for systems of interest, containing many atoms and vacuum. For large systems, generation of this unoccupied subspace, which is reliant upon iterative diagonalization techniques (usually via Conjugate Gradient), will become prohibitive. Another bottleneck is the formation and inversion of the dielectric matrix, which is an approximately N^4 operation, where N is the number of basis functions. Finally, the solution to the BSE is N^6 but with a small prefactor, allowing treatment of electron-hole interactions for systems of roughly 100 atoms at present. Alternative approximations and approaches, some of which avoid unoccupied states, are now being proposed, but the community is only in early stages of evaluating their efficiency and validating their efficacy.

9.2.3.3 HPC Requirements

Our production runs for DQMC and pump-probe typically run between 256 and 512 cores, while ED runs over a few thousand on Franklin. Each set of codes runs for 48 wallclock hours. For the pump-probe and DQMC codes, wallclock limits are not much of an issue since the codes give configuration outputs that can be used to restart jobs if statistics are not sufficient, for example. The ED code would benefit from larger RAM per CPU to increase the size of the overall vector that can be read into memory and manipulated. The DQMC and pump-probe DMFT runs do not typically produce large output files, while ED codes generally produce 0.01 TB per run which is later transferred to local machines for post-processing.

Production runs for MBPT with the GW/BSE formalism can run up to a few thousand cores for several hours. We are ultimately bound by I/O, large distributed matrix multiplications and inversions, and FFTs. Key inputs are the Kohn-Sham wavefunctions for many unoccupied states for 1-300 atoms, requiring 100-1,500 processors for up to 12 hours or more. Wallclock limits can be limiting for certain systems. Memory requirements are significant, up to 2GB/core; in some cases, more would be desirable, although advances in parallelization and scaling to more cores may reduce memory requirements somewhat in the future. Long queue times for jobs that use less than a few thousand cores and that request more than 30 total minutes can inhibit progress. A quick turnaround for tens of simultaneous jobs requiring 1,000 processors and a few hours per CPU would greatly enhance productivity. For the largest simulations, the 0.5 TB limit on scratch presents a barrier.

9.2.3.4 Access to 50x Resources

It is important to note that for many-body simulations, the complexity of the problems to be addressed will be resilient to efforts to simply scale up present computational approaches. Progress in this area will likely only occur if current algorithms are substantially modified. There are two important examples. For the case of Quantum Monte Carlo codes, which currently can be run over 1,000's of CPUs, a fundamental limitation involves the necessity to perform warm-up passes in each Markov process before measurements can be made, which is one of the reasons why QMC codes for photon spectroscopies do not show strong scaling. More relevantly, in the case of exact diagonalizations, since the size of the Hilbert space grows exponentially large, an increase of a factor of 50 may only increase the size of the cluster by one site. Finally, for DMFT-based codes for time-domain spectroscopies, the amount of data passing will present bottlenecks towards strong scaling.

In the short term, access to larger resources would allow us to perform many more simultaneous jobs and enhance our throughput. A 50x increase in the allocated CPU time and storage, rapidly and easily accessible with small several-thousand processor jobs of moderate length (3-6 hours), would allow us to validate the level of theory required to handle a variety of different material classes for larger systems and at higher levels of theory, including those most directly relevant to contemporary experiments in solar energy conversion. To reiterate, higher throughput for a larger number of medium-sized jobs would enhance impact significantly, while driving the development of more efficient and parallel methods.

9.2.3.5 Computational and Storage Requirements Summary

PI: Thomas Devereaux

	Current (2009)	**In 2013**
Computational Hours	4 M (700K at NERSC)	8 M
Parallel Concurrency	0.5 K – 10 K	1K – 20 K
Wall Hours per Run	48	72
Aggregate Memory	2 TB	10 TB
Memory per Core	2 GB	4 GB
I/O per Run	10 GB	20 GB
On-Line Storage Needed	1 TB	2 TB
Data Transfer	50 GB/week	200 GB/week
Archival Storage	1 GB	14 GB
CXI Detector Data @LCLS	48 TB	10 PB

PI: Jeffrey Neaton

	Current (2009)	**In 2013**
Computational Hours	5 M (2.6 at NERSC)	15 M
Parallel Concurrency	1,000	5,000
Wall Hours per Run	6	6
Aggregate Memory	200 GB	1 TB
Memory per Core	1 GB	2 GB
I/O per Run	20 GB	10 GB
On-Line Storage Needed	200 MB	10 GB
Data Transfer	15 GB/month	100 GB/month
Archival Storage	100 MB	10 GB

9.2.3.6 Support Services and Software

We need LAPACK, ScaLAPACK, and FFTW or other scalable FFT implementations for DFT and electron transfer calculations. The Multi-band Hubbard Exact Diagonalization methods require PARPACK eigensolvers plus ParMETIS for matrix reordering and partitioning. We would welcome NERSC performance analysis and improvement assistance for the version of the ED code without ParMETIS library calls where a significant fraction of time and poor scaling beyond a few hundred processors appears to result from the message passing associated with the distributed matrix-vector multiply.

9.2.3.7 Emerging HPC Architectures and Programming Models

A cost-effective, high-throughput, low-latency alternative to conventional architectures is GPU computing. It provides the added benefits of smaller physical and carbon footprint, ideal for deskside, high performance supercomputing with a small number of local users. We are starting a research program to understand how to effectively employ GPUs. GPUs would be useful in many aspects of our calculations, enabling, for example, storing and swapping configurations to limit the warm-up overhead in DQMC simulations, more memory per core to handle potentially larger vector sizes for ED, and carrying out simple BLAS and LAPACK matrix inversions and multiplications for pump-probe DMFT. We plan to purchase a local GPU cluster based on Nvidia Fermi, and would be interested in similar development clusters at NERSC if they become available.

9.2.3.8 Data Collection and Storage at LCLS and Relation to NERSC

The LCLS at Stanford is the world's first hard x-ray free electron laser. The pulses of x-ray laser light from LCLS will enable frontier new science in areas that include discovering and probing new states of matter, understanding and following chemical reactions and biological processes in real time, imaging chemical and structural properties of materials on the nanoscale, and imaging non-crystalline biological materials at atomic resolution.

The LCLS bases its technique on the ability to capture many 2-D diffraction patterns of a given target sample from many different angular orientations, and then the numerical reconstruction of the 3-D structure from the patterns. This can be done at picosecond time intervals and can be closely monitored and accessed in pump-probe type of explorations. The need to build up a large enough signal and to obtain full 3-D information may require about 10^5 - 10^6 diffraction patterns to be collected from a series of identical particles, corresponding to tens of terabytes of data. This volume of data could be collected in less than a day's operation and computing and data handling resources must be able to keep pace with this generation rate. In addition to the process of aligning different spectra, data processing based on iterative phase-retrieval methods from the over-sampled diffraction data set will require developments in theory, software and algorithms to handle the rapid flow of data.

While currently there is no plan of how data will be collected and shared with a wider community, it is worthwhile to mention in the context of NERSC in terms of potentially large flows of data structures from SLAC and other facilities (such as the planned FEL at LBNL and new light source at NSLS2 at Brookhaven).

10 Chemical Sciences

10.1 BES Chemical Sciences Overview

Mark Pedersen, Program Manager
Computational and Theoretical Chemistry
Office of Basic Energy Sciences

The BES Chemical Sciences, Geosciences, and Energy Biosciences Division supports research that explores fundamental aspects of chemical reactivity and energy transduction over an enormous range of scale and complexity. Phenomena are studied over spatial scales from the sub-nanometer, as defined by the structure of atoms and molecules, to kilometers, appropriate to the behavior of subsurface geological structures, and over time scales defined by the motions of electrons in atoms, attoseconds (10^{-18} seconds), to millennia over which geological change must be understood.

At the heart of this research lies the quest to understand and control chemical reactions and the transformation of energy at the molecular scale in systems ranging from simple atoms and molecules, to active catalysts, to complex biochemical or geochemical moieties. At the most fundamental level, the development and understanding of the quantum mechanical behavior of electrons, atoms, and molecules in the 20th century has now evolved into the ability to control and direct such behavior to achieve desired results, such as the optimal conversion of solar energy into electronic excitation in molecular chromophores or into the creation of multiple charge carriers in nanoscale semiconductors.

This Division also seeks to extend this era of 21st century control science to include the capability to tailor chemical transformations with atomic and molecular precision. Here, the goal is fully predictive capability for larger, more complex chemical systems, such as interfacial catalysis, at the same level of detail now known for simple molecular systems.

Finally, this Division seeks ultimately to extend a molecular level understanding and control to the emergent and highly non-equilibrium behavior of biological and geological systems through the application of modern experimental and computational tools.

The Division supports basic research that underpins a broad range of energy technologies. Research in chemistry has led to advances such as efficient combustion systems with reduced emissions of pollutants; new solar photoconversion processes; improved catalysts for the production of fuels and chemicals; and better separations and analytical methods for applications in energy processes, environmental remediation, and waste management. Research in geosciences results in advanced monitoring and measurement techniques for reservoir definition and an understanding of the fluid dynamics of complex fluids through porous and fractured subsurface rock. Research in the molecular and biochemical nature of photosynthesis aids the development of solar photo-energy conversion.

The Division also plays a major role in enabling the nanoscale revolution. The importance of nanoscience to future energy technologies is clearly reflected by the fact that all of the elementary steps of energy conversion (e.g., charge transfer, molecular rearrangement, and chemical reactions) take place on the nanoscale.

10.2 Chemical Sciences Case Studies

10.2.1 Quantum Monte Carlo for the Electronic Structure of Molecules

Principal Investigator: William A. Lester, Jr[1,2]
Contributors: Brian Austin[1,2], Dmitry Zubarev[1], Dominik Domin[1,2], Jarrod McClean[1], Jinhua Wang[1,3]
NERSC Repo: m208

[1]Department of Chemistry, University of California at Berkeley
[2]Chemical Sciences Division, Lawrence Berkeley National Laboratory
[3]State Key Laboratory of Superhard Materials, Jilin University

10.2.1.1 Summary and Scientific Objectives

This project uses Quantum Monte Carlo (QMC) to compute properties of chemical systems that are of major importance to basic energy sciences. QMC methods are computationally intensive, but their use is warranted when less demanding methods such as density functional theory lack the precision required to answer the question at hand or when discrepancies among experiments or other *ab initio* theories leave the issue unresolved. Unlike other quantum chemical methods, which depend strongly on the single-particle basis set and the extent of the many-particle representation of the wave function, the accuracy of QMC has only a weak and indirect dependence on the trial wave function. This allows QMC to resolve questions that cannot be adequately addressed otherwise. Projects currently pursued by this group include calculation of the potential energy curve for the interaction between lithium atoms and graphene sheets, computation of the O-H bond dissociation energy in phenol, and prediction of the S_0-S_1 excitation energy of a retinal protonated Schiff base.

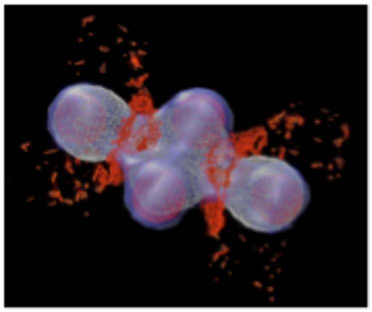


Figure 10-1 Electron Pair Localization Functionin a Lithium cluster.

As more computing resources become available, the systems that we study will increase in both number and size. The *O*(N) algorithms for wave function evaluation that have been developed in this group will leverage the expanded resources so that, within the next five years, we will be able to examine systems in the 250-350 atom range. Second, our QMC calculations may be extended to include both electronic and nuclear motion so that thermochemical properties can be computed immediately at the QMC level without requiring the Born-Oppenheimer approximation or the use of lower-level theories to determine molecular geometries or vibrational modes. In addition, we plan to increase the value of our QMC calculations by developing new methods of analyzing MC 'trajectories' using tools such as the electron-pair localization function (EPLF) shown in

Figure 1 to understand and explore exotic binding motifs or properties of electron correlation.

10.2.1.2 Methods of Solution

Our calculations use primarily the diffusion Monte Carlo (DMC) variant of QMC to determine the electronic structure of molecular systems. DMC is rooted in the time-dependent Schrodinger equation; solutions to this equation indicate that the coefficients of excited state wave functions decrease exponentially in imaginary time. An isomorphism between Schrodinger's equation and the diffusion equation allows the imaginary-time evolution of the wavefunction to be simulated by a random walk so that after sufficient simulation time, random walkers sample only the ground state. An estimate of the eigenvalue can then be obtained from Monte Carlo integration. The efficiency of the MC integration is improved by using a trial wave function as a basis for importance sampling.

Parallelization of the DMC algorithm is trivial because each walker's movement is independent of the others. Individual cores process only the walkers that have been distributed to them, so communication is needed only for averaging and occasional load balancing. The simplicity of this mode of parallelism allows excellent parallel performance using MPI, even for calculations involving tens of thousands of cores.

Nearly all of the DMC compute time is used to evaluate the trial wave function and its derivatives. Our trial wave functions use a linear combination of Slater determinants to describe the essential features of the electronic structure and to include static correlation effects. The molecular orbitals (MOs) used to compute the Slater determinant are computed from linear combinations of atom-centered basis functions. Either Gaussian or Slater-type basis functions may be used, depending on whether core electrons are replaced by pseudo-potentials. Short-range dynamical correlation is included by multiplying the determinants by a three-body Jastrow correlation function of the Schmidt-Moskowitz / Boys-Handy form.

Wavefunction parameters are optimized by minimizing the energy of the wavefunction or the variance of its local energies. The preferred approach minimizes the energy using the recently developed linearized wave function approximation, which involves diagonalization of a (potentially large) matrix of wave function derivatives. Alternatively, Conjugate Gradient methods can be used to minimize variance for a finite sample of walkers.

The primary code used for our group's QMC calculations is Zori. We have developed the Zori code with an emphasis on linear scaling algorithms that enable the treatment of larger molecules than was previously possible. Because the MC approach is inherently parallel, we have focused on serial efficiency to improve the performance of our code. Specifically, we use a combination of matrix compression and BLAS libraries to make the MO and correlation function evaluation routines linear scaling while maintaining high flop rates.

10.2.1.3 HPC Requirements

The largest calculations that we have performed to-date have involved 314 electrons. Calculations involving 50-75 electrons are routine. Jobs in this size range use less than 200 MB per core and require 1-2 GB of storage for checkpoint files. Total I/O for such jobs is roughly 50 GB.

Typical jobs require a small number of runs on 4,000 cores for 2-6 hours. This is significantly lower than our code can manage (Zori's scalability has been demonstrated using up to 24,576 cores on Franklin), but it provides a sensible balance between using a large number of walkers (cores) to minimize the statistical error per step and taking enough steps (wall-time) to ensure equilibration and proper sampling. Jobs in this size range also benefit from low charge factors and reasonable queue turnaround times.

The linear scaling algorithms that we have developed in the last 5 years have already enabled the treatment of larger systems than previously possible. Wavefunction optimization and ansatz improvement will be essential to pushing this limit higher. The use of more accurate trial wave functions will improve the efficiency of QMC calculations in a variety of ways. The reduced variance of the integrand decreases the number of points that must be sampled in order to reduce the error estimate to an acceptable level. The benefits of this effect are amplified by the reduced fluctuations of the walker weights. Improved wavefunctions will also reduce both the time-step error and the fixed-node error in DMC calculations. Greater understanding of many-body correlation effects is an auxiliary benefit of studying optimized wavefunctions.

A 50-fold increase in HPC resources would allow the simulation of systems with as many as 800-1,600 electrons. If effective core potentials are used, this is sufficient to model protein reaction centers or small nanoparticles using fully correlated electronic structure theory. Larger systems could be studied if the environment surrounding the QMC simulation is included via molecular mechanics (MM); the theory and code needed for such QMC/MM calculations are in a late stage of development in this group. Advances in this direction will require a roughly 50-fold increase in both CPU hours and storage. With minor code changes, memory requirements can be kept below 1-2 GB per core. Queuing and charge factor policies that encourage the highest possible concurrencies would also facilitate this research.

Additionally, increased HPC resources would make possible more accurate treatments of small molecules. Methods for optimizing molecular geometries and computing vibrational frequencies are available for many *ab initio* methods, but are missing from the QMC toolkit. With a larger HPC allocation we could begin to examine molecular geometries at the DMC level. In particular, DMC can be used without the Born-Oppenheimer approximation to sample the complete electron-nuclear wave function with full treatment of vibronic coupling and anharmonic vibrations. The ease of extending

QMC methods to both electronic and nuclear motion on equal footing is a unique benefit of the QMC formalism. A brute force approach to calculations of this type would require order of magnitude increases in CPU hours and storage, but minimal increases in memory per core. In addition to expanded computational resources, theoretical and algorithmic advances will facilitate this research. The efficiency of the method will hinge on the availability of accurate vibronic wave functions. Some exploration will be required to identify suitable forms for these wave functions and new code will be required for their evaluation and optimization. A major payoff for this approach would be the increased accuracy of computed thermochemical data.

10.2.1.4 Computational and Storage Requirements Summary

	Current	**Next 3-5 Years**
Computational Hours	4.5 M	100 M
Parallel Concurrency	4 K	32 K
Wall Hours per Run	2-6	15
Aggregate Memory	400 GB	32 TB
Memory per Core	0.1 GB	1 GB
I/O per Run	20 GB	2 TB
On-Line Storage Needed	1 GB	2 TB
Data Transfer	5 GB / month	50 GB / month
Archival Storage	100 MB	2 TB

10.2.1.5 Support Services and Software

Zori makes extensive use of the random number generators and linear algebra routines provided by GNU Scientific Library (GSL). The GSL is designed to take advantage of optimized BLAS routines (such as those included in Cray's LibSci) when these tools are available. The current version of Zori uses MPI for parallel communication and parallel I/O. All of Zori's input files are written in the XML format and parsing these files requires the libxml2 library.

10.2.1.6 Emerging HPC Architectures and Programming Models

The transparency of the high level parallelism inherent in QMC methods will provide a great deal of flexibility when modifying our code to take advantage of new HPC architectures. The loose coupling between nodes makes it possible to take advantage of cloud computing with little code modification. We also expect that the rate limiting steps in the evaluation of the wave function (evaluation of basis functions, molecular orbitals and correlation functions) will transfer well to GPUs. Several groups have reported order-of-magnitude speedup when their QMC codes were ported to GPUs and we intend to port Zori to GPUs as well.

10.2.2 Quantum Monte Carlo Studies of Solids

Principal Investigator: Burkhard Militzer, University of California, Berkeley
Contributors: Ken Esler[2], Kevin Driver[3], Ronald E. Cohen[4], Jeongnim Kim[2], David Ceperley[2]

[1]University of California, Berkeley
[2]University of Illinois at Urbana-Champaign
[3]Ohio State University
[4]Carnegie Institution of Washington
NERSC Repo: m744

10.2.2.1 Summary and Scientific Objectives

The goal of quantum Monte Carlo (QMC) calculations is to provide highly accurate descriptions of the ground-state properties of real materials with applications in energy science, physics, chemistry, geophysics, and materials science. With recent improvements in QMC algorithms and an increasing amount of CPU time available, we are now able to study solid materials of increasing complexity. Calculations with 512 valence electrons can be done and require between 50,000 and 200,000 core hours depending on the nature of the material and the level of accuracy that is required for the application. Calculations with 1,024 electrons and more have been done also. We expect this number to increase further because QMC simulations has been shown to run 10 to 15 faster on a GPU than on a CPU as we will be described in the last section of this case study.

The application and development of QMC methods is focused on systems where other computational methods have great difficulties. For systems with less than ~20 electrons, quantum chemistry methods work very well but they cannot be applied to larger systems, including all solids, because the CPU time requirements scale as $N^{6\ldots7}$. Density functional (DFT) methods provide remarkably good results for many classes of materials, given the comparatively small amount of CPU time invested. However, DFT can rarely predict the density of geo-materials with an error of less than 3%. In the case of silica (SiO_2), it predicts the wrong structure, stishovite, to be the ground state when the local density approximation is used. The later developed generalized gradient approximation correctly predicts the quartz structure to be the ground state but at the same time, it severally underestimates the bulk modulus of the material as shown in Figure 10-2.

So in DFT, one has to make a difficult choice of which functional to use. This is typically done by comparing with experiments but often, as in the case of silica, no perfect solution is available. Recent QMC calculations by Driver *et al.* (2010) in Figure 10-2 show that these inaccuracies are related to exchange and correlation effects that are missing in DFT but can be accurately described with QMC. The goal of this particular QMC study was to predict state of silica in Earth's mantle and characterize important phase transitions (Figure 8-3).

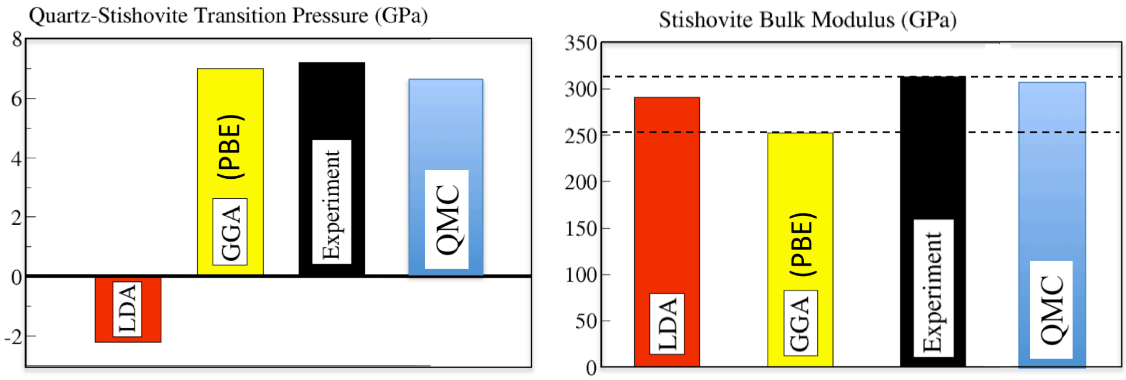


Figure 10-2. Left: Predictions of the quartz-to-stishovite transition pressure from different methods. LDA predicts the wrong ground state (stishovite). Right: Different calculation of the bulk modulus of stishovite SiO$_2$ phase. GGA underestimates it by 20%. Both QMC predictions are in good agreement with experiment. Adapted from K. P. Driver, R. E. Cohen, Z. Wu, B. Militzer, P. López Ríos, M. D. Towler, R. J. Needs, and J. W. Wilkins "Quantum Monte Carlo for minerals at high pressures: Phase stability, equations of state, and elasticity of silica," Proc. Nat. Acad. Sci. (2010).

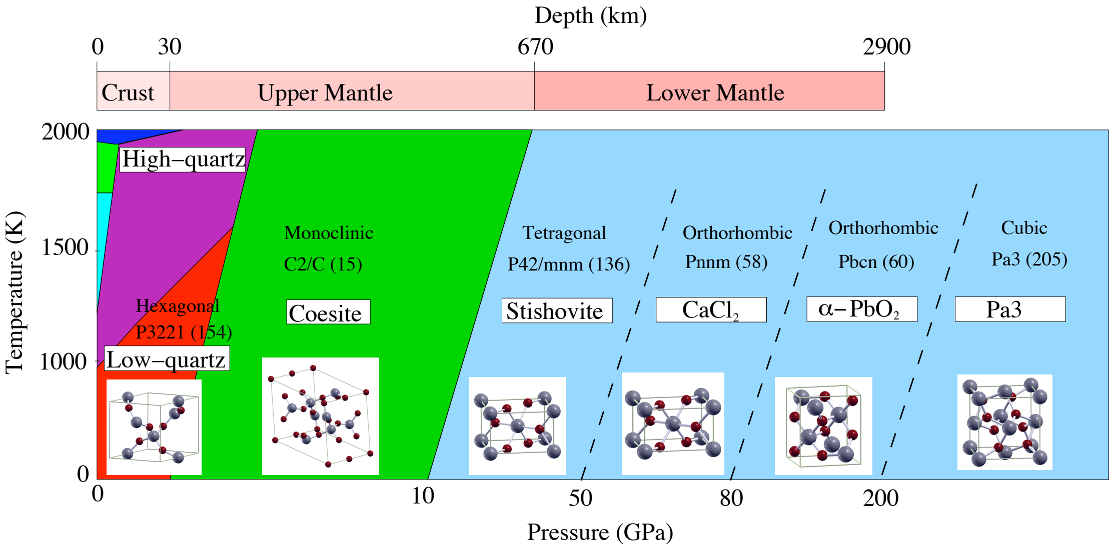


Figure 10-3. Phase diagram of SiO$_2$ silica with pressures in the Earth's mantle indicated in the upper scale.

10.2.2.2 Methods of Solution

The goal of *ab initio* electronic structure computation is to predict the properties of real materials with computational means alone rather than relying on input from experiment. Most of this task could be accomplished if we were able to exactly solve the Schrödinger equation for a system of electrons moving in a field of static nuclei. While exact solutions have been intractable for systems with more than a few electrons, we have a number of approximate techniques available that work well but differ significantly in accuracy and CPU requirements.

QMC is a stochastic method that treats electronic exchange and correlation effects explicitly and is therefore able to provide a much more accurate description of the ground-state properties of materials than density functional theory (DFT). DFT uses approximate functionals that cannot easily be improved in a systematic way. However, DFT has been remarkably successful in providing fairly accurate materials properties for the CPU time invested, and has contributed to the solution of many problems in different areas of science.

It should be pointed out that all DFT functionals are based on the seminal QMC calculations of the homogenous electron gas by Ceperley and Alder (1980). DFT combined with the local density approximation (LDA) describes the ground state of materials that naturally have a very inhomogeneous electronic density by integrating of all space where every parcel is treated like the homogeneous electron gas. Improvements have been made beyond LDA that incorporate density gradients or treat exchange effects more accurately. It is no longer necessary, however, to make this mean-field approximation, since we have arrived at the point where we can perform the QMC calculations directly for the material under consideration rather than relying on the mean-field calculations or approximate functionals. Typically, such QMC calculations use rather large supercells with many electrons in order to capture electronic correlation effects instead of relying on the Bloch approximation for non-interacting electrons.

10.2.2.3 HPC Requirements

With 50 times more CPU time, we could compute the phase diagrams of solid materials on a regular basis rather than selecting only very few structures and writing a separate computer time proposal for every single one. In some applications where DFT incorrectly predicts materials to be a metal or yields the wrong ground-state structure, we will have an accurate phase diagram for the first time that, at high pressure, will rival the best available experimental results in precision.

We would be able to simulate fluids with QMC that require a large number of atomic configurations will be sampled. So far, only fluid hydrogen has been study with QMC by the group of D. Ceperley.

Furthermore we would be able to study the interaction of structural distortions and orbital occupation in materials like FeO where small energy differences between different orbital and spin states have made it extremely difficult to study this geophysically very important material with DFT.

One also expects van der Waals and hydrogen bonds to be described more accurately with QMC than with DFT. Much more work and CPU time is needed however to fully explore the capability of the QMC method.

QMCPack has been shown to scale up to 100,000 cores by Ken Esler and consequently much bigger simulations will be done in the future.

10.2.2.4 Computational and Storage Requirements Summary

	Current (2009)	In 2013
Computational Hours	2 M (1.25 at NERSC)	20 M
Parallel Concurrency	64 – 512 cores	512 – 4096 cores
Wall Hours per Run	12	12
Aggregate Memory	128 – 1024 GB	1024 – 8182 GB
Memory per Core	2 GB	2 GB
I/O per Run	2 GB	4 – 16 GB
On-Line Storage Needed	0.5 TB	1 TB
Data Transfer	8 GB per week	20 GB per week
Archival Storage	1 TB	4 TB

10.2.2.5 Support Services and Software

No special services or software needed, all software we use is self-installed.

10.2.2.6 Emerging HPC Architectures and Programming Models

Through the diligent development work in QMCPACK by Esler, Kim, and Ceperley, we are able to report that QMC algorithms run 10 to 15 times faster on a GPU than they run on a quad-core Xeon processor using all four cores (Figure 8-4). This was made possible by a complete restructuring of the parallelization in QMC. The QMC algorithm requires that a large number of walkers sample the many-body wavefunction. Typically these walkers (~1,000) are distributed among different processors and their propagation proceeds almost independently until they need to be redistributed for load balancing.

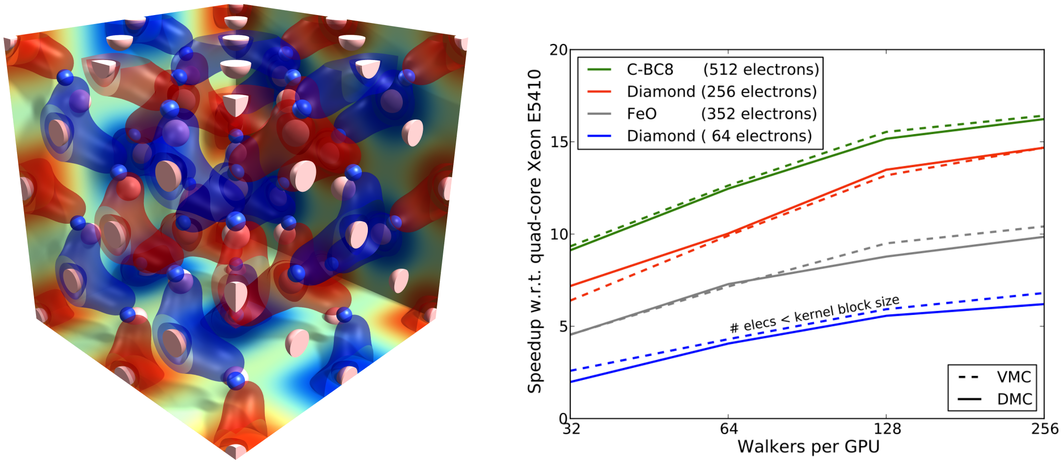
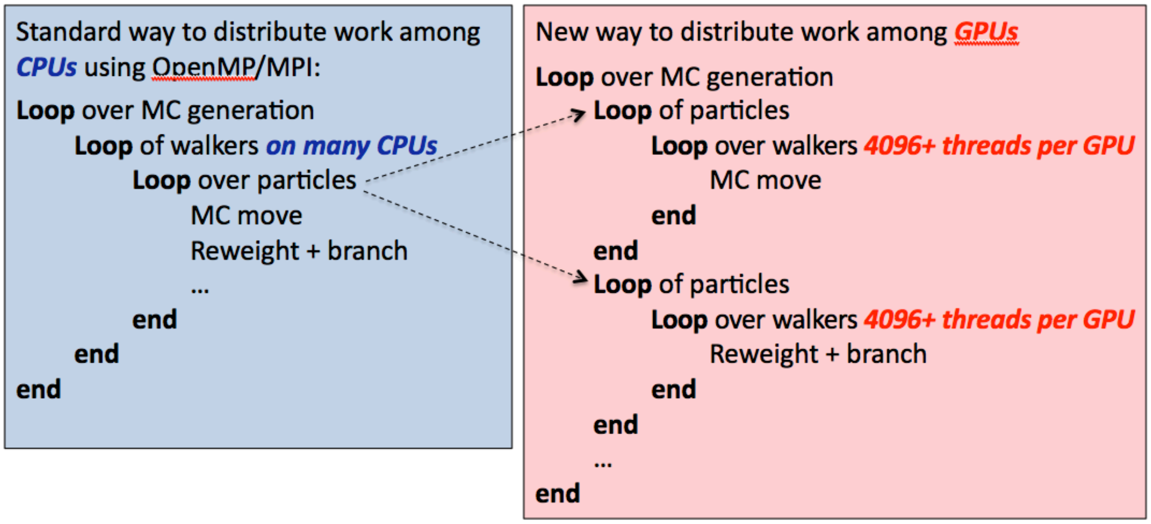


Figure 10-4. Left: Super cell of cubic boron nitride that can now be studied with QMC. Right: The code QMCPACK shows excellent scaling on a GPU and runs 10-15 times faster than on a quad-core Xeon processor using all four cores. (From K.P. Esler, J. Kim, D. M. Ceperley, "Fully accelerating quantum Monte Carlo simulations of real materials on GPUs", GP-GPU conference).

On a CPU, we propagate each electron in sequence for a single walker before proceeding to the next walker in order to make use of the large CPU cache. Efficient operation on a GPU, however, requires exposing as much parallelism as possible, so that we propagate all walkers in parallel. Flexibility in the QMC algorithm offers multiply ways to introduce parallelism and comparatively low communication makes it a very good candidate for GPU programming and for future multi-core architectures. Figure 10-5 illustrates the loop reordering that allowed Esler, Kim, and Ceperley to achieve excellent scaling on a GPU. Instead of distributing walkers among processors, many walkers are assigned to each GPU. The different steps in the calculation such as Monte Carlo moves, energy calculations, reweighting and branching are executed one after the other but each operation is executed for all walkers that are on assigned to a GPU. The approach of breaking up the algorithm in small pieces, each executed for all walkers at once, allows one to coalesce memory access and reduce the amount of shared memory required. We expect that many QMC future calculations will be performed on GPUs as such systems become available.



```
Loop over MC generation
    Loop of walkers on many CPUs
        Loop over particles
            MC move
            Reweight + branch
            ...
        end
    end
end
```

```
Loop over MC generation
    Loop of particles
        Loop over walkers 4096+ threads per GPU
            MC move
        end
    end
    Loop of particles
        Loop over walkers 4096+ threads per GPU
            Reweight + branch
        end
    end
    ...
end
```

Figure 10-5. Comparison of the traditional parallelization schemes on CPUs (left) with the new approach for GPUs (right).

10.2.3 Classical Molecular Dynamics Codes and Coupling of Length Scales

Principal Investigator: Peter Cummings, Vanderbilt University and Oak Ridge National Laboratory
Contributors: Normand Modine and Randall Cygan, Sandia National Laboratories
NERSC Repo: mp138

10.2.3.1 Summary and Scientific Objectives

Molecular dynamics is used broadly throughout the Basic Energy Sciences research portfolio. In materials sciences, classical molecular dynamics (CMD) is used to study, for example, the morphology and grain structure of solid-state materials, the dynamics of nanowires, ceramics, polymer melts and blends and hybrid organic-inorganic nanostructured materials. *Ab initio* molecular dynamics (AIMD) enables the investigation of reactive processes in systems where accurate reactive classical force fields have not been developed, albeit at a greatly increased computational cost. The exponential growth in computer power has increasingly enabled the application of AIMD to realistically complex systems over time scales relevant to scientifically and technologically important phenomena. Given expected continuing increases in computational power, an interesting and important challenge is to improve the approximations used in AIMD with the goal of achieving chemical accuracy, which is typically defined as average errors in the barriers that control dynamic processes less than the room-temperature thermal energy. For example, for a 100-1,000 atom extended system, where AIMD using the local density approximation (LDA) to Kohn-Sham density functional theory (DFT) can be performed using a teraflop computer, one could aim to perform a chemically accurate calculation using a petaflop computer. In chemical sciences, CMD is used to study, for example, conformations of molecules in solution, water and aqueous solutions, and self-assembly driven by weak interactions. At the nanoscale, researchers increasingly wish to create and model hybrid systems combining inorganic, organic and biological components (e.g., DNA-functionalized silicon chips, and inorganic-organic hybrid materials).

10.2.3.2 Molecular Dynamics Simulations in Materials Sciences

Classical and *ab initio* molecular dynamics are widely used to study dynamical behavior in materials. For example, the early-time structural response of materials to radiation damage produced by neutrons or ions involves the creation, clustering, and annealing of defects in a displacement cascade. These complex processes are best studied by molecular dynamics techniques. Another exciting possibility is the application of molecular dynamics based on the Time-Dependent Density Functional Theory (TDDFT) to study coupled electronic and ionic dynamics. Accurately resolving electronic dynamics typically requires sub-attosecond time steps, and therefore, millions of *ab initio* time-steps can be required to simulate processes of interest. This requires multi-month run times on hundreds or thousands of processors, and therefore provides a significant computational challenge.

Molecular electronics is one area of nanomaterials science in which classical molecular dynamics simulations are complementary to, and used in conjunction with, first principle methods. The breakthrough experiment designed to measure the current-voltage (I-V) characteristics of a single molecule was a molecular break junction experiment. In this experiment, a gold nanowire, onto which benzenedithiol (BDT) molecules are chemisorbed, is pulled apart, and the ends brought back together in order to trap an individual BDT molecule bonded to each end of the gold nanowire. Next, a voltage is applied to the gold wires and the current between the gold nanoscale electrodes is measured to determine the I-V curve.

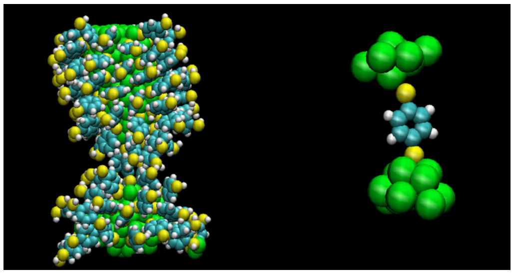


Figure 10-6. Images from a molecular dynamics simulation of the molecular break junction experiment. Left: gold nanowire with BDT molecules chemisorbed after rupture followed by bringing back together to trap a single BDT molecule bonded to each nanowire tip. Right: close-up of tip and BDT molecule bonded to both tips.

Molecular dynamics simulations have been used to study the characteristics of gold nanowire rupture and to create computationally configurations of Au-BDT-Au for which the electron transport (and hence I-V curve) can be measured. An example of an Au-BDT-Au configuration generated by molecular dynamics is shown in the Figure 10-6. Enroute to developing molecular dynamics simulations of molecular break junction formation, it was found – on the basis of 540 molecular dynamics simulations (30 simulations at all possible combinations of 3 elongations, 2 sizes and 3 temperatures) – that the rupture of a gold nanowire in vacuum exhibits a universal energy release mechanism (Pu, et al., *J. Am. Chem. Soc.*, 130 (2008), 17907). The mechanism governing the formation of break structures in Au nanowires continues to be explored, by conducting over 25,000 simulations (for a total simulated time of $\sim 10\ \mu s$) to ascertain the rate dependent mechanism with higher resolution as well as to explore the crystallographic structure of gold atoms localized around the break junction. These new simulations have yielded insight into the formation of polytetrahedral arrangements preceding the pre-rupture formation of monatomic wires.

10.2.3.3 Molecular Dynamics Simulations in Chemical Sciences

Molecular dynamics simulations are used very broadly in the chemical sciences, and the ability of molecular dynamics to study collective phenomena of increasing complexity and relevance grows in direct proportion to the availability of computing resources. As one example, we consider the case of nanoconfined fluids. For two decades, experiments by different research groups have disagreed on whether fluids composed of non-polar, symmetric molecules (such as linear or cyclic alkanes) undergo a sharp phase transition to an ordered solid phase when confined by molecularly smooth mica surfaces. Experimental evidence for a transition is seen when the number of layers of fluid between the surfaces is around 5-7 layers, depending on the nature of the nanoconfined fluid. Through a series of highly detailed molecular dynamics simulations, using an atomistically detailed model of both mica and the fluid (linear dodecane and cyclohexane), it has been conclusively shown that, indeed, these fluids undergo a transition from fluid to solid states at 5-7 layers of nanoconfined fluid, in agreement with some of the experimental findings. A completely unexpected finding is that the primary driving force for fluid-to-solid transition is the electrostatic interactions between the ions in the mica surface and the partial charges in the non-polar molecules. A snapshot of a configuration from the simulations is shown in Figure 10-7.

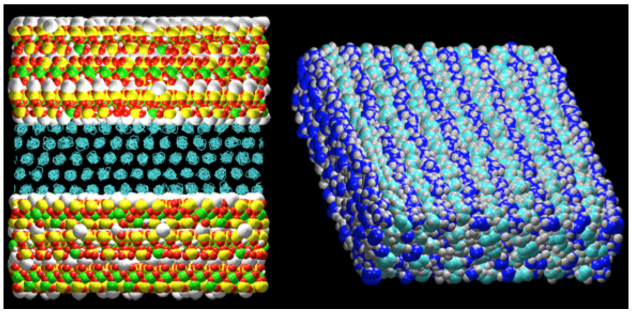


Figure 10-7. Equilibrated atomistic simulation (at constant pressure and temperature) of cyclohexane between mica sheets with a separation of 5 molecular diameters. Between 6 and 5 molecular diameters, the molecules undergo a rapid and abrupt transition to an FCC ordered solid-like structure (order parameter 0.90), shown on the right, where molecules are colored to accentuate the order.

10.2.3.4 Molecular Dynamics Simulations in Geosciences

The last decade has seen a significant increase in the number of computational studies involving the application of molecular dynamics in the geosciences. These investigations are wide ranging, and have a common emphasis on understanding the structure and behavior of natural materials at the temperatures and pressures associated with the Earth's crust and interior. Geochemical examples include the simulation of bulk minerals, silicate glasses, aqueous systems and other geological fluids, and the mineral-fluid interface where many important geological and environmental processes are controlled. Additionally, MD methods have allowed geophysicists to better evaluate structural changes in dense minerals that occur with the extreme pressures of the Earth's mantle and core, and to evaluate elastic and seismic properties and interpret discontinuities in the deep structure of the Earth. Molecular simulations are often the only means available to examine mineral and fluid behavior at such extreme conditions due to technical limitations in experimental methods. Two examples of CMD simulations relevant to the geosciences are shown in Figure 10-8. The periodic cell used in the simulation of the layered double hydroxide phase contains over one million atoms.

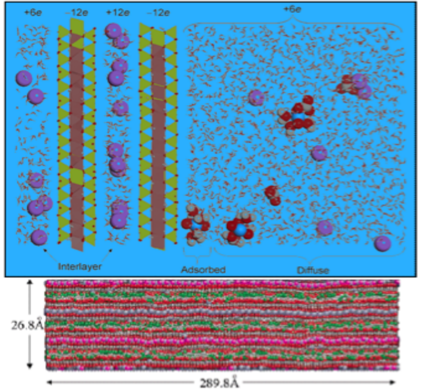


Figure 10-8. Top: snapshot from a large-scale MD simulation of the adsorption of hydrated uranyl (blue and red spheres) onto the basal surface of sodium montmorillonite clay used to predict the partitioning. Bottom: snapshot from a large-scale MD simulation of a layered double hydroxide material used to derive elastic properties and discriminate the mechanical response of the aluminum-magnesium (magenta) framework layer and hydrated chloride ions (green) of the interlayer.

Molecular dynamics, combined with geometry optimizations, have been used to examine the complex chemistry of multicomponent systems of natural phases that typically include significant disorder, vacancies, and complex surface and interfacial structures. The ubiquitous nature of nanocrystalline materials in crustal environments, characterized by relatively high surface areas, controls many geochemical processes including enhanced mineral reactivity. Significant electrostatic fields common to layered structure of clay minerals and many mineral-water systems present a challenge in the application and accuracy of molecular dynamics simulations. Accurate dissociation models for water molecules and protonation of surface functional groups are needed to develop predictive models for adsorption, dissolution, nucleation, and growth phenomena. Large-scale CMD simulations involving many tens to hundreds of thousands of atoms are required to capture the long-range electrostatics associated with the interaction of large organic materials and bio-molecules (e.g., crude oil, natural organic matter, etc.) at mineral surfaces and to improve the quality of enthalpy and free energy calculations (e.g., potential mean force).

10.2.3.5 Methods of Solution

Classical molecular dynamics simulations rely on the repetitive solution of Newton's equation of motion for (typically) a large assembly of atoms over an extremely large number of time steps. State-of-the-art simulations can involve many tens of thousands to millions of atoms in a periodic simulation cell with time steps on the order of femtoseconds (10^{-15} s) that are run for many nanoseconds (10^{-9} s), or in some extreme cases up to a microsecond (10^{-6} s), i.e., one billion time steps. Forces and energies are calculated for each time step based on a potential energy model or force field that generally uses parameterized analytical functions to evaluate pairwise interactions among all the atoms in the system. Force fields include expressions for bonded (i.e., stretch, bend, torsion and other molecular motions) and nonbonded (van der Waals, short-range repulsion, and Coulombic) interactions. Periodic boundaries needed to effectively simulate long-range electrostatic forces require efficient algorithms (e.g., Ewald, particle-mesh Ewald, and particle-particle mesh methods) to obtain suitable energy convergence and accurate forces and energies. Structural snapshots are saved to disk storage during the course of the MD simulation to provide an atomic trajectory of the system for subsequent analysis. Ultimately, CMD simulations can provide insight into the molecular control of structure, thermodynamics and transport properties, mechanics, rheology, vibrational and dynamical spectroscopy, and many other important physical and chemical properties of materials.

The molecular dynamics community is very diverse, with scientists from the materials (organic and inorganic systems), chemistry (organic, small-molecule and solvent/solute systems), biology (biological systems), as well as nanoscience and engineering communities, in which all of these systems (and hybrids of them) are of interest. The increasing complexity of the systems studied in the molecular dynamics community is driving it towards the use of standard codes, rather than in-house codes for each research group. In particular, in order to run large simulations that execute efficiently on the largest computers, molecular dynamics simulators are increasingly using codes that are efficient on large numbers of cores, such as LAMMPS, NAMD and GROMACS. With translators available between the biological simulations codes (CHARMM and AMBER) and LAMMPS, there is an increasing trend to use LAMMPS as a trajectory generator, combined with post-simulation analysis. Additionally, since LAMMPS is open-source, more advanced users have modified it to compute the required properties on the fly.

10.2.3.6 HPC Requirements

Although molecular dynamics simulations for relatively small chemical systems can be successfully completed on a single-processor machine such as a laptop computer, most research problems involving thousands to millions of atoms require high performance computing platforms to obtain suitable results in a reasonable time frame. Domain decomposition of the simulation cell among processors and data replication are often used for computational speedup. Because of the deterministic and time-evolution nature of molecular dynamics, such spatial methods are the only ones available for speedup. Efficient parallel codes such as LAMMPS, DL_POLY, GROMACS, NAMD, NWChem, and others are available to the scientific community and have been successfully used to model complex molecular systems in physics, chemistry, geochemistry, biochemistry, and materials science.

In contrast to CMD and the use of an empirical force field, quantum-based molecular dynamics methods rely on approximate solutions to the Schrödinger equation to obtain energies for each time step of the dynamics simulation. *Ab initio* molecular dynamics (AIMD), therefore provides a more accurate representation of the system interactions by calculating the potential energy of the system on the fly and enables the direct simulation of excited states and chemical reactions. Carr-Parrinello and AIMD methods are typically used to derive atomic trajectories but, due to the computational cost of electronic structure methods, these approaches are limited to relatively small system sizes (hundreds of atoms) and simulation times of tens of picoseconds (10^{-12} s).

Coarse-grained classical methods are often used to access larger-sized molecular systems for even longer simulation times. Rather than use atoms as the basic unit, coarse-grained models describe rigid units or beads that represent the fundamental structure of the molecule by grouping several heavy atoms into each bead. Several million beads can be easily modeled for tens of microseconds in these simulations. This approach has led to significant advances in understanding the structure and behavior of polymers and in evaluating large biomolecules and biological systems.

One significant barrier in the use of molecular dynamics is the large amount of disk storage required for the atomic trajectories derived from the simulations. For most CMD simulation studies it is necessary to store, at a minimum, the x,y,z coordinates for each atom in the system for a large number of time snapshots based on a prescribed time increment. Trajectory analysis involving, for example, evaluation of the mean-square displacements of atoms can then be used to derive diffusion coefficients and may require structural snapshots for tens of nanoseconds. In contrast, a power spectrum needs approximately 40 ps of an equilibrated trajectory to derive the vibrational behavior through the use of atomic velocity autocorrelation, but would necessarily require the velocity components for each atom. The frequency in capturing the structure varies depending on the detail needed in the analysis of the evolving system. Although molecular dynamics time steps are usually 0.5 to 2.0 femtoseconds to ensure dynamic stability, the resulting structures are stored less frequently. Nonetheless, a relatively small thousand-atom system can require over 4 GB of storage to obtain an accurate power spectrum, especially if the relatively rapid vibrational behavior of hydrogen is desired. More dynamical detail or long atomic correlations for a million-atom simulation may require storage of up to many tens of GB. Even with high performance LAN and WAN systems, there still remain practical issues in the rates and latency associated with the transfer of such large datasets. Developers of LAMMPS have recently added an option to perform correlation analysis during the course of the molecular dynamics simulation to reduce the trajectory storage requirements.

10.2.3.7 Computational and Storage Requirements Summary

	Current (2009)	**In 2013**
Computational hours:	6.5 M	~50 M
Parallel Concurrency:	100+ cores	1000+ cores
Wall hours per run:	24-maximum allowed	24-maximum allowed
Aggregate memory (GB)	150+	1500+
Memory per core (GB)	1.5	1.5
I/O per run (GB):	>= #atoms x 48 bytes (Cartesian positions and velocities) every 10-100 time steps	>= #atoms x 48 bytes (Cartesian positions and velocities) every 10-100 time steps
On-line storarge needed (GB):	0.01-1 TB	0.1-10 TB
Data transfer:	0.01-1 TB for each completed run	0.1-10TB
Archival storage (GB)	0.01-1TB for each system studied	0.1-10 TB for each system studied
	6 GB	30 GB

10.2.3.8 Support Services and Software

Because LAMMPS is distributed in source form, it requires several libraries (such as MPI and FFT routines); users need an optimally configured and compiled version of vanilla LAMMPS on the NERSC computers (currently supplied) plus makefiles for users who need to compile their own versions of LAMMPS modified for specific applications. Similar comments apply to GROMACS, NAMD and other standard molecular dynamics simulation codes. *Ab initio* molecular dynamics codes are generally not open source; nevertheless, optimized (for the NERSC machines) versions of the standard codes (VASP, CP2K, CPMD, etc) should be available for users with valid licenses for these codes.

Finally, it is important to highlight the difficulty in visualizing the results from molecular dynamics simulations involving hundreds of thousands to millions of atoms. High-end video cards for PC workstations such as those from NVIDIA provide the graphical processing power to examine such large systems, but often cannot provide the speed for manual manipulation of an extraordinarily large array of atoms, or provide practical animation of the atomic trajectory. Recent improvements in efficient visualization software (e.g., VMD, and AtomEye) have significantly improved upon these graphical needs.

10.2.3.9 Emerging HPC Architectures and Programming Models

Two new architectures are affecting CMD specifically. The first is graphical processor units (GPUs). There are a growing number of research groups implementing molecular dynamics on GPUs, with speed increases of 10-80 times that of conventional state-of-the art processors being reported. Standard codes (such as LAMMPS) are being ported to GPUs, and new codes (such as HOOMD), written specifically for GPUs, are being developed. While large speed gains of 10-80 times have been demonstrated for classical molecular dynamics of systems with short-ranged (non-electrostatic) interactions, the community is still waiting for an equally efficient implementation on GPUs that includes electrostatics. Additionally, GPU-to-GPU communication is slow enough that a parallel molecular dynamics domain-decomposed onto multiple GPUs is not practical today. Thus, for now, the main application of GPUs is for accelerating molecular dynamics of simulations with short-ranged interactions that contain less then several hundred thousand atoms (and so the calculation can be performed on a single GPU). As a consequence for NERSC, a battery of GPUs running ported versions of standard codes such a LAMMPS might absorb some of the demand for molecular dynamics cycles.

The second novel architecture of interest is the Anton computer designed by D. E. Shaw Research. Anton is the latest in a series of attempts to create a machine that is highly efficient at the hardware level in executing molecular dynamics. Predecessors include the GRAPE-MD machine, developed in Japan initially for solving galactic simulations (in which the interaction between objects is the 1/r gravitational attraction). Benchmarks indicate that on simulations with the same number of cores, Anton is 1-2 orders of magnitude faster than any other hardware/software combination. Anton is not available for purchase; however, D. E. Shaw Research will make one available to the general biology community through the NSF Pittsburgh Supercomputer Center, with time allocated by a peer-reviewed competitive proposal process run by the NIH. Those users who have access to Anton will have a significant speed advantage over their competitors, particularly in reaching long simulation times (μs and beyond).

10.2.4 Computational Studies at BNL of the Chemistry of Energy Production and Use

Principal Investigator: James T. Muckerman, Chemistry Department and Center for Functional Nanomaterials, Brookhaven National Laboratory
NERSC Repo: m783

10.2.4.1 Summary and Scientific Objectives

The principal goal of this work is to elucidate mechanisms for photochemical and photoelectrochemical fuels production with molecular catalysts using quantum chemical (including hybrid DFT) methods to aid in the design of improved catalysts. This is a problem of urgent programmatic interest to the DOE because of its relevance to renewable fuels and climate change, so that the aim of this project is to make as much progress as possible with current methodology.

This project focuses on processes catalyzed by transition metal complexes in solution. The absolute free energies in solution of a series of possible intermediates in the catalytic cycles for water oxidation, hydrogen production, and renewable hydride generation are calculated to identify the catalytic mechanism and the free energy profile along the reaction paths. Such mechanisms involve species that are capable of electron transfer reactions, proton transfer reactions, and (often) proton-coupled, electron-transfer (PCET) reactions. Once these intermediates are identified, we calculate the pK_a and E^o values for the acidity of their protons and their standard reduction potentials, respectively, to allow the construction of the Pourbaix diagram (a E vs. pH diagram analogous to a phase diagram that indicates the chemical form of the intermediate present in solution at a given pH and applied potential). The free energy profiles and the Pourbaix diagrams of the species along the catalytic cycle offer a means of understanding and manipulating the properties of the catalyst.

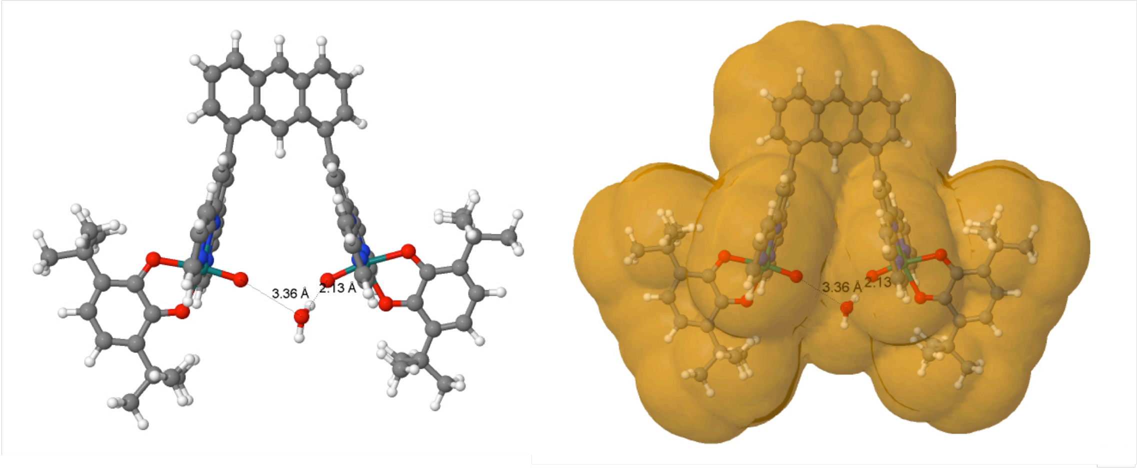


Figure 10-9. A possible intermediate in water oxidation by the Tanaka catalyst (left); and the same species shown in the solvent cavity in a polarizable continuum model of bulk water (right).

The calculation of the absolute free energy of a species, its pK_a value(s) and its standard reduction potential(s) involves several steps. We first calculate the optimized geometry (minimum electronic energy structure) of the species (including any explicit solvent molecules in its computational model) in the gas phase. Next we calculate all vibrational frequencies of the species in the gas phase at its optimized gas-phase structure in order to compute the zero-point energy, thermal energy and entropic corrections to its total electronic energy that define its absolute free energy in the gas phase. Finally, we reoptimize the geometry of the species in a polarizable continuum model of the bulk solvent to obtain its electronic energy in the solvent. The total free energy of the species in solution is then taken as $DG^o = E_{solv} - E_{gas} + DG^o_{ZP\&T}$, where $DG^o_{ZP\&T}$ is sum of all zero-point and thermal corrections to the gas-phase electronic energy. The calculations are carried out for a species with a given chemical formula by specifying its total charge and spin multiplicity, and often it is necessary to consider different spin states for a given charge state.

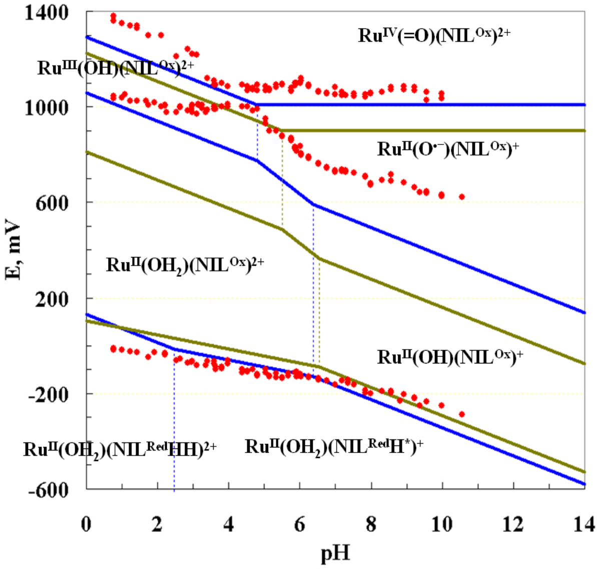


Figure 10-10. A Pourbaix diagram of [Ru^{II}(trpy)(NIL)(OH_2)]$^{2+}$ (where NIL is 2-iminoquinone), an analogue of the monomer of the Tanaka catalyst. The red points are experimental data, the blue lines correspond to the calculated diagram of the trans isomer, and the green lines that of the cis isomer. Horizontal lines corresponding to electron transfer reactions, vertical lines to proton transfer reactons, and sloped lines to proton-coupled electron-transfer reactions. The slope of –59 mV/pH times the number of protons transferred divided by the

We compute the free energy change, DG^o, for a given reaction step A + nH+ + me^- → B in solution as

$$DG^o_{solv} = G^o_{solv}(B) - G^o_{solv}(A) - nG^o_{solv}(H^+) - mG^o_{gas}(e^-) ,$$

where the standard state of the electron is in the gas phase with the "ion convention." If this is a reduction reaction, i.e., $n \neq 0$, then we express the free energy change with respect to the normal hydrogen electrode (NHE), $DG^o_{NHE} = ½\ G^o_{gas}(H_2) - G^o_{solv}(H^+) - G^o_{gas}(e^-)$. The standard reduction potentials are obtained from the relation $E^o = DG^o_{solv} / nF - DG^o_{NHE} / F$, where F is the Faraday constant. These same considerations can be used to calculation "hydricities", which are analogous to acidities, but refer to the ability of a species to donate a hydride ion.

Our current work is focused in three areas: (1) water oxidation by the so-called Tanaka catalyst and its analogues (see Figure 10-10); (2) the use of a photogenerated, renewable hydride donor (a model of NADPH) to reduce transition-metal bound carbonyl ligands (i.e., CO) via ionic hydrogenation, the most difficult step in the reduction of CO_2 to methanol; and (3) hydrogen production with hydrogenase-inspired catalysts, and the exploration of their applicability to hydrogen production in aqueous solution.

10.2.4.2 Methods of Solution

We employ quantum chemistry methodology augmented by self-consistent reaction field solvation models to compute the absolute free energy of each species of interest in solution. Our more sophisticated models employ explicit solvent molecules that interact through hydrogen bonding and other explicit electrostatic interactions not included in a polarizable continuum model of the bulk solvent. Electronic structure calculations are defined by the number of nuclei and electrons. Because of the size of the model systems, generally 70 to 200 atoms, high-level quantum chemistry methods such as coupled-cluster and multireference configuration interaction methods are not feasible. Hybrid density functional theory with a moderate-size basis is currently the highest quality calculation that we can perform. For purposes of making comparison of calculated species with those observed in experiments, we employ TD-DFT calculations to obtain UV-vis spectra of each intermediate considered.

There are obvious deficiencies in this level of theory, particularly in the identification of the lowest-energy spin state of species with so-called “non-innocent” ligands. These are ligands that are redox active, i.e, they can be reduced at potentials comparable to those that reduce the metal center, and there is some mixing of the metal dp orbitals with the ligand p* orbitals. Because it does not allow for multiconfigurational effects such as valence configuration interaction (e.g., the inclusion in the wavefunction of the $(\mathrm{HOMO})^2 \rightarrow (\mathrm{LUMO})^2$ excitation) even hybrid density functional theory tends to underestimate the stability of the closed-shell singlet states of such systems, and often erroneously predicts a triplet ground state. The same phenomenon can occur in calculations involving possible quartet or quintet states. In order to discern the true spin state of the system, we must frequently resort to Broken Symmetry (BS) DFT calculations in which the orbitals of a high-spin state are maximally localized (breaking spatial symmetry), and one or more spins are flipped to obtain a broken spin symmetry state with the minimum M_S value. This “state” is then optimized in a spin-unrestricted SCF calculation, and, in the case of a high-spin triplet, one of three possible outcomes can occur: (1) the highest occupied beta-spin orbital collapses to the corresponding alpha-spin orbital yielding the closed-shell singlet state; (2) the SCF converges to a BS state which has energy lower than the triplet state (and usually $\langle S^2 \rangle$ intermediate between the zero and two) indicative of the antiferromagnetically coupled spins of an open-shell singlet state; or (3) the BS state has energy higher than the triplet state indicative of the ferromagnetically coupled spins of a triplet state. This “poor man’s multireference configuration interaction” is not a spin eigenfunction, but yields the best available estimate of the true energy of an antiferromagnetically coupled state (or a state with significant valence configuration interaction).

Our tool of choice is the Gaussian quantum chemistry package, although it, unfortunately, runs effectively only in an SMP environment. Gaussian can be made to do all the calculations that we require, while other codes have certain deficiencies in this

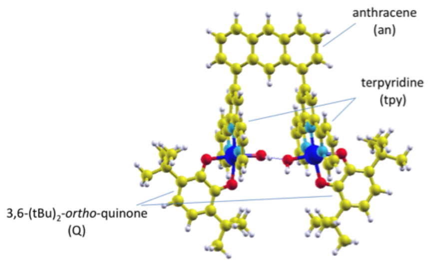


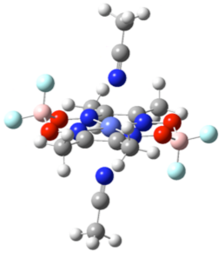
Figure 10-11. Resting state of the Tanaka (dimer) catalyst.

regard. GAMESS, for example, has deficiencies in the treatment of transition metals, with limited implementation of effective core potentials (ECPs) and poor initial wavefunction guesses that lead to convergence problems. NWChem has most of the required features, was especially written to scale well in an MPI environment, but is very slow. ORCA has an extremely effective implementation of Broken Symmetry calculations, but it offers no support for ECPs thus requiring all-electron calculations for our second-row transition-metal-containing systems.

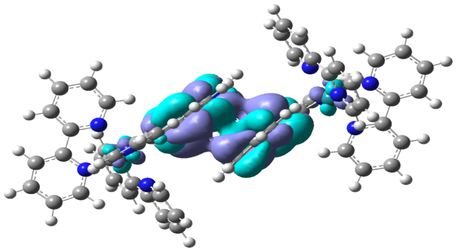
10.2.4.3 HPC Requirements

The typical size of our molecular catalyst calculations is up to 160 atoms and 1200 basis functions. The main theory we use is Density Functional Theory. The code is usually Gaussian03 and the parallel scheme is OpenMP. An alternative is to use the NWChem code, which is optimized for MPI, but it is very slow. For example, to match the wall-clock time of a job run with Gaussian03 using eight nodes on Bassi, one would have to use 256 nodes on Franklin.

An SMP architecture with at least 16 cores per node is urgently required for many of our calculations. With fewer than 16 we can work but this is a limitation. Consider the example of a geometry optimization on the $[Co^{II}(dmgBF_2)_2(CH_3CN)_2]^0$ doublet state, a proton reduction catalyst (at right), using B3LYP / 6-31+G(d,p). The geometry optimization of this species in a Polarizable Continuum Model (PCM) model of acetonitrile solution required 122 cycles and took 124 days to complete running as a standalone process on all 16-cores of a AMD64 Opteron quad-CPU, quad-core workstation with 64GB of memory and 1.5TB of local scratch disk. In contrast, using Bassi, with only eight processors on a node, it was not possible to complete a single optimization cycle within the time limit of the Bassi queue (48 hours).



The HOMO of p-stacking D,L-$\{[\mathbf{Ru\text{-}pbn}H^{\bullet}]^{2+}\}_2$ dimer (shown below), a critical intermediate in the photogeneration of an NADPH-like hydride donor, localizes the electrons in a bonding orbital between the π systems of the monomers. This system has 146 atoms and 948 basis functions. Using 16 AMD64 cores and 51.2GB of memory, the geometry optimization of this dimer in a PCM model of acetonitrile solution required nearly 187 days. This is another example of a job that could not be run on the architectures at NERSC at the time of this workshop.



The vibrational frequency calculation of the a derivative of the monomeric unit of this renewable hydride donor, $[Ru^{II}(bpy)_2(pbnH)]^{3+}$, consisting of 73 atoms, required over 71 days using four AMD64 cores and 12.8 GB of memory in a B3LYP / MWB28(Ru), 6-31G(d,p) calculation. Since we know of no way to checkpoint a vibrational frequency job before it finishes, this is another example of important work that cannot currently be carried out at NERSC.

The size of available SMP machines (in terms of number of processors per node) also limits the size of our simulation model design. If a larger SMP node were available, we could expand the size of our model catalytic systems from ten angstroms to ten nanometers.

With new resources, we would be able to treat solvent molecules quantum mechanically. This is significant for the treatment of water quantum mechanically, especially for water splitting catalysis. With a larger SMP computing infrastructure, we could explore even larger catalysts, e.g. molecular catalysts attached to electrodes, semiconductor photocatalysts, or self-assembled polymers. More throughput for small jobs is a key element for studying catalytic redox systems. One complete calculation is required for each oxidation state. Studying electrochemical catalysis requires investigating many possible oxidation states. Therefore, enhancing the throughput for calculating these small-to-medium sized jobs would be very beneficial to the progress of computational catalytic chemistry.

10.2.4.4 Computational and Storage Requirements Summary

	Current (2009)	**In 2013**
Main codes	Gaussian03, NWChem	Gaussian 09, NWChem
Computational Hours	620 K (300 at NERSC)	2.4 M
Parallel	16-256 MPI, 4-16 SMP	16-512 MPI, 8-32 SMP
Wall Hours per Run	48 hours	48 hours
Aggregate Memory	16 – 256 GB	32 – 512 GB
Memory per Core	1 GB	4 GB
I/O per Run	300 GB, 20 – 10 GB checkpoint files	1 TB, 100 – 500 GB checkpoint files
On-Line Storage Needed		1000 GB
Data Transfer		100 GB/year
Archival Storage	60 MB	2 TB, 400 files

10.2.4.5 Support Services and Software

The support of architectures that run Gaussian efficiently is crucial for our work. In addition to that, a major obstacle is the time limit of queue system. This is generally 48 hours or less for any queue implemented at NERSC. Our vibrational frequency calculations (required to obtain the free energy by including zero-point and thermal energy, and entropy corrections to the electronic energy) are often not able to finish in the 48-hour time frame, and we know of no way to checkpoint intermediate results. A special arrangement for a longer queue is required.

The number of jobs allowed to run simultaneously is also a limit for researchers who run less massively parallel calculations.

10.2.5 Reaction Dynamics in Complex Molecular Systems

Principal Investigator: Thomas Miller, California Institute of Technology
NERSC Repo: m822

10.2.5.1 Summary and Scientific Objectives

We employ path integral methods, rare-event sampling methods, and classical molecular dynamics to simulate reactive processes in complex molecular systems. Areas of primary focus include (i) proton-coupled electron transfer (PCET) dynamics in enzymes and photo-catalysts and (ii) long-timescale dynamics in protein-conducting transmembrane channels. The simulation methods are used to investigate the kinetics and regulation of these processes. This work provides improved understanding of the fundamental mechanisms that govern biomolecular energy conversion and protein transport.

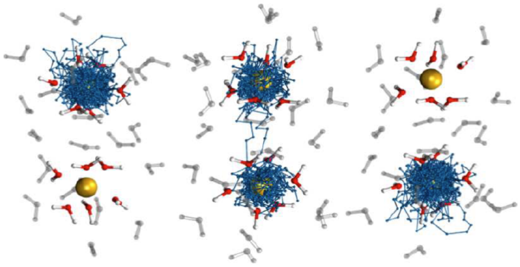


Figure 10-12. Direct simulation results for a benchmark electron transfer system in explicit water, using combining path integral methods with transition path sampling.

From the computational perspective, these applications are unified by the challenge of bridging dynamical length scales and timescales in molecular simulations. PCET dynamics feature the coupling of electron transfer, proton transfer, and solvent fluctuation dynamics; whereas protein translocation involves the coupling of channel gating motions with the transport of proteins across the membrane. The PCET studies face the additional challenge of simulating coupled quantum and classical dynamics. Path integral methods and rare-event sampling methods offer a computationally scalable approach to mitigating these challenges.

Specific areas of research being addressed by our current NERSC allocation include:

Specific Goal #1: Direct simulations of PCET dynamics in a symmetric, mixed-valence iron bi-imidazoline system.

Iron bi-imidazolines have been extensively studied as a prototype for PCET dynamics and as a model for the tyrosine reduction step in photosystem II. By combining path-integral molecular dynamics and with the transition path sampling method of studying chemical reactions, we will perform the first direct simulations of these reactions, we will determine the chronology of the ET/PT events, and we will explore the origin of the unexpectedly low experimental kinetic isotope effect in the PCET reaction rate. Using a modified version of the DL_POLY molecular simulation package, we will sample ~5,000 PIMD trajectories at an estimated cost of one million CPU hours.

Specific Goal #2: Regulation of protein translocation and membrane integration via the Sec translocon.

A critical step in the biosynthesis of many proteins involves either translocation across a cellular membrane or integration into a cellular membrane. Both processes proceed via the Sec translocase - a ubiquitous and highly conserved transmembrane channel. We will test the hypothesis that Sec-facilitated protein translocation and membrane integration is regulated by a mechanism in which the translocon acts as a substrate-controlled conformational switch between pathways for membrane integration and secretion. Using the string method in collective variables, we will characterize energetics, mechanism, and dynamics of protein translocation and membrane integration. To perform these calculations, a modified version of the NAMD package will be used to simulate calculations will simulate 40 independent replicas of the system along the transition path, each for a time of approximately 40 ns. The estimated cost will be two million CPU hours.

10.2.5.2 Methods of Solution

We typically perform large numbers (103-105) of nearly-independent MD or path integral molecular dynamics (PIMD) trajectories using molecular mechanics force fields and electron pseudopotentials. Typical trajectories require 106 timesteps, and systems sizes range from 10,000-150,000 atoms.

Calculations are performed using versions of the DL_POLY, GROMACS, and NAMD that we have modified to perform the path-integral and rare-event sampling techniques.

Heavy communication between processors is needed for the force-evaluation step of the MD and PIMD trajectories. Considerably less, but non-zero, communication is typically needed between the independent trajectories. Parallelization is achieved primarily using MPI.

10.2.5.3 HPC Requirements

Individual production trajectories for both PCET and protein translocation dynamics typically utilize between 32-128 cores for 6-12 wall-clock hours in systems such as Franklin and Hopper at NERSC and Jaguar at the OLCF. Methodological advances allow for the calculation reaction rate constants and mechanism from a large number of nearly independent trajectories, enabling efficient and large-scale parallelization. These individual trajectories can be bundled into combined jobs that utilize 5,012 production cores or more.

Access to 50x resources will allow us to:

1) A 50x increase in computational resources would allow for the use of *ab initio* potential energy surfaces (obtained from either DFT or embedded DFT methods) in our simulations. These *ab initio* potentials would offer improved accuracy and flexibility over the molecular mechanics force fields and electron pseudopotentials that are currently employed.

2) A 50x increase in computational resources will allow us to investigate larger and more complex systems. In the protein translocation problem, additional molecular motors are known to participate in the functioning of the Sec translocon, but they are neglected due to computational cost; the availability of more resources will allow us to perform new studies of the role of these collaborating macromolecules. Similarly, although PCET is observed in some relatively small organometallic systems, it is also central to the chemistry of large systems such as cytochrome c and photosystem II. With improved resources, these new application areas will be brought within reach.

10.2.5.4 Computational and Storage Requirements Summary

	Current (2009)	**In 2013**
Computational Hours	3 M	30 M
Parallel Concurrency	2K – 10K	10 K – 100 K
Wall Hours per Run	128 – 5K	5 K – 50 K
Aggregate Memory	6 – 12	12 – 24
Memory per Core	40 – 1024 GB	1 – 10 TB
I/O per Run	1 – 2 GB	1 – 2 GB
On-Line Storage Needed	4 GB	40 GB
Data Transfer	2 GB per job	8 GB per job
	20 GB/week	60 GB/week
Archival Storage	10 GB	50 GB

10.2.5.5 Support Services and Software

In order to develop such applications on "50X HPC resources" more CPU hours will be required with better throughput for medium-size jobs (queue policies) and better access to debug queues (i.e., larger core counts and longer run times) to allow for scaling studies of new algorithms, especially for parallel optimization algorithms. In addition, error checking and recovery service implementation will be essential for the larger long-running jobs.

10.2.5.6 Emerging HPC Architectures and Programming Models

MD, PIMD, and electronic structure calculations are proving amenable to architectures that employ GPUs. We are pursuing a research program to effectively employ GPUs in our calculations. In many cases, it appears that utilization of GPUs can be achieved without extensive modification of our codes and algorithms, given that our methods typically require large numbers of independent operations. Availability of test systems utilizing the new HPC architectures at NERSC will be very helpful, since it would allow consolidation and coordination of such efforts.

10.2.6 Computational Reacting Flow with Detailed Kinetics

Principal Investigator: Habib Najm, Sandia National Laboratories
Contributors: Jens Prager, Jaideep Ray, Cosmin Safta, Bert Debusschere, (Sandia National Laboratories)
Omar Knio, Johns Hopkins University
Mauro Valorani, University of Rome
Dimitris Goussis, National Technical University of Athens
Michael Frenklach, University of California at Berkeley
NERSC Repos: m914, m401

10.2.6.1 Summary and Scientific Objectives

The goal of this work is to advance the state of the art in understanding of chemically reacting flow. We focus on computations of flames with detailed chemical kinetic models for the oxidation of hydrocarbon fuels, where the objective is analyzing and understanding reacting flow structure. Our work provides improved understanding of the detailed structure of hydrocarbon flames and their interaction with transport processes in two-dimensional (2D) laboratory scale flows. Such understanding is important for building simplified flow-flame interaction models for more complex combustion systems.

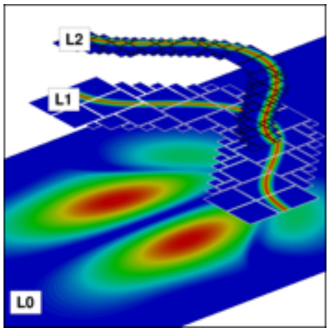


Figure 10-13. AMR computations of stoichiometric methane-air flame-vortex interaction using three mesh levels (L0,L1,L2). Methane-air chemistry is modeled using the GRI-Mech v3.1 chemical mechanism. Shown are the vorticity field on mesh level L0 and the HCO mass fraction field

The computational challenges in this work are driven by the complexity and stiffness of chemical models, and by the large range of length and time scales in these flows. We use highly resolved spatial meshes in 2D, low Mach number constructions, and operator-split time integration with implicit time integration of stiff chemical source terms. We have had an ongoing focus on the development and demonstration of scalable high-order adaptive mesh refinement (AMR) constructions for low Mach number reacting flow computations, addressing the length-scale range challenge. We focus on high-order spatial derivative stencils, interpolations, and filters, which are key to achieving scalable AMR performance on massively parallel computational platforms. We also develop and use computational singular perturbation (CSP) tools for analysis of reacting flow computational data, chemical model reduction, adaptive chemistry, and efficient explicit time integration of stiff chemical kinetic systems.

10.2.6.2 Methods of Solution

We use low Mach number formulations of the reacting flow equations. We solve the low Mach number Navier-Stokes equations using projection methods, involving an elliptic solve for the pressure field. This momentum solver is coupled with operator-split time integration of the species and energy equations. Operator splitting is employed to enable the use of optimized time integrators for transport and chemical source terms. Specifically, we use combinations of Runge Kutta (RK) and RK Chebyshev (RKC) methods for the transport terms, while using implicit backward difference formulae for time integration of the chemical source terms. Overall, our time integration constructions are second-order accurate.

We use both uniform and adaptive mesh constructions for spatial discretizations. In the uniform mesh context, we use second-order, centered finite-difference discretizations, in our "DFLAME" code. The projection scheme for the momentum equations employs a Fast Fourier Transform pressure solver. The code uses hybrid MPI/OpenMP parallelism. In the adaptive mesh context, we employ a block structured adaptive mesh hierarchy composed of uniform mesh patches. This code "CFRFS" is implemented in the context of the Common Component Architecture (CCA) framework. We use a combination of cell-centered/cell-averaged discretizations to achieve a fourth-order pressure projection algorithm for the Navier-Stokes equations on the base uniform mesh. We use preconditioned multigrid methods to solve the resulting system. We rely on centered fourth-order discretizations on the adaptive mesh hierarchy for the species and energy equations, employing high-order interpolations and filtering operations for prolongation/restriction operations at mesh patch boundaries. The overall construction is fourth-order in space.

10.2.6.3 HPC Requirements

We typically run our computations on Franklin, utilizing up to 8,000 cores in scalability testing, and up to 2,000 cores in production runs.

Our uniform mesh computational runs **today** employ meshes on the order of 1024x2048. We have run these studies focusing on 2-D edge flames employing detailed n-heptane oxidation kinetics, with up to 560 species in the chemical model. Our AMR computations have used 512x512 grid cells on the coarsest mesh level. Typically, about 25% of the computational domain uses mesh refinement, with two levels of refinement in the flame regions. We have used these computations to study 2-D premixed methane-air flames employing the GRI-Mech v3.0 kinetic mechanism (53 species). Generally, our computations have used up to 24 GB of memory per run, with each run taking up to 40 wall hours.

In the future our need to study more complex laboratory-scale reacting flows with detailed hydrocarbon kinetics will require significant enhancements in computational resources. In the next 3-5 years we plan to extend our current computations to include both steady and unsteady laminar lifted jet flames. We will target these flows initially with methane-air chemistry, moving to n-heptane, and iso-octane fuels. With this planned increase in physical domain size and chemical complexity, we expect significant increases in our computational requirements. In order to achieve significant speedup, we aim to run on up to 100,000 cores. This will require significant attention to load balancing and communication loads. We are presently at about 50% parallel efficiency at

8,000 cores with our uniform mesh code, employing strong scalability testing. The scalability of our AMR code at 1,000s of cores has yet to be studied. Clearly, we will have to identify key bottlenecks to speedup, and eliminate them in order to achieve meaningful scalability in the 100K core range. We expect these more complex computations to require up to 200 wall hours per run, and to use up to 2 TB of aggregate memory per run.

Access to 50x resources would allow us to: study n-heptane flames with domain sizes and flow topologies comparable to those we typically use for methane flames. This would enable *parametric* runs with detailed n-heptane models in laboratory-scale flames, allowing significant opportunities for understanding flame structure and dynamics. Yet, as indicated above, this will require immediate attention to load balancing bottlenecks in order to move beyond the 10K core range.

10.2.6.4 Computational and Storage Requirements Summary

	Current (2009)	**In 2013**
Computational Hours	150 K (55 at NERSC)	35 M
Parallel Concurrency	500-2,000	100,000
Wall Hours per Run	20-40	200
Aggregate Memory	24 GB	1-2 TB
Memory per Core	0.24 GB	0.9 GB
I/O per Run	20 – 100 GB, 0.2-6 GB checkpoint files	1 – 10 TB, 100-250GB checkpoint files
On-Line Storage Needed	30 GB	1.5 – 3 TB
Data Transfer	---	15 TB

10.2.6.5 Support Services and Software

We currently use codes written in both C++ and Fortran and rely on a number of external libraries, including both Chombo and GrACE for management of block-structured adaptive mesh refinement; *hypre* for high performance preconditioners and parallel multigrid methods; LAPACK for linear algebra; and DVODE/CVODE solvers (part of *sundials*) for time integration of ordinary differential equations. We have also relied on *craypat*. One of our code bases relies on the Common Component Architecture. Although in principle, this needs support for dynamic libraries, we have figured out ways around this when needed. Still, it's certainly preferred to have support for dynamic libraries.

10.2.6.6 Emerging HPC Architectures and Programming Models

While we are aware of the limitations of the current MPI/OpenMP parallelism paradigm, we have not yet explored the use of emerging HPC architectures, such as GPUs. Potentially, we envision a model where computationally intensive chemistry calculations are spawned to the GPU cores available on future supercomputing architectures.

Appendix A. Attendee Biographies

Brian Austin works with Jonathan Wertele's research group at the Berkeley Lab to run simulations for developing the Next Generation Light Source. As an undergraduate student at Reed College in Portland, Ore., Austin developed a strong interest in quantum chemistry. As a graduate student at U.C. Berkeley, Austin focused on enhancing the Quantum Monte Carlo (QMC) approach to computing molecular properties. He learned about high performance computing and parallel programming by helping to rewrite his research group's QMC code and running these codes on large-scale supercomputers at NERSC.

Hai-Ping Cheng is Professor of Physics and Chemistry at the University of Florida and a member of the Quantum Theory Project, the world's largest research center for theory, modeling, and computation of complex, novel molecular and materials systems. Her main research interest is in simulation of properties of large clusters and surface effects. She holds a Ph.D. from Northwestern University and was a visiting, guest, or research scientist at Los Alamos National Laboratory, NIST, and IBM Zurich.

Peter Cummings is the John R. Hall Professor of Chemical Engineering, Professor of Chemical & Biomolecular Engineering in the Department of Chemical & Biomolecular Engineering at Vanderbuilt University, and Principal Scientist, Center for Nanophase Materials Sciences, Oak Ridge National Laboratory. B. Math. (First Class Honors) University of Newcastle (Australia), 1976; Ph.D., Math. University of Melbourne (Australia), 1980

Randall T. Cygan is a Senior Member of the Technical Staff at Sandia National Laboratories. His research interests are varied, including investigations of mineral equilibria, chemical kinetics, surface chemistry of minerals, sorption and dissolution of minerals, shock metamorphism, and atomistic modeling of minerals and geochemical processes. He holds a Ph.D. from Pennsylvania State University.

James Davenport joined the Materials Sciences and Engineering Division in 2011 after many years at Brookhaven National Laboratory, where he served as Director of the Computational Science Center, Chair of the Department of Applied Science, and Associate Chair of the Physics Department. He is a Fellow of the American Physical Society and served as Chair of the Division of Materials Physics. He is co-author on more than 110 publications in areas such as density functional theory, electronic structure of metallic alloys, first principles molecular dynamics, hydrogen ab/adsorption, photoemission, and inverse photoemission.

Thomas Devereaux is currently a professor in the Photon Science Faculty at SLAC National Accelerator Laboratory and Stanford University and a member of the Pacific Institute for Theoretical Physics (PITP). He is the head of the X-ray Science and Techniques Group at the Stanford Institute for Materials and Energy Sciences (SIMES). SIMES is a joint institute between Stanford main campus and SLAC, a national laboratory, focusing on scientific foundation related to the energy challenge facing our society. He develops numerical methods and theories of photon-based spectroscopies of strongly correlated materials. Professor Devereaux received his Ph.D. in Physics from the University of Oregon in 1991, M.S. from University of Oregon in 1988, and B.S from New York University in 1986.

Andrew Felmy is a Laboratory Fellow in the Chemical Sciences Division at PNNL, Chief Scientist for Scientific Programs at the WR Wiley Environmental Molecular Sciences Laboratory. Dr. Felmy originally joined Battelle in March of 1980 as a Grade I Scientist. In 1982 he was promoted to Research Scientist and in 1983 went on educational leave to obtain his PhD in theoretical physical chemistry at the University of California, San Diego under Professor John H. Weare. Dr. Felmy received his PhD in 1988 and returned to Battelle. Since returning to Battelle, he has been promoted successively to Grade III Senior Research Scientist, Grade IV Staff Scientist, Technical Group Leader, Technical Group Manager of the Thermodynamic and Molecular Geochemistry Group, and is now the Acting Associate Director for Environmental Dynamics and Simulation in the EMSL. His research interests are in thermodynamics of Aqueous Electrolytes; development of Geochemical Models; thermodynamics of Surface Complexation Reactions; and utilization of molecular dynamics models for estimation of thermodynamic properties.

Anthony Ladd is Professor of Chemical Engineeringat the University of Florida where his research interests focus on the application of numerical simulations to complex systems containing a wide range of length scales and time scales. The goal is to model the essential physics of the problem in as simple and fundamental a way as possible.

Thomas Miller attended Texas A&M University as a major in chemistry and mathematics, where he graduatedwith honors in 2000 having won a number of awards and scholarships, including the Barry Goldwater Scholarship, the Beckham Award in Science, and the TAMU undergraduate mathematics competition. He received a British Marshall Scholarship to pursue graduate study in the U.K., which he used to obtain an M. Phil. from University College London in 2002. He then attended the University of Oxford on an NSF graduate research fellowship, earning a D. Phil. from Balliol College in 2005. His thesis work, which was performed in the research group of David Clary and entitled Quantum Simulation of Biological Molecules, was recognized with the Jowett Senior Exhibition Prize, the highest award for graduate scientific research from Balliol College. Tom returned to the U.S. for a joint postdoc with Bill Miller and David Chandler at UC Berkeley, where he joined the Helios project to study the fundamental science of solar energy conversion. He joined the Caltech faculty in June 2008. He has since received a Dreyfus New Faculty Award , a Sloan Research Fellowship, an NSF CAREER Award, and the American Chemical Society Hewlett-Packard Outstanding Junior Faculty Award.

Burkhard Militzer was born 1970 in Dresden, Germany. He received a diploma in physics from the Humboldt University at Berlin in 1996, and Ph.D. from the University of Illinois at Urbana-Champaign in 2000. After three years as post-doc at Lawrence Livermore National Lab, he worked for four years as associate staff member at the

Geophysical Laboratory of Carnegie Institution of Washington. In 2007, he joined UC Berkeley where he is Assistant Professor of Astronomy and
Assistant Professor of Earth and Planetary Science

Normand Modine is a Principal Member of the Technical Staff in the CINT Science Department at Sandia National Laboratories in Albuquerque, NM. He has sixteen years of experience in developing and applying computational techniques to understand the behavior of materials. His primary research interests are using computational techniques to research energy transfer processes, the dynamical behavior of defects, interfaces, and surfaces, and new methodologies for bridging length and time scales in nanoscale and nanostructured systems. He holds a Ph.D. and A.M. degrees in physics from Harvard University.

James Muckerman is a Senior Chemist in the Chemistry Department at Brookhaven National Laboratory. He holds a Ph.D. from University of Wisconsin. His research interests are in Theoretical dynamics of reactive collisions and photodissociation processes; Laser control of molecular dynamics; Grid methods in quantum mechanics; Combined quantum/classical approaches for treating the dynamics of large systems; Calculation of potential energy surfaces; Electronic properties of metal-containing clusters; Nanocatalysis.

Habib Najm is a Principal Member of the Technical Staff at Sandia National Laboratories in Livermore, CA. He received the MS and PhD degrees in Mechanical Engineering from MIT in 1986 and 1989, and the BE degree in Mechanical Engineering from the American University of Beirut in 1983. Before joining Sandia in 1993, Najm worked with the Semiconductor Process Design Center at Texas Instruments, on the development of sensors and control in semiconductor processing, and on design studies of thermofluid systems in semiconductor process technology. Najm's group at the Sandia Combustion Research Facility is involved in a range of computational reacting flow research funded by the US Dept. of Energy, Basic Energy Sciences/Chemical Sciences Division and the DOE SciDAC Computational Chemistry program. This work spans the development of algorithms for time integration and uncertainty quantification, distributed high-performance component software implementations, and computational studies of reacting flow with detailed hydrocarbon kinetics. The group also works on DARPA-funded research focused on detailed modeling and uncertainty quantification in electrochemical microfluid systems. Najm is co-author of over thirty archival journal articles and four US patents.

Jeffrey Neaton Jeffrey B. Neaton leads the Theory group at the Molecular Foundry in LBNL. Jeff received his Ph.D. in Physics from Cornell University in 2000, under the guidance of Neil W. Ashcroft. After a departmental postdoc in the Department of Physics and Astronomy at Rutgers University, he joined the Molecular Foundry at Lawrence Berkeley National Laboratory in 2003. His current research interests center on computational nanoscience, in particular the development and application of methods for calculating the structural, spectroscopic, and transport properties of inorganic and molecular nanostructures, particularly at interfaces and contacts. Present areas of interest include the electronic properties of the metal-organic interface, hybrid silicon-organic interfaces, and single-molecule junctions; self-assembly; nanoparticle assemblies; photovoltaics; hydrogen storage; ultrathin epitaxial films of transition metal oxides, such as ferroelectrics and multiferroics; and structural and electronic phases of light elements under pressure.

Greg Newman is a Senior Scientist at Lawrence Berkeley National Laboratory, Earth Science Division and Head of the Geophysics Department in the Earth Sciences Division. Prior to his appointment in January 2004, Dr. Newman worked nearly fourteen years at Sandia National Laboratories, Geophysical Technology Department. His interest, include large-scale, multi-dimensional, inverse and forward modeling problems arising in exploration geophysics, parallel computation and electromagnetic geophysics. He has over 20 years of experience in large-scale geophysical field simulation and computation. In 2000, Dr Newman was a Mercator Fellow at the Institute for Geophysics and Meteorology, University of Cologne, Federal Republic of Germany. The fellowship was awarded from the German National Science Foundation for a year of study in the Federal Republic of Germany. Studies at the Institute were directed on the formulation and implementation of 3D transient electromagnetic modeling and inversion algorithms for geophysical applications and lectures on the electromagnetic modeling and inversion. Dr. Newman was also affiliated with this institution from 1987-1989 as a Post Doctorate Appointee and an Alexander von Humboldt Fellow.

Mark Pederson is the program manager for Theoretical and Computational Chemistry. From 1996 until joining DOE, he was the section head in the “Theory of Molecules, Clusters and Nanoscale Devices” section at the Naval Research Laboratory (NRL). He also spent one year at Max-Planck-Institute (1992) and one year at NSF (2002) as a program director in Theoretical and Computational Chemistry. In 1986, Dr. Pederson joined the NRL as a National Research Council (NRC)-NRL postdoctoral researcher; he was hired as a permanent employee in 1988. Dr. Pederson’s research has concentrated on the development, testing, and application of massively parallel electronic structure methods for the density-functional-based simulation of molecules and clusters. He is a fellow of the American Physical Society and a member of the American Chemical Society and the Materials Research Society. Dr. Pederson has coauthored over 185 papers in the areas of electronic structure, vibrational spectroscopy, chemical-vapor deposition, fullerenes, molecular magnetism, and organic photovoltaics. He holds a Ph.D. in Theoretical Physics from University of Wisconsin.

G. Malcolm Stocks is a Lockheed Martin Corporate Fellow in the Theory Group of ORNL's Metals and Ceramics Division. He holds a Ph.D. degree in condensed matter physics from Sheffield University. Prior to joining ORNL in 1976, he was a lecturer at the University of Bristol, U.K., from 1972 to 1976. His research interests include materials science, high-performance heterogeneous scientific computing, and the development of scalable first-principles electronic structure algorithms and parallel algorithms.

Lin-Wang Wang received his BS in Physics from Shanghai Jiaotong University in 1985 and PhD in Solid State Physics from Cornell University in 1991. He worked at the National Renewable Energy Laboratory as a postdoctoral fellow and then a staff scientist from 1992 to 1999. Since 1999, he has been a staff scientist at Lawrence Berkeley National Laboratory. His main research interest is in material science simulations especially for semiconductor and nanoscale systems. He has developed various computational methods for thousand to million atom large scale electronic structure calculations. He is a recipient of the Chinese overseas outstanding young scientist award, and the 2008 ACM Gordon Bell award. He is a fellow of American Physical Society.

Appendix B. Workshop Agenda

Tuesday, February 9		
Time	Topic	Presenter
9:00	Join teleconference and WebEx conference	
9:30	Welcome, introductions, workshop goals, charge to committee	Yukiko Sekine, DOE-SC/ASCR
9:45	Workshop outline, logistics, format, procedures	Richard Gerber, NERSC
10:00	BES Program Office Research Directions	Mark Pederson, DOE / BES
10:30	NERSC Role in Basic Energy Sciences Research	Kathy Yelick, NERSC Director
11:30	Break	
11:50	Case Studies: Geosciences DOE POC: Nicholas Woodward	Andrew Felmy Greg Newman Tony Ladd
13:00	Case Studies: Density Functional Theory, DOE POC: Jim Davenport	G. Malcolm Stocks Hai-Ping Cheng Lin-Wang Wang
14:25	Break	
14:50	Case Studies: Quantum Monte Carlo DOE POC: Mark Pederson	Burkhard Militzer Mark Jarrell Brian Austin
16:10	Case Studies: Spectroscopy and Light-Matter Interactions DOE POC: Jim Davenport	Thomas Devereaux Jeffrey Neaton
17:40	Case Studies: Classical Molecular Dynamics Codes and Coupling of Length Scales DOE POC: Mark Pederson	Peter Cummings Normand Modine Randy Cygan
18:40	Case Studies: Chemical Physics Codes DOE POC: Mark Pederson	James Muckerman Thomas Miller Habib Najm
19:30	Adjourn for the day	

Wednesday, February 10		
9:30	Join teleconference and WebEx conference	
10:00	Summary of previous day's discussions	Harvey Wasserman
10:30	Case study format review; sample case study	Richard Gerber
10:50	Report schedule and process	Harvey Wasserman,
11:00	Q&A and general discussion	
11:30	Breakout session teleconferences	
12:30	Adjourn	

Appendix C. Abbreviations and Acronyms

AIMD	*Ab initio* Molecular Dynamics
ALCF	Argonne Leadership Computing Facility
AMR	Adaptive Mesh Refinement
ASCR	Advanced Scientific Computing Research
BDT	Benzenedithiol
BES	Basic Energy Sciences Office
BESAC	Basic Energy Sciences Advisory Committee
BLAS	Basic Linear Algebra Subroutines
BNL	Brookhaven National Laboratory
BSE	Bethe-Salpeter Equation
CHARMM	Chemistry at HARvard Molecular Mechanics
CMD	Classical Molecular Dynamics
CPM	Carr-Parrinello Method
CPMD	Carr-Parrinello Molecular Dynamics
CPU	Central Processing Unit (or processor)
CUDA	Compute Unified Device Architecture
DFT	Density Functional Theory
DMC	Diffusion Monte Carlo
DQMC	Determinant Quantum Monte Carlo
ECP	Effective Core Potentials
ED	Exact Diagonalization
EM	Electromagnetic
EPLF	Electron Pair Localization Function
ESnet	DOE's Energy Sciences Network
FFT	Fast Fourier Transform
FNAL	FermiLab National Accelerator Laboratory
FPGA	Field Programmable Gate Array
GAMESS	General Atomic and Molecular Electronic Structure System
GB	GigaByte
GPU	Graphical Processing Unit
GROMACS	GROningen MAchine for Chemical Simulations
GSL	GNU Scientific Library
GW	
HOMO	Highest Occupied Molecular Orbital
HPC	high-performance computing
HPSS	High Performance Storage System
I/O	Input Output
IDL	Interactive Data Language visualization software
INCITE	Innovative and Novel Computational Impact on Theory and Experiment
LANL	Los Alamos National Laboratory

LAMMPS	Large-scale Atomic/Molecular Massively Parallel Simulator
LB	Lattice Boltzmann (method)
LBNL	Lawrence Berkeley National Laboratory
LCLS	Linac Coherent Light Source
LDA	Local Density Approximations
LLNL	Lawrence Livermore National Laboratory
LS3DF	Linear Scaling 3-Dimensional Fragment (method)
MC	Monte Carlo (method)
MD	Molecular Dynamics
MO	Molecular Orbital
MPI	Message Passing Interface
MIMD	Multiple Instruction Multiple Data
MRT	Multi-Relaxation Time
NAMD	Not (just) Another Molecular Dynamics program
NERSC	National Energy Research Scientific Computing Center
NGF	NERSC Global Filesystem
NIH	National Institutes of Health
NSF	National Science Foundation
NSLS	National Synchrotron Light Source
OLCF	Oak Ridge Leadership Computing Facility
ORNL	Oak Ridge National Laboratory
OS	Operating System
PCET	Proton-Coupled Electron Transfer
PCM	Polarizable Continuum Model
PIMD	Path Integral Molecular Dynamics
PNNL	Pacific Northwest National Laboratory
QMC	Quantum Monte Carlo
SC	DOE's Office of Science
SCF	Self Consistent Field
SciDAC	Scientific Discovery through Advanced Computing
SLAC	SLAC National Accelerator Laboratory
SMP	Symmetric Multiprocessor, Shared Memory Multiprocessor
SNL	Sandia National Laboratories
TB	TeraByte
TDDFT	Time-Dependent Density Functional Theory
UPC	Unified Parallel C programming language
VASP	Vienna Ab-initio Simulation Package
XML	Extensible Markup Language

Appendix D. About the Cover

Left top: Results from a Quantum Monte Carlo calculation. The image shows a "One-particle slice" through an explicitly correlated many-body wave function for a derivative of the retinal chromophore. Coordinates for all but one electron were sampled from the many-body density and plotted as maroon (spin-up) and teal (spin-down) spheres. The red and blue surfaces are the 0.0001 isosurfaces of the wave function for the remaining (spin-up) electron. Nodes of the wavefunction (not shown) determine the accuracy of diffusion Monte Carlo and can be found between the isosurfaces. Image courtesy of Brian Austin, NERSC.

Middle top: A snapshot from a molecular dynamics simulation of 18 gold atoms placed on the surface of an alkanethiol self-assembled monolayer on the Au(111) surface. See Sabri Alkis, Chao Cao, Hai-Ping Cheng and Jeffrey L. Krause, "*Molecular Dynamics Simulations of Au Penetration through Alkanethiol Monolayers on the Au(111) Surface*," J. Phys. Chem. C, 2009, 113 (16), pp 6360–6366. Image courtesy of Hai-Ping Cheng and Eric Wang, University of Florida.

Right: A portion of the exciton wavefunction (the white isosurface) at the interface of a ZnS / ZnO nanorod. Simulations performed on a Cray XT4 at NERSC, also shown. Image courtesy of Lin-Wang Wang, LBNL. See Lin-Wang Wang, "*Computational challenges for nanostructure solar cells*," Energy & Environmental Science, 2009, 2, 944–955.

Middle bottom: A portion of a polar-coordinates plot of differential cross sections for the $O(^3P)+OH(^2\Pi) \rightarrow H(^2S)+O_2(^3\Sigma^-)$ reaction at 150meV of collision energy. See Jianyi Ma,Shi Ying Lin, Hua Guo, Zhigang Sun, Dong H. Zhang, and Daiqian Xie3, "*State-to-state Quantum Dynamics of the $O(^3P)+OH(^2\Pi) \rightarrow H(^2S)+O_2(^3\Sigma)$ Reaction*," Journal of Chemical Physics 133, 054302, August 7, 2010 (cover story). Image courtesy of Hua Guo, University of New Mexico.

Left bottom: A portion of the resulting image from a marine prospecting study with joint controlled source electromagnetic (CSEM) and magnetotelluric (MT) fields imaging of the Compos Basin offshore Brazil. The image demonstrates the ability to distinguish between a known oil field, a possible hydrocarbon trap, and brine. See Gregory A Newman and Michael Commer, "*Massively Parallel Electrical Conductivity Imaging of the Subsurface: Applications to Hydrocarbon Exploration,*" Journal of Physics: Conference Series 180 (2009) 012063. Image courtesy of Michael Commer, LBNL.

NERSC

U.S. DEPARTMENT OF
ENERGY
Office of Science

BERKELEY LAB
Lawrence Berkeley
National Laboratory