Molecular Science Computing Facility

The Molecular Science Computing Facility (MSCF) supports a wide range of computational activities in environmental molecular research, from benchmark calculations on small molecules to reliable calculations on large molecules, and from solids to simulations of large biomolecules and reactive chemical transport modeling. The MSCF provides an integrated production computing environment with links to external facilities within the U.S. Department of Energy (DOE), collaborating universities, and industry.

The MSCF provides computational resources for Computational Grand Challenge projects in environmental molecular science and basic and applied research areas that address the environmental problems and research needs facing the DOE and the nation. The Grand Challenge projects typically involve multiple investigators from universities, national laboratories, and industry working collaboratively as teams. These projects are usually granted allocations for three years, based on a fiscal year (FY). The MSCF supported nine three-year Grand Challenge projects during FY 2003. The average annual computer time allocation for Grand Challenge projects was 230,000 CPU hours. This average was down about 50% from previous years due to the new computer upgrade transition. The new system will have approximately four times the available hours over the previous IBM system.

The MSCF also supports smaller, short-term projects called MSCF Pilot Projects. The Pilot Projects are limited to a maximum of 75,000 processor hours and for a duration of one year. The MSCF supported 32 Pilot Projects during FY 2003 with an average allocation of 31,000 CPU hours. The Pilot Projects are typically directed at developing the capability to submit a Grand Challenge proposal in the future (e.g., a combination of theory/method or code development activities or calculations that provide the initial scientific basis of a Grand Challenge proposal).

The MSCF high-performance computing systems had 414 total users during FY 2003, of which 59% were external to Pacific Northwest National Laboratory (PNNL) and 41% were PNNL staff, postdoctoral fellows, or students. The distribution of the users by

Instrumentation & Capabilities

- **MPP2**. Production cluster of 980 HP rx2600 nodes, 1960 1.5 gigahertz IA64 processors, 450 terabytes local disk, 6.8 terabytes memory, 11.8 teraflops
- **Lustre**. Shared cluster filesystem, 53 terabytes
- **NWfs**. EMSL long-term data store, 85 terabytes
- **Network**. OC12 (600 MBit/sec) internet connection, Gigabit Ethernet MSCF backbone
- Graphics & Visualization Laboratory
- **NWVisus**. Visualization server, SGI Onyx 3400Graphics, 8 processors, 16 GB RAM, 2 Infinite Reality3 pipes, 144 GB disk, with a PanoramTech 3-screen monitor
- **Lionheart**. Visualization server, SGI Onyx, 8 processors, 4 Infinite Reality pipes, 40 GB disk, Serius digital video
- **Ogrady**. Parallel graphics Linux Dell cluster, 8 - 1.7 GHz Pentium3 processors with 16 GB RAM, 80 GB disk, gigabit switch, with Scalable Graphics Engine (SGE), IBM T221 high resolution display
- Digital video editing suite
- Electro Home stereo projection system
- Access GRID internet node
- Molecular Science Software Suite

type of institution (illustrated in Figure 1) is 46% from universities, 41% from PNNL, 10% from other DOE laboratories, 2% from industry, 1% from foreign national laboratories, and 0% from other government agencies.

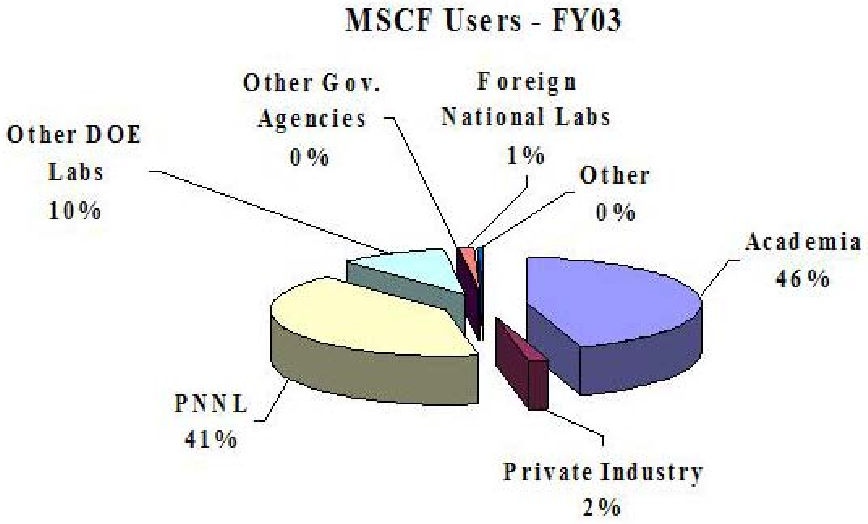

Figure 1. MSCF resource allocation by user affiliation.

User Computing in the MSCF

To optimally address the complex environmental problems facing DOE and the nation and to best utilize limited staff resources, the W.R. Wiley Environmental Molecular Sciences Laboratory (EMSL) follows the recommendation of the EMSL Science Advisory Committee to use a computational Grand Challenge approach to providing large blocks of resources to the user community. A call for proposals is issued annually, and teams of computational scientists respond with peer-reviewable proposals for system time allocations. Access for the Grand Challenge teams is for periods of one to three years. The following criteria are used when reviewing proposals:

- scientific merit
- appropriateness of the proposed method or approach
- relevance to the environmental problems and research needs of DOE and the nation
- technical competence of the investigators
- reasonableness and appropriateness of the proposed computer resources.

The process for the review of proposals and allocation of MSCF resources is shown in Figure 2. The request for proposals is open to all researchers, regardless of their institution or source of funding. For reference, the second request for Grand Challenge proposals involved 38 external scientific reviewers from leading universities and research institutions from around the world. Proposals received three reviews on average, and every proposal received at least two reviews.

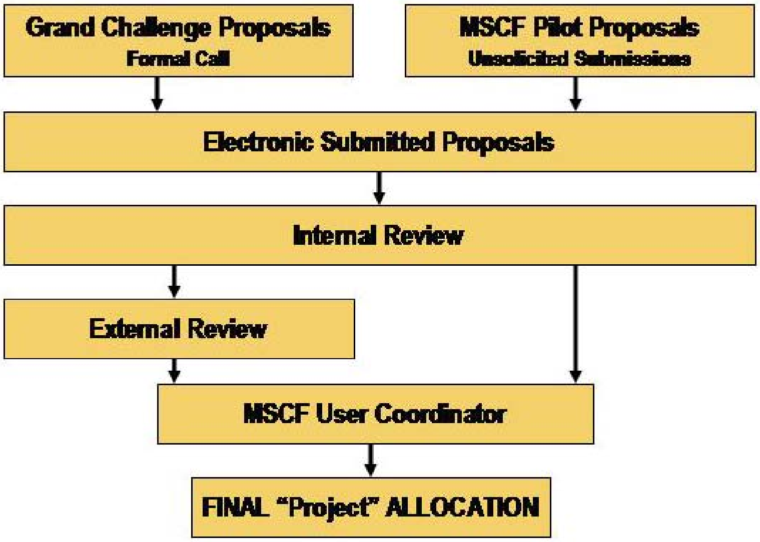

Figure 2. Review process and allocation proposal flow chart.

MSCF Resources

The MSCF provides a combination of production computing hardware and software resources and visualization tools to support the scientific research activities of the Computational Grand Challenge and EMSL Pilot Projects. The hardware and visualization resources are:

- The High-Performance Computing System-2 (MPP2), which provides a balanced supercomputer that has been tailored to meet the operational needs of EMSL users. The Hewlett-Packard (HP) supercomputer became operational in July 2003 with a theoretical peak performance of 11.8 teraflops with 6.8 terabytes of RAM and 256 terabytes of disk. The system allows EMSL users to perform remote visualizations and transfer data to remote storage via multiple Gigabit Ethernet connections.
- The EMSL's archive, which is located inside the MSCF. The archive uses a unique approach to disk storage by clustering many low-cost commodity disks to provide fault-tolerant, high-performance storage. The current EMSL archive has more than 35 terabytes of storage available to users and the ability to grow to over a petabyte. The unique software MSCF developed to manage the large pools of clusters disks, called NWfs, is made freely available to EMSL users.
- A Graphics and Visualization Laboratory (GVL), which provides production graphics and visualization facilities for the display and analysis of complex datasets from both experiments and simulations. The GVL contains four high-performance graphics stations based on Silicon Graphics Incorporated (SGI) technologies with high-speed Gigabit Ethernet connections to the production supercomputers and to the EMSL archive system in the MSCF, a digital video system integrated with the workstations to facilitate the display and capture of scientific data, and digital video editing equipment for the preparation of scientific presentations. The video system also is connected to the EMSL auditorium and to the World Wide Web to facilitate internet conferencing. An IBM Scalable Graphics Engine (SGE) is connected to the MSCF Dell Linux Cluster to provide the highest performance visualization capabilities.

Software resources include the Molecular Science Software Suite (MS3), a comprehensive, integrated set of tools that enables scientists to understand complex chemical systems at the molecular level. MS3 couples the power of advanced computational chemistry techniques with existing and rapidly evolving high-performance massively parallel computing systems with extensible problem-solving capabilities. The suite consists of three components: 1) the Northwest Computational Chemistry Software (NWChem), 2) the Extensible Computational Chemistry Environment (Ecce), and 3) the Parallel Software Development Tools (ParSoft).

NWChem Version 4.5. Version 4.5 of NWChem was released in July 2003. Full documentation and information is available on the NWChem website (http://www.emsl.pnl.gov/docs/nwchem). NWChem provides many methods to compute the properties of molecular and periodic systems by using standard Gaussian and plane-wave-based quantum mechanical descriptions of the electronic wave function or density. In addition, NWChem can perform classical molecular dynamics (MD) and free-energy simulations. These approaches may be combined to perform mixed quantum-mechanics and molecular-mechanics simulations.

NWChem is available on almost all high-performance computing platforms, workstations, PCs, and clusters of desktop or workgroup servers. NWChem development provides maximum efficiency on massively parallel processors; however, it has not been optimized for high-performance on single-processor desktop systems.

Extensible Computational Chemistry Environment (Ecce). Ecce is composed of a suite of distributed client/server UNIX-based X Window System applications that are seamlessly integrated. The resulting environment enables research scientists to transparently use complex computational modeling software accessing high-performance computing resources from their desktop workstations. There are six major components:

- The Calculation Manager, which aids in the organization and manipulation of computational chemistry studies. This tool allows an at-a-glance overview of the status of every calculation and easy access to key setup parameters and run statistics.
- The Molecule Builder, which is an intuitive, point-and-click tool that enables the building, visualization, modification, and manipulation of three-dimensional images of chemical systems.
- The Basis Set Tool, which enables the choice from over 230 predefined Gaussian basis sets or the ability to create new ones for use in *ab initio* electronic structure calculations.
- The Calculation Editor, which allows the user to choose input options using point-and-click and then generates the chemistry code-specific input.
- The Job Launcher, which is a tool for submitting a calculation to a computer for processing. The user may submit a calculation to any computer that has been registered as a compute server where the user has an account.
- The Calculation Viewer, which provides convenient access to current information for a single calculation during execution or after completion. It has many features for viewing and visualizing chemical system properties.

Parallel Software Tools and Libraries (ParSoft). ParSoft (also known as the Global Arrays [GA] toolkit) includes high-performance computing libraries and tools for applied parallel computing focused on interprocessor communications through the aggregate remote memory copy interface, high-performance input/output through the Parallel I/O tools and programming models for hierarchical memory systems through the GA and Memory Allocator tools. One of the primary goals for this work is to aid development of real scientific application codes.

Development of ARMCI has been supported by EMSL operations, the DOE2000 Global Arrays project, and recently by the Center for Programming Models for Scalable Parallel Computing.

MSCF Organization

The MSCF is divided into three groups: the Scientific Consulting Group, the Molecular Science Software Group, and the Computer Operations Group.

The Scientific Consulting Group provides an extremely diverse set of services for all users of the MSCF high-performance computers and the GVL. Scientists who need access to high-end computing equipment frequently have difficulties getting started. Problems range from logging in to getting user codes to run efficiently. User services include handling user proposal applications, following user progress of computational projects, managing proposal reviews for both Grand Challenge and EMSL Pilot Projects, helping with user accounts, providing general consulting support for MSCF software packages, supporting and maintaining software, managing the GVL, conducting training and user workshops, developing visualization software and high-quality visualizations, and producing websites.

The group manages the Grand Challenge and Pilot Project proposal process for the MSCF. This includes receiving the proposals, providing preliminary review for EMSL missions, preparing proposal packets for external peer review, evaluating process peer reviews, granting project allocation, and managing the allocation by using the MSCF QBank.

Consultants have various roles, including those of administrator, tutor, programmer, or research scientist, and field a variety of requests for support. During 2003, five scientific consultants responded to more than 600 email requests, and about 300 additional requests were handled over the telephone or during direct office visits.

Specific services include helping new users with accounts and providing general scientific consulting support for MSCF software packages. Consultants perform technical consulting activities for high-performance computing software, parallel programming, and high-performance system testing. Consultants also work directly with MS3 development teams to give customer feedback and test functionality.

Information about the use and configuration of the MSCF computational resources is critical to the user base. MSCF information is provided efficiently to users via the Internet through the MSCF home page. The MSCF website contains all necessary information on how to get started and request accounts. The website also provides details about the compute server configuration as well as documentation and web-based tutorials for the MS3. Staff expertise in developing the MSCF website is now being used in other web page development projects at PNNL. MSCF Staff serve on many different web page development teams as web page consultants.

Scientists generate enormous amounts of data either from computational resources or from the EMSL instruments. These data are usually complex and difficult to understand. To help scientists visualize data, we manage the MSCF GVL and provide support services. The GVL contains high-performance graphics compute servers for visualizing data. To help scientists present their findings, the GVL contains state-of-the-art multimedia equipment. The real-time digital video capture capability from the graphics compute servers allows for quick yet high-quality video production. Users can generate presentation media in any form—from video (all international video standards) to web-based animations. The group also provides basic video production services.

The consulting group also conducts MS3 training workshops and MSCF workshops, demonstrations, and tours.

The Molecular Science Software Group supports MS3. This effort includes developing high-performance versions of the software and new high-performance algorithms, responding to user requests for additional features, supporting and maintaining the software, diagnosing MS3 problems associated with computer vendor hardware and software, consulting on specific MS3 problems, distributing MS3 to remote sites, porting software to new architectures, and conducting training and user workshops.

The Molecular Science Software Group consists of EMSL- and PNNL-matrixed staff focused on developing next-generation molecular modeling software for newly evolving computer technologies, especially massively parallel computers. The group comprises computational chemists and computer scientists (with external collaborations to mathematicians) who work together to develop the MS3 software. This software is used by many of the MSCF Grand Challenges and has been distributed to over 800 sites worldwide. In addition to the development activities, this group is also responsible for training software users, conducting tutorial workshops and providing user support. Toward this end, several resources have been developed to facilitate user interactions. These include a website with manuals, download information, FAQs, known bugs, tutorials, and benchmark information; as well as a support queue for answering direct user questions and user email lists. During FY 2003, the group responded to more than 1100 queue requests with an initial response time of less than four working hours.

The Computer Operations Group runs the MSCF computer systems, including all the large supercomputing systems, the experimental computing system, and the EMSL archive. The group is responsible for the installation and upgrades of the MSCF supercomputers, operating system support, improving the quality of the system software for the users, allocation tracking, interfacing directly with the vendor, and providing system security and system-level support to users.

The Computer Operations Group operates, maintains, and advances the capabilities of the MSCF scientific computing systems. The group is responsible for the operation and implementation of the various production supercomputers in the MSCF and has developed system management, monitoring, allocation management, and scheduling capabilities. The primary focus is on providing high-quality, reliable production computing cycles on the production computer in the MSCF to support very large parallel calculations for Grand

Challenge projects. This year, the infrastructure upgrade continued with the installation of the final phase of the 11.8-teraflop HP supercomputer.

The MSCF Operations Group is working hard to create and disseminate novel solutions to the infrastructure problems that are unique to high-performance technical computing. One of the major outcomes this year is the NWLinux distribution. This Linux distribution is tailored to IA64 clusters and based on the software and system management techniques that were created to successfully install and administer MPP2. NWLinux is available for download at http://www.emsl.pnl.gov/nwlinux. A similar effort is under way to distribute NWfs, a Linux distribution tailored to large-scale storage clusters, such as those used in the MSCF. The MSCF Operations group hosted the first Lustre Users Group meeting in July 2003.

MSCF staff also administer the Experimental Computing Laboratory (ECL). In the ECL, new system, storage, and interconnect technologies are explored to help determine the direction that future EMSL and MSCF systems will take. Technologies explored in the ECL during 2003 included the SGI Altix 3000, Infiniband, Myrinet, and QSNet^{II}.

Upgrades

MSCF Computational Systems Upgrades

In 2003, the new Linux Cluster from HP was delivered to the MSCF in phases, and installation was completed in July. The system configuration has a theoretical peak performance of 11.8 teraflops and was officially declared to be the fifth-fastest supercomputer in the world in the November 2003 Top 500 list of supercomputers (http://www.top500.org). Pictured in Figure 3, the new system will be upgraded to the latest interconnect switch technology from Quadrics, QsNet^{II}. This upgrade is expected to be complete by April 2004.

Figure 3. HP supercomputer

The MSCF saw its computing capability grow by an order of magnitude with the successful deployment of the final phase of the HP MPP2 Linux supercomputing cluster, (Figure 4). The Phase 1 system replaced the IBM SP system, MPP1, in late 2002. Delivery of the new Phase 2 system began in February 2003. Nearly 120 new cabinets containing compute nodes, disk, and interconnect switches for Phase 2 were received and installed in the MSCF in early 2003. MSCF staff had installed the system software in January 2003 at HP's integration facility in Roseville, California, so that the new system could be operational in the shortest possible time. Early users were allowed on the Phase 2 system in April 2003, and the system was declared fully operational in July 2003. The Phase 1 system was incorporated into the Phase 2 system in August 2003. The final system configuration has a theoretical peak performance of 11.8 teraflops.

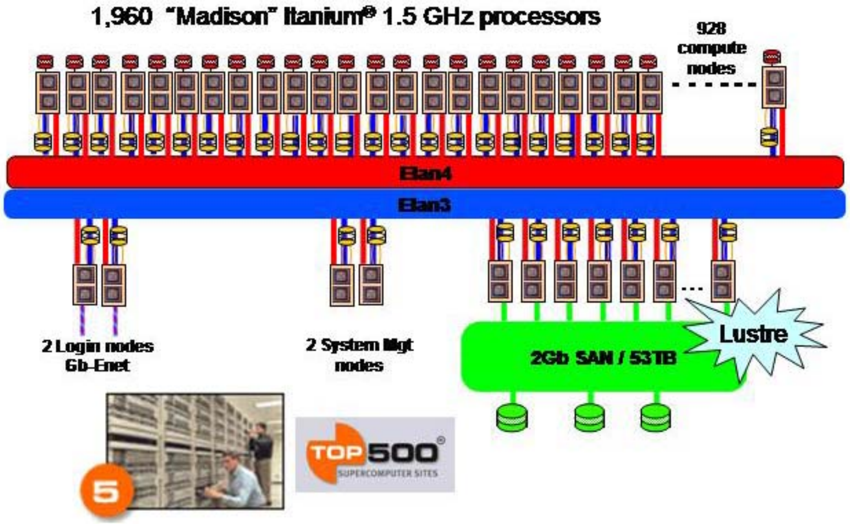

Figure 4. MPP2 Linux supercomputer cluster, 11.8 teraflops, 6.8 terabytes of memory.

The MSCF successfully implemented a 53-terabyte Lustre Lite filesystem. Lustre is a scalable, high-capacity, cluster-based parallel filesystem produced by Cluster Filesystems, Inc. Lustre has been championed primarily by DOE laboratories, but interest in Lustre is spreading throughout the high-performance and commercial communities. A high-performance shared filesystem is vital to high-performance parallel computation on MPP2, and Lustre fills this need for the MSCF. Lustre has run as fast as 3.2 gigabytes per second on MPP2.

The new storage system (NWfs) entered production in March 2003. NWfs is a Linux cluster with storage based on commodity ATA and Redundant Array of Independent Disks technology, providing high-capacity storage for 1/10th the cost of Enterprise SAN products. All of the EMSL mass spectroscopy data (nearly 15 terabytes) are stored in NWfs, and all of the data in the old scientific archive management system were successfully migrated into NWfs in May 2003. NWfs currently has 85 terabytes of raw disk, and of the 38 terabytes of that disk currently available to users, 19 terabytes are currently being used. Because of user demand, it is expected that NWfs will grow beyond 100 terabytes in 2004.

The MSCF is committed to grid computing, which will become essential to cross-site computation and data transfer. Both MPP2 and NWfs have been deployed on the DOE Science Grid, and MPP2 contributed computational cycles to Argonne's Genome Analysis and Database Update, the first grid-enabled genomics application.

NWChem

During 2003, the NWChem and ParSoft teams enabled the use of large numbers of processors to solve large computational challenges. For example, one Hartree-Fock iteration of 107 residues (part of the Ras-RasGAP system, which is important in cell signaling) was completed in three hours using 1400 processors of the MSCF HP machine. This is believed to be the largest calculation of its kind. Another example includes the ability to calculate MP2 energies and gradients for large water clusters, which provides unique information on the water-to-ice transition. Each one of these calculations requires five hours using 1024 processors of the HP cluster (see MSCF User Highlights).

With the change of the Elan3 network to Elan4 in the near future, we anticipate enabling even more capability and capacity computing at the MSCF using NWChem. Figure 5 shows a logarithmic plot of the wall clock time versus the number of processors used to execute the density-functional module of NWChem for the $Si_{75}O_{48}H_{66}$ molecule (3554 basis functions) on various platforms. This figure shows that even preliminary Elan4 results achieved a significant improvement (25 to 45% on moderate numbers of processors) in time to solution by upgrading the network. These results also clearly show the excellent performance of the new HP system in the MSCF.

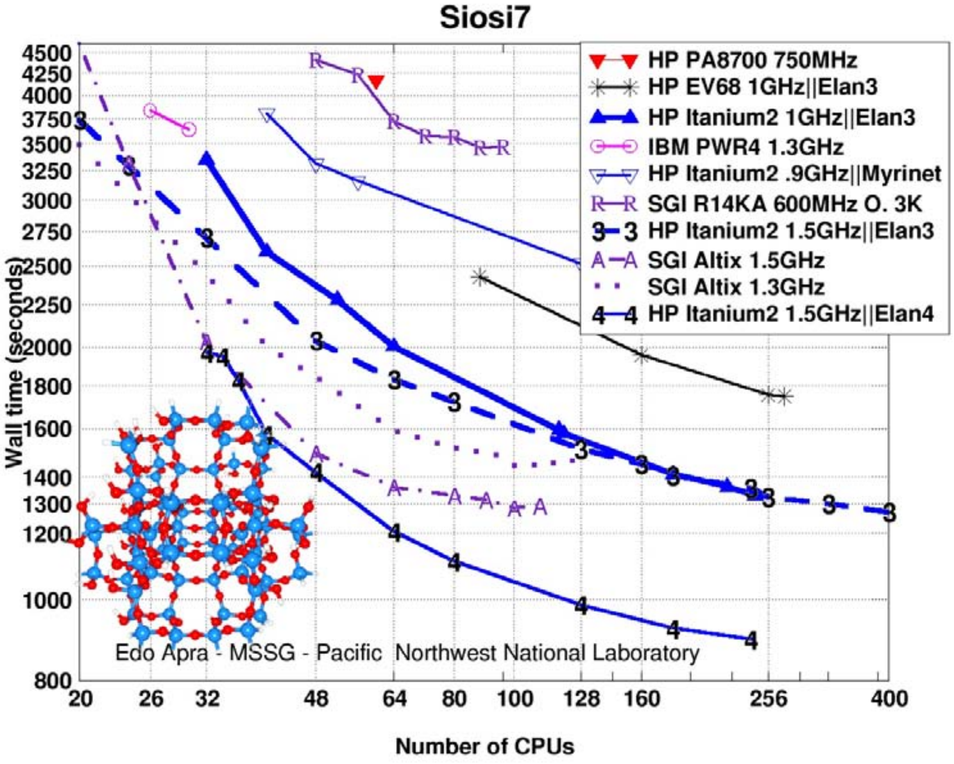

Figure 5. Wall clock time as a function of the number of processors used on various platforms for LDA calculations of 3554 basis functions.

In addition to the changes made in the code to enable large scientific computations, the NWChem team has been adding additional capability. In particular, new time-dependent density functional theory, electron transfer, and projector-augmented wave modules have been released in version 4.5. New many-body methods using the tensor contraction engine (TCE) and a new method for modeling the hopping of hydrogen atoms in MD systems also have been added to the development version.

The TCE is an independent code developed in Python, which takes mathematical equations for many-body wave functions represented in normal-ordered second quantization and produces a set of tensor equations that are further manipulated to produce a parallel implementation in software. In other words, the TCE is an automated code-generation mechanism to produce highly scalable software in an expedited manner. Since the project started, approximately one and half years ago, the TCE has produced more than a dozen parallel equations. Figure 6 shows a

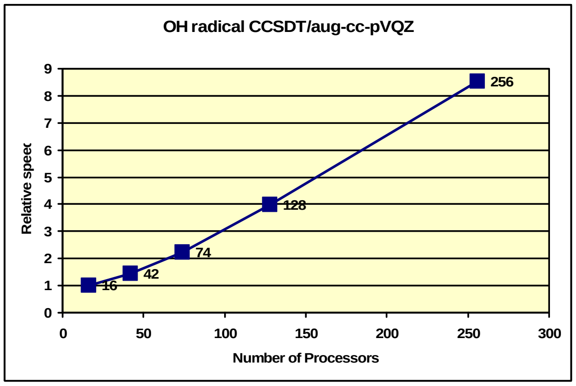

Figure 6. Relative speed as a function of the number of processors on the MSCF HP system for the OH radical using CCSDT/aug-cc-pVQZ.

sample scalability plot for the OH radical using the CCSDT/aug-cc-pVQZ method on the MSCF HP machine and shows an approximately 50% efficiency between using 16 and 256 processors. This is an amazing accomplishment for automatically generated code and it is anticipated that the software will only improve over time because new high-performance algorithms can be added to the TCE and then applied to all of the theoretical methods described above. The TCE work is in collaboration with other computational chemists and computer scientists located at Ohio State University, Oak Ridge National Laboratory, and Waterloo. The PNNL portion of the TCE is supported both through EMSL and through the Basic Energy Sciences division of DOE's Office of Science.

As part of a collaboration with Volkhard Helms and Tomaso Frigato, Max Planck Institute of Biophysics, Frankfurt, Germany, a new method allowing the simulation of proton transfer events in biomolecular systems has been implemented into the development version of NWChem. Called QHOP, this method will be available to users in a 2004 release of NWChem.

Ecce

Many modifications and improvements in Ecce have been accomplished this year, including the release of version 3.1 on August 15, 2003. This version of the code was widely promoted throughout the chemistry community, which resulted in over 100 unique downloads of Ecce within 60 days. Other accomplishments of Ecce during 2003 included:

- establishment of Molecular Builder as a stand-alone application
- greatly improved performance when working with large molecules in the Builder
- expanded graphical capabilities in Calculation Viewer, such as molecular orbital energy plots, tensor visualization support, and log or linear scale plots for graphs
- support for new NWChem electrostatic potential (ESP) and property run types
- capability for users to share controlled access to their data
- redesigned code registration user interface toolkit for use in the Python/Qt scripting environment
- Globus Toolkit (Argonne National Laboratory) v2.4 support
- POV-Ray (http://www.povray.org) support for creating publication-quality graphics
- redesigned Ecce website (http://ecce.emsl.pnl.gov).

The most significant accomplishments with respect to supporting MD calculations in 2003 have been the development of tools for editing and manipulating force fields. These include an ESP toolkit for setting up calculations used to fit partial charges to the electrostatic potential, and force field and topology editors that can be used to complete the force fields necessary to run a simulation. For most calculations, at least part of the system is not covered by standard force fields and the remaining parameters need to be supplied by the user. The topology editor provides a complete listing of all bonds, angles, and torsions in the system and makes it much easier for the user to identify what portions of the force field still need to be parameterized. The force field editor can be used to manipulate force field parameter files even in the absence of a chemical structure. Both these tools can be used to supplement the force field parameter files with the parameters needed to completely describe a chemical system.

ParSoft

Much of the EMSL work in ParSoft has concentrated on porting, tuning, and performance issues to the MSCF machine and on new algorithms that are required for improved performance. A beta release of version 3.3 was made available to the community in October 2003 that included

- nonblocking one-sided operations
- optimized matrix multiplication
- mirrored arrays (shared memory cache for GA data on SMP clusters)
- initial port to the Cray X1 (not yet optimized)
- support for the Mellanox Infiniband Network (not fully optimized)
- improvements in the disk resident array performance.

Many of the issues concerning the HP machine centered on the Quadrics libraries. Among these issues include command port overflows (which limited the number of processors that could be used for one GA-based job) and memory limitations in the libraries (which limited the amount of memory a particular job could access). Working with the Quadrics developers, we overcame these issues, which allowed the large NWChem jobs mentioned above. Continued improvements are being made to take full advantage of the Elan 4 libraries that have been made available on the machine only recently.

Reliable Electronic Structure Calculations for Heavy Element Chemistry: Molecules Containing Actinides, Lanthanides, and Transition Metals

WA de Jong,(a) *J Li,*(a) *BE Bursten,*(b) *JL Sonnenberg,*(b) *EJ Palmer,*(b) *T Yang,*(b)
MM Marino,(c) *WD Ermler,*(c) *C Zhan,*(d) *K Hirao,*(e) *IA Infante,*(f) *DA Dixon,*(g)
BP Hay,(g) *TK Firman,*(g) *Z Zhang,*(h) *RJ Harrison,*(i) *L Visscher,*(j) *SJ Van Gisbergen,*(j)
KG Dyall,(k) *PS Bagus,*(l) *AK Wilson,*(l) *S Matsika,*(m) *J Autschbach,*(n) *and*
B LeGuennic(n)

(a) W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington
(b) The Ohio State University, Columbus, Ohio
(c) University of Memphis, Memphis, Tennessee
(d) University of Kentucky, Lexington, Kentucky
(e) University of Tokyo, Japan
(f) Universiteit Amsterdam, Netherlands
(g) Pacific Northwest National Laboratory, Richland, Washington
(h) Lawrence Berkeley National Laboratory, University of California, Berkeley, California
(i) Oak Ridge National Laboratory, Oak Ridge, Tennessee
(j) Vrije Universiteit/Scientific Computing and Modeling, Amsterdam, Netherlands
(k) Independent Researcher, Portland, Oregon
(l) University of North Texas, Denton, Texas
(m) Temple University, Philadelphia, Pennsylvania
(n) State University of New York, Buffalo, New York

The goal of this project is to study the chemistry of heavy elements through the use of *ab initio* electronic structure calculations to improve our understanding of the behavior of actinides, lanthanides, and heavy transition metals, a group of molecules critical to U.S. Department of Energy (DOE) missions including environmental remediation.

In the first two years of the project, a wide variety of quantum chemical methods based on molecular orbital theory and density functional theory with the proper treatment of relativistic effects were used to interpret experimental data and to predict the chemistry of actinide, lanthanide, and transition metal compounds. Our results provide a firm theoretical basis for this area of chemistry, extending expensive experimental results into new areas of parameter space. The information obtained from calculations on actinide-containing molecules includes, but is not limited to, molecular structure, spectroscopic properties (infrared, nuclear magnetic resonance [NMR], ultraviolet-visible, electron spin resonance), complexation binding energies, and redox chemistry. The results of the calculations enable the characterization of the interaction of the actinide, lanthanide, and heavy transition metal ions with organic complexing agents present in nuclear waste stored at DOE sites and with anion ligands present in natural aqueous systems (e.g., carbonates). This work is advancing our understanding of the fate and transport of actinides in the environment and their interactions with new complexants, such as phosphates and amides, and is leading to the design of innovative *in situ* remediation technologies and novel separation systems. The results obtained from the calculations are a valuable supplement to current, very expensive experimental studies of the radioactive actinides, lanthanides, and radioactive heavy transition metal elements, and allow limited experimental data to be extrapolated to many other regimes of interest.

The radioactive actinide elements americium and curium are chemically similar to their lanthanide counterparts. Both are present in nuclear waste and, thus, present a challenge for actinide separation. Organodithiophosphinates (Figure 1, middle image) have great selectivity for americium over the lanthanide elements. We are using calculations to analyze the structural differences in americium-complexation relative to the lanthanides. This difference has been observed experimentally and has been suggested as the basis for this selectivity. Another compound with a potential application in actinide separation chemistry is the Kläui ligand that currently is being studied experimentally by colleagues at Pacific Northwest National Laboratory (PNNL). Our calculations provide insight into the binding ability of various derivatives of these ligands—for example, the binding of one or two Kläui ligands with uranyl (Figure 1, left image), uranium, and thorium to guide the design of new and better separation agents.

Developing an understanding of the potential reactions of uranyl and carbon tetrachloride is critical for the remediation of radioactive wastes at DOE's Hanford Site. In support of this critical need, our calculations are focused on determining the solvent-phase energetics to provide data for reliable fate and transport models for uranyl species. We have performed similar studies on the formation and solvation of uranyl halides (Figure 1, right image) and various other actinide-water complexes.

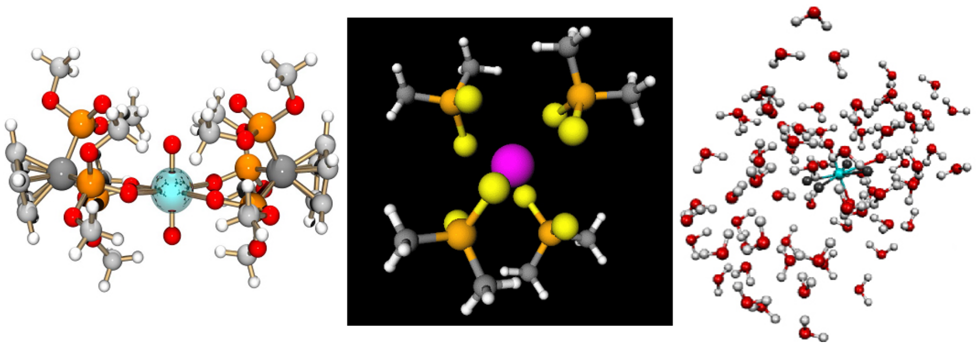

Figure 1. Molecular actinide systems studied in this project. Left: Uranyl binding with two Kläui ligands. Middle: Americium binding to organodithiophosphinate derivatives. Right: Uranyl fluoride molecule solvated in water.

Experiments performed at the University of Virginia and our calculations on CUO molecules in noble gas (argon and neon) matrices led to the discovery of actinide-noble gas complexes (Li et al. 2002). Another combined computational/experimental effort is the development of NMR spectroscopy as a tool for determining the speciation of nuclear waste stored in tanks at the Hanford Site. Our calculations on the NMR properties of uranyl carbonate and nitrate crystals predicted record wide signals for the uranium-bonded oxygen atoms, subsequently confirmed by experiments at PNNL.

Reference

Li J, BE Bursten, L Andrews, and BY Liang. 2002. "Noble Gas-Actinide Compounds: Complexation of the CUO Molecule by Ar, Kr, and Xe Atoms in Noble Gas Matrices." *Science* 295:2243-2245.

Biomolecular Simulation of Base Excision Repair and Protein Signaling

*TP Straatsma,[a] WK Apostoluk,[a] B Gopalan,[a] PE Smith,[b] JH Miller,[c]
A Aceves-Gaona,[c] MB Ernst,[c] ER Vorpagel,[d] T Frigato,[e] M Haranczyk,[f] SE Lee,[g]
RC Rittenhouse,[h] M Zacharias,[i] C Wong,[j] JA McCammon,[j] and W Sudholt[j]*
(a) Pacific Northwest National Laboratory, Richland, Washington
(b) Kansas State University, Manhattan, Kansas
(c) Washington State University Tri-Cities, Richland, Washington
(d) W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington
(e) Max Planck Institute of Biophysics, Frankfurt am Main, Germany
(f) University of Gdansk, Poland
(g) Massachusetts Institute of Technology, Cambridge, Massachusetts
(h) Walla Walla College, College Place, Washington
(i) International University Bremen, Germany
(j) University of California San Diego, La Jolla, California

Environmental factors such as ionizing radiation and oxidative stress contribute to continuous damage of cellular DNA, primarily in the form of oxidized bases, single-strand breaks, and losses of bases. These types of damage are the targets of the base excision repair mechanism enzymes, including polymerase-β. Failure to repair these damaged DNA sites in a correct and timely manner can result in cell death, carcinogenesis, or genetic disease. Resulting mutations in cell signal transduction enzymes can lead to uncontrolled cell proliferation or differentiation. For example, mutations in Ras, the molecular switch in several growth-factor signaling pathways, have been found in about 30% of human tumors. These signaling pathways often involve a chain of protein kinases that activate or deactivate each other through phosphorylation reactions, eventually controlling the activation of transcription factors in the cell nucleus.

Fidelity of Base Excision Repair Protein Human Polymerase β

A detailed understanding of the mechanism by which fidelity is assured by polymerase β (pol β) in the gap-filling DNA repair process is expected to require a complete and accurate description of the pol β active site in its active conformation just prior to the chemical nucleotidyl transfer step. Thus far, crystal structures obtained for pol β–gapped DNA–deoxyribonucleoside triphosphate complexes have been incomplete because of intentional structural alterations necessary to prevent the chemical step needed to obtain a stable structure from the midpoint of the pol β catalytic cycle. Using molecular dynamics (MD) methods and quantum chemical calculations, we have been able to more fully characterize the active conformation of the pol β active site with natural substrates, such as deoxycytidine triphosphate (dCTP), which is shown in Figure 1.

Effects of 8-Oxoguanine on DNA Bending

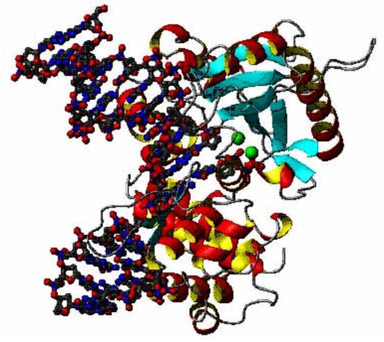
Figure 1. Closed conformation of the pol β (as ribbons)–gapped DNA (as balls and sticks)–dCTP ternary complex. The dCTP substrate and magnesium ions (in green) present in the active site can be seen in the center of the picture.

MD simulations were carried out on the fully solvated and cation-neutralized DNA oligonucleotide GGGAA***C***AACTAG:CTAGTT***G***TTCCC in its native form and with guanine in the central G_{19}:C_6 base pair replaced by 8-oxoguanine (8oxoG). The direction and magnitude of global bending (Figure 2) were assessed by a technique used previously to analyze simulations of DNA containing a thymine dimer. The presence of 8oxoG did not greatly affect the magnitude of DNA bending; however, bending directions that compress the major groove and expand the minor groove were significantly more likely when 8oxoG replaced G_{19}. Crystal structures of glycosylases bound to damaged-DNA substrates consistently show bending that expands the minor groove. We concluded that changes in bending dynamics that facilitate this expansion are part of the mechanism by which 8oxoG is recognized by the base excision repair pathway.

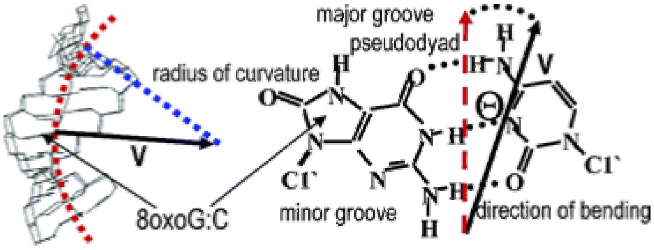

Figure2. Analysis of DNA bending due to 8-oxoguanine.

Simulation of Bulges in DNA and RNA

The goal of this ongoing project is to characterize the structure and dynamics of DNA that is chemically modified due to oxidative stress or irradiation. During the presently ongoing first phase of the project, MD simulations have been performed on a 12-base-pair B-DNA including water molecules and ions. The simulations have been performed for approximately 2 ns. The equilibrated simulation will serve as a reference system for simulations that include chemically modified nucleotide-bases. In addition, the equilibrated system will also be used to induce structural DNA deformations that are known to occur during DNA repair processes, such as minor groove opening and extra-helical base flipping. MD simulations of single unmatched extra bases (bulges) located at the center of otherwise double-stranded DNA and RNA were initiated. Such non-helical structures occur frequently in RNA and also in DNA as a result of DNA modification and damage. The simulations on a single adenine and a single uridine bulge in RNA have now been carried out for ~1.5 ns; however, because of the increased flexibility of the central bulge nucleotide compared to regular nucleic acids, the simulations are not yet fully equilibrated.

Protein Kinase A

An MD simulation on the apo form of protein kinase A was carried out up to ~3 ns. Snapshots from about 1 ns of this trajectory have been used for ligand docking experiments. We used Autodock 3.05 to dock balanol to these snapshots. The averaged binding energy is -12 ±2 kcal/mol. This value is quite close to the experimental values of -10 to -11 kcal/mol. For the scoring function used in Autodock 3.05, the electrostatic interactions contribute ~-2 kcal/mol to the binding energy. Although these results are consistent with those obtained by using a PB/SA (Poisson-Boltzmann/solvent-accessible area) model employed earlier, they differ quantitatively. The PB/SA model gave a significantly more favorable binding energy (-27 kcal/mol) and the electrostatics contributions (-0.9 kcal/mol) accounted for a smaller fraction of the total binding energy. However, the results from Autodock are in better agreement with a semi-empirical PB/SA model in which adjustable parameters were introduced to scale the results of the PB/SA model to fit existing experimental data. The semiempirical model gave a hydrophobic contribution of -16 kcal/mol, an electrostatics contribution of -1 kcal/mol, and other contributions amounting to ~2 kcal/mol.

Further study will include other derivatives of balanol and other ligands. Additional analyses will also be carried out to characterize the dynamics of the apo protein and their influence on molecular recognition.

Chemical Fate of Contaminants in the Environment: Chlorinated Hydrocarbons in the Groundwater

DG Truhlar,(a) CJ Cramer,(a) JD Thompson,(a) AW Jasper,(a) BJ Lynch,(a) NE Schultz,(a) C Zhu,(a) A Chakraborty,(a) YA Borisov,(b) K Morokuma,(c) IV Khavrutskii,(c) SJ Mo,(c) S Irle,(c) Q Wang,(c) G Zheng,(c) HA Witek,(c) J Liu,(c) A Dutta,(c) P Zhang,(c) D Quinonero,(c) K Omiya,(c) DG Musaev,(c) TP Straatsma,(d) M Dupuis,(d) BC Garrett,(d) and DT Chang(d)

(a) University of Minnesota, Minneapolis, Minnesota
(b) Russian Academy of Science, Moscow, Russian Federation
(c) Emory University, Atlanta, Georgia
(d) Pacific Northwest National Laboratory, Richland, Washington

Chlorinated hydrocarbons (CHCs), which can be carcinogens, are the most common contaminants found at hazardous waste sites and are the most prevalent contaminants on U.S. Department of Energy sites that were involved in the production of nuclear weapons. CHCs are resistant to biodegradation, but can degrade by abiotic processes such as hydrolysis, nucleophilic substitution, and dehydrochlorination. Unfortunately, disagreement still exists about the mechanisms and rates of many of the key reactions. Our work is focused on modeling the reactions involved in the degradation of CHCs in the groundwater. There are three primary goals for the research discussed in this paper: 1) development of a computational approach that will allow reaction pathways and rate constants to be accurately calculated; 2) development of more approximate approaches, evaluated against the more accurate approach, which will lay the groundwork for exploratory studies of more complex CHCs; and 3) application of these approaches to study the degradation pathways of CHCs in aqueous liquids. A major focus of this work is on the development of continuum solvation models, which are based on the solvation models of Cramer and Truhlar (1995) and are appropriate for calculations using density functional theory (DFT). Applications are focused on the initial elementary steps in reactions of hydroxide with chlorinated methane molecules, $CH_nCl_{(4-n)}$, and studies of reactions of hydroxide with chlorinate ethylene molecules, $C_2H_nCl_{(4-n)}$.

Significant progress has been made on development of continuum solvation methodology needed to simulate the reactions of hydroxide with chlorinated hydrocarbons. In our initial work in this area, the need to use diffuse functions to get accurate thermochemical simulations became clear. However, our solvation and charge models were not stable when diffuse functions were used. To solve this problem, we first derived a new type of population analysis that was stable to diffuse functions (Thompson et al. 2002), then we used this new type of population analysis to develop charge models that give stable charges with diffuse basis sets (Winget et al. 2002; Thompson et al. 2003). Finally, we are making excellent progress in using the new charge models to create new solvation models. The originally planned large computations will use this methodological development on this aspect of the project. Although it was more efficient to do the preliminary work on desktop workstations and computer servers, the Grand Challenge work on this part of the project using this methodology requires the supercomputers in the Molecular Science Computing Facility (MSCF) in the W.R. Wiley Environmental Molecular Sciences Laboratory (EMSL).

The use of the MSCF computers has been essential for the applications to the reactions of OH^- with chlorinated methane molecules, $CH_nCl_{(4-n)}$ with n=1-4. In previous work, we completed a hierarchy of calculations for the gas-phase reactions for all values of n and for micro-solvated reactions (with up to two water molecules) for n=1-3. Recent work on this project has focused primarily on the high-level electronic structure calculations (e.g., methods that include electron correlation such as second-order Møller-Plesset perturbation theory [MP2] and couple cluster methods) for the microsolvated reactions of carbon tetrachloride (n=4), which are a significant computational challenge and require the use of advanced software such as NWChem on the massively parallel computers in the MSCF. These calculations show a significant increase in the reaction barrier for the nucleophilic substitution (S_N2) reaction as the number of water molecules is increased.

To account for the effects of bulk solvation (i.e., increase of the amount of water molecules to a macroscopic number), we use a computationally tractable and efficient procedure for the calculation of potentials of mean force using mixed Hamiltonian models of electronic structure where quantum subsystems are described with computationally intensive *ab initio* wave functions. The mixed Hamiltonian is mapped into an all-classical Hamiltonian that is amenable to a thermodynamic perturbation treatment for the calculation of free energies. A small number of statistically uncorrelated (solute-solvent) configurations are selected from the Monte Carlo random walk generated with the all-classical Hamiltonian approximation. Those configurations are used in the averaging of the free energy using the mixed quantum/classical Hamiltonian. In Figure 1, the dashed and solid lines with circles are calculations of the energies for the gas-phase reaction using Hartree-Fock (HF) theory with a 6-31G* basis set (dashed curve with open circles) and MP2 with an aug-cc-pVDZ basis set (solid curve with filled circles). The energies are computed at geometries along a minimum energy path for the gas-phase reaction that is calculated using MP2/aDZ theory. The dashed line with open squares show the results of bulk solvation on the HF/6-31G* results using the quantum-mechanical/molecular-mechanical (QM/MM) approach described above. The most accurate value of the solvated potential energy curve is given by combining the MP2 results with those of the QM(HF)/MM and HF/6-31G* results, using a layered electronic structure approach, to give the solid curve with filled squares.

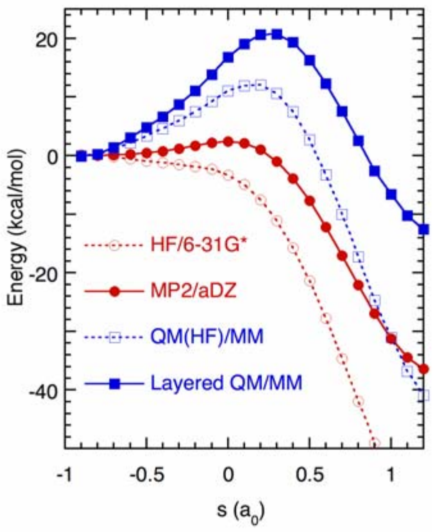

Figure 1. Application of this methodology is illustrated for the S_N2 substitution reaction of methyl chloride by hydroxide.

Another aspect of this work is the development of efficient DFT methods to study chemical reactivity. In another research effort at the University of Minnesota, we are developing new force fields for aluminum nanoparticles that are important to energy research because of their potential as high-energy materials. The computer hardware and software at Pacific Northwest National Laboratory (PNNL) has been essential to the latest stage of this work.

In particular, we have been studying the transition from small-molecule properties to their bulk limit in aluminum molecules, clusters, and nanoparticles, specifically the evolution of the binding energy per particle as the cluster gets bigger. We have also studied the multiplicities of the clusters. We have recently developed a new effective core potential method for use with hybrid DFT that provides accurate energies and geometries. This development was done on computers at the University of Minnesota, but we have been able to use the resulting technology to run large production runs on Al_{19}, Al_{43}, Al_{55}, Al_{79}, and Al_{87} on the PNNL computer system with NWChem. For nanoparticle-sized Al (more than 50 atoms), the capabilities of the PNNL software and hardware for this work are impressive. We have therefore been taking advantage of the parallel capabilities of NWChem and anticipate running even larger systems that will take full advantage of the computing power of EMSL for this kind of research.

References

Cramer CJ and DG Truhlar. 1995. "Continuum Solvation Bottles: Classical and Quantum Mechanical Supplementations." In *Reviews in Computational Chemistry*, eds. KB Lipkowitz and DB Boyd, vol. 6, pp. 1-72. John Wiley and Sons, New York.

Thompson JD, JD Xidos, TM Sonbuchner, CJ Cramer, and DG Truhlar. 2002. "More Reliable Partial Atomic Charges When Using Diffuse Basis Functions." *Physical Chemistry Communications* 5:117-134.

Thompson JD, CJ Cramer, and DG Truhlar. 2003. "Parameterization of Charge Model 3 for AM1, PM3, BLYP, and B3LYP." *Journal of Computational Chemistry* 24:1291-1304.

Winget P, JD Thompson, JD Xidos, CJ Cramer, and DG Truhlar. 2002. "Charge Model 3: A Class IV Charge Model Based on Hybrid Density Functional Theory with Variable Exchange." *Journal of Physical Chemistry A* 106:10707-10717.

Immobilization of Radionuclides in Oxides

H Jonsson,[(a)] F Vila,[(a)] K Tsemekhman,[(a)] A Arnaldsson,[(a)] L Arnadottir,[(a)]
MA Gabriel,[(a)] J Garza-Olguin,[(b)] F Gao,[(c)] EJ Bylaska,[(c)] LR Corrales,[(c)] S Hirata,[(d)]
RM Van Ginhoven,[(e)] and D Makarov[(f)]

(a) University of Washington, Seattle, Washington
(b) Universidad Autonoma Metropolitana, Iztapalapa, Irakium, Mexico
(c) Pacific Northwest National Laboratory, Richland, Washington
(d) W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington
(e) Commissariat a l'Energie Atomique, Gif Sur Yvette, France
(f) University of Texas, Austin, Texas

In this project, we are using computer simulations to study structural changes in oxide glasses that result from electronic excitations occurring over long time periods—for example, hydrolysis driven by radioactive decay.

At the most basic level, our project involves testing a methodology to establish what level of theory is needed to accurately reproduce electronic defects and chemical reactions in oxides, especially quartz and amorphous silica. We have recently reported an extensive study of excitons in silica clusters calculated by both wave function methods and density functional theory (DFT) (Van Ginhoven et al. 2003). It is clear from this and other studies we have performed that DFT calculations with functionals commonly used in condensed-phase simulations are inadequate for the study of excitons and electron holes. We have found that the B3LYP functional is much superior, but it is too computationally demanding when extended systems are studied because of the need to evaluate exact exchange. Instead, we have developed an algorithm and implementation in the NWChem software package for self-interaction correction (SIC) to DFT. This is the first implementation that we know of that includes self-consistency and incorporates periodic boundary conditions to allow simulations of extended systems. Furthermore, our implementation includes atomic forces and, therefore, enables us to calculate minimum energy structures and transition states. The computational cost is, in principle, only a factor of 3 to 5 over regular DFT calculations, but more optimization of the code is needed before that theoretical limit can be achieved.

The DFT-SIC calculations provide solutions to many of the problems associated with regular condensed-phase DFT calculations, and our code opens the possibility of addressing various questions that could not be computationally addressed in the past. Several test problems indicate that, with some simple modifications, the DFT-SIC approach can give comparable accuracy for both stable structures and transition states as the much more expensive B3LYP functional. We have applied the DFT-SIC method to several problems in oxides. One is the spin density of a hole created by substituting an Al-atom for a Si-atom in quartz. Regular DFT, using local-density-approximation (LDA) or generalized-gradient-approximation (GGA) functionals, shows a spin density that is delocalized over all four O-atom neighbors of the Al-atom. However, experimental results show that the spin density is localized on one of the O-atoms, as manifested, for example, by lengthening of only one of the Al-O bonds. The DFT-SIC calculations indeed show localization on only one of the O-atoms. Another example comes from triplet state excitons in quartz. With DFT/GGA calculations, the spin density of the lowest energy structure is distributed over all 72 atoms in the simulation cell. Localized excitons can, however, be found and have been studied in this project previously, but they are metastable with respect to the delocalized state.

With DFT-SIC, the localized excitons become lower in energy, which agrees with experimental findings. Finally, a similar reversal of the relative energy of localized and delocalized states is found for electron holes in amorphous silica. The DFT-SIC calculations show stable localized holes that agree with experimental results, while the holes delocalize at the DFT/GGA level of theory.

The relative energy and structure of different localized excitonic states in silica are still quite well described by DFT/GGA. We have used this level of theory to study the formation of defects in silica due to exciton formation. Ten different glass samples that are free of coordination defects have been generated (Figure 1a). The figure shows an unexpected damage process that was observed in one of the glasses. The damage was unexpected because no thermally activated process was required to form the defect. First, the system was excited to the triplet state, representing a fast process of initial excitation to some very high-energy singlet state and then intersystem crossing and trapping in the lowest triplet state. Because this is a long-lived state, the system was then relaxed in the triplet state, and a break in a Si-O bond was observed (Figure 1b). When the system was returned to the singlet state to represent radiative decay and then relaxed, the defect did not disappear, but rather developed into two coordination defects, 1) a fivefold coordinated Si-atom and 2) a small, strained, four-member ring consisting of two Si-atoms and two O-atoms (Figure 1c). The energy along this path was evaluated both in the singlet and triplet state, as shown in Figure 2.

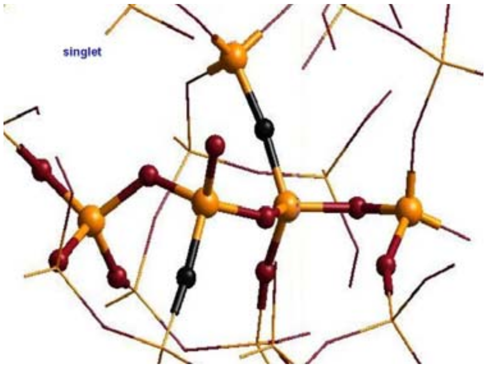

Figure 1a. Exciton-induced damage process in amorphous silica — Part of the 72 atom glass configuration. No coordination defects are present.

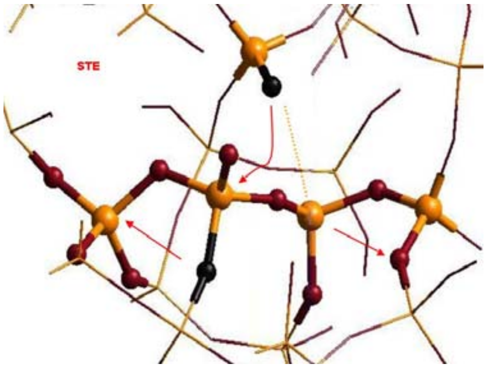

Figure 1b. Exciton-induced damage process in amorphous silica. Self-trapped triplet state exciton configuration. A Si-O bond has been broken.

Figure 1c. Exciton-induced damage process in amorphous silica. Relaxed singlet, ground-state configuration after decay of the exciton. A fivefold Si-atom (circled left) and a small, strained, four-membered ring (circled right) are now present.

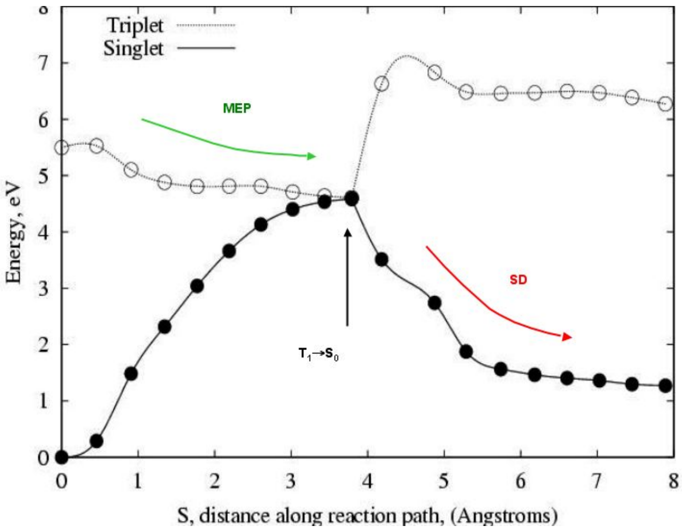

Figure 2. Exciton-induced damage process in amorphous silica. Energy of the singlet and triplet states for the configurational changes described in Figure 1a-c: first a minimum energy path on the triplet state surface, and then a steepest descent path on the singlet state surface.

We then searched for various transition paths from this state. A transition to the original defect-free structure involves a significant energy barrier of 0.76 eV. However, several other processes with similar activation energies could represent further migration of the defects. This could, therefore, be the first step in the formation of an interstitial O-atom defect. It is well known experimentally that radiation of glasses results in the formation of oxygen bubbles. Further simulation studies will address this issue and other defect formation mechanisms in glasses.

We have also studied the insertion of water molecules into glass samples that were free of coordination defects and found, surprisingly, that sites exist where the energy of the water molecule is comparable to the energy in liquid water. The barrier for dissociation of the molecule was found to be very small, on the order of 0.1 eV, and leads to a lower energy state with two hydroxyl groups. This represents an initial stage in hydrolysis of the glass. The importance of strain and anomalous Si-O rings has been studied to identify particularly reactive sites and sites where electronic defects such as excitons and holes are likely to self-trap.

Reference

Van Ginhoven RM, H Jonsson, KA Peterson, M Dupuis, and LR Corrales. 2003. "An *Ab Initio* Study of Self-Trapped Excitons in Alpha-Quartz." *Journal of Chemical Physics* 118(14):6582-6593.

Subsurface Multifluid Flow and Multicomponent Reactive Transport: Characterization of Processes and Properties

S Yabusaki,(a) M White,(a) DH Bacon,(a) V Freedman,(a) PC Lichtner,(b) AJ Valocchi,(c) SL Bryant,(d) T Arbogast,(d) M Wheeler,(d) RJ Lenhard,(e) P Jardine,(f) J Parker,(f) JP Gwo,(g) K Pruess,(h) and C Steefel(i)

(a) Pacific Northwest National Laboratory, Richland, Washington
(b) Los Alamos National Laboratory, Los Alamos, New Mexico
(c) University of Illinois, Urbana, Illinois
(d) University of Texas, Austin, Texas
(e) Idaho National Engineering and Environmental Laboratory, Idaho Falls, Idaho
(f) Oak Ridge National Laboratory, Oak Ridge, Tennessee
(g) University of Maryland, Baltimore County, Baltimore, Maryland
(h) Lawrence Berkeley National Laboratory, Berkeley, California
(i) Lawrence Livermore National Laboratory, Livermore, California

The cleanup of contaminated sites such as those resulting from U.S. Department of Energy (DOE) nuclear-materials production activities involves multiple physical and chemical processes operating on complex mixtures in strongly heterogeneous subsurface materials. This complex combination of factors makes direct observation of waste fate and transport difficult. Computer modeling and simulation of contaminant behavior that accurately represents those processes at the field scale would enable development of more efficient and cost-effective engineering strategies with less risk to biological receptors.

This project targets the characterization of processes and properties in multidimensional porous media as the unifying theme for understanding the fate and transport of contaminants in the subsurface. Our goal is to develop mechanistic process model representations and parameterizations that will lead to enhanced understanding and scientifically defensible predictions of subsurface behavior. Long-term predictions of contaminant fate are critical to the management of the DOE missions of cleanup and long-term stewardship of contaminated sites. The scope for this project addresses multidimensional laboratory and field experiments as well as site-specific, problem-driven site characterization, including history matching. A common theme for the simulations is the use of large numbers of grid cells (~105 to 107 cells) to resolve multiple scales of subsurface heterogeneity. An important aspect of supporting these activities will be the testing and evaluation of parallel programming tools, debugging environments, and visualization software.

A new constitutive theory for the migration of spreading and nonspreading, nonaqueous phase liquids (NAPLs) through the vadose zone was incorporated into the Subsurface Transport Over Multiple Phases (STOMP) multifluid simulator, and successfully validated. The model is currently being used to reconstruct the historical migration of carbon tetrachloride in variably saturated vadose zone sediments at DOE's Hanford Site in Washington State (Figure 1).

A hierarchical parallel job distribution environment designed for parameter estimation and uncertainty analysis was developed using the Condor-G workload management framework for networked workstations and the Globus grid computing toolkit. The environment enables desktop submission and control of massive numbers of forward model simulations required for uncertainty analyses and inverse modeling with the flexibility of using local pools of workstations or high-end computing platforms on the DOE Science Grid.

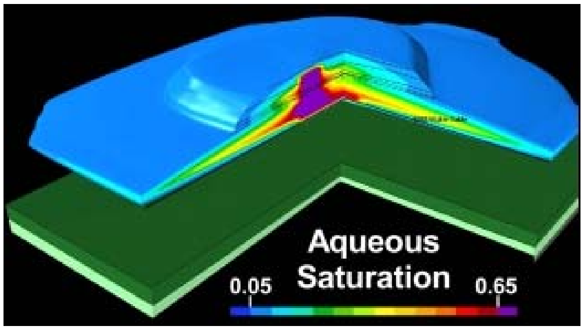

Figure 1. Simulation of the historical migration of carbon tetrachloride released from the Z-9 Crib in the variably saturated sediments of the Hanford Site vadose zone.

Three new capabilities were developed for the parallel Fortran Preprocessor (FP), which generates parallel code from a standard Fortran-90 source: 1) diagnostic feedback and consistency checking for distributed and shared arrays, 2) a utility to identify where results from the serial code differ from the generated parallel code, and 3) hierarchical communication that exploits on-node shared memory on symmetrical multiprocessing architectures. FP enabled the successful port of the STOMP and CRUNCH simulators to an entirely new 64-bit architecture, operating system, and compiler without modification of the source code.

Degradation Mechanism of Solar Cells in an Adverse Environment

Y Yan,[(a)] M Romero,[(a)] C Jiang,[(a)] MM Al-Jassim,[(a)] SJ Pennycook,[(b)] and G Duscher[(c)]
(a) National Renewable Energy Laboratory, Golden, Colorado
(b) Oak Ridge National Laboratory, Oak Ridge, Tennessee
(c) North Carolina State University, Raleigh, North Carolina

CdTe is a promising candidate for large-scale, thin-film solar cells because of its nearly ideal bandgap, high absorption coefficient, and ease of fabrication as a film. The structure of a typical CdTe/CdS solar cell consists of a glass superstrate, then a thin layer of SnO_2 (front contact), *n*-type CdS, *p*-type CdTe, and a back contact. Currently, low efficiency and degradation are the two major obstacles preventing further commercialization of this technology. It has been reported that the current-voltage characteristics of CdTe/CdS solar cells degrade when the cells are exposed to moisture and oxygen over extended periods of time. Currently, two scenarios have been proposed to explain the degradation mechanism. The first scenario is that water vapor may cause peeling of the front contact and oxidization of the back contact. The second proposed scenario is that water molecules interact with CdTe and then dissociate and produce H or O impurities, which diffuse into CdTe and alter its properties.

In FY 2003, we investigated the adsorption of water molecules on the CdTe(001) surface by density-functional, total-energy calculations within the generalized gradient approximation (GGA). We found that dissociative adsorption of water molecules on CdTe(001) is energetically unfavorable—water can only adsorb molecularly on Cd atoms. We also found that the formation of Cd and Te vacancies on the surface and electron and hole doping do not change the adsorption behavior of water on the CdTe(001) surface. Our results suggest that the first scenario could be the mechanism for moisture-induced degradation.

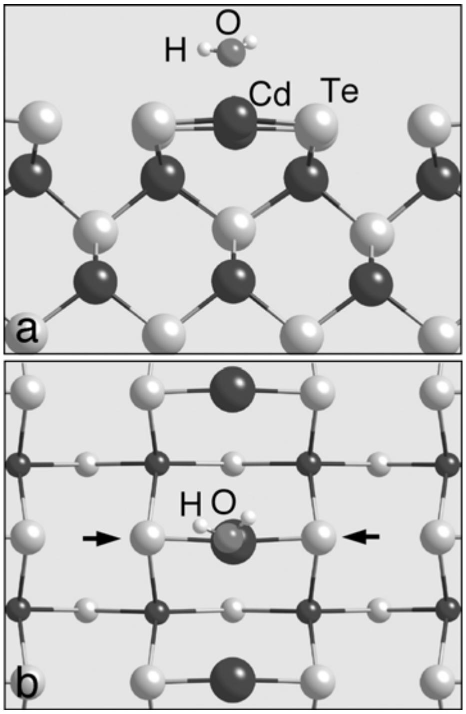

Figure 1. Side a) and top b) views of the structure of a water molecule adsorbed on the perfect (2×1) reconstructed CdTe(001) surface.

Figure 1 shows the side and top views of the structure of a water molecule adsorbed on the (2×1) reconstructed surface. The oxygen atom of the water molecule is at the top of a Cd atom on the surface. The black balls represent Cd atoms, and the grey balls represent Te atoms. In the top view, the large, medium, and small balls represent atoms of the first layer, second layer, and below the second layer, respectively. Adsorption of the water molecule causes rumpling of the CdTe(001) surface under the water molecule. The Cd atom under the O atom is pushed inward by 0.3 Å. The two Te atoms marked by two arrows are pushed downward slightly (~0.05 Å). The H-O-H bond angle is 106.8°, about 3° larger than that of a free water molecule (103.9°). The O-H bond length is 0.96 Å, about the same as that of a free water molecule. The Cd-O distance is 2.5 Å. The calculated adsorption energy is 0.22 eV, indicating that the adsorption could be just molecular, a conclusion that is confirmed from the charge-density plot of the O-Cd-Te plane. We further found that the adsorption of water molecules on the CdTe(001) surface can only be molecular, not dissociative.

We found that the formation of Cd and Te vacancies did not significantly change the adsorption energy of a molecule on the CdTe(0010) surface. Figure 2 shows the side and top views of the structure of a water molecule adsorbed on the Cd atom near the Te vacancy site on the (2×1) reconstructed surface. The adsorption energy was calculated to be 0.21 eV, which is comparable to the adsorption energy on the perfect surfaces.

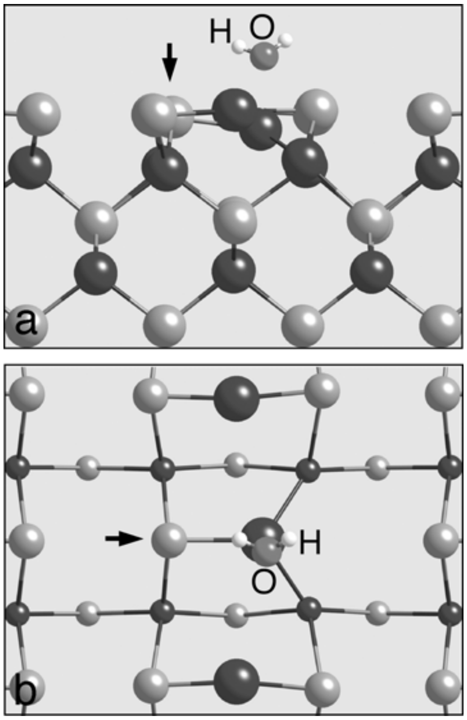

Figure 2. Side (a) and top (b) views of the structure of the adsorption of a water molecule on the Cd atom near the Te vacancy site on the (2×1) reconstructed surface.

We further used In- and Cu-doped CdTe to simulate the effects of electrons and holes, respectively, on the adsorption of water molecules on the CdTe(001) surface by substituting a Cd atom, which is located five layers from the surface, with a Cu atom or an In atom in our calculations. However, no significant change on the adsorption energies was observed.

Our results thus suggest that the first scenario is more likely the mechanism for moisture-induced degradation of a CdTe solar cell.

Models for Radiation-Induced Release of Cytokines by Autocrine-Cell Systems: Implications for Bystander Effects

J Miller,[a] F Zheng,[a] H Resat,[b] L Opresko,[b] D Springer,[b] and HS Wiley[b]
(a) Washington State University Tri-Cities, Richland, Washington
(b) Pacific Northwest National Laboratory, Richland, Washington

Experiments performed at the Gray Cancer Institute using a charged-particle microbeam and a soft x-ray microprobe have shown that levels of bystander-mediated cell killing are primarily determined by the amount of energy deposited in single cells rather than the number of targeted cells (Prise et al. 2003). We have proposed that autocrine processes, where capture of a ligand on a cell-surface receptor stimulates shedding of the same ligand, are the mechanism for this type of bystander effect. Community effects are a well-established feature of cellular systems when the spatial range of autocrine signals is on the micrometer scale (Wiley et al. 2003). Under these conditions, signals from the perturbation of a single cell can be quickly relayed to all of the cells in a high-density population. We are developing models that integrate data from a variety of sources to address the question of whether low doses of ionizing radiation stimulate this type of cell-to-cell communication. In this paper, we focus on the critical issue of the magnitude of shedding induced in cells that receive energy directly from low linear-energy-transfer (LET) radiation.

TGFα, an autocrine ligand of the epidermal growth factor receptor (EGFR), has been shown to mediate secondary activation of EGFR and downstream MAPK and JNK signaling pathways (Dent et al. 1999). Data on MAPK activation by 2 Gy of gamma irradiation delivered at a dose rate of 1.1 Gy/min are shown in Figure 1. Like Shvartsman et al. (2002), our models predicted secondary MAPK-activation kinetics that were more rapid than those observed by Dent et al. (1999) when the rate of protease activation was assumed to be proportional to the instantaneous prompt MAPK activity. The result shown by the solid curve in Figure 1 is based on the hypothesis that the rate of protease activation is proportional to the integral of the prompt MAPK activity multiplied by a Gaussian weighting function that introduces a time delay of approximately 2 hours. The biological basis for this mode of coupling MAPK activity to protease activation is unclear; however, we note that a dependence of the specificity of receptor tyrosine kinase signaling on the duration of extracellular regulated kinase activation is well established (Marshall 1995). The dashed curve in Figure 1 shows the magnitude and kinetics of shedding predicted by this model.

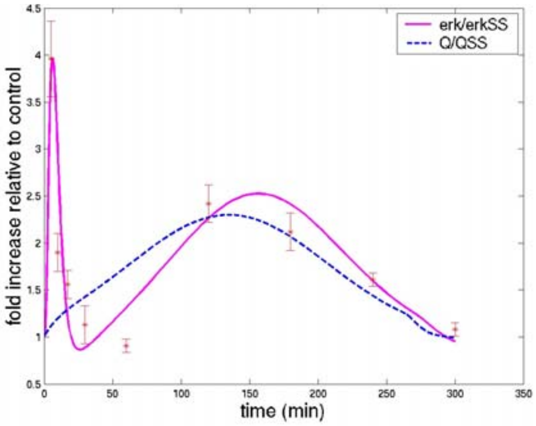

Figure 1. Kinetics of MAPK activation by 2 Gy of gamma irradiation.

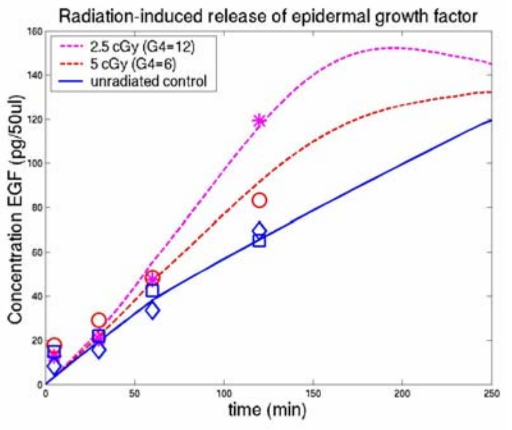

Figure 2. Radiation-induced shedding of EGFR by 2.5 and 5 cGy of gamma irradiation.

Direct observation of shedding under autocrine conditions is difficult due to the low probability that shed proteins escape into the bulk extracellular medium; hence, the use of antibodies that facilitate escape by blocking receptors is a common experimental practice. Data on shedding of EGFR obtained at Pacific Northwest National Laboratory by this method are shown in Figure 2 by the diamonds and asterisks. The solid curves show our model predictions under blocking conditions based on the same assumptions used to model the TGFα-mediated MAPK activation observed by Dent et al. (1999) at higher doses. The fit to data on shedding by un-irradiated human mammary epithetial cells/HCF cells (solid curve) was obtained with a constant shedding rate of 37 EGFR molecules per minute. Shedding induced in the same cell line by 2.5 cGy of gamma irradiation was explained by a time-dependent release rate with a shape similar to the dashed curve in Figure 1 and with a maximum value 1.8 times the shedding rate of un-irradiated control cells. Whether shedding of this magnitude by a hit cell can induce a community response among bystanders will depend on several factors, including ligand diffusivity, ligand-receptor affinity, steady-state level of free receptors, and cell plating density. Simulation techniques to investigate the interplay of these factors are under development.

References

Dent P, DB Reardon, JS Park, G Bowers, C Logsdon, K Valerie, and R Schmidt-Ullrich. 1999. "Radiation-Induced Release of Transforming Growth Factor Alpha Activates the Epidermal Growth Factor Receptor and Mitogen-Activated Protein Kinase Pathway in Carcinoma Cells, Leading to Increased Proliferation and Protection from Radiation-Induced Cell Death." *Molecular Biology of the Cell* 10(8):2493-2506.

Marshall CJ. 1995. "Specificity of Receptor Tyrosine Kinase Signaling: Transient Versus Sustained Extracellular Signal-Regulated Kinase Activation." *Cell* 80(2):179-185.

Prise KM, M Folkard, and BD Michael. 2003. "Bystander Responses Induced by Low LET Radiation." *Oncogene* 22(45):7043-7049.

Shvartsman SY, MP Hagan, A Yacoub, P Dent, HS Wiley, and DA Lauffenburger. 2002. "Autocrine Loops With Positive Feedback Enable Context-Dependent Cell Signaling." *American Journal of Physiology-Cell Physiology* 282(3):C545-C559.

Wiley HS, SY Shvartsman, and DA Lauffenburger. 2003. "Computational Modeling of the EGF-Receptor System: A Paradigm for Systems Biology." *Trends in Cell Biology* 13(1):43-50.

Relativistic Quantum Chemistry Studies of the Electronic Structures and Photoelectron Spectra of Negatively Charged Transition-Metal Complexes

J Li,[a] K Boggavarapu,[b] and L-S Wang[b,c]
(a) W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington
(b) Washington State University Tri-Cities, Richland, Washington
(c) Pacific Northwest National Laboratory, Richland, Washington

The central theme of this project is to elucidate electronic and geometrical features of various atomic clusters identified by anion photoelectron spectroscopy experiments. We have performed theoretical investigations on four topics in close collaboration with Wang's group at Washington State University Tri-Cities and the W.R. Wiley Environmental Molecular Sciences Laboratory. Following are brief summaries of these four projects.

An Unusual and Exceptionally Stable Au_{20} Tetrahedral Nanocluster

Experimental studies in Wang's group reveal that a series of naked gold clusters can be formed with Au_{20} being exceptionally stable and having a highest unoccupied molecular orbital-lowest unoccupied molecular orbital (HOMO-LUMO) energy gap around 1.8 eV. Through numerous structural searches and theoretical modeling of the electronic properties, we discovered that Au_{20} possesses an unusual tetrahedral pyramid structure (Figure 1). The calculated HOMO-LUMO gap, electron detachment energies, and excitation energies are in excellent agreement with those measured experimentally. We also found that this nanocluster has remarkably large second-order optical nonlinearity, and its materials are expected to have great potential in catalysis and opto-electric devices.

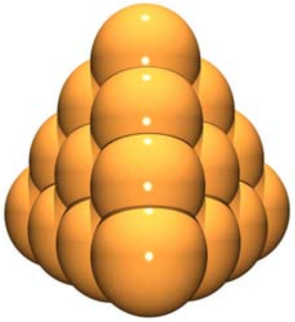

Figure 1. Optimized tetrahedral Au_{20} structure.

Observation and Identification of Icosahedral $Mo@Au_{12}$ and $W@Au_{12}$ Molecules

By using relativistic density function theory (DFT) calculations, we fully explored the electronic structures and energetics of the icosahedral, octahedral, and cubo-octahedral structures of W/Mo-Au_{12} clusters observed by Wang's group. The calculated icosahedral structures (Figure 2) of these clusters have the lowest energy among their isomers, and the theoretically simulated photoelectron spectra are in good agreement with the experimental results. These clusters are thus confirmed to possess an endohedral icosahedral structure with an extremely

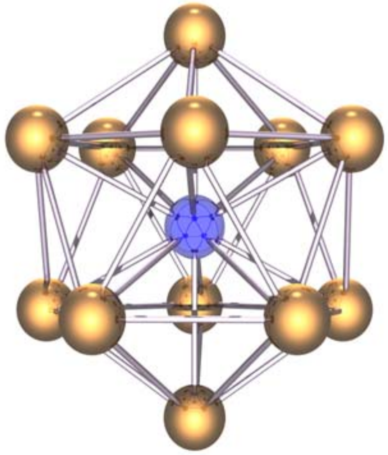

Figure 2. The structure of icosahedral $W@Au_{12}$.

stable electronic configuration, as previously predicted (Pyykko et al. 2002). This work is the first confirmation of icosahedral gold clusters.

Hydrocarbon Analogs of Boron Clusters

Using density functional geometry optimizations, we have determined the geometry structures of a series of boron clusters B_x (x = 10 - 16) discovered in Wang's group. In contrast to the elemental boron and boron compounds for which highly symmetric icosahedral structures dominate, small boron clusters are found to have quasi-planar structures, and exhibit aromaticity and antiaromaticity according to the Hückel rule. Our calculations revealed that the aromatic boron clusters possess circular shapes, whereas antiaromatic clusters are elongated, analogous to structural distortions of antiaromatic hydrocarbons. These planar boron clusters are the only series of molecules other than the hydrocarbons to exhibit size-dependent aromatic and antiaromatic characters in chemistry.

Structure and Bonding in $Fe\text{-}S_{n-}$ (n = 1-6) Clusters

We have undertaken computational studies of mono-iron-poly-sulfur systems that are essential in understanding electronic structures of Fe-S proteins. With experimental photoelectron spectroscopic results as guidelines, we explored various geometrical and electronic states of $Fe\text{-}S_{n-}$ clusters by DFT methods, and good agreement has been reached between the computed and experimental electron affinities. We discovered that Fe is not oxidized beyond its favorite +2 or +3 oxidation states as found in Fe-S proteins, and the geometry and electronic structure evolution of FeS_{n-} systems is quite different from the corresponding FeO_{n-} species.

Reference

Pyykko P and N Runeberg. 2002. "Icosahedral WAu_{12}: A Predicted Closed-Shell Species Stabilized by Aurophilic Attraction and Relativity and in Accord with the 18-Electron Rule." *Angewandte Chemie-International Edition* 41(12):2174-2176.

The Binding Energies of Isomers of Water Clusters: High-Level Electronic Structure and Empirical Potential Results

S Xantheas,[(a)] G Fanourgakis,[(a)] and E Aprà[(b)]
(a) Pacific Northwest National Laboratory, Richland, Washington
(b) W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington

The structures and energetics of water clusters play an important role in quantifying the interactions between water molecules. This molecular-level information assists in the development of empirical models used to study condensed phases, such as liquid water and ice, and the examination of transitions between the two phases.

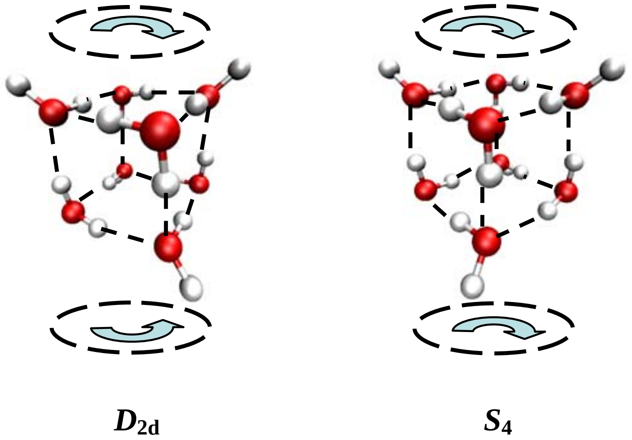

Figure 1. Structures of the D_{2d} and S_4 water octamer isomers showing the direction of the in-the-ring hydrogen bonded networks of the two constituent tetramers.

In this work, the second-order Møller-Plesset perturbation theory (MP2) complete basis set (CBS) limit for the binding energy of the two low-lying water octamer isomers of D_{2d} and S_4 (Figure 1) symmetry was estimated at -72.7 ±0.4 kcal/mol using the family of augmented correlation-consistent orbital basis sets of double through quintuple zeta quality. The largest MP2 calculation with the augmented quintuple zeta (aug-cc-pV5Z) basis set (2296 contracted gaussian basis functions) produced binding energies of -73.70 (D_{2d}) and -73.67 kcal/mol (S_4). The effect of higher correlation, computed at the CCSD(T) level of theory, were estimated to be <0.1 kcal/mol (Xantheas and Aprà 2004).

The newly established MP2/CBS limit for the water octamer was reproduced quite accurately by the newly developed all-atom polarizable, flexible-interaction potential (Burnham and Xantheas 2002). The TTM2-F binding energies of -73.21 kcal/mol (D_{2d}) and -73.24 kcal/mol (S_4) for the two isomers are just 0.5 kcal/mol (or 0.7%) larger than the MP2/CBS limit.

A similar theoretical approach has been used to study the larger $(H_2O)_{20}$ water clusters. Extensive MP2 calculations were used to accurately determine the most stable isomer (the edge-sharing isomer shown in Figure 2), compute binding energies (that once again were matched by the results of the TTM2-F potential), and generate vibrational frequencies that will be compared with experimental spectra.

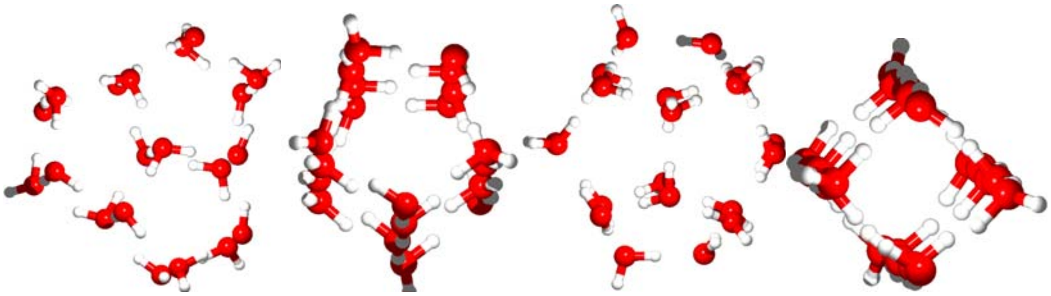

Figure 2. Isomers of the $(H_2O)_{20}$ water cluster: edge-sharing, face-sharing, dodecahedron and box-kite.

The largest MP2 calculation for these series of molecules was performed with the aug-cc-pVQZ basis set, for a total of 3440 basis functions. The computational requirements of this calculation required the current state-of-the-art in both electronic software and hardware.

All MP2 calculations were carried out using the NWChem code (Kendall et al. 2000) run on the newly acquired supercomputer at the Molecular Science Computing Facility in the W.R. Wiley Environmental Molecular Sciences Laboratory.

References

Burnham CJ and SS Xantheas. 2002. "Development of Transferable Interaction Models for Water: IV. A Flexible, All-Atom Polarizable Potential (TTM2-F) Based on Geometry-Dependent Charges Derived from an *Ab Initio* Monomer Dipole Moment Surface." *Journal of Chemical Physics* 116(12):5115-5124.

Kendall RA, E Aprà, DE Bernholdt, EJ Bylaska, M Dupuis, GI Fann, RJ Harrison, J Ju, JA Nichols, J Nieplocha, T Straatsma, TL Windus, and AT Wong. 2000. "High-Performance Computational Chemistry: An Overview of NWChem, a Distributed Parallel Application." *Computer Physics Communications* 128(1-2):260-283.

Xantheas SS and E Aprà. 2004. "The Binding Energies of the D_{2d} and S_4 Water Octamer Isomers: High-Level Electronic Structure and Empirical Potential Results." *Journal of Chemical Physics* 120(2):823-828.

Electronic Transfer Calculations for the Hemes of Flavocytochrome c3

M Dupuis,[(a)] ***TP Straatsma,***[(a)] ***T Frigato,***[(b)] ***DMA Smith,***[(c)] ***and ER Vorpagel***[(d)]
(a) Pacific Northwest National Laboratory, Richland, Washington
(b) Max Planck Institute of Biophysics, Frankfurt am Main, Germany
(c) Whitman College, Walla Walla, Washington
(d) W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington

This project is part of a program funded by the U.S. Department of Energy Office of Science for studies of advanced modeling and simulation of biological systems. The goal of this project is to provide the computational tools required for a detailed molecular-level characterization and understanding of intricate biological reactions such as the electron transfer through flavocytochrome c3 of *Shewanella frigidimarina*.

The first goal of this pilot project was to conduct a comprehensive comparison of density function theory functionals for modeling hemes, such as those in our cytochrome of interest. Comparing experimentally determined properties, such as spin state energy differences, iron-ligand bond distances, and Mössbauer quadrupole splitting of two different hemes in their reduced, oxidized, low-spin and high-spin states using several functionals, led to the conclusion that the B3LYP hybrid functional is indeed the best functional to use for modeling hemes and was the only one to predict all three experimental values well.

The second goal achieved in the project was calculation of electron transfer coupling elements for heme dimers using the electron transfer module of NWChem, implemented last year as part of another pilot project at the Molecular Science Computing Facility in the W.R. Wiley Environmental Molecular Sciences Laboratory. A very important and surprising discovery was that the orbital, which participates in the electron transfer [primarily Fe(d_{xy}) or Fe(d_{π})], strongly affects the magnitude of the electronic coupling between heme dimers and, therefore, significantly affects the overall electron transfer process. The cytochrome can modify the electron transfer by moving hemes closer together, but can also tune the electron transfer by changing the protein environment surrounding the heme so that one orbital is preferred over the other. Proper alignment of the electron transfer donor and acceptor orbitals on the iron atom is crucial (Figure 1).

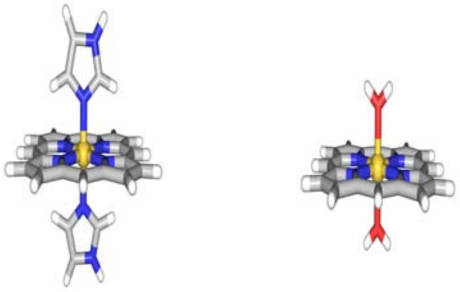

Figure 1. Two heme molecules studied, one with two imidazole ligands (left) and one with two water molecules (right) as ligands. B3LYP correctly predicted a preferred low-spin form with imidazole and a high-spin form with water.

We initially ran quantum-mechanical/molecular-mechanical calculations for the cytochrome hemes, which included the solvent and protein environment. We encountered spin contamination problems when using the fluorine pseudo-carbon link atoms in conjunction with open-shell hemes, but hydrogen link atoms worked better. This is not an ideal solution, however, so one of our future goals is to tune the pseudopotential fluorine link atom to minimize spin contamination.

Development of a Parallel Computational Environment for Imaging Science

DR Jones,[a] ER Jurrus,[b] KA Perrine,[b] BD Moon,[b] and GM Petrie[b]
(a) W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington
(b) Pacific Northwest National Laboratory, Richland, Washington

The rapid processing and analysis of multi-sensor imagery have important roles in many of the missions of the U.S. Department of Energy, including national security, energy resources, and environmental quality. This project contributes to those needs, as well as DOE's program area in the national stewardship for science, by providing the capability for cost-effective, real-time image processing of large-scale images.

Because of the size of high-resolution images (typically greater than 100 megapixels) from sources such as satellites with multiple sensors, most desktop computers do not have sufficient computing resources to process and display these images in a timely fashion. Many image-processing tasks can be split into parallel paths to reduce overall processing time, but this approach has required specialized hardware and customized software that have not been cost-effective enough to gain widespread use.

The goal of this project has been to create a parallel computational environment that could use clusters of off-the-shelf computing systems for image processing and display without the need to rewrite well-qualified image-processing algorithms for the parallel environment. The resulting software "workbench" demonstrates a scaleable, cost-effective approach that applies the power of parallel computing to reduce large-scale, image-processing times from hours to minutes or seconds, compared with single-processor solutions (Jones et al. 2003).

In FY 2001 and FY 2002, we prototyped the basic infrastructure for what we call the Parallel Computational Environment for Imaging Science (PiCEIS). While most other parallel imaging applications use a master-slave model with a serial cycle of input, output, and visualization, the images and data in PiCEIS are totally and persistently distributed on a distributed memory machine during the entire processing cycle. The PiCEIS code is portable to PC/Linux-based clusters as well as HP Linux (IA64), IBM SP, and SGI parallel computers.

In FY 2003, we undertook the following activities to optimize the evolving environment architecture: implementing parallel input/output for additional processing speed, completing new scripting capabilities by adding a Python interface for user-defined data flow, and testing performance on large datasets.

The input of large datasets is a performance bottleneck because their input may consume a disproportionate amount of processing time. We implemented a basic parallel input/output procedure and also an out-of-core capability based on Disk I/O and Disk Resident Arrays (Nieplocha et al. 1996) for out-of-core computations for images that will not fit into the computing platform's memory.

We moved from OpenDX to a Python-based architecture to gain flexibility and performance in creating and interacting with visualization tools. Modules for driving the IBM Scalable Graphics Engine (SGE) and IBM T221 high-resolution display were rewritten in Python.

In prior years, PiCEIS displayed test images of moderate size (300 megabytes) with linear parallel speedup on the 240-processor Linux cluster as well as the 512-processor IBM SP (using up to 48 CPUs) with a speedup factor of 75. Images that normally took more than an hour to process on a single processor were processed in a few seconds. In FY 2003, we tested very large test images (1.7 gigabytes or 256 megapixels) and obtained near-linear parallel speedup on the HP Linux cluster, maintaining a similar speedup factor of almost 70. We displayed these very large images in real time on the IBM SGE and T221 (Perrine and Jones 2003).

We completed the proof-of-concept of the PiCEIS image processing workbench with a prototype user interface. The interaction environment was expanded to enable simultaneous image processing of a large-scale image and overlay regions of interest. This real-time processing permits analysts to be flexible in their workflow. If general image inspection leads analysts to identify a specific area of interest, they can apply further processing algorithms in an area-of-interest window that updates as they move it (Figure 1).

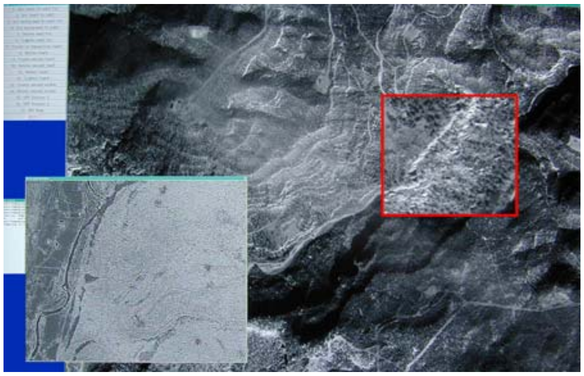

Figure 1. PiCEIS display of a 1.7-gigabyte satellite map with real-time image processing of details in regions of interest.

Further work is underway to optimize PiCEIS performance characteristics. For example, we are working to identify intelligent methods for breaking up an image for distributed processing based on the number of processors. We are also working toward smart pseudo-registration techniques to improve matching of edges when the parallel processed components of an image are stitched together.

References

Jones DR, ER Jurrus, BD Moon, and KA Perrine. 2003. "Gigapixel-Size, Real-Time Interactive Image Processing with Parallel Computers." In *8th International Workshop on High-Level Parallel Programming Models and Supportive Environments*, Nice, France, IEEE Computer Society Press, Los Alamitos, California. [Online]. Available URL: http://csdl.computer.org/comp/proceedings/ipdps/2003/1926/00/1926toc.htm.

Nieplocha J, RJ Harrison, and RJ Littlefield. 1996. "Global Arrays: A Non-Uniform Memory-Access Programming Model for High Performance Computers." *Journal of Supercomputing* 10(2):169-189.

Perrine KA and DR Jones. 2003. "Interactive Imaging Science on Parallel Computers: Getting Immediate Results." In *17th International Parallel & Distributed Processing Symposium*, Nice, France, IEEE Computer Society Press, Los Alamitos, California. [Online]. Available URL: http://csdl.computer.org/comp/proceedings/ipdps/2003/1926/00/1926toc.htm.

DOE Science Grid Operational Development and Testing

S Studham,[a] C Marasco,[a] DE Cowley,[a] BJ Pitre,[b] SB Yabusaki,[b] A Coleman,[b] CR Cole,[b] JD Myers,[b] JD Novotny,[c] KR Jackson,[c] JR Boverhof,[c] WE Johnston,[c] DL Olson,[c] VS Welch,[d] AA Rodriguez,[d] JM Schopf,[d] K Chanchio,[e] and PF LoCascio[e]

(a) W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington
(b) Pacific Northwest National Laboratory, Richland, Washington
(c) Lawrence Berkeley National Laboratory, University of California, Berkeley, California
(d) Argonne National Laboratory, Argonne, Illinois
(e) Oak Ridge National Laboratory, Oak Ridge, Tennessee

The U.S. Department of Energy (DOE) Science Grid shown in Figure 1 will support DOE science applications that require distributed collaborations, very large data volumes, unique instruments, and the incorporation of supercomputing resources into these environments (Johnston et al. 2000).

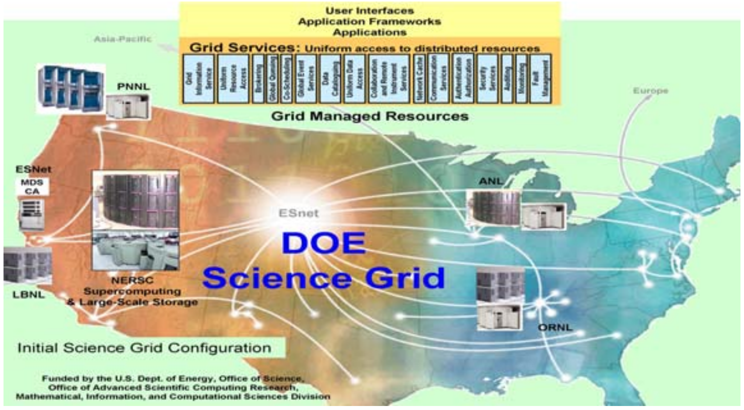

Figure 1. The Molecular Science Computing Facility (MSCF) is included in the deployment of the DOE Science Grid and is embracing grid middleware (embodied in the Globus toolkit) as a key technology to support large-scale computing and data transfer both inside and outside the W.R. Wiley Environmental Molecular Sciences Laboratory (EMSL). The grid infrastructure will integrate EMSL instruments, computers, and data more seamlessly than ever before and will allow a user to perform a single sign-on to authenticate to all grid resources.

Deployment on NWfs

The Globus toolkit was successfully deployed in 2003 on EMSL's new large-scale storage server, NWfs. This deployment makes the entire content of NWfs available to its users on the DOE Science Grid via the GridFTP protocol, thus enabling parallel data transfer, which can reduce transfer times to a fraction of that of a conventional FTP transfer. This improved capability paves the way for ultra-scale data transfers between Pacific Northwest National Laboratory (PNNL) and the other DOE Science Grid sites.

Bioinformatics on the DOE Science Grid

The Genome Analysis and Databases Update (GADU) system is an automated system for genome/sequence analysis. GADU uses the Globus grid toolkit (upon which the DOE Science Grid is implemented) to move data and schedule jobs on high-performance computer systems. In April, GADU used the DOE Science Grid to submit jobs to DOE systems, including the large EMSL supercomputer, MPP2, to analyze 59 microbial genomes in 11 hours. More than 200,000 individual Basic Local Alignment Search Tool (BLAST) jobs were run in the course of the analysis.

Firewall Policy Changes

Lessons learned during deployment in 2002 showed that network firewalls are a major consideration in Science Grid implementations, and that Science Grid administrators must work hand-in-hand with site security administrators. A major effort was undertaken in 2003 to modify PNNL network security with respect to firewalls in order to allow Science Grid connections to MSCF systems. This will smooth the process of allowing new Science Grid systems to gain access to MSCF resources.

Ucode-Globus Proof of Concept

Ucode (http://www.mines.edu/igwmc/freeware/ucode) is an inverse modeling code used by PNNL hydrologists to model groundwater flow at DOE's Hanford Site. In the original implementation, Ucode is run on dedicated Intel Architecture clusters that are configured to run multiple instances of the code in parallel. The individual Ucode instances are controlled by a PNNL-written program named taskr.pl, which was Science Grid-enabled in 2002.

Running the new taskr.pl locally and attempting to run it to remote sites taught us that firewall configuration and resource scheduling policies at supercomputing sites are of critical importance. Accordingly, new firewall policies relating to Science Grid services were adopted by PNNL as described above. In addition, taskr.pl was further modified to take advantage of the symmetrical multiprocessing compute nodes that are very common in high-performance computer clusters. This modification improves Ucode throughput by allowing its jobs to conform to a site's preferred processor geometries.

DOE Science Grid Helpdesk Infrastructure

RT Deployment

Support of the EMSL Support and Help Queue help browser, which had been deployed for Science Grid helpdesk support in 2002, was terminated in 2003, so the system was redeployed on a trouble ticket system named RT. RT is a far more configurable system that offers access via standard web browsers, solving the problem of widely deploying a custom client application. It will also facilitate new technology that will allow trouble tickets to be exchanged between Science Grid member sites.

Trouble Ticket Exchange

PNNL staff, in conjunction with staff members at Lawrence Berkeley National Laboratory, National Energy Research Supercomputing Center, Oak Ridge National Laboratory, National Center for Supercomputing Applications, and the International Virtual Grid Data Laboratory have created and agreed on a new trouble ticket-exchange standard. The distributed nature of the Science Grid means that help requests will need to be routed from site to site for resolution. The new trouble ticket-exchange method involves transforming user help requests into a common exchange format, then directing them to the site where they will need to be resolved. There, the request is again transformed from the common format into the native format of the site's help request system. This approach allows a site's local support staff to deal with the request using the same processes that they are accustomed to, thereby integrating grid support into normal site administration practices.

Reference

Johnston W, R Bair, I Foster, A Geist, and B Kramer. 2000. *The DOE Science Grid: Computing and Data Infrastructure for Large-Scale Science*. [Online]. Available URL: http://doesciencegrid.org.

Deploy Lustre as a Global Filesystem on the HPCS2 Supercomputer

ND Tenney,[(a)] *EJ Felix,*[(a)] *and PJ Braam*[(b)]
(a) W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington
(b) Cluster File Systems, Inc., Brookline, Massachusetts

Lustre is a parallel, object-based global file system. It was developed primarily by Cluster File Systems, Inc., but has been released under an open-source license that allows staff at Pacific Northwest National Laboratory (PNNL) to make improvements and fix software issues.

PNNL Lustre Deployment

Beginning in January 2003, the second phase of the PNNL supercomputer (MPP2) was being assembled in Roseville, California. PNNL staff members were onsite preparing the 53-TB SAN for test runs of early releases of the Lustre code. Lustre testing continued until June 2003. When the MPP2 went into production, Lustre also was deployed, being used as a global scratch file system. Speeds of up to 3.2 GB/sec have been observed during writes to the file system.

The file system consists of the aggregation of 1300 72-GB Fibre Channel disks enclosed in 33 HP Enterprise Virtual Arrays (EVAs), which provide hardware Redundant Array of Independent Disks (RAID), and controller failover capabilities for High Availability (HA). A 2-GB/sec Storage Area Network (SAN) provides HA access to the EVAs. Thirty-two HP rx2600 server computers service requests for access to the EVA storage. The remaining EVA is used to store the file system's metadata. File system requests arrive via the Quadrics Elan interface, which is the high-speed interconnect for the MPP2 supercomputer.

Lustre Architecture

The Lustre file system consists of a number of parts. A Lustre client provides the local system a method for accessing the data stored in the Lustre file system. It provides an interface to the client system that simply appears as if it were a file system on the local machine. An object storage target (OST) stores the actual data being placed in the Lustre file system. This data is often, but not always, "striped" across several OSTs in the Lustre file system. This is done to aggregate writes and reads to improve the performance of the file system. A metadata server (MDS) is the glue between the OST and the client. It stores the location of the real data on the OST's for client retrieval, as well as storing the file system attributes, such as file-creation time, file-modification time, and access privileges.

In the PNNL deployment of Lustre, there are 64 OSTs running on 32 server computers with failover pairs defined to improve stability of the file system. Each failover pair has full access to its partner's EVA storage. In addition, 2 MDS nodes are used, also in a failover configuration, with the backup node hosting the Lightweight Directory Access Protocol

(LDAP) server, which controls the failover configuration, as well as hosting the Lustre configuration for the MPP2 supercomputer. The remaining nodes in the supercomputer are configured as Lustre clients.

Lustre at PNNL

After MPP2 had been declared a production system, early users of the PNNL Lustre file system were successful in running applications that are often used on Molecular Science Computing Facility clusters, including: NWChem, ARMCI-based applications, file-system tests, and many others. PNNL staff developed several tools for monitoring the health and performance of the MPP2 Lustre file system (Figure 1).

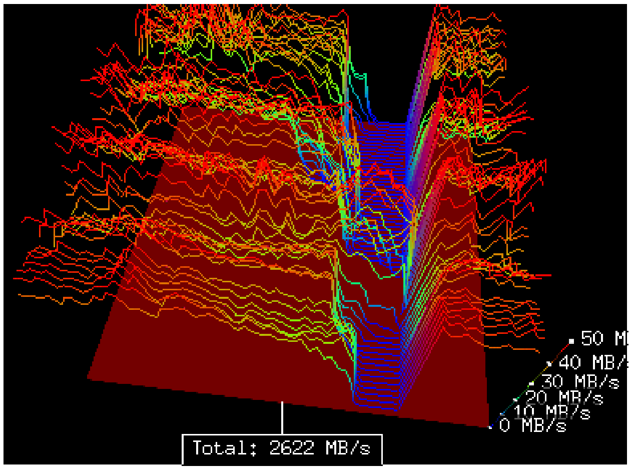

Figure 1. Graph of an NWChem process showing IO, a communication barrier, followed by more IO.

PNNL staff were also instrumental in the formation of a group organized to raise industry awareness of the Lustre file system and its applications to the computing industry. The Lustre Users Group, formed by PNNL and Lawrence Livermore National Laboratory, held its first conference at PNNL in July 2003. Attendees included representatives from several government laboratories and industries, and vendors interested in the high-performance computing industry. Detailed presentations of Lustre deployments were presented, including a number of presentations by PNNL staff regarding the PNNL deployment of Lustre.

User Projects

First Principles Methods for Predicting the Chemistry of Environmentally Relevant Systems

DT Mainz, WA Goddard, MS Diallo, WB Floriano, M Blanco, YH Jang, RJ Nielsen, DM Philipp, F Faglioni, V Molinero
California Institute of Technology, Pasadena, California
RA Friesner, BF Gherman, BD Dunietz, F Rittner, E Batista, M Baik, ME Wirstam, Z Zhou
Columbia University, New York, New York
RP Muller
Sandia National Laboratories, Albuquerque, New Mexico

Biomolecular Interfaces: Simulation of Sedimentation of Biomolecular Assemblies

J Kurzak, TE Warth, BW Beck, AD Haymet, GC Lynch, BM Pettitt, TC Rogala, K Wong, KM Dyer, RJ Cole, N Choudhury
University of Houston, Houston, Texas

Computational Studies in Molecular Geochemistry and Biogeochemistry

PE Smith
Kansas State University, Manhattan, Kansas
JR Rustad
University of California, Davis, California
DA Yuen, JW Halley
University of Minnesota, Minneapolis, Minnesota
GA Voth
University of Utah, Salt Lake City, Utah
AR Felmy, DA Dixon, M Dupuis, TP Straatsma, KM Rosso, EJ Bylaska
Pacific Northwest National Laboratory, Richland, Washington
RD Lins Neto
Laboratory of Physical Chemistry, Zurich, Switzerland
JH Weare
University of California San Diego, La Jolla, California

Simulation Studies of Protein-Protein Interactions

VH Helms
Saarland University, Saarbrucken, Germany

Development of a Parallel Computational Environment for Imaging Science

ML Rawls
Hanford High School, Richland, Washington
G Petrie, KA Perrine, ER Jurrus, BD Moon
Pacific Northwest National Laboratory, Richland, Washington
DR Jones
W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington

Computational Quantum Chemistry

JM Gonzales
Emory University, Atlanta, Georgia
HF Schaefer, NJ DeYonker, NR Brinkman, JD Larkin, C Pak, NA Richardson, KW Sattelmeyer, MS Schuurman, HL Woodcock
University of Georgia, Athens, Georgia
JP Kenny
Sandia National Laboratories, Livermore, California
RJ Harrison
Oak Ridge National Laboratory, Oak Ridge, Tennessee

Imaging Science and Technology Initiative

BD Moon
Pacific Northwest National Laboratory, Richland, Washington

MD Calculation of Gramicidin Conformational Dynamics

GS Harms
Universitat Wurzburg, Wurzburg, Germany

Computer Simulation of Single-Cell Irradiation Studies Using the PNNL Electron Microbeam

JH Miller, F Zheng
Washington State University, Richland, Washington

Lattice Boltzmann Methods

DR Rector
Pacific Northwest National Laboratory, Richland, Washington

Reliable Electronic Structure Calculations for Heavy Element Chemistry: Molecules Containing Actinides, Lanthanides, and Transition Metals

BE Bursten, JL Sonnenberg, EJ Palmer, T Yang
Ohio State University, Columbus, Ohio
MM Marino, WD Ermler
University of Memphis, Memphis, Tennessee
C Zhan
University of Kentucky, Lexington, Kentucky
K Hirao
University of Tokyo, Tokyo, Japan
IA Infante
Universiteit Amsterdam, Amsterdam, Netherlands
BP Hay, DA Dixon, TK Firman
Pacific Northwest National Laboratory, Richland, Washington
WA De Jong, J Li
W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington
Z Zhang
Lawrence Berkeley National Laboratory, University of California, Berkeley, Berkeley, California

RJ Harrison
Oak Ridge National Laboratory, Oak Ridge, Tennessee
L Visscher, SJ Van Gisbergen
Vrijie Universiteit/Scientific Computing and Modelling, Amsterdam, Netherlands
KG Dyall
Independent Researcher, Portland, Oregon
PS Bagus, AK Wilson
University of North Texas, Denton, Texas
S Matsika
Temple University, Philadelphia, Pennsylvania
J Autschbach, B LeGuennic
State University of New York at Buffalo, Buffalo, New York

Electronic Structure Calculations of the Hemes of Flavocytochrome Ifc3

TP Straatsma, M Dupuis
Pacific Northwest National Laboratory, Richland, Washington
ER Vorpagel
W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington
DM Smith
Whitman College, Walla Walla, Washington

Immobilization of Radionuclides in Oxides

J Garza-Olguin
Universidad Autonoma Metropolitana, Iztapalapa, Irakium, Mexico
F Gao, EJ Bylaska, LR Corrales
Pacific Northwest National Laboratory, Richland, Washington
S Hirata
W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington
RM Van Ginhoven
Commissariat a l'Energie Atomique, GIF SUR YVETTE, France
H Jonsson, F Vila, K Tsemekhman, A Arnaldsson, L Arnadottir, MA Gabriel
University of Washington, Seattle, Washington
D Makarov
University of Texas at Austin, Austin, Texas

A Computational Approach to Understanding Aerosol Formation and Oxidant Chemistry in the Troposphere

JS Francisco
Purdue University, West Lafayette, Indiana
R Bianco, JR Hynes, S Wang
University of Colorado, Boulder, Colorado
KA Peterson
Washington State University, Pullman, Washington

GK Schenter, SS Xantheas, LX Dang, BC Garrett, DA Dixon, SM Kathmann, GS Fanourgakis
Pacific Northwest National Laboratory, Richland, Washington
MA Miller
Foundation for Fundamental Research on Matter - Institute for Atomic and Molecular Physics, Amsterdam, Netherlands

Biomolecular Simulation of Base Excision Repair and Protein Signaling

PE Smith
Kansas State University, Manhattan, Kansas
JH Miller, A Aceves-Gaona, MB Ernst
Washington State University, Richland, Washington
TP Straatsma, WK Apostoluk, B Gopalan
Pacific Northwest National Laboratory, Richland, Washington
ER Vorpagel
W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington
T Frigato
Max Planck Institute of Biophysics, Frankfurt am Main, Germany
M Haranczyk
University of Gdansk, Gdansk, Poland
SE Lee
Massachusetts Institute of Technology, Cambridge, Massachusetts
RC Rittenhouse
Walla Walla College, College Place, Washington
M Zacharias
International University Bremen, Bremen, Germany
C Wong, JA McCammon, W Sudholt
University of California San Diego, La Jolla, California

Subsurface Multifluid Flow and Multicomponent Reactive Transport: Characterization of Processes and Properties

AJ Valocchi
University of Illinois, Urbana, Illinois
J Gwo
University of Maryland, Baltimore County, Baltimore, Maryland
SB Yabusaki, MD White, DH Bacon, A Coleman, JM Malard, Y Fang
Pacific Northwest National Laboratory, Richland, Washington
K Pruess
Lawrence Berkeley National Laboratory, University of California Berkeley, Berkeley, California
CI Steefel
Lawrence Livermore National Laboratory, Livermore, California
PC Lichtner
Los Alamos National Laboratory, Los Alamos, New Mexico
JC Parker, PM Jardine
Oak Ridge National Laboratory, Oak Ridge, Tennessee
RJ Lenhard
Idaho National Engineering and Environmental Laboratory, Idaho Falls, Idaho

M Rosing
Independent Researcher, Ft. Collins, Colorado
MF Wheeler, J Arbogast, SL Bryant
University of Texas at Austin, Austin, Texas

Chemical Fate of Contaminants in the Environment: Chlorinated Hydrocarbons in the Groundwater

DG Truhlar, CJ Cramer, JD Thompson, AW Jasper, BJ Lynch, NE Schultz, C Zhu, A Chakraborty
University of Minnesota, Minneapolis, Minnesota
YA Borisov
Russian Academy of Science, GSP-1, Moscow V-334, Russian Federation
K Morokuma, IV Khavrutskii, SJ Mo, S Irle, Q Wang, G Zheng, HA Witek, J Liu, A Dutta, P Zhang, D Quinonero, K Omiya, DG Musaev
Emory University, Atlanta, Georgia
TP Straatsma, M Dupuis, BC Garrett, DT Chang
Pacific Northwest National Laboratory, Richland, Washington

Degradation Mechanism of Solar Cells Under Adverse Environment

Y Yan, M Romero, C Jiang, MM Al-Jassim
National Renewable Energy Laboratory, Golden, Colorado
SJ Pennycook
Oak Ridge National Laboratory, Oak Ridge, Tennessee
G Duscher
North Carolina State University, Raleigh, North Carolina

Thermochemistry of Arsenic Compounds

DM Di Toro, JP Capitani
Manhattan College, Riverdale, New York

Homogeneous Photocatalysis of Water to Form Dihydrogen

SS Xantheas
Pacific Northwest National Laboratory, Richland, Washington

Large-Scale Code Development Applied to Building a Geometry Based Computational Framework for Biology and Ecology Applications

HE Trease
Pacific Northwest National Laboratory, Richland, Washington

Graphic Presentation and MD Simulations of Protein Interactions

HP Lu, X Tan
Pacific Northwest National Laboratory, Richland, Washington

Development and Analysis of the Hybrid Quantum Mechanics-Molecular Mechanics Approach in NWChem - Why Catalytic Antibodies Cannot Achieve the Level of Catalysis of Enzymes?

W Sudholt, KK Baldridge
University of California San Diego, La Jolla, California

Quantum Mechanics/Molecular Mechanics (QM/MM) Study of the Substrates Of Haloalkane Dehalogenase

PV Roberts, BT Kawakami
Stanford University, Stanford, California

Computational Thermochemistry and Benchmarking of Reliable Methods

DF Feller
Pacific Northwest National Laboratory, Richland, Washington

DOE Science Grid Operational Development and Testing

BJ Pitre, SB Yabusaki, A Coleman, CR Cole
Pacific Northwest National Laboratory, Richland, Washington
S Studham, DE Cowley
W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington
JD Novotny, KR Jackson, JR Boverhof, WE Johnston
Lawrence Berkeley National Laboratory, University of California Berkeley, Berkeley, California
VS Welch, AA Rodriguez, RA Bair
Argonne National Laboratory, Argonne, Illinois
K Chanchio, PF LoCascio
Oak Ridge National Laboratory, Oak Ridge, Tennessee

Ab Initio Characterization of Unreduced and Reduced (012) Surfaces of Hematite

M Dupuis, EJ Bylaska
Pacific Northwest National Laboratory, Richland, Washington

Magpie Annotation of Genome Sequences

MF Romine
Pacific Northwest National Laboratory, Richland, Washington

Relativistic Quantum Chemistry Studies of the Electronic Structures and Photoelectron Spectra of Negatively Charged Transition-Metal Complexes

K Boggavarapu
Washington State University, Richland, Washington
J Li
W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington
BM Elliott, AI Boldyrev
Utah State University, Logan, Utah

Parameterization and Upscaling for Field-Scale Transport Modeling

ML Rockhold, PD Meyer
Pacific Northwest National Laboratory, Richland, Washington

Ionization Energy Calculations of Deoxyribonucleotides

X Wang
Pacific Northwest National Laboratory, Richland, Washington
ER Vorpagel
W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington

Optical Microscopy Field Enhancement Simulation

NA Klymyshyn
Pacific Northwest National Laboratory, Richland, Washington

Electronic Transfer Calculations for the Hemes of Flavocytochrome c3

M Dupuis, TP Straatsma
Pacific Northwest National Laboratory, Richland, Washington
ER Vorpagel
W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington
T Frigato
Max Planck Institute of Biophysics, Frankfurt am Main, Germany
DM Smith
Whitman College, Walla Walla, Washington

Development of GASNet Communication System on the Quadrics Interconnect

DO Bonachea, CS Bell
University of California Berkeley, Berkeley, California

Molecular Modeling of Phospholamban

K Kuczera, Y Houndonougbo
University of Kansas, Lawrence, Kansas

Structure and Reactions of Novel Lithium Dialkylamide Aggregates and Mixed Aggregates

LM Pratt
Fisk University, Nashville, Tennessee

Benchmarking Atmospheric Models on BTC/Opus Itanium Cluster

RC Kelly
University Corporation for Atmospheric Research, Boulder, Colorado

Benchmarking Climate and Weather Modeling Codes on the PNNL Cluster

L Sankaran
Hewlett-Packard Company, Richardson, Texas

Benchmarking of IA-64

HJ Wasserman
Los Alamos National Laboratory, Los Alamos, New Mexico

Fuel Cell Concept Animation

MC Perkins
Pacific Northwest National Laboratory, Richland, Washington

Combat Service Support Automated Information Systems Interface

MC Perkins
Pacific Northwest National Laboratory, Richland, Washington

Mapping of H_2O Potential in the Ice-Like Medium

SS Xantheas
Pacific Northwest National Laboratory, Richland, Washington
V Buch
Hebrew University of Jerusalem, Jerusalem, Israel
J Sadlej
University of Warsaw, Warsaw, Poland

Effects of the Presence of Helium on Silicon Carbide

LR Corrales
Pacific Northwest National Laboratory, Richland, Washington
A Chartier, RM Van Ginhoven
Commissariat a l'Energie Atomique, GIF SUR YVETTE, France
C Meis
Commissariat a l'Energie Nucleair, Gif-sur-Yvette, France

Geometry Optimizations of Aqueous Aluminum Clusters

JR Rustad
University of California, Davis, California
WH Casey, AG Stack
University of California Davis, Davis, California

Determining the Phase Behavior of Polarizable Water Models

AZ Panagiotopoulos, PJ Lenart
Princeton University, Princeton, New Jersey

Immobilizing Enzymes for Useful Service

MC Perkins
Pacific Northwest National Laboratory, Richland, Washington

HP Itanium II Cluster Software Development and Application Porting

WA Hahn, JE Blanding, KS Poirier, N Srivastava
Hewlett-Packard Company, Marlboro, Massachusetts

Facility Planning at Battelle (3-D Architectural Rendurings)

MC Perkins
Pacific Northwest National Laboratory, Richland, Washington

Bioinformatics and Genome Analysis on DOESG Computational Resources
V Nefedova
Argonne National Laboratory, Argonne, Illinois

Neutrino Oscillation Illustration and Animation
MC Perkins
Pacific Northwest National Laboratory, Richland, Washington

Relativistic Many-Electron Theories
S Hirata
W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington
RJ Harrison, T Yanai
Oak Ridge National Laboratory, Oak Ridge, Tennessee
MA Nooijen, AA Auer
Princeton University, Princeton, New Jersey

New Approach to Extract and Visualize Discontinuities from 2D and 3D Seismic and Imagery Data Sets
GH Seedahmed, ER Jurrus, GG He
Pacific Northwest National Laboratory, Richland, Washington

Intrusion Detection Video
MC Perkins
Pacific Northwest National Laboratory, Richland, Washington

PNNL Past and Future Grounds Projects
MC Perkins
Pacific Northwest National Laboratory, Richland, Washington

Computational Fluid Dynamics Applied to Airflow Problems in Animal and Human Lungs
S Kabilan, K Beck, E Hoffman
University of Iowa, Iowa City, Iowa

Ab Initio Calculations of GTP Hydrolysis by Using NWChem and TAO Components
Y Alexeev
W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington
TL Windus
W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington

Local Motions in Protein Structures
HL Selzle
Institut fur Physikalische and Theoretische Chemie of the Technische Universitat Muechen, Garching, Germany
EW Schlag
Technical University Munich, Garching, Germany

Molecular Dynamics and QM/MM Studies of Macrolide Molecules

S Sekusak
Pliva Pharm. Ind. Inc., Analytical Centre, Zagreb, Croatia (Local Name: Hrvatska)
TP Straatsma
Pacific Northwest National Laboratory, Richland, Washington

Electronic Structure Calculations of Hydrogen Bonding in Biological Macromolecules

EJ Bylaska
Pacific Northwest National Laboratory, Richland, Washington
K Tsemekhman, D Baker, C Duarte, AV Morozov, WR Schief
University of Washington, Seattle, Washington

Image for March 2003 Issue of Resource Magazine

MC Perkins
Pacific Northwest National Laboratory, Richland, Washington
ER Vorpagel
W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington

Regional Scale Atmospheric Modeling

JD Fast, LK Berg
Pacific Northwest National Laboratory, Richland, Washington

Rapidly Deployable Sensor Network

ER Jurrus, BJ Burghard, VW Van Zandt
Pacific Northwest National Laboratory, Richland, Washington

Computational Study of Polycyclic Aromatic Hydrocarbons and Their Derivatives

M Vala, J Szczepanski
University of Florida, Gainesville, Florida
S Hirata
W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington

Computer Simulation of Radiation Effects in Zircon

R Devanathan
Pacific Northwest National Laboratory, Richland, Washington

Optimization of Peptide Identification from Tandem Mass Spectral Data

WR Cannon, JM Malard, KH Jarman, A Heredia-Langner
Pacific Northwest National Laboratory, Richland, Washington
DJ Baxter
W.R. Wiley Environmental Molecular Sciences Laboratory, Richland, Washington

Benchmarking Localized Coupled Cluster Methods in the NWChem Package

MA Nooijen, AA Auer
Princeton University, Princeton, New Jersey

Modeling Complex Bacterial Cell Systems through Development of Segregated Models with Spatial Localization

BD Wood
Oregon State University, Corvallis, Oregon
HE Trease, JK Fredrickson
Pacific Northwest National Laboratory, Richland, Washington

Computer Simulations of Chemical Reactions in Confined Environments and Mesoscopic Clusters

S Constas, Y Liu
University of Western Ontario, London, Ontario, Canada

Developing High-Performance Parallel Languages on the SGI Altix

DO Bonachea, W Chen, CS Bell
University of California Berkeley, Berkeley, California
JC Duell, CC Iancu
Lawrence Berkeley National Laboratory, University of California, Berkeley, California

Molecular Modeling Studies of Molecularly Imprinted Polymers Selective for Biosensor Materials

R Jacob, DE Mainwaring
Royal Melbourne Institute of Technology University, Melbourne, Australia

Gold: Acounting and Allocation Management

SM Jackson
Pacific Northwest National Laboratory, Richland, Washington

Holographic Imaging Radar Measurement System

MC Perkins
Pacific Northwest National Laboratory, Richland, Washington

Computational Investigation of the Fixed Charge Derivative Tris[(2,4,6-Trimethoxyphenyl) Phosphonium]

KA Herrmann, VH Wysocki
University of Arizona, Tucson, Arizona

Parallel Reactive Transport Modeling of Radioactive Waste Form Weathering Using the STORM Code

DH Bacon
Pacific Northwest National Laboratory, Richland, Washington

CFD Code Performance Testing

TE Michener
Pacific Northwest National Laboratory, Richland, Washington

NWChem and Ecce Software User Agreements

During 2003, 213 new NWChem/Ecce User Agreements were established, bringing the total number of User Agreements to 906. Both NWChem and Ecce are distributed to a worldwide community of researchers with varied interests from academic, industrial, and government sites. Examples include User Agreements with the Department of Chemistry at Emory University, the Supercomputing Center at the Korea Institute of Science and Technology Information, and the VTT Technical Research Centre of Finland.

Staff

Molecular Sciences Visualization and User Services

Donald R. Jones, Staff Scientist, Technical Lead
(509) 376-3013, dr.jones@pnl.gov

Toni P. Quackenbush, Senior Administrative Secretary
(509) 376-2767, toni@pnl.gov

Doug J. Baxter, Senior Research Scientist
(509) 376-3751, douglas.baxter@pnl.gov

Wibe A. de Jong, Senior Research Scientist
(509) 376-5290, wibe.dejong@pnl.gov

Jun Li, Senior Research Scientist
(509) 376-4354, jun.li@pnl.gov

Chris S. Oehmen, Senior Research Scientist
(509) 376-1481, christopher.oehmen@pnl.gov

Erich R. Vorpagel, Chief Scientist
(509) 376-0751, erich.vorpagel@pnl.gov

Molecular Science Computing Facility Computer Operations

R. Scott Studham, Technical Lead
(509) 376-8430, scott.studham@pnl.gov

Lisa Hobson, Administrative Secretary
(509) 376-8501, lisa.hobson@pnl.gov

David E. Cowley, Senior Research Scientist
(509) 376-9181, david.cowley@pnl.gov

Evan J. Felix, Research Scientist
(509) 376-1491, evan.felix@pnl.gov

Zach A. Garrett, Graduate Student
(509) 376-3239, zachary.garrett@pnl.gov

Scott M. Jackson, Senior Research Scientist
(509) 376-2205, scott.jackson@pnl.gov

Cindy Marasco, Senior Research Scientist
(509) 376-1241, cindy.marasco@pnl.gov

Ryan W. Mooney, Senior Research Scientist
(509) 376-3590, ryan.mooney@pnl.gov

Kevin D. Mote, Student
(509) 376-8501, kevin.mote@pnl.gov

Kenneth P. Schmidt, Technician III
(509) 376-4178, kenneth.schmidt@pnl.gov

Gary B. Skouson, Senior Research Scientist
(509) 376-5401, skouson@pnl.gov

Nathan D. Tenney, Research Scientist
(509) 376-1493, nathan.tenney@pnl.gov
Tim A. Witteveen, Senior Research Scientist
(509) 372-1363, timw@pnl.gov

Ryan P. Wright, Scientist
(509)376-3052, ryan.wright@pnl.gov

Molecular Sciences Software

NWChem

Theresa L. Windus, Staff Scientist, Technical Lead
(509) 376-4529, theresa.windus@pnl.gov

Jessica Foreman, Administrative Secretary
(509) 376-3412, jessica.foreman@pnl.gov

Edoardo Aprà, Senior Research Scientist
(509) 376-1280, edoardo.apra@pnl.gov

Yuri Alexeev, Postdoctoral Research Fellow
(509) 376-5152, yuri.alexeev@pnl.gov

Mahin Hackler, Scientist
(509) 376-2746, mahin.hackler@pnl.gov

So Hirata, Senior Research Scientist
(509) 376-6751, so.hirata@pnl.gov

Lisa Pollack, Postdoctoral Research Fellow
(509) 376-5152, lisa.pollack@pnl.gov

Marat Valiev, Senior Research Scientist
(509) 376-2514, marat.valiev@pnl.gov

Matrixed Staff

Eric Bylaska, Senior Research Scientist
(509) 376-9114, eric.bylaska@pnl.gov

Ecce

Gary Black, Senior Research Scientist
(509) 375-2316, gary.black@pnl.gov

Bruce Palmer, Senior Research Scientist
(509) 375-3899, bruce.palmer@pnl.gov

Karen Schuchardt, Staff Scientist
(509) 375-6525, karen.shuchardt@pnl.gov

ParSoft

Jarek Nieplocha, Staff Scientist
(509) 372-4469, jarek.nieplocha@pnl.gov

Manojkumar Krishnan, Research Scientist
(509) 372-4206, manojkumar.krishnan@pnl.gov

Bruce Palmer, Senior Research Scientist
(509) 375-3899, bruce.palmer@pnl.gov