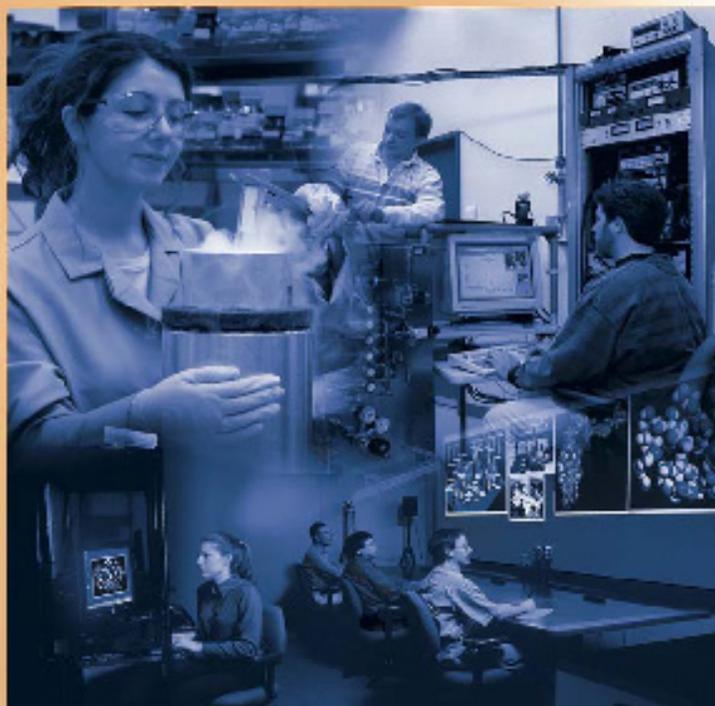




EMSL

**NWChem Meeting on Science Driven Petascale
Computing and Capability Development at EMSL**

January 25-26, 2007



Contents

Executive Summary	1
1. Background	2
2. Scope of the Meeting	4
3. Scientific Capability Needs	5
3.1. Kinetics and dynamics of chemical transformations in complex systems	5
<i>3.1.1 Prediction and Discovery of Reaction Pathways</i>	6
<i>3.1.2 Free energy profile and sampling of reaction pathways</i>	6
<i>3.1.3 Calculating accurate energetics</i>	7
3.2. Chemistry at interfaces and in the condensed phase	8
<i>3.2.1 Improved scaling characteristics for large molecular systems</i>	9
<i>3.2.2 Utilizing real space methodologies</i>	10
<i>3.2.3 Coupling multiple scales</i>	11
3.3 Spanning longer time regimes with molecular dynamics	12
<i>3.3.1 Simulating longer time scales</i>	12
4. Petascale Computing for Computational Chemistry	13
5. Open-Source NWChem Software	15
6. Summary	16
Appendix A: NWChem Overview	17
Appendix B: Meeting Agenda	24
Thursday, January 25, 2007	24
Friday, January 26, 2007.....	25
Appendix C: List of Participants	27
Appendix D: Presentation Abstracts	30
Appendix E: Poster Abstracts	37
Appendix F: Additional Input and Comments	43

*NWChem Meeting on Science Driven Petascale Computing and Capability
Development*

Meeting Report

Held January 25-26, 2007

**W.R. Wiley Environmental Molecular Sciences Laboratory
Richland, WA**

Executive Summary

On January 25, and 26, 2007, an NWChem meeting was held that was attended by 65 scientists from 29 institutions including 22 universities and 5 national laboratories. The goals of the meeting were to look at major scientific challenges that could be addressed by computational modeling in environmental molecular sciences, and to identify the associated capability development needs. In addition, insights were sought into petascale computing developments in computational chemistry.

During the meeting common themes were identified that will drive the need for the development of new or improved capabilities in NWChem. Crucial areas of development that the developer's team will be focusing on are 1) modeling of dynamics and kinetics in chemical transformations, 2) modeling of chemistry at interfaces and in the condensed phase, and 3) spanning longer time scales in biological processes modeled with molecular dynamics. Various computational chemistry methodologies were discussed during the meeting, which will provide the basis for the capability developments in the near or long term future of NWChem.

In addition to capability development, it will be essential for NWChem to effectively utilize petascale computing platforms in order for researchers to tackle large problem sizes and to get their results quicker. The NWChem software was developed 15 years ago, and has been shown to scale to thousands of processors (teraflops). An improved software architecture able to utilize the hardware advances and tens to hundreds of thousands of processors anticipated in the next ten years will need to be developed. Bringing NWChem to the petascale platforms will be a major effort that will require computational chemists, mathematicians and computer scientists to join forces.

Both new capability development and increasing NWChem's scalability to petaflop systems and beyond are major efforts, which combined require more resources and manpower than is currently available in the software development team at EMSL. Discussions were held regarding the NWChem software becoming open-source to enable an extensive collaborative capability and petascale development environment. There is an overwhelming support to go open-source, which may greatly increase the developer base of the software suite.

The needs identified during this meeting will provide a strategic direction for software development in the NWChem quantum chemistry software suite, and will ensure that the software capability will continue to provide EMSL and its user community with the cutting-edge capabilities they need on next-generation platforms.

1. Background

The W.R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a Department of Energy (DOE) national scientific user facility, provides its users with capabilities for research in environmental molecular sciences. One of the six facilities housed at the EMSL is the Molecular Science Computing Facility (MSCF), which provides the hardware and software capabilities tuned to ensure fastest time to solution for simulations in support of EMSL's science focus areas. In fiscal year 2007, over 80% of the MSCF projects required the use of, and about 50% of the computing cycles on the MSCF supercomputer are used to perform molecular simulations with the NWChem computational chemistry software.

NWChem is a one-of-a-kind computational chemistry software code developed as part of the EMSL construction project to provide users with the massively parallel and scalable computational chemistry software necessary to tackle complex questions in environmental molecular science. The software continues to be developed to provide new cutting-edge capabilities to address new scientific questions relevant to EMSL, and to ensure that the software will provide the fastest time-to-solution on the growing MSCF supercomputing resources and at other supercomputer centers. Capability development is user driven, and input is obtained through the MSCF Greenbook process, NWChem workshops, and through various discussions with, and requests from users directly.

NWChem is a large and complex code that consists of over 2 million lines of Fortran and C code, and provides many methods to compute the properties of molecular and periodic systems using standard quantum mechanical descriptions of the electronic wavefunction or density. Its classical molecular dynamics capabilities provide for the simulation of biological macromolecules and solutions. The object oriented programming model enables the various approaches to be combined to perform, for example, mixed quantum-mechanics and molecular-mechanics simulations. A more detailed overview of NWChem can be found in Appendix A.

The NWChem software is currently distributed to over 1750 sites world wide (70% academia and 75% of the sites outside the US) by means of an EMSL User Agreement. The use of the NWChem software has been cited in over 550 scientific publications, many in high ranking journals including Science and PNAS, and a significant number (over 300) of these publications utilized NWChem on the MSCF supercomputer at the MSCF. For example, in fiscal year 2006 about 100 publications cited the use of NWChem, and about 40 utilized the EMSL hardware capability.

Users have applied the capabilities of NWChem to a wide range of complex scientific problems. To show how NWChem impacts scientific research and discovery, two examples are briefly discussed, one related to hydrogen storage and one related to geochemistry simulating transition metals in the environment. Both studies were performed running the NWChem software on the EMSL supercomputer:

- Researchers from the University of Alabama modeled the thermodynamics cycle of hydrogen release and uptake for chemical boron-nitrogen hydrogen storage compounds, a critical part in the development of environmentally friendly hydrogen-based fuel cells as a power source. To obtain reliable predictions they performed high accuracy calculations with NWChem, routinely using a quarter (512 processors) of

the supercomputer. The results of this work were published in the top ranking Journal of Physical Chemistry A (volume 111, page 4411) in 2007.

- Researchers from the University of Calgary and the University of California did experiments on trivalent aluminum ions in biochemically and geochemically relevant solution environments such as natural bodies of water. Their experiments suggested a five-fold coordination, where a four- or six-fold coordination was expected. Simulations performed by a PNNL staff member, using NWChem's plane-wave density functional theory dynamics capability, confirmed the experimental discovery of a five-coordinate $\text{Al}(\text{H}_2\text{O})_4\text{OH}^{2+}$ ion as the predominant form under ambient conditions, in stark contrast with other trivalent metal aqua ions for which there is no evidence of stable pentacoordinate products. The results of this work were published in Science (volume 308, page 1450) in 2005.

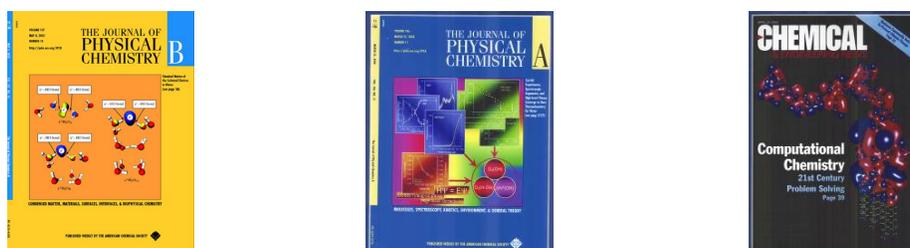


Figure 1. Some journal covers highlighting research done with the NWChem software.

NWChem is part of the Molecular Sciences Software Suite (MS³). In addition to NWChem, MS³ includes the Global Array Toolkit is the underlying parallel framework for NWChem that provides an efficient and portable shared-memory programming interface for distributed-memory computers, and the Extensible Computational Chemistry Environment (Ecce) that provides the user with a graphical user interface, scientific visualization tools, and underlying data management framework.

With the emergence of computing platforms that are a hundred teraflops, and with petascale computing platforms on the horizon, computational chemistry is on the verge of entering a new era of modeling. These huge computing resources will enable researchers to tackle scientific problems that are larger and more realistic than ever before, to include more of the complex dynamical behavior of nature, and to start asking new and different scientific questions. The next-generation supercomputing hardware, including those currently being procured by the MSCF, will likely consist of tens to hundreds of thousands of processors, a scale that could hardly be envisioned when NWChem was conceived 15 years ago. Both the changes in scientific needs and computing technology require that we examine our priorities and software capabilities to ensure that NWChem will provide its users with the cutting-edge capabilities for research in environmental molecular sciences they need on next-generation platforms.

2. Scope of the Meeting

The objectives of the NWChem developers and users meeting were to look at the major scientific questions that can be addressed by computational modeling in environmental molecular science in the next decade, and to identify the capability development needs that will allow scientists to address these questions. In addition, speakers and attendees were encouraged to discuss and provide insight into developments that could enable NWChem to effectively utilize petascale computing platforms. In addition, recommendations for bringing the NWChem software into the open-source environment were solicited. The meeting had 65 participants representing 29 institutions including 22 universities and 5 national laboratories.

The meeting consisted of sessions where scientific leaders in key programs and science areas along with key computational chemistry methodology developers, computer scientists, and mathematician discussed: (1) present and future scientific questions in environmental molecular sciences, (2) the development of new methods in computational chemistry, and (3) mathematics and computer science at the petascale. During the meeting the presentations were made available through live streaming over the internet, enabling various researchers to virtually participate in the meeting.

The insights of the meeting participants were collected in written form and through transcription of the discussion sessions. In addition, researchers that were unable to attend the meeting were provided with an opportunity to provide written feedback after the meeting. This information was summarized and provides the collective views of the participants on:

1. Major scientific questions to be addressed by computational chemistry in the next decade and new software capabilities needed to accomplish this.
2. Technical challenges to effectively utilize NWChem software on petascale computing resources.
3. Bringing NWChem software into the open-source environment.

This meeting report will serve as the cornerstone for the development of a long-term strategy for EMSL's NWChem quantum chemistry software suite, to ensure that the software capability will continue to provide EMSL and its user community with the cutting-edge capabilities they need on next-generation platforms.

3. Scientific Capability Needs

Computational chemistry is used to address scientific questions in many areas, including catalysis, solar energy, hydrogen economy, environmental remediation, and biology. Although the scientific questions can be different for each area, there are many common themes that drive the need for new and/or improved capabilities. By focusing the resources of the NWChem developer's team at EMSL on these common themes a larger return on investment can be obtained. Themes that were identified during the meeting are:

1. Kinetics and dynamics of chemical transformations
2. Chemistry at interfaces and in the condensed phase
3. Spanning longer time regimes with molecular dynamics

In the following three sections these common themes will be described. Following each theme some of the recent advancements in computational chemistry methodology and capability development will be discussed. In some cases multiple themes can benefit from these advancements. Other approaches discussed still have to mature but may lead to new ways of solving the scientific questions that are asked by the research community.

3.1. Kinetics and dynamics of chemical transformations in complex systems

To enable researchers to make major advances in areas of interest to DOE it is crucial to incorporate the kinetics and dynamics of chemical processes that drive many chemical reactions and transformations in solution or at interfaces. For example, a fundamental understanding of catalytic reaction mechanisms and the ability to accurately model reaction rates will allow scientists to predict and design catalytic activity and selectivity. This scientific challenge was outlined in the DOE Basic Energy Sciences Advisory Committee report "Opportunities for Catalysis in The 21st Century" and "Basic Research Needs for the Hydrogen Economy". The BES workshop report on the "Basic Research Needs for Solar Energy Utilization" discusses the need to gain insight into solar energy conversion processes, both in materials and biological systems, and to develop materials to harness the solar power requires the characterization of often complex chemical reaction pathways. Electron transfer reactions and redox chemistry play important roles in a number of problems that are important to DOE. For example, photocatalysis on TiO₂ surfaces can be used for the destruction of contaminants at DOE production sites, splitting of water, and mineralization of harmful organic compounds in polluted air and wastewaters. Electron transfer processes play a critical role in many biological processes. Dynamical reaction processes can exhibit complex behavior over time scales from femtoseconds to seconds that is often difficult to predict and/or identify. Understanding the kinetics and the dynamics of chemical processes will require an extensive exploration of the potential energy surface and a thorough understanding of the entropic, anharmonic effects, and possibly even quantum nuclear effects, of an ensemble of molecules. ***Obtaining accurate kinetic and thermodynamic results requires the capability to model an accurate potential energy landscape, to determine and discover reaction pathways, and to extract statistical information through sufficient sampling of the configuration space.***

3.1.1 Prediction and Discovery of Reaction Pathways

Accurate and efficient prediction of reaction pathways and transition states is essential to quantify and predict complex chemical reactions. A full suite of methods for reaction path calculations would be very valuable for efficient characterization of various possible reaction channels and exploration potential energy surfaces. The presence of a heterogeneous environment does present a host of challenging issues beyond the scope of standard techniques for finding reaction paths. Consequently, it is necessary to explicitly treat the large number of degrees of freedom and an abundance of possible reaction pathways. Modeling of the minimum energy path between the reactant and product can be accomplished using, for example, nudged elastic band (NEB) methodology. The NWChem software contains a very basic NEB implementation. More developments will be required for the efficient implementation of this method for large scale systems in finite temperature regime to determine the free energy surface. A further complication arises in the determination of a reaction pathway where only the starting point (reactant) is available. Such capability would be extremely important for studying complex chemical systems where there is a need to “discover” all possible reaction pathways. Useful methods that have been developed to tackle the “discovery” include the dimer method, hyperdynamics and metadynamics.

3.1.2 Free energy profile and sampling of reaction pathways

One of the emerging trends is the development of theories and methods to create a bridge between quantum mechanics modeling the reaction pathway and statistical mechanics, which enables the determination of the kinetics of a chemical transformation. Where once quantum chemists would only consider very few points on a reaction surface, such as the minima and the maxima, it is clear that this is insufficient to truly predict chemical reactivity, kinetics and dynamics. On the other hand, the statistical mechanics methods that have evolved largely use very empirical potentials that cannot describe true chemical reactions. NWChem does have some capabilities in this area through the coupling of its *ab initio* methodologies with the direct dynamics Venus code developed by Hase and collaborators at Texas Tech University.

Determining the free energy profile along a given reaction pathway, the so-called potential of mean force (PMF), is a central quantity for the analysis of reaction processes in complex environments. The PMF determination is computationally expensive because many energy evaluations (tens of thousands) are required for accurate statistical averaging. This presents an enormous computational burden for *ab initio* methodologies. Practical solutions to this problem need to be developed. An important development in this area is provided by an adaptive description of the chemical system where the level of theory varies depending on the particular region of the system. A particularly simple but yet very powerful example of such methodology is provided by combined quantum mechanical and molecular mechanics approach where chemical system is decomposed into two distinct types of regions; quantum mechanical (QM) region described by quantum-mechanical level of theory, and molecular mechanics (MM) region described by classical molecular mechanics methods. To improve the accuracy and applicability of QM/MM methods, a better description of interface region will be important. This involves more judicious schemes for choice of van der Waals parameters for QM atoms,

the Coulomb parameters for the MM interface atoms, as well truncation for covalent bonds crossing the interface. Another important aspect is the need for dynamical QM/MM partitioning to account for the dynamical exchange of particles in and out of the reaction region. Even with QM/MM-type of methods the expense of PMF calculations presents a formidable computational problem. One approach would be to devise appropriate strategies (often referred to as thermodynamical cycles) where bulk of statistical sampling is shifted to a more efficient and computationally cheaper representation of the molecular system. The PMF calculations are particularly well suited for petascale computing as the statistical sampling of large numbers of configurations (hundreds of thousands or even tens of millions) at multiple temperature regimes can be distributed to many groups of processors with a minimum amount of communication overhead.

3.1.3 Calculating accurate energetics

Modeling accurate kinetic and thermodynamic properties of chemical transformations requires the calculation of accurate potential energy landscapes as well as the making and breaking of bonds. (BES report on “Basic Research Needs for Clean and Efficient Combustion of 21st Century Transportation Fuels” discusses this need for combustion chemistry). To achieve chemical accuracy, or higher in the case of reaction rates and activation energies, in explorations of large portions of the ground- and excited-state potential energy surfaces, computationally expensive methods are needed, such as coupled-cluster singles and doubles with perturbative triples corrections (CCSD(T)) and its excited state counterparts. At present this is achievable for molecular systems with tens of atoms and a thousand basis functions, but it will be crucial to accomplish this feat for large complex systems such as condensed phases and interfaces. In addition, high-accuracy methods combined with molecular mechanics in high-level QM/MM simulations are giving researchers a unique chance to model chemical reactions in realistic environments.

The validation of models and new approximate methodologies used to reduce the computational effort in calculations require access to accurate benchmark results. Benchmark calculations serve as a standard to quantify the uncertainty in various approximate computational models and methods, especially where direct experimental validation cannot be obtained. The continued expansion of computing resources will enable researchers to obtain the necessary benchmark data larger molecular systems with increasingly accuracy computational chemistry methods. In addition, these high accuracy calculations can achieve chemical accuracy (<1 kcal/mol) and also serve as a validation for experimental data.

An alternative methodology would be Quantum Monte Carlo (QMC), which is an advanced many-body approach for high accuracy electronic structure calculations. While much progress has been made in the last decade, QMC still has a way to mature. It has been applied to a number of systems such as molecules, clusters and solids with up to a few hundreds of valence electrons. It typically provides about 95% of the electron-electron correlation energy and 1-2% errors for energy differences. The method has many desirable features compared to other high accuracy methods such as CCSD(T) including:

- Current implementations of QMC scale as N^3 , although work has been done to achieve linear scaling characteristics. This is significantly better than other high accuracy methods such as CCSD(T) which formally scales as N^7 .
- Zero point motions can be introduced for ions, and thermal effects can be introduced for both electrons and ions, albeit at a significant computational cost.
- QMC has been utilized to compute forces as well as excited states with the same symmetry as the ground state is described, while the calculation of optical and electronic response properties is still in its infancy.
- It is a suitable method for both molecules and solids.
- It is very (embarrassingly) parallel as a result of the Monte-Carlo approach.

In its current state it is not a black box method and direction (an initial set of nodal points) is typically needed from a more traditional *ab initio* calculation such as Hartree-Fock, MP2 or select CI (currently codes like NWChem can help by providing complete information about *ab initio* wavefunctions, including molecular orbital coefficients, CI coefficients, CSF determinants, etc.). However, it seems to have all the essential ingredients to become the method of choice for the future to calculate accurate energetics.

Accurate computational chemistry modeling of chemical transformations, or reaction pathways involving (transitions between) complex electronic states, will require the inclusion of relativistic effects, including spin-orbit coupling. For molecular systems containing heavy elements the inclusion of relativity is essential to obtain accurate insights in its redox chemistry. Redox chemistry is important in the containment and cleanup of nuclear wastes at DOE production sites. The ability to predict and control oxidation states of actinide elements, which control their solubility, is critical for disposal issues at DOE production sites, from performance assessment calculations of the safety of nuclear waste repositories to the remediation of subsurface contamination. In the latter case, two science areas directly impacted are DOE's environmental cleanup mission and DOE's Global Nuclear Energy Partnership, which requires the development of technologies that enable recycling and consumption of long-lived radioactive waste.

3.2. Chemistry at interfaces and in the condensed phase

With increased computational resources, researchers will be able to perform simulations of more realistic molecular models. Depending on the scaling characteristics of the computational method used, molecular systems of tens to hundreds of atoms are now routinely studied. However, the drive in the next five to ten years will be to model (the formation of) "real" experimentally observed condensed phase systems. The condensed phase often is a complex molecular system, combining the solid state, solution phase, and the gas-phase (see for example Figure 2).

The BES and DOE Office of Advanced Scientific Computing Research (ASCR) report on "Theory and Modeling in Nanoscience" expresses that "Within five to ten years, there must be robust tools for quantitative understanding of structure and dynamics at the nanoscale, without which the scientific community will have missed many scientific opportunities as well as a broad range of nanotechnology applications". Similar statements were made in the report from the BES Subcommittee on Theory an

Computation “Opportunities for Discovery: Theory and Computation in Basic Energy Sciences”. At the nanoscale, properties such as reactivity may be dependent on the size and shape of the molecular system. Understanding these size dependencies, and the nucleation, growth, and self-assembly of nanosized molecular systems can lead to fundamental insight in the nucleation processes in aerosol chemistry, self-assembling monolayers on surfaces, and proteins. Beyond the nanoscale, there is an interest in studying the mechanical and macroscopic properties of materials, for example the formation and progression of cracks. While atomistic simulations have provided great insight into the basic mechanisms of processes like crack formation, diffusion and phase transformations in solids, the large number of atoms in any realistic macroscopic structure is simply too large to simulate directly.

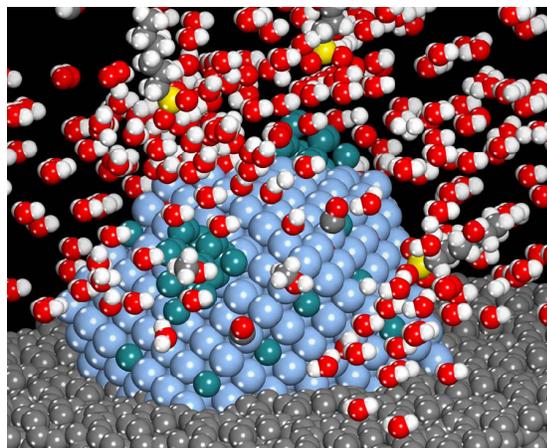


Figure 2: Complex molecular systems, with interactions at various interfaces. Graphics courtesy of Prof. Neurock (U. Virginia)

Much of the chemistry in the condensed phase happens at the boundaries of different phases, e.g. at the interface. Computational chemistry software provides researchers with tools that enable them to understand, control, and predict chemical processes and properties in condensed-phase environments and at interfaces will be crucial to answer many scientific questions. Application areas can be found in, for example, catalysis and environmental remediation. On example would be bioremediation where the interaction and charge transfer between a biological system and a surface needs to be modeled. The BES report on “Basic Research Needs for Geosciences: Facilitating 21st Century Energy Systems” outlines the needs for simulation tools “...to understand molecular-scale structure, function, energetics and dynamics processes at mineral-water interfaces”. The need to study of condensed phases with actinides under extreme conditions, such as temperature and radiation fields, for heavy elements with chemical transformations is discussed in the BES report “Basic Research Needs for Advanced Nuclear Energy Systems”, which sees “the scientific challenge is to develop a predictive capability for modeling solutions and interfacial phenomena for actinide-containing systems under extreme conditions of temperatures, pH, and high radiation fields.”

Modeling the chemistry in the condensed phase and at complex interfaces requires the use of methods that scale effectively with the system size, and the development of approaches that enable the coupling of scales ranging from atoms to macroscopic systems.

3.2.1 Improved scaling characteristics for large molecular systems

With scaling characteristics for quantum chemistry methods of N^3 for density functional theory (DFT) to N^6 and higher for coupled cluster (CC) methods, it is clear that an order of magnitude in increase in computing resources will not provide the research community with capabilities to study molecular systems that are as large as needed to study realistic

systems. To carry out simulations on biomolecular systems or nanomaterials requires the development of methodologies that better scaling behavior with the system size studied. Within the DFT framework work is underway to develop linear scaling methodologies. For example, the combination of Gaussian basis sets and Finite Elements enables the scaling behavior to reach $N^{1.0} \sim N^{1.2}$ for large molecular system sizes. Within the CC framework formulations employing localized orbital basis sets (instead of canonical Hartree-Fock orbitals) provide substantially better scaling (scale linearly in a natural way for one dimensional systems) and are ideal candidates for large scale CC simulations. An efficient parallel application of the non-iterative CC methods accounting for the effect of higher order excitations will result in a dramatic shift of the accuracies especially for the systems being beyond the reach of the current implementations of the local CC formalism.

3.2.2 Utilizing real space methodologies

It is widely appreciated that to use computational methods for the design of materials encompassing a wide assortment of elements from the Periodic Table, highly efficient solvers based as closely as possible on accurate quantum mechanics are needed. The most widely used solvers today are based on either LCAO Gaussian basis sets or the Fast Fourier transform. Both these solvers are highly efficient in serial and they have been parallelized to an intermediate numbers of processors (~100's processors). However, these methods have been shown to scale poorly for much larger numbers of processors. These inefficiencies are due, in part, to the fact that the quantum mechanical operators (H, ρ , S, etc.) require at least $O(\log P)$ communications per processor.

Solvers based on real space discretizations have the potential to use significantly more processors than the best fast solvers of today. Real space methods can in principle limit the communications per processor to $O(1)$ by only using local information to define the quantum mechanical operators. There have been a number of efforts to develop fast *ab initio* solvers based on real space (vs. Fourier or LCAO expansions) solutions to the DFT equations. Uniform finite difference grids coupled with multilevel solvers has led to significant progress in the calculation of large systems with large numbers of processors (~10K processors). While these methods are often robust enough for predicting structural properties their accuracy is often not good enough for use in *ab initio* molecular dynamics simulations. More recently, a complementary multiresolution approach using a multiwavelet basis for solving DFT equations has been developed. To date, none of these algorithms have been implemented in dynamical calculation. Another approach uses multilevel eigenvalue solver based on structured adapted mesh refinement and finite element grids. This algorithm could be used to solve problems with length scales much shorter than can be solved by FFT or uniform grid methods. Nevertheless, there were several problems with the solver that made it less efficient for AIMD than FFT approaches. For example, the energy evaluation does not lead to a smooth function. This would be disaster for force evaluations as needed in a dynamics simulation. This will be a concern for the development of the solvers based on real space discretizations.

A straightforward way to solve this problem is to use high-order FE integration, since high-order integrations describe smooth functions very efficiently. However, high-order FE is computationally complicated and results in operators that are not $O(N)$. To address

these problems several strategies could be pursued. One promising strategy is to use a mixed basis set method, in which the smooth part of the wavefunctions is described by a plane-wave expansion and the stiff part of the wavefunction is describe by a low order FE integration. Another strategy is to experiment with core mesh generation technologies. For instance, one could explore the use of optimal mesh movement for supplementing (and sometimes replacing) refinement to model multiple spatial scales. Moving-mesh techniques have the advantage of avoiding topological changes to the underlying mesh, and can be implemented using very simple data structures. One could also pursue developing quality tetrahedral volume meshes that are built around existing structures such as discrete atomic locations, partial lattice sub-meshes, or complete triangulated surfaces of molecular assemblies.

3.2.3 Coupling multiple scales

Traditionally, the modeling of macroscopic properties is based on continuum models. This model breaks down when studying properties that have their origin at the atomic scale (dislocations, grain boundaries) or when studying systems at the nanoscale. Nanoscale particles are still small enough to be strongly influenced by the atomic electronic structure. At the same time atomic scale distortions (for example, adhesion of a molecule to the surface) can influence the properties of the nanoparticle as a whole. The temporal domain in nanoscale systems can range from the frequency of individual atomic vibrations to hydrodynamic relaxation times over distances of nanometers to micrometers. However, the sheer number of atoms and degrees of freedom required to model properties in any realistic mesoscale or macroscale structure, using fully *ab initio* or atomistic methodologies, is too large to simulate directly using current or future available computing resources.

One approach is the use of mixed continuum/atomistic quasi-continuum methods, which has shown considerable promise in bridging the range of scales involved. In regions that behave like a continuum the method uses continuum assumptions to reduce degrees of freedom, while at the same time using atomistic details in the regions where it is required. In molecular dynamics simulations of microseconds are needed to model biological processes such as protein folding which requires an exploration of conformational space. More work is needed to extend these methods to model dynamics, finite temperature systems, and solutions. New approaches are needed for integrating out fast degrees of freedom from the slow ones to produce lower dimensional models that accurately reflect the dynamics and can be simulated over long periods of time.

Many of the current multiscale methods are also limited by their dependence on highly approximate empirical interatomic potentials to describe material behavior. The advanced materials applications that we are targeting (organometallics, transition metal oxides and ceramics) are typically chemically complex displaying changing bond patterns with composition environment. For such materials, empirical potentials are simply not accurate enough to capture the properties that make them important to advanced applications. Quantum chemistry codes like NWChem could provide complete information about accurate energetics and forces for chemically complex materials needed for these multiscale methodologies.

3.3 Spanning longer time regimes with molecular dynamics

The three dimensional crystal structures of biomolecules and their complexes provide a static view of the different conformational states of these systems, but do not provide information regarding the dynamics during transitions between the different states and the reactions that they induce. Transitions between distinct conformational states are driven by factors such as binding of substrates, embedding in their environment such as membranes, association in protein complexes, and dimerization and oligomerization processes. The time scale of these transitions is not known precisely but it is certainly greater than the time scale currently available to computer simulations of such systems. Accurately predicting entropic factors Molecular dynamics simulations in the time range up to tens to hundreds of nanoseconds can be routinely carried out, showing incipient motions and small amplitude conformational transitions, but they are still not sufficiently long to even reveal large conformational changes as suggested by some of the crystal structures. *In order to study biologically relevant phenomena it is necessary to significantly enhance our capability to simulate for significantly longer simulation times.*

3.3.1 Simulating longer time scales

One way is to develop a much improved parallel implementation of molecular dynamics simulations that can provide a faster time-to-solution. However, the reduction in the time-to-solution for simulations on large numbers (thousands to tens of thousands) of processors is limited due to the physical limitations of the interprocessor network. Alternatively, new techniques such as multigraining and parallel tempering should be considered. Using multigrain techniques longer time scales can be reached by treating groups of atoms as single particles, effectively reducing the number of particles and interactions when compared to single atoms currently used in molecular dynamics simulations. Parallel tempering, or replica exchange (effectively running multiple semi-independent runs simultaneously with limited communication between them), provides a mechanism to effectively sample the complete conformational space.

4. Petascale Computing for Computational Chemistry

In addition to the NWChem software capability development, it is essential to make sure that current and future capabilities available in the NWChem software will be able to scale and make effective use of the next-generation hardware architectures. Scalable quantum chemistry software will enable researchers to (1) tackle larger molecular complexes with greater accuracy, and (2) obtain their results with a faster time-to-solution. Both the new capability development and the effort to maintain scalability of the software are major efforts requiring significant resources.

Petascale computing hardware will be available in the next couple of years. Currently, large systems with tens and hundreds of thousands of processors providing hundreds of teraflops to petaflops of operations power are available to the scientific community. It is conceivable that the next-generation hardware at the EMSL will start to approach this size in the near future. To build systems beyond the petaflop range, an increase of one or two orders of magnitude in the number of processors can be expected.

The availability of these large computing resources will enable computational chemistry researchers to tackle scientific problems that are larger (tens of thousands of atoms) and more realistic than ever before, to include more of the complex dynamical behavior of nature, and to start asking new and different scientific questions that simply could not be addressed before. For example, to model the interactions and/or reactions at an interface a large complex system of atoms and molecules in both the solution phase and in the surface itself needs to be considered. However, computational chemistry software is not ready to effectively utilize the huge number of processors that are available.

The NWChem software was developed 15 years ago, and has been shown to scale to thousands of processors for certain algorithms, and the expectation is that those algorithms could scale to 10,000 processors when larger problem sizes are calculated. It is important to recognize that not all methods and algorithms will scale to petaflop systems. However, some of the scientific methodologies discussed in the previous section could benefit from multiple levels of parallelism, for example when incorporating multiple coupled scales, dynamics, and especially when sampling the configuration space for the statistical mechanics needed to model the kinetics of chemical reactions.

To take full advantage of (near) peta-scale architectures, and achieve high computational performance on such computer hardware, it will be necessary to significantly improve the efficiency of current high performance scientific software. Since NWChem was developed as a massively parallel scientific software package, achieving additional improvement for next generation architectures will be challenging. However, as with other mature modeling and simulation codes, NWChem was designed and implemented for massively parallel computers with a ‘simple’ and homogeneous structure, i.e. single processor nodes with a homogeneous node-to-node communication pattern. The next generation computer architectures are likely to have multiple core processors, multiple processor nodes, and hierarchical network fabrics. In addition improvements in CPU clock speed have far outpaced those in memory, memory bandwidth, and the interprocessor network, which dramatically changes the balance of the next generation architectures. These hardware developments and the change in balance have

consequences for the design of highly efficient scientific software, some of which will be discussed below.

It is clear that for software to scale to the large numbers of processors predicted for future supercomputing systems, communication tolerant algorithms need to be developed (i.e. less communication and more computation). Data motion becomes increasingly expensive as the relative performance (latency and bandwidth) of memory and interprocessor networks continues to fall further behind to the performance of the individual processors (see Figure 3). There are two main strategies to overcome this obstacle:

- Trade computations for communication. At a certain point, recomputation of essential data will be more efficient than storing it globally and retrieving it from remote processors.
- Overlapping communication with computation. Much of the communication costs can be hidden by simultaneously performing communication and computation operations. This approach is only effective if the granularity of the computation is coarse enough (takes enough cycles) to allow the communication operation to finish. This can be challenging, for example in molecular dynamics. Various computer science groups are developing software and strategies to effectively hide communication (Tarragon, Charm++, DMCS, Mobile Object Layer, and SMARTS), though this is still not a solved problem.

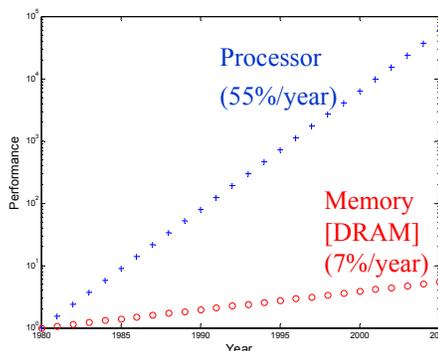


Figure 3: Increases in processor performance greatly exceed those of memory.

Development of the next-generation NWChem software capable of running efficiently on systems with tens of thousands of processors (and beyond) is a major effort that will have to include scientists from various disciplines, including computational chemists, mathematicians, and computer scientists. Expertise needs to be drawn from groups in various disciplines that are working on the development of new advanced numerical methods and scalable computational chemistry algorithms, multi-level parallelism, and are working on the development of new ways to effectively utilize large numbers of processors, network topologies, and hierarchical memory structures. To make NWChem ready for the petascale architectures of the future will require benchmarking, software infrastructure and architecture redesign, and performance tuning with emphasis on multiple levels of parallelism at the task level, the function level, and the communication and memory levels.

The primary target is to deliver a next-generation NWChem, which will be taking full advantage of the greatly enhanced hardware capabilities that will be made available the next high performance compute platform to be installed in the MSCF.

5. Open-Source NWChem Software

As mentioned before in the previous two sections, both new capability development and increasing the scalability of NWChem's large and complex code to petaflop systems and beyond are major efforts, which combined will require more resources and man-power of than is currently available in the software development team at EMSL. At the NWChem developers and users meeting the question was posed if the software should be made available and distributed as open-source, providing one way to increase the developers base. The three main arguments for going open-source are:

1. NWChem software developers at EMSL will need additional resources, both in man-power and funding, to develop the new and more complex capabilities and improve scalability on peta-scale platforms, and provide support and maintenance. Many developers interested in improving NWChem's scalability, or adding new capabilities, are reluctant to do so because of the user agreement restrictions.
2. Administration and distribution of these user agreements is becoming more difficult and costly every year due to the ever increasing number of new and returning sites.
3. Various funding agencies are mainly interested in providing programmatic funding for open-source software products, possibly opening avenues to obtain additional resources needed for the development efforts.

Making NWChem open-source will enable an extensive collaborative capability and petascale development environment. NWChem can become a real community code and the participation of external developers will increase. New and improved capabilities developed externally will be available at EMSL, further enhancing the software suite that EMSL can provide to its user community. Increased collaborative research and development will ensure that NWChem will be able to make effective use of the next-generation hardware architectures. Elimination of the user agreement will greatly increase the ease of access to the software. In addition, funding required maintaining the user agreements can be redirected to the development of new science capabilities in NWChem. One downside of going open-source is that (some) control over the software distribution and development has to be relinquished.

The participants at the meeting were in agreement there making NWChem software available under an open-source license would be beneficial to the computational chemistry community. Open-source licensing could be done using licenses such as the GNU General Public License (GPL), though various open-source licenses available (see for example <http://www.opensource.org/>) should be explored and a license should be chosen that is not too restrictive to the user community. Some discussions were focused on maintaining a high quality standard for the NWChem software, and how releases of the software could be handled in the open-source environment. A suitable distribution model could be the development of an NWChem consortium.

6. Summary

At the NWChem developers and users meeting presentations covered (1) major scientific questions that could be addressed by computational chemistry in the next decade and new software capabilities needed to accomplish this, (2) technical challenges to effectively utilize NWChem software on petascale computing resources, and (3) NWChem software and the open-source environment. Within the science areas three common themes were identified that will drive the need for new and/or improved capabilities: (1) kinetics and dynamics of chemical transformations, (2) chemistry at interfaces and in the condensed phase, and (3) spanning longer time regimes with molecular dynamics. These three areas will form the basis for focused capability development by the NWChem software developer's team.

The main theme throughout the meeting was the need for capabilities that enable scientists to accurately describe the kinetics and dynamics of chemical transformations in complex systems. This will require the capability to model an accurate potential energy landscape, to determine and discover reaction pathways, and to extract statistical information through sufficient sampling of the configuration space. Modeling the chemistry in the condensed phase and at complex interfaces requires the use of methods that scale effectively with the system size, and the development of approaches that enable the coupling of scales ranging from atoms to macroscopic systems. In order to study biologically relevant phenomena with molecular dynamics methodologies it is necessary to significantly enhance our capability to simulate for significantly longer simulation times.

Petascale quantum chemistry software will enable researchers to (1) tackle larger molecular complexes with greater accuracy, and (2) obtain their results with a faster time-to-solution. It is essential to make sure that current and future capabilities available in the NWChem software will be able to scale and make effective use of these next-generation hardware architectures.

Both the new capability development and the need to maintain scalability of the software on next-generation architectures are major efforts requiring resources, man-power, and expertise that currently exceed those available at the software development team at EMSL. A possible path forward is for the NWChem software to become an open-source development platform to enable an extensive collaborative capability and petascale development environment. This was unequivocally supported by the participants at the meeting this.

Appendix A: NWChem Overview

NWChem Introduction

NWChem is cutting-edge software that offers an extensive array of highly scalable, parallel computational chemistry methods needed to address a wide range of large, challenging scientific questions. As one of the U.S. Department of Energy's premier computational chemistry tools, NWChem is continuously developed at DOE's William R. Wiley Environmental Molecular Sciences Laboratory (EMSL) at Pacific Northwest National Laboratory (PNNL) in Richland, Washington.

NWChem aims to provide its users with computational chemistry tools that are scalable both in their ability to treat large scientific computational chemistry problems efficiently, and in their use of available parallel computing resources from high-performance parallel supercomputers to conventional workstation clusters. The software is available free of charge, subject to the terms and conditions of the EMSL Software User Agreement (<http://www.emsl.pnl.gov/docs/nwchem/download.html>).

NWChem is part of the Molecular Sciences Software Suite (MS³). In addition to NWChem, MS³ includes the Global Array Toolkit that provides an efficient and portable shared-memory programming interface for distributed-memory computers, and the Extensible Computational Chemistry Environment (Ecce) that provides the user with a graphical user interface, scientific visualization tools, and an underlying data management framework.

Science and NWChem

NWChem enables researchers to run highly scalable, parallel computations on large, challenging scientific problems. Initially, the problems of interest were focused on environmental issues, but NWChem has recently been used to solve large scientific problems in many different areas, including the examination of metal clusters, biological systems, nanostructures, and materials.

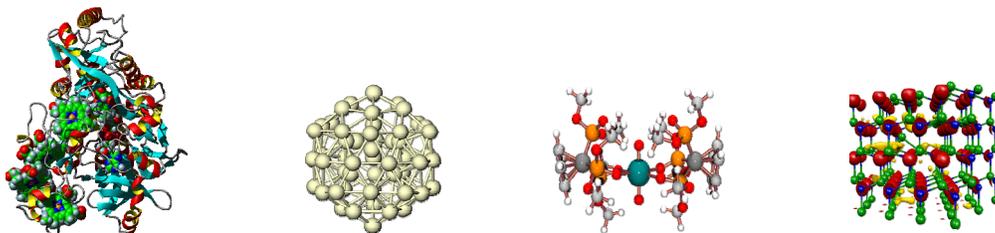


Figure 4. With NWChem, researchers can tackle molecular systems including biomolecules, nanostructures, actinide complexes, and materials.

NWChem offers an extensive array of computational chemistry methods needed to address scientific questions that are relevant to reactive chemical processes occurring in our everyday environment—photosynthesis, protein functions, and combustion, to name a few. They include a multitude of highly correlated methods, density functional theory (DFT) with many exchange-correlation functionals, plane-wave DFT with exact

exchange and Car-Parrinello, molecular dynamics with AMBER and CHARMM force fields, and combinations of them.

The NWChem computational chemistry software package runs large chemistry problems efficiently and is used by thousands of people worldwide to investigate questions about chemical processes by applying theoretical techniques to predict the structure, properties, and reactivity of chemical and biological species ranging in size from tens to millions of atoms.

Comprehensive Suite of Capabilities

NWChem provides many methods for computing the properties of molecular and periodic systems using standard quantum mechanical descriptions of the electronic wavefunction or density. Its classical molecular dynamics capabilities provide for the simulation of macromolecules and solutions, including the computation of free energies using a variety of force fields. These approaches may be combined to perform mixed quantum-mechanics and molecular-mechanics simulations.

The specific methods for determining molecular electronic structure, molecular dynamics, and pseudopotential plane-wave electronic structure and related attributes are listed in the following sections.

Molecular Electronic Structure

Methods for determining energies and analytic first derivatives with respect to atomic coordinates include the following:

- Hartree-Fock (RHF, UHF, high-spin ROHF)
- Gaussian orbital-based density functional theory (DFT) using many local and non-local exchange-correlation potentials (LDA, LSDA)
- second-order perturbation theory (MP2) with RHF and UHF references
- complete active space self-consistent field theory (CASSCF)

Analytic second derivatives with respect to atomic coordinates are available for RHF and UHF, and closed-shell DFT with all functionals.

The following methods are available to compute energies only:

- iterative CCSD, CCSDT, and CCSDTQ methods and their EOM-CC counterparts for RHF, ROHF, and UHF references
- active-space CCSDt and EOM-CCSDt approaches
- completely renormalized CR-CCSD(T), and CR-EOM-CCSD(T) correction to EOM-CCSD excitation energies
- locally renormalized CCSD(T) and CCSD(TQ) approaches
- non-iterative approaches based on similarity transformed Hamiltonian: the CCSD(2)_T and CCSD(2) formalisms.
- MP2 with RHF reference and resolution of the identity integral approximation MP2 (RI-MP2) with RHF and UHF references
- selected CI with second-order perturbation correction

For all methods, the following may be performed:

- single point energy calculations
- geometry optimization with constraints (minimization and transition state)
- molecular dynamics on the fully *ab initio* potential energy surface
- automatic computation of numerical first and second derivatives
- normal mode vibrational analysis in Cartesian coordinates
- ONIOM hybrid calculations
- Conductor-Like Screening Model (COSMO) calculations
- electrostatic potential from fit of atomic partial charges
- spin-free one-electron Douglas-Kroll calculations
- electron transfer (ET)
- vibrational SCF and DFT

At the SCF and DFT level of theory various (response) properties are available, including NMR shielding tensors and indirect spin-spin coupling.

Quantum Mechanics/Molecular Mechanics (QM/MM)

The QM/MM module in NWChem provides a comprehensive set of capabilities to study ground and excited state properties of large-molecular systems. The QM/MM module can be used with practically any quantum mechanical method available in NWChem. The following tasks are supported:

- single point energy and property calculations
- excited states calculation
- optimizations and transition state search
- dynamics
- free energy calculations

Pseudopotential Plane-Wave Electronic Structure

The NWChem Plane-Wave (NWPW) module uses pseudopotentials and plane-wave basis sets to perform DFT calculations. This method's efficiency and accuracy make it a desirable first principles method of simulation in the study of complex molecular, liquid, and solid-state systems. Applications for this first principles method include the calculation of free energies, search for global minima, explicit simulation of solvated molecules, and simulations of complex vibrational modes that cannot be described within the harmonic approximation.

The NWPW module is a collection of three modules:

- PSPW (PSeudopotential Plane-Wave) – A gamma point code for calculating molecules, liquids, crystals, and surfaces.
- Band – A band structure code for calculating crystals and surfaces with small band gaps (e.g. semi-conductors and metals).
- PAW (Projector Augmented Wave) – a gamma point projector augmented plane-wave code for calculating molecules, crystals, and surfaces

These capabilities are available:

- constant energy and constant temperature Car-Parrinello molecular dynamics (extended Lagrangian dynamics)
- LDA, PBE96, and PBE0, exchange-correlation potentials (restricted and unrestricted)
- SIC, pert-OEP, Hartree-Fock, and hybrid functionals (restricted and unrestricted)
- Hamann, Troullier-Martins, Hartwigsen-Goedecker-Hutter norm-conserving pseudopotentials with semicore corrections
- geometry/unitcell optimization, frequency, transition-states
- automated wavefunction initial guess with linear combinations of atomic orbitals
- constraints needed for potential of mean force (PMF) calculation
- wavefunction, density, electrostatic, Wannier plotting
- band structure and density of states generation

Molecular Dynamics

The NWChem Molecular Dynamics (MD) module can perform classical simulations using the AMBER and CHARMM force fields, quantum dynamical simulations using any of the quantum mechanical methods capable of returning gradients, and mixed quantum mechanics molecular dynamics simulation and molecular mechanics energy minimization.

Classical molecular simulation functionality includes the following methods:

- single configuration energy evaluation
- energy minimization
- molecular dynamics simulation
- free energy simulation (MCTI and MSTP with single or dual topologies, double-wide sampling, and separation-shifted scaling)

The classical force field includes the following elements:

- effective pair potentials
- first-order polarization
- self-consistent polarization
- smooth particle mesh Ewald
- twin-range energy and force evaluation
- periodic boundary conditions
- SHAKE constraints
- constant temperature and/or pressure ensembles
- dynamic proton hopping using the Q-HOP methodology
- advanced system setup capabilities for biomolecular membranes

Scaling NWChem to large numbers of processors

The NWChem software has been shown to scale to large numbers of processors available at various supercomputer sites across the United States. High accuracy methods such as coupled cluster and MP2 perturbation theory have been shown to scale to at least 1024 processors. EMSL's supercomputer was capable of running a coupled cluster with perturbative triples calculation on eight water molecules on 1840 processors for 37 hours, running at 63% of peak efficiency, utilizing the high speed interconnect and memory available. Density functional theory calculations have also been shown to scale, as can be seen in figure 5. Additional benchmark results, for example for the molecular dynamics module in NWChem can be found on the NWChem web pages.

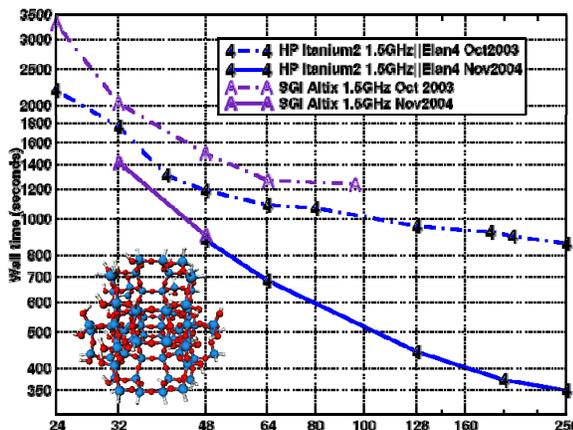


Figure 5. Scalability of density functional theory energy calculation for $\text{Si}_{75}\text{O}_{148}\text{H}_{66}$ cluster on the EMSL supercomputer.

Partnering with NWChem

NWChem was designed in a modular fashion to make it easier to develop new modules, to extend current modules, and interface between modules and with other programs. Programs that have an interface with NWChem include VENUS (Texas Tech University), POLYRATE, and Python. A list of collaborators can be found on the NWChem web site.

Platform Availability

NWChem is designed to provide the best time-to-solution on high-performance parallel supercomputers, but also performs well on desktops, clusters of conventional workstations. It is uniquely tailored to be scalable to thousands of processors, both in its ability to treat large problems efficiently and in its use of available parallel computing sources to obtain results faster. NWChem runs on a wide variety of architectures, interconnects, and operating systems. A full list of platforms and compilers is available on the NWChem website (<http://www.emsl.pnl.gov/docs/nwchem/nwchem.html>).

Access to New Science in NWChem Version 5.0

With the new NWChem 5.0 release many new capabilities have now become available to the user community. These new capabilities within the highly scalable framework of NWChem provide computational chemists with a unique tool to perform cutting-edge science, as described in the following sections.

High Accuracy Is Scalable to Large Problem Sizes

Major improvements and optimizations resulting in a factor 2-4 speedup in the TCE code and the closed-shell CCSD(T) code now enable NWChem users to address scientific problem sizes that could not be addressed before. For example, recent open-shell CCSD(T) calculations performed at PNNL employed more than 600 basis functions. Larger open-shell systems with over a 1000 basis functions should be possible.

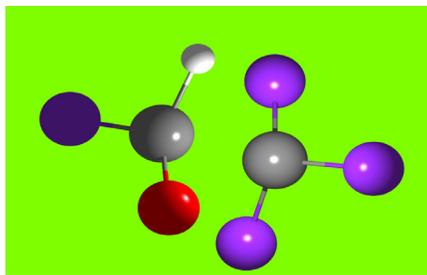


Figure 6. Open-shell CCSD(T) on CF_3CHFO with 606 basis functions

Extensive QM/MM Capabilities Are Seamlessly Integrated

The new QM/MM module provides a seamless integration between molecular mechanics and most quantum-mechanical theories in NWChem. It boasts an extensive array of capabilities geared toward comprehensive description of ground and excited state, and dynamical properties of large molecular systems for chemistry and biology.

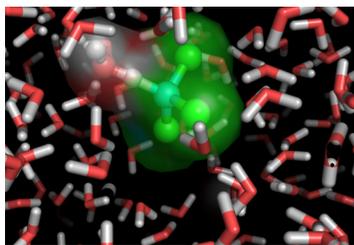
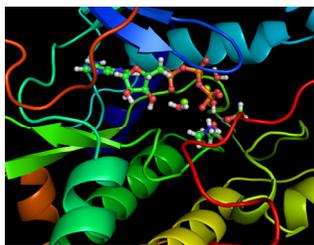


Figure 7. Applying QM/MM: Modeling catalytic properties of protein kinases (left), and simulating reactions in aqueous solutions (right).

For example, QM/MM was used to model the catalytic properties in proteins.

High Accuracy Is Achievable for Excited States

Equation of Motion (EOM) CCSD(T), combined with completely renormalized techniques, provides an ideal framework for large-scale calculations of singly and partially doubly-excited states. These high-accuracy methods can be combined with QM/MM to provide a unique capability in NWChem.

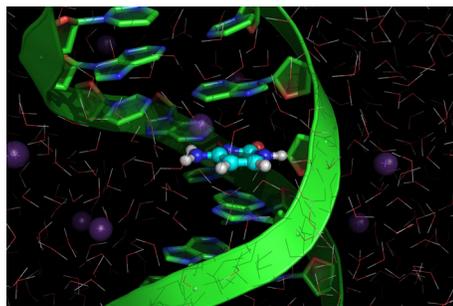


Figure 8. Photostability of cytosine in the native DNA environment, utilizing NWChem's new combined EOM-CCSD(T) and QM/MM capability.

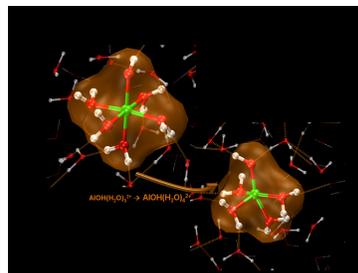
Plane Wave Has Exact Exchange

The NWChem Plane-Wave module now enables users to use exact exchange and the Self-Interaction Correction (SIC) within its framework for complex molecular, liquid, and solid-state systems. In addition, great improvements in parallel performance allow users to study larger systems at longer time scales with a faster time-to-solution.

As a scientific example, NWPW Car-Parrinello simulations supported the experimental discovery of a five-coordinate $\text{Al}(\text{H}_2\text{O})_4\text{OH}^{2+}$ ion as the predominant form under ambient

conditions, in stark contrast with other trivalent metal aqua ions for which there is no evidence of stable pentacoordinate products (see Science **308**, 1450 [2005]).

Figure 9. Unique NWPW simulations provide evidence of a five-coordinate $\text{Al}(\text{H}_2\text{O})_4\text{OH}^{2+}$ ion



Molecular Dynamics Allows Protons to Hop

The Molecular Dynamics (MD) module of NWChem now includes the capability for dynamic proton hopping using the new, efficient Q-HOP methodology developed by the Volkhard group at the University of Saarland, Germany. This technique relies on classical MD simulations in which stochastic instantaneous transfer between neighboring proton accepting groups is determined by an easily computable probability. This probability is based on a modified version of Transition State Theory incorporating tunneling at a semi-classical level.

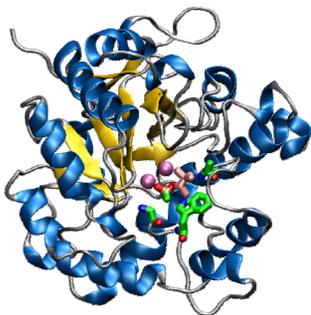


Figure 10. MD simulations provide insight into why mutations in the active site of wild-type enzyme organophosphorous hydrolase result in a 4 orders-of-magnitude increase of catalytic activity

Relevant Web Sites

NWChem <http://www.emsl.pnl.gov/docs/nwchem/nwchem.html>

GA Toolkit <http://www.emsl.pnl.gov/docs/global/>

Ecce <http://ecce.emsl.pnl.gov/index.shtml>

Appendix B: Meeting Agenda

Thursday, January 25, 2007

TIME	TOPIC	SPEAKER
7:30-8:00	Registration	
8:00-8:15	Welcome and announcements	Allison Campbell (EMSL)
SESSION I: NWCHEM AND THE MOLECULAR SCIENCE COMPUTING FACILITY		
8:15-8:50	NWChem Status and Future Directions	Bert de Jong (EMSL)
8:50-9:10	Highlights from Computational Research at EMSL	Erich Vorpagel (EMSL)
SESSION II: CHEMICAL REACTIONS AND DYNAMICS		
9:10-9:50	Computing Needs to Model Chemical Transformations: A DOE/BES Perspective	Bruce Garrett (PNNL)
9:50-10:10	Informal Discussions	
10:10-10:50	Toward More Realistic Simulations for Large-size Reactive Systems by Combining Quantum Mechanics and Molecular Mechanics in New Ways	Hai Lin (U. Colorado Denver)
10:50-11:30	NWChem in 2022	Robert Harrison
11:30-12:00	Panel: Computational needs for science area	Chair: Bert de Jong
12:00-13:30	Working Lunch with Topic: Open-source NWChem	
SESSION III: METHODS FOR LARGE SCALE SIMULATIONS		
13:30-14:40	<u>Bill Lester was unable to fly in due to extreme fog. Prof. Mitas kindly agreed to provide an intro to QMC</u> Quantum Monte Carlo advances: pfaffian wavefunctions, topology of fermion nodes and QMC/MD methods	Lubos Mitas (North Carolina State U.)

14:40-15:00	Informal Discussions	
15:00-15:40	Toward a practical DFT for large systems	Kimihiko Hirao (U. Tokyo)
15:40-16:20	Component Architectures for Quantum Chemistry: Forging New Capabilities and Insights	Curtis Janssen (Sandia Nat. Lab.)
16:20-16:50	Panel: Large scale simulations on petascale computing resources	Chair: Eric Bylaska
17:00-18:00	Poster session	Chair: Dunyou Wang

Due to extreme fog and freezing fog, some of our speakers are unable to fly in and participate in person at the meeting. Absent speakers were replaced by scientists from EMSL and PNNL. The presentations of the absent speakers are available as additional information, and were included in the report.

Friday, January 26, 2007

TIME	TOPIC	SPEAKER
SESSION IV: BIOLOGICAL SYSTEMS AND INTERFACES		
8:00-8:30	NWChem Molecular Dynamics Simulations: Current Developments and Applications	Tjerk Straatsma (PNNL)
8:30-9:00	Can Biomolecular Simulations Benefit from Petascale Computing? Issues and Perspectives	Roberto Lins (PNNL)
9:00-9:20	Large scale biomolecularsimulations with NWChem QM/MM module	Marat Valiev (EMSL)
9:20-9:40	Informal Discussions	
SESSION V: MATERIALS AND TRANSPORT		
9:40-10:20	Polaron transport in metal oxides: A Frontier for Computational Chemistry	Kevin Rosso (PNNL)
10:20-11:00	Real-space method with multigrid acceleration (RMG): parallelization, scaling and applications to electronic structure and quantum transport problems	Wenchang Lu (North Carolina State U.)
11:00-11:30	Panel: Computational needs for science areas	Chair: Tjerk Straatsma

11:30-13:00	Working Lunch with topic: New and emerging science areas	
SESSION VI: MATHEMATICS AND COMPUTER SCIENCE FOR LARGE SCALE SIMULATIONS		
13:00-13:40	An inner technical look at Interactive Supercomputing's Star-P for Parallel Computing with MATLAB(r) and python and other clients	Alan Edelman (MIT)
13:40-14:20	Toward Solving the Multiparticle Schrodinger Equation via an Unconstrained Sum of Slater Determinants	Gregory Beylkin (U. Colorado Boulder)
14:20-14:30	Informal Discussions	
SESSION VII: GENERATION OF PARALLEL SOFTWARE FOR HIGH ACCURACY METHODS		
14:30-15:10	Large molecule applications of coupled-cluster theory: Parallel Implementations and Natural Linear Scaling	Rod Bartlett (U. Florida)
15:10-15:50	Compiler/Runtime Optimizations for the Tensor Contraction Engine	Saday Sadayappan (Ohio State U.)
15:50-16:20	Panel: Needs for computer science and mathematics developments	Chair: Bert de Jong
16:20-16:30	Closing and adjourn	

Appendix C: List of Participants

Edgar Arcia

Tri-Cities Prep
earcia@tcprep.org

Edo Apra

Oak Ridge National Laboratory
aprae@ornl.gov

Jochen Autschbach

University at Buffalo, State of NY
jochena@nsm.buffalo.edu

Rod Bartlett

University of Florida
bartlett@qtp.ufl.edu

Doug Baxter

Pacific Northwest National Laboratory
douglas.baxter@pnl.gov

Gregory Beylkin

University of Colorado, Boulder
beylkin@boulder.colorado.edu

Gary Black

Pacific Northwest National Laboratory
gary.black@pnl.gov

Petia Bobadova-Parvanova

Central Michigan University
petia@phy.cmich.edu

Adrian Brown

NASA Ames Research Center
abrown@arc.nasa.gov

Eric Bylaska

Pacific Northwest National Laboratory
eric.bylaska@pnl.gov

Bill Cannon

Pacific Northwest National Laboratory
bill@pnl.gov

Donald Camaioni

Pacific Northwest National Laboratory
donald.camaioni@pnl.gov

Allison Campbell

Pacific Northwest National Laboratory
allison.campbell@pnl.gov

Tom Cheatham III

University of Utah
tec3@utah.edu

Aurora Clark

Washington State University
auclark@wsu.edu

Huib van Dam

CCLRC Daresbury Laboratory
h.j.j.vandam@dl.ac.uk

Nathaniel Deskins

Pacific Northwest National Laboratory
nathaniel.deskins@pnl.gov

Ram Devanathan

Pacific Northwest National Laboratory
ram.devanathan@pnl.gov

Michel Dupuis

Pacific Northwest National Laboratory
michel.dupuis@pnl.gov

Robert Eades

IBM
raeades@us.ibm.com

Alan Edelman

Massachusetts Institute of Technology
edelman@mit.edu

Andy Felmy

Pacific Northwest National Laboratory
ar.felmy@pnl.gov

Bruce Garrett
Pacific Northwest National Laboratory
bruce.garret@pnl.gov

John George
IBM Federal
johngeo@us.ibm.com

Koushik Ghosh
IBM Federal
kghosh@us.ibm.com

Bojana Ginovska
Washington State University-Tri Cities
bojana.ginovska@pnl.gov

Jeff Hammond
University of Chicago
jhammond@uchicago.edu

Robert Harrison
Oak Ridge National Laboratory
harrisonrj@ornl.gov

Kimihiko Hirao
University of Tokyo
hirao@qcl.t.u-tokyo.ac.jp

Florian Janetzko
University of Calgary
f.janetzko@ucalgary.ca

Curtis Janssen
Sandia National Laboratory
cljanss@sandia.gov

Bert de Jong
Pacific Northwest National Laboratory
bert.dejong@pnl.gov

Sebastien Kerisit
Pacific Northwest National Laboratory
sebastien.kerisit@pnl.gov

Karol Kowalski
Pacific Northwest National Laboratory
karol.kowalski@pnl.gov

Manajkumar Krishanan
Pacific Northwest National Laboratory
manoj@pnl.gov

Bill Lester
University of California
walester@lbl.gov

Jose Lopez
Texas Tech University
jose.g.lopez@ttu.edu

Jun Li
Pacific Northwest National Laboratory
jun.li@pnl.gov

Hai Lin
University of Colorado
hai.lin@cudenver.edu

Roberto Lins
Pacific Northwest National Laboratory
roberto.lins@pnl.gov

Wenchang Lu
North Carolina State University
luw@nemo.physics.ncsu.edu

Lubos Mitas
North Carolina State University
lmitas@unity.ncsu.edu

Chris Mundy
Pacific Northwest National Laboratory
chris.mundy@pnl.gov

Jarek Nieplocha
Pacific Northwest National Laboratory
jarek.nieplocha@pnl.gov

Matt Neurock
University of Virginia
mn4n@virginia.edu

Patrick Nichols
Pacific Northwest National Laboratory
Patrick.nichols@pnl.gov

Mohamed Osman
Washington State University
osman@eecs.wsu.edu

Vencislav Parvanov
Pacific Northwest National Laboratory
vencislav.parvanov@pnl.gov

Fabrizio Petrini
Pacific Northwest National Laboratory
fabrizio.petrini@pnl.gov

Monty Pettit
University of Houston
pettitt@uh.edu

Juan Piernas
Pacific Northwest National Laboratory
juan.piernascanovas@pnl.gov

Julia Rice
IBM & Science Advisory Committee
julia@almaden.ibm.com

Robert Rittenhouse
Walla Walla College
rittro@wwc.edu

Kevin Rosso
Pacific Northwest National Laboratory
kevin.rosso@pnl.gov

Jim Rustad
University of California, Davis
rustad@geology.ucdavis.edu

Saday Sadayappan
Ohio State University
sadayappan.1@osu.edu

Thereza Soares
Pacific Northwest National Laboratory
tasoares@pnl.gov

Vinod Tipparaju
Pacific Northwest National Laboratory
vinod@pnl.gov

Tjerk Straatsma
Pacific Northwest National Laboratory
tps@pnl.gov

Arun Venkatnathan
Pacific Northwest National Laboratory
arun.venkatnathan@pnl.gov

Erich Vorpapel
Pacific Northwest National Laboratory
erich.vorpapel@pnl.gov

Matthew Wander
Stony Brook University
mwander@notes.cc.sunysb.edu

Dunyou Wang
Pacific Northwest National Laboratory
dunyou.wang@pnl.gov

Sergio Wong
University of California, San Diego
swong@mccammon.ecse.edu

Yan Zhao
University of Minnesota
yshao@chem.umn.edu

Appendix D: Presentation Abstracts

The full presentations are available in electronic form on the NWChem meeting CDROM.

Bert de Jong

Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory

“NWChem Status and Future Directions”

The NWChem development team is committed to providing researchers, at the EMSL and worldwide, with the software resources they need for discovery and technological innovation in computational molecular sciences. In this presentation an overview of the NWChem software will be given, the current status of NWChem will be discussed, and some future directions will be highlighted.

Erich Vorpapel

Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory

“Highlights from Computational Research at EMSL”

The EMSL is committed to providing researchers with the resources they need for discovery and technological innovation in the environmental molecular sciences to support DOE and the nation. The purpose of this presentation is to provide a broad overview of accomplishments made by users of the Molecular Science Computing Facility. A set of highlight slides will be presented from recently completed, and some on going Computation Grand-Challenge projects.

Bruce C. Garrett

Chemical & Materials Sciences Division, Pacific Northwest National Laboratory

“Computing Needs to Model Chemical Transformations: A DOE/BES Perspective”

Providing a detailed understanding of the factors controlling bond-breaking/making processes in condensed phases will impact a number of areas important to DOE’s mission, ranging from catalysis, which is important in energy related applications and in environmental systems, to acid-base chemistry and redox processes important in the environment, and to charge transfer processes in energy production systems (e.g., solar cells and fuel cells), radiation biology, and bioremediation. Computational chemistry is well positioned to provide the fundamental understanding needed to realize control of chemical transformations. Predictions of reaction mechanisms and rate constants require:

- accurate electronic structure theories to obtain information about the potential energy surfaces,
- statistical mechanical approaches for sampling the relevant configurations contributing to the reaction,
- dynamical theories for calculating the overall rates of reaction, and
- multiscale theories, such as master equations, for inferring overall rate coefficients and their pressure dependences from the above.

Although challenges do still exist, significant progress has been made in all of these areas and reliable predictions of gas-phase reaction rates is now possible. A major scientific challenge for the future is to develop computational tools that provide the understanding required to control chemical reactions in condensed phase environments and at complex interfaces at the same level of detail we understand and control gas-phase reactions today. This presentation will overview recent advances and future needs in statistical mechanical and dynamical approaches, which when efficiently coupled with electronic structure methods in NWChem will allow us to harness petascale computing to address the important challenge of understanding how to control condensed-phase chemical transformations.

Hai Lin¹ and Donald G. Truhlar²

¹University of Colorado at Denver and Health Sciences Center, ²University of Minnesota

“Toward More Realistic Simulations for Large-size Reactive Systems by Combining Quantum Mechanics and Molecular Mechanics in New Ways”

We will present our recent progress in methodology and computer programs for the modeling of large-size reactive systems by combining quantum mechanics (QM) and molecular mechanics (MM). These methods can be applied for multiscale modeling or whenever one needs to treat a portion of a large system at a higher level than the whole system. Two kinds of advances will be discussed, and some future developments will be outlined. The first set of advances to be discussed includes polarized-boundary[1] and flexible-boundary QM/MM schemes, which account for self-consistent mutual polarization and charge transfer between the QM and MM moieties to build more seamless connections between them. These methods are especially powerful when combined with the redistributed charge algorithm[2] for treating the QM-MM boundary.

The second kind of advance involves new techniques for dynamics simulations, including

(1) adaptive partitioning schemes,[3] which allow atoms to switch between the QM and MM subsystems or to change their QM or MM characteristics during the trajectory propagation or any molecular dynamics simulation. and (2) QM/MM-based multiconfiguration molecular mechanics,[4,5] which is highly efficient in generating potential energy surfaces for dynamics calculations at the level of variational transition state theory with multi-dimensional tunneling contributions or at other levels of dynamical simulation. These advancements will contribute to more accurate, more efficient, and more realistic molecular modeling and simulations in the areas of chemistry, biology, and material sciences.

Acknowledgments. We are grateful for the help of Andreas Heyden, Oksana Tishchenko, Yan Zhang, and Yan Zhao. This work is supported by the Department of Energy, the National Science Foundation, the Office of Naval Research, Research Corporation, the U. S. Army Research Office, and the Minnesota Supercomputing Institute.

[1] Zhang, Y.; Lin, H.; Truhlar, D. G., *J. Chem. Theory Comput.*, submitted.

[2] Lin, H.; Truhlar, D. G., *J. Phys. Chem. A* **2005**, *109*, 3991.

[3] Heyden, A.; Lin, H.; Truhlar, D. G., *J. Phys. Chem. B*, in press.

[4] Lin, H.; Zhao, Y.; Tishchenko, O.; Truhlar, D. G., *J. Chem. Theory Comput.* **2006**, *2*, 1237.

[5] Lin, H.; Truhlar, D. G., *Theor. Chem. Acc.* **2007**, in press [available in On-line First at DOI 10.1007/s00214-006-0143-z]

William A. Lester, Jr.†

Kenneth S. Pitzer Center for Theoretical Chemistry, University of California at Berkeley

“Linear-Scaling Evaluation of the Local Energy in Quantum Monte Carlo”

For atomic and molecular quantum Monte Carlo calculations, most of the computational effort is spent in the evaluation of the local energy. We have developed a scheme for reducing the computational cost of the evaluation of the Slater determinants and correlation function for the correlated molecular orbital ansatz. A sparse representation of the Slater determinants makes possible efficient evaluation of molecular orbitals. A modification to the scaled distance function facilitates a linear scaling implementation of the Schmidt-Moskowitz-Boys-Handy (SMBH) correlation function that preserves the efficient matrix multiplication structure of the SMBH function. For the evaluation of the local energy, these two methods lead to asymptotic linear scaling with respect to the size of the molecule.

†Also, Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720.

Lubos Mitas*

North Carolina State University

“Quantum Monte Carlo advances: pfaffian wavefunctions, topology of fermion nodes and QMC/MD methods”

Quantum Monte Carlo (QMC) is an advanced many-body approach for high accuracy electronic structure calculations. It has been applied to a number of systems such as molecules, clusters and solids with up to a few hundreds of valence electrons. It typically provides about 95% of the electron-electron correlation energy and 1-2% errors for energy differences. Recent developments open new avenues to attack the only approximation involved (the fixed-node approximation) by understanding the topology of fermion nodes. In particular, we explicitly prove that for $d > 1$ fermion ground states have the minimal number of two nodal cells for arbitrary system size under very general conditions. We show that pairing wavefunctions, such as the Bardeen-Cooper-Schrieffer and pfaffian wavefunctions, have this important property and enable to decrease the fixed-node errors in a very compact manner. Finally, I will mention recent progress in coupling of QMC with the molecular dynamics, ie, development of the dynamical $T > 0$ approach within the correlated wavefunctions framework.

* in collaboration with L.K. Wagner, M. Bajdich, G. Drobný, K.E Schmidt (Arizona State U.), J.C. Grossman (UC Berkeley).

Kimihiko Hirao

University of Tokyo

“Toward a practical DFT for large systems”

Owing to the theoretical developments and high-speed computers, quantum chemistry can now describe the properties of small to medium size molecules with chemical accuracy (2 kcal/mol or 0.1eV) comparable to those of experiment. The focus of theoretical chemists' interest is moving from accurate investigations of small molecules to high-speed computations of large systems. We are entering a new period when the computer simulations can be carried out for large systems, e.g. biomolecules and nanomaterials.

Density Functional Theory (DFT) may be the only tool that enables us to carry out accurate simulations for larger systems with reasonable computational cost. If practical DFT is developed, which can handle biomolecules and nanomaterials, we can enlarge greatly the scope of computational chemistry.

I will talk on an efficient algorithm to evaluate the Coulomb integrals, dual-level approach to DFT, and accurate description of van der Waals interactions.

Robert J. Harrison

University of Tennessee and ORNL

“NWChem in 2022”

NWChem is now about 15 years old. It is mostly written in a 30-year old programming language. If you look around, you'll see that most of us are 40+. In the past 15 years, in addition to our aging, much has changed in computational science and computational chemistry in particular. We have machines with 100+K processors, and an ever increasing array of new and complex algorithms for fast and accurate calculations on large molecular systems. Looking ahead 15 more years to 2020, the pace of change will even greater, and we can anticipate computers with literally 100s of millions of processors and exciting theoretical or algorithmic breakthroughs that revolutionize our discipline. How do we shepherd NWChem forward into this new era and keep it at the forefront of both scientific capability and computational

performance? What can we contribute that is of greatest value to our scientific community and is also well aligned with the anticipated missions of our major funding agency?

I don't have answers to these questions, but I do have some opinions and ideas to contribute to discussion.

Monty Pettit

University of Houston

“Simulations in Biomaterials and Biology”

Biology and bionanotechnology have intrinsic scales of heterogeneity. The lack of homogeneity provides significant challenges in simulation of events and thermodynamics. The accuracy and scaling of current methods is an important consideration if we are to make efficient use of peta-scale computational resources. In particle dynamics, both at the atomic and meso scale, long range forces consume most of the computational effort. Fast multipole methods have better computational scaling than fourier Ewald methods for handling long range forces. A major limitation of fast summation methods in the form of fast multipole algorithms has been the number of particles required to break even against other more traditional methods to handle long range forces in particle simulations. We have a new load balanced parallel implementation of a non-adaptive version of Greengard and Rokhlin's fast multipole method for distributed memory architectures with focus on applications in molecular dynamics. We introduce a novel load balancing and communication overlapping scheme. Our implementation includes periodic boundary conditions calculations with lattice sums and facilitates multiple time stepping techniques without sacrificing determinism of computation and demonstrates strong scaling.

Thomas Cheatham III

University of Utah

“The future of biomolecular simulation at the petascale: Plague or panacea?”

For the past decade, biomolecular simulators have been riding a wave of excitement brought about by advances in the methods and simulations that now enable routine simulation of moderate scale biomolecular systems (~10-100K atoms) on the 10-100 ns time scale on terascale machines. The emergence of petascale computing fuels further excitement and the expectation that we can jump three orders of magnitude in scale, for example to improve the granularity/accuracy of QM/MM simulations, make multiple microsecond MD simulations of biomolecules routine, or to enable realistic calculations of very large biomolecular assemblies. The reality is that this simple scaling from the terascale to the petascale hides the technical (and political) realities. To enable our community, there needs to be a concerted effort to improve the codes, develop new approaches, and improve the programming models. This is not easy as we have been lulled into complacency by the uniformity of the machines (i.e. clusters with Myrinet or Infiniband and MPI) and algorithms (such as particle mesh Ewald and its 3D FFT for explicitly solvated molecular dynamics simulations and generalized Born methods for implicit solvent). Contrary to what was seen in the late 80's and early 90's where a variety of heterogeneous machines existed, each with distinct programming models and therefore specially tailored versions of the MD codes, in the last decade everything has been about clusters and simplifying/generalizing the codes. History repeats and we come back into the era of heterogeneous machines (with limited memory bandwidth, in-core vs. on-board vs. inter-node communications, and special purpose hardware such as FPGA's) that promise to complicate the situation enormously as we strive for the petascale. We will discuss the embarrassingly non parallel nature of biomolecular simulation and engage dialogue on the question of how we may make use of the new era of capacity/capability computing for biomolecular simulation.

Matthew Neurock

Departments of Chemical Engineering and Chemistry, University of Virginia

“First Principles Elucidation and Design of Catalytic and Electrocatalytic Materials: Progress, Challenges and Perspectives”

Heterogeneous catalysis is at the heart of over 90% of all chemical transformations of molecules into useful products. Catalytic processes are responsible for the production of the energy necessary to power our homes and cars, the purification of the air that we breathe; the fabrication of the materials we use each day, the manufacturing of the foods that we eat, and the synthesis of pharmaceutical intermediates. Knowledge of the atomic structure at and near the active sites of the catalyst and how it influences reactivity could revolutionize our ability to design more active and selective catalysts. Significant advances in both theory and simulation have occurred over the past decade thus making theoretical chemistry an invaluable partner to experiment in this endeavor. Theory and simulation can be used to begin to predict properties of different catalytic surfaces. The nature and strength of the adsorbate-surface bonds that form are important in dictating its surface chemistry. The influence of the local molecular environment within the vicinity of the active site, however, can be just as important. We can model the atomic structure along with the effects of the local molecular topography about the active site. We can also begin to probe ways to exploit them in the design of active metal particles. There are, however, significant limitations as the result of the inaccuracies of the quantum mechanical methods used and perhaps more importantly the size of the models that can be constructed.

In this talk, we describe the use of *ab initio* methods along with *ab initio*-based dynamic and kinetic methods to simulate the catalytic performance over supported metal particles. We show that it is now possible to track the nature of the active surface site along with the local “molecular” environment about the active surface ensemble. We specifically probe the influence of surface coverage, bimetallic alloys, and the molecular networks that form at a liquid/metal interface on catalytic activity. This talk will focus on the application of these tools to simulating the synthesis of oxygenates and the electrocatalytic oxidation of methanol and CO. We present some of the opportunities as well as the challenges.

Wenchang Lu

Center for High Performance Simulation and Department of Physics, North Carolina State University

“Real-space method with multigrid acceleration (RMG): parallelization, scaling and applications to electronic structure and quantum transport problems”

Quantum simulations of materials with thousands of atoms present challenging issues for computational physicists and chemists, even when using density functional theory. I will describe the essential elements and some applications of a real-space multigrid (RMG) approach, which has been developed for performing very large calculations on massively parallel computers. In this method, all the wave functions, charge density, and potentials are represented on grids in real space. Multigrid techniques provide preconditioning and convergence acceleration at all length scales and therefore leads to particularly efficient algorithms. The data are distributed over the processors quite evenly and lead to simple and efficient parallelization over thousands of processors. Both ultrasoft and norm conserving pseudopotentials are implemented, using a sequence of grid resolutions when needed.

By expanding the DFT total energy in variationally-optimized non-orthogonal orbitals strictly localized in overlapping localization regions, it is possible to perform accurate calculations for over 2000 atoms on massively parallel computers while using only a minimal basis. The variational optimization is performed for each atomic configuration, preserving the accuracy and the compactness of the basis. The minimal localized basis is particularly useful for studying quantum transport in nanoscale devices. Due to the localization and the structure of the quantum transport

problem, the Green's functions needed for the calculation of electron transmission become block-diagonal and can be iteratively computed in $O(N)$ time.

I will also discuss a hybrid quantum molecular dynamics method in which the central region is treated by density functional theory and the surrounding environment by a modified Thomas-Fermi (TF) approximation. The hybrid method enables quantum mechanical simulations for fairly large protein fragments in solution, since most of the water molecules are treated by the inexpensive TF method.

The methodology will be illustrated by several applications, including: (i) determination of surface atomic structures through their optical signatures, and (ii) quantum transport properties of organic molecules sandwiched between semiconducting or metallic leads.

In collaboration with V. Meunier, M. Hodak, F. Ribeiro, Q. Zhao, G. Schmidt and J. Bernholc.

Alan Edelman

MIT and Interactive Supercomputing

“An inner technical look at Interactive Supercomputing's Star-P for Parallel Computing with MATLAB(r) and python and other clients”

Star-P is a unique technology offered by Interactive Supercomputing after nurturing at MIT. Star-P through its abstractions is solving the ease of use problem that has plagued supercomputing.

Some of the innovative features of Star-P are the ability to program in MATLAB, hook in task parallel codes written using a processor free abstraction, hook in existing parallel codes, and obtain the performance that represents the HPC promise. All this is through a client/server interface. Other clients such as Python or R could be possible. The MATLAB, Python, or R becomes the "browser."

Parallel computing remains challenging, compared to serial coding but it is now that much easier compared to solutions such as MPI. Users of MPI can plug in their previously written codes and libraries and continue forward in Star-P.

Gregory Beylkin¹, Martin J. Mohlenkamp², and Fernando Perez¹

¹University of Colorado at Boulder, ²Ohio University

“Toward Solving the Multiparticle Schrodinger Equation via an Unconstrained Sum of Slater Determinants”

We are developing a method for solving the multiparticle Schrödinger equation using an unconstrained sum of Slater determinants. Current methods impose additional structural constraints on the determinants, such as orthogonality between orbitals or an excitation pattern. In our approach no prior information is needed.

We use an integral formulation of the problem, a Green's function iteration, and a fitting procedure based on the computational paradigm of separated representations. The core procedure is the construction and solution of a matrix-integral system derived from antisymmetric inner products.

Our ultimate goal is to complement existing methodology with general-purpose, automatically adaptive methods; this talk will describe the status of the project and the initial numerical results.

Rod Bartlett, Norbert Flocke, Tom Hughes, Victor Lotrich, Erik Deumens
University of Florida

“Large molecule applications of coupled-cluster theory: Parallel Implementations and Natural Linear Scaling”

Coupled-cluster (CC) theory is widely recognized as the most applicable, predictive quantum chemical method for medium sized molecules. However, its application is severely restricted by molecular size. One step is to make such CC calculations run efficiently in parallel. The ACES Q.C. group has invented SIAL (super instruction assembler language) to make it possible to develop CC programs that scale very well with more than 240 processors. This approach really does remove the details of the memory handling and message passing from the quantum chemical programmer, allowing the programmer to focus on the theory. Parallelization helps, but further extension requires more method development. Based upon localized orbitals and transferability, we have invented the natural linear scaled CC method for application to large molecules and polymers.

Saday Sadayappan
Ohio State University

“Compiler/Runtime Optimizations for the Tensor Contraction Engine”

The Tensor Contraction Engine (TCE) is a domain-specific parallelizing compiler that transforms a high-level domain-specific language into parallel code with transparent handling for large sparse out-of-core arrays. The parallel code generated by the TCE executes over a parallel global-address-space runtime layer implemented over ARMCI/GA that enables locality-aware load-balancing. This talk will discuss several of the compiler and runtime optimizations implemented in the TCE system.

Curtis L. Janssen
Sandia National Laboratories

“Use of the Common Component Architecture Approach”

We review the use of the Common Component Architecture approach within the quantum chemistry domain to tackle the software engineering challenges which arise as advanced algorithms are adopted and growing numbers of software packages are integrated to study complex, coupled physical phenomena. The development of common interfaces has allowed the adoption of advanced optimization solvers and high-level interchangeability of quantum chemistry packages. Components have been created which manage multiple levels of parallelism, providing much more efficient usage of parallel machines. Early efforts towards low-level integration of chemistry packages are examined. The ability to share intermediate data expands the capabilities available to any one software package, thereby enabling the rapid development of advanced methods. New methods for the study of reactions involving heavy elements, which depend on our component environment, are highlighted.

Appendix E: Poster Abstracts

Yan Zhao

University of Minnesota

“DFT Study on the Interaction Potentials of Coronene Dimers”

State-of-the-art density functional theory has been applied to generate potential energy curves for the sandwich, T-shaped, and parallel-displaced configurations of the large prototype of aromatic pi-pi interactions, the coronene dimer. Results were obtained using a newly developed density functional, M06-2X, with different augmented double zeta and triple zeta basis sets. The high-quality estimates of the DFT potential energy curves for the coronene dimer presented here provide a better understanding of how the strength of pi-pi interactions varies with distance and orientation of the aromatic rings.

Huib Van Dam

CCLLRC Daresbury Laboratory (England)

“Development of TDDFT Gradients”

Overview of the development of TDDFT gradients for the analytical calculation of forces in molecules in the excited state. The technology enables the efficient optimisation of molecular structures of molecules in an excited state. It also opens the door to excited state dynamics as well as the calculation of first order properties. The poster outlines the current developments and presents their current status. It also invites discussions on applications.

Petia Bobadova-Parvanova

Formerly Cherry L. Emerson Center for Scientific Computation, Emory University

“Hydrogenation of Dinitrogen Coordinated in Dinuclear Metallocene Complexes”

The design and synthesis of novel catalysts, capable of hydrogenating molecular nitrogen under mild conditions, is of great fundamental and industrial interest. The almost-century old Haber-Bosch process requires extreme conditions to utilize molecular nitrogen and hydrogen to produce ammonia. Given the fact that more than 100 million tons of ammonia are produced every year, finding a new catalyst that could reduce the required temperature and pressure would be of great economic advantage. Although a new catalyst has been sought for more than 70 years, a break-through discovery has not been done yet. However, new dinuclear metallocene-N₂ complexes have been synthesized and reactions of N₂ cleavage and hydrogenation at mild conditions have been observed. These reactions have shown promising possibilities for solving the long-standing problem, but have raised several very intriguing questions.

The results of computational studies of dinitrogen hydrogenation at different dinuclear metallocene-N₂ complexes will be presented and the reasons behind the remarkably different experimental reactivity of these complexes will be explained in terms of different properties and intramolecular interactions. The mechanism of the dinitrogen hydrogenation reaction will be compared and the necessary conditions that would lead to successful dinitrogen hydrogenation will be discussed. It will be demonstrated how first-principle electronic structure calculations can answer puzzling questions that emerge from experimental results and help designing new compounds with desired properties.

Vencislav Parvanov , Donald Camaioni , Maciej Gutowski*, Nancy Hess, Wendy Shaw, Abhi Karkamkar, Tom Autrey
Pacific Northwest National Laboratory

“Modeling of Rotational Behavior of Solid State Ammonia Borane”

This work is ongoing investigation of properties of ammonia borane BH_3NH_3 . As material ammonia borane is a solid powder. For all published structures of this material at all temperatures and phases is typical to observe dihydrogen bonds forming network perpendicular to BN bonding. Rotation around BN bonds was often suggested, considering the bond strengths in solid state. Rotational barriers were measured as separate motions of NH_3 and BH_3 groups using NMR and neutron scattering techniques. Measured activation energies for these motions from different sources show much higher barrier for BH_3 of 21-25 kJ/mol than NH_3 8-13kJ/mol. We present simple cluster model for calculating these barriers. Our results show that rotation of Borane part has higher barrier than experimentally measured, when considered as independent. We suggest a correction coming from the trend of this molecule to keep the more stable internal conformation. Thus calculated barriers match exactly the experimental value.

Patrick Nichols , Wibe de Jong , Eric Bylaska
Pacific Northwest National Laboratory

“Two component DFT methods to study the Stability of Actinide Oxidation States”

The behavior of actinide oxides in solution is largely determined by the stability of the various oxidation states for that particular element. A comprehensive understanding of this aspect of actinide materials will be crucial to future waste processing, separation and storage efforts. This work will present methodologies to facilitate the study of actinide complexes in solution and interfacial interactions through the use of ab initio molecular dynamics. The incorporation of spin-orbit and relativistic scalar pseudopotentials in NWChem is a crucial first step. The final goal will allow the use of ZORA and higher order relativistic approximations.

James Rustad
University of California, Davis

“Ab Initio Molecular Dynamics Calculation of Isotope Fractionation Between Borate-Boric Acid in Aqueous Solution”

Ab initio molecular dynamics calculations are used here to calculate vibration frequencies for $\text{B}(\text{OH})_3(\text{aq})$ and $\text{B}(\text{OH})_4^-(\text{aq})$. We show that previous calculations have either underestimated or omitted altogether a major fractionating vibrational mode. The new results indicate that the ^{11}B partitions into $\text{B}(\text{OH})_4^-$ in water, in contrast to recent experimental measurement of the fractionation factor. The discrepancy appears to result from using finite-temperature vibrational frequencies in the standard harmonic expression for the fractionation factor. While our results connect the measured spectrum to previous harmonic electronic structure calculations, they indicate that harmonic frequencies must be extracted from experimental vibrational spectra before they can be used in the standard expressions.

Jeff R. Hammond^a, Karol Kowalski^b, Wibe de Jong^b

^aUniversity of Chicago, ^bPacific Northwest National Laboratory

“Coupled-Cluster Linear Response Properties for Very Large Systems Using New Functionality Within NWChem”

Coupled-cluster linear response theory, recently added to NWChem using the TCE, is used to study dipole polarizabilities of extended aromatic systems as large as pentacene and with more than 700 basis functions. We evaluate the accuracy of various density-functionals for these systems using CCSD (coupled-cluster singles and doubles). Although the failure of DFT to describe electric properties of extended systems is known, the coupled-cluster results allow us to quantify this breakdown better than with experiment. A systematic comparison of the Sadlej, Dunning and Pople basis sets is also performed.

Edoardo Aprà

Oak Ridge National Laboratory

“NWChem Performance Analysis”

We analyze benchmark results of the NWChem computational chemistry code on parallel computers. Benchmark data was collected applying Quantum Mechanical methods (e.g., Density-Functional Theory, second-order Moller-Plesset theory MP2) on molecules of increasing size (corresponding to increasing computational complexity in terms of aggregate use of memory and disk resources). Analysis of benchmark data will focus both on serial and parallel performance.

Jochen Autschbach

State University of New York at Buffalo

“Linear TDDFT response computations with NWChem: Optical rotatory dispersion”

We have recently implemented a TDDFT linear response module in the NWChem code. The program module allows for calculations of frequency-dependent electric polarizabilities as well as optical rotation. In order to avoid unphysical singularities near electronic excitations an empirical damping constant can be specified in the input. The "modified velocity gauge" by Pedersen et al. has been implemented for calculating gauge-origin independent optical rotations. Benchmark data will be presented which showcases the scaling of the program in parallel computations. Calculated optical rotatory dispersion (ORD) curves for a number of organic molecules and transition metal complexes will be compared to experiment. In general, the agreement with experiment for the anomalous ORD near resonances is good.

Florian Janetzko^a, Andreas M. Köster^b, Dennis R. Salahub^a

^aUniversity of Calgary, ^bDepartamento de Quimica, CINVESTAV

“Development of the Cyclic Cluster Model for KS-DFT Methods and its Application to Covalent Periodic Systems”

The Cyclic Cluster Model (CCM) offers an alternative approach for the simulation of extended systems like polymers, surfaces or crystals. It combines the advantages of the Free Cluster Model (FCM) and the Supercell Model (SCM), and the results of CCM calculations can be directly compared with FCM results, since the same methodology (e.g. basis sets, functionals etc.) can be used. In this presentation we compare the three different models and give a detailed description of the CCM. The cyclic cluster formalism for first-principle KS-DFT methods is presented and the implementation of the CCM in the Kohn-Sham density functional theory program (KS-DFT) deMon2k1 is discussed. The deMon2k CCM was applied to

covalent carbon-based systems periodic in one, two and three dimensions. Results of deMon2k CCM simulations of transpolyacetylene, graphene and diamond are presented. The optimized structures and electronic properties are compared with results of FCM calculations and available experimental data from the literature.

Bojana Ginovska^a, Donald Camaioni^b, Michel Dupuis^b

^aWashington State University Tri-Cities, ^bPacific Northwest National Laboratory

“Reaction Paths and Excited States in $\text{H}_2\text{O}_2+\text{OH}\rightarrow\text{HO}_2+\text{H}_2\text{O}$ ”

The mechanism of the hydrogen abstraction reaction $\text{H}_2\text{O}_2+\text{OH}\rightarrow\text{HO}_2+\text{H}_2\text{O}$ in gas phase was studied, using DFT (MPW1K) level of theory. We located 2 pathways for the reaction, both going through the same intermediate complex $\text{OH-H}_2\text{O}_2$, but via two distinct transition state structures that differ by the orientation of the hydroxyl hydrogen relative to the incipient hydroperoxy hydrogen. In one case, these hydrogens are on same side of the plane made by the 3 oxygen atoms (Transition state A) and in the other these hydrogens are on opposite sides of the plane (Transition state B). The precursor complex is 6.3 kcal mol⁻¹ more stable than the reactants. Transition states A and B are respectively 0.9 and 1.5 kcal mol⁻¹, above the reactants. The complex formed on the product side of the reaction is 41.1 kcal mol⁻¹ below the energy of the reactants. The first two excited states were calculated for selected points of both pathways using time-dependent DFT, multiconfigurational quasi-degenerate-perturbation theory (MCQDPT2/ CASSCF) and equation of motion coupled cluster singles, doubles model (EOM-CCSD) EOMCCSD energies and completely renormalized EOM-CCSD(T)(IA) correction. An avoided crossing between the two excited states was found on both reaction pathways, on the product side of the barrier to H-transfer on the ground state surface, near the transition states. On the first excited state surfaces, there is a barrier for H abstraction of 19.6 kcal mol⁻¹ for transition states A and 22.2 kcal mol⁻¹ for transition states B, with respect to the precursor complex. The hydrogen transfer on the second excited state surface is barrierless. Our TDDFT calculations show that in the precursor complex, the first excited state is 0.31 eV, and the second excited state 3.33 eV above the ground state. The respective excited states for the successor complex are 1.18 eV and 6.5 eV above the ground state. The vertical excitation energies at the transition state A for the first and the second excited state are 0.74 eV and 2.43 eV respectively. For transition state B, these energies are 0.55 eV and 2.18 eV. The values for the energies given above are not zero-point corrected.

Jose G. Lopez^a, Grigoriy Vayner^a, U.Lourderaj^a, Srirangam V. Addepalli^a, William L. Hase^a, Shuji Kato^b, Wibe A. de Jong^c, Theresa L. Windus^c

^aTexas Tech University, ^bUniversity of Colorado, ^cPacific Northwest National Laboratory

“Direct Dynamics Trajectory Study of $\text{F}^- + \text{CH}_3\text{OOH}$ Reactive Collisions”

In this poster we present results of a theoretical investigation of the reaction dynamics of methyl hydroperoxide with F^- . In order to study this reaction, we performed direct dynamics trajectory simulations at the B3LYP/6-311+G(d,p) level of theory using the general chemical dynamics program VENUS1 interfaced with the electronic structure theory software package NWChem. 2 The trajectories were initiated in the reactant and transition state regions of the $\text{F}^- + \text{CH}_3\text{OOH}$ potential energy surface and were propagated for a maximum of 4 ps. The simulation results are in excellent agreement with a previous experimental study. 3 We observe two product channels, $\text{HF} + \text{CH}_2\text{O} + \text{HO}^-$ and $\text{HF} + \text{CH}_3\text{OO}^-$. The reaction path followed by the trajectories that form the main reaction products, $\text{HF} + \text{CH}_2\text{O} + \text{HO}^-$, occurs via an ECO2 mechanism and differs from that defined by the intrinsic reaction coordinate. This suggests that the trajectories follow a path controlled by the dynamics instead of the path of steepest descent.

Matthew C. F. Wander¹ , Sebastien Kerisit², Kevin M. Rosso², Martin A. A. Schoonen¹

¹Department of Geosciences and Center for Environmental and Molecular Science (CEMS), Stony Brook University, ²Pacific Northwest National Laboratory

“Kinetics of Triscarbonato Uranyl Reduction to Insoluble U⁺⁴ by Aqueous Ferrous Iron: Theoretical Study”

Uranium is a pollutant whose mobility is tied to its oxidation state. While U(VI) in the form of the uranyl cation is readily reduced by a range of natural reductants, but carbonate greatly reduces its reduction potential. Very little is known about the elementary processes involved in uranium reduction from U(VI) to U(V) to U(IV) in general. In this study, we examine the theoretical kinetics of ET from ferrous iron to triscarbonato uranyl in aqueous solution. A combination of molecular dynamics (MD) simulations and density functional theory (DFT) electronic structure calculations using NWchem are employed to compute the ET parameters that enter into Marcus' model, including the thermodynamic driving force, reorganization energies, and electronic coupling matrix elements. MD simulations predict that two ferrous iron atoms will bind in an inner-sphere fashion to the three-membered carbonate ring of tricarbonato uranyl, forming the charge-neutral Fe₂UO₂(CO₃)₃(H₂O)₈ complex. The first ET step converting U(VI) to U(V) is predicted by DFT to occur at a rate on the order of 1 s⁻¹. The second ET step converting U(V) to U(IV) is predicted to be significantly endergonic involving the catalytic transfer of two protons prior to the ET and would occur at a much slower rate of 1x10⁻¹⁸M/s⁻¹. Therefore, U(V) is a kinetically stabilized end-product in this ET system.

N. Aaron Deskins , Michel Dupuis

Pacific Northwest National Laboratory

“Electron/Hole Transport in TiO₂ from First Principles”

This work focuses on the intrinsic electron/hole transport in stoichiometric TiO₂. Charge hopping is described by a polaron model, whereby a negative/positive polaron is localized at a Ti³⁺/O⁻ site and hops to an adjacent Ti⁴⁺/O²⁻ site. Polaron hopping is described via Marcus theory formulated for polaronic systems and quasi-equivalent to Emin/Holstein/Austin/Mott theory. We obtain the relevant parameters in the theory (namely the activation energy δG^* , the reorganization energy λ , and the electronic coupling matrix elements V_{ab}) for selected crystallographic directions in rutile and anatase, using periodic DFT+U and Hartree-Fock cluster calculations. The DFT+U method was required to correct the well-known electron self-interaction error in DFT in the calculation of polaronic wavefunctions. Our results give non-adiabatic activation energies of similar magnitude in rutile and anatase, all near ~ 0.3 eV for electron hopping and ~ 0.5 eV for hole hopping. The electronic coupling matrix element, V_{ab} , was determined to be largest for electron polaron hopping parallel to the c direction in rutile and indicative of adiabatic transfer (thermal hopping mechanism) with a value of 0.20 eV, while the other directions investigated in both rutile and anatase gave V_{ab} values about one order of magnitude smaller and indicative of diabatic transfer (tunneling mechanism) in anatase. Adiabatic transfer (large V_{ab}) was predominantly seen for bulk hole transport. Results also show a larger activation energy for hole transport on the rutile (110) surface compared to bulk.

Arun Venkatnathan, Ram Devanathan, Michel Dupuis

Chemical and Materials Science Division, Fundamental Science Directorate, Pacific Northwest National Laboratory

“Characterization of nanostructure of solvated polymer electrolyte membrane using atomistic simulations”

Polymer Electrolyte Membrane fuel cells (PEMFC) play a key role in a hydrogen economy due to their high efficiency and minimal pollution and are promising candidates for various domestic and industrial

applications. However, limitations in their performance under desirable fuel cell operating conditions of temperature points to the need of developing new membrane materials for use in PEMFC. The design and development of new membrane materials requires an understanding of the nanostructure of these membranes, their chemical structure and reactivity, and how these influence proton conductivity under low hydration conditions. As a first step we carried out classical molecular dynamics simulations to examine the structure of Nafion™ (DuPont) membranes under varying conditions of membrane hydration and temperature. The nanostructure of hydrated Nafion membrane, vehicular transport of the solvated protons and water molecules along with structural and dynamical properties of the hydrated Nafion under different conditions will be presented.

This work is supported by the Division of Chemical Sciences, Office of Basic Energy Sciences, U.S. Department of Energy (DOE). Battelle operates the Pacific Northwest National Laboratory for the DOE.

Appendix F: Additional Input and Comments

The additional input and comments provided subsequent to the meeting have been collected on the NWChem meeting CDROM. Contributions have been received from:

1. Prof. Scott Baden, University of California San Diego
2. Prof. Theresa Windus, Iowa State University
3. Dr. Huub van Dam, CCLRC Daresbury Laboratory (UK)
4. Prof. Lubos Mitas, North Carolina State University
5. Dr. Erich Vorpagel, PNNL
6. Dr. Matthew Wander, Stonybrook University
7. Dr. Aaron Deskins, PNNL
8. Dr. Daryl Clerc, ArTek Product Development, Inc.
9. Prof. Sidney Yip, MIT
10. Prof. Nathan Baker, Washington University in St. Louis
11. Prof. Bill Hase, Texas Tech University
12. Prof. Michael Green, City College of New York