NWChem Status and Future Directions

Bert de Jong
TGL, High Performance Software Development Group
Pacific Northwest National Laboratory
bert.dejong@pnl.gov

Background

MS³
MOLECULAR SCIENCE SOFTWARE SUITE

- NWChem is part of the Molecular Science Software Suite

- Developed as part of the construction of EMSL
- Designed and developed to be a highly efficient and portable **Massively Parallel** computational chemistry package
- Provides computational chemistry solutions that are scalable with respect to chemical system size as well as MPP hardware size

MS³

MOLECULAR SCIENCE SOFTWARE SUITE

NWCHEM

HIGH-PERFORMANCE COMPUTATIONAL CHEMISTRY SOFTWARE

Computational chemistry software suite designed to perform Quantum and Classical Mechanics calculations on parallel supercomputers

GA TOOLS

PARALLEL COMPUTING LIBRARIES AND SOFTWARE TOOLS

Efficient and portable shared-memory programming interface for distributed-memory computers

ECCE

EXTENSIBLE COMPUTATIONAL CHEMISTRY ENVIRONMENT

Graphical user interface, scientific visualization tools, and underlying data management framework







People

Technical Lead and **NWChem Task Lead:** Bert de Jong

Global Array Task Lead: Jarek Niepłocha

Ecce Task Lead: Gary Black

NWChem developers:
E. Bylaska, K. Kowalski, T. Straatsma, M. Valiev, D. Wang

GA Tools Developers:
M. Krishan, V. Tipparaju, B. Palmer

Ecce Developers:
T. Elsethagen, B. Palmer, M. Peterson, K.Schuchardt, L. Sun

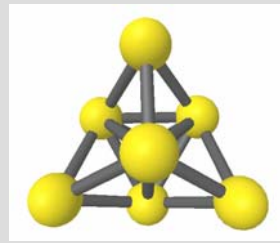




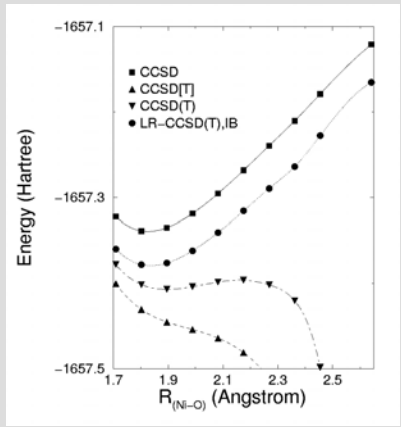



Karol Kowalski

- Areas of interest:
  - Description of correlation effects in molecules and nuclei.
  - Development of the novel coupled-cluster (CC) methods based on the Method of Moments of Coupled-Cluster Equations.
  - Excited-state methods: globally/locally renormalized single- and multi-reference coupled-cluster approaches.
  - Pushing the limits: acquiring the spectroscopic accuracies in large scale CC calculations.



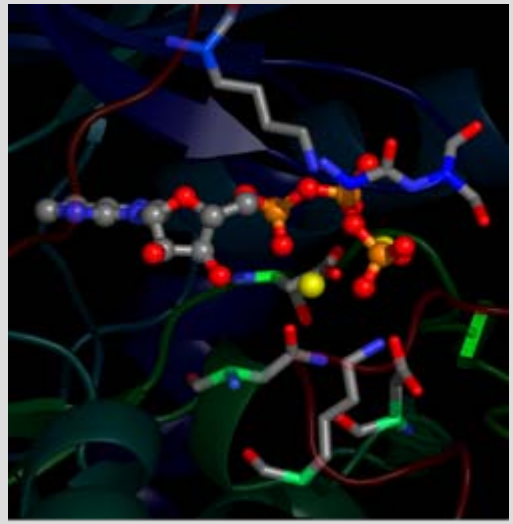
$$E = \sum_{n=0}^{N} \sum_{J_n} \frac{\left\langle \Phi \middle| (e^{S^+} M_{J_n}(m_A) e^{T_{D_n}^{(A)}})_C \middle| \Phi \right\rangle}{\Gamma_I(J_n) + \left\langle \Phi \middle| (e^{S^+} \Gamma_{II}(J_n) e^{T_{D_n}^{(A)}})_C \middle| \Phi \right\rangle}$$



- Education and experience:
  - Ph.D. Physics, Nicolaus Copernicus University
  - Joined PNNL in 2004

Marat Valiev

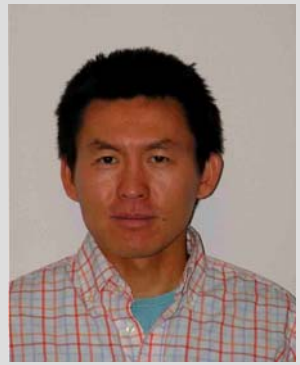
- Areas of Interest
  - Field Theoretical Approaches to Density Functional Theory
  - Mixed local and plane wave basis ab-initio methods
  - Combined quantum mechanics/molecular mechanics methods (QM/MM)
  - Ab-Initio Molecular Dynamics
  - Large scale simulations of complex enzymatic reactions
  - Structure and dynamics of ion solvation



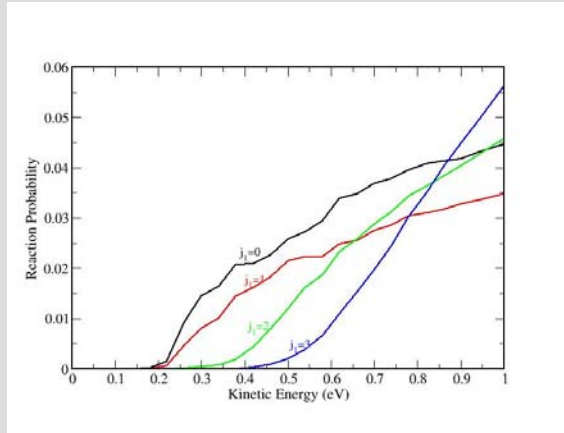
Active Site in cAPK kinase

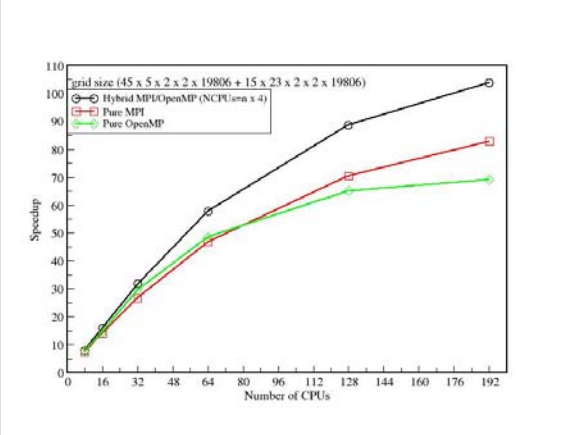
- Education and Experience
  - Ph.D. Physics, University of Connecticut
  - Joined PNNL in 2003

Dunyou Wang



- Areas of interest:
  - Quantum Dynamics Calculations on Chemical Reactions in Gas Phase
  - Chemical Reactions in Chemistry-Climate Interactions
  - Theoretical Studies on Radiation-induced Damages to DNA bases
  - High-end parallel computing using MPI, OpenMP paradigms and GA tools

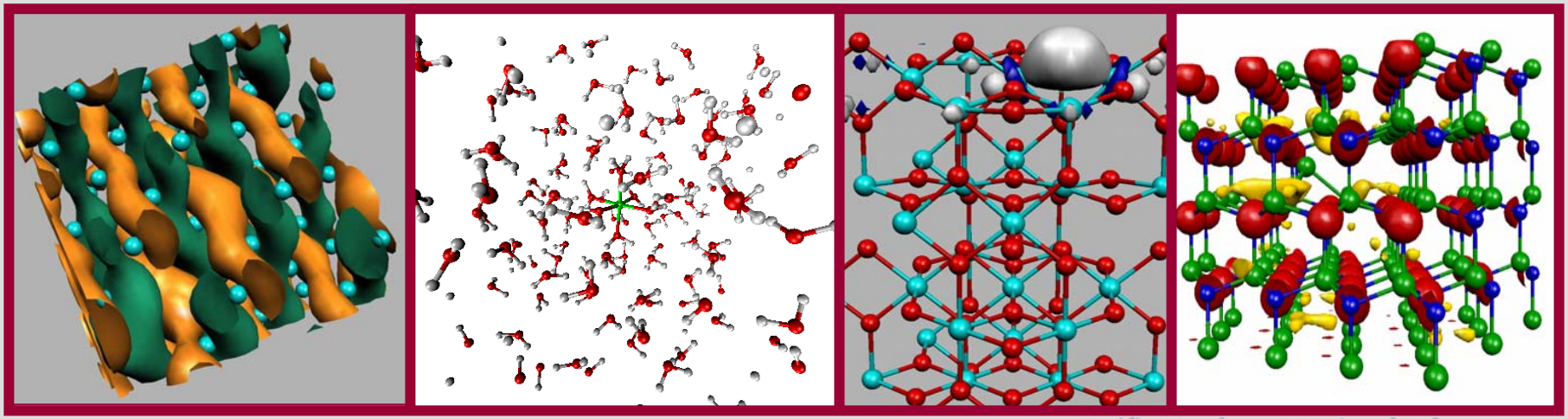




- Education and experience:
  - Ph.D. Computational Physical Chemistry, New York University
  - Joined PNNL in 2006

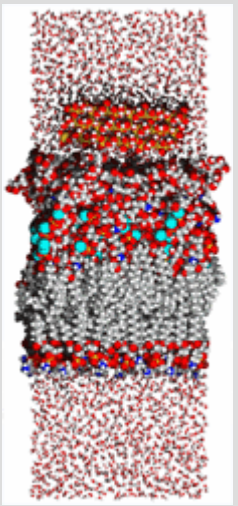
Eric Bylaska

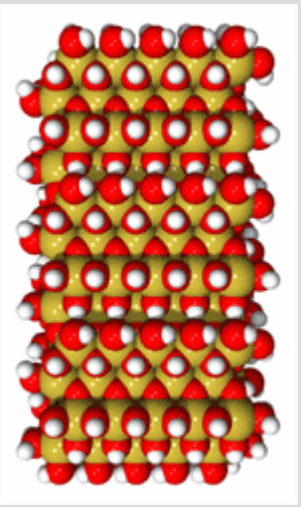
- Areas of Interest
  - High performance computational chemistry
  - Pseudopotential Plane-Wave (PSPW) and Projector Augmented Wave (PAW) first principles methods for molecules, clusters, liquids, surfaces, and solids
  - Studies of metal solvation and transport
  - Studies of degradation pathways of organochlorine compounds
- Education and Experience
  - Ph.D., Physical Chemistry, University of California, San Diego
  - Joined PNNL in 1998



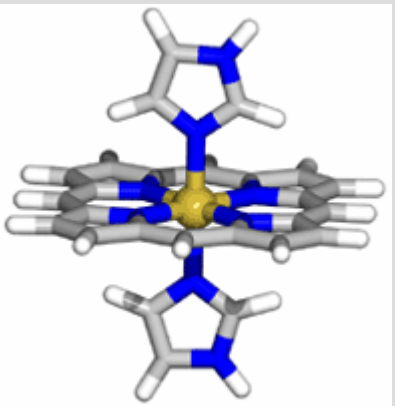
T. P. Straatsma

- Areas of interest:
  - Molecular Dynamics methods applied to chemical and biomolecular systems
  - Microbial membrane mediated geochemical processes at mineral surfaces .
  - Transport processes across microbial membranes.
  - Development and application of novel methods to study complex enzymatic reactions



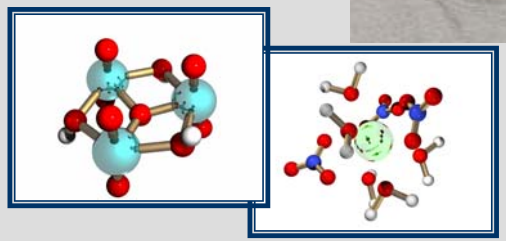


- Education and experience:
  - Ph.D., Mathematics and Natural Sciences, University of Groningen
  - Joined PNNL in 1994
  - Manager of the Computational Biology Bioinformatics group



Bert de Jong

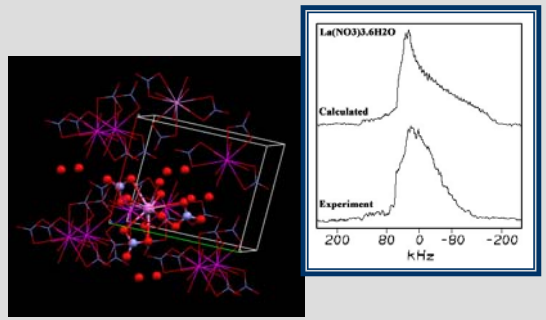
- Areas of interest:
  - Development of relativistic computational chemistry software
  - Development of (response) properties
  - Interactions between theory and experiment; how can theory contribute to experimental interpretations?
    - Science driven development and application



$$\hat{H}^{tf}_{DK} = \sum_i^n E_i + \sum_i^n V^{tf}_{eff}(i) + \sum_{i<j}^n \frac{1}{r_{ij}} + \sum_{I<J}^{nuc} \frac{Z_I Z_J}{R_{IJ}}$$

$$V^{tf}_{eff}(i) = -\sum_i A_i V(i) A_i + \sum_i A_i R_i V(i) R_i A_i - \sum_i W_i(i) E_i W_i(i) - \frac{1}{2}\sum_i \left\{ (W_i(i))^2 E_i + E_i (W_i(i))^2 \right\}$$

- Education and experience:
  - Ph.D. Chemistry, University of Groningen
  - Joined PNNL in 2000



Additional staff

- Postdoc: Partick Nichols
  - Development of spin-orbit ZORA within the plane-wave framework
  - Funded through BES-HEC (PI: de Jong)

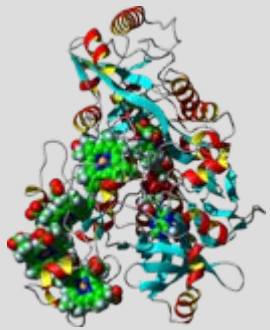
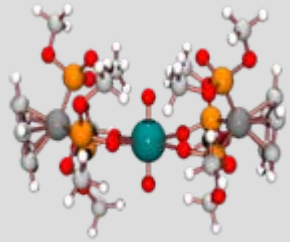
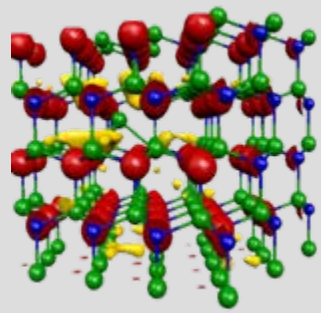
- New hire process underway for a Density Functional Theory developer
  - Interfacial chemistry
  - Linear scaling for large biomolecules
  - Innovations in time-dependent DFT
  - Hopefully we'll be hiring soon !

NWChem Overview

- More than 2,000,000 lines of code (mostly FORTRAN)
  - About half of it computer generated (Tensor Contraction Engine)
- A new version is released on a yearly basis:
  - Addition of new modules and bug fixes
  - Ported the software to new hardware
  - Increased performances (serial & parallel)
- Freely available after signing a user agreement
- World-wide distribution (downloaded by 1600+ groups)
  - 70% is academia, rest government labs and industry
- 70-75% cycles on the EMSL Molecular Science Computing Facility hardware use NWChem

NWChem Capabilities

- Provides major modeling and simulation capability for molecular science
  - Broad range of **molecules**, including **bio**molecules, **nano**particles and heavy elements
  - Electronic structure of molecules (non-relativistic, relativistic, ECPs, first and second derivatives)
  - Increasingly extensive **solid state** capability (DFT plane-wave, CPMD)
  - Molecular dynamics, molecular mechanics

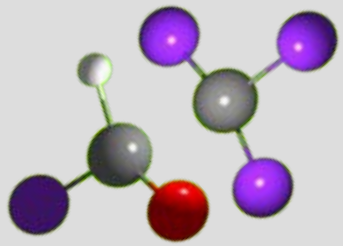





NWChem Platform Capabilities

- Emphasis on modularity and portability
- Performance characteristics – designed for MPP
  - Single node performance comparable to best serial codes
  - Scalability to 1000's of processors
- Portable – runs on a wide range of computers
  - MSCF supercomputer to PC with Windows
  - Various operating systems, interconnects, CPUs

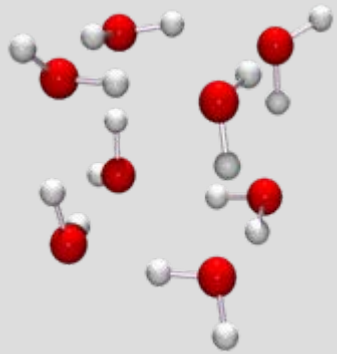
NWChem: A state-of-the-art capability High Performance and Scalability of Correlated Methods

- Work by: Karol Kowalski
- TCE Open-shell CCSD(T) handles problem sizes larger than any other code
- Provides unique science capability



CF_3CHFO:
1212 spin-orbitals, 57 electrons

- Work by: Bert de Jong
- Closed-shell CCSD(T) on 1840 processors achieves 63% of 11 TFlops peak efficiency

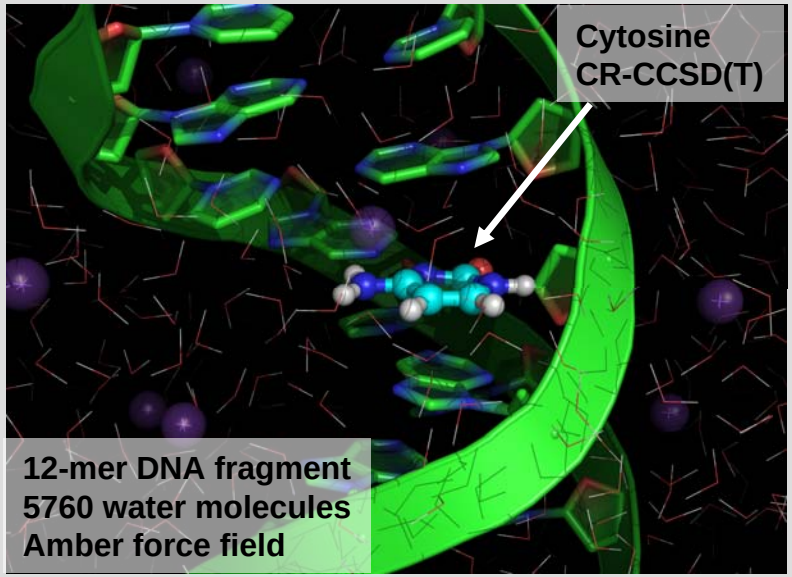


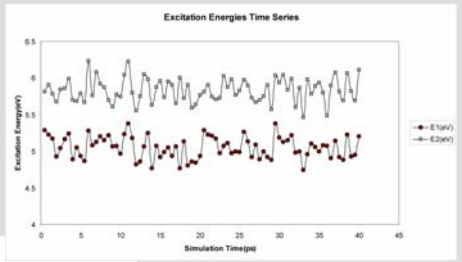
8-water cluster:

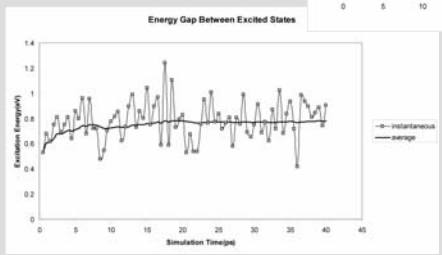
1376 basis functions, 32 electrons

NWChem: High-Impact Science Calculations of Excited States In DNA

- Work by: Valiev and Kowalski
- Unique science using hybrid of QM/MM and CC Methods
- Simulations to calculate excited states for cytosine in the native DNA environment.
- Results show that protein environment affects the excitation, leading to a 0.4 eV stabilization of the $\pi\pi^*$ excited state relative to gas phase calculations
- *JCP* and *J Phys Chem A*

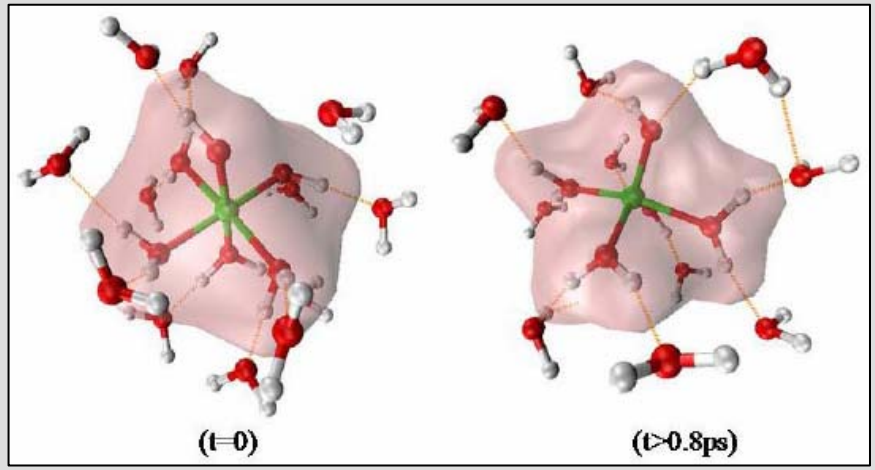






NWChem: Plane Wave Calculations Confirm Five-fold Coordination for Al^{III} in Water

- Work by: Eric Bylaska
- Experimental results after two years of work suggested a five-fold coordination for Al^{III} in geochemically critical conditions
- Sophisticated plane-wave DFT dynamics simulations confirmed proposed mechanism

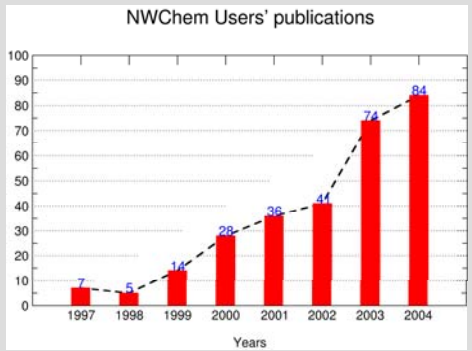


Swaddle, Rosenqvist, Yu, Bylaska, Phillips, Casey. ***Science*, 2005**

NWChem Used in High-Impact Science Publications

	Using MSCF	Using offsite	Total pubs
Top 5 Journals	20	22	42
Top 10 Journals	29	32	61
All FY06 Papers	39	59	98

- About 400 publications using NWChem



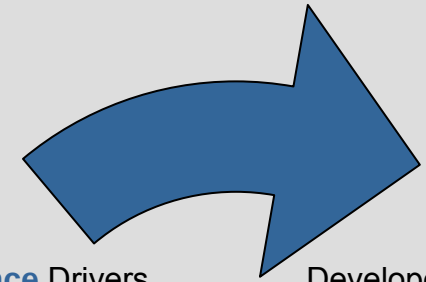
NWChem 5.0 released September 2006

- Release in conjunction with Ecce 4.0 and GA Tools 4.0
- New capabilities
  - New QM/MM module integrated with all QM methods
  - Vibrational SCF available for all QM methods to incorporate anharmonic effects
  - New coupled cluster methods (EOM, CR, LR, Active-Space) allow high accuracy for difficult and multi-reference cases
  - Combine electron transfer with proton hopping in classical MD – critical to Biogeochemistry Grand Challenge
  - Hybrid functionals and Self-Interaction Correction in plane-wave DFT expand its applicability
  - Kinetic Monte Carlo for Dynamic Nucleation Theory
  - Interface with VENUS for direct dynamics

NWChem 5.0 (continued)

- Major improvements in capabilities
  - Reduced memory usage and speedup in TCE and CCSD(T) allows access to very large problem sizes
  - Speedup in plane-wave methodologies
- Combined with GA Tools 4.0 available on new platforms
  - Including Cray XT3 (ORNL), IBM BlueGene
- Combined with Ecce 4.0, users can now set up molecular dynamics calculations through graphical user interface

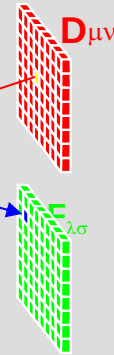
NWChem Capability Development Strategy



$$\rho(\mathbf{x}_Q) = \sum_{\mu\nu} D_{\mu\nu}\, \chi_\kappa(\mathbf{x}_Q)\, \chi_\nu(\mathbf{x}_Q)$$

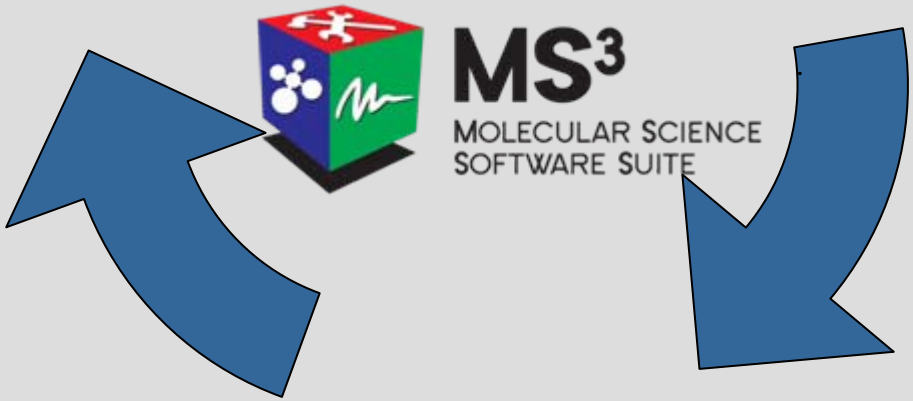
$$F_{\lambda\sigma} \mathrel{+}= \sum_Q w_Q\, \chi_\lambda(\mathbf{x}_Q)\, V^{xc}[\rho(\mathbf{x}_Q)]\, \chi_\sigma(\mathbf{x}_Q)$$

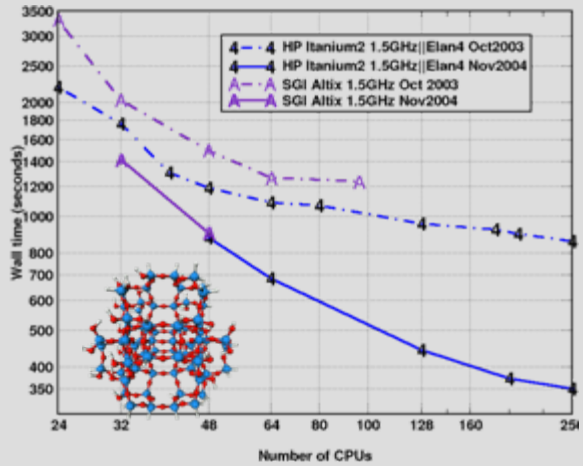
```
my_next_task =
SharedCounter()
  do i=1,max_i
    if(i.eq.my_next_task)
then
      call ga_get()
        (do work)
      call ga_acc()
      my_next_task =
SharedCounter()
    endif
  enddo
barrier()
```



Evolving **Science** Drivers.
Input gathered from:
1) Feedback from users
2) MSCF Greenbook

Developers of the 3 teams discuss the **implementation** of the evolving Science Drivers.
Develop new capabilities.
Porting to new architectures,
Performance tuning on the MSCF hardware.





New SW features/methods made available
to users through a yearly **release** process.
Testing phase preliminary to the **distribution** phase.
Move to **support/maintenance** phase.

GA Toolkit 4.0 **NWChem 5.0** **Ecce 4.0**

NWChem Capability Development Drivers

- Science at EMSL and in MSCF Greenbook
- Scientific needs of experimentalists in EMSL
- User needs
  - Workshops
  - Surveys
  - General requests
- Availability of large computational resources

NWChem Future Capability Developments

- Parallel scaling to tens of thousands of processors
- New properties, analysis, longer time scales in molecular dynamics
- Extending relativistic and interfacial capabilities for heavy element chemistry
  - ($300K BES Heavy Elements Program Funding)
- Linear and quadratic response at CCSD(T) level
  - Properties such as (hyper)polarizability, analytic gradients & hessians
- Reduced scaling and cost with respect to system size (Order-N)
- Quadratic response and frequency-dependent linear response
  - Polarizability, Circular Dichroism, Optical rotation, Raman intensities
- Coupled cluster response properties
  - Frequency-dependent polarizability
- Improved models to include interfaces and solvent effects
- Semi-empirical capability to handle large systems quickly
- Expanding dynamics capabilities, Monte Carlo/Simulated annealing

Community involvement is crucial…

- NWChem developers at EMSL have only limited resources
  - Focus on capabilities important to EMSL
  - Providing support and maintenance
- NWChem has broader developer base scattered around the US (globe)
- Get the computational chemistry community involved more actively in development
  - Developer workshops
  - Invite potential developers to EMSL for a week or two to get over initial barrier

Fostering new developers and developments in NWChem

- Jeff Hammond (U. Chicago)
  - DOE Computational Science Graduate Fellow
  - Developed the basic functionality to calculate frequency dependent response properties at the CCSD level of theory
- Prof. Jochen Autschbach (U. Buffalo)
  - Invited for two week stay to gain the skills to develop new property capabilities
  - Currently implementing Optical Rotation and other properties for density functional theory
- The new capabilities will be available in the FY07 NWChem release

NWChem meeting: Science driven petascale computing and capability development (1)

- Petascale computing hardware is around the corner !
- Large systems already exist:
  - IBM Blue Gene/L at LLNL: 131,000 procs, 367 Tflops
  - Red Storm at NNSA: 26,544 procs, 127 Tflops
  - SGI Colombia at NASA: 10,000 procs, 61 Tflops
  - Jaguar ORNL: 10,424 procs, 54 Tflops
- Next generation MSCF supercomputer probably ~10,000 processors
- Petaflops coming shortly at LANL, ORNL, ANL, …

NWChem meeting: Science driven petascale computing and capability development (2)

- Petascale computational chemistry should be science driven
- Petascale resources will enable chemistry researchers to
  - enter a new era of modeling
  - tackle scientific problems that larger and more realistic than ever before
  - to include more of the complex dynamical behavior of nature
  - to start asking new and different scientific questions
- Scientific questions will drive choice of methods and algorithms that need to scale to petaflop systems

Towards petascale computational chemistry

- Computational chemistry software not ready for petascale computing yet
  - BES/BER/ASCR need applications effectively utilizing the hardware
  - NWChem does scale to teraflops, but not to petaflops
- It's important to recognize that not all current methods and algorithms **will** scale to petaflop systems
- But, not all of them **have** to scale to petaflop systems either
  - Do we want to do a 100,000 processor DFT calculation ?
  - Or do we need integrate over a hundred 1,000 processor DFT calculations to analyze dynamical properties ?

Goals of this meeting

- Get insight into the computational chemistry science that will utilize petascale computing
  - Science that needs single petascale runs
  - Science that requires multi-scale approaches
    - Dynamics (trajectories / Monte Carlo)
    - Exploration of energy and property surfaces (gradient/hessian)
- Get insight into new methods, computer science, and mathematics for petascale computing
  - Computational chemistry models, theories, algorithms and their limitations
  - New mathematics and multiscale developments

Outline of this meeting

- Morning sessions focus on science
  - Chemical reactions and dynamics
  - Biological systems and interfaces
  - Materials and transport
- Afternoon sessions focus on new methods, computer science, and mathematics
  - Methods for large scale simulations
  - Computer science, mathematics, and software generation for large scale simulations

Outline of this meeting

- Panel discussions with participants
  - If you have items you would like to have discussed, please write them down and bring it to registration before panel session
- Working lunch
  - Open discussion on open-source NWChem
  - Open discussion on new and emerging science areas (broad, what has not been covered)
- Informal discussions
  - Stretch your legs, coffee, talk to NWChem developers,….

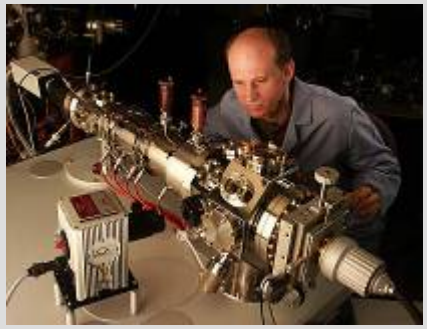
Final product of this meeting

- Report will be written
  - All your input will be compiled and become part of the report
  - Draft will be distributed for comments and feedback
- Strategic plan will be developed to ensure that NWChem will continue to provide EMSL users and the NWChem user community with the cutting-edge capabilities they need on next-generation platforms

WR Wiley Environmental Molecular Sciences Laboratory

A national scientific user facility integrating experimental and computational resources for discovery and technological innovation






William R. Wiley
EMSL
Environmental Molecular Sciences Laboratory