Scientific Innovation Through Integration

EMSL

NWChem: Pushing the scientific envelope on large computing platforms

Bert de Jong
High Performance Software
Molecular Science Computing Facility

Background

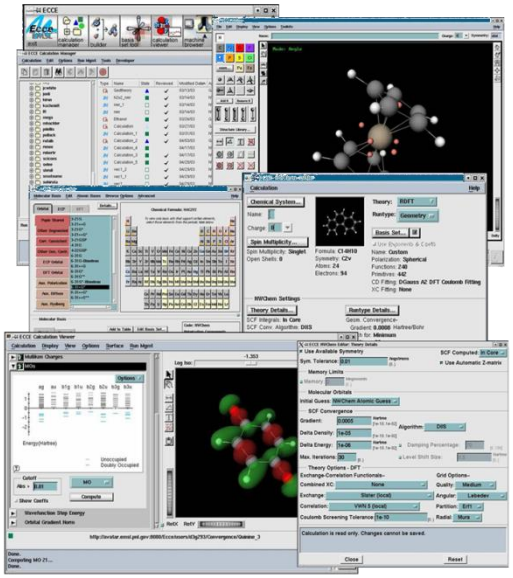
- NWChem is part of the Molecular Science Software Suite

MS³
MOLECULAR SCIENCE SOFTWARE SUITE

NWCHEM
HIGH-PERFORMANCE COMPUTATIONAL CHEMISTRY SOFTWARE

GA TOOLS
PARALLEL COMPUTING LIBRARIES AND SOFTWARE TOOLS

ECCE
EXTENSIBLE COMPUTATIONAL CHEMISTRY ENVIRONMENT



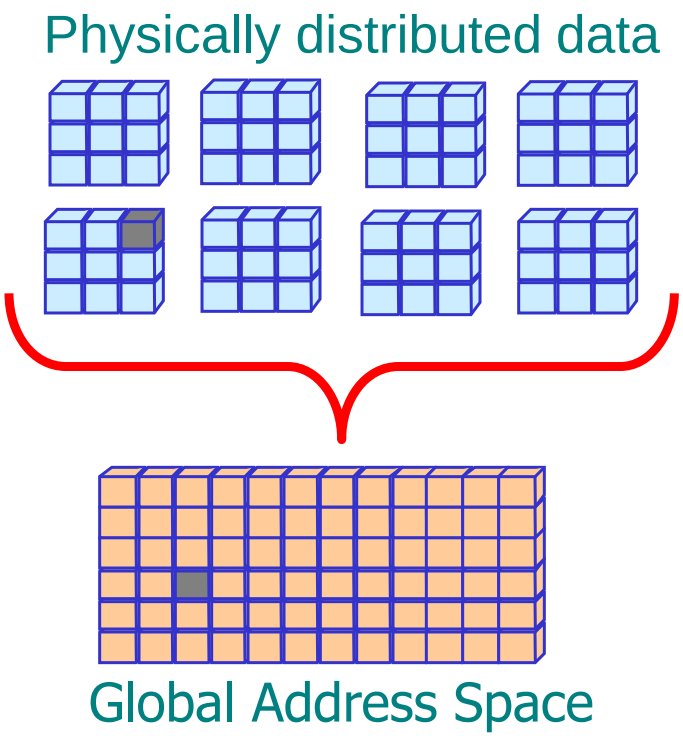
- Designed and developed to be a highly efficient and portable **Massively Parallel** computational chemistry package
- Provides computational chemistry solutions that are scalable with respect to chemical system size as well as MPP hardware size

NWChem Overview

- Originally designed for parallel architectures
  - Scalability to 1000's of processors (part even to 10,000)
- Emphasis on modularity, portability, and integration
- Portable – runs on a wide range of computers
  - Supercomputer to Mac or PC with Windows
  - Now runs efficiently on IBM BlueGene, Cray XT, InfiniBand
- **Freely** available after signing a user agreement
  - Going away in NWChem 6.0
- World-wide distribution (downloaded by 2800+ groups)
  - 70% is academia, rest government labs and industry

Parallel Scalability with the Global Array Toolkit

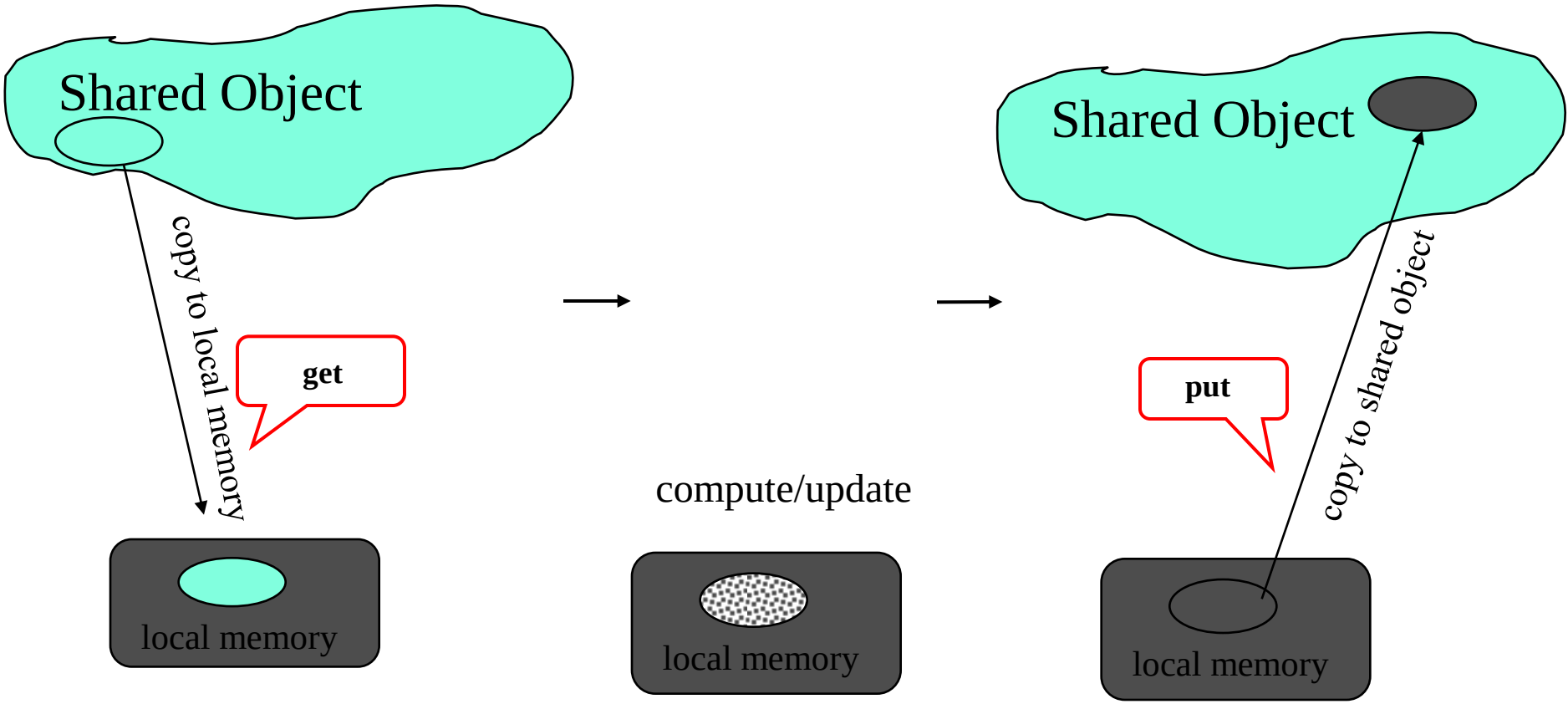
- NWChem users Global Array Toolkit to scale and perform on parallel architectures
- Programming model: Distributed dense arrays that can be accessed through a shared memory-like style



- Programmer sees one globally accessible data structure
  - I.e. accessing A(4,3) rather than buf(7) on task 2
- Hides much of the necessary parallel infrastructure
  - Programmer needs to be aware of locality of data for scalability

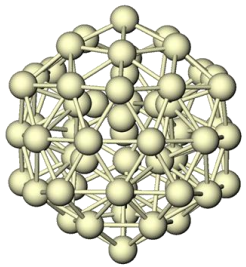
Global Array Model of Computations

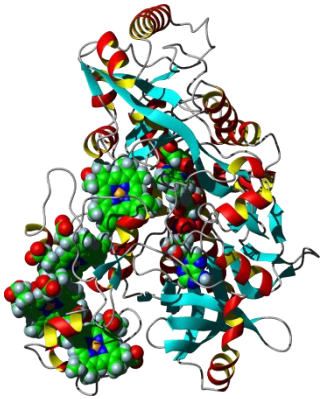
- General use model in NWChem:

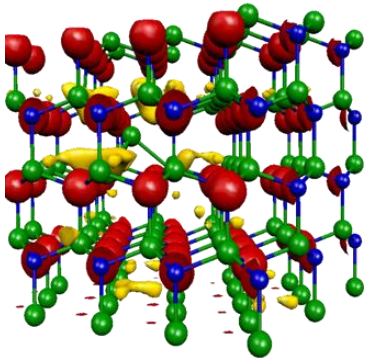


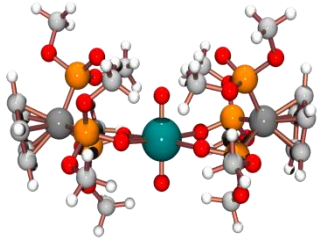
NWChem Science Capabilities

- 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)
  - Extensive **solid state** capability (DFT plane-wave, CPMD)
  - Molecular dynamics, molecular mechanics









NWChem's transition

- Started as construction project, mainly focused on development
- Now is major production code with scientists pushing for new capabilities in addition to scalability
- Going forward need to balance capability and scalability developments
  - Doing both requires lots of bodies
  - We need collaborative development environment beyond EMSL/PNNL

NWChem's core developer team

Bert de Jong
Team lead
Properties/Relativity

Karol Kowalski
High accuracy

Niri Govind
Density functional theory

Looking for a junior and senior hire

Eric Bylaska
Plane wave methods

Tjerk Straatsma
Molecular dynamics

Marat Valiev
QM/MM

Huub van Dam
DFT/HPC

Pacific Northwest
NATIONAL LABORATORY

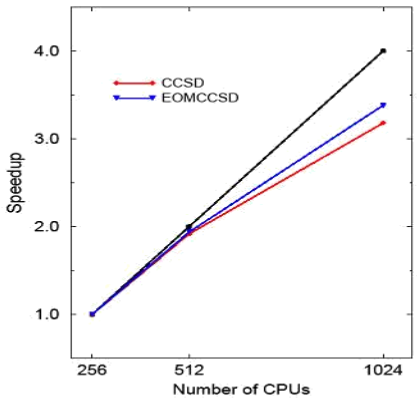
NWChem capabilities overview

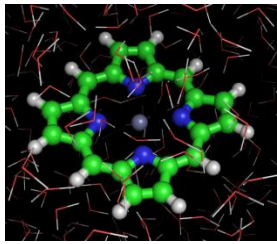
- NWChem brings a full suite of methodologies to solve large scientific problems
  - High accuracy methods (MP2/CC/CI)
  - Gaussian based density functional theory
  - Plane wave based density functional theory
  - Molecular dynamics (including QMD)
  - Integrated and working together

- Will not list those things standard in most computational chemistry codes
  - *http://www.emsl.pnl.gov/docs/nwchem/nwchem.html*

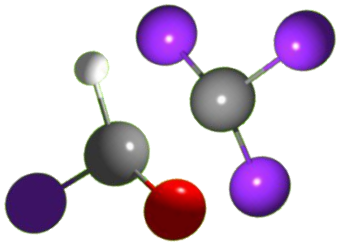
NWChem: high accuracy (1)

- Coupled cluster
  - Closed shell coupled cluster [CCSD and CCSD(T)]
  - Tensor contraction engine (TCE)
    - Spin-orbital formalism with RHF, ROHF, UHF reference
    - CCD, CCSDTQ, CCSD(T)/[T]/t/…, LCCD, LCCSD
    - CR-CCSD(T), LR-CCSD(T)
    - EOM-CCSDTQ for excited states
    - MBPT(2-4), CISDTQ, QCISD
    - Working on multireference CC



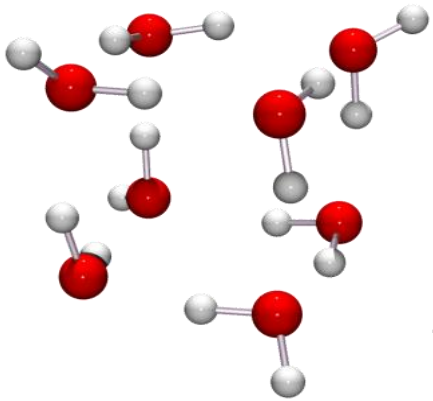


CC performance: Zn-porphyrin, 395 bf, 124 elec, no symmetry



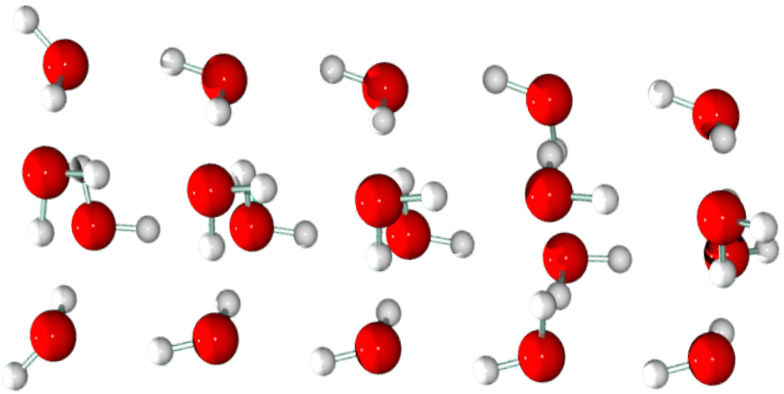
CF_3CHFO CCSD(T) with TCE: 606 bf, 57 electrons, open-shell

NWChem: Gordon Bell Finalist



2007

2009



8 water molecules on EMSL's MPP2 achieve 7 TFlops (63% of peak) on 1,840 processors

24 water molecules on ORNL's Jaguar achieve 1.31 PFlops (over 50% of peak) on 225,000 processors

ORNL – PNNL Collaboration

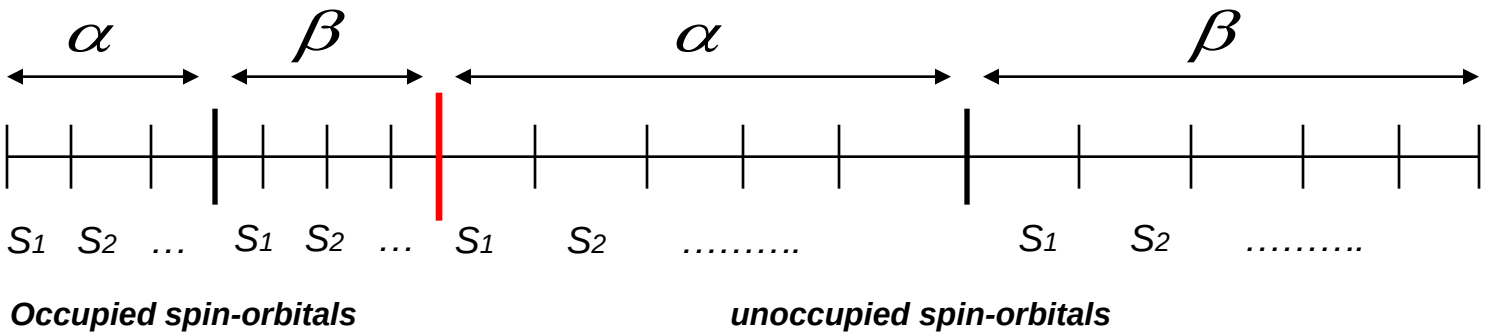
Lead Edo Apra

- **Improving our understanding of water's properties**

NWChem: Tensor Contraction Engine

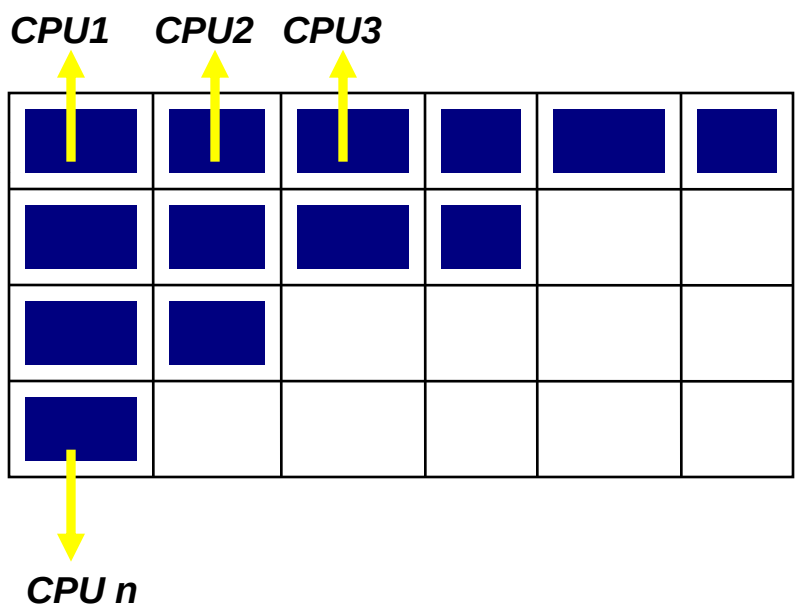
Started by Robert Harrison and So Hirata (lots of DGEMMs)

Tile structure:



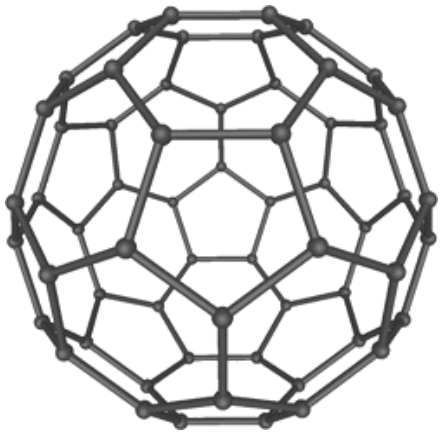
Tensor structure (based on the tile structure):

$$T_a^i \Rightarrow T_{[p_n]}^{[h_m]}$$



NWChem: TCE examples

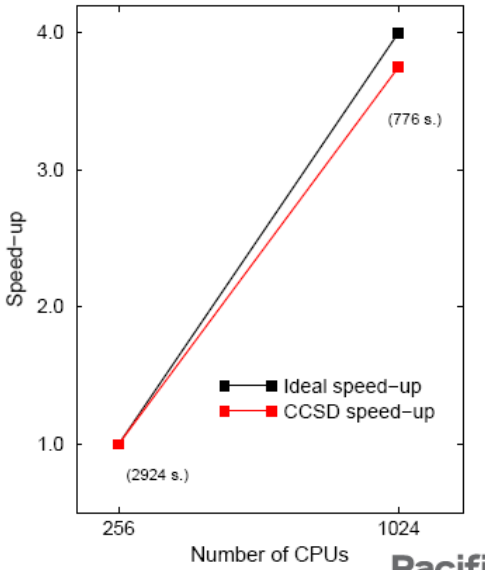
- Latest with TCE
  - Response properties
    - Frequency dependent linear response up to CCSDT
    - Quadratic response properties for CCSD
  - (EOM-) CCSD gradients
    - Being implemented by Pen-Dong Fan



1080 basis set functions, 1.5 billion T_2 amplitudes

Static and frequency dependent polarizabilities of C_{60} in Å^3

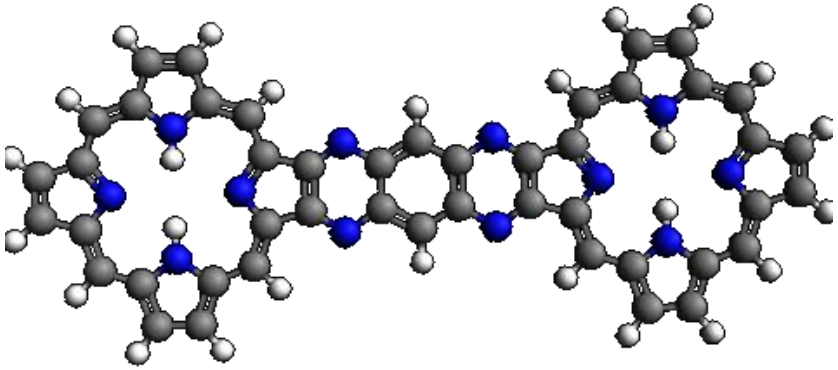
Wavelength (nm)	Full CCSD	Expt.
∞	82.20	76.5 8
1064	83.62	79 4
532	88.62	



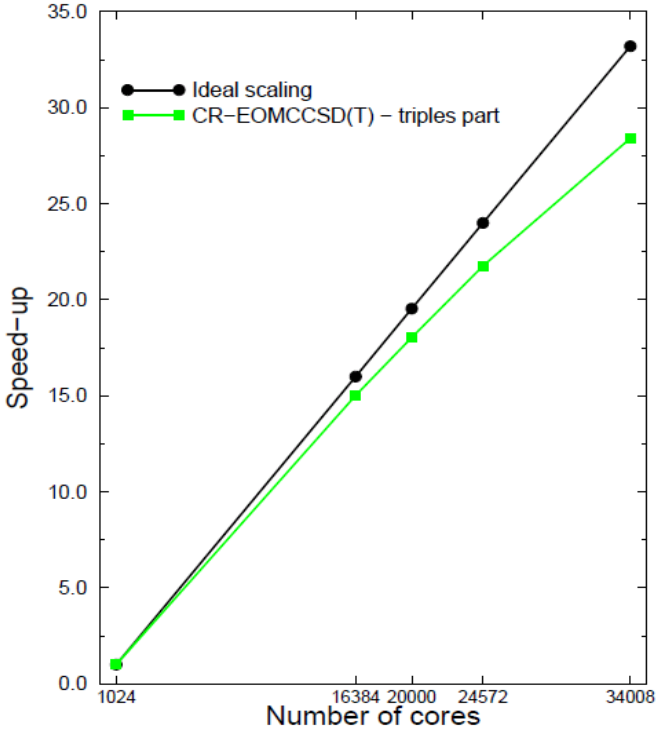
K. Kowalski, J.R. Hammond, W.A. de Jong, A. Sadlej, JCP

NWChem - New Science with High Accuracy Methods

- Designing light harvesting systems and molecular switches requires a highly accurate and comprehensive understanding of the behavior of electrons in the system.



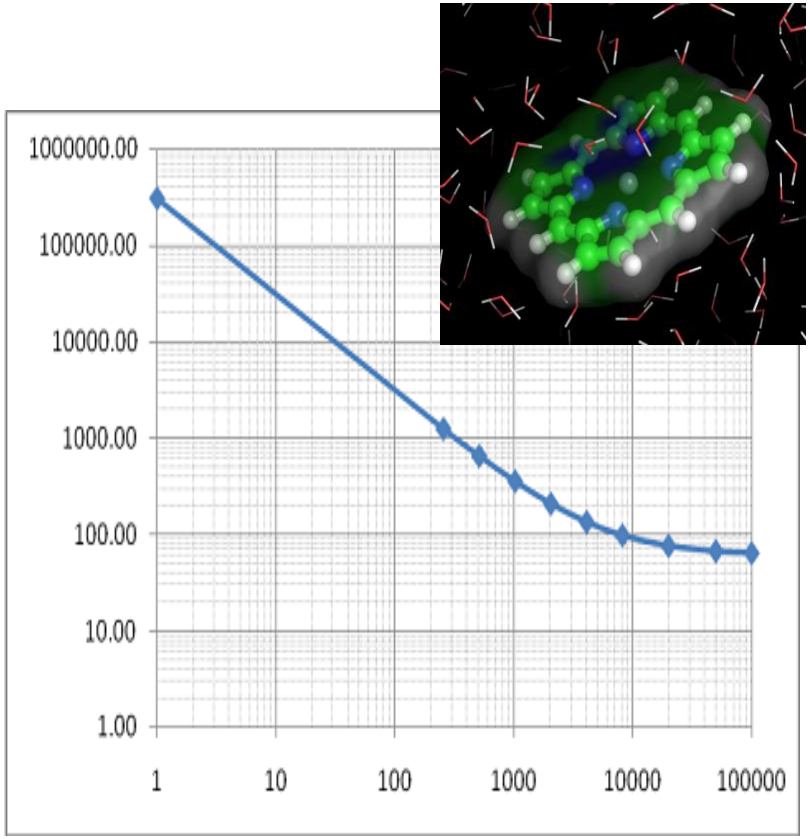
P_2TA molecule forms a basis for molecular switches and wires in nanoelectronics



Ideal scaling of high accuracy CR-EOMCCSD(T) triples for the P_2TA molecule (648 basis functions)

NWChem - High Accuracy to the Petascale

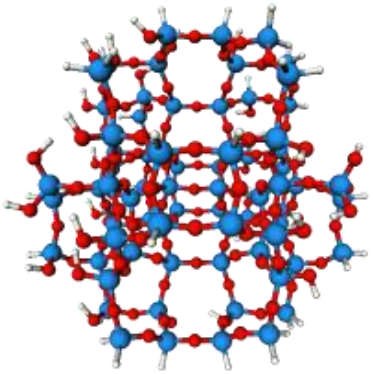
- Improving the parallel efficiency for coupled cluster algorithms
  - Balancing communication and operations
  - Better task scheduling based on knowledge of computational costs and overhead
  - Heterogeneous computing using processor groups
  - Improving data-localization reduces communication



Performance model for the light harvesting Zn-porphyrine molecule

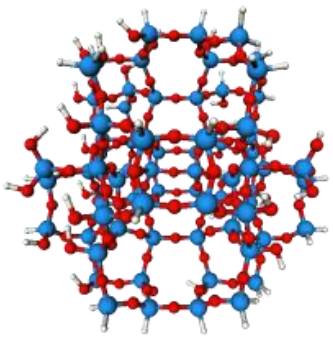
NWChem: DFT

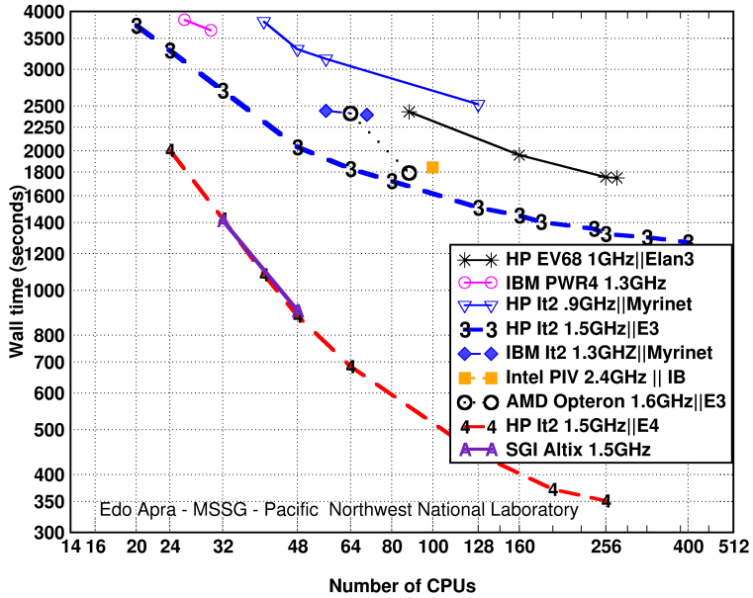
- Density functional theory
  - Wide range of local and non-local exchange-correlation functionals
    - Truhlar's M05, M06 (Yan Zhao)
    - Hartree-Fock exchange
    - Meta-GGA functionals
    - Self interaction correction (SIC) and OEP
  - Spin-orbit DFT
    - ECP, ZORA, DK
  - Constrained DFT
    - Implemented by Qin Wu and Van Voorhis
  - TDDFT for excited states (Soon: spin-orbit TDDFT)
  - Coulomb attenuated (LC) functionals (Andzelm, Baer, Govind)



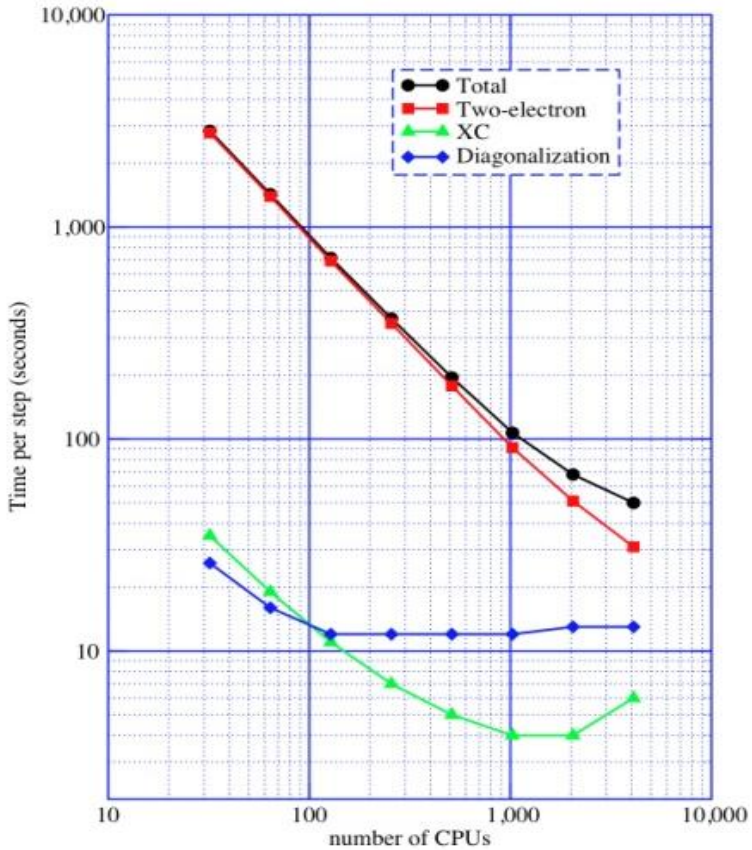
$Si_{75}O_{148}H_{66}$

NWChem: Current DFT can scale to 4K





DFT performance: $Si_{75}O_{148}H_{66}$, 3554 bf, 2300 elec, Coulomb fitting



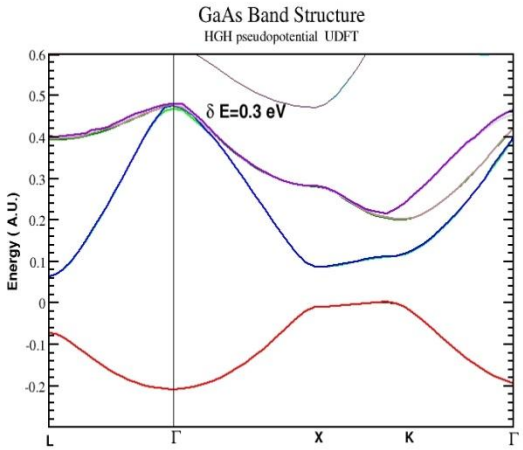
Overall and component timings from Gaussian basis set based calculations of C_{240} on up to 8K processors

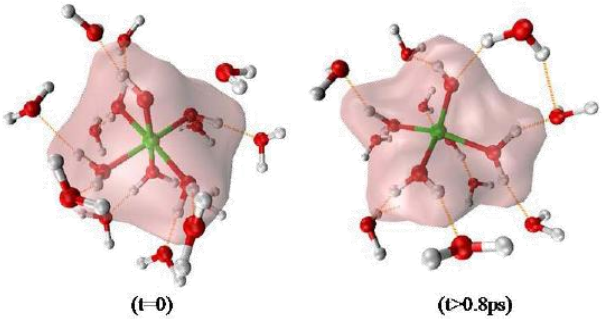
Pacific Northwest NATIONAL LABORATORY

NWChem: plane wave (1)

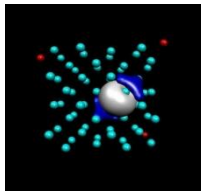
- Plane wave density functional theory
  - Gamma point pseudopotential and projector augmented wave
  - Band structure (with spin-orbit ZORA)
  - Extensive dynamics functionality with Car-Parrinello
  - AIMD QM/MM molecular dynamics, e.g. SPC/E,CLAYFF solid state MD
  - Various exchange-correlation functionals
    - LDA, PBE96, PBE0 ,(B3LYP)
    - Exact exchange
  - SIC and OEP

Spin-Orbit splitting in GaAs





Car-Parrinello provides evidence for five-coordinate $Al(H_2O)_4OH^{2+}$
Swaddle et al, ***Science***, **2005**

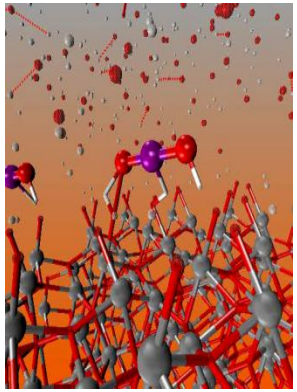


SIC localization

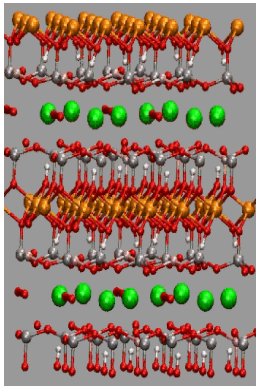
NWChem: plane wave (2)

- Can handle charged systems
- A full range of pseudopotentials and a pseudopotential generator
- A choice of state-of-the-art minimizers

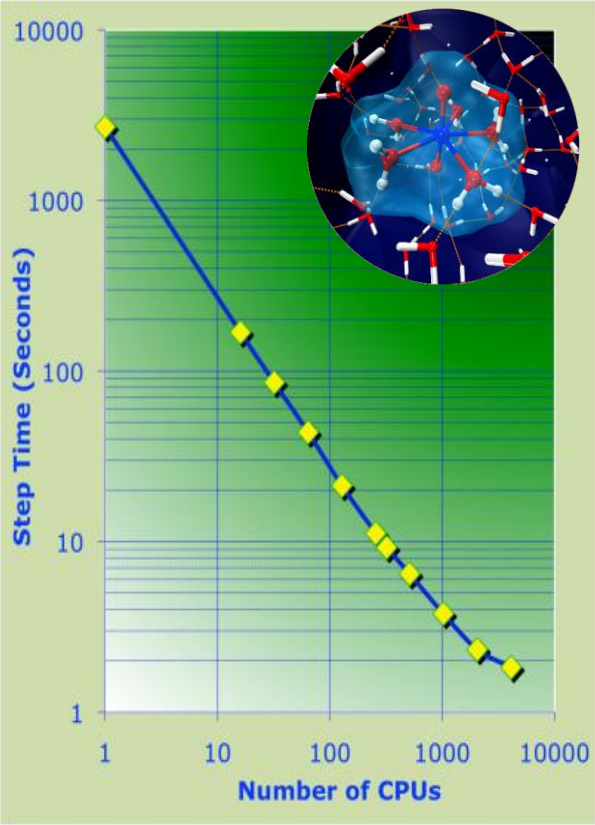
- Will be able to do QM/MM



Molecules interacting with a reactive surface



Complex materials and interfaces



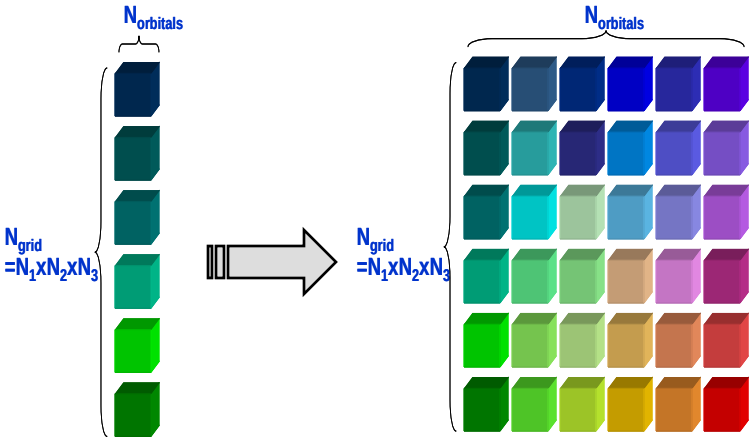
Car-Parrinello plane wave performance, PBE96 GGA Functional, -300 K thermostat, 0.121 fs time step, 122 water molecules-15.6 Å box

NWChem - Hard Petaflop Scaling of Plane Waves

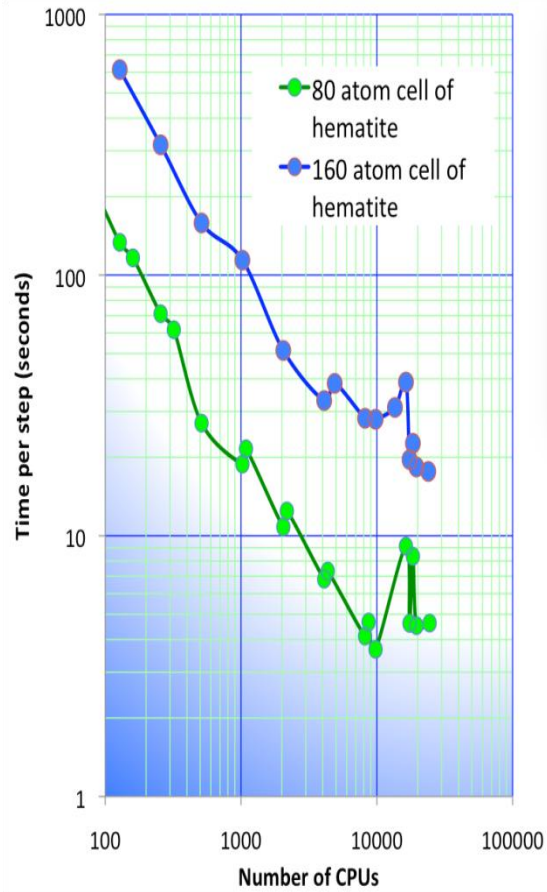
- Smart data distribution and smart communication algorithms enable hybrid-density functional theory to scale to large numbers of processors

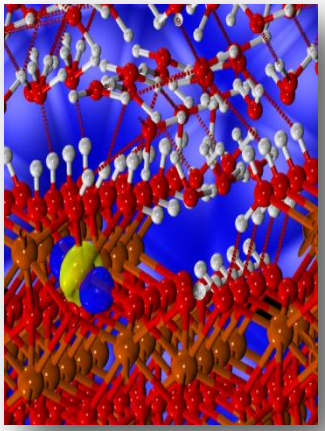
Parallel Data Distribution

Old New



Gygi et al. Supercomputing '06



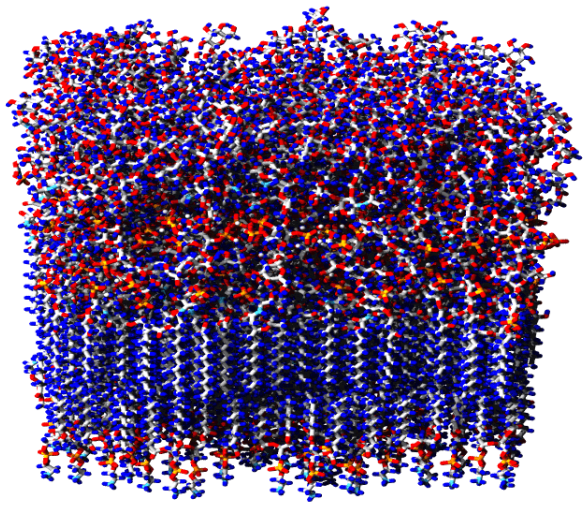


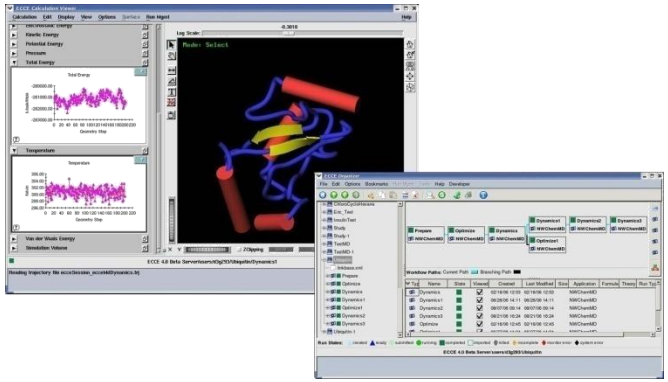
Hybrid density functional theory plane wave calculation on hematite scaling on NERSC's Franklin

Pacific Northwest NATIONAL LABORATORY

NWChem: molecular dynamics (1)

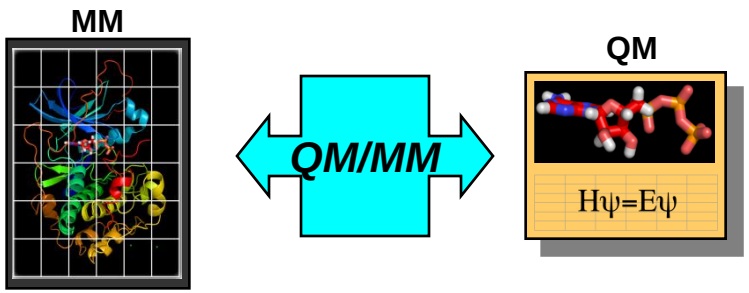
- Molecular dynamics
  - Charm and Amber force fields
  - Various types of simulations:
    - Energy minimization
    - Molecular dynamics simulation including *ab initio* dynamics
    - Free energy calculation
    - Multiconfiguration thermodynamic integration
  - Electron transfer through proton hopping (Q-HOP), i.e. semi-QM in classical MD
    - Implemented by Volkhard group, University of Saarland, Germany
  - Set up and analyze runs with Ecce

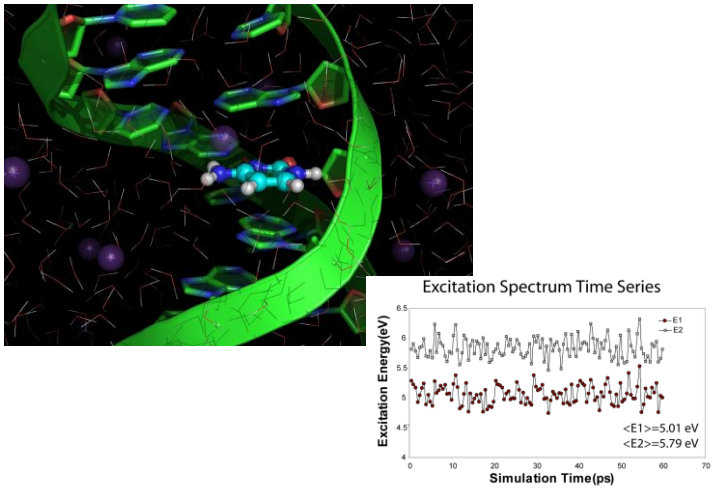


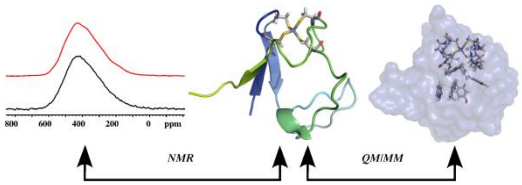


NWChem: extensive QM/MM (1)

- Seamless integration of molecular dynamics with coupled cluster and DFT
  - Optimization and transition states
  - QM/MM Potential of Mean Force
  - Modeling properties at finite temperature
    - Excited States with EOM-CC
    - Polarizabilities with linear response CC
    - NMR chemical shift with DFT





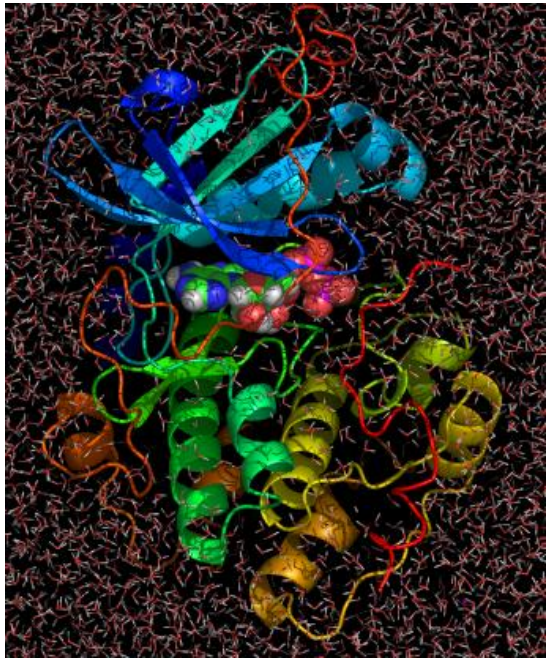


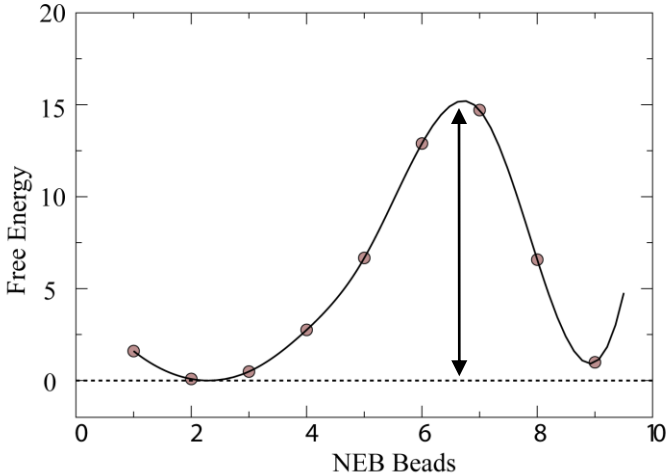
- Also connected to solid state MD module

QM/MM CR-EOM-CCSD(T) Excited State Calculations of cytosine base in DNA, Valiev et al., ***JCP*** *125 (2006)*

NWChem: extensive QM/MM (2)

- QM/MM for pathways (change in computing model)
  - NEB-QM/MM approach for Reaction Pathway Calculations
  - Free energy calculation





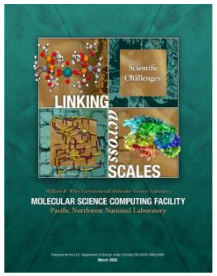
~15 kcal/mol reaction barrier

Reaction pathway and free energy barrier determination for the phosphorylation reaction in cAPK protein kinase using NEB QM/MM approach, *Valiev et al JPC B. 111 (2007)*

NWChem: other functionality

- Other functionality available in NWChem
  - NMR shielding and indirect spin-spin coupling
  - COSMO
  - ONIOM
  - Relativity through spin-orbit ECP, ZORA, and DK
    - Open-shell spin-orbit coupled cluster
  - Electron transfer
  - Vibrational SCF and DFT for anharmonicity
  - Module for dynamic nucleation theory Monte Carlo
    - Windus (Iowa State Univ.), Kathman (PNNL)
  - Interface with VENUS for chemical reaction dynamics
  - Interface with POLYRATE, Python
  - Interface with NBO

NWChem Software Refresh Strategy

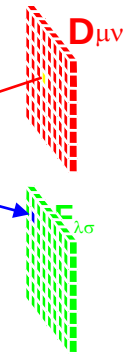


Evolving Science Drivers.
Input gathered from:
1) Feedback from users
2) EMSL Greenbook

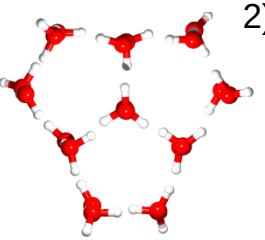
$$\rho(\mathbf{x}_Q) = \sum_{\mu\nu} D_{\mu\nu}\, \chi_\kappa(\mathbf{x}_Q)\, \chi_\nu(\mathbf{x}_Q)$$

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

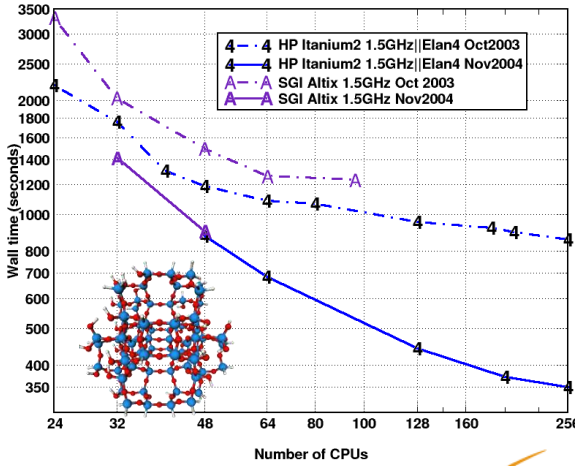
```
my_next_task =
SharedCounter()
  do i=1,max_i
    if(i.eq.my_next_task)
then
      call ga_get()
        (do work)
      call ga_acc()
      my_next_task =
SharedCounter()
    endif
  enddo
barrier()
```



Implementation of the science needs.
Porting to new architectures,
Performance tuning on the hardware.
Support/maintenance phase.



MS³
MOLECULAR SCIENCE SOFTWARE SUITE



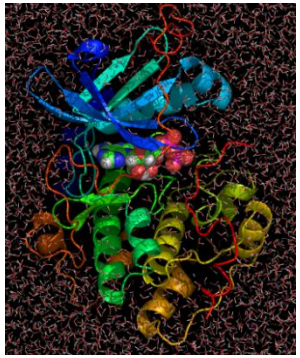
New SW features/methods made available
to users through a yearly release process.
Testing phase preliminary to the
distribution phase.

Pacific Northwest
NATIONAL LABORATORY

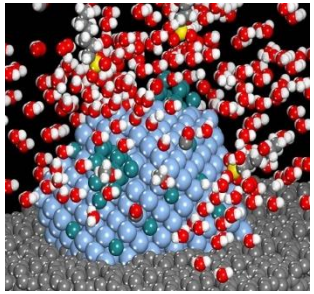
Future Scientific Directions

EMSL

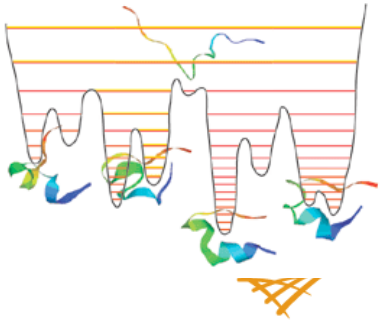
- Kinetics and dynamics of chemical transformations
  - Predict and design catalytic activity and selectivity by accurately modeling reaction mechanisms and rates
  - Fundamental understanding of solar energy conversion processes through complex chemical reaction pathways in materials and biological systems



- Chemistry in condensed phase and at interfaces
  - Understand and predict chemical processes at surfaces and interfaces
  - Nucleation and growth at nanoscale and beyond



- Spanning longer times for biological systems
  - Biologically relevant processes induced by large conformational changes (transitions or reactions)
  - Time scale of changes greater than the time scale currently simulated (~100 nanoseconds)



Pacific Northwest
NATIONAL LABORATORY

Moving Beyond Petascale, working towards Exascale

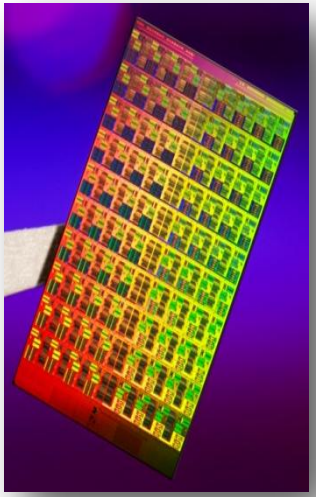
EMSL

- Efficient NWChem on petascale platform and beyond enables our users
  - Tackle larger molecular complexes with greater accuracy
  - Integration of methods and dynamics
  - Obtain their results with a faster time-to-solution
- Goal is to enable NWChem to run effectively on petascale platforms and beyond
  - Understanding scalability issues in computational chemistry
  - Software infrastructure and architecture (re)design
    - 100,000s of complex compute nodes
      - Each with multiple (heterogeneous) cores
    - Computing paradigm is changing drastically
    - Actor based models for efficient load balancing
    - Applications need fault resilience and/or tolerance

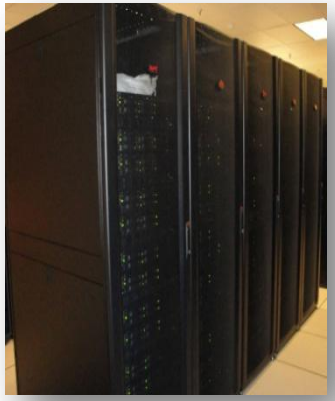
Pacific Northwest
NATIONAL LABORATORY

NWChem - Chemistry on New Architectures

- Large multicore compute nodes
  - Heterogeneous nodes and/or heterogeneous computing in a node
- GPGPU
  - Rapidly evolving field
- Heterogeneous networks
  - Align data movement with network topology/bandwidth



Intel's Teraflop Research Chip has 80 cores and 1 TFlops performance



Barracuda: NWChem development cluster funded by Recovery Act

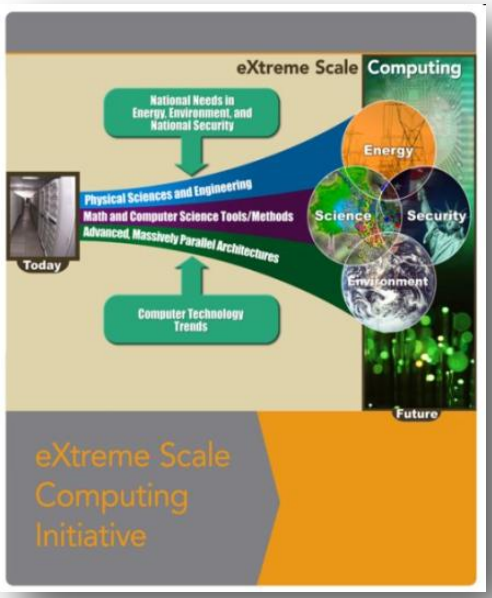
NVIDIA Tesla S1070
4 TFlops in 1U

Pacific Northwest NATIONAL LABORATORY

NWChem: Achieving scalability will take a large team

- Extreme scaling requires a collaborative multidisciplinary development team
  - Computational chemists
  - Computer scientists
  - Mathematicians

- PNNL's eXtreme Scale Computing Initiative leads the way
  - Scalable computational chemistry methods
  - Fault tolerance at the application level
  - Scalable software infrastructure



NWChem: Going Open-Source

- NWChem consortium to deliver code and infrastructure for computational chemistry community to build upon
- Greatly expands user base through ease of access by students
- Expanding developers base by leveraging development and resources of universities and national labs
- Establish more collaborative development environment
- License will be Educational Community License (ECL)

Fostering new developers and developments in NWChem

- Jeff Hammond (PD at Argonne)
  - 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

- Prof. Lasse Jensen (Penn State U)
  - New developer, very excited
  - Working on the development of a Raman module, Surface Enhanced Raman Spectroscopy (SERS) application

Summary

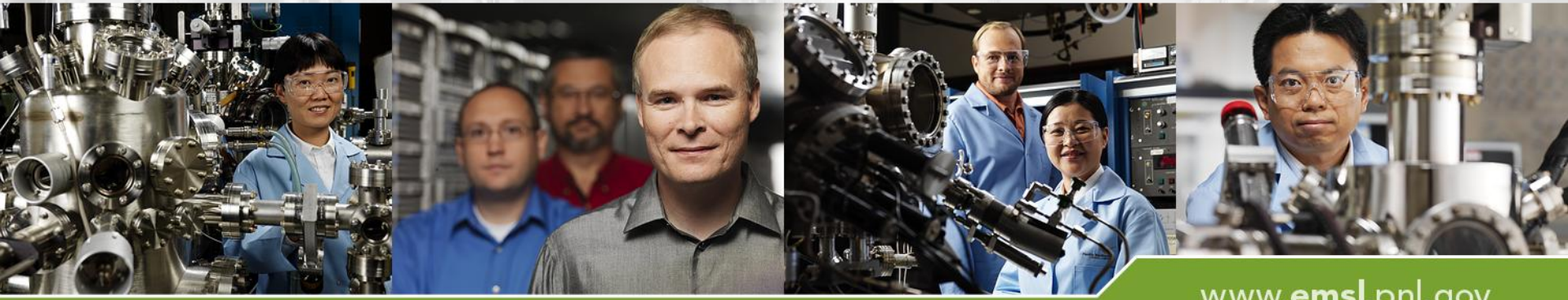
- NWChem provides users with a wide range of capabilities to solve scientific problems
  - High accuracy to DFT to molecular mechanics
  - Scalable with science problem and computing resources
  - Various capabilities can be coupled together
  - Freely available

- Point of contact: Bert de Jong, bert.dejong@pnl.gov

- http://www.emsl.pnl.gov/docs/nwchem/nwchem.html

Environmental Molecular Sciences Laboratory

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



www.emsl.pnl.gov

EMSL | U.S. DEPARTMENT OF ENERGY

Pacific Northwest NATIONAL LABORATORY