EMSL

Basic introduction of NWChem software

www.emsl.pnl.gov

Pacific Northwest NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

U.S. DEPARTMENT OF ENERGY

Background

EMSL

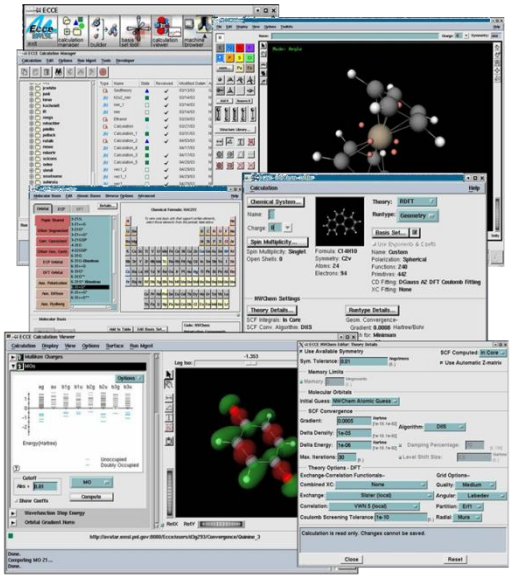
- NWChem is part of the Molecular Science Software Suite

MS^3
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

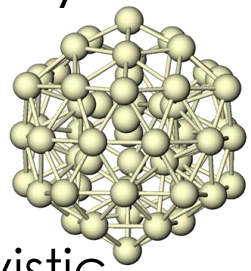
EMSL

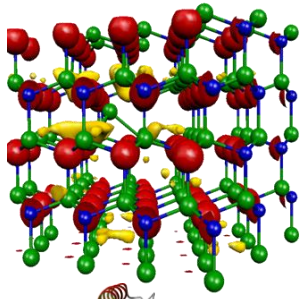
- Originally designed for parallel architectures
  - Scalability to 10,000's of processors (partly even to 100,000)
- Emphasis on modularity, portability, and integration
- Portable – runs on a wide range of computers
  - Supercomputer to Mac or PC with Windows
  - Now runs efficiently on IBM BlueGene, Cray XT, InfiniBand
- Uses Global Arrays/ARMCI for parallelization
- **NWChem 6.1 is open-source and freely available**

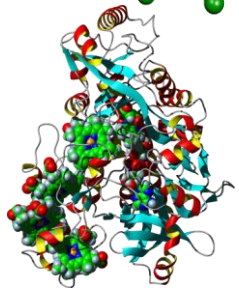
NWChem Science Capabilities

EMSL

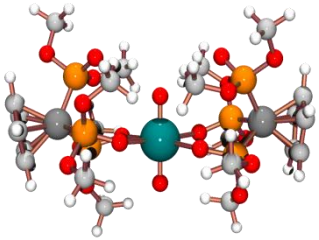
- Provides major modeling and simulation capability for molecular science
  - Broad range of **molecules**, including **bio**molecules, **nano**particles and heavy elements
  - Electronic structure of molecules (non-relativistic, relativistic, ECPs, first and second derivatives)
  - **Solid state** capability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







- World-wide distribution
  - 70% is academia, rest government labs and industry
- About 140/year publications citing NWChem



NWChem's core developer team

Bert de Jong
Team lead
Properties/Relativity

Karol Kowalski
High accuracy

Niri Govind
Density functional theory

Ken Lopata
EMSL Wiley Postdoc

Eric Bylaska
Plane wave methods

Edoardo Aprà
DFT & HPC

Marat Valiev
QM/MM

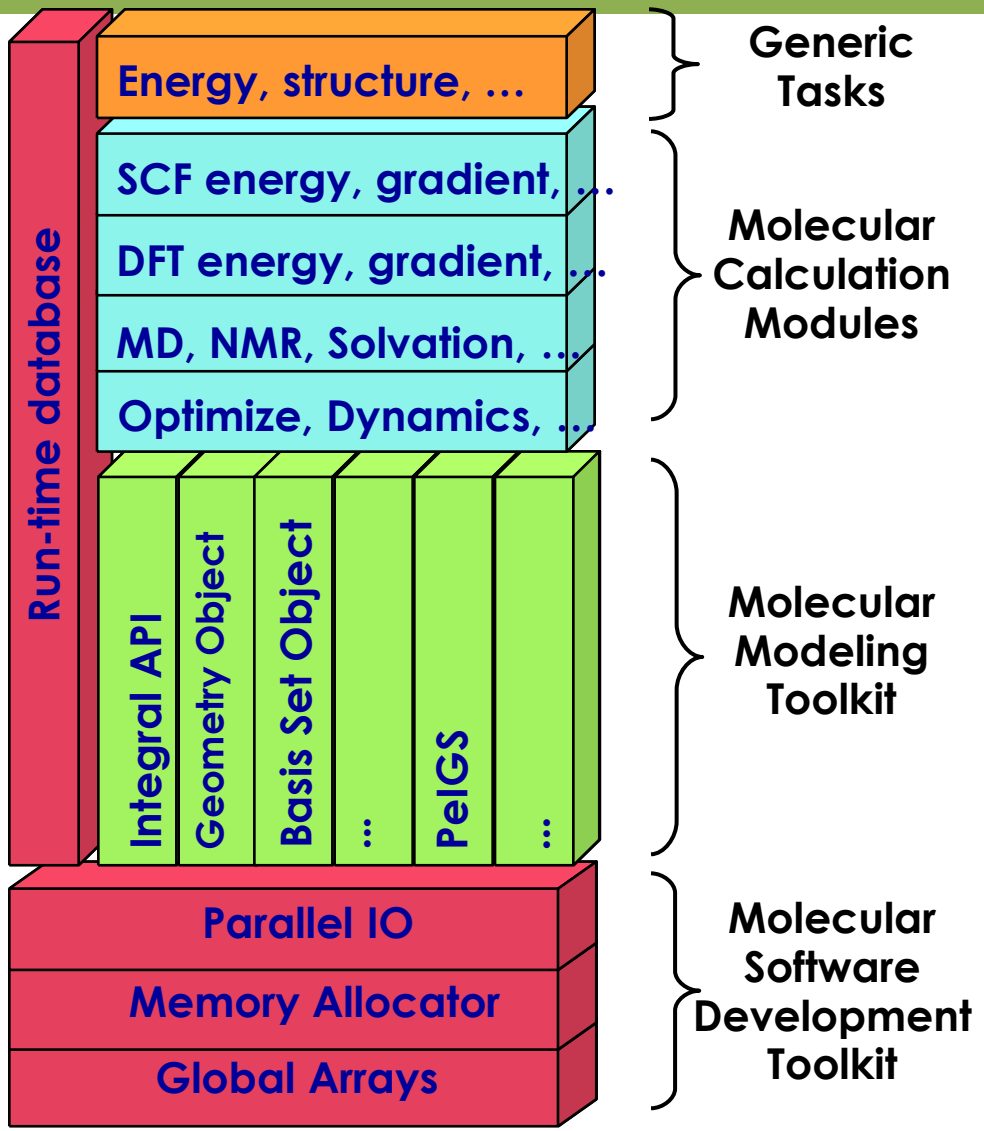
Huub van Dam
DFT/HPC

NWChem capabilities overview

EMSL

- NWChem brings a full suite of methodologies to solve large scientific problems
  - Gaussian-based DFT/TDDFT
    - Ground & Excited States, Optimization, Properties (NMR, Electric field gradient, linear response,…)
  - Plane wave based DFT
    - Car-Parinello MD (CPMD), Band Structure, Optimization, etc.
  - High Accuracy Methods → MP, CC, EOMCC
    - Ground & Excited States
  - Molecular Dynamics, Molecular Mechanics
  - Integrated Methodologies → QM/MM
  - Scripting → Python

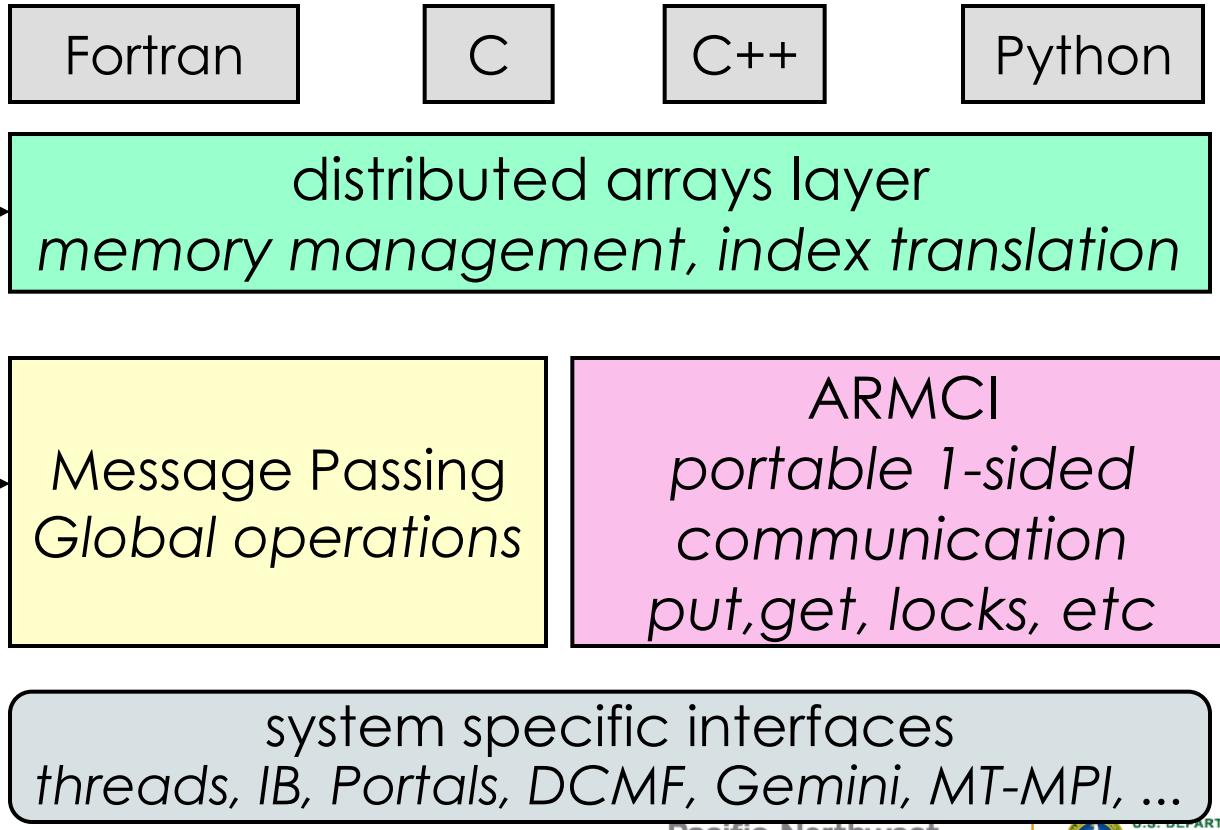
NWChem Structure



- **Object-oriented design**
  - abstraction, data hiding, APIs
- **Parallel programming model**
  - non-uniform memory access, Global Arrays, MPI
- **Infrastructure**
  - GA, Parallel I/O, RTDB, MA, …
- **Program modules**
  - communication only through the database
  - persistence for easy restart

Structure of GA

Application programming language interface

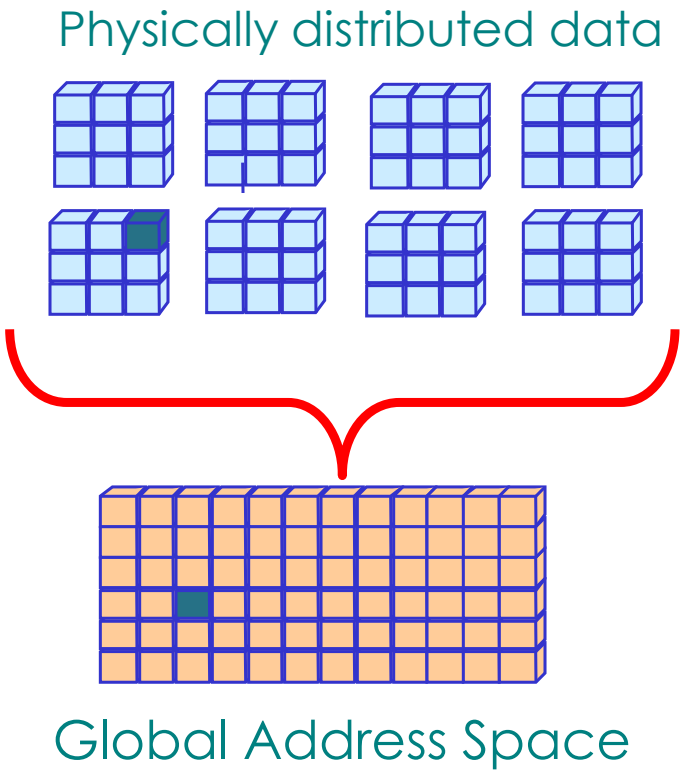


Global Arrays and MPI are completely interoperable Code can contain calls to both libraries.

Global Arrays

EMSL

- **Distributed dense arrays** that can be accessed through a shared memory-like style
- **High level abstraction** layer for the application developer (that's me!)
- **One-sided** model = no need to worry and send/receive



single, shared data structure/ global indexing

e.g., access A(4,3) rather than buf(7) on task 2

NWChem: Gaussian DFT (1)

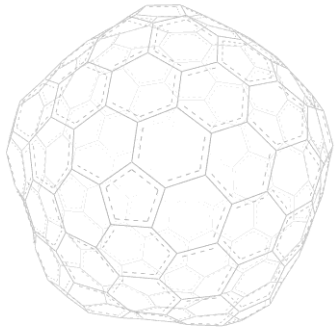
EMSL

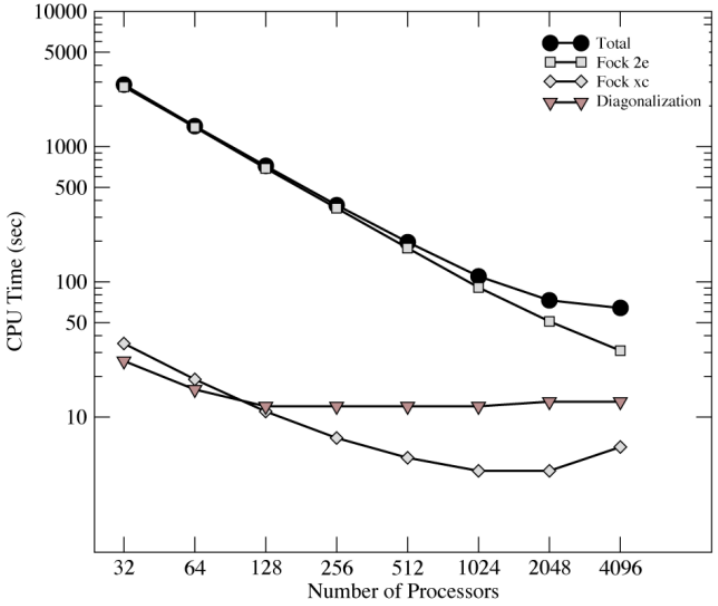
- Gaussian based DFT → Finite systems (molecules, clusters, nanostructures)
  - Wide range of local and non-local exchange-correlation functionals
    - Traditional xc functionals
    - Wide range of hybrid functionals (B3LYP, PBE0, BeckeH&H…)
    - HF Exchange
    - Meta-GGA functionals
    - Minnesota functionals (M05, M06)
    - SIC and OEP
    - Range separated functionals
    - DFT + D implementation (long-range empirical vdW)
    - Double hybrid functionals
  - Spin-orbit DFT
    - ECP, ZORA, DK
  - Constrained DFT
  - TDDFT for excited states → Optical spectroscopy
  - Various properties (NMR, Linear response,…)

NWChem: Gaussian DFT (2)

EMSL

- Calculation on C_{240}
  - PBE0 functional, 6-31G*
  - Direct integral evaluation
  - Size 3600 basis functions
- Timings for different components of the Kohn-Sham matrix construction
  - Fock 2e – two electron integrals
  - Fock xc – the DFT contribution
  - Diagonalization – eigenvector solve
- Scalability limited by diagonalization
- Fock_2e cost can be dramaticc reduce by using Coulomb-fitting





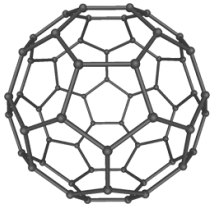
NWChem: High Accuracy Methods (1)

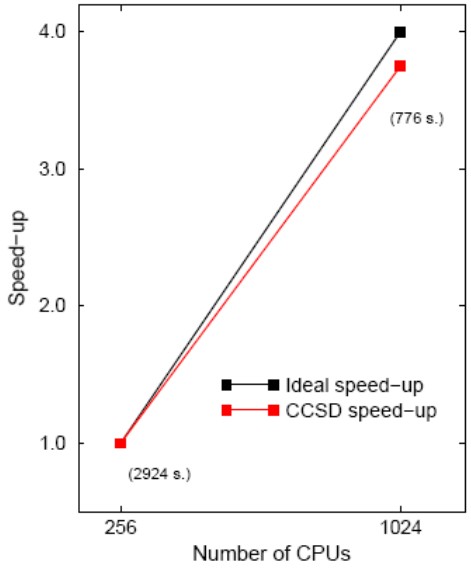
EMSL

- Coupled Cluster
  - Closed shell coupled cluster [CCSD and CCSD(T)]
  - Tensor contraction engine (TCE)
    - Spin-orbital formalism with RHF, ROHF, UHF reference
    - CCSD,CCSDT, …
    - CCSD(T), CR-CCSD(T), …
    - EOMCCSD,EOMCCSDT
    - Linear response CC (polarizabilities, hyperpolarizabilities)
    - Active-space CCSDt/EOMCCSDt

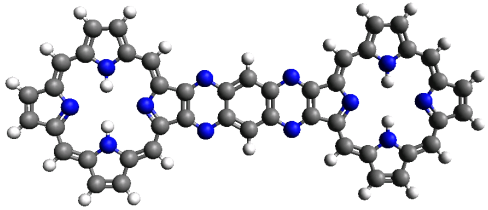
NWChem: High Accuracy Methods (2)

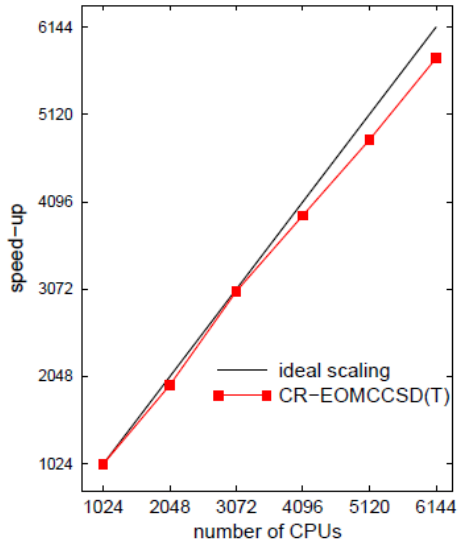
- Extensive development of scalable algorithms





CCSD calculation of C_{60}
(1080 basis set functions)



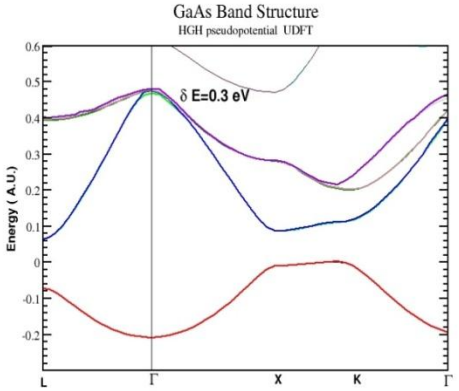


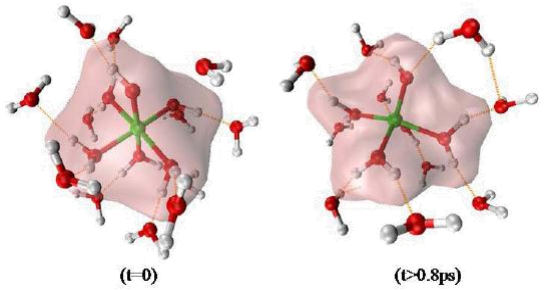
CR-EOMCCSD(T) calculation of the Porphyrin dimer linked by a tetraazaanthracene bridge (P_2TA)

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

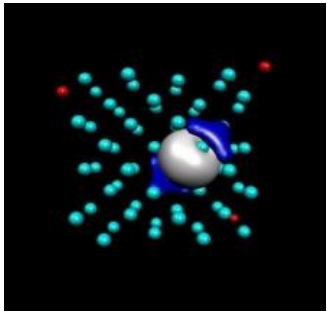
Spin-Orbit splitting in GaAs





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

SIC localization



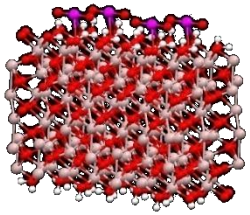
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

U.S. DEPARTMENT OF ENERGY

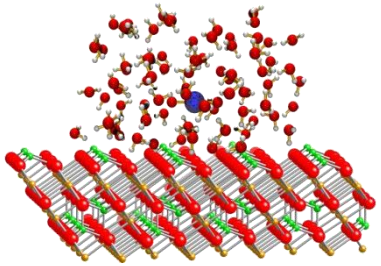
NWChem: Plane wave (2)

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

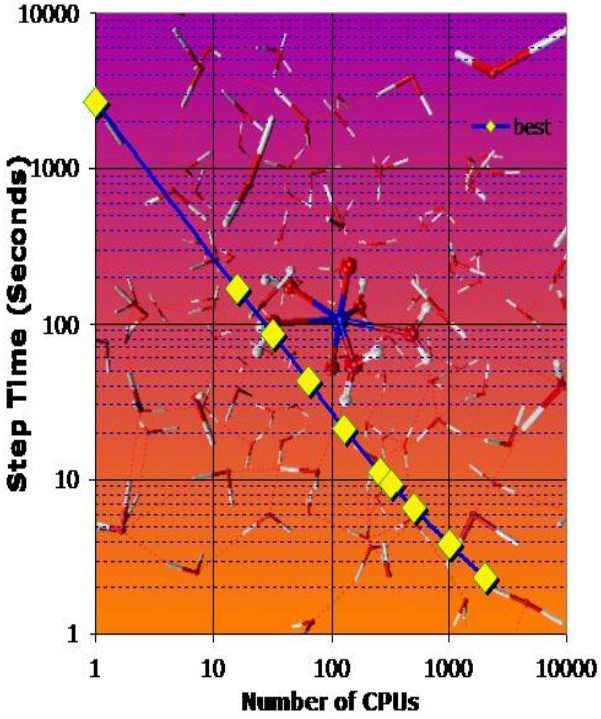
- Can also do plane-wave QM/MM



Uranyl on a hydroxylated Al_2O_3 surface



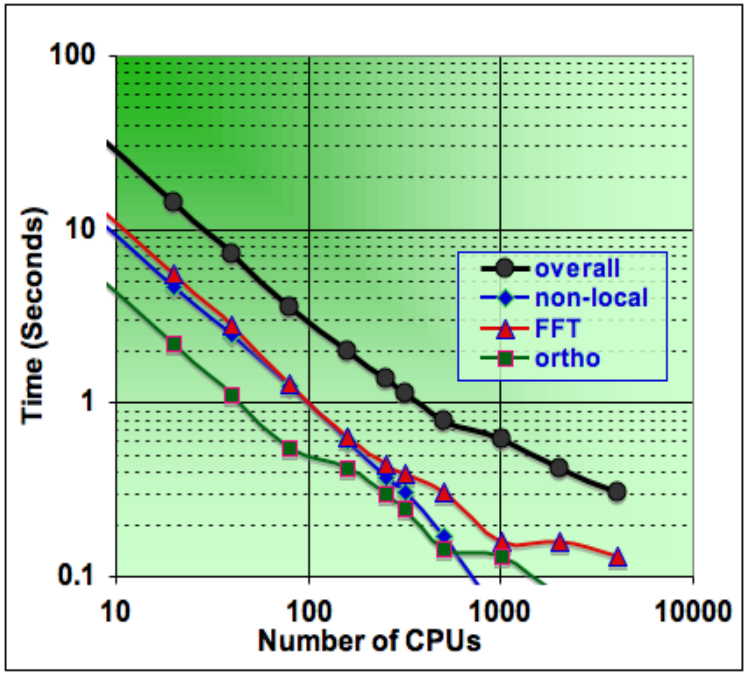
Uranyl in solution interacting with iron oxide



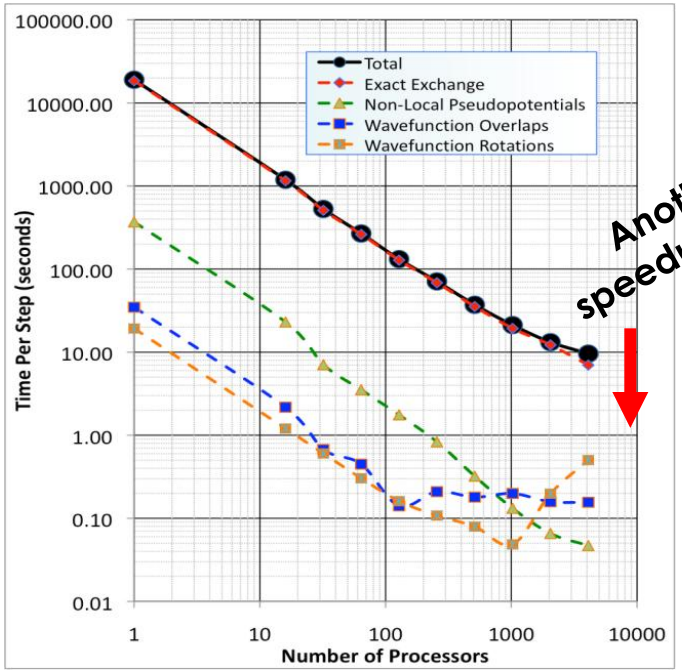
Car-Parrinello plane wave performance, PBE96 GGA Functional, -300 K thermostat, 0.121 fs time step, 122 water molecules-15.6 Å box

NWChem: Plane wave (3)

- Extensive work done to develop parallel plane wave algorithm for hybrid-DFT solvers
  - Results below obtained on NERSCs Franklin machine



DFT calculation on $Nb_{10}O_{28}^{6-}$– O(Ne)

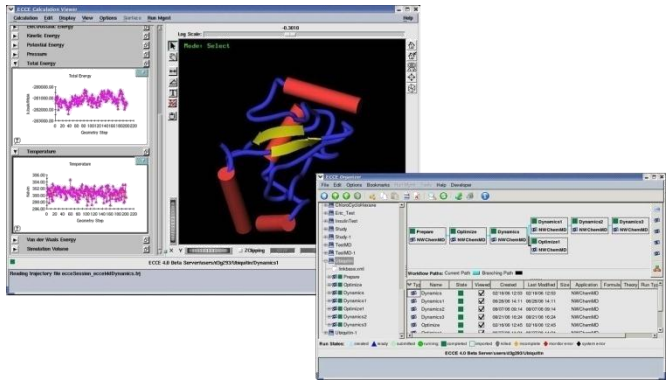


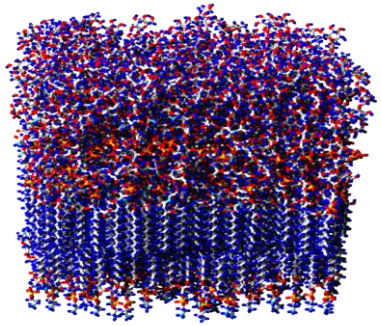
Hybrid DFT calculation on 80 atom cell of hematite– O(Ne*Ne)

NWChem: Molecular Dynamics

EMSL

- Molecular dynamics
  - Charmm 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 Helms group, University of Saarland, Germany
  - Set up and analyze runs with ECCE





Pacific Northwest NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY

Proudly Operated by Battelle Since 1965

NWChem: Hybrid QM/MM

EMSL

- 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 EOMCC, TDDFT
    - Polarizabilities with linear response CC
    - NMR chemical shift with DFT
- QM/MM for pathways
  - NEB-QM/MM approach for Reaction Pathway Calculations
  - Free energy calculation

NWChem: other functionality

EMSL

- Other functionality available in NWChem
  - NMR shielding and indirect spin-spin coupling
  - COSMO
  - ONIOM
  - Relativity through spin-orbit ECP, ZORA, and DK
  - Electron transfer
  - Vibrational SCF and DFT for anharmonicity
  - Module for dynamic nucleation theory Monte Carlo
  - Interface with VENUS for chemical reaction dynamics
  - Interface with POLYRATE, Python
  - Interface with NBO

NWChem Input Basics

EMSL

- Minimal input (all defaults)

```
geometry
   n    0.00 0.00 0.00
   n    0.00 0.00 1.08
end

basis
   n library cc-pvdz
end

task scf
```

- Performs a closed-shell SCF on the N_2 molecule

Geometry Input: Units

- Input can be in Angstrom or atomic units

```
geometry  # units are in angstroms
  C   0  0      0
  H   0  0.9885 -0.4329
  H   0 –0.9885  0.4329
end
```

OR

```
geometry units au  # change units to a.u.
  C   0  0      0
  H   0  1.868 -0.818
  H   0 -1.868  0.818
end
```

Geometry Input: Symmetry

- Water molecule with C_{2v} symmetry

```
geometry units au #input using symmetry
  C   0  0      0
  H   0  1.868 -0.818
  symmetry c2v
end
```

- C_{60} with I_h symmetry

```
geometry  #bonds = 1.4445 and 1.3945 Angstrom
    symmetry Ih
    c -1.2287651    0.0    3.3143121
end
```

Geometry Input: autosym and autoz

- By default NWChem will:
  - Attempt to find symmetry if none is specified
  - Attempt to build a z-matrix from cartesian coordinates (for the geometry optimization)
  - Center the molecule in the reference frame

```
geometry noautoz noautosym nocenter
   C   0  0      0
   H   0  0.9885 -0.4329  #Angstroms
   H   0 –0.9885  0.4329
end
```

Geometry Input: zmatrix

- Geometry can be specified using a z-matrix format

```
geometry
   zmatrix
     O
     H1 O 0.95
     H2 O 0.95 H1 108.0
   end
end
```

- Distances and angles can be specified with variables too (see documentation)

Geometry Input: zmatrix

- Distances and angles can be specified with variables

```
geometry
   zmatrix
     O
     H1 O doh
     H2 O doh H1 ahoh
     variables
       ahoh 108.0
       doh 0.95
   end
end
```

Geometry Input: zcoord

- Forcing internal coordinates

```
geometry
  Si      0.0000E+00  0.0000E+00  0.0000E+00
  H      -0.9436E+00 -0.8807E+00  0.7319E+00
  H       0.7373E+00 -0.8179E+00 -0.9932E+00
  H      -0.7835E+00  0.1038E+01 -0.7137E+00
  Si      0.1699E+01  0.1556E+01  0.1695E+01
  H       0.7715E+00  0.2377E+01  0.2511E+01
  H       0.2544E+01  0.6805E+00  0.2539E+01
  H       0.2514E+01  0.2381E+01  0.7713E+00
end

### fix the Si-Si distance to 4.0 angstroms ###
geometry adjust  # initial state
  zcoord
    bond 1 4 4.00 r constant
  end
end
```

Geometry Input: system

- Crystal lattice, used in plane wave code, for 3-D periodic systems (crystals)

```
geometry units angstroms center noautosym noautoz print
  system crystal
    lat_a 3.625d0      #diamond
    lat_b 3.625d0
    lat_c 3.625d0
    alpha 90.0d0
    beta  90.0d0
    gamma 90.0d0
  end
  C     -0.50000d0  -0.50000d0  -0.50000d0
  C      0.00000d0   0.00000d0  -0.50000d0
  C      0.00000d0  -0.50000d0   0.00000d0
  C     -0.50000d0   0.00000d0   0.00000d0
  C     -0.25000d0  -0.25000d0  -0.25000d0
  C      0.25000d0   0.25000d0  -0.25000d0
  C      0.25000d0  -0.25000d0   0.25000d0
  C     -0.25000d0   0.25000d0   0.25000d0
end
```

Basis Set Input: Using libraries

EMSL

- Atoms can be defined by symbol and name

```
basis
     O library cc-pvdz
     H library cc-pvdz file /usr/d3g681/nwchem/libraries/
end
```

- * can be used to state that all atoms in the system should be using the same basis set type

```
basis
      * library cc-pvdz
end
```

Basis Set Input: Explicit basis sets

EMSL

- Basis set input can be done with exponents and coefficients

```
basis spherical
  H s
    13.0100 0.019685
    1.9620   0.137977
    0.4446   0.478148
    0.1220   0.501240
 H s
    0.1220   1.000000
 H p
    0.7270 1.000000
end
```

Basis Set Input: Explicit basis sets

EMSL

- Libraries and explicit input can be used together

```
basis spherical
   * library cc-pvdz
H p
   0.007270 1.000000
end
```

Task Input

- Task directive tells NWChem what it should do

```
task scf                          # default is energy
task scf energy

task dft optimize
task dft saddle
task ccsd frequencies

task pspw optimize

task md dynamics
```

Task Input

- Tasks are preformed in sequence as listed in input

```
task scf energy

task dft optimize ignore     # ignore if failed, go to next task
task dft saddle
task ccsd frequencies
```

Restarting a calculation

- To restart NWChem will need certain files, that should be saved in permanent directory

```
start ne
permanent_dir /users/me
geometry
   ne 0 0 0
end
basis
   ne library cc-pvdz
end
task scf
```

```
restart ne
permanent_dir/users/me
scf
   thresh 1e-8
end
task scf
```

Setting memory and charge keyword

EMSL

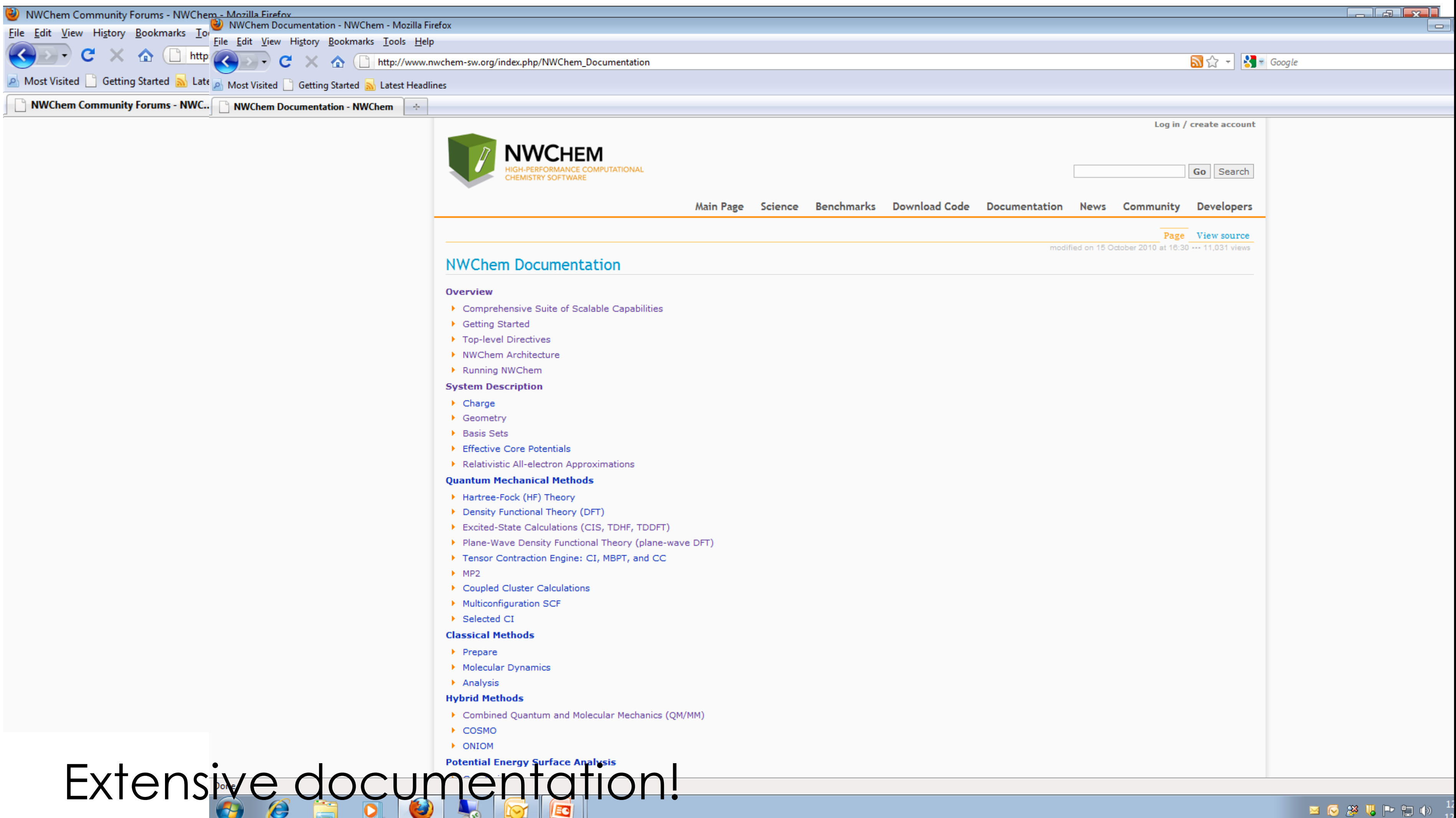
- If NWChem fails with an error asking for more memory, you can set it explicitly

```
memory 2400 mb
```

- Remember, memory is per processor!
- By default, molecules have a neutral charge (0)

```
charge -1
```

NWChem web pages



Extensive documentation!

NWChem website

EMSL

http://www.nwchem-sw.org

NWChem Mailing List

- Archive at http://www.emsl.pnl.gov/docs/nwchem/nwchem-support/
- To subscribe

  mail majordomo@emsl.pnl.gov

  subscribe nwchem-users
- To post

  mail nwchem-users@emsl.pnl.gov

EMSL

Questions … ?

Pacific Northwest NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY

Proudly Operated by Battelle Since 1965