Basic introduction of NWChem software
Background

- NWChem is part of the Molecular Science Software Suite
- 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 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**
Provides major modeling and simulation capability for molecular science

- Broad range of molecules, including biomolecules, nanoparticles 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

- 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
• 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

NWChem Structure

Run-time database

Energy, structure, ...

SCF energy, gradient, ...

DFT energy, gradient, ...

MD, NMR, Solvation, ...

Optimize, Dynamics, ...

Generic Tasks

Molecular Calculation Modules

Integral API

Geometry Object

Basis Set Object

Parallel IO

Memory Allocator

Global Arrays

Molecular Modeling Toolkit

Molecular Software Development Toolkit

Parallel I/O

Memory Allocator

Global Arrays
Structure of GA

Application programming language interface

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

- Fortran
- C
- C++
- Python

**distributed arrays layer**

memory management, index translation

Message Passing

Global operations

ARMCI

portable 1-sided communication

put, get, locks, etc

system specific interfaces

threads, IB, Portals, DCMF, Gemini, MT-MPI, ...

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

- **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

Physically distributed data

- single, shared data structure/
- global indexing
- e.g., access A(4,3) rather than buf(7) on task 2

Global Address Space
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 (1)
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 dramatically reduce by using Coulomb-fitting
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
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

Car-Parrinello provides evidence for five-coordinate Al(H$_2$O)$_4$OH$^{2+}$
Swaddle et al, Science, 2005

Spin-Orbit splitting in GaAs
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

Car-Parrinello plane wave performance, PBE96 GGA Functional, -300 K thermostat, 0.121 fs time step, 122 water molecules-15.6 Å box
Extensive work done to develop parallel plane wave algorithm for hybrid-DFT solvers

Results below obtained on NERSC's Franklin machine

DFT calculation on $\text{Nb}_{10}\text{O}_{28}$ & O(No)

Hybrid DFT calculation on 80 atom cell of hematite & O(No*No)
Molecular dynamics
- Charmm and Amber force fields
- Various types of simulations:
  - Energy minimization
  - Molecular dynamics simulation including \textit{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
NWChem: Hybrid QM/MM

- 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
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

- 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 N₂ molecule
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
```
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)
Distances and angles can be specified with variables

```plaintext
geometry
  zmatrix
  O
  H1 O doh
  H2 O doh H1 ahoh
variables
  ahoh 108.0
  doh 0.95
end
end
```
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
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
```
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 can be done with exponents and coefficients

<table>
<thead>
<tr>
<th>basis spherical</th>
</tr>
</thead>
<tbody>
<tr>
<td>H s</td>
</tr>
<tr>
<td>13.0100 0.019685</td>
</tr>
<tr>
<td>1.9620 0.137977</td>
</tr>
<tr>
<td>0.4446 0.478148</td>
</tr>
<tr>
<td>0.1220 0.501240</td>
</tr>
<tr>
<td>H s</td>
</tr>
<tr>
<td>0.1220 1.000000</td>
</tr>
<tr>
<td>H p</td>
</tr>
<tr>
<td>0.7270 1.000000</td>
</tr>
</tbody>
</table>

end
Basis Set Input: Explicit basis sets

- Libraries and explicit input can be used together

```plaintext
basis spherical
  * library cc-pvdz
H  p
  0.007270 1.000000
end
```
Task directive tells NWChem what it should do

task scf
 task scf energy

# default is energy

task dft optimize
 task dft saddle
 task ccisd frequencies

task pspw optimize

task md dynamics
Tasks are performed in sequence as listed in input:

```
task scf energy

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

task ccisd 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
```
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

Science that can be done
Information about scalability and performance
Download your own version
Detailed documentation and user manual
Questions can be asked on the community forums

Extensive documentation!
NWChem website

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
Questions ... ?