Introduction To NWChem

By

Benjamin J. Lynch

October 26, 2006

Outline

- Background
- Methods available
- Creating input files
- Running calculations at MSI

Background

- Computational chemistry package
- Developed at Pacific Norwest National Laboratory
- Free

NWChem mailing list

- Many NWChem users have the same questions you do.
- The authors of NWChem are very responsive to questions sent to this mailing list.
- To subscribe:

Compose an email to majordomo@emsl.pnl.gov

The body of the message should be:

```
subcribe nwchem-users
```

When do you want to use NWChem?

When you have a machine with a lot of memory (>256 GB)

When you want to run plane-wave calculations

When you are running into disk I/O problems with other programs

When you need to run a CCSDTQ calculation, or one of the many other theories that aren't available elsewhere

When you are using the **IBM BladeCenter**

Computational Chemistry on the BladeCenter

- Coming this Spring!
- 1140 cores
- 5.9 TFlops
- 30 TB of Disk
- IB interconnect

Methods Available

- HF/DFT
- SODFT
- MP2
- CASSCF

- TDDFT
- CCSD(T)
- CCSDTQ
- Non-canonical MP2, MP3, MP4

- Relativistic effects
- Plane-wave basis sets
- Classical molecular dynamics
- CCSDTQ dipoles

- Classical dynamics

Input Format

```
start water
Title “H2o energy”

geometry
…
end

Basis
 * library 3-21G
end

MP2; freeze atomic; END

Task mp2 energy
```

Start a new calculation (not a restart job)

Each directive is specified by a directive name, options, and ends with the word “end”

Input is not case sensitive

Semicolons can replace new lines.

One or more tasks are specified

Input Format

is comment character

directives are read and tasks are executed
in the order they appear in the input file.

Some of the input directives

echo	Copy the contents of the input file into the output
scratch_dir	Set the location of the scratch files
permanent_dir	Location of important summary files and database file (for restart)
title	An optional title can be given to keep track of calculations

start	Start a new calculation
restart	Restart a calculation using information from the .db file
memory	Specify the maximum memory usage

mp2	Control mp2-related options
ccsd	Define ccsd-related options
dft	Define exchange-correlation functional and grid
tce	Specify options for a calculation using the Tensor Contraction Engine

geometry	List of atoms and coordinates
scf	Specify the type of wavefunction and convergence options.
basis	Define a basis set for the system
task	Execute a specified task
python	Embed python code to control NWChem

charge	Set the charge of the system
set	Modify value in the .db file
unset	Delete a value in the .db file
stop	Stop NWChem
dirdyvtst	Run POLYRATE calculation using methods in NWChem

.db files

- When you run nwchem, it will create a file named input.db
- This file contains restart information
- Data in the database can be accessed using the **`set`** and **`unset`** directives

GEOMETRY directive

The geometry can be specified in Cartesians or z-matrix.

```
geometry
  C  0.0   0.0  0.0
  H  0.742 0.0  0.0
end
```

Z-matrix example

```
geometry
  zmatrix
    O
    H1 1 r1
    H2 1 r1 2 a1
   variables
    r1    0.95
    a1    108.0
  end
end
```

Fixing a variable

```
geometry
  zmatrix
    O
    H1 1 r1
    H2 1 r1 2 a1
   variables
    a1   108.0
   constants
    r1   0.95
  end
end
```

Units

```
geometry units atomic
  C   0.0   0.0   0.0
  H   1.384 0.0   0.0
end

geometry units pm
  C   0.0   0.0   0.0
  H   73.256 0.0   0.0
end
```

Default units are angstroms

Symmetry

- Defining symmetry occurs within the geometry directive

```
geometry
  C 0.0 0.0 0.0
  …
  symmetry D6h
end
```

Symmetry example

```
geometry units au
   C 1.855 1.855 0
   H 3.289 3.289 0
   symmetry D6h
end
```

Only the symmetry-unique atoms are listed for this benzene molecule

- The symmetry keyword is sometimes required

```
geometry units au
   symmetry c2v
   O 1.4 0.0 0
end
```

This is an O_2 molecule

The symmetry keyword is required for some calculations on linear molecules.

crystals

The lattice parameters are specified in the geometry directive.
The coordinates for atoms in the unit cell are then specified.

```
geometry
  system crystal
  lat_a 5.00
  lat_b 5.00
  lat_c 5.00
  alpha 90.0
  beta  90.0
  gamma 90.0
  end
  Na     1.0   1.0   1.0
  Cl     1.0   1.0  -1.0
  Na    -1.0   1.0   1.0
  …
```

BASIS directive

```
basis
  * library 3-21G
end
```

This will specify the 3-21G basis set from the NWChem library for all atoms

```
basis
  C library 3-21G
  O library 6-31G
end
```

This will specify the 3-21G basis set from for carbon atoms and
The 6-31G basis set for oxygen atoms

Defining your own basis set

```
basis
  hydrogen s
      13.0     0.019
      1.96     0.138
      0.44     0.478
      0.12     0.501
  hydrogen p
     0.141     1.0
end
```

SCF related options

`SEMIDIRECT, DIRECT`	(integral storage method)
`Thresh`	(convergence threshold)
`MAXITER`	(maximum SCF iterations)

```
scf
  MAXITER 30
  doublet
  uhf
end
```

ECP

- Effective core potentials are specified very similar to the way that basis sets are defined.

```
ecp
  * library crenbl_ecp
end
```

ECPs can also be explicitly defined

```
ecp
  O nelec 2                  # replace 2 core
  electrons
  O ul                       # d
    1    80.0     -1.60
    1    30.0     -0.40
    2    1.96     -0.066
  O s                        # s - d
    0    0.92      0.395
    0   28.65      2.52
    2    9.30      17.04
  O p                        # p - s
    2    52.34     27.97
    2    30.72    -16.496
end
```

Tensor Contraction Engine (TCE)

- Developed by some folks at Ohio State and Oak Ridge National Lab.
- Generates “low-level” FORTRAN code for high-level mathematical expressions.
- Used in NWChem to generate code to perform various high levels of *ab initio* theory.

Tensor Contraction Engine (TCE)

- TCE section in NWChem allows many calculations, including:
  - MP2
  - CCD
  - CCSD
  - QCISD
  - MP4
  - CISDTQ
  - CCSDTQ
  - EOM-CCSDTQ

All levels are available from RHF, UHF, and ROHF reference functions

TCE directive syntax

```
tce
   CCSDT
end

task tce energy
```

Tensor Contraction Engine (TCE)

- When should the TCE be used?
  - When you can't do the calculation otherwise
- **NWChem has more than 1 way to do the same calculation**
  - **The original MP2, CCSD, and CCSD(T) code in NWChem is more efficient than using the TCE (as of version 4.7)**
  - **Original CCSD code can only handle RHF reference wavefunctions, TCE can handle RHF, UHF, or ROHF.**

A few notes about CC calculations

- CCSDT is slightly less accurate than CCSD(T) for many applications and it's much more expensive
- CCSDTQ is slightly more accurate compared to CCSD(T) and it's incredibly expensive
  - CCSDTQ uses > 80000 lines of Fortran

Scaling of electronic structure calculations

Method	Scaling
Hartree-Fock	N^4
Hybrid DFT	N^4
MP2	N^5
CCSD	N^6
CCSD(T)	N^7
CCSDT	N^8
CCSDTQ	N^{10}

Plane Waves

- Instead of using atom-centered gaussian functions, you can also use plane waves.
- Plane waves provide a systematic way to improve a basis set
- Plane-waves are lousy at describing the rapidly-changing electron density near the nucleus.
  - Pseudopotentials are almost always used, even for 1st-row atoms (B, C, N, O, F)
- These methods are often used for calculations on bulk solids

Plane Waves

```
nwpw
   energy_cuttoff 40.0
   xc_pbe96
   ewald_ncut 8
end

task pspw energy
```

Pseudopotential generator

```
PSPW
  PSP_GENERATOR
  ELEMENT “Fe”
  CHARGE 26.0
  …
  END
END
```

DFT Calculations

```
DFT
   GRID FINE
   XC MPW1K
END

TASK DFT energy
```

DFT Functionals available

ACM

B3LYP

MPW1K

XPBE96

And 39 other GGA and meta-GGA functionals (in version 4.7)

Correlation energy

- By default, **NO ORBITALS** are frozen

```
mp2
    freeze atomic
End
ccsd
    freeze atomic
end
```

Freezing a few orbitals

```
mp2
    freeze core 1
end
ccsd
    freeze virtual 2
end
```

ROMP4 correlation energy

```
scf
  rohf
end

tce
  mp4
end

task tce energy
```

CCSDTQ using TCE

nproc	walltime	speedup
1	2242.5	
2	1168.0	1.9
4	708.8	3.2
6	508.6	4.4

Specifying memory usage

memory total 1000 mb

memory heap 250 stack 250 global 500 mb

MEMORY

Individual routines will allocate memory

Program will crash if it gets to a routine that asks for too much memory

NWChem usually will give a good error message.

If you are running the job in parallel, there will be many MPI errors that will follow

I/O

- Large QM calculations require tens or hundreds of gigabytes of temporary storage.
- The default I/O scheme in NWChem is to use a Global Array virtual file.
  - No writing to disk
  - Information is kept in memory, and distributed across all the nodes available.
- Several other disk I/O schemes are available if your calculation requires more memory.

Python

Python is an interpreted language (like Perl).

Python can be used to control NWChem, manipulate values from the .db file, and anything else you can program in python.

Python in NWChem

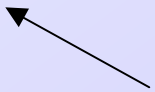
```
python
  print ‘Hello World!’
end
```

All python code is shifted left by as many spaces as there are in the first line when it is sent to the python interpreter.

Python example

```
python
   print 'Hello World!'
   print 'Hello again!'
end
```

print 'Hello World!'
print 'Hello again!'



This is sent to the interpreter

Another Python example

```
python
                print 'Hello World!'
                print 'Hello again!'
end
```

```
print 'Hello World!'
print 'Hello again!'
```

The same lines are sent to the interpreter

Python

```
python
   print 'Hello World!'
 print 'Hello again!'
end
```

NWChem will die.

Indentation must be >= that of the first line

nwchem routines can be called from your python script

```
python
   task_energy('ccsd')
```

running an energy calculation

```
   rtdb_put("mynum",42)
   print rtdb_get("mynum)
end
```

storing/retrieving numbers from the NWChem database file

Multi-step jobs

To run multiple calculations, simply insert additional task directives, and change other variables when needed. Calculations will be performed in the order they appear in the input file.

```
task DFT energy
task scf energy
```

DFT calculation

HF calculation

Multi-step jobs

```
mp2; freeze atomic; end
task mp2 energy
mp2; freeze core 0; end
task mp2 energy
```

NWChem at MSI

Login to blade, altix, or regatta

```
module load nwchem
```

```
qnwchem -m 600 mb -p 2 myinput.nw
```

Interactive run (always 1 processor)

```
module load nwchem

rnwchem myjob.nw >& myjob.out
```

Production and Test runs on blade

```
module load nwchem
qnwchem -p 256 myinput.nw

qnwchem -q test -p 4 myinput.nw
```

```
qnwchem will use p/4 nodes and will
  always request all available memory on
  the node (~7 GB)
```

Updates

- NWChem version 4.7 is currently installed. Version 5.0 will be installed soon. Watch your email from cc-list@msi.umn.edu

- Subscribe to the cc-list if you are not already on the list. Send a message to cc-list-request@msi.umn.edu and type the word **`subscribe`** in the body of the message.

Questions ?

Email: help@msi.umn.edu