Computational Chemistry
An introduction to the computational chemistry package, Gaussian 03
Matt McKenzie
LSU–HPC / LONI
Topics

- Basic Theory
  - Gaussian specific

- Gaussian job creation and submission

- Analysis of output
Computational Chemistry Landscape

Quantum Mechanics
- Gaussian 03
- Gamess
- Qumax

Thermodynamics
- NAMD
- LAMMPS
- Gromacs
- Amber
- AutoDock

Statistical Mechanics
- CPMD
- QM/MM

PINY, NWChem
Pre–req. = knowledge of quantum mechanics
Goal: Obtain a fundamental knowledge of \textit{ab initio} calculations using Gaussian
  ◦ Cannot cover the entire computational \textit{ab initio} topic

How wave functions are constructed

Levels of theory used

Focus on small molecules, gas phase
Theory

- Solutions to the Schrödinger equation
  \[(\hat{H} - E)\Psi(r_1, r_2, \ldots, r_N) = 0\]

- Ways to solve:
  - Numerical: Approx. partial derivatives of \(\Psi\) as finite differences of values of \(\Psi\)
  - Expansion Method: model \(\Psi\),
    \[\Psi(r_1, r_2, \ldots, r_N) \approx \sum_i c_i \eta_i(r_1, r_2, \ldots, r_N)\]
    - The set of \(\eta_i\) forms a basis representation of \(\Psi\)
Expansion Method
\[(\hat{H}-E)\Psi(r_1,r_2,\ldots,r_N)=0\]
\[\Psi(r_1,r_2,\ldots,r_N) \approx \sum_i c_i \eta_i(r_1,r_2,\ldots,r_N)\]
The set of \(\eta_i\) forms a basis representation of \(\Psi\)

From dictionary.com,

Basis
1. the bottom or base of anything; the part on which something stands or rests.
2. anything upon which something is based; fundamental principle; groundwork.
3. the principal constituent; fundamental ingredient.
4. a basic fact, amount, standard, etc., used in making computations, reaching conclusions, or the like: The nurse is paid on an hourly basis. He was chosen on the basis of his college grades.
5. Mathematics. a set of linearly independent elements of a given vector space having the property that every element of the space can be written as a linear combination of the elements of the set.
\[\Psi(r_1,r_2,\ldots,r_N) \approx \sum_i c_i \eta_i(r_1,r_2,\ldots,r_N)\]
Latin Origin: step, place one stands on, pedestal
Theory – Basis sets

- \( \Psi(r_1, r_2, \ldots, r_N) \approx \sum_i c_i \eta_i(r_1, r_2, \ldots, r_N) \)
- Instead of the whole system, let’s focus on a molecular orbital

\[
\varphi_i = \sum_{n=1}^{N} c_{ni} \eta_n
\]

- Increase \( \eta = \) higher quality of the wave function
- Increase \( \eta = \) higher computational cost
Slater-Type Orbitals

\[ \eta^{STO}(r) = x^l y^m z^n e^{-\zeta r} \]

Similar to angular momentum \( L = l + m + n \)

Gaussian Type Orbital

\[ \eta^{GTO}(r) = x^l y^m z^n e^{-\alpha r^2} \]

\( \zeta \) & \( \alpha \) are the orbital exponents.

The radial 'size'

A large \( \alpha \) means the electron is held tight to the nucleus.
A small \( \alpha \) means the electron can roam far from the nucleus.
STO vs. GTO

STO
- Correct cusp at \( r \to 0 \)
- Desired decay \( r \to \infty \)
- Correctly mimics H orbitals
- Natural choice

GTO
- Wrong
- Wrong
- Gaussian \( \times \) gaussian
- Mathematically & computationally easier
Contracted Gaussian Functions (CFG)

- Product of 2 gaussians is another gaussian
- Integration is easy (versus STO’s)
- Loss of accuracy BUT increase the number of gaussians used
- STO–nG, where \( n = \# \text{ GTOs used} \)

\[
\eta_{CGF}^{\tau} = \sum_{1}^{n} d_{n\tau} \eta^{GTO}_{n}
\]
\[
\phi_i = \sum_{n=1}^{N} d_{ni} \eta_n \\
\eta^{GTO}(r) = x^l y^m z^n e^{-\alpha r^2} = \phi^{GF}
\]

\[
\phi_{2s}(r) = \sum_{i=1}^{3} d_{2s} \phi_{2s}^{GF}(r, \alpha_{2s}) + d'_{2s} \phi_{2s}^{GF}(r, \alpha'_{2s}) \\
\phi_{2s}(r) = -0.395\phi_{2s}^{GF}(r, 3.664) + 1.215\phi_{2s}^{GF}(r, 0.771) \\
+ 1.00\phi_{2s}^{GF}(r, 0.195)
\]
Extended Basis Set: Split Valence

- Minimal basis set: STO–3G
  - A single CGF for each atomic orbital
- Double–Zeta (DZ) Basis Sets
  - Multiple basis functions corresponding to each atomic orbital = greater flexibility
  - Inert core orbitals: single CGF
  - Valence orbitals: double set of CGF’s

\[ \varphi_{2s}(r) = d \varphi_{2s}^{STO}(r, \zeta_1) + d \varphi_{2s}^{STO}(r, \zeta_2) \]
Extended Basis Sets (cont.)

3–21G

- Core = 1s orbital, a sum of 3 gaussians
- Each valence orbital is split into 2 parts
- Valence = 2s & 2p orbitals:
  - Inner part a sum of 2 gaussians
  - Outer part is 1 gaussian

- Triple–Zeta (TZ)
  - Inner = 1 CGF
  - Valence = 3 CGF’s
  - 6–311G
Polarization Functions:

- Distortion of molecular orbitals due to the molecular environment
- Functions with a higher angular momentum than any occupied atomic orbital
  - p–functions for H–He **
  - d–functions for Li–Ca *
  - f–functions for transition metals
  - 6–31G(d,p) = 6–31G** = polarization functions added to elements H–Ca
Diffuse Functions:
- Electrons are bound looser in anions, neutral molecules w/ unshared pairs, zwitterions
- Wave function tail is important

- Remember,
- \( \phi = \text{Constants} \times \exp[-\alpha r^2] \)
- For longer tails, this means
- \( 0.1 > \alpha > 0.01 \)
- Large radial extent

- 6–31 ′ ′G ( *’s here for polar.)
- + diffuse functions on first row
- ++ diffuse function on H
## Other Basis Sets

These are “about” equivalent

<table>
<thead>
<tr>
<th>Pople split valence</th>
<th>Pople valence triple zeta</th>
<th>Dunning correlation consistent</th>
</tr>
</thead>
<tbody>
<tr>
<td>6–31G</td>
<td>6–311G</td>
<td>cc–pvdz</td>
</tr>
<tr>
<td>6–31+G*</td>
<td>6–311+G*</td>
<td>cc–pvtz</td>
</tr>
<tr>
<td>6–31++G**</td>
<td>6–311++G**</td>
<td>cc–pvqz</td>
</tr>
</tbody>
</table>

- **Diffuse functions**
  - Long-distance interactions
  - Anions

- **Polarization functions**
  - Angular charge distribution
  - Flexibility
The variational principle.
- If $\Psi$ is the true wavefunction, then for any model antisymmetric wavefunction $\Phi$
  $$E(\Phi_{\text{model}}) > \text{true } E(\Psi)$$

- To find the HF w.f.:
  - Minimize the energy w.r.t. changes in the orbitals
    - $\Phi_i \rightarrow \phi_i + \delta \phi_i$

- Adjustable parameters, the $c_{ni}$’s, the orbital coeff’s

- $E_{\text{HF}}$ forms an upper bound, $E_{\text{exact}} - E_{\text{HF}} = E_{\text{correlation}}$
  - HF neglects electron correlation
Simplistic & typical routine
Self-consistent field method

From *.gbs

Next section

Self-consistent = Iterative, step by step

Formulation orbital combo’s
Adjust -311

Field (a.k.a. mean field) where each particle is subjected to a mean field (average) created by all the other particles

Calculate Properties

End

Yes

SCF Converged?

No

Input 3D Coordinates of atomic nuclei

Initial Guess Molecular Orbitals (1-electron vectors)

Fock Matrix Formation

Fock Matrix Diagonalization
Molecular Geometry Optimization

- Hessian matrix is the square matrix of second-order partial derivatives of a function; that is, it describes the local curvature of a function of many variables.

\[ H_{ij} = \frac{\partial^2 E}{\partial x_i \partial x_j} \]

- All H eigenvalues positive = Minimum
- All H eigenvalues negative = Maximum
- Mix of pos. & neg. eigenvalues = Saddle
Molecular Geometry Optimization Algorithms

- **Energy only**
  - Simplest
    - Might not correspond to the true minimum

- **Gradient following (1\textsuperscript{st} order)**
  - Line search, backtracking, gradient descent (steepest descent)
  - Less robust, gradient evaluation at every iteration step

- **Higher order**
  - Newton–Raphson, Berny, RFO, GDIIS, ODIIS
  - Hessians & higher order derivations at every step
  - Powerful, more time consuming, VERY dependent on initial configuration / guess
Higher orders of theory
Configuration Interaction

• HF theory is limited by its neglect of electron–electron correlation.
  – Artificially, electrons interact with a SCF, not individual $e’s$.
  – Most higher ordered levels of theory begin with HF formulation
    of the w.f.
• “Full CI” includes the Hartree–Fock ground–state determinant and
  all possible variations.

$$
\Psi = a_0 \Psi_0 + \sum_{s=1}^{\infty} a_s \Psi_s
$$

– The wavefunction becomes where $s$ includes all combinations of
  substituting electrons into HF virtual orbitals.
  • CISD(T) = Config Inter. of singlets, doublets; triplets combinations
    • The $a_s$’s are optimized; not so practical if $s$ goes to infinity!
– Partial CI’s are useful, and takes less time... ~4 months?
Levels of Theory

Semi-empirical methods
MNDO, AM1, PM3

Hartree–Fock
HF–SCF

Excitation Hierarchy
CIS, CISD, CISDT
CCS, CCSD, CCSDT

Multiconfigurational HF
MCSCF, CASSCF

Perturbation Hierarchy
MP2, MP3, MP4

Multireference Perturbation
CASPT2, CASPT3

Full CI
Pople’s Diagram
HCl Potential Energy Scan, comparison of levels of theory
Basis Set: aug-cc-pvdz

$\Delta E = \text{CCSD(T)} - xx$ (Kcal/mol)

HF: 112.931
No é corr.!!

MP2: 12.881

MP3: 2.672

1 a.u. =
1 hartree =
627.5095 Kcal/mol
Gaussian runtime

- Understand how the input file is constructed
- Specifying atomic coordinates
- How to submit your job to a HPC
Defining Atomic Coordinates

- Cartesian Coordinates
  - René Descartes
  - Atom $x y z$

- Z matrix
  - Internal Coordinate System
    - `newatom`
    - `newatom atom1 bond-length`
    - `newatom atom1 bond-length atom2 bondangle`
    - `newatom atom1 bond-length atom2 bondangle atom3 dihedral-angle`
This atom is connected to atom1 having a bond-length and with atom2 has a bondangle and with atom3 has a dihedral angle
Molden

Select Point:
- First
- Next
- Prev
- Nul

Miscellaneous:
- Dens. Mode
- Read File
- ZMAT Editor
- PostScript

Draw Mode:
- Solid
- Stick Color
- Shade
- Perspect.
- Label
- Backbone

Render Forces:
- Forces
- Dim, Scale
- THiscale

Calculate:
- Distance
- Angle
- Dihedral

Convergence:
- SCF Error
- Energy Error

Status line:
No coordinates found!
Gaussian Job Details

%chk=water.chk
%NProcLinda=2
%NProcShared=4
# mp2/6-31g* opt=tight freq

Title Line
Title
Charge Multi.
Coordinates
Blank line !

Checkpoint file
# nodes
# ppn
Job description
Title

%chk=water.chk
%NProcLinda=2
%NProcShared=4
# mp2/6-31g* opt=tight freq

Title Line

<blank line>
Gaussian Job Details

%chk=water.chk
%NProcLinda=2
%NProcShared=4
# mp2/6-31g* opt=tight freq
<blank line>
Title Line
<blank line>
0 1
O
H 1 1.0
H 1 1.0 2 104.5
<blank line>

Variables section
Defining used variables

If defined by calculation type: bond stretches, isotopes, continue (serial job processing)
Add “+gaussian-03” to your .soft file
Resoft

On IBM P5 systems
Command is
- runG03 inputfile nprocs queue
- runG03 input 8 MP5L
- (LL) ‘llclass’ for queue types
Gaussian on Dell x86 maxchines

Add the softenv key & resoft
Sample submission script, ‘qsub subscript’

```bash
#!/bin/tcsh
#PBS -A your_allocation       # specify the allocation. Change it to your allocation
#PBS -q checkpt              # the queue to be used.
#PBS -2 nodes=1:ppn=4        # Number of nodes and processors
#PBS -l walltime=1:00:00      # requested Wall-clock time.
#PBS -o g03_output           # name of the standard out file to be "output-file".
#PBS -j oe                   # standard error output merge to the standard output file.
#PBS -N g03test              # name of the job (that will appear on executing the qstat command).
set NPROCS=`wc -l $PBS_NODEFILE |gawk '{print $1}'`
setenv GAUSS_SCRDIR /scratch/$USER
set NODELIST = ( -vv -nodelist "" `cat $PBS_NODEFILE `"" -mp 4)
setenv GAUSS_LFLAGS "$NODELIST"
# cd to the directory with Your input file
cd ~USERNAME/g03test
# Change this line to reflect your input file and output file
g03 < test.inp > g03job.out
```
Fighting the Queue

- Not all simulations are equal!
- Benchmark your simulation
  - Walltime vs. # processors
  - Walltime vs. # nodes (w/ half procs idle)
  - Find the optimum configuration
  - Decrease walltime to 1 hour
- Walltime slightly greater than time needed
- THEN: Tailor your submission script to reflect YOUR needed walltime & # nodes
Fighting the Queue

- Check & monitor the load on other HPCs
- NOTE: Gaussian license is site specific

- Researchers tend to mainly work with a few clusters and submit different job types/sizes
## Analysis— some pointers

<table>
<thead>
<tr>
<th>Theory</th>
<th>Basis Set</th>
<th>Ground State Bond Length</th>
<th>% Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(hartrees)</td>
<td>(Angstroms)</td>
</tr>
<tr>
<td>HF</td>
<td>3-21G*</td>
<td>-457.981</td>
<td>1.2671</td>
</tr>
<tr>
<td>HF</td>
<td>6-31G*</td>
<td>-460.06</td>
<td>1.2662</td>
</tr>
<tr>
<td>HF</td>
<td>6-311G*</td>
<td>-460.087</td>
<td>1.2713</td>
</tr>
<tr>
<td>HF</td>
<td>cc-pvDZ</td>
<td>-460.09</td>
<td>1.2776</td>
</tr>
<tr>
<td>HF</td>
<td>cc-pvTZ</td>
<td>-460.107</td>
<td>1.2672</td>
</tr>
<tr>
<td>HF</td>
<td>aug-cc-pvDZ</td>
<td>-460.093</td>
<td>1.2767</td>
</tr>
<tr>
<td>HF</td>
<td>aug-cc-pvTZ</td>
<td>-460.108</td>
<td>1.2675</td>
</tr>
<tr>
<td>B3LYP</td>
<td>3-21G*</td>
<td>-458.709</td>
<td>1.2888</td>
</tr>
<tr>
<td>B3LYP</td>
<td>6-31G*</td>
<td>-460.796</td>
<td>1.2895</td>
</tr>
<tr>
<td>mp2</td>
<td>6-311G**</td>
<td>-460.244</td>
<td>1.2734</td>
</tr>
<tr>
<td>CCISD</td>
<td>3-21G*</td>
<td>-458.145</td>
<td>1.2841</td>
</tr>
<tr>
<td>CCISD</td>
<td>6-31G*</td>
<td>-460.218</td>
<td>1.2854</td>
</tr>
<tr>
<td>CCISD</td>
<td>cc-pVTZ</td>
<td>-460.371</td>
<td>1.2749</td>
</tr>
<tr>
<td>Experiment</td>
<td></td>
<td></td>
<td>1.2746</td>
</tr>
</tbody>
</table>

### Geometry optimization of HCl

Is CCISD/cc-pVTZ the best?

Is bond length a good judgment of accuracy of the model?
References

- http://www.gaussian.com
- http://www.gaussian.com/g_ur/keywords.htm
  - Gaussview
- http://www.cmbi.ru.nl/molden/molden.html
- http://www.cscs.ch/molekel/