

Electronic Structure Calculations in Quantum Chemistry

Alexander B. Pacheco

User Services Consultant
LSU HPC & LONI
sys-help@loni.org

HPC Training
Louisiana State University
Baton Rouge
Nov 16, 2011



- 1 Introduction
- 2 Ab Initio Methods
- 3 Density Functional Theory
- 4 Semi-empirical Methods
- 5 Basis Sets
- 6 Molecular Mechanics
- 7 Quantum Mechanics/Molecular Mechanics (QM/MM)
- 8 Computational Chemistry Programs
- 9 Exercises
- 10 Tips for Quantum Chemical Calculations



What is Computational Chemistry?

- **Computational Chemistry** is a branch of chemistry that uses computer science to assist in solving chemical problems.
- Incorporates the results of theoretical chemistry into efficient computer programs.
- Application to single molecule, groups of molecules, liquids or solids.
- Calculates the structure and properties of interest.
- Computational Chemistry Methods range from
 - 1 Highly accurate (*Ab-initio*, DFT) feasible for small systems
 - 2 Less accurate (semi-empirical)
 - 3 Very Approximate (Molecular Mechanics) large systems



Theoretical Chemistry can be broadly divided into two main categories

- 1 Static Methods \Rightarrow Time-Independent Schrödinger Equation

$$\hat{H}\Psi = E\Psi$$

- ◆ Quantum Chemical/*Ab Initio* /Electronic Structure Methods
- ◆ Molecular Mechanics

- 2 Dynamical Methods \Rightarrow Time-Dependent Schrödinger Equation

$$i\hbar \frac{\partial}{\partial t} \Psi = \hat{H}\Psi$$

- ◆ Classical Molecular Dynamics
- ◆ Semi-classical and *Ab-Initio* Molecular Dynamics



- Provide a brief introduction to Electronic Structure Calculations in Quantum Chemistry.
 - 1 Overview of Quantum Chemical methods.
 - 2 What kind of calculations can we carry out?
 - 3 What experimental properties can we study/understand?
 - 4 How to create input files?
 - 5 Tips and Tricks to run calculations?



- *Ab Initio*, meaning "from first principles", methods solve the Schrödinger equation and does not rely on empirical or experimental data.
- Beginning with fundamental and physical properties, calculate how electrons and nuclei interact.
- The Schrödinger equation can be solved exactly only for a few systems
 - ◆ Particle in a Box
 - ◆ Rigid Rotor
 - ◆ Harmonic Oscillator
 - ◆ Hydrogen Atom
- For complex systems, *Ab Initio* methods make assumptions to obtain approximate solutions to the Schrödinger equations and solve it numerically.
- "Computational Cost" of calculations increases with the accuracy of the calculation and size of the system.

What can we predict with *Ab Initio* methods?

- Molecular Geometry: Equilibrium and Transition State
- Dipole and Quadrupole Moments and polarizabilities
- Thermochemical data like Free Energy, Energy of reaction.
- Potential Energy surfaces, Barrier heights
- Reaction Rates and cross sections
- Ionization potentials (photoelectron and X-ray spectra) and Electron affinities
- Frank-Condon factors (transition probabilities, vibronic intensities)
- Vibrational Frequencies, IR and Raman Spectra and Intensities
- Rotational spectra
- NMR Spectra
- Electronic excitations and UV-VIS spectra
- Electron density maps and population analyses
- Thermodynamic quantities like partition function

Ab Initio Theory

- **Born-Oppenheimer Approximation:** Nuclei are heavier than electrons and can be considered stationary with respect to electrons. Also known as "clamped nuclei" approximations and leads to idea of potential surface
- **Slater Determinants:** Expand the many electron wave function in terms of Slater determinants.
- **Basis Sets:** Represent Slater determinants by molecular orbitals, which are linear combination of atomic-like-orbital functions i.e. basis sets

- Solve time-independent Schrödinger equation

$$\hat{H}\Psi = E\Psi$$

- For many electron system:

$$\hat{H} = \underbrace{-\frac{\hbar^2}{2} \sum_{\alpha} \frac{\nabla_{\alpha}^2}{M_{\alpha}}}_{\hat{T}_n} - \underbrace{\frac{\hbar^2}{2m_e} \sum_i \nabla_i^2}_{\hat{T}_e} + \underbrace{\sum_{\alpha > \beta} \frac{e^2 Z_{\alpha} Z_{\beta}}{4\pi \epsilon_0 R_{\alpha\beta}}}_{\hat{V}_{nn}} - \underbrace{\sum_{\alpha, i} \frac{e^2 Z_{\alpha}}{4\pi \epsilon_0 R_{\alpha i}}}_{\hat{V}_{en}} + \underbrace{\sum_{i > j} \frac{e^2}{4\pi \epsilon_0 r_{ij}}}_{\hat{V}_{ee}}$$

$\underbrace{\hspace{15em}}_{\hat{V}}$

- The wave function $\Psi(R, r)$ of the many electron molecule is a function of nuclear (R) and electronic (r) coordinates.
- Motion of nuclei and electrons are coupled.
- However, since nuclei are much heavier than electrons, the nuclei appear fixed or stationary.

- Born-Oppenheimer Approximation: Separate electronic and nuclear motion:

$$\Psi(R, r) = \psi_e(r; R)\psi_n(R)$$

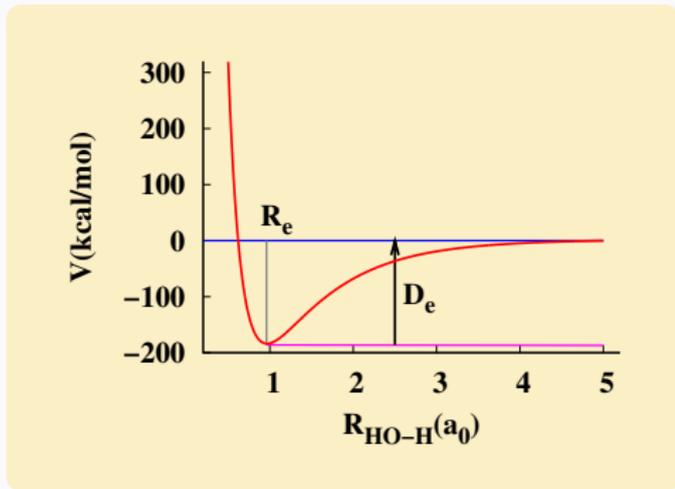
- Solve electronic part of Schrödinger equation

$$\hat{H}_e\psi_e(r; R) = E_e\psi_e(r; R)$$

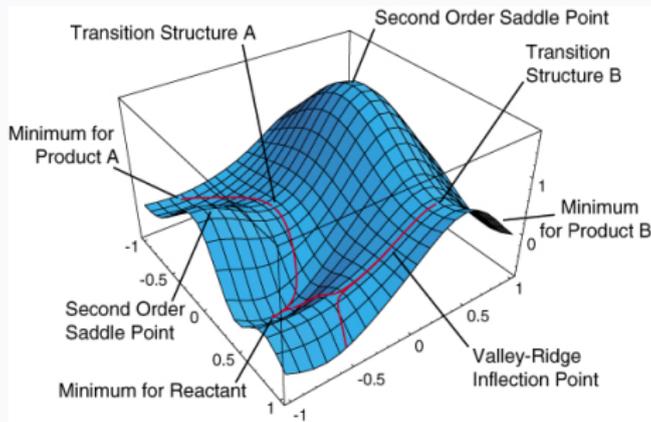
- BO approximation leads to the concept of potential energy surface

$$V(R) = E_e + V_{nn}$$

- The electronic potential is a function of nuclear coordinates.
- In Molecular Dynamics, the nuclei move along this energy surface obeying Newton's Laws of Motion.



- The potential energy surface (PES) is multi-dimensional ($3N - 6$ for non-linear molecule and $3N - 5$ for linear molecule)
- The PES contains multiple minima and maxima.
- Geometry optimization search aims to find the global minimum of the potential surface.
- Transition state or saddle point search aims to find the maximum of this potential surface, usually along the reaction coordinate of interest.



Picture taken from Bernard Schlegel's course slide at <http://www.chem.wayne.edu/~hbs/chm6440/>



- Geometry optimization is used to find minima on the potential energy surface, with these minimum energy structures representing equilibrium structures.
- Optimization also is used to locate transition structures, which are represented by saddle points on the potential energy surface.
- Optimization to minima is also referred to as energy minimization.
- During minimization, the energy of molecules is reduced by adjusting atomic coordinates.
- Energy minimization is done when using either molecular mechanics or quantum mechanics methods, and it must precede any computational analyses in which these methods are applied.
- For example, geometry optimization can be used to
 - 1 characterize a potential energy surface
 - 2 obtain a structure for a single-point quantum mechanical calculation, which provides a large set of structural and electronic properties
 - 3 prepare a structure for molecular dynamics simulation - if the forces on atoms are too large, the integration algorithm may fail.
- These energies apply to molecules in a hypothetical motionless state at 0 K. Additional information is needed to calculate enthalpies (e.g., thermal energies of translation, vibration, and rotation) and free energies (i.e., entropy).



- The electronic Hamiltonian (in atomic units, $\hbar, m_e, 4\pi\epsilon_0, e = 1$) to be solved is

$$\hat{H}_e = -\frac{1}{2} \sum_i \nabla_i^2 - \sum_{\alpha,i} \frac{Z_\alpha}{R_{i\alpha}} + \sum_{i>j} \frac{1}{r_{ij}} + \sum_{\alpha>\beta} \frac{Z_\alpha Z_\beta}{R_{\alpha\beta}}$$

- Calculate electronic wave function and energy

$$E_e = \frac{\langle \psi_e | \hat{H}_e | \psi_e \rangle}{\langle \psi_e | \psi_e \rangle}$$

- The total electronic wave function is written as a Slater Determinant of the one electron functions, i.e. molecular orbitals, MO's

$$\psi_e = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) & \phi_2(1) & \cdots & \phi_N(1) \\ \phi_1(2) & \phi_2(2) & \cdots & \phi_N(2) \\ \cdots & \cdots & \cdots & \cdots \\ \phi_1(N) & \phi_2(N) & \cdots & \phi_N(N) \end{vmatrix}$$

- MO's are written as a linear combination of one electron atomic functions or atomic orbitals (AO's)

$$\phi_i = \sum_{\mu=1}^N c_{\mu i} \chi_{\mu}$$

$c_{\mu i} \Rightarrow$ MO coefficients

$\chi_{\mu} \Rightarrow$ atomic basis functions.

- Obtain coefficients by minimizing the energy via Variational Theorem.
- Variational Theorem: Expectation value of the energy of a trial wavefunction is always greater than or equal to the true energy

$$E_e = \langle \psi_e | \hat{H}_e | \psi_e \rangle \geq \varepsilon_0$$

- Increasing $N \Rightarrow$ Higher quality of wavefunction \Rightarrow Higher computational cost



The most popular classes of ab initio electronic structure methods:

- Hartree-Fock methods
 - ◆ Hartree-Fock (HF)
 - Restricted Hartree-Fock (RHF): singlets
 - Unrestricted Hartree-Fock (UHF): higher multiplicities
 - Restricted open-shell Hartree-Fock (ROHF)
- Post Hartree-Fock methods
 - ◆ Møller-Plesset perturbation theory (MPn)
 - ◆ Configuration interaction (CI)
 - ◆ Coupled cluster (CC)
- Multi-reference methods
 - ◆ Multi-configurational self-consistent field (MCSCF)
 - ◆ Multi-reference configuration interaction (MRCI)
 - ◆ n-electron valence state perturbation theory (NEVPT)
 - ◆ Complete active space perturbation theory (CASPTn)



- 1 Wavefunction is written as a single determinant

$$\Psi = \det(\phi_1, \phi_2, \dots, \phi_N)$$

- 2 The electronic Hamiltonian can be written as

$$\hat{H} = \sum_i h(i) + \sum_{i>j} v(i,j)$$

where $h(i) = -\frac{1}{2}\nabla_i^2 - \sum_{i,\alpha} \frac{Z_\alpha}{r_{i\alpha}}$ and $v(i,j) = \frac{1}{r_{ij}}$

- 3 The electronic energy of the system is given by:

$$E = \langle \Psi | \hat{H} | \Psi \rangle$$

- 4 The resulting HF equations from minimization of energy by applying of variational theorem:

$$\hat{f}(x_1)\phi_i(x_1) = \varepsilon_i\phi_i(x_1)$$

where ε_i is the energy of orbital χ_i and the Fock operator f , is defined as

$$\hat{f}(x_1) = \hat{h}(x_1) + \sum_j [\hat{J}_j(x_1) - \hat{K}_j(x_1)]$$



- 1 $\hat{J}_j \Rightarrow$ Coulomb operator \Rightarrow average potential at x due to charge distribution from electron in orbital ϕ_i defined as

$$\hat{J}_j(x_1)\phi_i(x_1) = \left[\int \frac{\phi_j^*(x_2)\phi_j(x_2)}{r_{12}} dx_2 \right] \phi_i(x_1)$$

- 2 $\hat{K}_j \Rightarrow$ Exchange operator \Rightarrow Energy associated with exchange of electrons \Rightarrow No classical interpretation for this term.

$$\hat{K}_j(x_1)\phi_i(x_1) = \left[\int \frac{\phi_j^*(x_2)\phi_i(x_2)}{r_{12}} dx_2 \right] \phi_j(x_1)$$

- 3 The Hartree-Fock equations are solved numerically or in a space spanned by a set of basis functions (Hartree-Fock-Roothan equations)

$$\phi_i = \sum_{\mu=1}^K C_{\mu i} \tilde{\phi}_{\mu} \quad S_{\mu\nu} = \int dx_1 \tilde{\phi}_{\mu}^*(x_1) \tilde{\phi}_{\nu}(x_1)$$

$$\sum_{\nu} F_{\mu\nu} C_{\nu i} = \varepsilon_i \sum_{\nu} S_{\mu\nu} C_{\nu i} \quad F_{\mu\nu} = \int dx_1 \tilde{\phi}_{\mu}^*(x_1) \hat{f}(x_1) \tilde{\phi}_{\nu}(x_1)$$

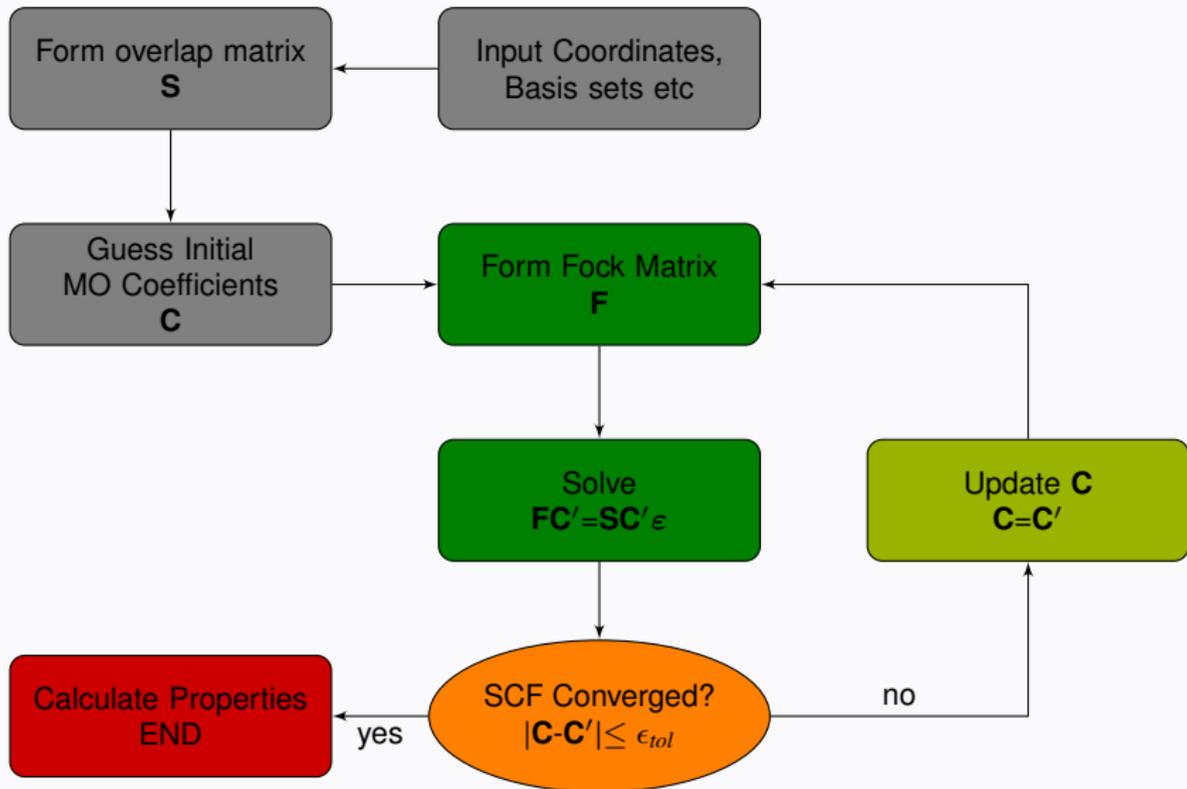
$$\mathbf{FC} = \mathbf{SC}\varepsilon$$



- 1 The Hartree-Fock-Roothan equation is a pseudo-eigenvalue equation
- 2 C 's are the expansion coefficients for each orbital expressed as a linear combination of the basis function.
- 3 Note: C depends on F which depends on $C \Rightarrow$ need to solve self-consistently.
- 4 Starting with an initial guess orbitals, the HF equations are solved iteratively or self consistently (Hence HF procedure is also known as self-consistent field or SCF approach) obtaining the best possible orbitals that minimize the energy.

SCF procedure

- 1 Specify molecule, basis functions and electronic state of interest
- 2 Form overlap matrix S
- 3 Guess initial MO coefficients C
- 4 Form Fock Matrix F
- 5 Solve $FC = SC\epsilon$
- 6 Use new MO coefficients C to build new Fock Matrix F
- 7 Repeat steps 5 and 6 until C no longer changes from one iteration to the next.





- ◆ Methods that improve the Hartree-Fock results by accounting for the correlation energy are known as **Post Hartree-Fock methods**
- ◆ The starting point for most Post HF methods is the Slater Determinant obtained from Hartree-Fock Methods.
- ◆ **Configuration Interaction (CI) methods:** Express the wavefunction as a linear combination of Slater Determinants with the coefficients obtained variationally

$$|\Psi\rangle = \sum_i c_i |\Psi_i\rangle$$

- ◆ **Many Body Perturbation Theory:** Treat the HF determinant as the zeroth order solution with the correlation energy as a perturbation to the HF equation.

$$\hat{H} = \hat{H}_0 + \lambda \hat{H}'$$

$$\varepsilon_i = E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \dots$$

$$|\Psi_i\rangle = |\Psi_i^{(0)}\rangle + \lambda |\Psi_i^{(1)}\rangle + \lambda^2 |\Psi_i^{(2)}\rangle \dots$$

- ◆ **Coupled Cluster Theory:** The wavefunction is written as an exponential ansatz

$$|\Psi\rangle = e^{\hat{T}} |\Psi_0\rangle$$

where $|\Psi_0\rangle$ is a Slater determinant obtained from HF calculations and \hat{T} is an excitation operator which when acting on $|\Psi_0\rangle$ produces a linear combination of excited Slater determinants.



Scaling Behavior	Method(s)
N^3	DFT
N^4	HF
N^5	MP2
N^6	MP3,CISD,CCSD,QCISD
N^7	MP4,CCSD(T),QCISD(T)
N^8	MP5,CISDT,CCSDT
N^9	MP6
N^{10}	MP7,CISDTQ,CCSDTQ
$N!$	Full CI

N = Number of Basis Functions



- Density Functional Theory (DFT) is an alternative to wavefunction based electronic structure methods of many-body systems such as Hartree-Fock and Post Hartree-Fock.
- In DFT, the ground state energy is expressed in terms of the total electron density.

$$\rho_0(r) = \langle \Psi_0 | \hat{\rho} | \Psi_0 \rangle$$

- We again start with Born-Oppenheimer approximation and write the electronic Hamiltonian as

$$\hat{H} = \hat{F} + \hat{V}_{ext}$$

where \hat{F} is the sum of the kinetic energy of electrons and the electron-electron interaction and \hat{V}_{ext} is some external potential.



- Modern DFT methods result from the Hohenberg-Kohn theorem
 - The external potential V_{ext} , and hence total energy is a unique functional of the electron density $\rho(r)$

$$\text{Energy} = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \equiv E[\rho]$$

- The ground state energy can be obtained variationally, the density that minimizes the total energy is the exact ground state density

$$E[\rho] > E[\rho_0], \text{ if } \rho \neq \rho_0$$

- If density is known, then the total energy is:

$$E[\rho] = T[\rho] + V_{ne}[\rho] + J[\rho] + E_{nn} + E_{xc}[\rho]$$

where

$$E_{nn}[\rho] = \sum_{A>B} \frac{Z_A Z_B}{R_{AB}} \qquad V_{ne}[\rho] = \int \rho(r) V_{ext}(r) dr$$

$$J[\rho] = \frac{1}{2} \int \frac{\rho(r_1)\rho(r_2)}{r_{12}} dr_1 dr_2$$



- If the density is known, the two unknowns in the energy expression are the kinetic energy functional $T[\rho]$ and the exchange-correlation functional $E_{xc}[\rho]$
- To calculate $T[\rho]$, Kohn and Sham introduced the concept of Kohn-Sham orbitals which are eigenvectors of the Kohn-Sham equation

$$\left(-\frac{1}{2}\nabla^2 + v_{\text{eff}}(r)\right)\phi_i(r) = \varepsilon_i\phi_i(r)$$

Here, ε_i is the orbital energy of the corresponding Kohn-Sham orbital, ϕ_i , and the density for an "N"-particle system is

$$\rho(r) = \sum_i^N |\phi_i(r)|^2$$

- The total energy of a system is

$$E[\rho] = T_s[\rho] + \int dr v_{\text{ext}}(r)\rho(r) + V_H[\rho] + E_{xc}[\rho]$$



- T_s is the Kohn-Sham kinetic energy which is expressed in terms of the Kohn-Sham orbitals as

$$T_s[\rho] = \sum_{i=1}^N \int dr \phi_i^*(r) \left(-\frac{1}{2} \nabla^2 \right) \phi_i(r)$$

v_{ext} is the external potential acting on the interacting system (at minimum, for a molecular system, the electron-nuclei interaction), V_H is the Hartree (or Coulomb) energy,

$$V_H = \frac{1}{2} \int dr dr' \frac{\rho(r)\rho(r')}{|r - r'|}$$

and E_{xc} is the exchange-correlation energy.

- The Kohn-Sham equations are found by varying the total energy expression with respect to a set of orbitals to yield the Kohn-Sham potential as

$$v_{\text{eff}}(r) = v_{\text{ext}}(r) + \int \frac{\rho(r')}{|r - r'|} dr' + \frac{\delta E_{xc}[\rho]}{\delta \rho(r)}$$

where the last term $v_{xc}(r) \equiv \frac{\delta E_{xc}[\rho]}{\delta \rho(r)}$ is the exchange-correlation potential.

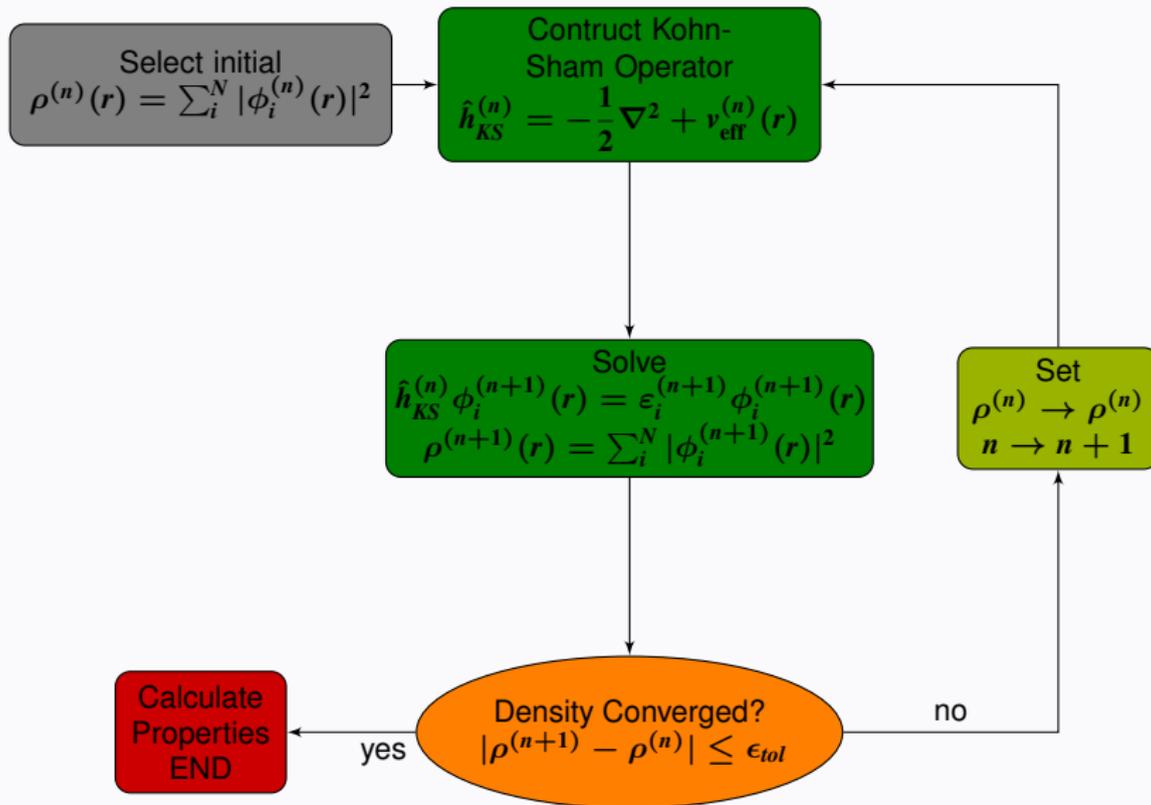


- The exchange-correlation potential, and the corresponding energy expression, are the only unknowns in the Kohn-Sham approach to density functional theory.
- There are many ways to approximate this functional E_{xc} , generally divided into two separate terms

$$E_{xc}[\rho] = E_x[\rho] + E_c[\rho]$$

where the first term is the exchange functional while the second term is the correlation functional.

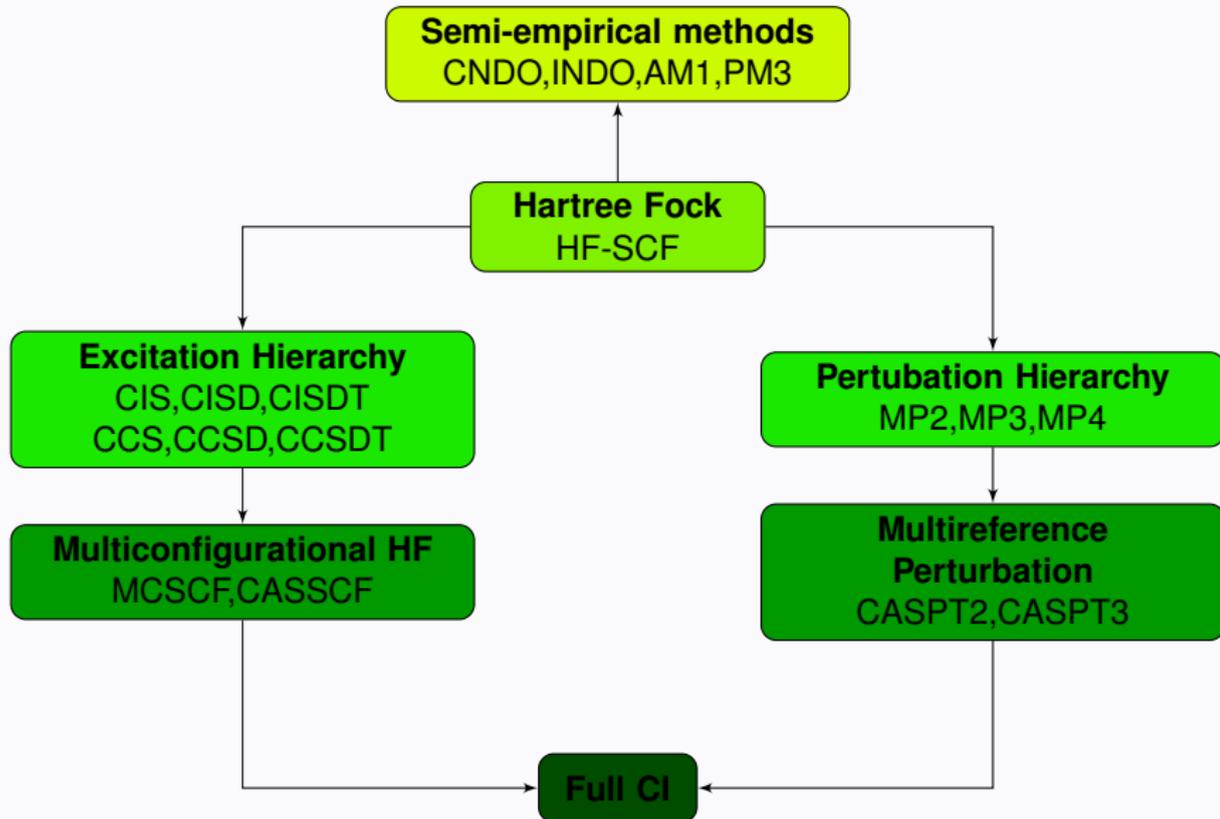
- Quite a few research groups have developed the exchange and correlation functionals which are fit to empirical data or data from explicitly correlated methods.
- Popular DFT functionals (according to a recent poll)
 - ◆ PBE0 (PBE/PBE), B3LYP, PBE, BP86, M06-2X, B2PLYP, B3PW91, B97-D, M06-L, CAM-B3LYP
 - <http://www.marcelswart.eu/dft-poll/index.html>
 - <http://www.ccl.net/cgi-bin/ccl/message-new?2011+02+16+009>





- Semi-empirical quantum methods:
 - ◆ Represents a middle road between the mostly qualitative results from molecular mechanics and the highly computationally demanding quantitative results from *ab initio* methods.
 - ◆ Address limitations of the Hartree-Fock calculations, such as speed and low accuracy, by omitting or parametrizing certain integrals
- integrals are either determined directly from experimental data or calculated from analytical formula with *ab initio* methods or from suitable parametric expressions.
- Integral approximations:
 - ◆ Complete Neglect of Differential Overlap (CNDO)
 - ◆ Intermediate Neglect of Differential Overlap (INDO)
 - ◆ Neglect of Diatomic Differential Overlap (NDDO) (Used by PM3, AM1, ...)

Semi-empirical methods are fast, very accurate when applied to molecules that are similar to those used for parametrization and are applicable to very large molecular systems.



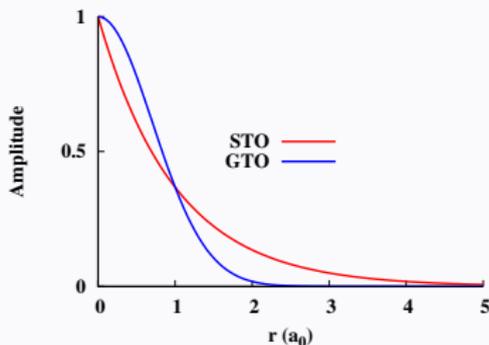


- Slater type orbital (STO) or Gaussian type orbital (GTO) to describe the AO's

$$\chi^{\text{STO}}(r) = x^l y^m z^n e^{-\zeta r}$$

$$\chi^{\text{GTO}}(r) = x^l y^m z^n e^{-\xi r^2}$$

where $L = l + m + n$ is the total angular momentum and ζ, ξ are orbital exponents.





Why STO

- Correct cusp at $r \rightarrow 0$
- Desired decay at $r \rightarrow \infty$
- Correctly mimics H orbitals
- Natural Choice for orbitals
- Computationally expensive to compute integrals and derivatives.

Why GTO

- Wrong behavior at $r \rightarrow 0$ and $r \rightarrow \infty$
- Gaussian \times Gaussian = Gaussian
- Analytical solutions for most integrals and derivatives.
- Computationally less expensive than STO's

Pople family basis set

- 1 Minimal Basis: STO-nG
 - ◆ Each atom optimized STO is fit with n GTO's
 - ◆ Minimum number of AO's needed
- 2 Split Valence Basis: 3-21G, 4-31G, 6-31G
 - ◆ Contracted GTO's optimized per atom.
 - ◆ Valence AO's represented by 2 contracted GTO's
- 3 Polarization: Add AO's with higher angular momentum (L)
 - ◆ 3-21G* or 3-21G(d), 6-31G* or 6-31G(d), 6-31G** or 6-31G(d,p)
- 4 Diffuse function: Add AO with very small exponents for systems with diffuse electron densities
 - ◆ 6-31+G*, 6-311++G(d,p)

Correlation consistent basis set

- ◆ Family of basis sets of increasing sizes.
- ◆ Can be used to extrapolate basis set limit.
- ◆ cc-pVDZ: Double Zeta(DZ) with d's on heavy atoms, p's on H
- ◆ cc-pVTZ: triple split valence with 2 sets of d's and 1 set of f's on heavy atom, 2 sets of p's and 1 set of d's on H
- ◆ cc-pVQZ, cc-pV5Z, cc-pV6Z
- ◆ can be augmented with diffuse functions: aug-cc-pVXZ (X=D,T,Q,5,6)

Pseudopotentials or Effective Core Potentials

- ◆ All Electron calculations are prohibitively expensive.
- ◆ Only valence electrons take part in bonding interaction leaving core electrons unaffected.
- ◆ Effective Core Potentials (ECP) a.k.a Pseudopotentials describe interactions between the core and valence electrons.
- ◆ Only valence electrons explicitly described using basis sets.
- ◆ Pseudopotentials commonly used
 - Los Alamos National Laboratory: LanL1MB and LanL2DZ
 - Stuttgart Dresden Pseudopotentials: SDDAll can be used.
 - Stevens/Basch/Krauss ECP's: CEP-4G, CEP-31G, CEP-121G
- ◆ Pseudopotential basis are "ALWAYS" read in pairs
 - Basis set for valence electrons
 - Parameters for core electrons

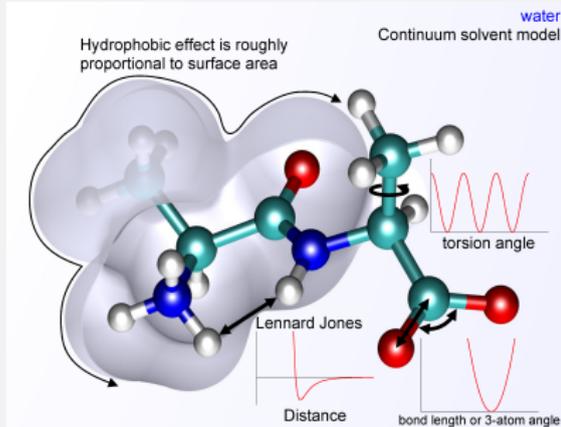


- The potential energy of all systems in molecular mechanics is calculated using force fields.
- Molecular mechanics can be used to study small molecules as well as large biological systems or material assemblies with many thousands to millions of atoms.
- All-atomistic molecular mechanics methods have the following properties:
 - ◆ Each atom is simulated as a single particle
 - ◆ Each particle is assigned a radius (typically the van der Waals radius), polarizability, and a constant net charge (generally derived from quantum calculations and/or experiment)
 - ◆ Bonded interactions are treated as "springs" with an equilibrium distance equal to the experimental or calculated bond length
- The exact functional form of the potential function, or force field, depends on the particular simulation program being used.

General form of Molecular Mechanics equations

$$\begin{aligned}
 E &= E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}} + E_{\text{vdW}} + E_{\text{elec}} \\
 &= \frac{1}{2} \sum_{\text{bonds}} K_b (b - b_0)^2 && \text{Bond} \\
 &+ \frac{1}{2} \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 && \text{Angle} \\
 &+ \frac{1}{2} \sum_{\text{dihedrals}} K_\phi [1 + \cos(n\phi)]^2 && \text{Torsion} \\
 &+ \sum_{\text{nonbonds}} \left\{ \begin{array}{l} \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \text{ van der Waals} \\ + \frac{q_1 q_2}{Dr} \text{ Electrostatics} \end{array} \right.
 \end{aligned}$$

- Commonly used force fields: AMBER, CHARMM, Drieding, UFF, TIP3P, TIP4P



Picture taken from
http://en.wikipedia.org/wiki/Molecular_mechanics

- What do we do if we want simulate chemical reaction in large systems?
- Quantum Mechanics(QM): Accurate, expensive ($\mathcal{O}(N^4)$), suitable for small systems.
- Molecular Mechanics(MM): Approximate, does not treat electrons explicitly, suitable for large systems such as enzymes and proteins, cannot simulate bond breaking/forming

- Methods that combine QM and MM are the solution.
- Such methods are called Hybrid QM/MM methods.
- The basic idea is to partition the system into two (or more) parts
 - 1 The region of chemical interest is treated using accurate QM methods eg. active site of an enzyme.
 - 2 The rest of the system is treated using MM or less accurate QM methods such as semi-empirical methods or a combination of the two.

$$\hat{H}_{\text{Total}} = \hat{H}_{\text{QM}} + \hat{H}_{\text{MM}} + \hat{H}_{\text{QM-MM}}^{\text{int}}$$



ONIOM: Divide the system into a real (full) system and the model system. Treat the model system at high and low level. The total energy of the system is given by

$$E = E(\text{low, real}) + E(\text{high, model}) - E(\text{low, model})$$

Empirical Valence Bond: Treat any point on a reaction surface as a combination of two or more valence bond structures

$$H(\mathbf{R}, \mathbf{r}) = \begin{vmatrix} H_{11}(\mathbf{R}, \mathbf{r}) & H_{12}(\mathbf{R}, \mathbf{r}) \\ H_{21}(\mathbf{R}, \mathbf{r}) & H_{22}(\mathbf{R}, \mathbf{r}) \end{vmatrix}$$

Effective Fragment Potential: Divide a large system into fragments and perform *ab initio* or DFT calculations of fragments and their dimers and including the Coulomb field from the whole system.

Software	QB	Eric	Louie	Oliver	Painter	Poseidon	Philip	Tezpur
CPMD	✓	✓	✓	✓	✓	✓		✓
GAMESS	✓	✓	✓	✓	✓	✓	✓	✓
Gaussian		✓	✓	✓	✓		✓	✓
NWCHEM	✓	✓	✓	✓	✓	✓		✓
WIEN2K	✓	✓	✓	✓	✓	✓		

Software	Bluedawg	Pelican	Pandora
CPMD	✓		✓
GAMESS			✓
Gaussian	✓	✓	✓
NWCHEM	✓	✓	✓



- Commercial Software: Q-Chem, Jaguar,CHARMM
- GPL/Free Software: ACES, ABINIT, Octopus
- http://en.wikipedia.org/wiki/Quantum_chemistry_computer_programs
- <http://www.ccl.net/chemistry/links/software/index.shtml>
- <http://www.redbrick.dcu.ie/~noel/linux4chemistry/>

Job Type	Gaussian	GAMESS	NWCHEM
	# keyword	runtyp=	task
Energy	sp	energy	energy
Force	force	gradient	gradient
Geometry optimization	opt	optimize	optimize
Transition State	opt=ts	sadpoint	saddle
Frequency	freq	hessian	frequencies, freq
Potential Energy Scan	scan	surface	✓
Excited State	✓	✓	✓
Reaction path following	irc	irc	✓
Molecular Dynamics	admp, bomd	drc	dynamics, Car-Parrinello
Population Analysis	pop	pop	✓
Electrostatic Properties	prop	✓	✓
Molecular Mechanics	✓	✓	✓
Solvation Models	✓	✓	✓
QM/MM	oniom	✓	qm3



- Site specific license
 - 1 Gaussian 03 and 09
 - LSU Users: Eric, Pandora, Pelican, Philip, Tezpur
 - Latech Users: Painter, Bluedawg
 - 2 Gaussian 03
 - ULL Users: Oliver, Zeke
 - Tulane Users: Louie, Ducky
 - Southern Users: Lacumba
 - 3 UNO Users: No License
- Add +gaussian-03/+gaussian-09 to your .soft file and resoft
- If your institution has license to both G03 and G09, have only one active at a given time.
- Only LSU has a LINDA license to run Gaussian on multiple nodes.



- Input files for GAMESS, GAUSSIAN and NWCHEM are written in free format.
- Molecule description in either Z-Matrix format or Cartesian Coordinates.
- Gaussian: Need to specify number of processors to be used in input file
`%NProcShared`

Gaussian Input: h2o-opt-freq.com

%chk=h2o-opt-freq.chk	checkpoint file
%mem=512mb	amount of memory
%NProcShared=4	number of smp processors
	blank line
#p b3lyp/6-31G opt freq	Job description
	blank line
H2O OPT FREQ B3LYP	Job Title
	blank line
0 1	Charge & Multiplicity
O	Molecule Description
H 1 r1	in Z-matrix format
H 1 r1 2 a1	with variables
	blank line
r1 1.05	variable value
a1 104.5	
	blank line



```
#!/bin/tcsh
#PBS -q checkpt
#PBS -l nodes=1:ppn=4
#PBS -l walltime=24:00:00
#PBS -o g09_testjobs.out
#PBS -j oe
#PBS -V
#PBS -N h2o-opt-freq

setenv WORK_DIR $PBS_O_WORKDIR
cd $WORK_DIR
setenv OMP_NUM_THREADS 4
set NPROCS='wc -l $PBS_NODEFILE |gawk '//print $1''

g09 < h2o-opt-freq.com > h2o-opt-freq.log
```

GAMESS Input: h2o-opt-freq.inp

\$CONTRL SCFTYP=RHF RUNTYP=OPTIMIZE

Job Control Data

COORD=ZMT NZVAR=0 \$END

\$STATPT OPTTOL=1.0E-5 HSEND=.T. \$END

Geometry Search Control

\$BASIS GBASIS=N31 NGAUSS=6

Basis Set

NDFUNC=1 NPFUNC=1 \$END

\$DATA

Molecular Data Control

H2O OPT

Job Title

Cnv 2

Molecule Symmetry group and axis

O

Molecule Description

H 1 rOH

in Z-Matrix

H 1 rOH 2 aHOH

rOH=1.05

Variables

aHOH=104.5

\$END

End Molecular Data Control

```
#!/bin/bash
#PBS -q checkpt
#PBS -l nodes=1:ppn=4
#PBS -l walltime=01:00:00
#PBS -V
#PBS -j oe
#PBS -o gamess-water.out
#PBS -N gamess-exam1

export WORKDIR=$PBS_O_WORKDIR
export NPROCS='wc -l $PBS_NODEFILE | gawk '//print $1''
export SCRDIR=/work/$USER/scr/
if [ ! -e $SCRDIR ]; then mkdir -p $SCRDIR; fi
rm -f $SCRDIR/*

cd $PBS_O_WORKDIR
rungms h2o-opt-freq 01 $NPROCS h2o-opt-freq.out $SCRDIR
```

NWCHEM Input: h2o-opt-freq.nw

```

title "H2O"                                     Job title
echo                                             echo contents of input file
charge 0                                       charge of molecule
geometry                                       geometry description in
  zmatrix                                       z-matrix format
  O
  H 1 r1
  H 1 r1 2 a1
  variables                                     variables used with values
  r1 1.05
  a1 104.5
end                                             end z-matrix block
end                                             end geometry block
basis noprint                                  basis description
  * library 6-31G
end
dft                                             dft calculation options
  XC b3lyp
  mult 1
end
task dft optimize                             job type: geometry optimization
task dft energy                               job type: energy calculation
task dft freq                                 job type: frequency calculation
  
```



```
#!/bin/bash
#PBS -q checkpt
#PBS -l nodes=1:ppn=4
#PBS -l walltime=0:30:00
#PBS -V
#PBS -o nwchem_h2o.out
#PBS -e nwchem_h2o.err
#PBS -N nwchem_h2o

export EXEC=nwchem
export EXEC_DIR=/usr/local/packages/ \
  nwchem-5.1-mvapich-1.0-intel-10.1/bin/LINUX64/
export WORK_DIR=$PBS_O_WORKDIR
export NPROCS='wc -l $PBS_NODEFILE |gawk '//print $1''

cd $WORK_DIR
mpirun_rsh -machinefile $PBS_NODEFILE -np $NPROCS \
  $EXEC_DIR/$EXEC $WORK_DIR/h2o-opt-freq.nw >& \
  $WORK_DIR/h2o-opt-freq.nwo
```



Goals

- Create Gaussian (or GAMESS/NWChem) Input files for
 - 1 Optimization
 - 2 Scan: relaxed and optimized
 - 3 Properties: Electrostatics, MOs, etc
 - 4 AIMD

Assignment

- Molecule: $[NH_3 - H - NH_3]^+$
- Method/Basis: B3LYP/6-311++G(D,P)
- Job Type: Geometry Optimization + Frequency
- Scan along the $N - H - N$ axis by moving the H , bonded to one N to the other and analyze and discuss the Potential Energy Curve.
- Population Analysis: Calculate and visualize MO's and electrostatic potential around the molecule.
- Optional: Run an AIMD simulation for at least 2ps and obtain a spectra. Compare with the harmonic spectra.



Choice of Basis Set

- STO-3G is too small.
- 6-31G* and 6-31G** give reasonable results.
- For greater accuracy, use correlation consistent basis sets e.g. cc-pVTZ
- For anions and probably excited states, use basis sets with diffuse functions (aug, +). e.g. 6-31+G*, aug-cc-pVTZ

GAMESS Basis Sets

- In GAMESS, you can create a file containing basis sets that you want to use
- Define `EXTBAS` variable which points to the basis set file
- See pseudo basis example
- In input line, if you name your basis set as `STTGRD`, then add `$BASIS`
`EXTFIL=.T. GBASIS=STTGRD $END`



Choice of Method

- Always pick DFT over HF
- In general: HF < DFT ~ MP2 < CCSD < CCSD(T)
- Pay attention to scaling behavior

SCF Convergence Issues

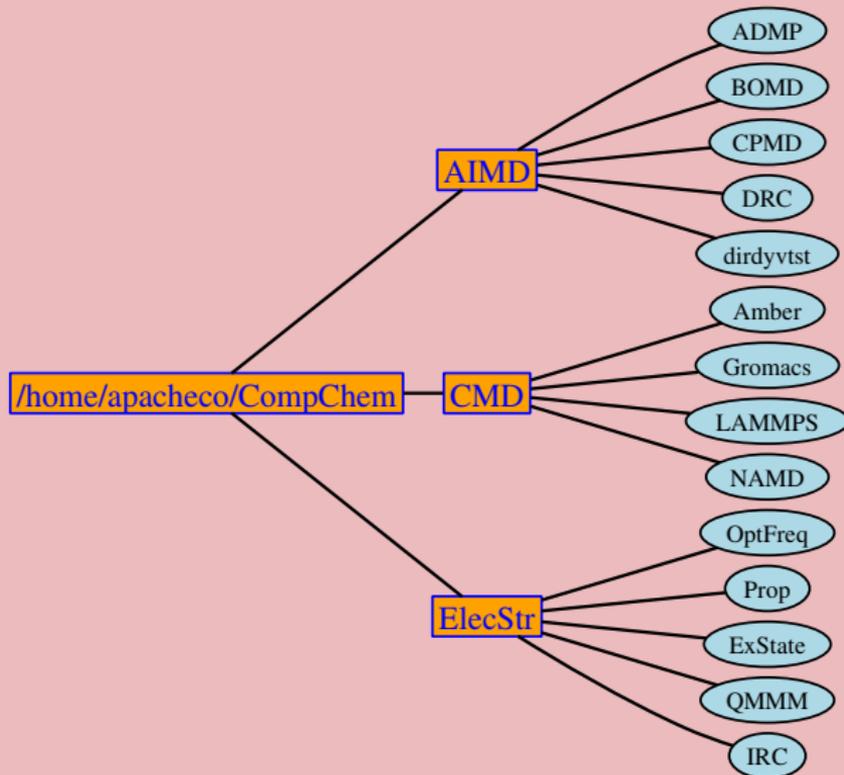
- Has SCF (HF and DFT) really converged? Important if you use iop(5/13) in Gaussian route card.
- If SCF doesn't converge:
 - 1 Increase maximum number of SCF iterations.
 - GAMESS: max 200 SCF iterations cannot be increased further.
 - 2 Use smaller basis set as an initial guess.
 - 3 Try level shifting
 - 4 Use forced convergence method:
 - Gaussian: SCF=QC, XQC or DM and item 1 above
 - GAMESS: SOSCF



Geometry Optimizations

- Many problems in computational chemistry are optimization problems: i.e., finding the "stationary points" where a multidimensional function has vanishing gradients.
- The energy as a function of nuclear coordinates. Minima, transition states may be of interest.
- Make sure that the geometry optimization actually converges.
- Run a frequency calculation to check whether the geometry is a local minima (zero imaginary frequencies) or a transition state (only one imaginary frequency)
- Tighten convergence criterion to remove unwanted imaginary frequencies.
- Having more than $3N-6$ ($3N-5$ for linear) frequencies implies that you are not at a minimum. Double check and tighten convergence if necessary.

On LONI Linux Systems





Useful Links

- GAMESS: <http://www.msg.chem.iastate.edu/gamess>
- Gaussian: <http://www.gaussian.com>
- NWCHEM: <http://www.nwchem-sw.org>
- Basis Set: <https://bse.pnl.gov/bse/portal>

Further Reading

- David Sherill's Notes at Ga Tech: <http://vergil.chemistry.gatech.edu/notes/index.html>
- David Young's Notes on CCL: <http://www.ccl.net/cca/documents/dyoung/>
- Mark Tuckerman's Notes at NYU:
<http://www.nyu.edu/classes/tuckerman/quant.mech/index.html>
- Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory, A. Szabo and N. Ostlund
- Introduction to Computational Chemistry, F. Jensen
- Essentials of Computational Chemistry - Theories and Models, C. J. Cramer
- Exploring Chemistry with Electronic Structure Methods, J. B. Foresman and A. Frisch
- ◆ Modern Electronic Structure Theory, T. Helgaker, P. Jorgensen and J. Olsen (Highly advanced text, second quantization approach to electronic structure theory)