Introduction to Computational Chemistry

Alexander B. Pacheco

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

LONI Workshop Series
Southern University, Baton Rouge
February 18, 2011
Outline

1. Introduction
2. Ab Initio Methods
3. Density Functional Theory
4. Semi-empirical Methods
5. Molecular Mechanics
6. Basis Sets
7. Molecular Dynamics
8. Computational Chemistry Programs
Outline

1. Introduction
2. Ab Initio Methods
3. Density Functional Theory
4. Semi-empirical Methods
5. Molecular Mechanics
6. Basis Sets
7. Molecular Dynamics
8. Computational Chemistry Programs
Computational Chemistry is a branch of chemistry that uses principles of computer science to assist in solving chemical problems. It uses the results of theoretical chemistry, incorporated into efficient computer programs. Application to single molecule, groups of molecules, liquids or solids. Calculates the structure and properties such as relative energies, charge distributions, dipole and multipole moments, spectroscopy, reactivity, etc.

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
What is Computational Chemistry II

Theoretical Chemistry: broadly be divided into two main categories

1. Static Methods $\Rightarrow$ Time-Independent Schrödinger Equation
   - Quantum Chemical or *Ab Initio* or Electronic Structure Methods
   - Molecular Mechanics

2. Dynamical Methods $\Rightarrow$ Time-Dependent Schrödinger Equation
   - Classical Molecular Dynamics
   - Semi-classical and *Ab-Initio* Molecular Dynamics
Outline

1. Introduction
2. Ab Initio Methods
3. Density Functional Theory
4. Semi-empirical Methods
5. Molecular Mechanics
6. Basis Sets
7. Molecular Dynamics
8. Computational Chemistry Programs
Ab Initio Methods

- *Ab Initio* meaning "from first principles" methods solve the Schrödinger equation and does not rely on empirical 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.
- "Cost" of calculations increases with the accuracy of the calculation and size of the system.

High Performance Computing @ Louisiana State University - http://www.hpc.lsu.edu February 18, 2011
Ab Initio Methods

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 Methods

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 the 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.
Born-Oppenheimer Approximation I

- Solve time-independent Schrödinger equation

\[ \hat{H} \Psi = E \Psi \]

- For many electron system:

\[
\hat{H} = -\frac{\hbar^2}{2} \sum_{\alpha} \frac{\nabla^2_{\alpha}}{M_{\alpha}} - \frac{\hbar^2}{2m_e} \sum_{i} \nabla^2_{i} + \sum_{\alpha>\beta} \frac{e^2 Z_{\alpha} Z_{\beta}}{4\pi\epsilon_0 R_{\alpha \beta}} - \sum_{\alpha,i} \frac{e^2 Z_{\alpha}}{4\pi\epsilon_0 R_{\alpha i}} + \sum_{i>j} \frac{e^2}{4\pi\epsilon_0 r_{ij}}
\]

\[ \hat{H} = \hat{T}_n + \hat{T}_e + \hat{V}_{nn} + \hat{V}_{en} + \hat{V}_{ee} \]

- 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 II

- 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 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.
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)

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

- \( c_{\mu i} \) ⇒ MO coefficients
- \( \chi_\mu \) ⇒ atomic basis functions.

Obtain coefficients by minimizing the energy via Variational Theorem.

Variational Theorem: Expectation value of the energy 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 \( N \) ⇒ Higher quality of wavefunction ⇒ Higher computational cost
Ab Initio Methods

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)
Hartree-Fock

1. Wavefunction is written as a single determinant
\[ \Psi = det(\phi_1, \phi_2, \cdots \phi_N) \]

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:
\[ \hat{f}(x_1) \chi_i(x_1) = \varepsilon_i \chi_i(x_1) \]

where \( \varepsilon_i \) is the energy of orbital \( \chi_i \) and the Fock operator \( f \), is defined as
\[ \hat{f}(x_1) = \hat{h}(x_1) + \sum_j \left[ \hat{J}_j(x_1) - \hat{K}_j(x_1) \right] \]
Hartree-Fock

1. \( \hat{J}_j \Rightarrow \) Coulomb operator \( \Rightarrow \) average potential at \( x \) due to charge distribution from electron in orbital \( \chi_i \) defined as

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

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

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

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

\[
\chi_i = \sum_{\mu=1}^{K} C_{\mu i} \tilde{\chi}_\mu
\]

\[
\sum_{\nu} F_{\mu \nu} C_{\nu i} = \varepsilon_i \sum_{\nu} S_{\mu \nu} C_{\nu i}
\]

\[
\text{FC} = SC\varepsilon
\]
The Hartree-Fock-Roothan equation is a pseudo-eigenvalue equation. 

C’s are the expansion coefficients for each orbital expressed as a linear combination of the basis function.

Note: \( C \) depends on \( F \) which depends on \( C \) \( \Rightarrow \) need to solve self-consistently.

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.
1. In Hartree-Fock theory, electron motions of independent of each other i.e. uncorrelated.

2. However, this is not true. For two electrons with same spin $| \Psi_1(r_1) \alpha(\omega_1) \Psi_2(r_2) \alpha(\omega_2) \rangle$, the probability of finding electron 1 at $r_1$ and electron 2 at $r_2$

$$ P(r_1, r_2) dr_1 dr_2 = \frac{1}{2} \left( |\Psi_1(r_1)|^2 |\Psi_2(r_2)|^2 + |\Psi_1(r_2)|^2 |\Psi_2(r_1)|^2 ight. $$

$$ \left. - [\Psi_1^*(r_1) \Psi_2(r_1) \Psi_2^*(r_2) \Psi_1(r_2)] + [\Psi_2^*(r_1) \Psi_1(r_1) \Psi_1^*(r_2) \Psi_2(r_2)] \right) dr_1 dr_2 $$

Now $P(r_1, r_1) = 0 \Rightarrow$ No two electrons with same spins can be at the same place $\Rightarrow$ "Fermi hole"

3. Same-spin electrons are correlated while different spin electrons are not.

4. Energy difference between HF energy and the true energy is the correlation energy

$$ E_{corr} = E_0 - E_{HF} $$
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.

\[ \hat{H} = \hat{H}_0 + \lambda \hat{H}'; \varepsilon_i = E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \cdots \]

\[ |\Psi_i\rangle = |\Psi_{i}^{(0)}\rangle + \lambda |\Psi_{i}^{(1)}\rangle + \lambda^2 |\Psi_{i}^{(2)}\rangle \cdots \]

**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.
<table>
<thead>
<tr>
<th>Scaling Behavior</th>
<th>Method(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N^4</td>
<td>HF</td>
</tr>
<tr>
<td>N^5</td>
<td>MP2</td>
</tr>
<tr>
<td>N^6</td>
<td>MP3, CISD, CCSD, QCISD</td>
</tr>
<tr>
<td>N^7</td>
<td>MP4, CCSD(T), QCISD(T)</td>
</tr>
<tr>
<td>N^8</td>
<td>MP5, CISDT, CCSDT</td>
</tr>
<tr>
<td>N^9</td>
<td>MP6</td>
</tr>
<tr>
<td>N^10</td>
<td>MP7, CISDTQ, CCSDTQ</td>
</tr>
</tbody>
</table>
1. Introduction
2. Ab Initio Methods
3. Density Functional Theory
4. Semi-empirical Methods
5. Molecular Mechanics
6. Basis Sets
7. Molecular Dynamics
8. Computational Chemistry Programs
Density Functional Theory

- 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

\[
\hat{H} = \hat{F} + \hat{V}_{\text{ext}}
\]

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

- Modern DFT methods result from the Hohenberg-Kohn theorem

1. The external potential $V_{\text{ext}}$, and hence total energy is a unique functional of the electron density $\rho(r)$

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

2. 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:

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

where

$$E_{\text{nn}}[\rho] = \sum_{A>B} \frac{Z_A Z_B}{R_{AB}}$$

$$V_{\text{ne}}[\rho] = \int \rho(r) V_{\text{ext}}(r) dr$$

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

- 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, $\varepsilon_i$ is the orbital energy of the corresponding Kohn-Sham orbital, $\phi_i$, and the density for an "N"-particle system is

$$\rho(r) = \sum_{i=1}^{N} |\phi_i(r)|^2$$

- The total energy of a system is

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

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

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

- $v_{\text{ext}}$ 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.
Density Functional Theory

- 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.

- Popular DFT functionals (according to recent poll)

  - PBE0 (PBEPBE), B3LYP, PBE, BP86, M06-2X, B2PLYP, B3PW91, B97-D, M06-L, CAM-B3LYP


  - http://www.ccl.net/cgi-bin/ccl/message-new?2011+02+16+009
Semi-empirical Methods

- 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.
- Integral approximations:
  - Complete Neglect of Differential Overlap (CNDO)
  - Intermediate Neglect of Differential Overlap (INDO)
  - Neglect of Diatomic Differential Overlap (NDDO) (Used by PM3, AM1, ...)
- Integrals are either determined directly from experimental data or calculated from analytical formula with *ab initio* methods or from suitable parametric expressions.

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.
Heirarchy of Methods

Semi-empirical methods
CNDO, INDO, 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
Outline

1. Introduction
2. Ab Initio Methods
3. Density Functional Theory
4. Semi-empirical Methods
5. Molecular Mechanics
6. Basis Sets
7. Molecular Dynamics
8. Computational Chemistry Programs
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.
Generally the bond and angle terms are modeled as harmonic potentials centered around equilibrium bond-length values derived from experiment or *ab initio* calculations.

Morse potential can be used for an accurate reproduction of vibrational spectra but at a higher computational cost.

The dihedral or torsional terms typically have multiple minima and thus cannot be modeled as harmonic oscillators.

The non-bonded terms are much more computationally costly to calculate in full.

Modelled using a short range van der Waals interactions usually Lennard-Jones potential and a long range or electrostatic interaction which has a functional form of the Coulomb potential.

Generally a cutoff radius is used to speed up the calculation so that atom pairs whose distances are greater than the cutoff have a van der Waals interaction energy of zero.
General form of Molecular Mechanics equations

\[ E = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{vdW}} + E_{\text{elec}} \]

\[ = \frac{1}{2} \sum_{\text{bonds}} K_b (b - b_0)^2 \]

\[ + \frac{1}{2} \sum_{\text{angles}} K_{\theta} (\theta - \theta_0)^2 \]

\[ + \frac{1}{2} \sum_{\text{dihedrals}} K_{\phi} [1 + \cos(n\phi)]^2 \]

\[ + \sum_{\text{non-bonds}} \left\{ \frac{12}{r} - \frac{6}{r^6} \right\} \]
Outline

1. Introduction
2. Ab Initio Methods
3. Density Functional Theory
4. Semi-empirical Methods
5. Molecular Mechanics
6. Basis Sets
7. Molecular Dynamics
8. Computational Chemistry Programs
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 \( \zeta, \xi \) are orbital exponents.
### Basis Sets II

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

#### Why GTO
- Wrong behavior at $r \to 0$ and $r \to \infty$
- Gaussian × Gaussian = Gaussian
- Analytical solutions for most integrals and derivatives.
- Computationally less expensive than STO’s
## Pople family basis set

### Minimal Basis: STO-nG
- Each atom optimized STO is fit with n GTO's
- Minimum number of AO's needed

### Split Valence Basis: 3-21G, 4-31G, 6-31G
- Contracted GTO's optimized per atom.
- Valence AO's represented by 2 contracted GTO's

### 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)

### 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 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
Outline

1. Introduction
2. Ab Initio Methods
3. Density Functional Theory
4. Semi-empirical Methods
5. Molecular Mechanics
6. Basis Sets
7. Molecular Dynamics
8. Computational Chemistry Programs
Obtained from solutions of the time-dependent Schrödinger equations.

Classical Molecular Dynamics: Using "predefined potentials" usually force fields from Molecular Mechanics to propagate the nuclei which are treated as classical particles obeying Newton’s Laws of motions.

Ab Initio Molecular Dynamics: also known as First Principles/Direct/Quantum MD. Here the forces acting on the nuclei are computed from electronic structure calculations "on-the-fly" as the molecular dynamics trajectory is generated. Nuclei are treated classical with Newton’s Laws of motion for nuclei are derived from the TDSE.

Quantum Dynamics: Full Quantum treatment of all particles. Here the nuclei are treated as wavepackets whose motion is coupled to that of electrons obtained from electronic structure calculations.

Cost and Accuracy of calculations:
Classical MD > AIMD > Quantum Dynamics
The full Hamiltonian for a molecular system is

\[ \hat{H} = -\sum_{i} \frac{\hbar^2}{2M_i} \nabla_i^2 + \frac{-\hbar^2}{2m_e} \nabla_i^2 + V_{n-e}(\mathbf{r}, \mathbf{R}) \]

\[ \hat{H} = H_e(\mathbf{r}, \mathbf{R}) \]

Separate Electronic and Nuclear degrees of freedom

\[ \Phi(\mathbf{r}, \mathbf{R}, t) \approx \Psi(\mathbf{r}, t) \chi(\mathbf{R}, t) \exp \left[ \frac{i}{\hbar} \int_{t_0}^{t} dt' \tilde{E}_e(t') \right] \]

where \( \Psi \) and \( \chi \) are normalized separately and the last term is a phase factor

\[ \tilde{E}_e = \int d\mathbf{r} d\mathbf{R} \Psi^*(\mathbf{r}, t) \chi^*(\mathbf{R}, t) H_e \Psi(\mathbf{r}, t) \chi(\mathbf{R}, t) \]
Obtain coupled equations for electronic and nuclear wavefunction:

**Time-Dependent Self-Consistent Field (TDSCF) method**

\[ i\hbar \frac{\partial \Psi}{\partial t} = - \sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 \Psi + \left\{ \int dR \chi^* (R, t) V_{n-e}(r, R) \chi(R, t) \right\} \Psi \]

\[ i\hbar \frac{\partial \chi}{\partial t} = - \sum_I \frac{\hbar^2}{2M_I} \nabla_I^2 \chi + \left\{ \int dr \Psi^* (r, t) H_e(r, R)^- (r, t) \right\} \chi \]

Define nuclear wavefunction as

\[ \chi(R, t) = A(R, t) \exp \left[ iS(R, t)/\hbar \right] \]

where \( A \) and \( S \) are real.

Solve the time-dependent equation for nuclear wavefunction and take classical limit \( (\hbar \to 0) \) to obtain the Hamilton-Jacobi equation.

\[ \frac{\partial S}{\partial t} + \sum_I \frac{\hbar^2}{2M_I} (\nabla_I S)^2 + \int dr \Psi^* H_e \Psi = 0 \]
S is the mechanical action which is related to the momentum as
\[ P_I \equiv \nabla_I S \]

We can obtain the Newtonian equation of motion as
\[ M I \ddot{R}_I = \frac{dP_I}{dt} = -\nabla_I \int dR \Psi^* H_e \Psi = -\nabla_I V_e^E (R, t) \]

Thus, the nuclei move according to classical mechanics in an effective potential \( V_e^E \) due to the electrons. \(^1\)

For the electronic wavefunction in the TDSCF equations, take classical limit by replacing the nuclear wavefunctions by delta functions
\[ \prod_I \delta (R_I - R_I(t)) \] centered on the instantaneous positions \( R_I(t) \) which results in a time-dependent equation for electrons
\[ i\hbar \frac{\partial \Psi}{\partial t} = -\sum_i -\frac{\hbar^2}{2m_e} \nabla_i^2 \Psi + V_{n-e}(r, R(t)) = H_e(r, R(t)) \Psi(r; R; t) \]
This approach which incorporate feedback between the classical and quantal degrees of freedom in both directions is called the "Ehrenfest Molecular Dynamics".

If the electronic wavefunction is described in terms of electronic states or determinants $\Psi_k$

$$\Psi(r, R; t) = \sum_{k=0}^{\infty} c_k(t) \Psi_k(r; R)$$

then transition between electronic states are included. Useful for describing **Non-Adiabatic Dynamics, Charge Transfer dynamics, Electron Transfer. Trajectory Surface Hopping** is a derivative of this method designed to address the drawbacks of Ehrenfest dynamics.
If the choice of basis set \( \{ \Psi_k \} \) is the adiabatic basis obtained from solving the time-independent electronic Schrödinger equation and we consider only the ground state wavefunction \( \Psi_0 \), then nuclei move on a single potential energy surface

\[
H_e \Psi_0 = E_0 \Psi_0; \quad V^E_e = \int d\mathbf{r} \Psi_0^* H_e \Psi_0
\]

In this limit the Ehrenfest potential is identical to the ground state Born-Oppenheimer potential and the method is known as \textbf{Born-Oppenheimer Molecular Dynamics}. 
If the Ehrenfest potential \( V_e^E \) is approximated to a global potential surface

\[
V_e^E \approx V_e^{\text{approx}} (R_I) = \sum_{I=1}^{N} v_1(R_I) + \sum_{I>J}^{N} v_2(R_I, R_J) + \sum_{I>J>K}^{N} v_3(R_I, R_J, R_K) + \cdots
\]

in terms of a truncated expansion of many-body contributions, then the electronic degrees of freedom are replaced by interaction potentials \( \{v_n\} \).

Thus the problem is reduced to purely classical mechanics once the \( \{v_n\} \) are determined usually Molecular Mechanics Force Fields. This is most commonly known as Classical Molecular Dynamics.

---

Electronic energy obtained from

- Molecular Mechanics $\Rightarrow$ Classical Molecular Dynamics
  1. LAMMPS
  2. NAMD
  3. Amber
  4. Gromacs

Ab-Initio Methods $\Rightarrow$ Quantum or Ab-Initio Molecular Dynamics

1. Born-Oppenheimer Molecular Dynamics: Gaussian, GAMESS
2. Extended Lagrangian Molecular Dynamics: VASP, CPMD, Gaussian (ADMP), NWCHEM(CPMD), QChem (curvy-steps ELMD)
3. Time Dependent Hartree-Fock and Time Dependent Density Functional Theory: Gaussian, GAMESS, NWCHEM, QChem
4. Multiconfiguration Time Dependent Hartree(-Fock), MCTDH(F)
5. Non-Adiabatic and Ehrenfest Molecular Dynamics, Multiple Spawning, Trajectory Surface Hopping
6. Wavepacket Methods: Gaussian (QWAIMD)
### Classical Molecular Dynamics

**Advantages**
1. Large Biological Systems
2. Long time dynamics

**Disadvantages**
1. Cannot describe Quantum Nuclear Effects

### Ab Initio and Quantum Dynamics

**Advantages**
1. Quantum Nuclear Effects

**Disadvantages**
1. $\sim 100$ atoms
2. Full Quantum Dynamics ie treating nuclei quantum mechanically: less than 10 atoms
3. Picosecond dynamics at best
Outline

1. Introduction
2. Ab Initio Methods
3. Density Functional Theory
4. Semi-empirical Methods
5. Molecular Mechanics
6. Basis Sets
7. Molecular Dynamics
8. Computational Chemistry Programs
Computational Chemistry Programs

Software: Systems Installed

- **AMBER**: x86 clusters, Ducky, Lacumba, LSU HPC
- **Desmond**: QueenBee
- **DL_POLY**: x86 clusters
- **Gromacs**: x86 clusters, P5 clusters, Philip
- **LAMMPS**: x86 clusters, P5 clusters, LSU HPC
- **NAMD**: x86 clusters, P5 clusters except Lacumba, Tezpur, Pelican
- **OpenEye**: x86 clusters
- **CPMD**: x86 clusters, P5 clusters, Tezpur
- **GAMESS**: x86 clusters
- **Gaussian**: x86 clusters except QueenBee and Poseidon, P5 Clusters except Neptune, LSU HPC
- **NWCHEM**: x86 clusters, P5 clusters, LSU HPC except Philip
- **Piny_MD**: x86 clusters, P5 clusters, Pelican
Computational Chemistry Programs

- 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/
Using Gaussian on LONI Systems

- **Site specific license**
  1. **Gaussian 03 and 09**
     - LSU Users: Eric
     - 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.
Example Job submission script on Intel x86

#!/bin/tcsh
#PBS -A your_allocation
# specify the allocation. Change it to your allocation
#PBS -q checkpt
# the queue to be used.
#PBS -l 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).

# setup g03 variables
source $g03root/g03/bsd/g03.login
set NPROCS=`wc -l $PBS_NODEFILE |gawk '//{print $1}'`
setenv GAUSS_SCRDIR /scratch/$USER
# cd to the directory with Your input file
cd ~apacheco/g03test
# Change this line to reflect your input file and output file
g03 < g03job.inp > g03job.out

Linda Acess

set NODELIST = ( -vv -nodelist '"' 'cat $PBS_NODEFILE '"' "" -mp 4)
setenv GAUSS_LFLAGS " $NODELIST "
g03l < g03job.inp > g03job.out
#!/bin/tcsh
# @ account_no = your_allocation
# @ requirements = (Arch == "Power5")
# @ environment = LL_JOB=TRUE ; MP_PULSE=1200
# @ job_type = serial
# @ node_usage = shared
# @ wall_clock_limit = 12:00:00
# @ initialdir = /home/apacheco/g03test
# @ class = checkpt
# @ error = g03_$ (jobid).err
# @ queue

# setup g03 variables
source $g03root/g03/bsd/g03.login
# setup and create Gaussian scratch directory
setenv GAUSS_SCRDIR /scratch/default/$USER
mkdir -p $GAUSS_SCRDIR
# cd to the directory with Your input file
cd ~apacheco/g03test
# Change this line to reflect your input file and output file
g03 < g03job.inp > g03job.out
Sample Input

%chk=h2o-opt-freq.chk
%mem=512mb
%nProcShared=4

#p b3lyp/6-31G opt freq

H2O OPT FREQ B3LYP

0 1
O
H 1 r1
H 1 r1 2 a1

r1 1.05
a1 104.5

Input Description

checkpoint file
amount of memory
number of smp processors
blank line
Job description
blank line
Job Title
blank line
Charge Multiplicity
Molecule Description
  Z-matrix format
    with variables
blank line
  variable value
blank line
Using GAMESS on LONI Systems

- Add `+gamess-12Jan2009R1-intel-11.1` (on Queenbee) to your `.soft` and `resoft`.

### Job submission script

```bash
#!/bin/bash
#PBS -A your_allocation
#PBS -q checkpt
#PBS -l nodes=1:ppn=4
#PBS -l walltime=00:10:00
#PBS -j oe
#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 $WORKDIR
rungms h2o-opt-freq 01 $NPROCS h2o-opt-freq.out $SCRDIR
cp -p $SCRDIR/$OUTPUT $WORKDIR/
```
Sample Input

```
$CONTRL SCFTYP=RHF RUNTYP=OPTIMIZE
   COORD=ZMT NZVAR=0 $END
$STATPT OPTTOL=1.0E-5 HSSEND=.T. $END
$BASIS GBASIS=N31 NGAUSS=6
   NDFUNC=1 NPFUNC=1 $END
$DATA
H2O OPT
Cnv 2

O
H 1 rOH
H 1 rOH 2 aHOH

rOH=1.05
aHOH=104.5
$END
```

Input Description

```
Job control data

geometry search control
6-31G** basis set

molecular data control
Title
Symmetry group and axis

molecule description in z-matrix

variables

end molecular data control
```
Using NWChem on LONI Systems

- Add `+nwchem-5.1.1-intel-11.1-mvapich-1.1` (on Queenbee) to your `.soft` and `resoft`

---

Job submission script

```bash
#!/bin/sh
#
#PBS -q checkpt
#PBS -M apacheco@cct.lsu.edu
#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 >& $WORK_DIR/h2o-opt-freq.nwo
```
Sample Input

```bash
title "H2O"

echo

charge 0

geometry
zmatrix
 O
 H 1 r1
 H 1 r1 2 a1
variables
 r1 1.05
  a1 104.5
end
end

basis noprint
 * library 6-31G
end

dft
 Xc b3lyp
  mult 1
end
task dft optimize
task dft energy

task dft freq
```
<table>
<thead>
<tr>
<th>Job Type</th>
<th>Gaussian # keyword</th>
<th>GAMESS runtyp=</th>
<th>NWChem task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>sp</td>
<td>energy</td>
<td>energy</td>
</tr>
<tr>
<td>Force</td>
<td>force</td>
<td>gradient</td>
<td>gradient</td>
</tr>
<tr>
<td>Geometry optimization</td>
<td>opt</td>
<td>optimize</td>
<td>optimize</td>
</tr>
<tr>
<td>Transition State</td>
<td>opt=ts</td>
<td>sadpoint</td>
<td>saddle</td>
</tr>
<tr>
<td>Frequency</td>
<td>freq</td>
<td>hessian</td>
<td>frequencies, freq</td>
</tr>
<tr>
<td>Potential Energy Scan</td>
<td>scan</td>
<td>surface</td>
<td>✓</td>
</tr>
<tr>
<td>Excited State</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Reaction path following</td>
<td>irc</td>
<td>irc</td>
<td>✓</td>
</tr>
<tr>
<td>Dynamics</td>
<td>admp, bomd</td>
<td>drc</td>
<td>✓</td>
</tr>
<tr>
<td>Population Analysis</td>
<td>pop</td>
<td>pop</td>
<td>✓</td>
</tr>
<tr>
<td>Electrostatic Properties</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Molecular Mechanics</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Solvation Models</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

✓ → method exists and keyword requires more than one option
Population Analysis in NWCHEM requires a PROPERTY ... END block with various options for different properties

```plaintext
property
dipole
mulliken
end
```

Frequency calculations in GAMESS at end of optimization is carried out by adding HSSEND=.T. keyword in the STATPT control line

```plaintext
$STATPT HSSEND=.T. $END
```

- Excited State Calculations include TDHF, TDDFT, CIS, CC methods

**QM/MM Methods**

- Gaussian: ONIOM
- GAMESS: Effective Fragment Potential, $EFRAG block
- NWCHEM: task qmmm
Dynamics Calculations:

- **Gaussian:**
  - BOMD: Born-Oppenheimer Molecular Dynamics
  - ADMP: Atom centered Density Matrix Propagation (an extended Lagrangian Molecular Dynamics similar to CPMD) and ground state BOMD

- **GAMESS:**
  - DRC: Direct Dynamics, a classical trajectory method based on "on-the-fly" ab-initio or semi-empirical potential energy surfaces

- **NWCHEM:**
  - Car-Parrinello: Car Parrinello Molecular Dynamics (CPMD)
  - DIRDYVTST: Direct Dynamics Calculations using POLYRATE with electronic structure from NWCHEM
Related HPC Tutorials

- **Fall Semester**
  - Introduction to Gaussian/Electronic Structure Methods

- **Spring Semester**
  - MD: Programming to Production
    - April 6th
  - Introduction to CPMD/Ab Initio Molecular Dynamics
    - April 27th
Useful Links

- Amber: http://ambermd.org
- Desmond: http://www.deshawresearch.com/resources_desmond.html
- DL_POLY: http://www.cse.scitech.ac.uk/ccg/software/DL_POLY
- Gromacs: http://www.gromacs.org
- NAMD: http://www.ks.uiuc.edu/Research/namd
- CPMD: http://www.cpmd.org
- GAMESS: http://www.msg.chem.iastate.edu/gamess
- Gaussian: http://www.gaussian.com
- NWChem: http://www.nwchem-sw.org
- PINY_MD: http://homepages.nyu.edu/~mt33/PINY_MD/PINY.html
- 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
- 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
- Molecular Modeling - Principles and Applications, A. R. Leach
- Computer Simulation of Liquids, M. P. Allen and D. J. Tildesley
- Modern Electronic Structure Theory, T. Helgaker, P. Jorgensen and J. Olsen (Highly advanced text, second quantization approach to electronic structure theory)