Introduction to Computational Chemistry Packages

Alexander B. Pacheco

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

LONI HPC Workshop
University of Louisiana - Lafayette
Lafayette
Nov 9, 2011

Outline

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 $\Longrightarrow$ Time-Independent Schrödinger Equation

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

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

2. Dynamical Methods $\Longrightarrow$ Time-Dependent Schrödinger Equation

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

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

- *Ab Initio*, meaning "from first principles", methods solve the Schrödinger equation and does not rely on empirical or experimental data.
- Begining 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 Electronic Structure 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 know as "clamped nuclei" approximations and leads to idea of potential surface.

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

- Slater Determinants: Expand the many electron wave function in terms of Slater determinants.

$$\psi_e(\mathbf{r}; \mathbf{R}) = (N!)^{-1/2}\mathcal{A}\prod_{i=1}^{N}\phi_i(\mathbf{r}; \mathbf{R})$$

- Basis Sets: Represent Slater determinants by molecular orbitals, which are linear combination of atomic-like-orbital functions i.e. basis sets

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

- Variational Theorem: Expectation value of the energy of a trial wavefunction is always greater than or equal to the true energy

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

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:

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

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

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

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

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

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

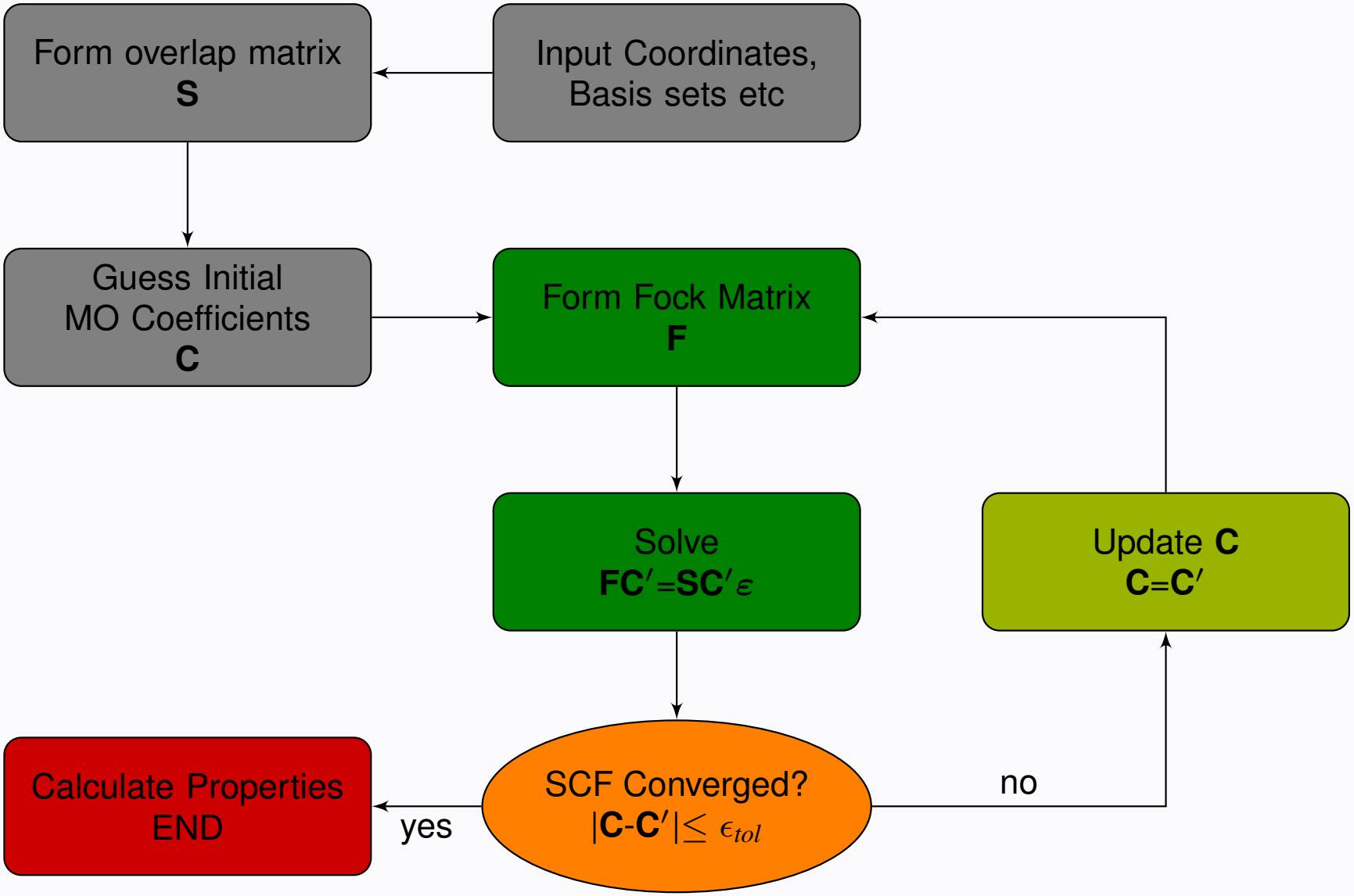
$$\mathbf{FC} = \mathbf{SC}\varepsilon$$

1. The Hartree-Fock-Roothan equation is a pseudo-eigenvalue equation
2. $\mathbf{C}$'s are the expansion coefficients for each orbital expressed as a linear combination of the basis function.
3. Note: $\mathbf{C}$ depends on $\mathbf{F}$ which depends on $\mathbf{C} \Rightarrow$ 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 $\mathbf{S}$
3. Guess initial MO coefficients $\mathbf{C}$
4. Form Fock Matrix $\mathbf{F}$
5. Solve $\mathbf{FC} = \mathbf{SC}\varepsilon$
6. Use new MO coefficients $\mathbf{C}$ to build new Fock Matrix $\mathbf{F}$
7. Repeat steps 5 and 6 until $\mathbf{C}$ no longer changes from one iteration to the next.

SCF Flow Chart



Post Hartree-Fock Methods

- 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 obtain from Hartree-Fock Methods.
- **Configuration Interaction (CI) methods**: Express the wavefunction as a linear combination of Slater Determinants with the coeffcients 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 pertubation 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.

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.

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

Hohenberg-Kohn theorem

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

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

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:

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

where

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

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

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

and two unknowns, 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

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

DFT

- The total energy of a system is

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

- T_s is the Kohn-Sham kinetic energy

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

- 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,

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

  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

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

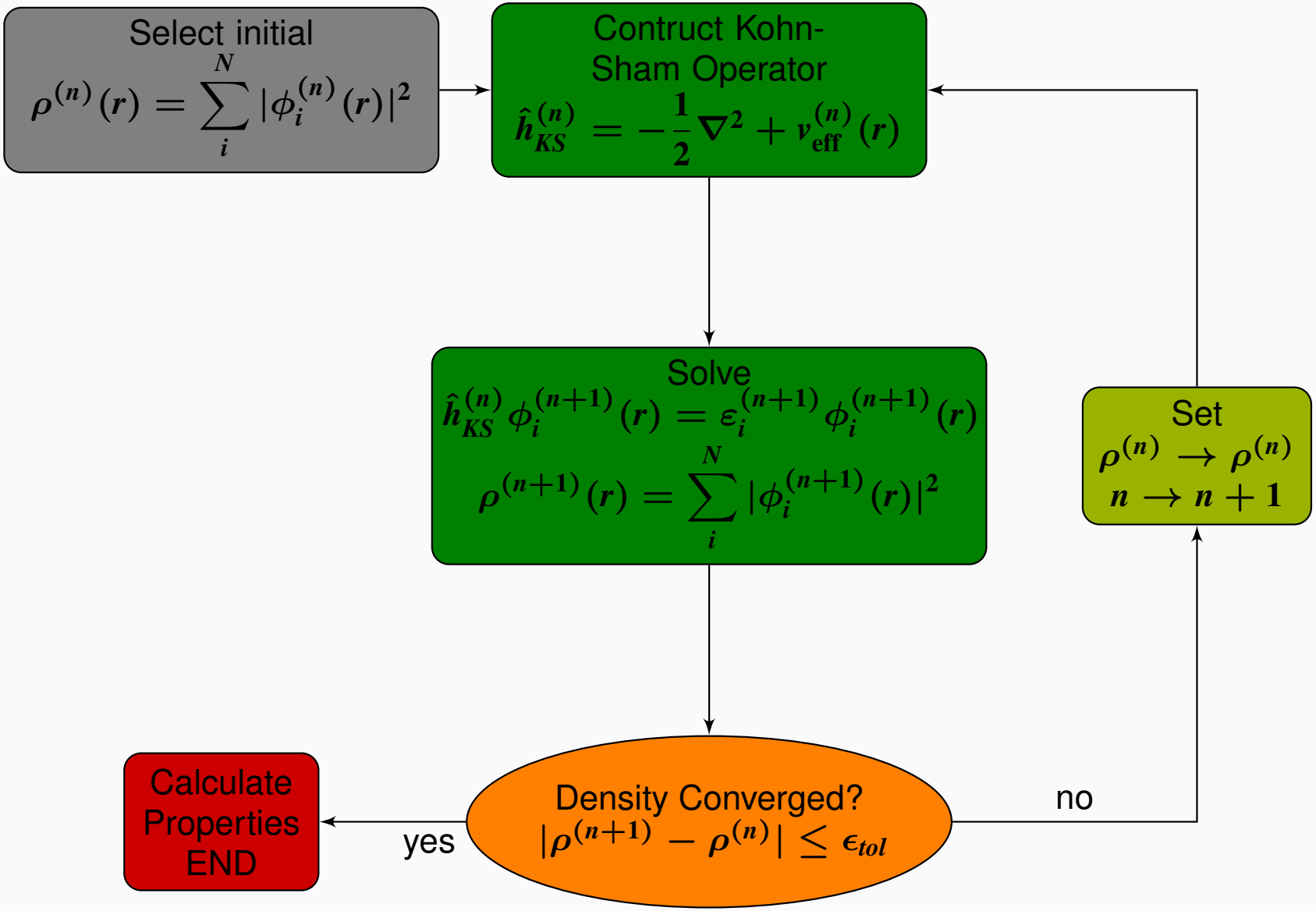
  where the last term $v_{\text{xc}}(r) \equiv \dfrac{\delta E_{\text{xc}}[\rho]}{\delta \rho(r)}$ is the exchange-correlation potential.

- 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

$$E_{\mathrm{xc}}[\rho] = E_{\mathrm{x}}[\rho] + E_{\mathrm{c}}[\rho]$$

  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 explicity correlated methods.
- Popular DFT functionals (according to a recent poll)
  - PBE0 (PBEPBE), 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`

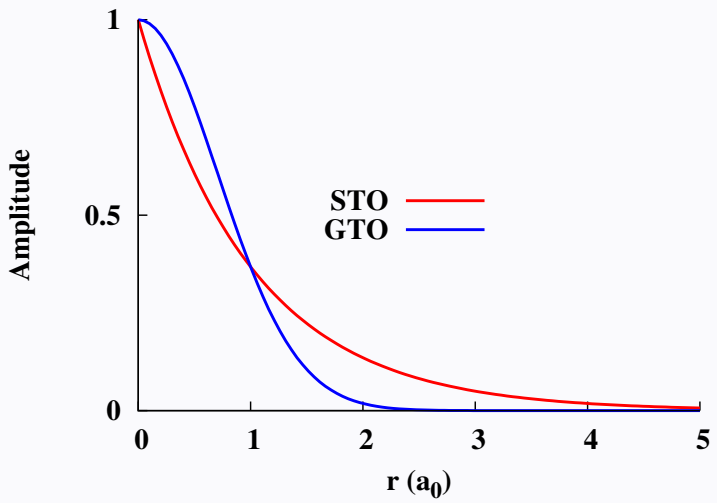
DFT Flow Chart



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

$$\phi^{\mathrm{STO}}(r) = x^l y^m z^n e^{-\zeta r}$$
$$\phi^{\mathrm{GTO}}(r) = x^l y^m z^n e^{-\xi r^2}$$

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



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 $\times$ 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
  - Stuttgard 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

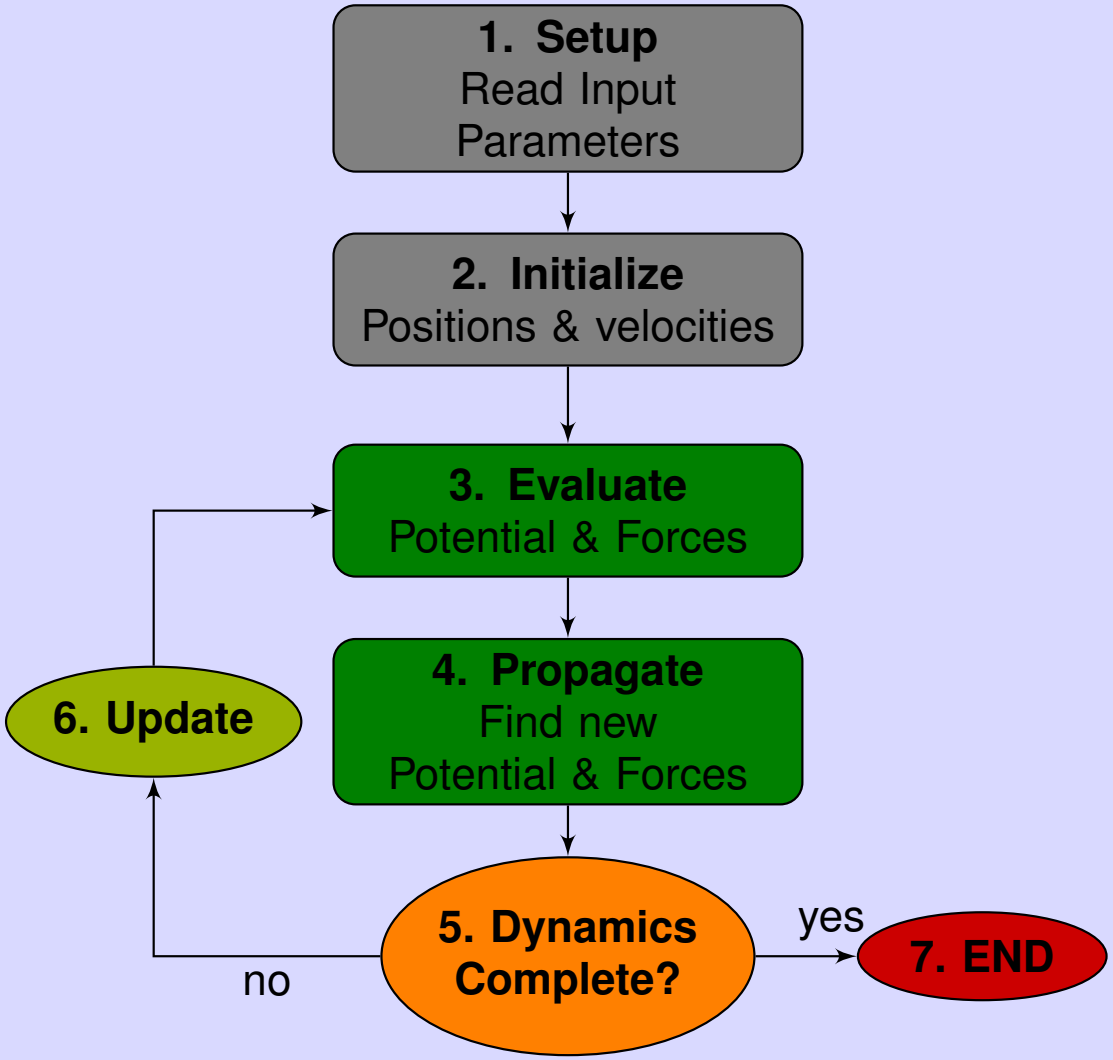
- Molecular Dynamics is a computer simulation of physical movements of atoms and molecules.
- The atoms and molecules are allowed to interact and their trajectories are determined by numerically solving the Newton's equation of motion.

$$\mathbf{F}_i = -\boldsymbol{\nabla}_i V = m_i \mathbf{a}_i$$

- The fundamental principle of Molecular Dynamics is the **Ergodic Hypothesis**: *the time average of a process is equal to an average over the statistical ensemble*.
- Ab-Initio Molecular Dynamics: Atoms and Molecules move in an average electronic potential obtained from ab-initio methods.
- Classical Molecular Dynamics: Atoms and Molecules move in an empirical potential or one obtained from Molecular Mechanics force fields.

General Schematic for MD Program

1. Setup: Read input parameters
2. Initialize: Obtain initial positions and velocities
3. Evaluate: Potential Energy and Forces on nuclei
4. Propagate nuclei using an appropriate time integration algorithm
5. Check if Dynamics is complete.
6. If incomplete update variables and goto Step 3.
7. If complete end dynamics or carry out all required analysis.



- Born-Oppenheimer Molecular Dynamics (BOMD): Electronic potential and nuclear forces are obtain on-the-fly during the dynamics.

$$M_I \ddot{\mathbf{R}}_I = -\nabla_I \min_{\Phi_0} \langle \Phi | H_e | \Psi \rangle$$

- Extended Lagrangian Molecular Dynamics (ELMD): Extend the Lagrangian by adding kinetic energy of fictitious particles and obtain their equation of motions from Euler-Lagrange equations.

$$\mathcal{L} = \hat{T} - \hat{V} \qquad \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{R}}_I} = \frac{\partial \mathcal{L}}{\partial \mathbf{R}_I}$$

1. Car-Parrinello Molecular Dynamics (CPMD)

$$\mathcal{L}_{\text{CPMD}} = \sum_I \frac{1}{2} M_I \dot{\mathbf{R}}_I^2 + \sum_i \frac{1}{2}\mu_i \langle \dot{\phi}_i | \dot{\phi}_i \rangle - \langle \Phi_0 | H_e | \Phi_0 \rangle + \sum_{i,j} \mathbf{\Lambda}_{i,j} \left(\langle \phi_i | \phi_j \rangle - \delta_{i,j} \right)$$

2. Atom-centered Density Matrix Propagation (ADMP)

$$\mathcal{L}_{\text{ADMP}} = \frac{1}{2}\text{Tr}(\mathbf{V}^T \mathbf{M} \mathbf{V}) + \frac{1}{2}\mu \text{Tr}(\dot{\mathbf{P}}\dot{\mathbf{P}}) - E(\mathbf{R}, \mathbf{P}) - \text{Tr}[\mathbf{\Lambda}(\mathbf{P}\mathbf{P} - \mathbf{P})]$$

3. Curvy-steps ELMD (cs-ELMD)

$$\mathcal{L}_{\text{csELMD}} = \sum_I \frac{1}{2} M_I \dot{\mathbf{R}}_I^2 + \frac{1}{2}\mu \sum_{i<j} \dot{\Delta}_{ij} - E(\mathbf{R}, \mathbf{P}); \quad \mathbf{P}(\lambda) = e^{\lambda \Delta} \mathbf{P}(0) e^{-\lambda \Delta}$$

- 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

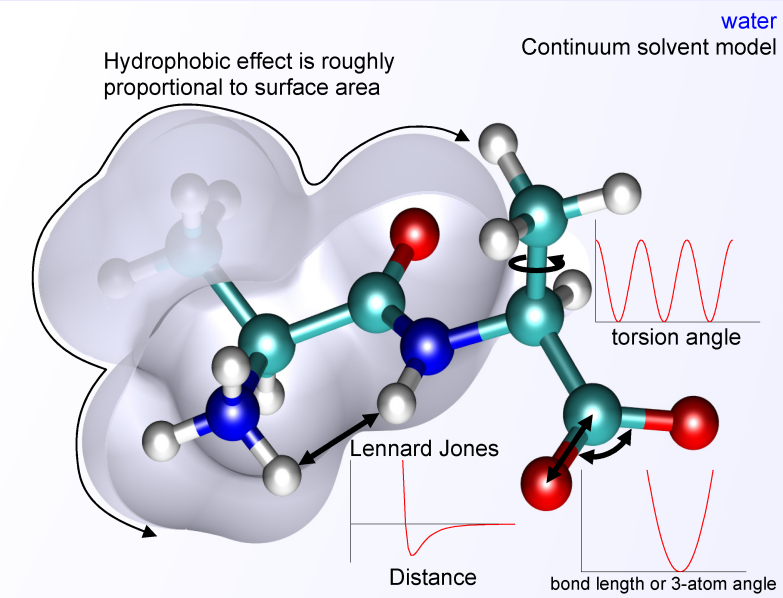
$$E = E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}} + E_{\text{vdW}} + E_{\text{elec}}$$

$$= \frac{1}{2}\sum_{\text{bonds}} K_b (b - b_0)^2 \qquad \text{Bond}$$

$$+ \frac{1}{2}\sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 \qquad \text{Angle}$$

$$+ \frac{1}{2}\sum_{\text{dihedrals}} K_\phi \left[1 + \cos(n\phi)\right]^2 \qquad \text{Torsion}$$

$$+ \sum_{\text{nonbonds}} \begin{cases} \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] & \text{van der Waals} \\ + \frac{q_1 q_2}{Dr} & \text{Electrostatics} \end{cases}$$



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

Outline

Software Installed on LONI Systems

Software	QB	Eric	Louie	Oliver	Painter	Poseidon
Amber	✓	✓	✓	✓	✓	✓
Desmond	✓					
DL_Poly	✓	✓	✓	✓	✓	✓
Gromacs	✓	✓	✓	✓	✓	✓
LAMMPS	✓	✓	✓	✓	✓	✓
NAMD	✓	✓	✓	✓	✓	✓
OpenEye	✓	✓	✓	✓	✓	✓
CPMD	✓	✓	✓	✓	✓	✓
GAMESS	✓	✓	✓	✓	✓	✓
Gaussian		✓	✓	✓	✓	
NWCHEM	✓	✓	✓	✓	✓	✓
Piny_MD	✓	✓	✓	✓	✓	✓

Gaussian

- One of the most popular packages in Computational Chemistry.
- Wavefunction based modeling of electronic structure of chemical systems.
- Model Chemistry: Molecular Mechanics (AMBER, Dreiding, UFF force fields), HF (RHF,UHF,ROHF), DFT, MP(2-5), CIS, CISD, CCSD, CCSD(T), G1,G2,CBS, CASSCF,GVB
- Basis Sets: Large Library of basis sets, can also include basis set in input file
- Capability: Energy, Gradients, Geometry Optimization, Transition State search, Frequency analysis, Solvation methods, ONIOM, IRC for reaction path, ADMP and BOMD for Molecular Dynamics.

- 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.

Gaussian Input

%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

- Wavefunction based modeling of electronic structure of chemical systems.
- Model Chemistry: HF,DFT,MP2,CI,CCSD(T),AM1,PM3
- Basis Sets: Most not all are available, others can be read from input file.
- Capability: Geometry optimization, Transition State search, Frequency calculations, MEP for reaction, Direct Dynamics.

GAMESS Input

```
$CONTRL SCFTYP=RHF RUNTYP=OPTIMIZE                    Job Control Data
   COORD=ZMT NZVAR=0 $END
$STATPT OPTTOL=1.0E-5 HSSEND=.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
```

- NWChem provides many methods for computing the properties of molecular and periodic systems using standard quantum mechanical descriptions of the electronic wavefunction or density.
- Classical molecular dynamics capabilities provide for the simulation of macromolecules and solutions, including the computation of free energies using a variety of force fields.
- Model Chemistry: Hartree-Fock (RHF,UHF,ROHF), DFT, MP2, CASSCF, CCSD,CCSDT,CCSDTQ etc
- Methods: Single Point Energies, Geometry Optimization, Molecular Dynamics, numerical first and second derivatives, Vibrational Analysis, ONIOM, COSMO, Electron Transfer, vibrational SCF and DFT, Pseudopotential Plane-Wave Electronic Structure, Molecular Dynamics.
- Basis Sets: Vast Library at `https://bse.pnl.gov/bse/portal`

NWCHEM Input

```
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
```

Job Types and Keywords

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	✓	qmmm

CPMD

- Car Parrinello Molecular Dynamics
- Typical problems CPMD is used for
  - Liquid Structures
  - Polarization effects
  - Bond breaking/formations
    - Proton transfer
  - Dynamic/thermal properties (e.g. diffusion)
  - Metadynamics
  - QM/MM
  - Path Integrals
  - TDDFT

CPMD Input

- Divided into sections
- Only sections pertaining to your simulated model at hand need to be present.
  - Common sections: INFO, CPMD, SYSTEM, ATOMS, DFT
  - Special sections: PIMD, PATH, RESP, LINRES, TDDFT, PROP, HARDNESS, CLASSIC, BASIS, VDW, QMMM
- Sections are defined `&Section_Name` and followed by `&END`, for example,

```
&ATOMS
  (information)
&END
```

- Each section has their own keywords
- Lines that do not match known keywords are ignored
- KEYWORDS HAVE TO BE IN ALL CAPS
  - kEYWORDS starting with a lower case character are ignored
  - Useful feature to re-use the input file
- Order of keywords is arbitrary

CPMD Input

```
&CPMD
  OPTIMIZE GEOMETRY XYZ
  CONVERGENCE ORBITALS
    1.0d-7
  CONVERGENCE GEOMETRY
    5.0d-4
&SYSTEM
  SYMMETRY
    SIMPLE CUBIC
  CELL
    16.00 1.0 1.0 0.0 0.0 0.0
  CUTOFF
    60.0
&END
&ATOMS
  *H_MT_LDA.psp
    LMAX=S
      2
      8.800 8.000 8.000
      7.200 8.000 8.000
&END
```

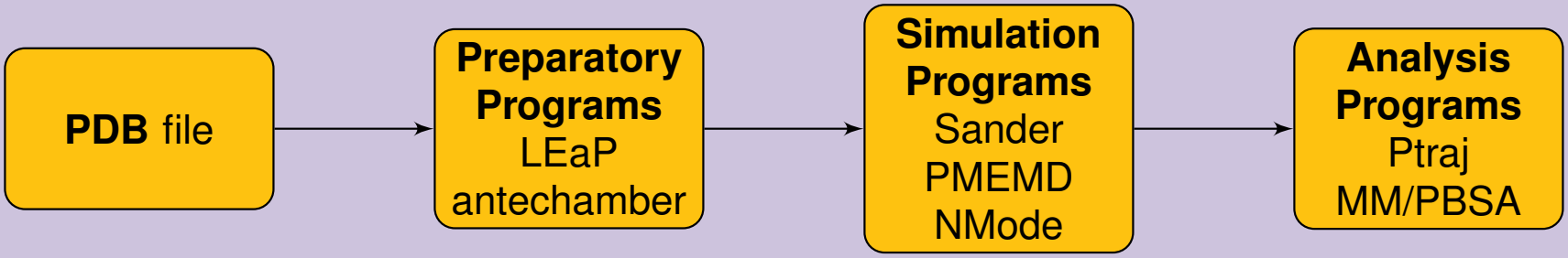
What is AMBER?

- AMBER, Assisted Model Building with Energy Refinement refers to two things
  1. A collective name for a suite of programs that allow users to carry out molecular dynamics simulation
  2. A set of molecular mechanical force fields for the simulation of biomolecules

Capabilities

- Classical MD (NVT,NPT,etc)
- Force Fields
- QM/MM
- Free Energy Calculations
- Structural and Trajectory analysis
- Parallelize dynamic codes
- Enhanced sampling (replica exchange MD)
- Explicit Solvent Models with PME
- Implicit Solvent Models with PB or GB approach

1. obtain and edit initial structure
2. prepare input parameter and topology file
3. run simulations and save trajectory
4. analyze output and trajectory files



Topology information includes

- atom types are assigned to identify different elements and different molecular orbital environments
- charges are assigned to each atom
- connectivities between atoms are established

Parameter information includes

- force constants necessary to describe the bond energy, angle energy, torsion energy, nonbonded interactions (van der Waals and electrostatics)
- other parameters for setting up the energy calculations (GB radii, FEP parameter sets)

LEaP

- Includes a tex-based interface - tleap and a graphical user interface - xleap
- Capabilities
  - Read AMBER force field information
  - Read and write files containing structural information (i.e.PDB files)
  - Construct new residues and molecules
  - Link together residues and create nonbonded complexes of molecules
  - Place counterions around a molecule; Solvate molecules; Mdoify internal coordinates within a molecule
  - Generate files that contain topology and parameters for AMBER
  - Set atomic charges, identify the position of disulphide bridges, delete bonds, addition of atoms, ions, etc ...

loading PDB file, adding the disulfide cross links and saving files

```
source leaprc.ff03
bpti = loadPdb 6pti.mod.pdb
bond bpti.5.SG bpti.55.SG
bond bpti.14.SG bpti.38.SG
bond bpti.30.SG bpti.51.SG
saveAmberParm bpti prmtop prmcrd
quit
```

Amber 11 Manual Page 17

Sander

- Sander: Simulated Annealing with NMR-Derived Energy Restraints
  - Energy minimization, molecular dynamics and NMR refinements
  - Free energy calculations (Umbrella Sampling; SMD; etc)
  - QM/MM implementation (EVB; semi-empirical/AMBER)
  - Polarizable force field (AMOEBA)
  - Enhanced Sampling (REMD; LES, etc)
- Usage: `sander -i mdin -o mdout -p prmtop -c inpcrd -r restrt`

```
mdin:  input control data for minimization/MD run.
mdout:  output file for user readable state info and diagnostics
prmtop:  molecular topology, force fields etc
inpcrd:  initial coordinates and velocities
restrt:  restart filename
```

200 steps of minimization, generalized Born solvent model

```
&cntrl
maxcycle=200,
imin=1,
cut=12.0,
igb=1,
ntb=0,
ntpr=10,
/
```

Amber 11 Manual Page 17

Run Sander

```
sander -i min.in -o 6pti.min1 -c prmcrd -r 6pti.min1.xyz
```

LAMMPS

- LAMMPS stands for Large-scale Atomic/Molecular Parallel Simulator.
- LAMMPS is a classical molecular dynamics code that models an ensemble of particles in a liquid, solid, or gaseous state designed to run efficiently on parallel computers.
- It can model atomic, polymeric, biological, metallic, granular, and coarse-grained systems using a variety of force fields and boundary conditions.
- LAMMPS can model systems with only a few particles up to millions or billions.
- LAMMPS is designed to be easy to modify or extend with new capabilities, such as new force fields, atom types, boundary conditions, or diagnostics.
- LAMMPS runs efficiently on single-processor desktop or laptop machines, but is designed for parallel computers.
- It is an open-source code, distributed freely under the terms of the GNU Public License (GPL).

- LAMMPS doesn't
  1. Build molecular systems
  2. Assign force-dield coefficients auto-magically
  3. Compute lots of diagnostics on-the-fly
  4. Visualize your output
- LAMMPS version "4 May 2011" is installed on all LONI Dell Linux Clusters.
- Add the appropriate soft keys to your .soft file
  - +lammps-4May11-intel-11.1-mvapich-1.1
- Command line options for running LAMMPS
  - -in inputfile: specify input file
  - -log logfile: specify log file
  - -partition MxN L: Run on (MxN)+L processors with M partitions on N processors each and 1 partition with L processors
  - -screen file: Specify a file to write screen information

LAMMPS Input

- Reads an input script in ASCII format one line at a time.
- Input script consists of 4 parts
  1. Initialization: Set parameters that need to be defined before atoms are created or read-in from a file.
     `units, dimension, newton, processors, boundary, atom_style, atom_modify`
  2. Atom definition: `read_data, read_restart, lattice, region, create_box, create_atoms, replicate`
  3. Settings: Once atoms and molecular topology are defined, a variety of settings can be specified: force field coefficients, simulation parameters, output options, etc.
     `pair_coeff, bond_coeff, angle_coeff, dihedral_coeff, improper_coeff, kspace_style, dielectric, special_bonds, neighbor, neigh_modify, group, timestep, reset_timestep, run_style, min_style, min_modify, fix, compute, compute_modify, variable`
  4. Run a simulation: A MD is run using the run command. Energy minimization (molecular statics) is performed using the minimize command. A parallel tempering (replica-exchange) simulation can be run using the temper command.
- `http://lammps.sandia.gov/doc/Section_commands.html`

LAMMPS Input for 3d Lennard-Jones Melt

units lj	style of units used for a simulation
atom_style atomic	what style of atoms to use in a simulatio
lattice fcc 0.8442	Lattice structure
region box block 0 20 0 20 0 20	define simulation region
create_box 1 box	create simulation box
create_atoms 1 box	create atoms on lattice
mass 1 1.0	
velocity all create 3.0 87287	
pair_style lj/cut 2.5	Formula for Pairwise Interaction
pair_coeff 1 1 1.0 1.0 2.5	pairwise force field coefficients
neighbor 0.3 bin	pairwise neighbor lists
neigh_modify every 20 delay 0 check no	
fix 1 all nve	constant NVE time integration
dump id all atom 10 dump.melt	For visualization
thermo 50	output thermodynamics data
run 2500	Simulation Steps

- NAMD is a parallel molecular dynamics code designed for high-performance simulation of large biomolecular systems.
- Based on Charm++ parallel objects, NAMD scales to hundreds of processors on high-end parallel platforms and tens of processors on commodity clusters using gigabit ethernet.
- NAMD uses the popular molecular graphics program VMD for simulation setup and trajectory analysis, but is also file-compatible with AMBER, CHARMM, and X-PLOR.

What is needed to run NAMD?

- Protein Data Bank (pdb) file which stores atomic coordinates and/or velocities for the system.
- Protein Structure File (psf) which stores structural information of the protein, such as various types of bonding interactions.
- A force field parameter file. A force field is a mathematical expression of the potential which atoms in the system experience. CHARMM, X-PLOR, AMBER, and GROMACS are four types of force fields, and NAMD is able to use all of them. The parameter file defines bond strengths, equilibrium lengths, etc.
- A configuration file, in which the user specifies all the options that NAMD should adopt in running a simulation.

NAMD configuration file

- A NAMD configuration file contains a set of options and values which determine the exact behavior of NAMD, what features are active or inactive, how long the simulation should continue, etc.
- The following parameters are *required* for every NAMD simulation

  `numsteps`, `coordinates`, `structure`, `parameters`, `exclude`, `outputname` and one of the following three: `temperature`, `velocities`, `binvelocities`.

NAMD Configuration File for decalanin

```
numsteps 10000                    protocol params

coordinates alanin.pdb            initial config
temperature 300K
seed 12345

outputname ./alanin               output params
binaryoutput no

timestep 1.0                      integrator params

structure alanin.psf              force field params
parameters alanin.params
exclude scaled1-4
1-4scaling 1.0
switching on
switchdist 8.0
cutoff 12.0
pairlistdist 13.5
margin 0.0
stepspercycle 20
```

- GROMACS (Groningen Machine for Chemical Simulations) is a versatile package which performs molecular dynamics, i.e. simulates the Newtonian equations of motion for systems with hundreds to millions of particles.
- It is primarily designed for biochemical molecules like proteins and lipids that have a lot of complicated bonded interactions, but since GROMACS is extremely fast at calculating the nonbonded interactions (that usually dominate simulations) many groups are also using it for research on non-biological systems, e.g. polymers.
- GROMACS calling structure is quite complicated, and it is not recommended that users proceed without at least becoming familiar with the program.
- You will need a protein structure file (pdb file) and a file containing energy minimization data.

Files at /home/apacheco/CompChem/CMD/GROMACS

1. Generate topology file using pdb2gmx

```
pdb2gmx_d -f cpeptide.pdb -o cpeptide.gro -p cpeptide.top
```

2. Define the water box size

```
editconf_d -f cpeptide.gro -o cpeptide.gro -d 0.5
```

3. Solvate the peptide

```
genbox_d -cp cpeptide.gro -cs -o cpeptide_b4em.gro -p cpeptide.top
```

4. Energy Minimization, step 1: preprocess the input files.

```
grompp_d -f em -c cpeptide_b4em -p cpeptide -o cpeptide_em
```

5. Energy minimization, step 2

```
mdrun_d -s cpeptide_em -o cpeptide_em -c cpeptide_b4pr -v
```

6. Position Restrained MD, step 1: generate the binary topology file

```
grompp_d -f pr -c cpeptide_b4pr -r cpeptide_b4pr -p cpeptide -o cpeptide_pr
```

7. Position Restrained MD, step 2: run the simulation

```
mdrun_d -s cpeptide_pr -o cpeptide_pr -c cpeptide_b4md -v
```

8. MD Simulation, step 1: generate the binary topolgy file

```
grompp_d -f md -c cpeptide_b4md -p cpeptide -o cpeptide_md
```

9. MD simulation, step 2: run the simulation

```
mdrun_d -s cpeptide_md -o cpeptide_md -c cpeptide_after_md -v
```

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 + Freqeuncy
- 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 $\sim$ 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.