Introduction to Atomistic Simulations

- introduction to the levels of theory and their application to research
- atomistic software currently available (free and licensed, HPRC/LMS)
- general overview of the resources required for a variety of atomistic simulations (quantum vs classical)
  - cores/cpus
  - memory
  - disk
  - Service Units (SUs)
- job file preparation
  - LSF
  - SLURM
Introduction to Atomistic Simulations

- Quantum Mechanics
  - \textit{ab initio}
  - Density-Functional Theory (DFT)
  - Semi-Empirical
- Classical Mechanics
  - molecular mechanic (MM)
  - molecular dynamics (MD)
- Prediction of properties
  - Structural, chemical, physical, biological, etc.
- Understand, explain, and possibly predict chemical processes
- Tools to gain insight at the atomic and molecular level
  - Microscopic $\leftrightarrow$ Macroscopic

Microscopic $\leftrightarrow$ Macroscopic

\begin{itemize}
  \item Time
    \begin{itemize}
      \item $>$min
      \item \textit{s}
      \item \textit{µs}
      \item \textit{ns}
      \item \textit{ps}
      \item \textit{fs}
    \end{itemize}
  \item Length Scale
    \begin{itemize}
      \item Ångstroms
      \item nm
      \item \textit{µm}
      \item mm
      \item m
    \end{itemize}
\end{itemize}
Atomistic Simulations

- Quantum Mechanics
  - \textit{ab initio} - based on first principles
    - Hartree-Fock Theory (HF or SCF)
    - Møller-Plesset Perturbation Theory (MPn; \(n = 2, 3, 4, \ldots\))
    - Configuration Interaction (CI; CIS, CISD, CISDT, \ldots)
    - Coupled-Cluster (CC; CCD, CCSD, CCSD(T), CCSD(TQ), \ldots)
    - Complete Active Space Self Consistent Field (CASSCF)
    - Multi-Reference Configuration Interaction (MRCI)
    - and many more
  - Density Functional Theory (DFT)
    - B3LYP, BP86, B3PW91, mPW1PW91, PBE, M06, TPSS, oB97x-D
    - and many more
  - Semi-empirical
    - AM1
    - PM3, PM5, PM7
    - and many more

Quantum Mechanics

- \textit{ab initio} methods
  - Schrödinger Equation
    - \(H\Psi = E\Psi\); time-independent Schrödinger equation.
  - Applicable to any system, in principle.
  - Can model bond breaking and formation
  - Used for benchmark values
  - Can only be used for small system (normally < 200 atoms)
  - Computationally expensive
    - Scaling: \(N^n; n = 2, 3, 4, 5, 6, \ldots\)
  - Commonly used codes
    - Gaussian 09, GAMESS-US, Spartan, NWChem
    - Q-Chem, MOLPRO, Dalton, GAMESS-UK, CRYSTAL
    - and many more
Quantum Mechanics

- Density Functional Theory (DFT)
  - Total energy of a system depends only on the electron density
    \[ E_{\text{tot}} = f[\rho(x,y,z,s)] \]
  - Applicable to any system, in principle.
  - Can model bond breaking and formation
  - Includes electron correlation with little cost compared to \textit{ab initio} methods
  - Exact functional is not known
- Commonly used software for DFT
  - Gaussian 09, Jaguar, DMol3, Turbomole, Amsterdam Density Functional (ADF), GAMESS-US, NWChem, MOLPRO, Spartan, GAMESS-UK, CRYSTAL, and many more

Quantum Mechanics

- Semi-empirical Methods
  - AM1, PM3, PM5, PM7, PM7-TM, SAM1, etc.
  - Approximate solution to the Schrödinger equation
  - Replaces the expensive integrals with parameters
  - Applicability is limited by available parameterization
    - Mostly used for 1\textsuperscript{st} row main group elements
    - Limited applicability to transition metals
  - Can model bond breaking/formation
  - Can model much larger systems than \textit{ab initio} or DFT
- Commonly used software
  - MOPAC 2012, Gaussian 09, Chem3d, AMPAC, VAMP, and many more
Reaction Mechanism Calculations

<table>
<thead>
<tr>
<th>Entry</th>
<th>X</th>
<th>R1</th>
<th>Lewis acid</th>
<th>3a % yield</th>
<th>5a-% yield</th>
<th>10a-% yield</th>
<th>11a-% yield</th>
<th>E/Z ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>OPh</td>
<td>TMS</td>
<td>BF3OEt2</td>
<td>&lt;5</td>
<td>NA</td>
<td>62</td>
<td>&lt;5i</td>
<td>&lt;5</td>
</tr>
<tr>
<td>2</td>
<td>OPh</td>
<td>TMS</td>
<td>TiCl4</td>
<td>&lt;5</td>
<td>NA</td>
<td>61</td>
<td>&lt;5</td>
<td>&lt;5</td>
</tr>
<tr>
<td>3</td>
<td>OPh</td>
<td>TMS</td>
<td>MgBr2·OEt2</td>
<td>&gt;19:1</td>
<td>19</td>
<td>18</td>
<td>&lt;5</td>
<td>&lt;5</td>
</tr>
<tr>
<td>4</td>
<td>OPh</td>
<td>TMS</td>
<td>ZnCl2</td>
<td>20.1</td>
<td>16</td>
<td>17</td>
<td>&lt;5</td>
<td>&lt;5</td>
</tr>
<tr>
<td>5</td>
<td>SPh</td>
<td>TBS</td>
<td>TiCl4</td>
<td>&lt;5</td>
<td>NA</td>
<td>57</td>
<td>&lt;5</td>
<td>&lt;5</td>
</tr>
<tr>
<td>6</td>
<td>SPh</td>
<td>TES</td>
<td>ZnCl2</td>
<td>32</td>
<td>21</td>
<td>26</td>
<td>&lt;5</td>
<td>&lt;5</td>
</tr>
<tr>
<td>7</td>
<td>SPh</td>
<td>TES</td>
<td>ZnCl2</td>
<td>70</td>
<td>26</td>
<td>&lt;5</td>
<td>&lt;5</td>
<td>&lt;5</td>
</tr>
<tr>
<td>8</td>
<td>SPh</td>
<td>TES</td>
<td>ZnCl2</td>
<td>70</td>
<td>26</td>
<td>&lt;5</td>
<td>&lt;5</td>
<td>&lt;5</td>
</tr>
</tbody>
</table>

Cunxiang Zhao, T. Andrew Mitchell, Ravikrishna Vallakati, Lisa M. Pérez, and Daniel Romo
DOI: 10.1021/ja209163w
Reaction Mechanism Calculations

Original assignment for spectra a)

(E)-2a is lower in energy than (E)-2a-ZnCl₂

NMR Calculations

<table>
<thead>
<tr>
<th></th>
<th>(E)-2a</th>
<th>(E)-2a-ZnCl₂</th>
<th>(E)-2a-ZnCl₂*</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>exp[*]</td>
<td>calc[a]</td>
<td>exp[ex]</td>
</tr>
<tr>
<td>H₆</td>
<td>5.46</td>
<td>5.66</td>
<td>5.56</td>
</tr>
<tr>
<td>H₄</td>
<td>7.33–7.35</td>
<td>7.47</td>
<td>7.63–7.64</td>
</tr>
<tr>
<td>H₂</td>
<td>7.54–7.57</td>
<td>7.53</td>
<td>7.83–7.87</td>
</tr>
<tr>
<td>H₁</td>
<td>8.42–8.44</td>
<td>8.59</td>
<td>8.85–8.89</td>
</tr>
</tbody>
</table>
Molecular Modeling Workshop

UV/Vis spectra via TD-DFT

Molecular Mechanics (Classical) \[ F = \frac{ma}{m} = m\frac{d^2r}{dt^2} ; \quad F = -\frac{\partial V}{\partial r} \]

- Newton's equations
- The potential is approximated by an empirical function force field that is fitted to approximately reproduce known interactions
- Applicability is limited by the availability of parameterization
- Generally, the connectivity of atoms cannot change during the simulation
  - Generally, not suitable for reaction mechanisms
  - Can predict relative energies of different conformational states of material
  - And much more
Molecular Mechanics

- The molecule is considered to be a collection of atoms held together by simple elastic or harmonic forces.

- Force Field - A mathematical expression that describes the dependence of the energy of a molecule on the coordinates of the atoms in the molecule.

- Force Field Energy Expression:
  
  \[
  E = E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}} + E_{\text{oop(out-of-plane)}} + E_{\text{non-bond}} + E_{\text{other}}
  \]

**Force Field Terms - Bond Term**

<table>
<thead>
<tr>
<th>Bond Stretch</th>
<th>Harmonic</th>
<th>Cubic</th>
<th>Quartic</th>
<th>Morse</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[E = \frac{K}{2}(R - R_0)^2]</td>
<td>[E = \frac{K}{2}(R - R_0)^2[1+d(R - R_0)]]</td>
<td>[E = \frac{K}{2}(R - R_0)^2[1+c(R - R_0)+d(R - R_0)^4]]</td>
<td>[E = D(\exp(-\alpha(R - R_0^2)) - \frac{1}{2})^2 - S]</td>
</tr>
<tr>
<td>Parameters</td>
<td>(K) : kcal/mol (\AA^2)</td>
<td>(K, c, d) : kcal/mol (\AA^2)</td>
<td>(K, c, d) : kcal/mol (\AA^2)</td>
<td>(\alpha, S, D) : kcal/mol</td>
</tr>
<tr>
<td>Units</td>
<td>(\AA)</td>
<td>(\AA)</td>
<td>(\AA)</td>
<td>(\AA)</td>
</tr>
<tr>
<td>Force Field</td>
<td>Bond Term</td>
<td>Bond Term</td>
<td>Bond Term</td>
<td>Bond Term</td>
</tr>
<tr>
<td>Software</td>
<td>Unbonded</td>
<td>COMPASS</td>
<td>COMPASS</td>
<td>COMPASS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Notes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- where \(\alpha = \frac{\sqrt{2K}}{2L_0}\) and \(S\) is a constant shift value.
Force Field Terms – Angle Term

**Harmonic**

\[ E = \frac{K_a}{2} \left( \theta - \theta_0 \right)^2 \]

\[ K_a = \text{kl/mol/Å}^2 \]

<table>
<thead>
<tr>
<th>\θ_0</th>
<th>\theta</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>a</td>
<td>p</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>\theta</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>rad</td>
<td>p</td>
</tr>
</tbody>
</table>

**Cubic**

\[ E = \frac{K_a}{2} \left( \theta - \theta_0 \right)^3 \left[ 1 + d(\theta - \theta_0) \right] \]

\[ K_a = \text{kl/mol/Å}^2 \]

<table>
<thead>
<tr>
<th>\θ_0</th>
<th>\d_0</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>a</td>
<td>p</td>
</tr>
</tbody>
</table>

**Quartic**

\[ E = \frac{K_a}{2} \left[ (\theta - \theta_0)^4 \right] \left[ 1 + c(\theta - \theta_0) + d(\theta - \theta_0)^2 \right] \]

\[ K_a = \text{kl/mol/Å}^2 \]

<table>
<thead>
<tr>
<th>\θ_0</th>
<th>\c_0</th>
<th>\d_0</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>a</td>
<td>a</td>
<td>p</td>
</tr>
</tbody>
</table>

**Cosine Harmonic**

\[ E = \frac{K_a}{2} \left( \frac{1}{\cos(\theta - \theta_0)} \right)^2 \]

\[ \theta_0 = 0, E = K_a \left( 1 - \cos(\theta) \right) \]

<table>
<thead>
<tr>
<th>\θ_0</th>
<th>\θ</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>rad</td>
<td>p</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>\θ</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>rad</td>
<td>p</td>
</tr>
</tbody>
</table>

**Cosine Periodic**

\[ E = C \left[ 1 - \left( \frac{1}{2} \right)^n \cos(n\theta) \right] \]

\[ C = \text{kl/mol} \]

<table>
<thead>
<tr>
<th>\θ</th>
<th>\n</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>p</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>\n</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>p</td>
</tr>
</tbody>
</table>

**Torsion**

**Dihedral**

\[ E = \frac{1}{2} \sum \left( \frac{\theta}{\theta_0} - 1 \right)^2 \]

Where \θ = angle between the 3X and 3Y planes

<table>
<thead>
<tr>
<th>\θ_0</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>p</td>
</tr>
</tbody>
</table>

**Skew Dihedral**

\[ E = \frac{1}{2} \sum \left( \frac{\theta}{\theta_0} - 1 \right) \cos(\theta) \]

where \θ = initial phase shift

<table>
<thead>
<tr>
<th>\θ_0</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>p</td>
</tr>
</tbody>
</table>

**Inversion**

\[ E = \frac{1}{2} \sum \left( \frac{\theta}{\theta_0} - 1 \right)^2 \]

Where \θ_0 = the angle of the three umbrella inversion

<table>
<thead>
<tr>
<th>\θ_0</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>p</td>
</tr>
</tbody>
</table>

**Umbrella**

\[ E = C \frac{\theta^2}{\theta_0^2} \]

<table>
<thead>
<tr>
<th>\θ_0</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>p</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>\θ</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>rad</td>
<td>p</td>
</tr>
</tbody>
</table>

**Bend**

\[ E = C \frac{\theta^2}{\theta_0^2} \]

<table>
<thead>
<tr>
<th>\θ_0</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>p</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>\θ</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>rad</td>
<td>p</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>\θ</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>rad</td>
<td>p</td>
</tr>
</tbody>
</table>
**Force Field Terms – Van der Waals Terms**

\[ E = E_{v d W} \]

**LJ 9 4**
\[ E = E^{9/4} \left( \frac{r_{ij}}{r_{ij}^{0}} \right) \]
\[ r_{ij}^{0} \text{ is Lennard-Jones } \]
\[ \sigma_{ij} \text{ is Lennard-Jones } \]
\[ \text{Custom} \]

**LJ 9 6**
\[ E = E^{9/6} \left( \frac{r_{ij}}{r_{ij}^{0}} \right) \]
\[ r_{ij}^{0} \text{ is Lennard-Jones } \]
\[ \sigma_{ij} \text{ is Lennard-Jones } \]
\[ \text{COMPASS Custom} \]

**LJ 12 4**
\[ E = E^{12/4} \left( \frac{r_{ij}}{r_{ij}^{0}} \right) \]
\[ r_{ij}^{0} \text{ is Lennard-Jones } \]
\[ \sigma_{ij} \text{ is Lennard-Jones } \]
\[ \text{Custom} \]

**LJ 12 6**
\[ E = E^{12/6} \left( \frac{r_{ij}}{r_{ij}^{0}} \right) \]
\[ r_{ij}^{0} \text{ is Lennard-Jones } \]
\[ \sigma_{ij} \text{ is Lennard-Jones } \]
\[ \text{Draxing Universal Model with Merz-Kramers Custom} \]

**Exponential**
\[ E = E^{12} \left( \frac{r_{ij}}{r_{ij}^{0}} \right) \]
\[ r_{ij}^{0} \text{ is Lennard-Jones } \]
\[ \sigma_{ij} \text{ is Lennard-Jones } \]
\[ \text{Custom} \]

**Morse**
\[ E = E^{12} \left( \frac{r_{ij} - r_{0}}{r_{ij}^{0}} \right) \]
\[ r_{ij}^{0} \text{ is Lennard-Jones } \]
\[ \sigma_{ij} \text{ is Lennard-Jones } \]
\[ \text{Custom} \]

**Buckingham**
\[ E = E^{12} \ln \left( \frac{r_{ij}^{0}}{r_{ij}} \right) \]
\[ r_{ij}^{0} \text{ is Lennard-Jones } \]
\[ \sigma_{ij} \text{ is Lennard-Jones } \]
\[ \text{Custom} \]

**Soft Harmonic**
\[ E = \frac{1}{2} k \left( \frac{r_{ij} - r_{ij}^{0}}{r_{ij}^{0}} \right)^{2} \]
\[ r_{ij}^{0} \text{ is Lennard-Jones } \]
\[ \sigma_{ij} \text{ is Lennard-Jones } \]
\[ \text{Custom} \]
Molecular Mechanics

- Force Fields differ in their parameters, terms and the method of development
  - Class I - simple functional form with data fitted to quantum mechanical calculations and/or experiment (AMBER, CHARMM, etc)
  - Class II - more complicated functional form using cross terms and data fitted to quantum mechanical calculations and/or experiment (CFF, PCFF, etc)
  - Class III - new generation force fields that incorporate polarizability (AMOEBA, AMBER ff02, CHARMM Drude, etc)
- Rules Based - covers most of the periodic table – UFF, DREIDING, etc
  - Fundamental quantities are derived for each atom type: electronegativity, hardness, atomic radius, etc.
  - Forcefield parameters are derived at runtime using a series of theoretically or empirically derived rules
- Specialist - developed for a particular family of compounds flourinated polymers, zeolites, etc.
- Reaction Forcefields - ReaxFF

Common Force Fields

- MM2/MM3/MM4: Molecular Mechanic Force field for small organic molecules
- CHARMM: Chemistry at Harvard Macromolecular Mechanics
- AMBER: Assisted Model Building with Energy Refinement
- GROMOS: GROningen MOlecular Simulation
- OPLS: Optimized Parameters for Liquid Simulation
- CFF: Consistent Force Field
- CVFF: Valence Consistent Force Field
- MMFF94: Merck Molecular Force Field 94
- DREIDING: Generic rules based force field
- UFF: Universal Force Field
- ReaxFF: Speciality force-field to allow bond breaking
Minimization

- Minimize the potential energy
  \[ E = E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}} + E_{\text{covalent}} + E_{\text{nonbond}} + E_{\text{other}} \]

![Energy Landscape with Local and Global Minima](image)

Molecular Dynamics

- Minimization methods will only optimize your molecule to the closest local minimum
- Methods to find the global minimum
  - Systematic conformational search
    - Very time consuming and essentially impossible for anything but the smallest of molecules
    - \( \text{CH}_3(\text{CH}_2)_{n+1}\text{CH}_3 \)
      - \( n=1 \): 3 possible configurations, <1 sec
      - \( n=2 \): 243, 1 min
      - \( n=10 \): 59,049, 2 hour
      - \( n=15 \): 14,348,907, 100 days
  - Molecular Dynamics, Random Sampling, Monte Carlo, Hybrid Monte Carlo/Dynamics methods
    - Samples the potential energy surface by perturbing the geometry
Molecular Dynamics

- Methods to find the global minimum
  - High temperature dynamics
    - Simulated annealing
    - Quench dynamics

![Diagram of Molecular Dynamics](image)

Energy supplied to the minimized structure at the start of the simulation

Molecular Dynamics

- Molecular Dynamics Variations
  - Constant Volume - Constant Temperature (NVT)
  - Constant Volume - Constant Energy (NVE)
  - Constant Pressure – Constant Temperature (NPT)

- construct a set of velocities based on the ensemble being used
- Velocities satisfy the Maxwell-Boltzmann distribution
- Each run will start with a different random seed

- Allow atoms to move for one time step
- Calculate the force on the atoms - forcefield
- Calculate the acceleration $F=ma$
- Calculate the new velocity
- Calculate the new position
- Repeat for as many time steps as desired

\[ r(t+\delta t) = r(t) + \delta t \cdot v(t) + \frac{1}{2} \delta t^2 a(t) \]

\[ v(t+\delta t) = v(t) + \frac{1}{2} \delta t (a(t) + a(t+\delta t)) \]

\[ r(t+\delta t) = r(t) + \delta t \cdot v(t) + \frac{1}{2} \delta t^2 a(t) \]
Molecular Dynamics

- Choosing a time step
  - Your time step should be a factor of 10 smaller than the fastest process in your system.
    - Molecular motions such as rotations and vibrations are on the order of $10^{-11}$ to $10^{-14}$ s
    - Therefore, a time step of 1 fs ($10^{-15}$ s) or less must be used for most systems.
  - You can increase your time step by restricting the fastest processes
    - SHAKE or RATTLE algorithms restrict the vibrational motion of the molecule of interest
    - Therefore, a time step of 2-3 fs can be used with the SHAKE or RATTLE algorithm
    - There are some modified shake algorithms that claim they are stable up to time steps of 8 fs
- Most simulations are on the order of picoseconds ($10^{-12}$ s) or nanoseconds ($10^{-9}$ s)
- Protein folding tripzip2 (12-residue protein) folds on the order of 2.5 µs ($10^{-6}$ s)

Molecular Simulation of ab Initio Protein Folding for a Millisecond Folder NTL9(1−39)

- http://folding.stanford.edu/English/Papers#ntoc6
- Gromacs(GPU version) and Folding@Home were used for the simulations with implicit solvation

http://youtu.be/gFcp2Xpd29I
Conformational Searches

Ion mobility-mass spectrometry peptide map of bovine hemoglobin. Two low-energy MD calculated structures are assigned to peptide signals within the plot: (1) LLGNVLVVVLAR and (2) LLVVYYPWTQR. The two peptide projections shown are 15 (top) and 10 Å (bottom).


Monte Carlo – RIS Metropolis Monte Carlo

- In outline, an RMMC simulation proceeds as follows:
  - Perform an energy minimization on the molecule
  - Randomly select a rotatable backbone bond.
  - Select a random torsion value for this bond between -180 and +180 degrees.
  - Rotate the bond to its new torsion value and compute the new energy of the chain.
  - Generate a random number, R, between 0 and 1. If \( \exp\left(-\frac{(E_{\text{new}} - E_{\text{old}})}{kT}\right) > R \), keep the new torsion value. Otherwise, restore the old value.
  - Repeat until the desired number of iterations has been performed.
Molecular Modeling Workshop

Molecular Dynamics Steps

- Prepare the Molecule
- Minimization
  - Required!!!!!!
- Heating
  - Raise the temperature of the system from 0°C to the desired temperature.
  - Many programs/researchers skip this step
- Equilibration
  - Ensure that the system is stable (NVE)
- Production
  - Run simulation to collect pertinent data
    - NVT, NVE, NPT, etc

Software – Quantum Mechanics

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Licensing</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABINIT</td>
<td>QM (Molecular and Periodic Systems)</td>
<td>Site License*</td>
<td>Free</td>
</tr>
<tr>
<td>AMPAC</td>
<td>QM (Semi-empirical)</td>
<td>Site License</td>
<td>$</td>
</tr>
<tr>
<td>GAMESS-US</td>
<td>QM</td>
<td>Site License*</td>
<td>Free</td>
</tr>
<tr>
<td>Gaussian</td>
<td>QM</td>
<td>Site License</td>
<td>$</td>
</tr>
<tr>
<td>MOLPRO</td>
<td>QM (specializing in high-level calculations)</td>
<td>Site License</td>
<td>$$</td>
</tr>
<tr>
<td>NBO</td>
<td>Wavefunction analysis program</td>
<td>Site License</td>
<td>$</td>
</tr>
<tr>
<td>NWChem</td>
<td>QM</td>
<td>Site License*</td>
<td>Free</td>
</tr>
<tr>
<td>ORCA</td>
<td>QM specializing in spectroscopic properties</td>
<td>Site License*</td>
<td>Free</td>
</tr>
<tr>
<td>Quantum ESPRESSO</td>
<td>QM solid state and surfaces</td>
<td>Site License</td>
<td>Free</td>
</tr>
<tr>
<td>SIESTA</td>
<td>QM specializing in electron transport and Solids</td>
<td>Site License*</td>
<td>Free</td>
</tr>
<tr>
<td>VASP</td>
<td>QM specializing in QMD and ultra-soft ECPs</td>
<td>Restricted Licenses</td>
<td>$$</td>
</tr>
</tbody>
</table>

module spider *keyword*
Quantum Mechanics – Computational Cost

<table>
<thead>
<tr>
<th>Level of Theory</th>
<th>Approx. Scaling</th>
</tr>
</thead>
<tbody>
<tr>
<td>HF</td>
<td>$N^3$ iterative</td>
</tr>
<tr>
<td>Density Functional Theory</td>
<td>$N^3$ iterative</td>
</tr>
<tr>
<td>MP2</td>
<td>$N^3$ non-iterative</td>
</tr>
<tr>
<td>MP3, MP4(SDQ)</td>
<td>$N^5$ non-iterative</td>
</tr>
<tr>
<td>CISD, CCSD, QCISD</td>
<td>$N^6$ iterative</td>
</tr>
<tr>
<td>MP4(SDTQ)</td>
<td>$N^7$ non-iterative</td>
</tr>
<tr>
<td>MP5</td>
<td>$N^8$ non-iterative</td>
</tr>
<tr>
<td>CISDT, CCSDT, QCISDT</td>
<td>$N^8$ iterative</td>
</tr>
<tr>
<td>MP6</td>
<td>$N^{10}$ non-iterative</td>
</tr>
<tr>
<td>MP7</td>
<td>$N^{10}$ non-iterative</td>
</tr>
<tr>
<td>CISDTQ, CCSDTQ, QCISDTQ</td>
<td>$N^{10}$ iterative</td>
</tr>
</tbody>
</table>

$N =$ number of basis functions

Quantum Mechanics – Resources required

- In general, expect quantum-level calculations to require a lot of resources: cpu, memory, and disk.
- Level of theory and type of calculation is extremely important in determining the resources needed.
  - Post-SCF (not HF or DFT) will be very cpu, memory and disk intensive.
- Some levels of theory will parallelize better than others.
  - Post-SCF (not HF or DFT) historically have not parallelized well.
- Some types of calculations will parallelize better than others.
- A level of theory may run efficiently and parallelize well in one program but horribly in another program.
Resources required
Will your calculation run with more than 1 core?
Does it require shared memory?
How many cores should I use? How much memory do I need? Disk space?

<table>
<thead>
<tr>
<th>Cores</th>
<th>GGA Functional</th>
<th>Hybrid Functional</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>106</td>
<td>1571</td>
</tr>
<tr>
<td>32</td>
<td>60</td>
<td>1004</td>
</tr>
<tr>
<td>64</td>
<td>41</td>
<td>621</td>
</tr>
<tr>
<td>128</td>
<td>25</td>
<td>566</td>
</tr>
<tr>
<td>256</td>
<td>22</td>
<td>685</td>
</tr>
</tbody>
</table>

Parallelization efficiency
Benchmarking to determine required resources

- Get guidance from group members if possible: Ask group members if someone has tested the efficiency of the code and level of theory on the machine that you are using and hope that they are willing to share their knowledge.
- See if there is guidance in the program manual
- Run your own “benchmarking” calculations
  - Run short representative jobs using increasing number of cores and calculate the speed-up using the WALLTIME
- Caveats
  - Overhead may dominate the walltime for very short jobs
  - Competition with other users on the system may skew your results
  - Multi-node job timing can be effected by job placement
- IMPORTANT: Just because all of the cpus are busy for your parallel job does NOT mean that you are running an efficiently!
- Seek guidance from HPRC (help@hprc.tamu.edu)

### Software – Molecular Mechanics/Dynamics

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Licensing</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMBER</td>
<td>Molecular Mechanics/Dynamics (MM/MD)</td>
<td>Site License</td>
<td>$</td>
</tr>
<tr>
<td>GROMACS</td>
<td>MM/MD specializing in speed and coarse-grained simulations</td>
<td>Site License*</td>
<td>Free</td>
</tr>
<tr>
<td>LAMMPS</td>
<td>MM/MD</td>
<td>Site License*</td>
<td>Free</td>
</tr>
<tr>
<td>NAMD</td>
<td>MM/MD</td>
<td>Site License*</td>
<td>Free</td>
</tr>
</tbody>
</table>
Molecular Mechanics (classical) – Resources required

- In general, molecular mechanics (classical) calculations will not need a lot of memory or disk space.
- The larger the system the more cores you will need to run in a reasonable amount of time.
- Most simulations have thousands of atoms
- Really big job… namd 64 million atoms… http://www.ks.uiuc.edu/Research/HIV/
  http://dx.doi.org/10.1038/nature12162

Parallelization Efficiency
What about GPUs?

- Quantum level calculations are not well suited for running on GPUs
- Classical calculations are well suited for running on GPUs
  - Generally, only the non-bonding interactions are calculated using the GPU and all other calculations are performed using the CPUs

ADA Gaussian Job file

```bash
#!/bin/bash
#
#BSUB -L /bin/bash -W 10:00 -n 20 -R span[ptile=20] -J test -e test.job.e%J -o test.job.o%J
#BSUB -R "rusage[mem=2700]"
#BSUB -M 2700

cd $TMPDIR
export g09root=/software/lms/g09_D01
 ./g09root/g09/bsd/g09.profile

echo -P 20 > Default.Route
echo -M 40GB >> Default.Route

module purge
env
date
g09 < $LS_SUBCWD/test.com > $LS_SUBCWD/test.log
date
exit
```
Terra Gaussian Job file

```bash
#!/bin/bash
#
#SBATCH -J test -e test.job.e%j -o test.job.o%j
#SBATCH -t 24:00:00 -n 28 --mem=56G
#SBATCH --ntasks-per-node=28 # always specify 1 node for gaussian jobs

cd $TMPDIR
export g09root=/sw/group/lms/sw/g09_D01
    . $g09root/g09/bsd/g09.profile

module purge
eval

g09 < $SLURM_SUBMIT_DIR/test.com > $SLURM_SUBMIT_DIR/test.log
date
exit
```

Terra NWChem Job file

```bash
#!/bin/bash
#SBATCH -J 601          # job name
#SBATCH --o 601.job.o%j
#SBATCH --e 601.job.e%j
#SBATCH --n 84              # total number of mpi tasks requested
#SBATCH --mem=56000
#SBATCH --ntasks-per-node=28
#SBATCH -t 24:00:00        # run time (hh:mm:ss)

cd $SLURM_SUBMIT_DIR
module purge
module load intel/2016b

export NWCHEM_TOP=/sw/group/lms/nwchem-6.6
export NWCHEM_BASIS_LIBRARY=/sw/group/lms/nwchem-6.6/src/basis/libraries/

mpirun $NWCHEM_TOP/bin/LINUX64/nwchem 601.nw &> 601.out

date
exit
```