Introduction to Computational Chemistry

CH20238 / CH30239

Dr Claire L. McMullin and Dr Vera Krewald

WH 1.24 *WH 1.28*

Computational Chemistry

- ... is not just chemistry done "*using*" a computer

- As computing power has advanced and there are many more "user-friendly" software packages almost anyone who fancies a go can run a calculation (even if some of them maybe shouldn't!)

- The main goal of this course is to gently introduce the different methodologies available, present their advantages and disadvantages, to identify the best methods to use depending on what the chemistry question / system is
  - *… meaning you can judge the suitability of computational studies when you come across them*

Objectives

- Understand the concept of an energy surface and how it relates to chemical behaviour and reactivity, and the role of computational modelling in bridging between theory and experiment

- Demonstrate basic understanding of energy minimisation, geometry optimisation, relative energies, isodesmic relations, vibrational frequencies and molecular dynamics

- Appreciate the operational working of molecular mechanics, molecular orbital (Hartree-Fock and semi-empirical) and density functional theory methods within a typical computational chemistry software package and the significance of basis sets and empirical parameters in context

- Appreciate the advantages and disadvantages of popular computational methods for different problems

Topics

- **Molecular mechanics** and **empirical force-field** methods

- *Potential energy surfaces*, *energy minimisation* and *geometry optimisation*

- Practical application of **molecular orbital** and **density functional** methods

- *Electron density*, *electronic charge* and *electrostatic potential*

- *Molecular vibrations*, *zero-point energy*, *free energy*

- *Transition structures* and *reactivity*

- **Molecular dynamics**

Your Next Five Weeks

10 Lectures

      W19-23; every Thursday (11:15 4E 3.10) and Friday (13:15 CB 2.6)

- Molecular Mechanics (MM)
- Semi-Empirical Methods
- Periodic Boundary Conditions (PBC)
- Molecular Dynamics (MD)

3 Lectures (CLM)

- Hartree-Fock (HF)
- Density Functional Theory (DFT)
- Basis Sets and Pseudopotentials

5 Lectures (VK)

- Computed vs. Real Life

2 Lectures (CLM)

Your Next Five Weeks

Workshops

W22-24; either Wednesday (12:15 CB 3.1) or Friday (14:15 CB 5.1)

- W22   Multiple Choice Question Practice

- W23   *Student Drop-In?*

- W24   Evaluating Methods Practice

Textbooks

- Level 4 of the Library "541.28" – under the Group Theory books!

  - Introduction to Computational Chemistry   **1** (and online)
    - Frank Jensen, *Wiley* (2017 3rd Edition, 2007 2nd Edition)

  - Molecular Modeling Basics   **4** (and online)
    - Jan Jensen, *Routledge* (2010)

  - Essentials of Computational Chemistry: Theories and Models   **1**
    - Christopher Cramer, *Wiley* (2004 2nd Edition)

  - Molecular Quantum Mechanics   **4**
    - Peter Atkins & Ronald Friedman, *Oxford University Press* (2010)

Textbook

- Less relevant textbooks:

    - Chemical Applications of Molecular Modelling   **1**
        - Jonathan Goodman, *RSC* (1998)

    - Molecular Modeling of Inorganic Compounds   **1**
        - Peter Comba, Trevor Hambley & Bodo Martin, *Wiley* (2009)

    - Structural Methods in Molecular Inorganic Chemistry   **1**
        - David Rankin, Norbert Mitzel & Carole Morrison, *Wiley* (2013)

- Advanced reading:

    - Modern Quantum Chemistry – Introduction to Advanced Electronic Structure Theory
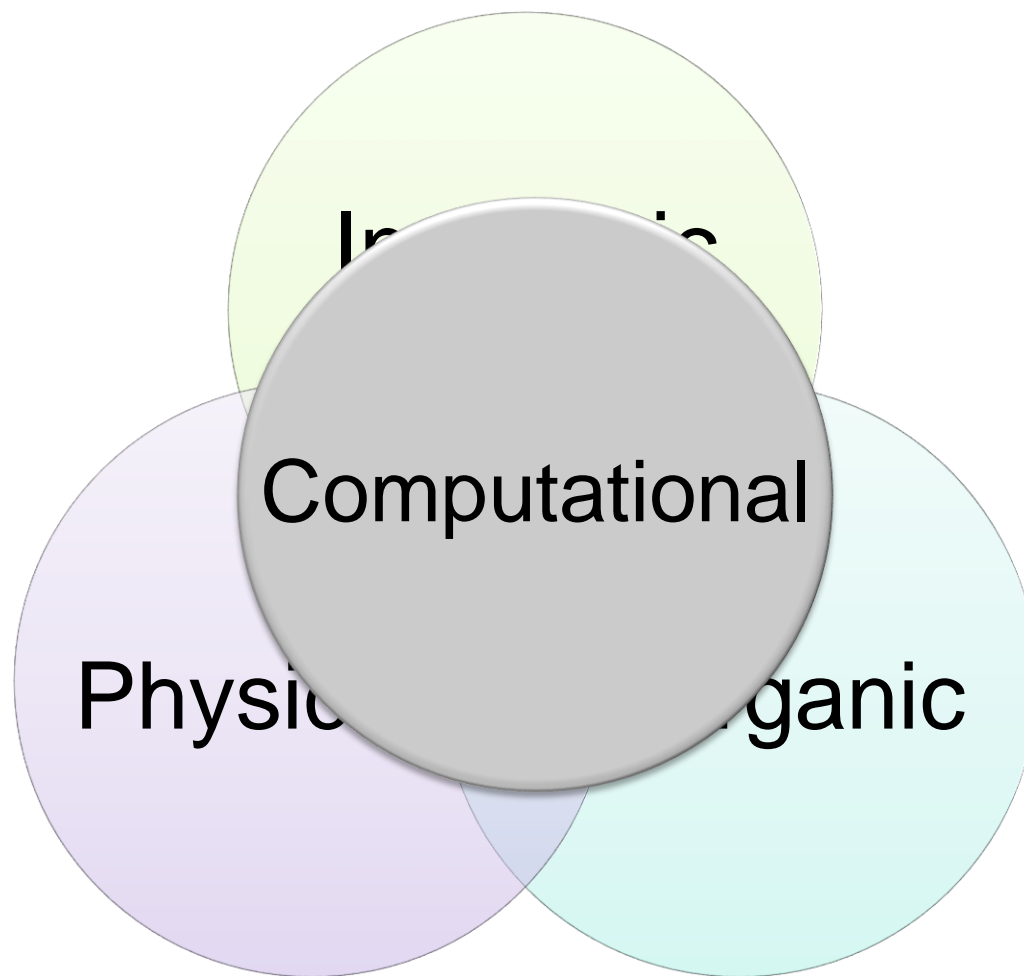   - Attila Szabo & Neil S. Ostlund, *Dover Publications Inc.* (1989)

Examinations

- Exam   100 %

- Three Sections:

    1. Multiple Choice Questions

    2. "Medium"-length Questions

    3. Essay-like Question

As this is the first time the unit has run, we've made a mock paper for you – because there aren't any past papers!

Research Themes at UoBath

Research Themes at UoBath

- Catalyst design for renewable resources

- Metal-organic frameworks and metal-protein interactions

- Biological enzyme mechanisms

- Solid-state batteries and perovskite solar cells

- N_2 / H_2O catalyst properties (redox, magnetic, spectroscopic)

- Organometallic reaction mechanisms; e.g. cross coupling and C-H activation

- Ion transport through solids (Li batteries)

- Structure / stability of ceramics and minerals

- Reactions in solution and enzyme active sites

Antoine Buchard

Rob Deeth

Carmen Domene

Saiful Islam

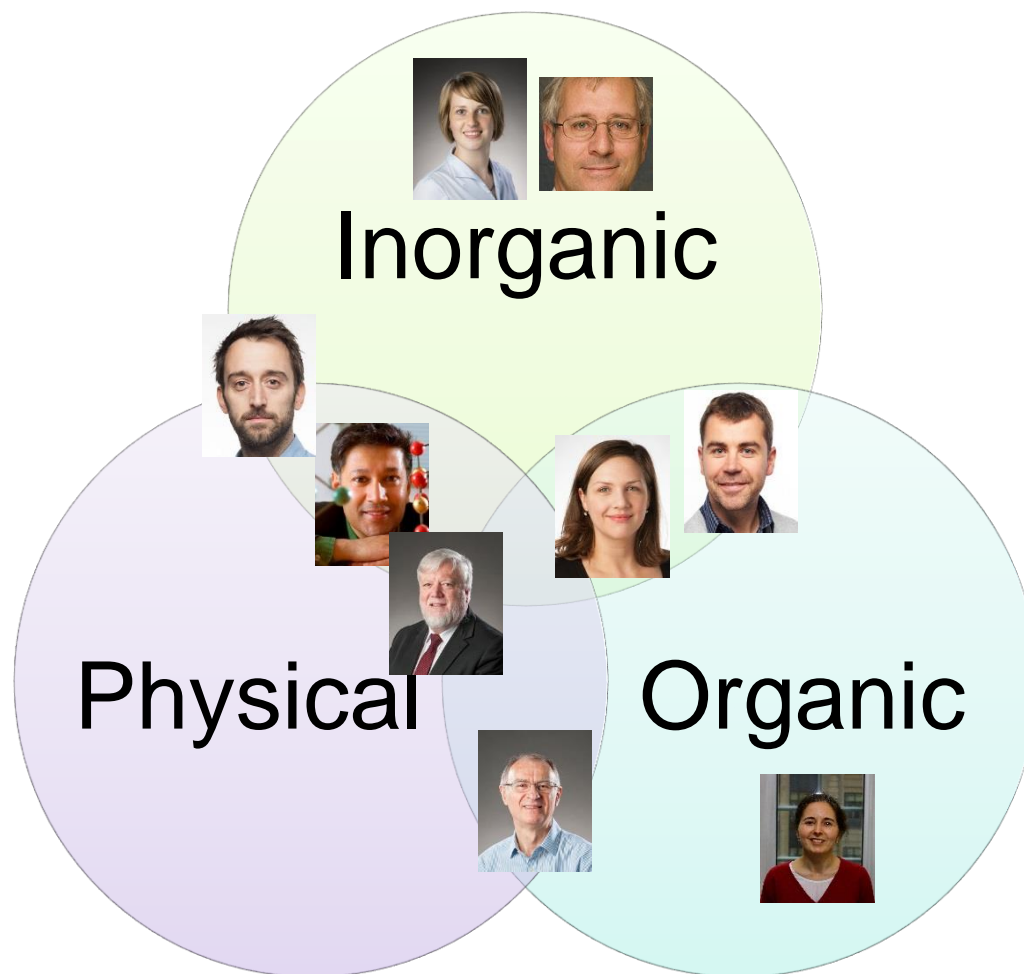**Vera Krewald**

Claire McMullin

Ben Morgan

Steve Parker
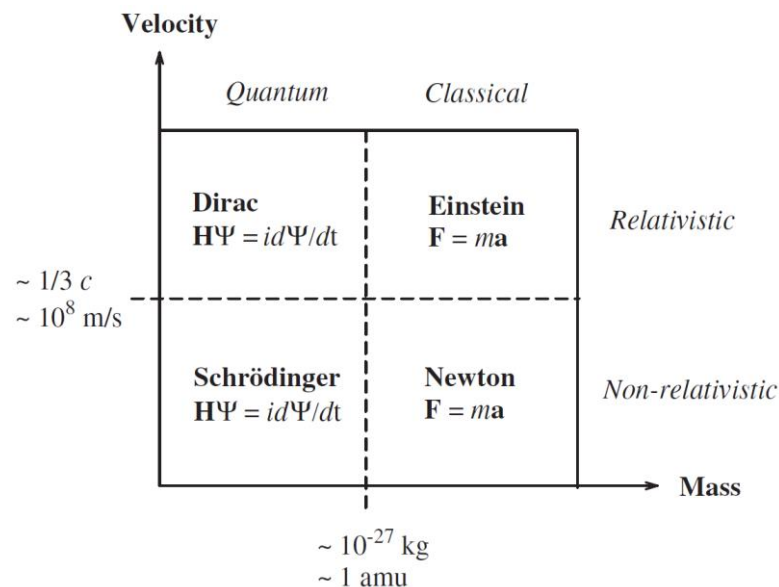
Ian Williams

Research Themes at UoBath

Previously, In Quantum Mechanics…

- Prof. Steve Parker's CH20147 lectures at the end of S1

- Topics to familiarise yourself with for this course:

  - Wave-particle duality
  - Wavefunctions, operators, eigenvalues
  - Particle in a box
  - $\widehat{H}\boldsymbol{\Psi} = \mathrm{E}\boldsymbol{\Psi}$
  - Zero point energy
  - Ground states and excited states
 - Variation method
  - Expectation value

Dynamical Equations

- Mathematical form depends on <u>mass</u> and <u>velocity</u> of the particle(s)
- Classical mechanics is ***deterministic*** – it can be integrated over time
- QM is ***probabilistic*** – which is square of the wave function (Ψ^2)

- Newton = heavy and slow
- Einstein = velocity ~ speed of light
  - increasing mass relative to rest mass

- Borderline mass = a proton (a H nucleus)
- Electrons much lighter = only QM
- Light particles have <u>*wave*</u> and <u>*particle*</u> characteristics
- Hydrogen atoms are special – sometimes they need QM corrections (e.g. tunnelling through barriers)

Velocity

| Quantum | Classical |

Dirac
$H\Psi = id\Psi/dt$

Einstein
$F = ma$

Relativistic

~ 1/3 c
~ 10^8 m/s

Schrödinger
$H\Psi = id\Psi/dt$

Newton
$F = ma$

Non-relativistic

Mass

~ 10^{-27} kg
~ 1 amu

Potential Energy Surface (PES)

- Connection between experimental data and molecular modelling

- Reactants, products, stable intermediates and transition states (TS) are found along a reaction coordinate (RC) on the PES

Molecular Modeling Basics, Jan Jensen, *Routledge* (2010)
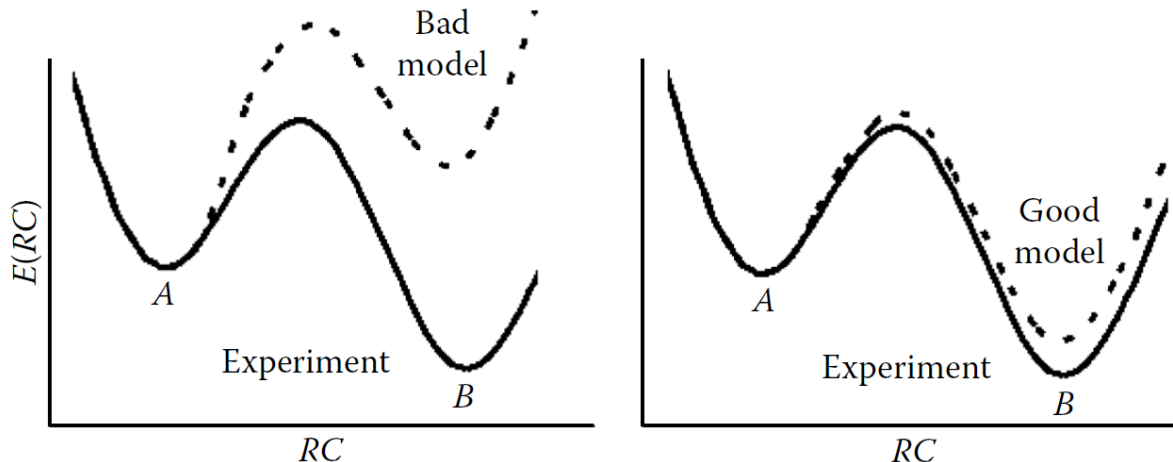
Potential Energy Surface (PES)

- Connection between experimental data and molecular modelling

- Reactants, products, stable intermediates and transition states (TS) are found along a reaction coordinate (RC) on the PES

- Stationary points; when gradient is zero $$\frac{dE}{dR} = 0$$

- Second energy derivatives identify what type of stationary point has been found:

  - $\frac{d^2E}{dR^2} > 0$; minima   stable intermediates

  - $\frac{d^2E}{dR^2} < 0$; maxima   TSs

Remember: a TS has a negative (imaginary) frequency as it is a maximum
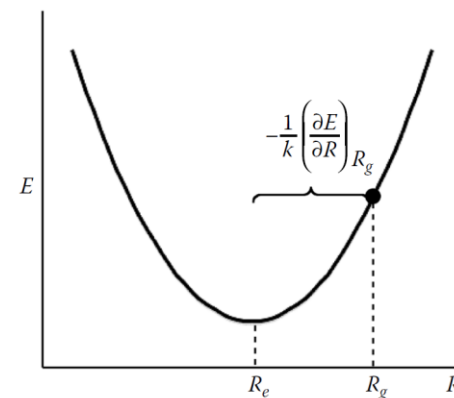
What is an Optimisation?

- Finding the best position (x, y, z) for all atoms in the molecule / structure

- Trying to get to the bottom of the "well"

- E.g. for a quadratic PES

  - $E = \frac{1}{2}k(R - R_e)^2$

  - R_e is the value of R at which the energy is lowest – the **equilibrium geometry**
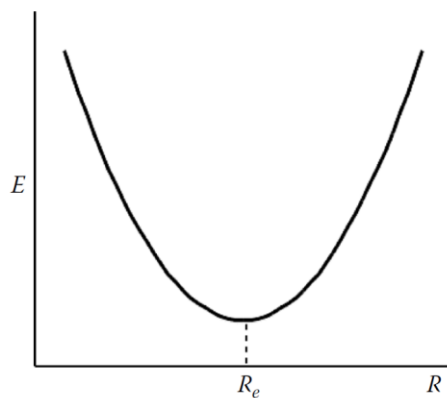

Work out gradient

Tells you which direction the energy is *decreasing*

Take small steps in that direction

$$-\frac{1}{k}\left(\frac{\partial E}{\partial R}\right)_{R_g}$$

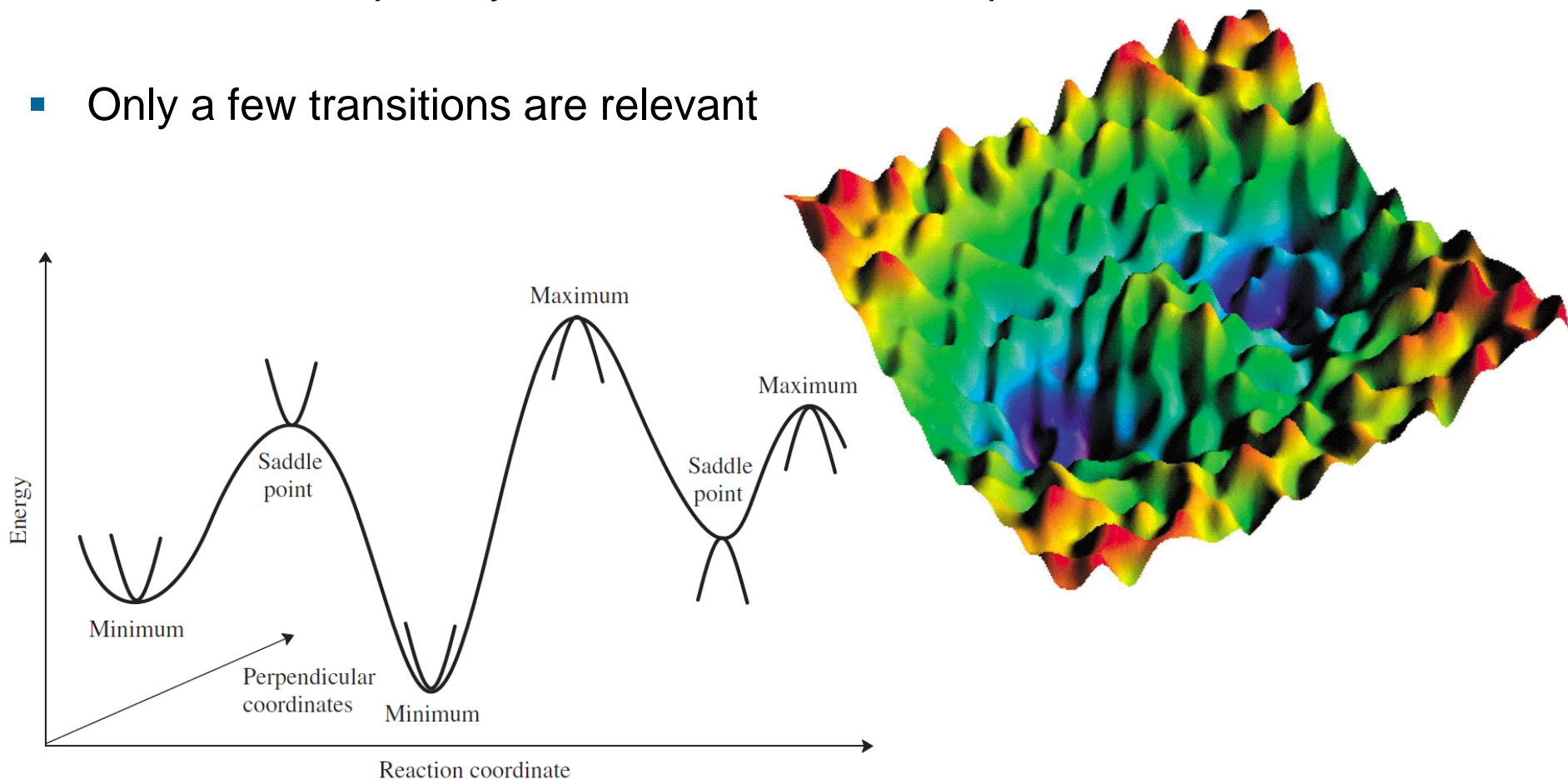**Molecular Modeling Basics**, Jan Jensen, *Routledge* (2010)

Minimum Energy Pathway

- This connects a transition state with its two closest minima

- The guess structure is the TS displaced along the imaginary frequency mode

- The resulting collection of structures is known as the intrinsic reaction coordinate (IRC)

Molecular Modeling Basics, Jan Jensen, *Routledge* (2010)
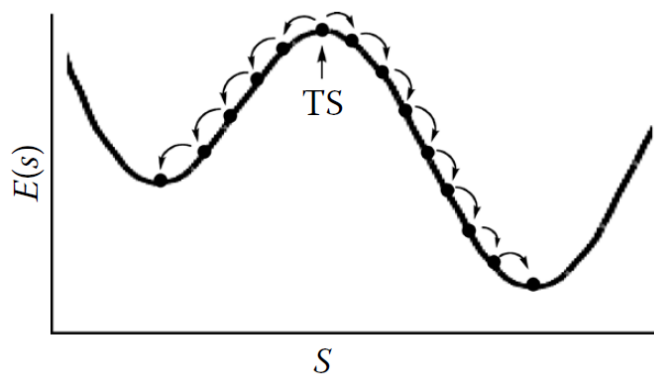
Multidimensional Surface

- PES for a complex system is dense in saddle points

- Only a few transitions are relevant

Introduction to Computational Chemistry
Frank Jensen, *Wiley* (2017)

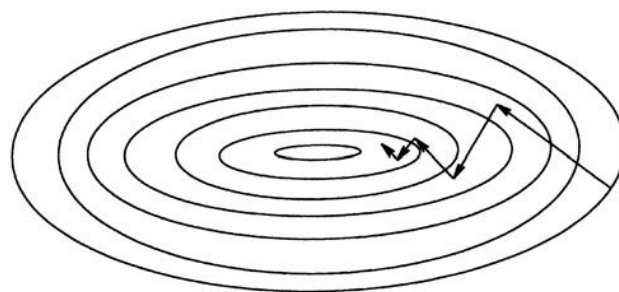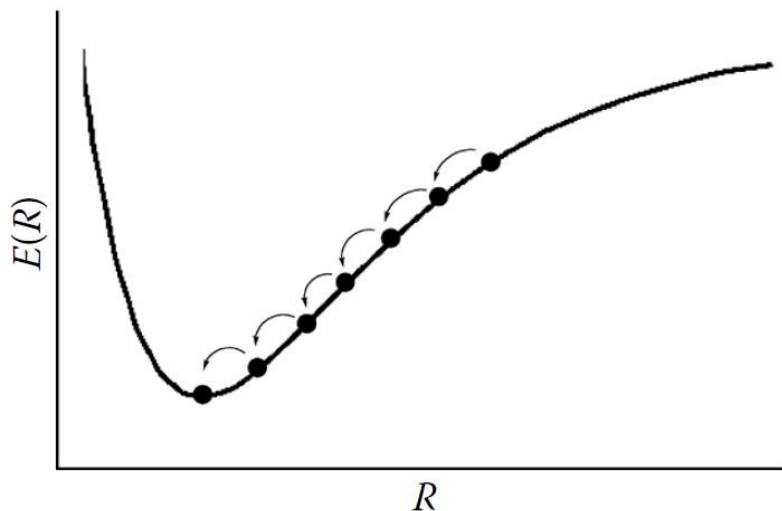**Annu. Rev. Phys. Chem.**, P. G. Bolhuis *et al*, *53*, 291-318 (2002)

Steepest Descent

- 1st order method
- Moves downhill to calculate forces (gradients)
- Relieves strain in built model
- Slow convergence
- Can't locate TSs

$$\underline{x}_{k+1} = \underline{x}_k + \lambda_k \underline{s}_k$$

New co-ord.

Initial co-ord.

Direction

Stepsize

Means you are minimising the energy to find the closest minimum to the <u>starting geometry</u>

Quadratic / Newton-Raphson step

- 2nd order method
- Reduces computational cost
  - One instead of many steps
  - Works out k (2nd derivative of E)
  - Uses the Hessian matrix
- Can locate TSs as Hessian provides uphill direction
- Works if the surface is quadratic
  - i.e. close to the minimum
  - but not if far away

$$\underline{x}_{k+1} = \underline{x}_k - gH^{-1}$$

New co-ord.   Initial co-ord.   inverse of Hessian

gradient

The Hessian matrix is the 3N x 3N diagonalised square matrix of second-order partial derivatives of a function

**Molecular Modeling Basics**, Jan Jensen, *Routledge* (2010)
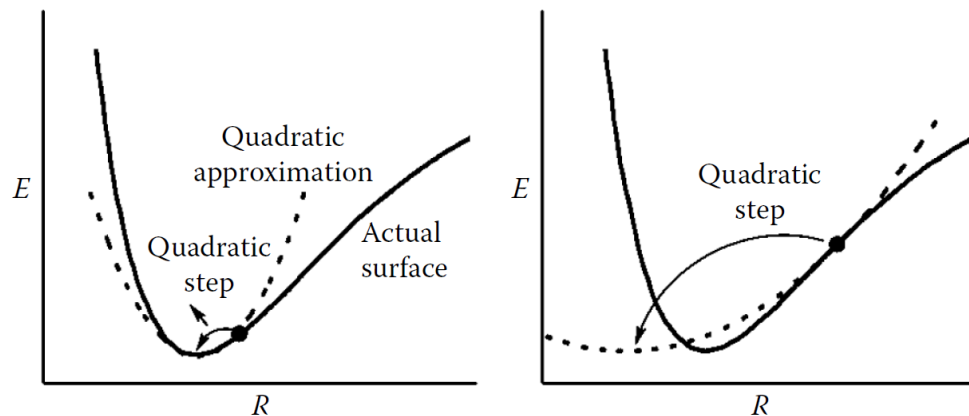
Geometry Optimisation

- So we use an energy minimisation to find the optimal geometry for a structure

- Optimisations end either when the stationary point has been found (calculation has converged) or the program runs out of steps
  - Number of steps is often linked to number of atoms in the system
  - Each step explores one degree of freedom of the structure

- If the PES is flat, likely will require more steps to converge
  - "Flat" means small gradients and so a smaller stepsize
  - PES tends to be nonquadratic too - so can't use Newton-Raphson

Global Minimum

- Danger that optimisation may trap you in a local minimum – not the global minimum

- Can try many different starting geometries
    - i.e. a **conformational search**

- Programs typically achieve this by changing dihedral angles of rotatable bonds

Molecular Modeling Basics, Jan Jensen, *Routledge* (2010)
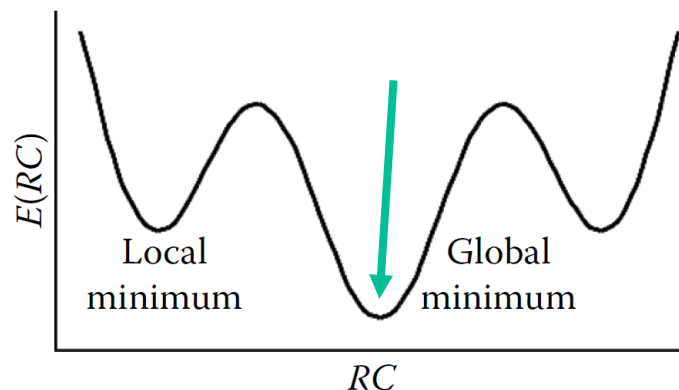
Monte Carlo Method

- A class of computational algorithms, that involve repeated random sampling to compute results (to satisfy Boltzmann distribution)

- Especially useful when trying to simulate systems with many coupled degrees of freedom; *i.e.* fluids, disordered materials, solids and cellular structures

- For models with significant uncertainty in the input

- Unbiased approach to conformational searching
  - A random "kick" is given to coordinates of a randomly chosen atom / molecule

Monte Carlo (MC)

1. Energy comparison made between structure (E_1) and random distortion (E_2)
2. If $E_2 < E_1$, then the second geometry is a better candidate for the global minimum and excepted
3. Process 1-2 repeated
4. If $E_2 > E_1$, then the second geometry is a worst candidate and is rejected if geometry 2 is much higher in energy

- Energy of system not fixed – selected according to Boltzmann probability

$$P(x) = \frac{\exp(-U(x)/kT)}{\int \exp(-U(x)/kT)dX}$$

Zero-Point Energy

- At the equilibrium structure the atoms are (hypothetically) at rest and free from the effects of vibration

- Energy difference between vibrational ground state and the equilibrium structure
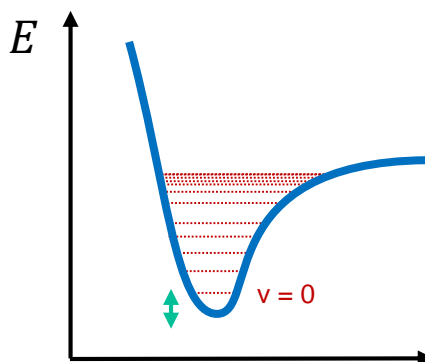
$$E_{ZPE} = \sum_i \frac{1}{2} \hbar \omega_i$$

- "Simply" add up the $\frac{1}{2} \hbar \omega_i$ contribution from each vibration mode i
  - Multiple by Avogadro's number to report energies in units of kJ mol^{-1}

- Modes that contribute the most to the ZPE correction have a higher wavenumber e.g. hydrogen bond modes – due to reduced mass term

Z Matrix

- Describes the geometry of a molecule not by Cartesian coordinates (x, y, z), but by the atom <u>connectivity</u>
  - An *internal coordinate representation*

- It identifies each atom in a molecule by a bond distance, bond angle and dihedral angle in relation to other atoms in the molecule

- Can allow symmetry to be enforced upon the molecule
  - Speeds up parts of the calculation

- Example: CH_4 (methane)

```
C   0.000000   0.000000   0.000000
H   0.628736   0.628736   0.628736
H  -0.628736  -0.628736   0.628736
H  -0.628736   0.628736  -0.628736
H   0.628736  -0.628736  -0.628736
```

```
C
H 1 1.089000
H 1 1.089000 2 109.4710
H 1 1.089000 2 109.4710 3 120.0000
H 1 1.089000 2 109.4710 3 -120.0000
```

Computing Power

Quality of e- correlation method

Slow High quality

Fast Low quality

Quality of basis set
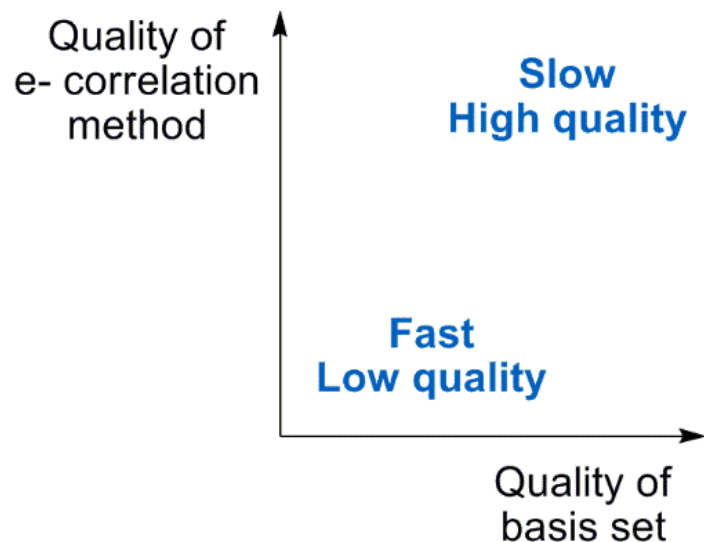
The "cost" of a calculation isn't defined by how expensive the machine is that you use, but by **how long** the calculation will take to run – i.e. how many processors will be used and how long you have to wait before convergence. This is what we mean when we refer to the computational cost!

Computing power keeps increasing:

When I was doing my PhD (2007-2010) the biggest system I worked on had ~ 80 atoms and took 2 weeks to optimise using DFT.

Now running on the HPC in Bath (2018) I can optimise a 150 atom structure in a day!

Increasing Computational Quality

- Increase quality of electron correlation method – increases computing cost
    - Depending on method it scales N^3 (DFT), N^4 (HF) or even higher, when N is the number of atoms in the system

- Increase quality of basis set (more functions used to describe electrons / orbitals) – increases computing cost
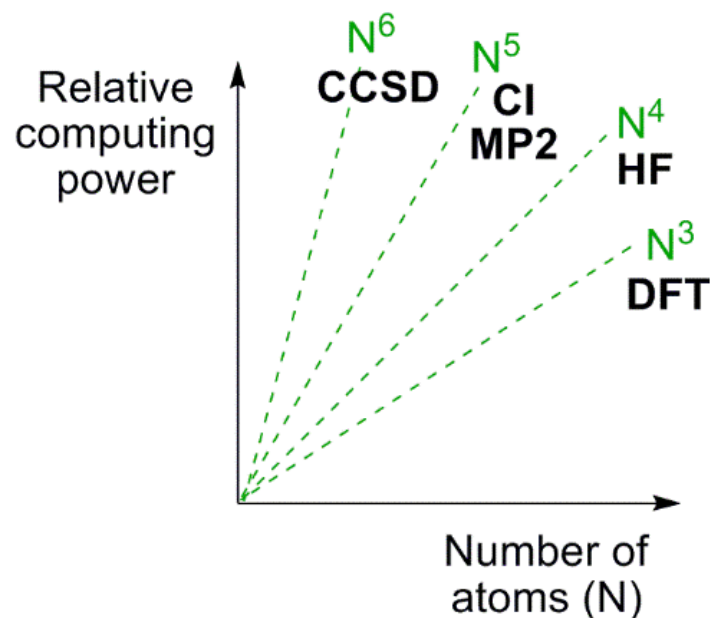
DFT = Density Functional Theory
HF = Hartree Fock
CI = Configuration Interaction
MP2 = Møller Plesset 2
CCSD = Coupled Cluster (Single Double)

Jacob's Ladder

- A staircase to heaven (from a dream of Jacob described in the Old Testament)

- Also refers to improvements in density functional approximations (DFT) to the exchange-correlation energy

- Accuracy increases as you go up the ladder

HEAVEN OF CHEMICAL ACCURACY

unoccupied $\{\phi_i\}$	generalized RPA
ε_x	hyper-GGA
τ and/or $\nabla^2 n$	meta-GGA
∇n	GGA
n	LSD

HARTREE WORLD
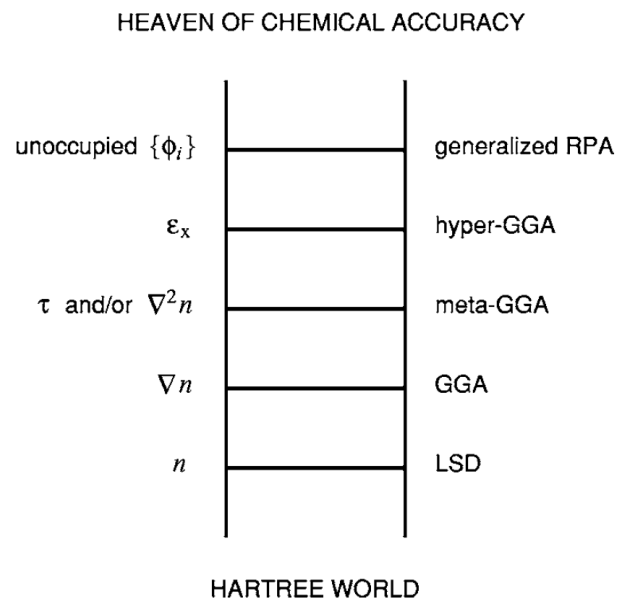
One Method Does Not Fit All

- Different questions require different methods…

- Variables to consider:
    - The size of the system
    - Accuracy required
    - What questions you are trying to answer
    - Solid, liquid or gas?
    - Electronic, structural or dynamic?

Programs

- Lots of different software packages that use specific or a variety of methodologies:

  - SPARTAN
  - Gaussian / GaussView
  - Jaguar
  - ORCA
  - Turbomole
  - PCModel
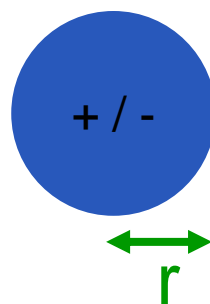 - CHARMM

Molecular Mechanics

What is Molecular Mechanics?

- Classical Newtonian mechanics used to model molecular systems
- Potential Energy calculated using force fields
- Can model small molecules as well as large biological systems (N > 1000)

- Simplifications:

  - Atom = point charge

    + / -

    r

  - Bond = spring
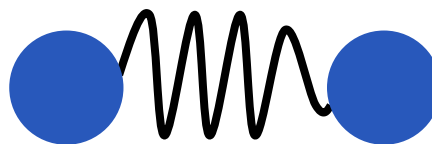
No electrons – so can not treat a chemical reaction

The Forcefield

$$V(r) = \sum_{bonds} \frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{angles} \frac{k_i}{2}(\theta - \theta_i)^2 + \sum_{torsions} \frac{V_n}{2}(1 + \cos(n\omega - \gamma))$$

Total Potential Energy

BONDS

ANGLES

DIHEDRAL

$$+ \sum_{i=1}^{N} \sum_{j=1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \right) + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

Lennard Jones potential (VdW)

Electrostatic term

The Forcefield

Total Potential Energy

$$V(r) = \sum_{bonds} \frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{angles} \frac{k_i}{2}(\theta - \theta_i)^2 + \sum_{torsions} \frac{V_n}{2}(1 + \cos(n\omega - \gamma))$$

stretching   *bending*   *rotation*

$$+ \sum_{i=1}^{N} \sum_{j=1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \right) + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

repulsive   *attractive*   *charge*

NON-BONDING INTERACTIONS

V_{bond}

- l_i is bond length

- $l_{i,0}$ is equilibrium bond length

- k_i is force constant

- Harmonic term

$$\sum_{bonds} \frac{k_i}{2} (l_i - l_{i,0})^2$$

Problem: bonds are "springs" so does not model bond dissociation
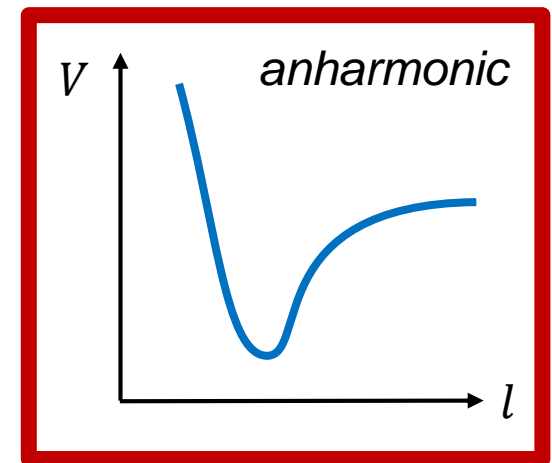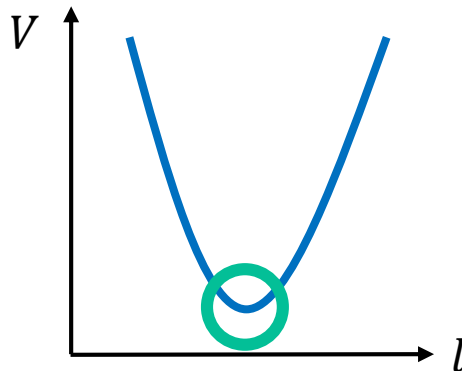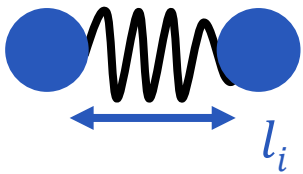
V$_{\text{angle}}$

- θ is bond angle

- θ_i is equilibrium bond angle

- Also a harmonic term

- Less energy needed to distort angle from equilibrium value – hence force constants are smaller

$$\sum_{angles} \frac{k_i}{2}(\theta - \theta_i)^2$$
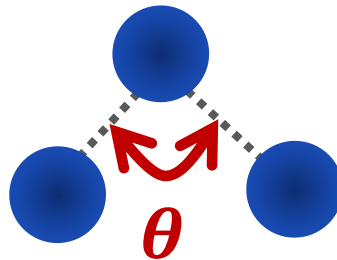
V$_{\textbf{dihedral}}$

- n is the multiplicity

- γ is the phase factor

- ω is the torsion angle

$$\sum_{torsions} \frac{V_n}{2}(1 + \cos(n\omega - \gamma)$$
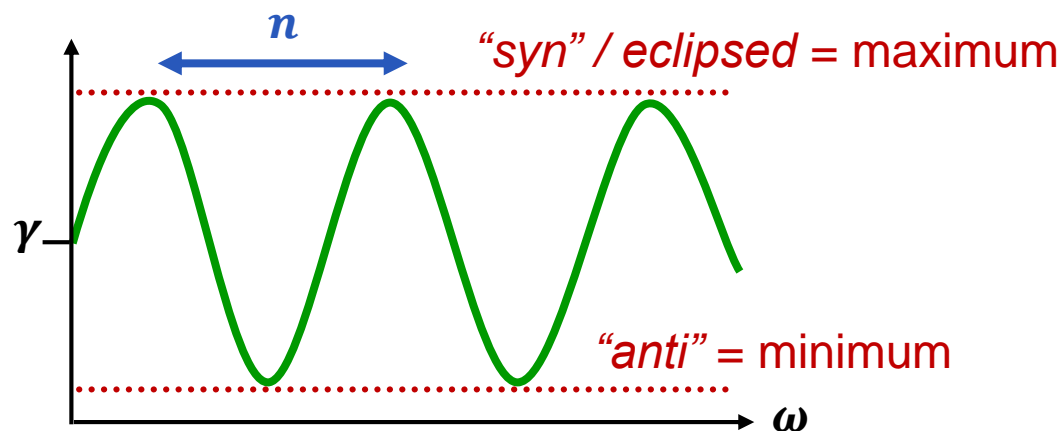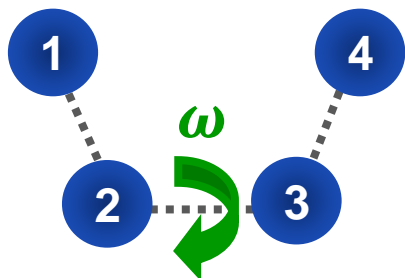
Looking down / along the 2-3 bond:

$0°$ **syn**

$180°$ **anti**

$360°$ **syn**
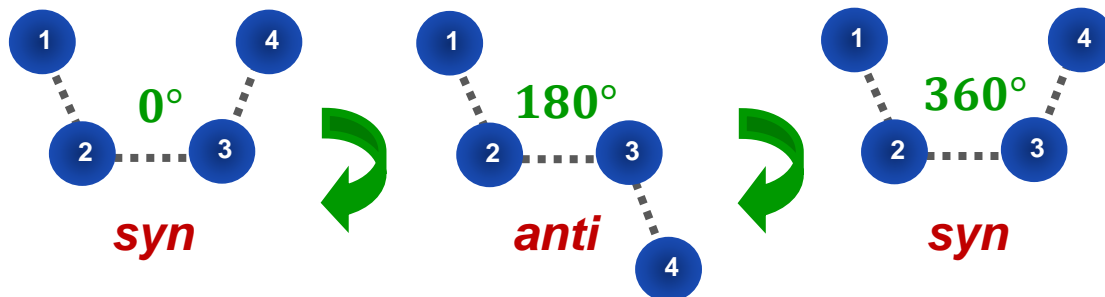
Improper Torsions

- Improper torsions use the cosine potential to restrain angle between 0°-180°

$$\sum_{improper} \frac{V_2}{2}[1+\cos(2\omega - 180)]$$

- Or a harmonic function

$$\sum_{improper} \frac{k_w}{2}(\omega - \omega_0)^2$$

- **Energy of out of plane functions**

- Not a normal 1-2-3-4 connectivity, but 3 atoms bonded to an atom in same plane e.g. sp^2 centre such as carbon in carbonyl or aromatic rings

    - A "normal" torsion parameter doesn't keep planarity at 0°
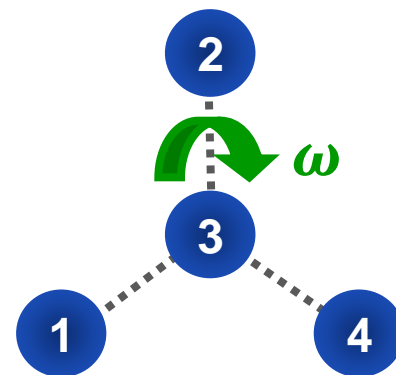
V~VdW~
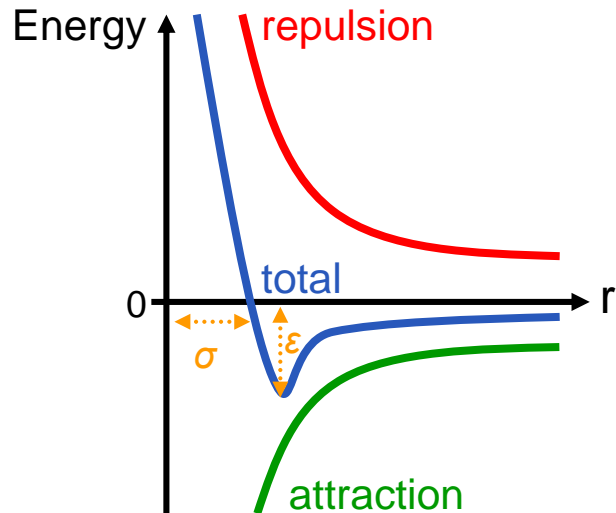
Lennard-Jones 12,-6 potential

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \right)$$

- r_{ij} is distance between two atoms *i* and *j*

- ε_{ij} is the well depth (of the energy minimum)

- σ_{ij} is the collision diameter (where interaction energy = 0)

V_{VdW}
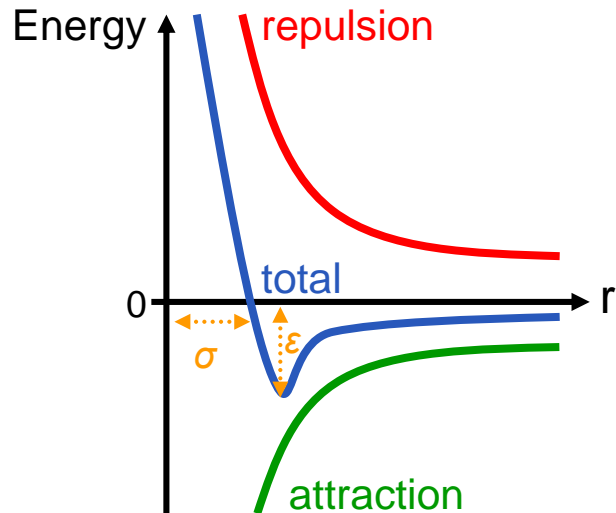
Lennard-Jones 12,-6 potential

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \right)$$

- **Attraction** (dispersion) and short-range **repulsion**

- **Dispersion interactions ~ 1 / r^6**
  - Attractive due to instantaneous dipoles from fluctuations of electronic distributions

- **Repulsive interactions ~ 1 / r^{12}**
  - At very short distances molecules repel each other – insignificant over large distances.
  - Pauli Principle: ↓e$^-$ density between nuclei = ↑ nuclear-nuclear repulsion

V_{electrostatic}

- q_i and q_j are the charges of the atoms

$$\frac{q_i q_j}{4\pi \varepsilon_0 r_{ij}}$$

- ε_{ij} molecular dielectric expression – accounts for attenuation of electrostatic interaction by the environment

- Coulombic potential

- Influence of polarity on energy and structure

- A linearly varying distance-dependent dielectric $\frac{1}{r_{ij}}$ accounts for increase in environmental bulk as the separation distance between interacting atoms increases
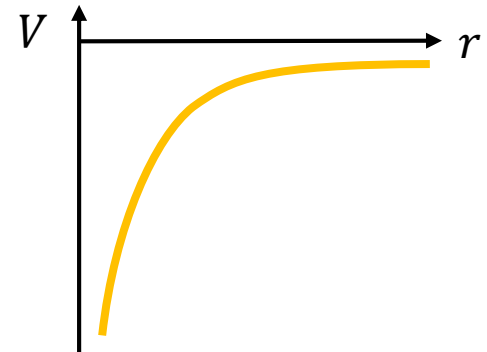
Empirical Potential Energy Function

Bonds
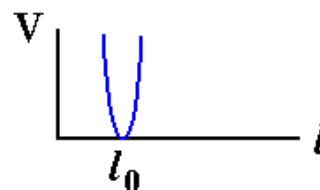
Angles
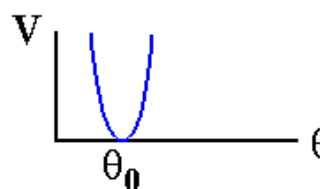
Improper Dihedrals
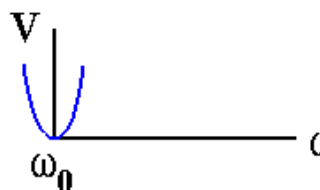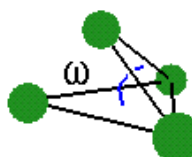
Torsions
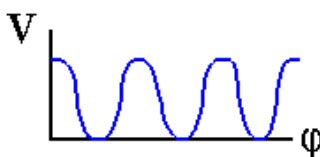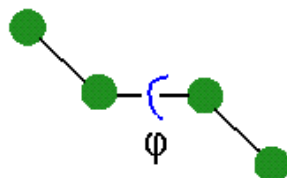
Electrostatics
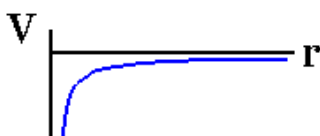
van der Waals
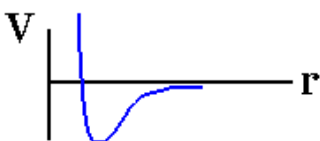
The MM Forcefield is Collectively:

- Potential Functions
- Force constants
- VdW multipliers
- Other constant terms
- Equilibrium bond, angle & dihedral values
- Partial charge values
- Atomic masses and radii

Lets actually have a look at one…

```
      ################################
      ##                            ##
      ##   Force Field Definition   ##
      ##                            ##
      ################################


forcefield              MMFF94

bondunit                71.9663
bond-cubic              -2.0
bond-quartic            2.333333        !! (7/12) * bond-cubic^2
angleunit               0.021922
angle-cubic             -0.007
str-bndunit             2.51210
torsionunit             0.5
vdwtype                 BUFFERED-14-7
radiusrule              CUBIC-MEAN
radiustype              R-MIN
radiussize              DIAMETER
epsrule                 HHG
vdw-14-scale            1.0
chg-14-scale            0.75
dielectric              2.0


      ################################
      ##                            ##
      ##   Literature References    ##
      ##                            ##
      ################################


T. A. Halgren, "Representation of van der Waals (vdW) Interactions in
Molecular Mechanics Force Fields: Potential Form, Combination Rules,
and vdW Parameters", JACS, 114, 7827-7843 (1992)


      ################################
      ##                            ##
      ##   Atom Type Definitions    ##
      ##                            ##
      ################################


atom        1    C    "Alkyl Carbon"        6    12.000    4    4    1    1    1
atom        2    C    "Olefinic Carbon"     6    12.000    3    4    2    2    1
atom        3    C    "Carbonyl Carbon"     6    12.000    3    4    3    3    1
atom        4    C    "Acetylenic Carbon"   6    12.000    2    4    4    4    1
atom        5    H    "Nonpolar Hydrogen"   1     1.008    1    1    5    5    5
```

```
40        ############################
41        ##                        ##
42        ##  Atom Type Definitions  ##
43        ##                        ##
44        ############################
45
46
47   atom      1    C    "Alkyl Carbon"          6    12.000    4    4    1    1    1
48   atom      2    C    "Olefinic Carbon"       6    12.000    3    4    2    2    1
49   atom      3    C    "Carbonyl Carbon"       6    12.000    3    4    3    3    1
50   atom      4    C    "Acetylenic Carbon"     6    12.000    2    4    4    4    1
51   atom      5    H    "Nonpolar Hydrogen"     1     1.008    1    1    5    5    5
52   atom      6    O    "Divalent Oxygen"       8    15.995    2    2    6    6    6
53   atom      7    O    "Carbonyl Oxygen"       8    15.995    1    2    7    7    6
54   atom      8    N    "Amine Nitrogen"        7    14.003    3    3    8    8    8
55   atom      9    N    "Imine Nitrogen"        7    14.003    2    3    9    9    8
56   atom     10    N    "Amide Nitrogen"        7    14.003    3    3   10   10    8
57   atom     11    F    "Fluorine"              9    18.998    1    1   11   11   11
58   atom     12    Cl   "Chlorine"             17    34.969    1    1   12   12   12
59   atom     13    Br   "Bromine"              35    78.918    1    1   13   13   13
60   atom     14    I    "Iodine"               53   126.900    1    1   14   14   14
61   atom     15    S    "Sulfide -S-"          16    31.972    2    2   15   15   15
62   atom     16    S    "Thiocarbonyl S=C"     16    31.972    1    2   16   16   15
63   atom     17    S    "Sulfoxide Sulfur"     16    31.972    3    6   17   17   15
64   atom     18    S    "Sulfone, Sulfate"     16    31.972    4    6   18   18   15
65   atom     19    Si   "Silicon"              14    27.977    4    4   19   19   19
66   atom     20    C    "Csp3 in 4-Ring"        6    12.000    4    4   20   20    1
67   atom     21    H    "Hydroxyl Hydrogen"     1     1.008    1    1   21   21    5
68   atom     22    C    "Csp3 in 3-Ring"        6    12.000    4    4   22   22    1
69   atom     23    H    "Amine Hydrogen"        1     1.008    1    1   23   23    5
70   atom     24    H    "Carboxyl Hydrogen"     1     1.008    1    1   24   24    5
71   atom     25    P    "Phosphate"            15    30.994    4    4   25   25   25
72   atom     26    P    "Phosphine"            15    30.994    3    3   26   26   25
73   atom     27    H    "Imine Hydrogen"        1     1.008    1    1   27   27    5
74   atom     28    H    "Amide Hydrogen"        1     1.008    1    1   28   28    5
75   atom     29    H    "Phenolic Hydrogen"     1     1.008    1    1   29   29    5
76   atom     30    C    "Csp2 in 4-Ring"        6    12.000    3    4   30    2    1
77   atom     31    H    "Water Hydrogen"        1     1.008    1    1   31   31   31
78   atom     32    O-   "Carboxylate Oxygen"    8    15.995    1    2   32    7    6
79   atom     33    H    "Sulfoxy Hydrogen"      1     1.008    1    1   33   21    5
80   atom     34    N+   "Quaternary Nitrogen"   7    14.003    4    3   34    8    8
81   atom     35    O-   "Oxide at Metal"        8    15.995    1    2   35    6    6
82   atom     36    H    "Hydrogen on N+"        1     1.008    1    1   36   36    5
83   atom     37    C    "Aromatic Carbon"       6    12.000    3    4   37    2    1
84   atom     38    N    "Pyridine Nitrogen"     7    14.003    2    3   38    9    8
85   atom     39    N    "Pyrrole Nitrogen"      7    14.003    3    3   39   10    8
86   atom     40    N    "N-C=C or N-C=N"        7    14.003    3    3   40   10    8
87   atom     41    C    "Carboxylate Carbon"    6    12.000    3    4   41    3    1
88   atom     42    N    "Nitrile Nitrogen"      7    14.003    1    3   42   42    8
89   atom     43    N    "Sulfonamide Nitrogen"  7    14.003    3    3   43   10    8
90   atom     44    S    "Thiophene Sulfur"     16    31.972    2    2   44   16   15
```

```
      ################################
      ##                          ##
      ##  Van der Waals Parameters  ##
      ##                          ##
      ################################


*13. MMFFVDW.PAR: This file supplies parameters for van der Waals interactions.
*
*          Copyright (c) Merck and Co., Inc., 1994, 1995, 1996
*                         All Rights Reserved
*
* E94  - From empirical rule (JACS 1992, 114, 7827)
* C94  - Adjusted in fit to HF/6-31G* dimer energies and geometries
* X94  - Chosen in the extension of the paratererization for MMFF94
*        by analogy to other, similar atom types or, for ions, by
*        fitting to atomic radii (and sometimes to association energies
*        for hydrates)
*
*  power    B      Beta    DARAD     DAEPS
    0.25   0.2    12.      0.8       0.5
*
*         type  alpha-i   N-i      A-i      G-i    DA
*-------------------------------------------------------------
vdwmmff     1   1.0500   2.4900   3.8900   1.2820   -
vdwmmff     2   1.3500   2.4900   3.8900   1.2820   -
vdwmmff     3   1.1000   2.4900   3.8900   1.2820   -
vdwmmff     4   1.3000   2.4900   3.8900   1.2820   -
vdwmmff     5   0.2500   0.8000   4.2000   1.2090   -
vdwmmff     6   0.7000   3.1500   3.8900   1.2820   A
vdwmmff     7   0.6500   3.1500   3.8900   1.2820   A
vdwmmff     8   1.1500   2.8200   3.8900   1.2820   A
vdwmmff     9   0.9000   2.8200   3.8900   1.2820   A
vdwmmff    10   1.0000   2.8200   3.8900   1.2820   A
vdwmmff    11   0.3500   3.4800   3.8900   1.2820   A
vdwmmff    12   2.3000   5.1000   3.3200   1.3450   A
vdwmmff    13   3.4000   6.0000   3.1900   1.3590   A
vdwmmff    14   5.5000   6.9500   3.0800   1.4040   A
vdwmmff    15   3.0000   4.8000   3.3200   1.3450   A
vdwmmff    16   3.9000   4.8000   3.3200   1.3450   A
vdwmmff    17   2.7000   4.8000   3.3200   1.3450   -
vdwmmff    18   2.1000   4.8000   3.3200   1.3450   -
vdwmmff    19   4.5000   4.2000   3.3200   1.3450   -
vdwmmff    20   1.0500   2.4900   3.8900   1.2820   -
vdwmmff    21   0.1500   0.8000   4.2000   1.2090   D
vdwmmff    22   1.1000   2.4900   3.8900   1.2820   -
vdwmmff    23   0.1500   0.8000   4.2000   1.2090   D
vdwmmff    24   0.1500   0.8000   4.2000   1.2090   D
vdwmmff    25   1.6000   4.5000   3.3200   1.3450   -
vdwmmff    26   3.6000   4.5000   3.3200   1.3450   A
vdwmmff    27   0.1500   0.8000   4.2000   1.2090   D
```

```
436    ################################
437    ##                            ##
438    ##  Bond Stretching Parameters  ##
439    ##                            ##
440    ################################
441
442
443    bond        1    1    4.2580    1.5080    C94
444    bond        1    2    4.5390    1.4820    C94
445    bond        1    3    4.1900    1.4920    C94
446    bond        1    4    4.7070    1.4590    X94
447    bond        1    5    4.7660    1.0930    C94
448    bond        1    6    5.0470    1.4180    C94
449    bond        1    8    5.0840    1.4510    C94
450    bond        1    9    4.7630    1.4580    C94
451    bond        1   10    4.6640    1.4360    C94
452    bond        1   11    6.0110    1.3600    #C94
453    bond        1   12    2.9740    1.7730    C94
454    bond        1   13    2.5290    1.9490    E94
455    bond        1   14    1.7060    2.0900    E94
456    bond        1   15    2.8930    1.8050    C94
457    bond        1   17    2.8410    1.8130    X94
458    bond        1   18    3.2580    1.7720    X94
459    bond        1   19    2.8660    1.8300    #X94
460    bond        1   20    4.6500    1.5040    C94
461    bond        1   22    4.2860    1.4820    E94
462    bond        1   25    2.9800    1.8100    #X94
463    bond        1   26    2.7900    1.8300    #X94
464    bond        1   34    3.8440    1.4800    #C94
465    bond        1   35    7.9150    1.3070    X94
466    bond        1   37    4.9570    1.4860    C94
467    bond        1   39    6.1140    1.4450    C94
468    bond        1   40    4.9220    1.4460    C94
469    bond        1   41    3.8300    1.5100    #C94
470    bond        1   43    3.9710    1.4720    X94
471    bond        1   45    3.8440    1.4800    X94
472    bond        1   46    3.8130    1.4820    X94
473    bond        1   54    4.2670    1.4610    C94
474    bond        1   55    4.6460    1.4540    C94
475    bond        1   56    4.1660    1.4530    C94
476    bond        1   57    4.6690    1.4610    E94
477    bond        1   58    4.3290    1.4510    E94
478    bond        1   61    4.8450    1.4240    X94
479    bond        1   62    4.4560    1.4440    X94
480    bond        1   63    4.4810    1.4710    E94
481    bond        1   64    4.5180    1.4690    E94
482    bond        1   67    4.1880    1.4590    E94
483    bond        1   68    4.2170    1.4790    C94
484    bond        1   72    2.9560    1.8010    X94
485    bond        1   73    2.6080    1.8390    X94
486    bond        1   75    2.5470    1.8580    E94
```

```
942    ################################
943    ##                            ##
944    ##  Angle Bending Parameters  ##
945    ##                            ##
946    ################################
947
948
949    angle       1    1    1    0.8510   109.6080   C94
950    angle       1    1    2    0.7360   109.4450   C94
951    angle       1    1    3    0.7770   107.5170   C94
952    angle       1    1    4    1.0060   110.2650   E94
953    angle       1    1    5    0.6360   110.5490   C94
954    angle       1    1    6    0.9920   108.1330   C94
955    angle       1    1    8    0.7770   108.2900   C94
956    angle       1    1    9    1.1360   108.1940   E94
957    angle       1    1   10    1.0500   109.9600   C94
958    angle       1    1   11    1.2250   108.3130   C94
959    angle       1    1   12    1.0560   108.6790   C94
960    angle       1    1   13    1.0780   106.8200   E94
961    angle       1    1   14    0.9800   109.9450   E94
962    angle       1    1   15    0.7430   107.3970   C94
963    angle       1    1   17    1.0890   108.5780   E94
964    angle       1    1   18    1.0930   109.3150   E94
965    angle       1    1   19    0.7550   115.4360   E94
966    angle       1    1   20    1.0210   108.6590   E94
967    angle       1    1   22    1.0010   110.1250   E94
968    angle       1    1   25    0.8030   112.3560   X94
969    angle       1    1   26    0.8330   109.8790   E94
970    angle       1    1   34    1.1790   106.4930   C94
971    angle       1    1   37    0.7560   108.6170   C94
972    angle       1    1   39    0.9270   109.1700   C94
973    angle       1    1   40    1.1300   108.6780   E94
974    angle       1    1   41    0.3300    98.4220   C94
975    angle       1    1   43    1.1350   108.0190   E94
976    angle       1    1   45    1.1970   105.0280   E94
977    angle       1    1   54    1.1730   106.4240   E94
978    angle       1    1   55    1.1500   107.6040   E94
979    angle       1    1   56    1.1990   110.3710   C94
980    angle       1    1   57    1.0120   109.9000   E94
981    angle       1    1   58    1.1790   106.3270   E94
982    angle       1    1   61    1.1250   109.3110   E94
983    angle       1    1   63    1.0060   110.0580   E94
984    angle       1    1   64    0.9880   111.0640   E94
985    angle       1    1   67    1.2160   104.5570   E94
986    angle       1    1   68    1.0180   107.1950   C94
987    angle       1    1   73    1.1600   104.6580   E94
988    angle       1    1   78    1.0120   109.8500   E94
989    angle       1    1   80    0.9470   113.3270   E94
990    angle       1    1   81    1.1080   109.8370   E94
```

| strbnd | angle stretches |
| opbend | out of plane bends |

```
        ###########################################
        ##                                       ##
        ##   Torsions                            ##
        ##                                       ##
        ###########################################


torsion        1    1    1    1     0.1030 +1    0.6810 -2    0.3320 +3    C94
torsion5       1    1    1    1     0.1440 +1   -0.5470 -2    1.1260 +3    C94
torsion        1    1    1    2    -0.2950 +1    0.4380 -2    0.5840 +3    C94
torsion        1    1    1    3     0.0660 +1   -0.1560 -2    0.1430 +3    C94
torsion        1    1    1    5     0.6390 +1   -0.6300 -2    0.2640 +3    C94
torsion        1    1    1    6    -0.6880 +1    1.7570 -2    0.4770 +3    C94
torsion5       1    1    1    6     0.0000 +1    0.0000 -2    0.0540 +3    C94
torsion        1    1    1    8    -1.4200 +1   -0.0920 -2    1.1010 +3    C94
torsion5       1    1    1    8     0.0000 +1   -0.1580 -2    0.3230 +3    C94
torsion        1    1    1   11     0.5930 +1    0.6620 -2    1.1200 +3    C94
torsion        1    1    1   12    -0.6780 +1    0.4170 -2    0.6240 +3    C94
torsion        1    1    1   15    -0.7140 +1    0.6980 -2    0.0000 +3    C94
torsion        1    1    1   34    -0.6470 +1    0.5500 -2    0.5900 +3    C94
torsion        2    1    1    5     0.3210 +1   -0.4110 -2    0.1440 +3    C94
torsion        3    1    1    3     0.4430 +1    0.0000 -2   -1.1400 +3    C94
torsion        3    1    1    5    -0.2560 +1    0.0580 -2    0.0000 +3    C94
torsion        3    1    1    6    -0.6790 +1   -0.0290 -2    0.0000 +3    C94
torsion        5    1    1    5     0.2840 +1   -1.3860 -2    0.3140 +3    C94
torsion        5    1    1    6    -0.6540 +1    1.0720 -2    0.2790 +3    C94
torsion        5    1    1    8    -0.7440 +1   -1.2350 -2    0.3370 +3    C94
torsion        5    1    1   10     0.0000 +1    0.0000 -2    0.4270 +3    C94
torsion        5    1    1   11     0.0000 +1    0.5160 -2    0.2910 +3    C94
torsion        5    1    1   12     0.6780 +1   -0.6020 -2    0.3980 +3    C94
torsion        5    1    1   15     1.1420 +1   -0.6440 -2    0.3670 +3    C94
torsion        5    1    1   25     0.0000 +1    0.0000 -2    0.2950 +3    X94
torsion        5    1    1   34     0.6920 +1   -0.5300 -2    0.2780 +3    C94
torsion        5    1    1   37     0.0000 +1    0.0000 -2    0.3890 +3    C94
torsion        5    1    1   39     0.0000 +1    0.0000 -2    0.2780 +3    C94
torsion        5    1    1   41     0.0000 +1    0.0000 -2   -0.1410 +3    C94
torsion        5    1    1   56     0.0000 +1    0.0000 -2    0.3240 +3    C94
torsion        5    1    1   68     0.0000 +1    0.0000 -2    0.1360 +3    C94
torsion        6    1    1    6     0.4080 +1    1.3970 -2    0.9610 +3    C94
torsion5       6    1    1    6     0.3130 +1   -1.0350 -2    1.6310 +3    C94
torsion        8    1    1    8     1.0550 +1    0.8340 -2    0.0000 +3    C94
torsion       11    1    1   11    -0.3870 +1   -0.5430 -2    1.4050 +3    C94
torsion       12    1    1   12     0.0000 +1    0.0000 -2    0.8930 +3    C94
torsion       15    1    1   15    -0.1770 +1    0.0000 -2    0.0490 +3    C94
torsion        1    1    2    1     0.4190 +1    0.2960 -2    0.2820 +3    C94
torsion        1    1    2    2    -0.4940 +1    0.2740 -2   -0.6300 +3    C94
torsion        1    1    2    5     0.0750 +1    0.0000 -2    0.3580 +3    C94
torsion        2    1    2    2    -0.2930 +1    0.1150 -2   -0.5080 +3    C94
torsion        2    1    2    5     0.3010 +1    0.1040 -2    0.5070 +3    C94
torsion        3    1    2    1     0.5650 +1   -0.5540 -2    0.2340 +3    C94
torsion        3    1    2    2    -0.5770 +1   -0.4820 -2   -0.4270 +3    C94
```

```
        ###########################################
        ##                                       ##
        ##   Atomic Partial Charge Parameters    ##
        ##                                       ##
        ###########################################


mmffchrg       1      0.0000    0.0000    0.0000    E94
mmffchrg       2     -0.1350    0.0000    0.0000    E94
mmffchrg       3     -0.0950    0.0000    0.0000    E94
mmffchrg       4     -0.2000    0.0000    0.0000    E94
mmffchrg       5     -0.0230    0.0000    0.0000    E94
mmffchrg       6     -0.2430    0.0000    0.0000    E94
mmffchrg       7     -0.6870    0.0000    0.0000    E94
mmffchrg       8     -0.2530    0.0000    0.0000    E94
mmffchrg       9     -0.3060    0.0000    0.0000    E94
mmffchrg      10     -0.2440    0.0000    0.0000    E94
mmffchrg      11     -0.3170    0.0000    0.0000    E94
mmffchrg      12     -0.3040    0.0000    0.0000    E94
mmffchrg      13     -0.2380    0.0000    0.0000    E94
mmffchrg      14     -0.2080    0.0000    0.0000    E94
mmffchrg      15     -0.2360    0.0000    0.0000    E94
mmffchrg      16     -0.4750    0.0000    0.0000    E94
mmffchrg      17     -0.1910    0.0000    0.0000    E94
mmffchrg      18     -0.1180    0.0000    0.0000    E94
mmffchrg      19      0.0940    0.0000    0.0000    E94
mmffchrg      20     -0.0190    0.0000    0.0000    E94
mmffchrg      21      0.1570    0.0000    0.0000    E94
mmffchrg      22     -0.0950    0.0000    0.0000    E94
mmffchrg      23      0.1930    0.0000    0.0000    E94
mmffchrg      24      0.2570    0.0000    0.0000    E94
mmffchrg      25      0.0120    0.0000    0.0000    E94
mmffchrg      26     -0.1420    0.0000    0.0000    E94
mmffchrg      27      0.0940    0.0000    0.0000    E94
mmffchrg      28      0.0580    0.0000    0.0000    E94
mmffchrg      29      0.2070    0.0000    0.0000    E94
mmffchrg      30     -0.1660    0.0000    0.0000    E94
mmffchrg      31      0.1610    0.0000    0.0000    E94
mmffchrg      32     -0.7320    0.5000   -0.5000    E94
mmffchrg      33      0.2570    0.0000    0.0000    E94
mmffchrg      34     -0.4910    0.0000    1.0000    E94
mmffchrg      35     -0.4560    0.5000   -1.0000    E94
mmffchrg      36     -0.0310    0.0000    0.0000    E94
mmffchrg      37     -0.1270    0.0000    0.0000    E94
mmffchrg      38     -0.4370    0.0000    0.0000    E94
mmffchrg      39     -0.1040    0.0000    0.0000    E94
mmffchrg      40     -0.2640    0.0000    0.0000    E94
mmffchrg      41      0.0520    0.0000    0.0000    E94
mmffchrg      42     -0.7570    0.0000    0.0000    E94
```
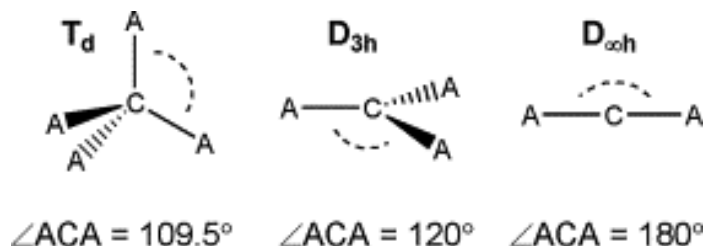
Forcefield names

- MM2
- MM3
- MMFF94
- AMBER
- CHARMM
- TINKER
- UFF

What is the name of the MM forcefield that SPARTAN uses?

MM Problems for TM Complexes

1. "Unique Labelling Problem"
   - How to describe the angular geometry at metal?
     - Simple for carbon – 3 common geometries (tetrahedral, trigonal planar and linear with angles 109.5°, 120° and 180°)

$\angle ACA = 109.5°$ $\angle ACA = 120°$ $\angle ACA = 180°$

   - Common TM complex geometries have multiple reference angles for the same A-M-A triad

$\angle AMA = 90°$ $\angle AMA = 90°$
$\angle AMA = 180°$ $\angle AMA = 120°$
                            $\angle AMA = 180°$
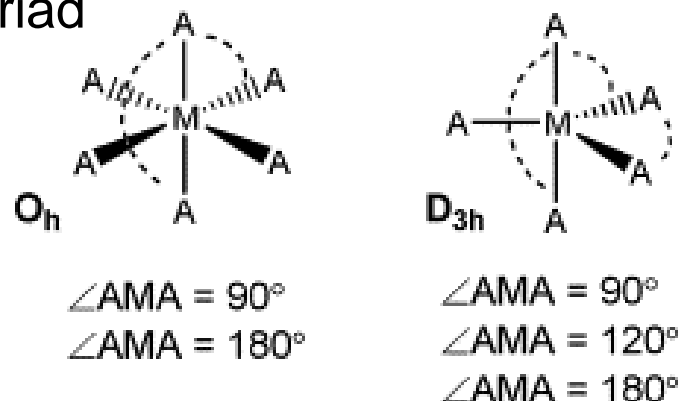
MM Problems for TM Complexes

2. Can't model ligand field splitting

   ▪ No concept of strong or weak ligands / metals

   ▪ Or high and low spin

     ▪ No evaluation of octahedral vs. tetrahedral vs. square planar

3. Doesn't capture trans effect and influence

4. Doesn't capture Jahn-Teller distortions

5. Can't describe bonding of π-donor and π-acceptor ligands

> These are all **stereoelectronic** effects… The forcefield parameterisation is not molecule specific but general and so doesn't account for these electronic effects that lead to structural and thermodynamic variation.

MM Uses

- Energy minimisation
- Geometry optimisation
    - 1st order – uses 1st derivatives of energy $\frac{dV}{dR}$
    - 2nd order – uses 1st and 2nd order derivatives $\frac{dV^2}{d^2R}$
- Conformation searches
- Computer Aided Design
- Molecular Dynamics calculations

Computer Aided Design

Ligand = small molecule

- *"in silico"*
- Used in drug discovery
    - Relies on knowledge of the three-dimensional structure of target and charges in the surrounding environment
    - High throughput screening

- Structure-based approach
    - Protein-Ligand docking
    - Pharmacophore
    - Ligand design

- Ligand-based approach
    - Similarity (or not) to previously known active ligands
    - Molecular descriptors
    - Quantitative structure-activity relationships

Wikipedia
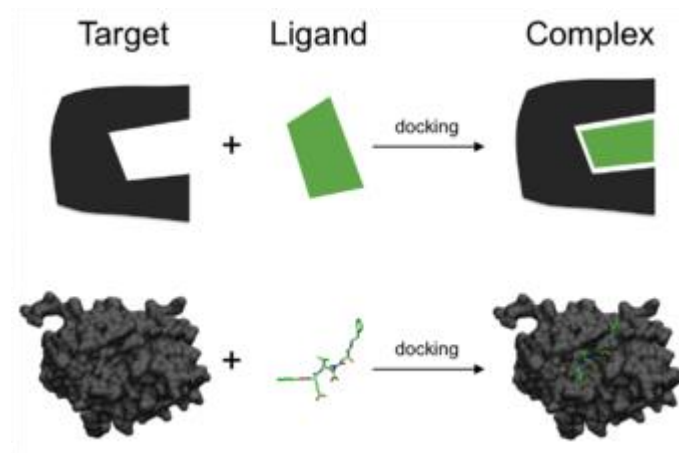
QM/MM

- MM calculations take significantly less time than QM methods
    - But ignore electrons and orbitals

- QM/MM = combination of the two, where the reactive parts of the system are treated with QM

- Scales as N^2

Computational Approaches in Physics (2016)
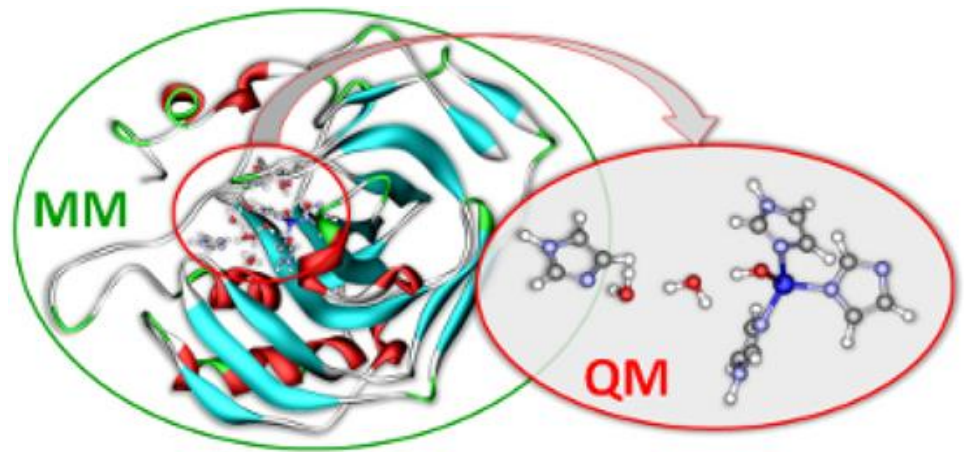
QM/MM

- Partition the system into two regions
    - Active site   requires accurate QM calculation
    - Rest of system   approximate and fast MM calculation

- Partitioning correctly is the tricky part… where should the boundary be?
    - Use "capping" – replacing large groups with H atoms in QM region

- Chemistry Nobel prize in 2013 awarded to Warshel and Levitt for this topic (and also Karplus for Molecular Dynamics)

Will show examples in the final two lectures

Ab Initio

Ab Initio

- QM uses Schrödinger wave equation to compute the electronic structure of a molecule

- Describes the wave function in terms of the Hamiltonian operator ($\widehat{H}$), which contains **kinetic** ($\widehat{T}$) and **potential** ($\widehat{V}$) energy operators for a general N-particle system

- $\widehat{H}$ operator is equal to the energy (E) of the system

$$\widehat{H}\Psi = \left[\widehat{T} + \widehat{V}\right]\Psi = E\Psi$$

Molecular Hamiltonian

- The kinetic and potential operators are then split further into five terms to give the Molecular Hamiltonian:

Kinetic energy due to nuclei in system

Potential energy arising from Coulombic electron-nucleus attraction

Potential energy due to nuclei-nuclei repulsions

$$\widehat{H} = \widehat{T}_n + \widehat{T}_e + \widehat{V}_{en} + \widehat{V}_{ee} + \widehat{V}_{nm}$$

Kinetic energy due to each electron

Potential energy due to electron-electron repulsion

- Solving this equation is only possible for the simplest of systems as the wavefunction depends simultaneously on the coordinates of ALL the electrons and nuclei!

Born-Oppenheimer Approximation

- Separates the electronic and nuclear motions of the system

- Treat the nuclei as stationary
    - Electrons move much faster than nuclei, so they appear stationary and we ignore their kinetic motion

- Means electronic structure can be calculated with the nuclear coordinates and not their momentum to give a potential energy surface

- Excludes nuclear kinetic energy term ($\hat{T}_n$) – gives the electronic Hamiltonian ($\hat{H}_{el}$), and the electronic Schrödinger equation:

$$\hat{H}_{el} = \hat{T}_e + \hat{V}_{en} + \hat{V}_{ee} + \hat{V}_{nm}$$

Hartree Fock (HF)

- Calculates each electron in an orbital independently

- Views other electrons as a mean field – neglecting electron-electron correlation

- Starting point for QM approaches

HF will be covered in more detail from lecture 4 onwards

- By adding parameters to HF you are creating *semi-empirical* methods…

Semi Empirical

Semi Empirical (SE)

- Attempts to solve the electronic structure calculation for a many-body system

- **Adds approximations based on experimental parameters**

- *[Ab initio methods are generated without reference to experimental data]*

- For many years with HF calculations, computational chemists were limited to the study of very small (5-10 atom) molecules – due to lack of computational power

- Frustrated QM chemists derived approximate versions of HF that were faster but included empirical parameters derived from experiment to help correct the deficiencies introduced by their approximations

Semi Empirical (SE)   Part 2

- These SE methods retain a QM component and can be used to describe <u>bond making and breaking</u>

- Half way house between accuracy of QM methods and speed of classical simulations

- SE = approximate QM methods, as still seeking solutions to Schrödinger's equation…

  - … but with a series parameterisation schemes (often derived from experimental data) to dramatically cut computational cost

Semi Empirical (SE)   Part 3

- Remember technique is based on parameterisation so quality of results obtained only reflects the quality of the parameterisation scheme
  - Experimentally observed data;
    - Structures
    - Ionisation energies
    - Heats of formation

- SE methods were widely popular – have now been displaced by DFT

- Increase in computational power over time has also benefited SE:
  - Can now perform relatively long MD simulations at a SE level of theory

SE Approximations

- Neglect all electrons except the valence electrons

- Parameters required for every element in a system but not for bond lengths and angles

- Advantages
  - Simplification to integrals so that time required to compute increases only as N^3
  - Hence can study large molecules
  - Experimental results include effects of electron correlation, so some of this is implicit in the calculations
    - Also a disadvantage - difficult to assess the errors and not clear how great an allowance is made

Hückel Approximation

- First SE method developed
  - Was for conjugated π systems

- HMO = Hückel molecular orbital theory

- Each molecular orbital is written as a linear combination of atomic orbitals (LCAO)

- Gives rise to a set of secular equations, which have non trivial solutions
  - Set selected matrix elements to zero and parameterize non zero elements

Neglecting Overlaps

- Pople and Dewar (and co-workers) made significant progress in treating general molecular systems with SE methods

- Treat valence electrons explicitly
  - The valence-electron Hamiltonian H_v is:

$$H_{\mathrm{V}} = \sum_{i=1}^{N_{\mathrm{V}}} h_i^{\mathrm{V}} + \frac{1}{2} j_o \sum_{i,j}^{N_{\mathrm{V}}} \frac{1}{r_{ij}}$$

- Most primitive approach = CNDO (Complete Neglect of Differential Overlap)

What is an Overlap?

$$\Psi = c_1\phi_1 + c_2\phi_2$$

$$\frac{\int \Psi^*\hat{H}\Psi d\tau}{\int \Psi^*\Psi d\tau} = \frac{\int [c_1\phi_A + c_2\phi_B]\hat{H}[c_1\phi_A + c_2\phi_B]d\tau}{\int [c_1\phi_A + c_2\phi_B]^2 d\tau}$$

$$= \frac{c_1^2 H_{AA} + 2c_1c_2 H_{AB} + c_2^2 H_{BB}}{c_1^2 S_{AA} + 2c_1c_2 S_{AB} + c_2^2 S_{BB}}$$

$$= \frac{c_1^2 H_{AA} + 2c_1c_2 H_{AB} + c_2^2 H_{BB}}{c_1^2 + 2c_1c_2 S + c_2^2}$$

- H is energy, S probability
- H_{AA} = **Coulomb integral**: interaction energy of e⁻ with its' "home" AO
- H_{AB} = **Resonance integral**: interaction energy between orbital A and e⁻ of orbital B
- S = **Overlap integral**: $\int \phi_1\phi_2 d\tau$

ϕ_1 ϕ_2
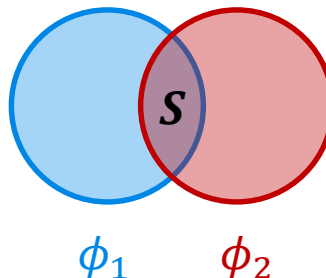
S goes from 0 → 1

$$S_{AA} = S_{BB} = 1$$
$$S_{AB} = S_{BA} = S$$

Differential Overlap

- What is a differential overlap?

    - The two-electron repulsion integral

    - Number of orbitals used in calculation = N, number of two-electron repulsion integrals scales as N^4, but after removing the differential overlaps scales as N^2

- So, the two-electron integrals are set to zero

- Neglects all 3-centre and 4-centre two-electron integrals

- Pariser-Parr-Pople (PPP) and Complete Neglect of Differential Overlap (CNDO) methods are examples of Zero Differential Overlap (ZDO) approximations

Intermediate Neglect of Differential Overlap (INDO)

- Intermediate neglect is not applied when A = B = C = D
  - i.e. when all four basis functions are on the same atom

- These one-centre exchange integrals are important for explaining the splitting between electronic states that comes from the same electronic configuration, so INDO gives vastly improved results over CNDO (especially spectroscopically)

- Although CNDO and INDO give reasonable equilibrium geometries when compared to experiment, they give poor results (as do HF-SCF methods) when compared with experimental quantities such as standard enthalpies of formation

- Methods = INDO, MINDO, ZINDO, SINDO

Neglect of Diatomic Differential Overlap (NDDO)

- From Pople

- Does not apply "ZDO" when A = B and C = D
  - When the basis functions for the first electron are on the same atom and the basis functions for the second electron are on the same atom

- This relies on fact that many of two-electron integrals involving basis functions centered on more than two atoms are very small – so can be neglected

- Gives dramatic computational time saving

- Methods = MNDO, AM1, PM3

Modified Neglect of Differential Overlap (MNDO)

- Tends to give covalent bonds that are too short
- Activation energies too high
- Poorly predicted proton affinities

- Added in additional parameters for nuclear repulsion energy

- Was all done *ad hoc*, some parameters fitted against a subset of data

Austin Model 1 (AM1)

- MNDO refitted simultaneously against larger set of experimental data

- Dewar developed improved MNDO called Austin model 1 (AM1) named after the University Texas at Austin where it was developed

- Applies Koopmans theorem to get better enthalpies and ionisation energies than MNDO

- But can get some very peculiar geometries for H-bonds

Parameterisation Model 3 (PM3)

- MNDO refitted again to give Parameterization Model 3 (PM3)

- Is the third parameterisation of the MNDO method
  - version 1 = MNDO, version 2 = AM1

- Gives better bond lengths, ionisation energies and enthalpies of formation than previous two versions

- There are some cases where PM3 is worse than MNDO though
  - *E.g.* all *sp*^{3} nitrogens predicted to be pyramidal!

- PM3 fixed some problems with AM1 but introduced others, so on balance if one method performs very poorly for your set of molecules it is worth trying the other

Semi Empirical in the Future

- Continues to be developed – always room for improvement!

- Current methods do not use *d* orbitals in their basis sets – so rubbish for most *d* metal compounds
    - Or any *p* elements where we know *d* orbitals are important

- Both MNDO and AM1 have had further iterations where attempts to extend parameterisation to include *d* orbitals:
    - MNDO/d and AM1/d

Periodic Boundary Conditions

Solids

- An infinite three dimensional-periodic array

- Electronic energy levels are more complicated than isolated molecules

- How can we treat this computationally when we've been previously looking at discrete individual molecules?

- Have to set boundary conditions, similar in the way that crystallography identifies a unit cell, ie. an independent "block" that repeats to make up the macro structure

- Approach is called **Periodic Boundary Conditions** (PBC)

Reciprocal Lattice and Brillouin Zone

- Unit cell = includes all the unique symmetry of a crystal / solid
  - The entire lattice can be built by repetition of the unit cell in three dimensions
  - A repeating "brick" of 7 potential 3D shapes

- Reciprocal lattice = Fourier transformation of the real-space lattice

- Brillouin zone = crystallographic primitive cell in reciprocal space
  - Allows us to classify how waves propagate through a material
  - Relates to allowed lattice vibrations and electronic wave functions

Band Structure

- Not isolated energy levels

- Is a continuum of <u>occupied orbitals</u> (*valence band*) and <u>vacant orbitals</u> (*conduction band*)
  - Collectively called **band structure**

- Energy gap between the valence and conduction bands defines whether the material is:
  - Insulator   *large band gap*
  - Conductor   *no gap*
 - Semi-conductor   *intermediate band gap*

k Vectors and Points

- Simplify the 3D Brillouin zone by slicing it to create 2D representations

- k vectors link k points along pathways in the Brillouin zone
  - Generates 2D band structure diagrams

- x-axis is the k vector showing k point positions, y-axis represents energy

- Permitted electronic levels correspond to occupied (valence) and unoccupied (conduction) bands

- Γ is the mid-point of the reciprocal lattice

- Band gap is shortest crossing from HOMO to LUMO

http://cmt.dur.ac.uk/sjc/thesis_dbj/node75.html
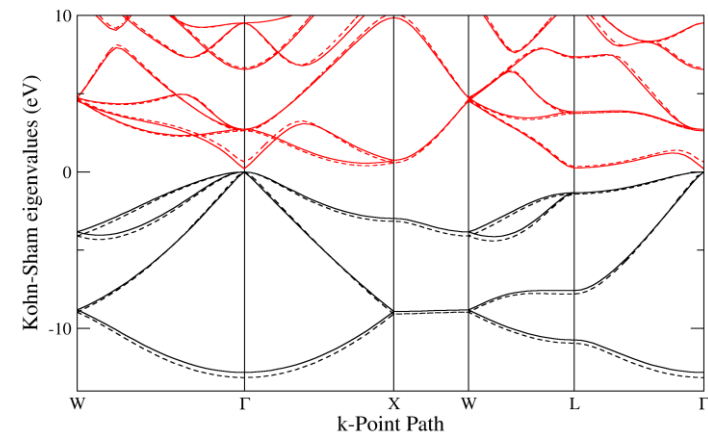
Setting the Boundaries

- A calculation on an infinite array of molecules is achieved by just treating the atoms shown in the asymmetric unit of the unit cell
    - All other molecules of the array are replicas of these parts

- Boundaries are "soft"
    - Should an atom cross one of the boundaries it will reappear at the opposite side
    - By its replica moving into the central cell to take its place…

- PBCs stop electrons falling off the end of the model
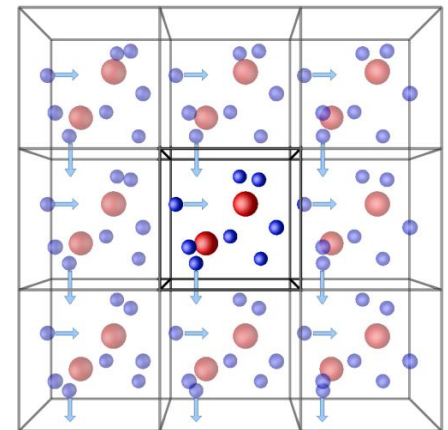
Plane Waves

- Electrons modeled as waves propagating through the model using continuous sinusoidal-type functions
    - Called **plane waves**

- For reciprocal lattice vectors **G**;

$$\Psi(r) = \sum_{G} c_{\boldsymbol{G}} \exp(i\boldsymbol{G}r)$$

    - $c_{\boldsymbol{G}}$ are basis set weighting coefficients
    - $\exp(i\boldsymbol{G}r)$ is the representation of a plane wave at any position r

- Collection of plane waves used with different wavelengths to capture the range of electrons and their kinetic energies
    - Slow valence electrons
    - Fast core electrons

Cutoffs

- The greater the range of plane waves used = better the basis set

- Quality of plane-wave basis set controlled in simulation by defining the energy cut-off

- This is the value chosen so that the lowering of the total energy by the addition of any further plane waves to the basis set is negligible

- Linear combination of plane waves effective in modeling the behaviour of valence electrons
  - Core electron density varies rapidly so require many plane waves to describe the core wavefunction – simplify by using a pseudopotential

- **Delocalised basis sets model kinetic energies of valence e⁻ and atomic pseudopotential functions to mimic effects of the core e⁻**

Optimising a Solid-State Structure

- **<u>Lattice Dynamics</u>**
  - Harmonic approximation
  - Each atom of asymmetric unit is displaced by small amounts from equilibrium position
  - Resulting energy points fitted to parabolas to obtain force constants k
  - Combine with reduced masses → get harmonic vibrational frequencies
- **<u>Molecular Dynamics</u>**
  - Convert calculated forces on atoms to velocities and accelerations via Newton's equations of motion
  - Simulation generates configurations of the molecular system connected to time
  - Statistical treatment leads to time-dependent properties
  - Have to manually assign vibrational modes to their symmetry species

Problems with PBC

- The Slater determinant in Fock-theory applied to plane-wave basis set is very large when approximating $\hat{H}$
    - Impedes performance of calculation
    - Currently can only use density functional theory

- MD simulation requires manual symmetry assignment to vibrational modes

- Lattice Dynamics assumes that system behaves harmonically

cp2k.org
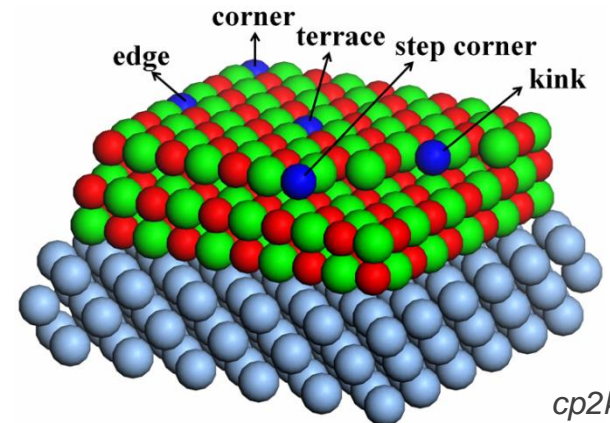
Molecular Dynamics

Molecular Dynamics (MD)

- Molecules are always moving – even at the absolute zero of temperature

- Ideally, molecules shouldn't be treated as a collection of static conformations

- Molecular Dynamics aims to simulate movement and motion

- Conformational searches can extract information about which conformer geometries have the lowest energies
  - But would not show if they slowly interconverted
    - … a long MD simulation would!

- Karplus won the Nobel prize in 2013 for work in this area

Conformational Searching

- Conformational searching will provide a list of the global minimum and local minima

- Populations then described by a Boltzmann distribution
    - Two assumptions:
        1. Conformations are in equilibrium
        2. Energy barriers to interconversion are high
            - If barriers are low then molecules may spend most of time converting instead of being minima!

Molecular Modeling Basics, Jan Jensen, *Routledge* (2010)
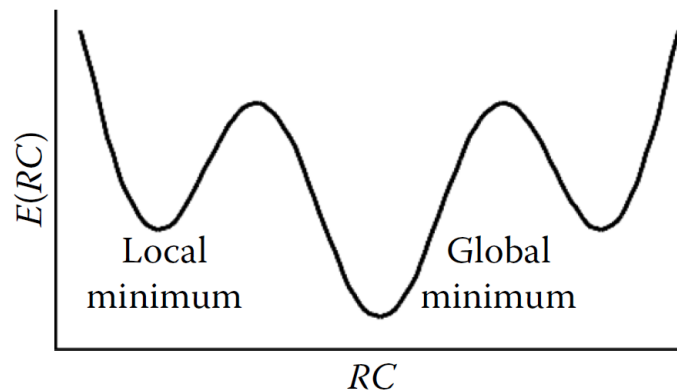
Boltzmann Distribution

- In statistical mechanics this is a <u>probability distribution</u> that gives the probability that a system will be a certain state (*i.e.* conformation) as a function of the state's energy and the temperature of the system

$$p_i = \frac{e^{\frac{-\varepsilon_i}{kT}}}{\sum_{j=1}^{M} e^{\frac{-\varepsilon_j}{kT}}}$$

- The denominator is known as the canonical partition function; Q (or Z)

$$p_i = \frac{1}{Q} e^{\frac{-\varepsilon_i}{kT}}$$

p_i = probability of state i
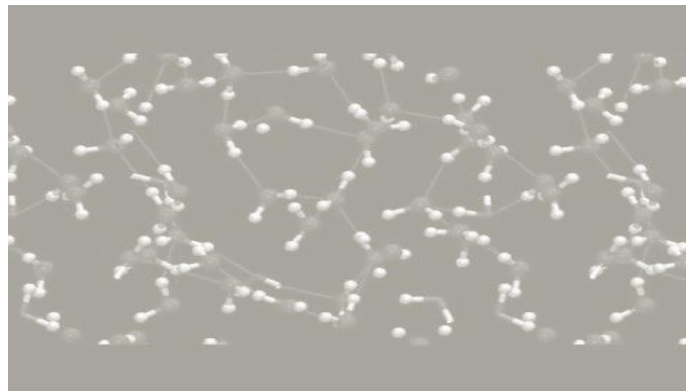
ε_i = energy of state i

k = Boltzmann constant,

T = temperature of the system

M = number of states accessible to the system

Basic Algorithm

- Divide time into discrete time steps
  - ~ 10^{-15} seconds each

- At each step:
  - Compute the forces acting on each atom
  - Move the atoms a little bit
    - Updating the position and velocity of each atoms using Newton's laws of motion

Newton's Laws of Motion

- **First law**: The velocity of a body remains constant unless the body is acted upon by an external force

$$\sum \boldsymbol{F} = 0 \Rightarrow \frac{d\boldsymbol{v}}{dt} = 0$$

- **Second law**: The acceleration, $\boldsymbol{a}$, of a body is parallel and directly proportional to the net force, $\boldsymbol{F}$, and inversely proportional to the mass, m

$$\boldsymbol{F} = m\frac{d\boldsymbol{v}}{dt} = m\boldsymbol{a}$$

- **Third law**: The mutual forces of action and reaction between two bodies are equal, opposite and co-linear

$$\sum \boldsymbol{F}_{a,b} = \sum \boldsymbol{F}_{b,a}$$

Equations of Motion

- Since velocity (v) is the derivative of position (x)
- And acceleration (a) is the derivative of velocity
- We can write the equations of motion as:

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{v}$$

$$\frac{d\boldsymbol{v}}{dt} = \frac{F(\boldsymbol{x})}{m}$$

- For N atoms, we have 3N position coordinates and 3N velocity coordinates

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \delta_t \boldsymbol{v}_i$$
$$\boldsymbol{v}_{i+1} = \boldsymbol{v}_i + \delta_t F(\boldsymbol{x}_i)/m$$

- Where δ_t is the time step

Parameters for MD

- A starting structure

- The temperature
  - Amount of energy given to the structure depends on the temperature
  - Energy is divided between movement and potential energy
  - Equally divided between all the atoms

- The step size
  - A short time that needs to be shorter than the shortest time in which anything interesting can happen to the molecule!

- The length of the run
  - As long as possible! Depends on the information required…
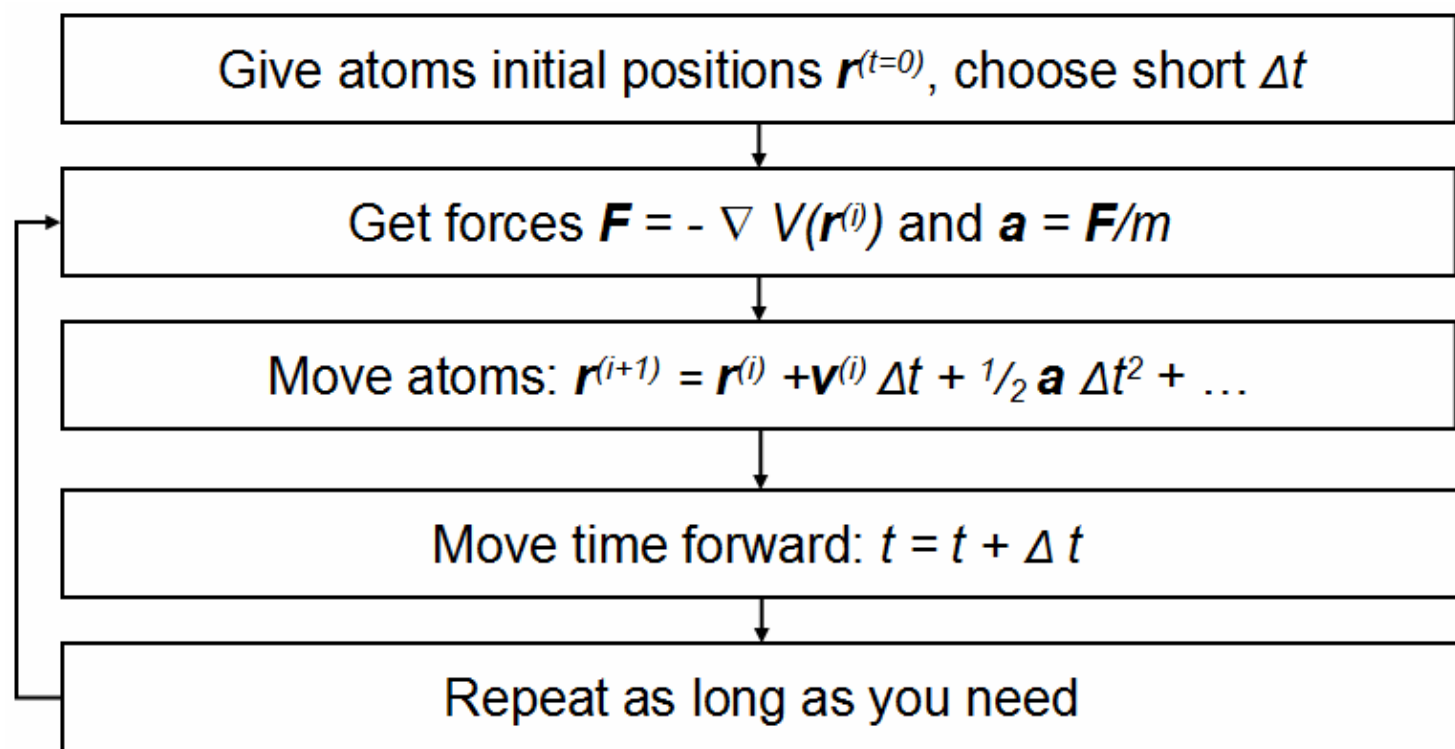
Differences between MD and MC

Property	MD	MC
Basic information needed	Gradient	Energy
Particles moved in each step	All	One
Coordinates	Cartesian	Any
Constraints	Difficult	Easy
Atomic velocities	Yes	No
Time dimension	Yes	No
Deterministic	(Yes)	No
Sampling	Physical	Non-physical
Natural ensemble	NVE	NVT

Constant Energy

- Total energy of the system can change from being mainly kinetic energy (KE) to mainly potential energy (PE)

- At the bottom of a "well" / minimum there state has a high KE and low PE

- Therefore, simulation will spend most of its time in high PE states as they correspond to time when the system is moving slowly

- This is not an ideal description of molecular behaviour
  - Produces a microcanonical ensemble of states:
    - *isolated system with constant energy, no energy or mass exchanged with the rest of the universe*

Constant Temperature

- Temperature corresponds to system's kinetic energy

- As calculation proceeds more energy is added to cool it down
  - System is in a high energy conformation

- Energy is removed from the system when it heats up
  - System is in a low energy conformation

- Produces a canonical ensemble of states:
  - *system in thermal equilibrium, can only exchange heat with the outside*

Microcanonical Ensemble

- System isolated from changes in moles (**N**), volume (**V**) and energy (**E**).
- Adiabatic process with no heat exchange
- Total energy is conserved - exchange of potential and kinetic energy
- For every time-step, each particle's position and velocity may be integrated – giving a trajectory
- Therefore, given the initial positions and velocities, we can calculate all future (or past) positions and velocities

  - N – constant number of particles
  - V – constant volume
  - E – constant Energy

- Temperature of system not fixed

NVE

Canonical Ensemble

- Moles (**N**), volume (**V**) and temperature (**T**) are conserved
- Energy of endothermic and exothermic processes is exchanged with a thermostat
- Probability distribution over microcanonical ensembles
- Energy exchanged with a large reservoir
- Total energy of system and reservoir combined remains constant, given by the Boltzmann distribution
- *i.e.* a heat bath

- NVT = canonical ensemble
  - N – constant number of particles
  - V – constant volume
  - T – constant temperature

NVT

Isothermal-Isobaric Ensemble

- Moles (**N**), pressure (**p**) and temperature (**T**) are conserved
- Thermostat and barostat are needed
  - For a constant number of particles either the volume or pressure can be fixed, but not both
- Corresponds to an open flask at ambient temperature and pressure
- For the simulation of biological membranes, isotropic pressure control is not appropriate, and instead either constant membrane area (Np_AT) or constant surface tension ($Np_\gamma T$)

NpT

Grand Canonical

- Volume (**V**), energy (**E**) and chemical potential (**μ**) are conserved
- A constant chemical potential (μ) is incommensurable with a constant number of particles
  - *open system, can exchange both energy and mass with the outside*

VEμ

Ensembles Summary

N	p	V	T	E	μ	Acronym	Equilibrium	Name
X		X	X			NVT	A has minimum	Canonical (MC)
X		X		X		NVE	S has minimum	Micro-canonical (MD)
X	X		X			NpT	G has minimum	Isothermal-Isobaric
		X	X		X	VEμ	(pV) has maximum	Grand-canonical

Simulated Annealing

- Makes random changes to a structure to find global minimum

- Random configuration and velocity for each atom

- Traces the movement of the molecule as the energy is slowly removed

- Works provided energy is removed infinitely slowly
  - More rapid rate of cooling will remove the certainty of success

Molecules in Solution

- A very large number of solvent molecules are required to solvate even the smallest of systems

- We want to know the solvent's effect on the solute molecule not necessarily all possible arrangements of the system

- To provide useful data we consider a sufficient number of solvent molecule arrangements (a representative sample) produced by MD simulations

- Particularly suited to studying protein conformations or other large molecules

Stochastic Dynamics

- Focus interest on small part of system
    - *E.g.* active site of a large protein or molecule in a large amount of solvent

- A full MD simulation would take a prohibitively long time – so "cheat" by constraining part of the system of lesser interest and only allowing the interesting part to move

- Difficulty is what to do an interface between moving and non-moving parts
    - Hence stochastic dynamics – used to transfer energy in this situation

MD Summary

- Molecules are not static but continuously move
    - Closer to reality than extracting information from a forcefield

- Investigates low lying regions of PES – not just minima

- Can limit searches to accessible conformations rather than all possible conformations

- Can find entropy (S), free energy (G) as well as internal energy (U)

MD Challenges

- Non-bonded interactions involve weak interactions between all particles in the system, which is normally the bottleneck in speed of MD simulations

- Get temperature-related phenomenon (rises) due to the small number of atoms that are used in MD simulations
  - *E.g.* when undergoing exothermic conformational changes, binding, or depositions of substrates on surfaces
  - Can cause the substrate to be vaporized in the model

- Complexity prevents it being used for simpler problems, such as conformation analysis of small molecules *in vacuo*

Computed vs. Real Life

Lectures 9 and 10

Week 23