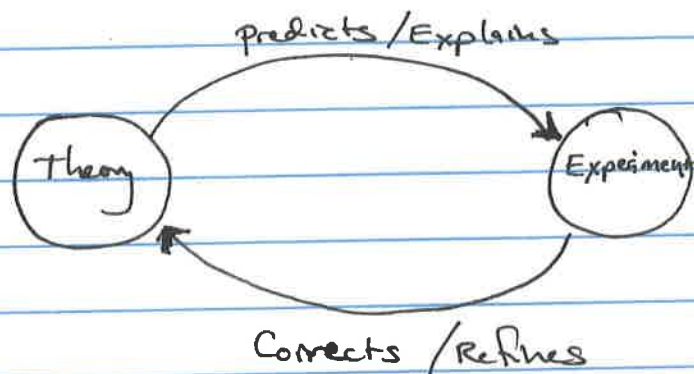
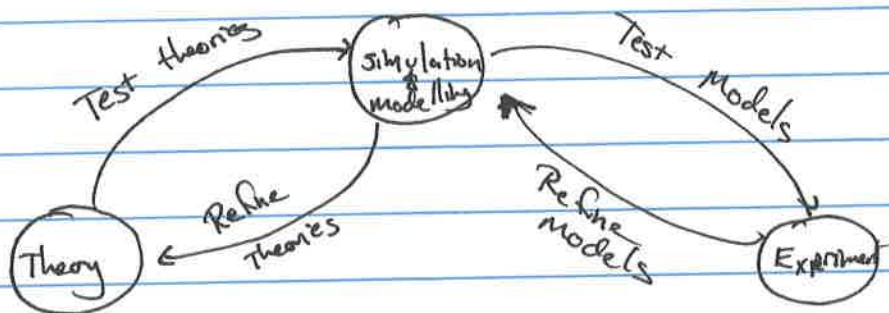


①

How science used to work:



How science works today:



(computer simulation is sometimes called the 3rd leg of science - the 4th is "Big Data")

Why is computer simulation so valuable in Chemistry?

- Measurements under extreme conditions that are not available experimentally (or which would be prohibitive)
- Details of molecular motion & structure that would be impossible or difficult to study experimentally
- Descriptions of events that are too fast, or too slow to study experimentally

To do all of these things, we need to know 2
the limitations of our computer models.

For the purposes of most of chemistry, the most complete description of a system we will ever need is the non-relativistic Schrödinger equation

$$H(\vec{R}, \vec{r}) \psi(\vec{R}, \vec{r}) = E_{\text{tot}} \psi(\vec{R}, \vec{r})$$

$\vec{R}$ = coordinates of the atomic nuclei (3 for water)

$\vec{r}$ = coordinates of the electrons (10 for water)

$\psi(\vec{R}, \vec{r})$ = wavefunction of the entire system

E_{tot} = total Energy

↑
a function of
39 coordinates!

$H(\vec{R}, \vec{r})$ = the "Hamiltonian" for this system (includes all contributions to the total energy)

$$= K_{\text{nuc}} + K_{\text{el}} + V(\vec{R}, \vec{r})$$

K_{nuc} = kinetic energy of nuclei

K_{el} = kinetic energy of electrons

$V(\vec{R}, \vec{r})$ = potential of interaction for all of the nuclei & electrons

$$V(\vec{R}, \vec{r}) = \frac{1}{4\pi\epsilon_0} \left[\sum_{a=1}^{N-1} \sum_{b=a+1}^N \frac{Z_a Z_b}{R_{ab}} + \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{e^2}{|r_i - r_j|} - \sum_{a=1}^N \sum_{j=1}^n \frac{Z_a e}{|R_a - r_j|} \right]$$

all are coulombic (charge-charge) interactions

These equations are too complicated for general use, so we must simplify.

③

The Born-Oppenheimer approximation: Since electrons are much lighter than atomic nuclei, they move much faster, so we can break the Schrödinger equation into ~~two~~ parts:

$$(K_{el} + V(\vec{R}, \vec{r}))$$

$$① \quad H_{el}(\vec{r}; \vec{R}) \psi(\vec{r}; \vec{R}) = E_{el}(\vec{R}) \psi(\vec{r}; \vec{R})$$

$\psi(\vec{r}; \vec{R})$ is the electronic wave function - for water this is a function of the 30 electronic positions ($\vec{r}$). It is parametrically dependent on the nuclear positions ($\vec{R}$)

So we solve the electronic problem assuming that the nuclei are nailed down in fixed positions, and we get:

$E_{el}(\vec{R})$ = the electronic energy (depends on where the nuclei are!)

$$② \quad H_{nuc}(\vec{R}) \phi(\vec{R}) = E \phi(\vec{R})$$

$\phi(\vec{R})$ is the nuclear wave function

$$H_{nuc}(\vec{R}) = K_{nuc} + E_{el}(\vec{R})$$

This last equation assumes that the electrons are whipping around so fast that the nuclei see only the average field of the electrons!

(4)

Some notes:

- E is an approximation of E_{tot} , and it includes contributions from vibrational, rotational, and translational motion of the molecule.

- $\psi(\vec{r}, \vec{R}) \approx \psi(\vec{r}; \vec{R}) \phi(\vec{R})$

$\hat{=}$ the decoupling of electronic & nuclear motion breaks down when we have light atoms (H) moving quickly.

A summary:

① $H_{\text{el}}(\vec{r}; \vec{R}) \psi(\vec{r}; \vec{R}) = E_{\text{el}}(\vec{R}) \psi(\vec{r}; \vec{R})$

② $H_{\text{nuc}}(\vec{R}) \phi(\vec{R}) = E \phi(\vec{R})$

Equation ① is a hard problem. It requires a large amount of computation using ab initio quantum codes, (Gaussian, Qchem, GAMESS) or other approximate methods (MNDO, MOPAC, DFT)

Equation ② gives the quantum nuclear behavior. It also requires an immense amount of computation for more than ~ 6 atoms

How do we get around this issue?

Nuclei are relatively heavy, and therefore classical dynamics can be used:

$$F = ma \Rightarrow -\frac{dE_{\text{el}}(\vec{R})}{d\vec{R}_i} = m_i \frac{d^2 \vec{R}_i}{dt^2}$$

(5)

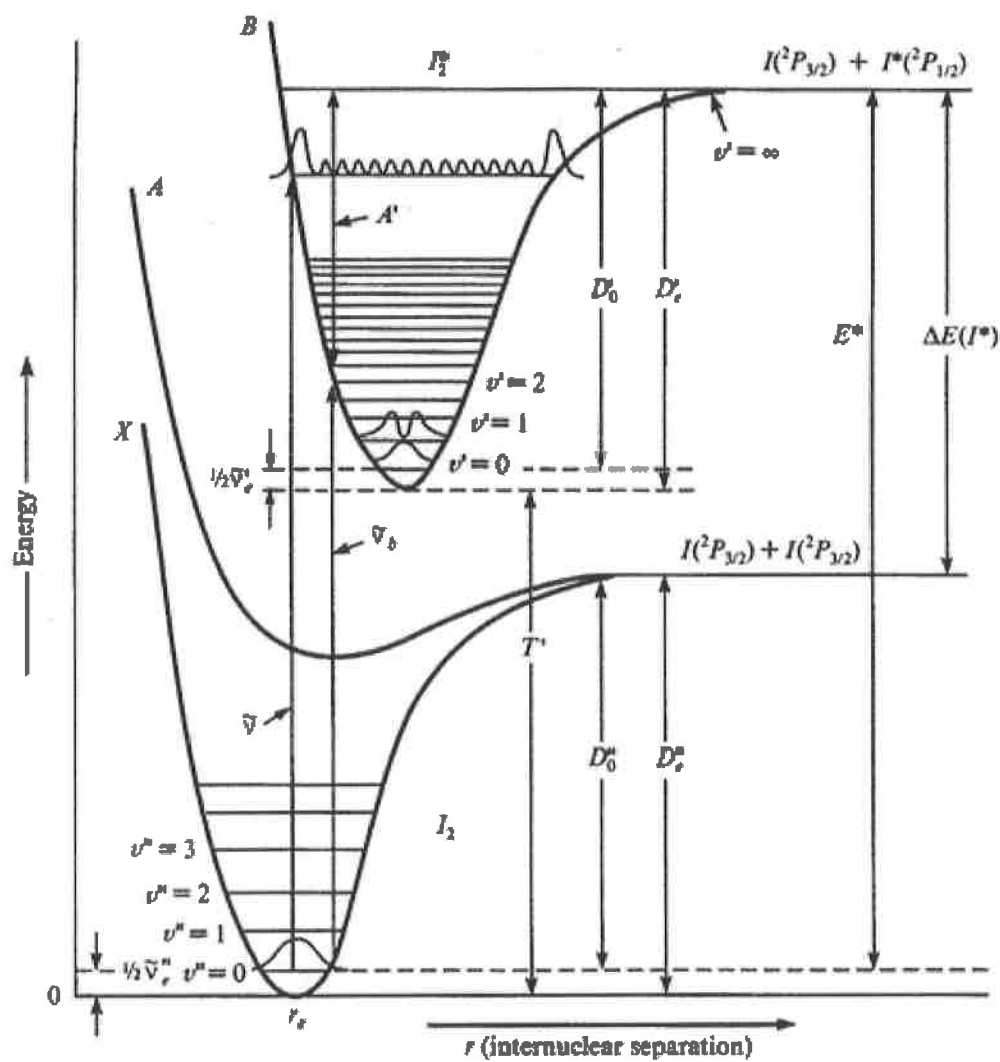
Solving
$$-\frac{\partial E_{el}(\vec{R})}{\partial R_i} = m_i \frac{d^2 R_i}{dt^2}$$

for the nuclei moving classically on a potential energy surface derived from electronic QM is sometimes called Born-Oppenheimer molecular dynamics.

We can further approximate $E_{el}(\vec{R})$ with a set of simple functions that model the internal behavior in a molecule. This approximation is called molecular mechanics or a force field.

Summary of approximations so far:

1. Born-Oppenheimer : nuclei move in an average field of the electrons
2. Nuclei move on a single potential energy surface (that is a single electronic state)
3. The PES (potential energy surface) can be modeled with simple functions (i.e. a force-field)
4. Nuclear motion can be described well-enough with classical mechanics!



(6)

We're largely in Ch. 4 in Leach today.

$$(1) \quad H_{el}(\vec{r}; \vec{R}) \psi(\vec{r}; \vec{R}) = E_{el}(\vec{R}) \psi(\vec{r}; \vec{R})$$

$$(2) \quad \underbrace{H_{nuc}(\vec{R})}_{K_{nuc} + E_{el}(\vec{R})} \phi(\vec{R}) = E \phi(\vec{R})$$

$E_{el}(\vec{R})$ comes from expensive calculations

We'd like to come up with approximate models for the electronic problem that build in most of the chemistry we know:

← approximated by empirical fit

$$V(\vec{R}) \approx E_{el}(\vec{R})$$

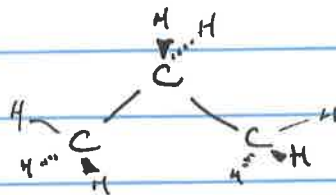
Force Fields: Analytic expressions (& their parameters) describing approximate potential energy surfaces

$$V(\vec{R}) = V_{\text{bonded}} + V_{\text{vdw}} + V_{\text{electrostatic}} + V_{\text{HB}} + V_{\text{constraints}}$$

V_{bonded} deals with the internal motions of molecules and assumes that bond lengths, angles, etc. tend to be near reference or equilibrium positions

$$V_{\text{bonded}} = \sum_{\text{bonds}} \frac{k_{\text{bond}}}{2} (l - l_0)^2 + \sum_{\text{angles}} \frac{k_{\text{angle}}}{2} (\theta - \theta_0)^2 + \sum_{\text{torsions}} \left[\sum_n \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) \right]$$

Consider ~~ethane~~ Propane:



Bond lengths & angles are "hard" degrees of freedom
(we can't deform very much without a large energy penalty)

Torsions (dihedrals) are "soft"

Let's go into some detail here:



A "Harmonic" bond: $V_{\text{bond}} = \frac{k_{ij}}{2} (|\vec{r}_{ij}| - r_{ij}^0)^2$

The relevant physical measurement is the distance:

$$|\vec{r}_{ij}| = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2}$$

k_{ij} and r_{ij}^0 are parameters that should be the same for any bonded pairs of the same atomic types.

Questions:

How many bonds are there in Propane?

(10)

How many types of bonds are there in propane?

(2)

How many parameters are needed to describe the bonds in propane?

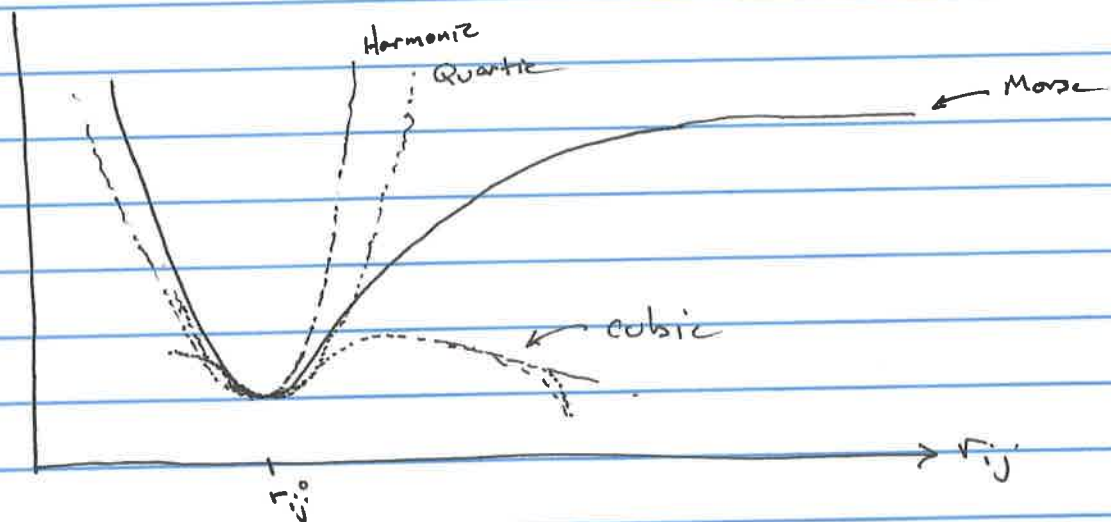
(4)

(8)

We might decide that harmonic springs aren't a good enough description of bonding:

$$\text{Morse: } V_{\text{bond}} = D_{ij} [1 - e^{-a_{ij}(r_{ij} - r_{ij}^0)}]^2$$

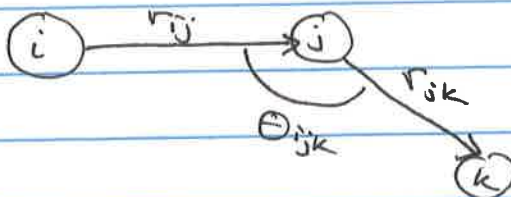
$$\text{Anharmonic: } V_{\text{bond}} = \underbrace{\frac{k_{ij}}{2}(r_{ij} - r_{ij}^0)^2}_{\text{cubic}} \underbrace{\left[1 - k'_{ij}(r_{ij} - r_{ij}^0) - k''_{ij}(r_{ij} - r_{ij}^0)^2 \dots \right]}_{\text{quartic}}$$



In general Morse is more accurate than Harmonic, but, it does let the bonds stretch to unrealistic lengths with only small restoring forces. ← can result in some unphysical results.

More accurate functional forms also usually require more parameters.

(9)

Bends

$$V_{\text{bend}} = \frac{k_{ijk}}{2} (\theta_{ijk} - \theta_{ijk}^0)^2 \quad \leftarrow \text{harmonic bend}$$

$$\theta_{ijk} = \cos^{-1} \left(\frac{\vec{r}_{ij} \cdot \vec{r}_{jk}}{|\vec{r}_{ij}| |\vec{r}_{jk}|} \right)$$

k_{ijk} and θ_{ijk}^0 are constants that are the same for any bending triplet of the same atomic types (even in reversed order)

Questions:

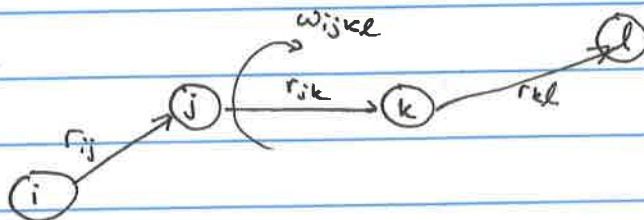
How many bends are there in propane? (18)

How many types of bend are there in propane? (3)

How many parameters are needed to describe the bends in propane? (6)

What is θ_{ijk}^0 for $\text{H}-\text{C}(\text{sp}^3)-\text{H}$?

for $\text{H}-\text{C}(\text{sp}^3)-\text{C}(\text{sp}^3)$?

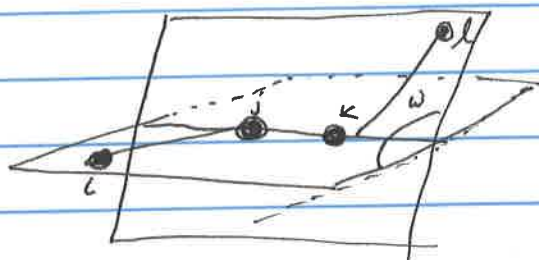
Torsions

$$V_{\text{torsion}} = \sum_{n=0}^3 \frac{V_{ijkl}^n}{2} \left[1 + \cos(n w_{ijkl} - \gamma_{ijkl}^n) \right]$$

$$\cos w_{ijkl} = \frac{(\vec{r}_{ij} \times \vec{r}_{jk}) \cdot (\vec{r}_{jk} \times \vec{r}_{kl})}{|\vec{r}_{ij} \times \vec{r}_{jk}| |\vec{r}_{jk} \times \vec{r}_{kl}|}$$

(10)

This can be hard to visualize, but imagine folding a piece of paper along the $j-k$ bond, this gives two planes, one with atom i & one with atom l :



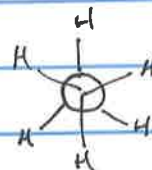
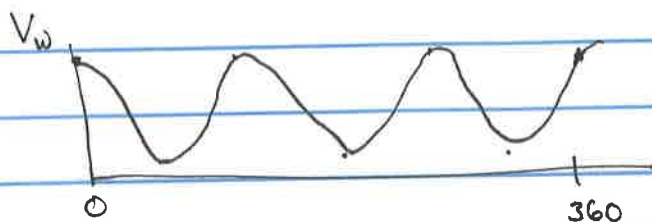
ω is the angle between those 2 planes. To find this angle, we construct vectors normal to each plane:

$$\vec{r}_{ij} \times \vec{r}_{jk} \quad \& \quad \vec{r}_{jk} \times \vec{r}_{kl}$$

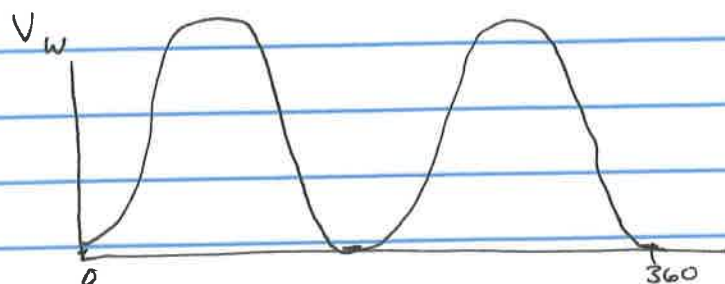
The torsion angle is then a dot-product problem!

Examples: 2 tetrahedral (sp^3) carbons connected by a single bond

$$V_{\omega} = \frac{V_s}{2} [1 + \cos(3\omega)]$$



2 sp^2 carbons connected via a double bond:



(11)

Because individual torsions can have character of many of these contributions, we can include lots of parameters. Here's a typical MM2 style torsion:

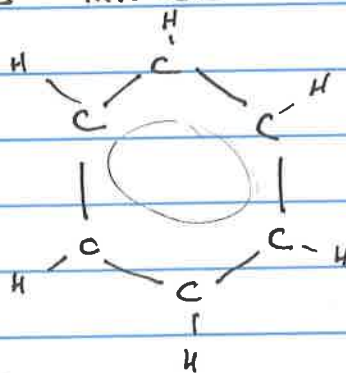
$$\begin{aligned}
 V(\omega) &= \frac{V_1}{2}(1 + \cos \omega) + \frac{V_2}{2}(1 - \cos(2\omega)) + \frac{V_3}{2}(1 + \cos(3\omega)) \\
 &\quad \uparrow \text{just a shift by } 180^\circ, \text{ so:} \\
 &= \frac{V_1}{2}(1 + \cos(\omega)) + \frac{V_2}{2}(1 + \cos(2\omega + \pi)) + \frac{V_3}{2}(1 + \cos(3\omega)) \\
 &= \sum_{n=1}^3 \frac{V_n}{2}(1 + \cos(n\omega + \gamma_n)) \quad \leftarrow \text{general form}
 \end{aligned}$$

Questions: How many torsions are there in propane? (18)
 How many types of torsion? (2)
 How many parameters? effectively: 4

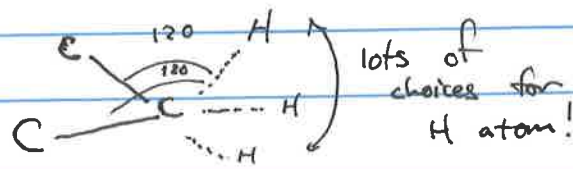
There's one more potential that is often added to sp^2 -hybridized atoms involved in π -bonding

Consider Benzene:

We know that the $C(sp^2)$ atom types are going to strongly favor 120° bond angles



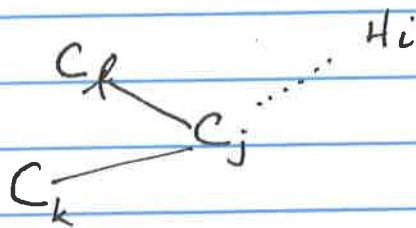
Edge-on, the Hydrogens have lots of choices to satisfy this, however:



What we need to enforce planarity is an out-of-plane bending or inversion potential.

This is an ad hoc term used to correct local classical potentials using chemical information that is derived from long-range electronic properties.

There's no real agreement on the best way to do it, either:

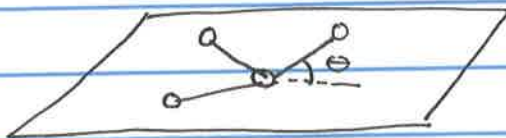


Lots of older force fields define "improper torsions" where the connectivity is not along bonds:

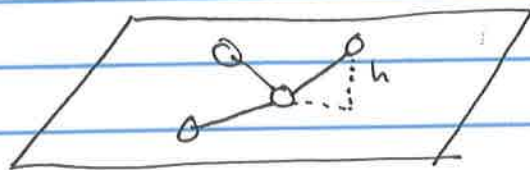
$$V_{\text{improper}}(\omega_{ijke}) = k(1 - \cos 2\omega_{ijke})$$

This is a potential that tries to keep the angle between the $i-j-k$ and $j-k-l$ planes at 0° or 180°

Other choices:

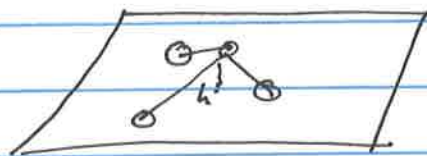


Bend of one satellite relative to plane of central + 2 satellites



Height of one satellite out of plane

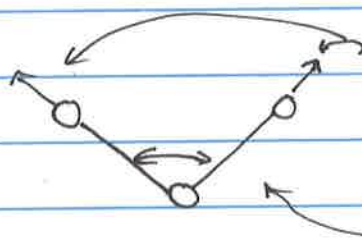
There are central-atom descriptions also:



← Central atom height out of plane of 3 satellites.

In general improper torsions, oop bends, inversions make force fields messy and sometimes introduce unintended consequences (e.g. frequency shifts)

Cross Terms:



stretching these bonds should make it easier to bend this bend

$$V(l_1, l_2, \theta) = \frac{k_{l_1 l_2 \theta}}{2} \left[(l_1 - l_1^0) + (l_2 - l_2^0) \right] (\theta - \theta_0)$$

↗ sometimes called a "stretch-bend" potential

Also: ~~Stretch-Stretch~~ Stretch-Stretch : $V(l_1, l_2) = \frac{k_{l_1 l_2}}{2} (l_1 - l_1^0)(l_2 - l_2^0)$

Stretch-torsion : $V(l, \omega) = k(l - l^0)(1 + \cos n\omega)$

If a force field uses only harmonic functions for stretching & bending and has no cross terms we call it a class I force field

Force fields with anharmonicity and coupling terms is called a class II force field

A short interlude on Parametrization

Once the functional forms are settled

- Determine k_{ij} , r_{ij}^0 , k_{ijk} , θ_{ijk}^0 , etc.
- Different sets of parameters can give the same energy for any given structure. Which is correct?
- Dependencies exist between terms.

How is it done?

- Using Experimental structural & thermodynamic data
- Using QM calculations

Transferability

If we tune parameters for propane, will they work for dodecane? For C_{60} ? For Proteins?

Goal: one set of parameters, valid for many systems

We don't want to reparameterize each time we study a new system.

We want predictive ability

A major part of transferability is Atom Types

Atom Types: Example $C(sp^3)$, $C(sp^2)$, $C(sp)$ are treated as different kinds of atoms

We parametrize each atom type in a different way based on connectivity

$$V(\vec{R}) = V_{\text{intra}} + V_{\text{vdw}} + V_{\text{elect}} + V_{\text{H-Bonding}} + V_{\text{constraint}}$$

↑
interaction potential

$$V_{\text{intra}} = V_{\text{bonds}} + V_{\text{bends}} + V_{\text{torsion}} + V_{\text{oop}} + V_{\text{cross}}$$

↑ out of plane ↑ cross terms

Non-Bonded interactions : $V_{\text{vdw}} + V_{\text{elect}}$

2 ways of handling these:

- All-Atom
 - interactions between molecules is a sum of interactions between constituent atoms
 - interaction sites are usually on atomic positions, but occasionally extra sites are added
- United-Atom
 - Some atoms (usually H) are combined with bonded heavy atoms
 - Reduces computational cost ($\sim \frac{1}{2}$ of atoms are H)
 - Allows longer time steps in MD
 - H-atom positions are often missing in experimental structures
 - Disadvantages - lose explicit H-bonding
 - lose some steric effects
 - some analysis becomes more difficult

Name	Canonical Reference	Primary Uses	Interesting features
MM2/MM3	"Molecular Mechanics. The MM3 Force Field (interatomic potential) for Hydrocarbons. 1". <i>J. Am. Chem. Soc.</i> 111 8551-8565.	Alkanes	
AMBER	Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995). "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules". <i>J. Am. Chem. Soc.</i> 117 : 5179–5197	Proteins	99SB variant is most widely used for proteins. Amber relies on TIP water models.
CHARMM	Mackerrill, A.D., Jr.; Brooks, B.; Brooks, C. L., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. (1998). "CHARMM: The Energy Function and Its Parameterization with an Overview of the Program". In Schleyer, P.v.R.; et al. The Encyclopedia of Computational Chemistry 1. Chichester: John Wiley & Sons. pp. 271–277.	Proteins / Nucleic Acids / Lipids	
GROMOS	Schmid N., Eichenberger A., Choutko A., Riniker S., Winger M., Mark A. & van Gunsteren W., "Definition and testing of the GROMOS force-field versions 54A7 and 54B7", <i>European Biophysics Journal</i> , 40(7) , (2011), 843–856	Proteins / Lipids	
OPLS	Jorgensen WL, Maxwell DS, Tirado-Rives J (1996). "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids". <i>J. Am. Chem. Soc.</i> 118 (45): 11225–11236	Organic Liquids	
UFF	UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations by A.K. Rappe, C.J. Casewit, K.S. Colwell, W.A. Goddard III, W.M. Skiff, <i>J. Am. Chem. Soc.</i> 114 (1992) 10024–10035		Full coverage of the periodic table!
TraPPE	Full list here: http://chem-siepmann.oit.umn.edu/siepmann/trappe/index.html	Liquid / Gas Phase equilibria	
MMFF	Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94, Thomas A. Halgren, <i>J. Comp. Chem.</i> ; 1996 ; 490-519	wide-range of organic chemistry and drug discovery applications	optimized using compounds in the CSD

ReaxFF	van Duin, Adri C. T.; Dasgupta, Siddharth; Lorant, Francois; Goddard, William A. (2001). "ReaxFF: A Reactive Force Field for Hydrocarbons". <i>The Journal of Physical Chemistry A</i> 105 (41): 9396–9409	Chemical reactions	
Martini	Marrink, Siewert J.; de Vries, Alex H.; Mark, Alan E. (1 January 2004). "Coarse Grained Model for Semiquantitative Lipid Simulations". <i>The Journal of Physical Chemistry B</i> 108 (2): 750–760.	Coarse-grained FF for lipids (also proteins and carbohydrates)	Four heavy atoms map to a single site
GAFF	http://ambermd.org/antechamber/gaff.html	Rational Drug Design	
GLYCAM	http://glycam.org/docs/forcefield/	Carbohydrates	

Electrostatic (Coulombic) interactions

$$V_{el} = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

q_i = partial charge on site i

q_j = partial charge on site j

r_{ij} = distance

ϵ_0 = permittivity of free space

$$V(r) = 332 \frac{q_i q_j}{\epsilon r_{ij}} \quad \frac{\text{kcal}}{\text{mol}}$$

← when q_i & q_j are measured in electrons & r_{ij} is in Å

ϵ is the dielectric constant of intervening medium

$\epsilon \approx 1$ in a vacuum

≈ 2 in non-polar liquid hydrocarbons

≈ 80 in liquid H₂O

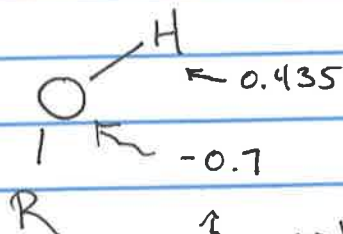
Partial Charges can be very tricky to assign

We're interested in the net charge on each nucleus but electrons are not distributed equally.

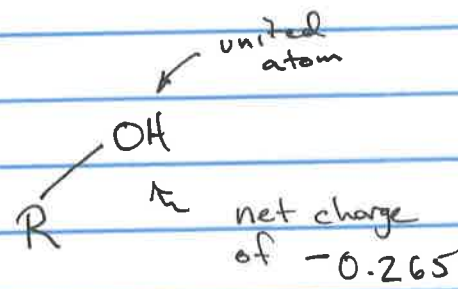
The charges in a force field will be taken from Mulliken population analysis, CHelpG, RESP, etc.

Higher order moments

Consider an alcohol moiety:



partial charges in Amber



but also a point dipole

Dipoles measure polarity of a site

This is a vector quantity (magnitude & direction)

For 2 charges of equal magnitude q but opposite sign, separated by a vector $\vec{r}$



$$\vec{\mu} = q\vec{r}$$

The direction (by convention) is from $\oplus$ to $\ominus$ charge.

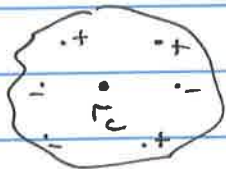
With multiple charges

$$\mu_x = \sum_i q_i (x_i - x_c)$$

charge at site i

relative x-coordinate of site i

x-component of dipole



The dipole depends on the choice of origin for charged species and is usually referenced to the center-of-mass or center-of-charge.

Dipole-Dipole interactions

$$V(R) = \frac{|\mu_i||\mu_j|}{4\pi\epsilon_0 r_{ij}^3} \left(\hat{\mu}_i \cdot \hat{\mu}_j - 3(\hat{\mu}_i \cdot \hat{r}_{ij})(\hat{\mu}_j \cdot \hat{r}_{ij}) \right)$$

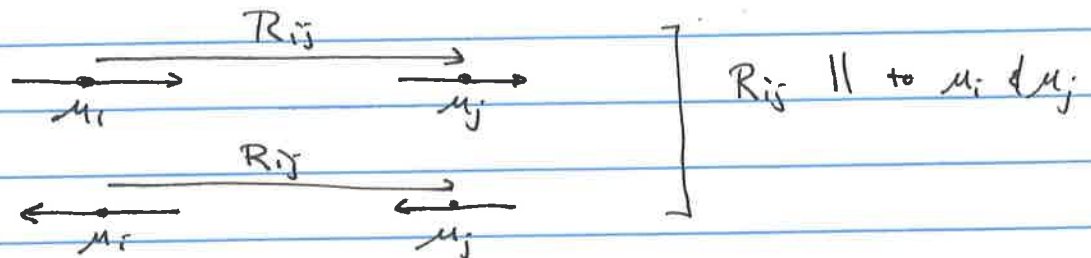
$\hat{\mu}$ = unit vector

$$\hat{\mu}_i = \frac{\vec{\mu}_i}{|\vec{\mu}_i|}$$

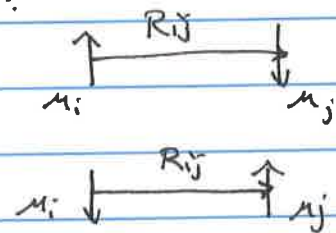
$$V(R) \sim \frac{1}{R^3} \leftarrow \text{overall distance dependence}$$

Orientational Dependence

The Dipole-Dipole interaction favors Aligned head-to-tail interactions:



Anti-aligned perpendicular interactions:



Ion-Dipole interactions scale like $V(R) \sim \frac{1}{R^2}$

Vander Waals or induced-dipole induced dipole interactions are relatively weak, but they are ubiquitous! Even neutral atoms have London dispersion forces.

The attractive nature can be derived from QM using perturbation theory or classically using Drude models for charge motion. In either case

$$V_{\text{dispersion}} \approx -\frac{1}{R^6}$$

← attractive

← distance dependence

(19)

Exchange-repulsion is the dominant repulsive force at shorter ranges, and this typically goes like $e^{-\alpha R}$

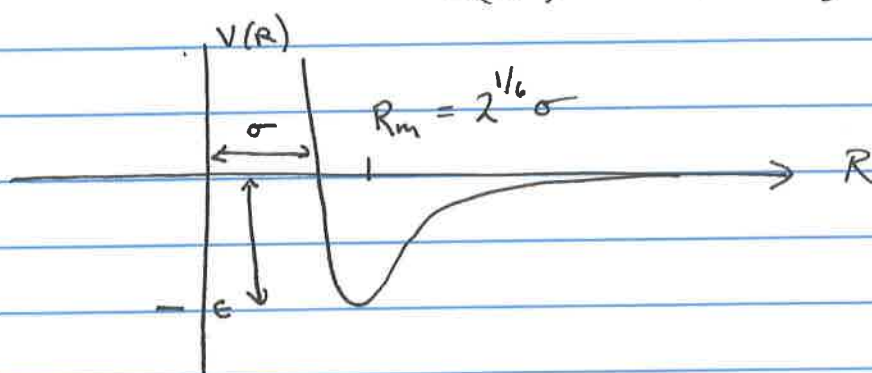
The Buckingham Potential tries to mimic the correct behavior for both dispersion & exchange-repulsion.

$$V(R) = \epsilon \left[\frac{6}{\alpha-6} e^{-\alpha \left(\frac{R}{R_m} - 1 \right)} - \frac{\alpha}{\alpha-6} \left(\frac{R_m}{R} \right)^6 \right]$$

Unfortunately this potential badly misbehaves at ^{very} short distances

The Most common (dispersion + exchange-repulsion) model is the Lennard-Jones potential

$$V_{LJ}(R) = 4\epsilon \left[\left(\frac{\sigma}{R} \right)^{12} - \left(\frac{\sigma}{R} \right)^6 \right]$$



Equivalently :

$$V_{LJ} = \epsilon \left[\left(\frac{R_m}{R} \right)^{12} - 2 \left(\frac{R_m}{R} \right)^6 \right]$$

or

$$V = \frac{A}{R^{12}} - \frac{C}{R^6}$$

The $\frac{1}{R^{12}}$ term is completely ad hoc. Actual repulsions are probably a bit weaker, but this is the dominant model!!

Other less common ways of handling dispersion + exchange-repulsion

(19.5)

Mie:
$$V(r) = \frac{n}{n-m} \left(\frac{n}{m}\right)^{n/n-m} \epsilon \left[\left(\frac{\sigma}{r}\right)^n - \left(\frac{\sigma}{r}\right)^m \right]$$

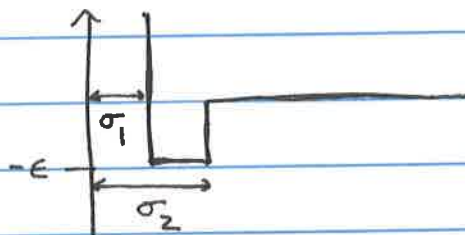
↑
generalization of LJ

Sutherland:
$$V(r) = \begin{cases} -\epsilon \left(\frac{\sigma}{r}\right)^6 & r > \sigma \\ \infty & r \leq \sigma \end{cases}$$

Ideal Gas: $V(r) = 0$

Hard sphere:
$$V(r) = \begin{cases} 0 & r > \sigma \\ \infty & r \leq \sigma \end{cases}$$

Square-Well or "Go" model :



"Mixing Rules" : Force fields may be simple to parameterize for a fixed set of atom types, but all of the cross interactions multiply the effort dramatically

One common heuristic is the use of Mixing rules:

$$\sigma_{AB} = \frac{1}{2} (\sigma_{AA} + \sigma_{BB})$$

$$\epsilon_{AB} = \sqrt{\epsilon_{AA} \epsilon_{BB}}$$

← These are the Lorentz-Berthelot rules, which are common, but not universal.