This semester

Andrew Torda, April 2010

- Models – mostly proteins
  - from detailed to more abstract models

Books

- None necessary
  - for my group and Prof Rarey
    - “Molecular Modelling: Principles and Applications” Leach, Andrew
  - some later material (Monte Carlo) from
    - "Understanding Molecular Simulation", Frenkel and Smit

Grand Plan

- Models for proteins (mostly)
  - from detailed to less details
- Energies
- Dynamics
- Basis of them
- How to work with them
  - simulations, calculations
    - protein folding, evolution

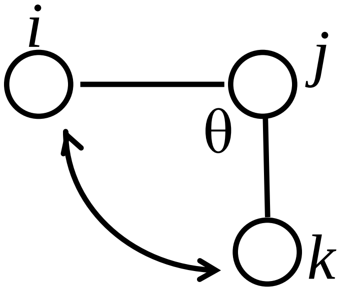
- What kind of model is appropriate for different problems ?

Remind Me

- List of lecture and übung topics

Some questions

- 3 bonded atoms – push atom *k* left
  - which atoms will feel a force ?



- I have a multiple sequence alignment
  - the conserved sites are most important
  - is this true ?

- Can you compare the free energy of two conformations of
  - a drug
  - a protein ?

- to be answered during the semester

Do I have to memorise all the formula ?

- Very few
  - Coulombs law
  - example form of energy for bonds and angles
  - Boltzmann distribution
  - definition of entropy
  - relationship of free and potential energy

- Most other examples will be derived

HORROR

- today
  - lots of lectures

Atomistic Energy Models

Andrew Torda, April 2010

- Why do we need models ?
- Previous lectures – need for low energy configurations
  - really needs definition of energy
- Can we define energy ?
  - for very simple systems yes
  - for more complex systems
    - only approximations
- Need to know when approximations are small and when bad
    - are charges on atoms like fixed charges ?
      - not for chemical reactions
    - are bonds like springs ?

Is energy sufficient ?

Does the world care about potential energy ?

- no. Really cares about free energy

  $G = U - TS$ (usually speak of ΔG)

Approach to free energy

- calculate potential energy "U" with a model
- get entropy "S" from some sampling method (often implicit)

Need good models for energy

Definitions

- This topic is classical / atomistic
- Often referred to as “molecular mechanics”
- quantum effects are not reproduced

Different levels of models

	ab initio	very detailed
QM		
	semi-empirical	
atomistic		
coarse grain continuous		
lattice		
big molecule as single point		very coarse (colloids, diffusion)

Why we like atomistic models

- intuitive
    - how do we draw structures ? store coordinates ?
- atoms sometimes correspond to measurable properties
    - x-ray crystallography, NMR
- predicting some dynamics
- interactions between proteins, proteins + ligands

What do we want to be able to do ?

- Often to simulate a system
    - Monte Carlo, molecular dynamics (MD)
    - details later, for MD we need
    - $F = ma$ (Newton), $a = F/m$ gives us acceleration and
        - $F = -dU / dr$ or better $\vec{F} = -dU \big/ d\vec{r}$
        - important.. friendly energies have proper derivatives

Force fields in general

- What is a force field ?
  - the set of equations / formulae that tell us about the force acting on a particle
  - classic example
    - I have charge
    - if I bring another charge near, it feels a force due to the first
    - what is the equation telling me about the energy ?

$$U(r_{ij}) = \left(\frac{1}{4\pi\varepsilon_0}\right)\frac{q_i q_j}{r_{ij}}$$

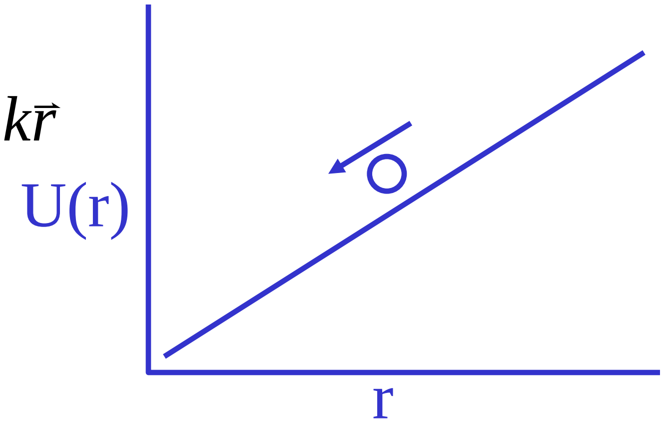
    - and the force…

Force fields, energies, derivatives

- Important rule $\vec{F} = -dU/d\vec{r}$

- a toy example, one dimension $U(\vec{r}) = k\vec{r}$
- force is $\vec{F} = -dU/d\vec{r}$

$$= -k$$

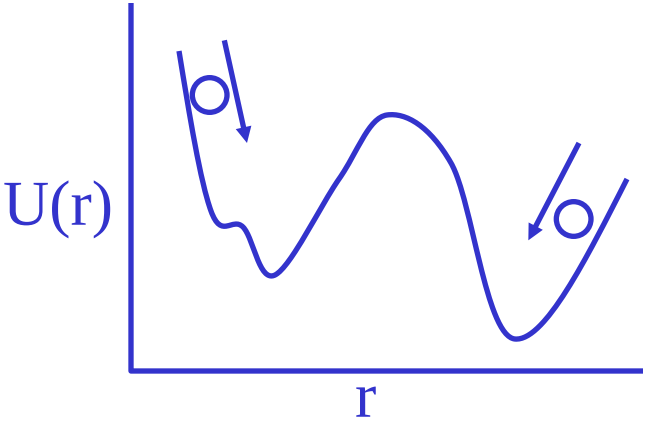


- how do we want force ?
  - in *x*, *y*, *z* terms $\vec{r}$

- really $F_x = -\partial U/\partial x$

$$F_y = -\partial U/\partial y$$

$$F_z = -\partial U/\partial z$$



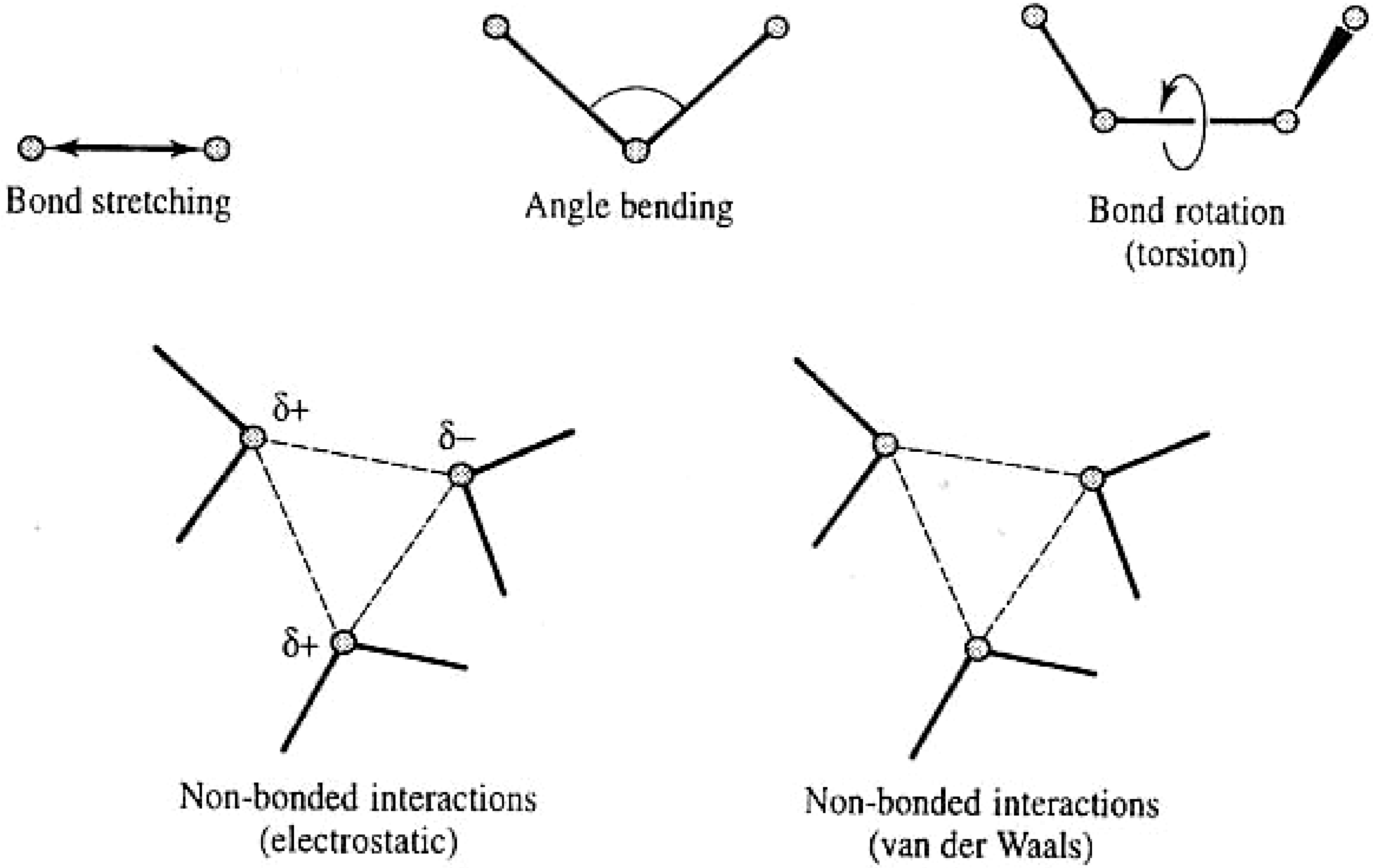
A simple force

- With an energy $U(r_{ij}) = \left(\frac{1}{4\pi\varepsilon_0}\right)\frac{q_i q_j}{r_{ij}}$

- force is $\vec{F} = -dU/d\vec{r} = -\left(\frac{1}{4\pi\varepsilon_0}\right)\frac{q_i q_j}{r_{ij}^2}\hat{r}_{ij}$

- Rule
  - if the derivative of energy is non-zero
    - there is a force
  - exam questions about bond angles, dihedrals, …

protein force field (picture)



Bond stretching

Angle bending

Bond rotation
(torsion)

Non-bonded interactions
(electrostatic)

Non-bonded interactions
(van der Waals)

picture from Leach, A.R., Molecular Modelling Principles and Applications

Protein force fields scary equation

$$
\begin{aligned}
U(r) = & \sum_{i=1}^{Nbond} \frac{k_i}{2}\left(r_i - r_{i,0}\right)^2 && \text{bonds} \\
+ & \sum_{i=1}^{Nangle} \frac{k_i}{2}\left(\theta_i - \theta_{i,0}\right)^2 && \text{angles} \\
+ & \sum_{i=1}^{Ndihedral} k_i\left(1+\cos\left(n\varphi_i - v_0\right)\right) && \text{dihedrals} \\
+ & \sum_{i=1}^{Natom} \sum_{j=i+1}^{Natom} 4\varepsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right) + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} && \text{non} - \text{bonded}
\end{aligned}
$$

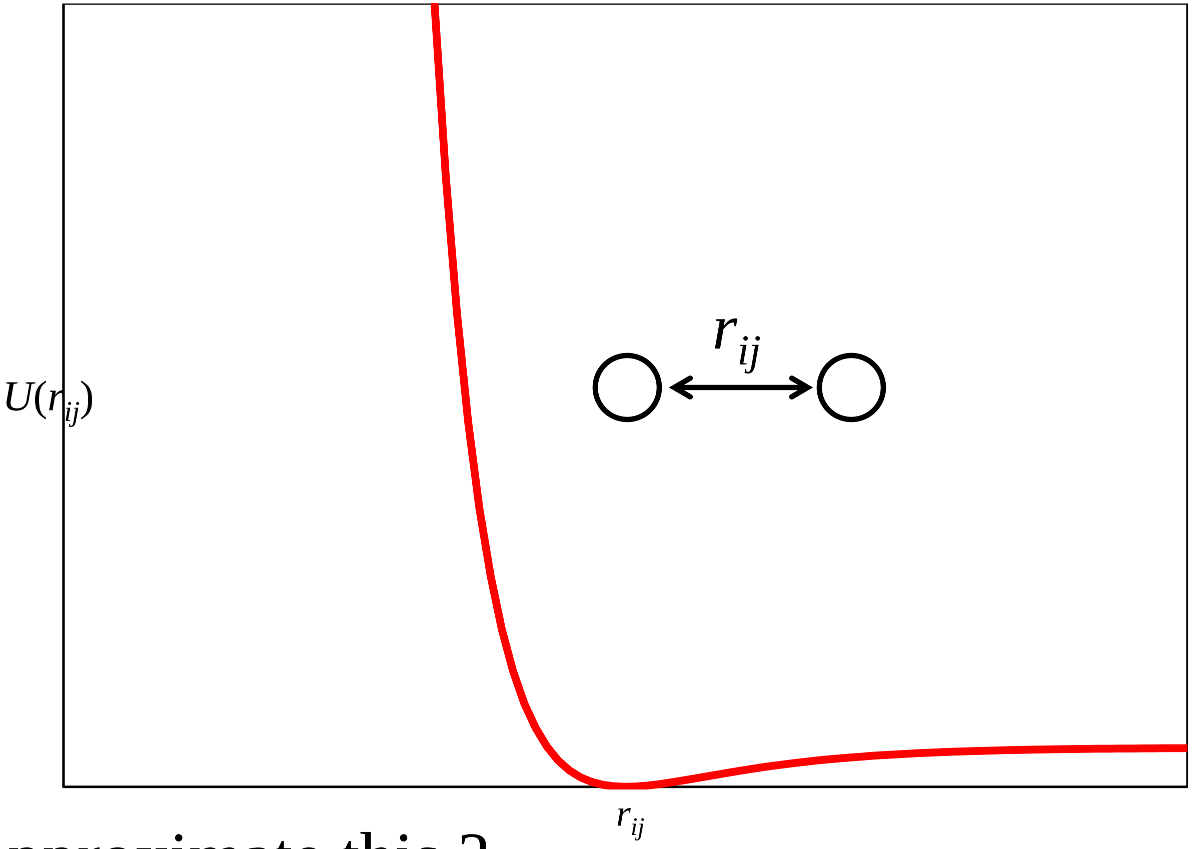
Is this the truth ?

- not a bad model

Bonds

What are bonds really ?

- as we pull two particles apart, what happens ?

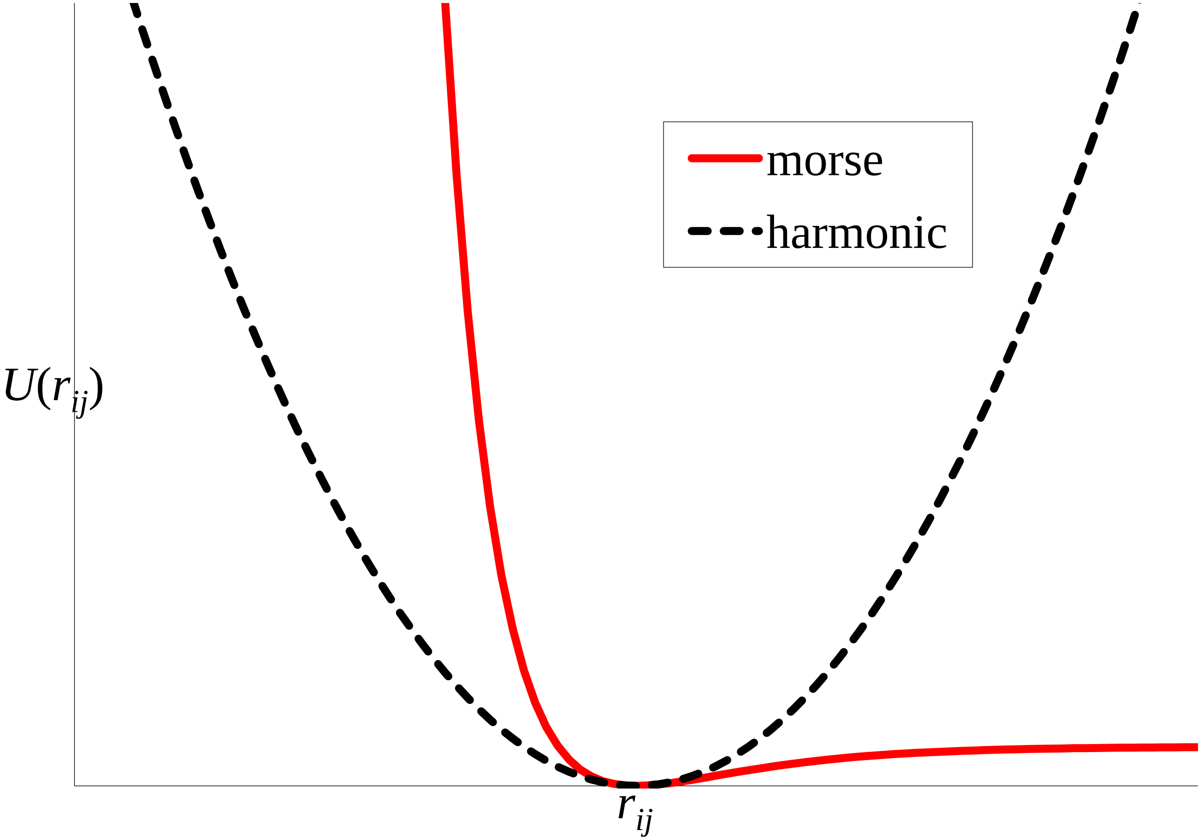


Can we approximate this ?

- this formula has exp() terms [cpu expensive]

Bonds - approximate

- Is the black line a good approximation ?

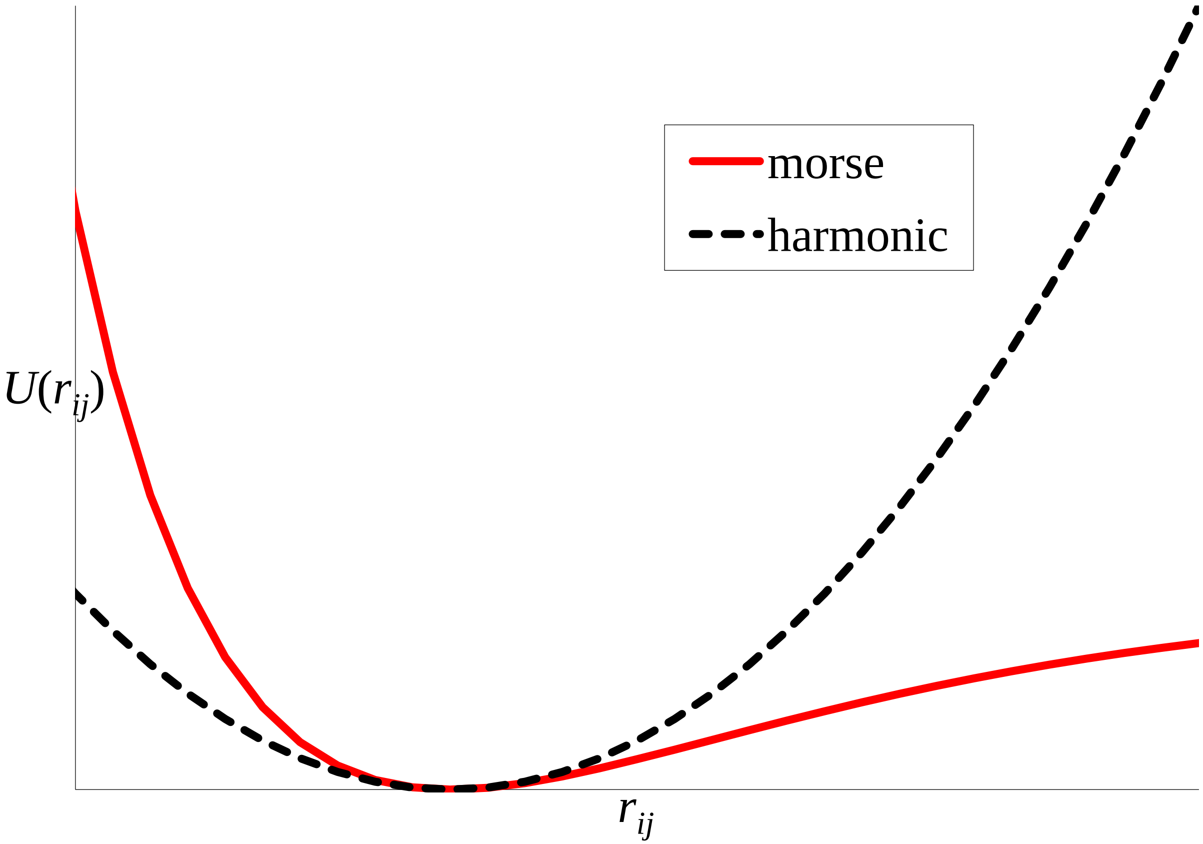


$$U(r_{ij}) = \frac{k}{2}\left(r_{ij} - r_0\right)^2$$

Bonds – good enough

Bonds don't stretch much

- harmonic is often good enough



Why do we write $k/2(r_{ij}-r_0)^2$?

Bonds – the end

- We normally write $U(r_{ij}) = k/2\ (r_{ij} - r_0)^2$

$$\vec{F}_i = -\frac{k_{ij}}{2}\frac{dU\left(\left|\vec{r}_{ij}\right| - r_{ij,0}\right)^2}{d\vec{r}_{ij}}$$

$$= -k_{ij}\left(\left|\vec{r}_{ij}\right| - r_{ij,0}\right)\vec{r}_{ij}$$

- so first form looks like Hookes law

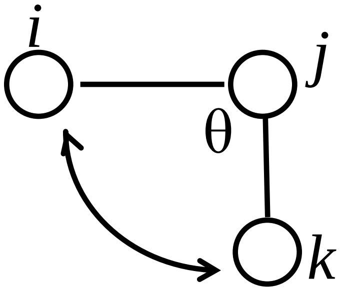
Do bonds matter ?

- fluctuations very small at room temp (< 0.1 Å)
- our structures are not so accurate / we simulate to look at coarse features
- often treated as rigid joints (maybe more in MD lectures)

Angles

- Angles are not as boring as bonds

$$U_{angle}\left(\vec{r}_i, \vec{r}_j, \vec{r}_k,\right) = \frac{k}{2}\left(\theta_{ijk} - \theta_0\right)^2$$



or

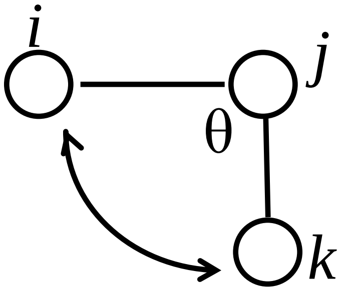
$$U_{angle}\left(\vec{r}_i, \vec{r}_j, \vec{r}_k,\right) = \frac{k}{2}\left(\cos\theta_{ijk} - \cos\theta_0\right)^2$$

but to get forces is messy

(use chain rule)

Why is this difficult ?

- We use physical models like



$$U_{angle}\left(\vec{r}_i, \vec{r}_j, \vec{r}_k,\right) = \frac{k}{2}\left(\cos\theta_{ijk} - \cos\theta_0\right)^2$$

- force is obvious
  - looking at picture
  - in terms of cos *θ*
  - in terms of *θ ?*
  - we want *x*, *y*, *z*

$$F_{angle}\left(\vec{r}_i\right) = {}^{-\partial U(\vec{r}_i)}\Big/_{\partial\left(\vec{r}_i\right)}$$

$$= \frac{-\partial U(\vec{r}_i)}{\partial\cos\theta}\,\frac{\partial\cos\theta}{\partial\left(\vec{r}_i\right)}$$

$$= \frac{-\partial U(\vec{r}_i)}{\partial\cos\theta}\,\frac{\partial\cos\theta}{\partial\theta}\,\frac{\partial\theta}{\partial\left(\vec{r}_i\right)}$$

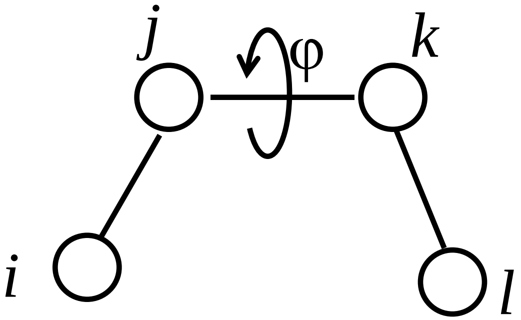
Angle forces

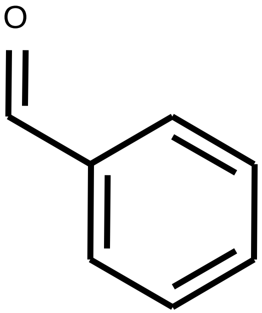
$$
\begin{aligned}
F_{angle_i} &= \frac{-\partial U_{angle}(\vec{r}_i)}{\partial \vec{r}_i} \\
&= \frac{-\partial U_{angle}(\vec{r}_i)}{\partial \cos\theta_{ijk}} \frac{\partial \cos\theta_{ijk}}{\partial \vec{r}_i} \\
&= -k\left(\cos\theta_{ijk} - \cos\theta_0\right)\left(\frac{\vec{r}_{kj}}{r_{kj}} - \frac{\vec{r}_{ij}}{r_{ij}}\cos\theta_{ijk}\right)\frac{1}{r_{ij}}
\end{aligned}
$$

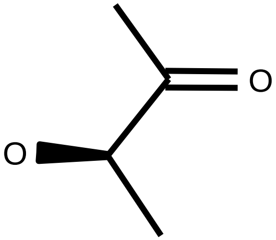
- the other atoms ?
  - similar expression for F_j
  - $F_k = -(F_i + F_j)$

dihedral / torsion angles

- basically…
- but details will vary



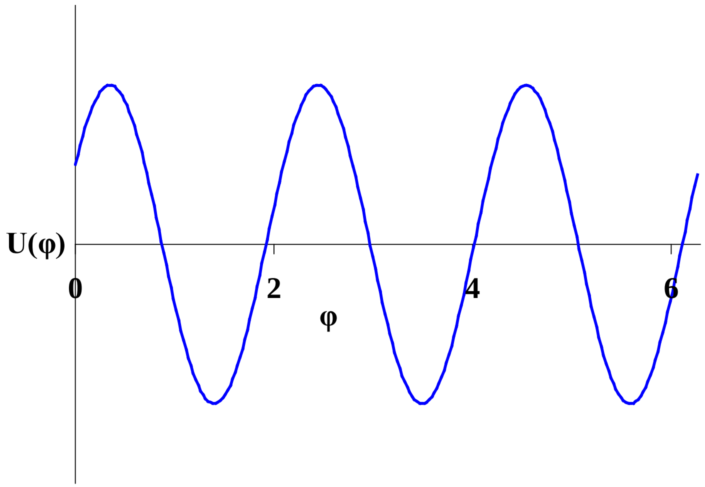


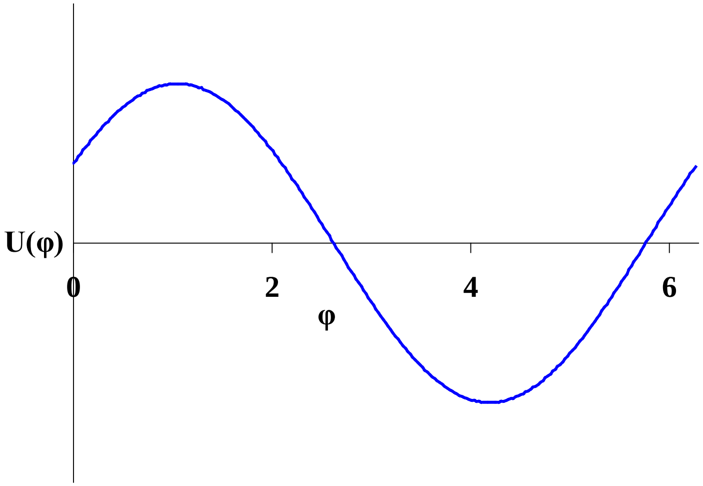


- how large are energy barriers ?
- how many minima are there ?

Form of dihedral term

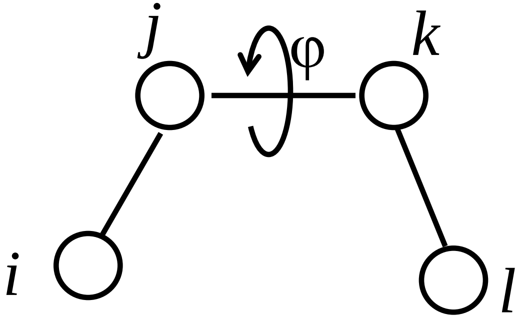
Maybe something has three minima or one minimum





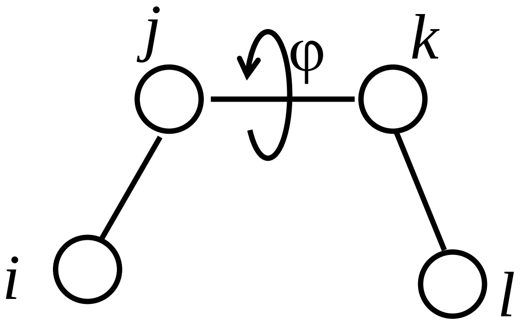
What are the causes ?

- do *i* and *l* interact ?
- electron clouds from *j* and *k* ?
- model $k_i(1+\cos(n\varphi_i - v_0))$



model for dihedral angles

$$U(\vec{r}) = k_i\left(1 + \cos\left(n\varphi_i - v_0\right)\right)$$



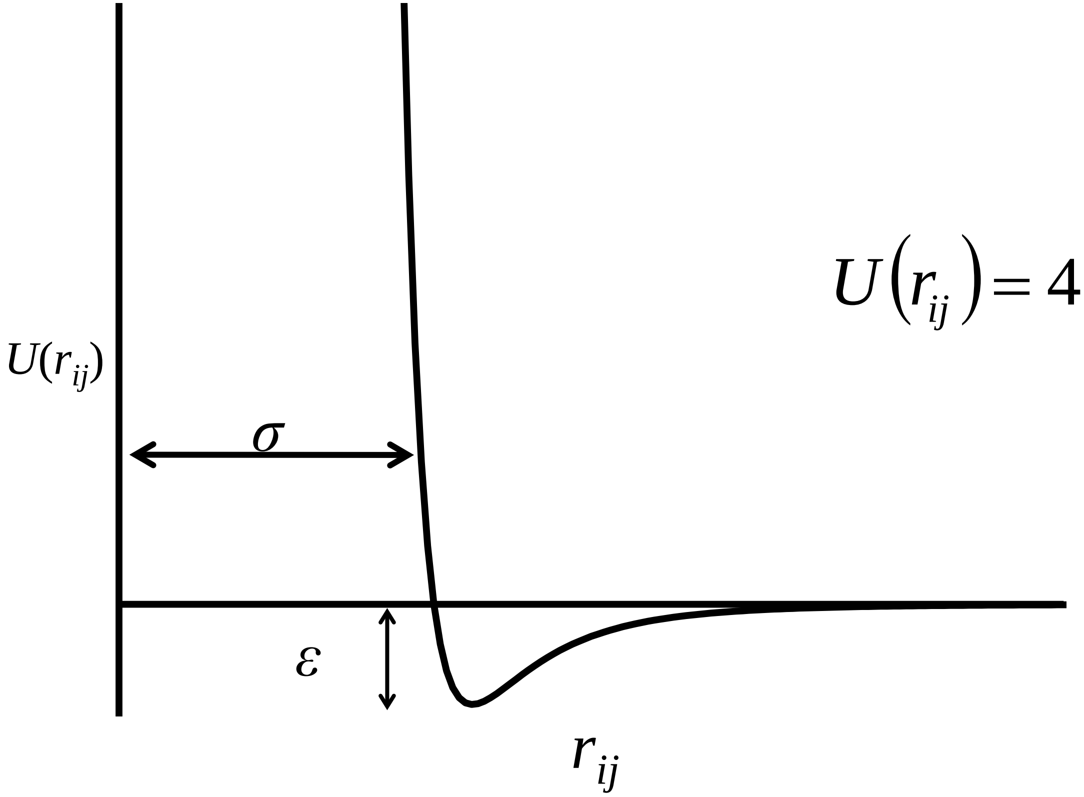
- Properties ?
    - *n* controls multiplicity
        - *n*=3 butane
        - *n*=2 peptide bond
- at room temperature, barriers are similar to *kT* so
    - atoms do rotate, but there are preferences (from spectroscopy)
- how good is the model ?
    - some rotamers are preferred (need other terms)
- What do forces look like ? scary (much trigonometry)
    - intuitively easy, maths messy- think of *j* and *k*

Non bonded forces

- van der Waals / Lennard-Jones / dispersion + attraction
- electrostatic
- why are they separated from others ?
  - bonds, angles and dihedrals
    - you know the participants in advance
  - non-bonded
    - atoms can move to and from each other

van der Waals

- what do we know in advance ?
  - “inert” gases do form liquids (atoms like each other)
  - atoms do not sit on top of each other



$$U(r_{ij}) = 4\varepsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right)$$

$$\text{distance of minimum energy} = 2^{1/6}\sigma$$

van der Waals – how good is it ?

$$U\left(r_{ij}\right) = 4\varepsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right)$$

- repulsive and attractive look similar
  - quite different
- attractive part ?
  - electrons wobble and talk to neighbours – induce charges
  - r^{-6} pretty good
- repulsive part ?
  - what do atoms look like ?
  - really exponential, r^{-12} is very convenient

Lennard-Jones terms – how real

- how real is it ?
  - good for liquid argon
    - diffusion, transport…
- Lennard-Jones fluids
- nice features
  - ε, σ = “well depth” and size
  - ε, σ specific for atom pairs
    - σ for H is tiny, for C is much bigger
- typical exam question
  - dimensions of energy are ML^2T^{-2}
  - if the units of σ are Angstrom Å (dimension L)
    - what are the units (or dimensions) of ε ?

Electrostatics

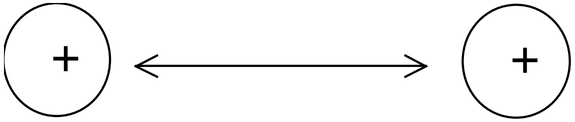
- Coulombs law
  - Sounds easy

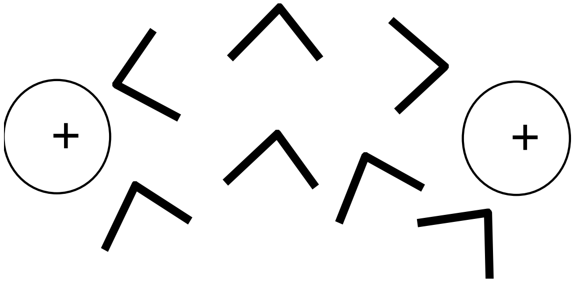
$$U\left(r_{ij}\right)=\frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

sometimes

$$U\left(r_{ij}\right)=\frac{q_i q_j}{D r_{ij}}$$

- where D is dielectric constant
- why is it difficult ?
  - what is D ?
  - with and without water ?
  - intervening protein ?





  - more on solvent models later

Model can be made better / worse

Is this model good or bad ? fast or slow ?

What will it be used for ?

- molecular dynamics simulations, energy minimising, rarely a simple energy evaluation
- very cpu intensive

Two conflicting goals

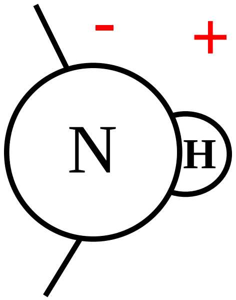
- make model cheaper, but maintain quality
- add details to make model better

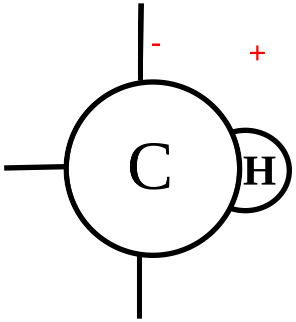
Cost of model ?

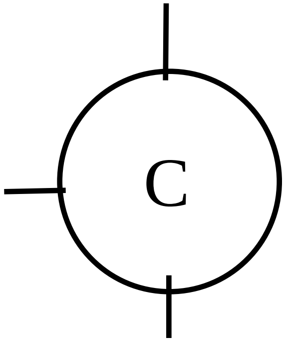
- bonds, angles, dihedrals $O(n)$
- non-bonded n particles interact with n particles $O(n^2)$
- probably ≈ 90% of time spent on non-bonded interactions

Cheaper models - United atoms / heavy atoms

- When does a proton matter ?
  - charge interactions, H bonds
- When is a proton not interesting ?
  - most aliphatic and aromatic
  - hardly a charge / tiny radius
- Do we even need the hydrogen ?
  - Use a "united atom"
    - mass – easy 12 + 1
    - charge ? nothing
    - radius slightly larger
- Rule
  - Use explicit H in polar groups
  - absorb everywhere else
  - CH, CH_2, CH_3

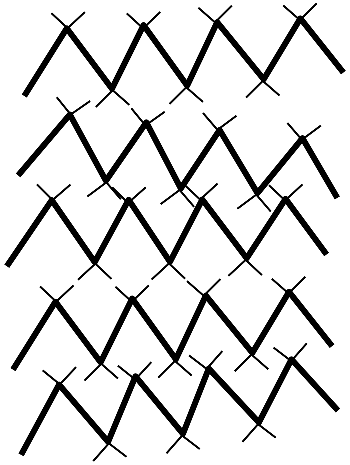






United atoms how bad ?

- Gain of united atoms ?
    - roughly halve number of atoms
    - Costs ?
        - dynamics ? no problem
        - structure ? no real problem
- Problems ?
    - some small effects can be seen in certain systems
        - lipids
    - maybe some effects in proteins



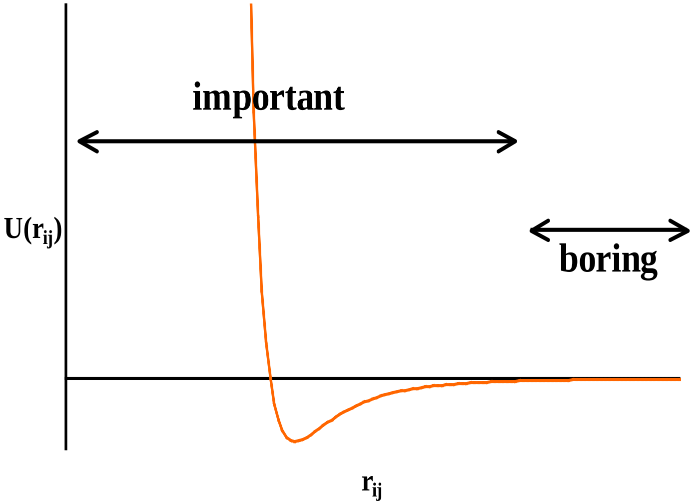
Cutoffs (cheaper models)

How important is an interaction ?

- some are not interesting (bonds)
- some depend where you are

$$U(r_{ij}) = 4\varepsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right)$$

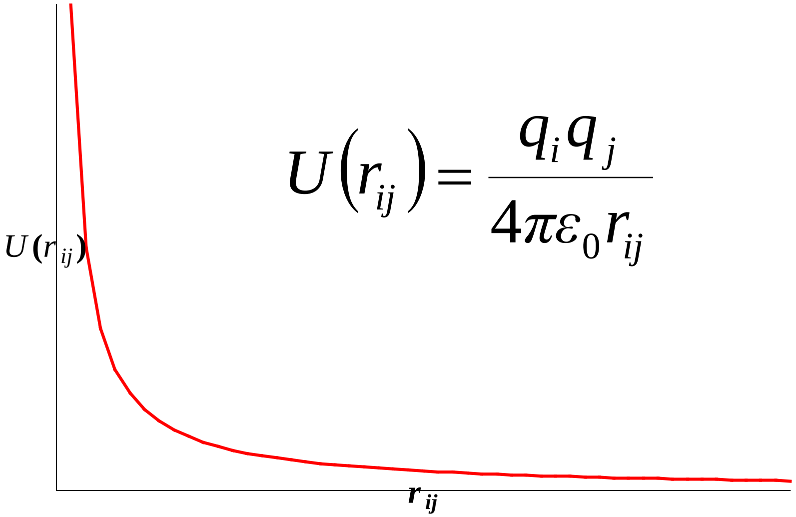
  - r^{-6} and r^{-12} become small quickly
  - if $r_{ij} > 6\text{-}8$ Å, $U_{LJ}(r_{ij}) = 0$
    - not a problem



- more difficult.. for other terms

cutoffs - problems

- electrostatics
- r^{-1} shrinks slowly
- but eventually, we should be able to ignore
- bigger cutoff (10 – 15 Å)



$$U(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

- in practice
  - use one cutoff for all non-bonded calculations
- problems
  - subtle – look at derivative

Possible improvements

Better bonds ?

- more sophisticated than harmonic $(r\text{-}r_0)^2$
- not very interesting

L-Jones better than r^{-12} ?

- not worth worrying about

Electrostatics

- polarisation ?
  - completely lacking from model so far (fixed partial charges)
  - we know it is important
  - popular, difficult

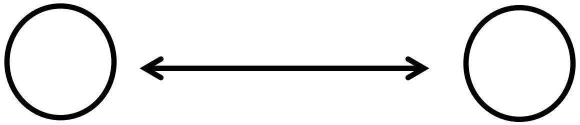
Water

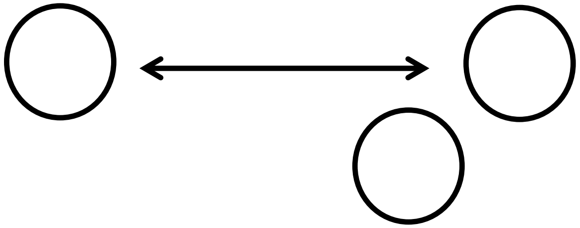
- model so far is *in vacuo* ... more later

Problems and fixes in model

Model is fundamentally wrong

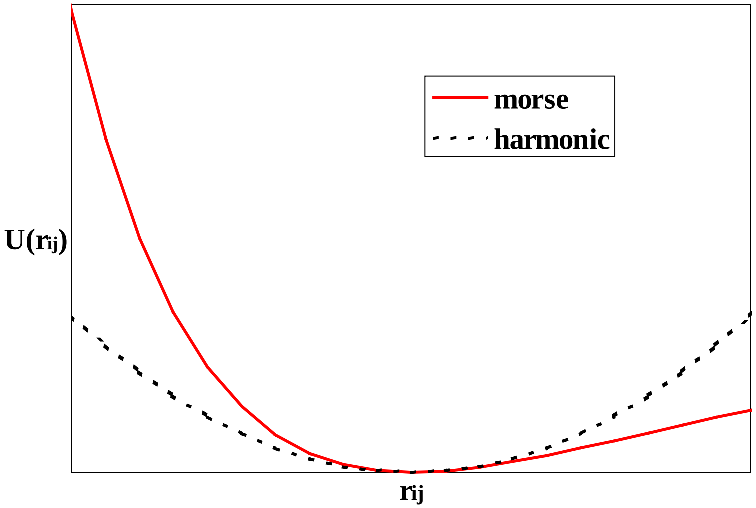
- think about forms of functions
- all two-body based, compare...





•can we represent interactions with two body form ?

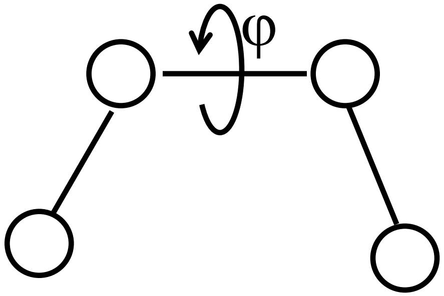
- •can be good for a small range
- •remember earlier picture...
- •what works well at 300K may not work at 600K



Why problems may not be seen

Model is not perfect, why do simulations work ?

- lots of parameters, not independent
- example
  - what controls density of a polar fluid
    - charges ?
    - Lennard-Jones terms ?
    - both
- complicated example
  - rate of rotation..
    - torsional term ***
    - size of atoms and barriers (ε, σ) **
    - angles ? a bit
  - errors in one part of force field compensated elsewhere



Testing force fields

What should a force field do ?

- if I simulate a protein
  - it should not blow up
    - necessary / not sufficient
- reproduce energy changes
- global minimum (free) energy should agree with experiment

Dream properties

Perfect model of physics would work in all cases

Transferability

- atomic parameters same from protein to protein
- from protein to organic molecule
- across temperature ranges ?

Disappointments /difficulties

Special systems / special problems

- highly charged systems
  - DNA
    - solvent and charges
- lipids
  - repetitive nature emphasises some problems

Meaning of disappointment

- simulate a protein and it falls apart
- it implodes
- density of a system is wrong
- energetic predictions are wrong
- dynamic predictions are wrong
- smaller structural predictions are wrong

Parameters

Force field / model has lots of parameters

- charge, mass, ε, σ, bondlengths, angles, …

Sources

- literature
    - mass
- partial charges ?
    - high level calculations
- measurements on small molecules
    - bond lengths, geometry
- trial and error (example)
    - simulate a liquid and
        - reduce σ to increase density
        - decrease ε to make it boil more easily

Parameters are a compromise

Model is not perfect

- internal compensation
- compromise example
  - partial charges are not really fixed
  - depend on environment + geometry
- make a decision and adjust others to work in important area

Atomistic force field summary

- very good model for much
  - protein dynamics
  - interactions
  - structure
- main model
  - ignore water
  - 3 kinds of bonded interactions
  - 2 non-bonded
  - many frills possible, some necessary
- conservative force field
  - $U(r_{ij})$ energy depends on coordinates only
    - no time component
- energy continuous
  - useful
- derivative always defined
  - very useful

Where next ?

- Better models (not much)
- Fixing worst aspects
- Simpler models

- Simulation
- first some rules from statistical mechanics