

This semester

Models – mostly proteins

- from detailed to more abstract models

Some simulation methods

Books

None necessary – for my group and Prof Rarey

- “Molecular Modelling: Principles and Applications” Leach, Andrew (good, not cheap)
- Essentials of Computational Chemistry : Theories and Models / Christopher J. Cramer (ebook, in library) bits of chap 2 & 3

Grand Plan

- Models for proteins (mostly)
 - from detailed to less details
- Energies
- Dynamics

How to work with them

- simulations, calculations
 - protein folding, evolution
- What kind of model is appropriate for different problems ?

Themes

- atomistische Energie-Modelle
- Energie / Freie Energie
- Monte Carlo (MC)
- Molecular Dynamics Simulations (MD)
- Applications and Extensions of MC and MD
- Advanced Monte Carlo
- Water
- Coarse Grain / Mesoscopic / low resolution methods
- Lattice Methods
- Evolution
- Protein Folding

Organization

Im Prinzip

- 2 × Vorlesung
- 1 × Vorlesung
- 1 × Übung

In der Praxis

- 2 × Vorlesung + 2 × Vorlesung or
- 2 × Vorlesung + 2 × Übung

People

- Timur Olzhabaev (later in semester)

Zeitplan

- Mo 10:15
- Di 12:15

Sprache ?

Organization

Klausurtermine in Stine (29 Juli, 25 Sep)

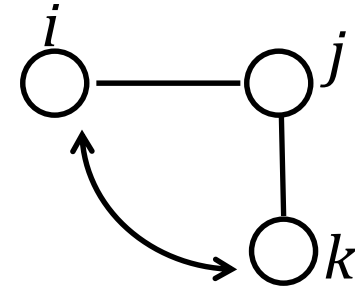
Andrew der Pedant - bei Klausuren

- energie/potenzielle energie/freie energie
- absolute Werte / relative Werte
- grossere / kleinere, mehr positiv / mehr negative
- Vektoren, Skalar
- Vorzeichnungen...

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

Why are we here ?

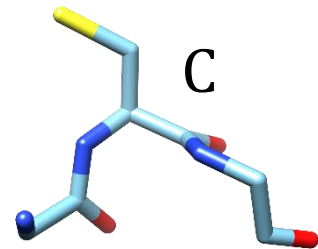
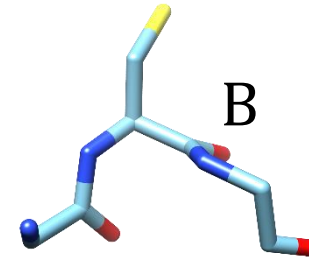
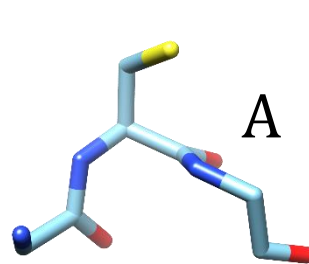
- observables
- distributions
- two ways to get at distributions
 1. from Boltzmann relation (later)
 2. from dynamics simulations
- dynamics simulations
 - need forces
- forces need energies

Observables

- \mathcal{A} is density, coordinates, something you can measure
- what you see is $\langle \mathcal{A} \rangle$ the average over time and molecules $\langle \mathcal{A} \rangle = \frac{1}{\tau} \int_0^\tau \mathcal{A}_t dt$
- simulating gives us a sampling over time / molecules

Can we just average over values ?

- 3 rotamers in a protein
- they are not equally likely
 - A is most common, C is high energy and not seen (no contribution)



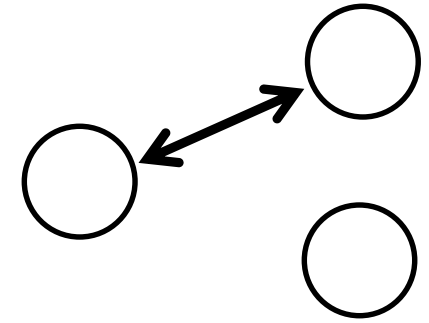
How to get the correct averages ?

- simulate the system and do the averaging over every \mathcal{A} in the simulation
- how would you simulate ?

Simulations need forces and energies

Distributions depend on energies

- we need energies



Dynamics simulations need forces

- forces come from energies $F = \frac{-dU}{dr}$

So we spend weeks discussing energies

Atomistic Energy Models

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 ?
- are bonds like springs ?

Is energy sufficient ?

Does the world care about potential energy ?

- No. Really cares about free energy

$$G = U - TS \quad (\text{usually speak of } \Delta G)$$

Approach to free energy

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

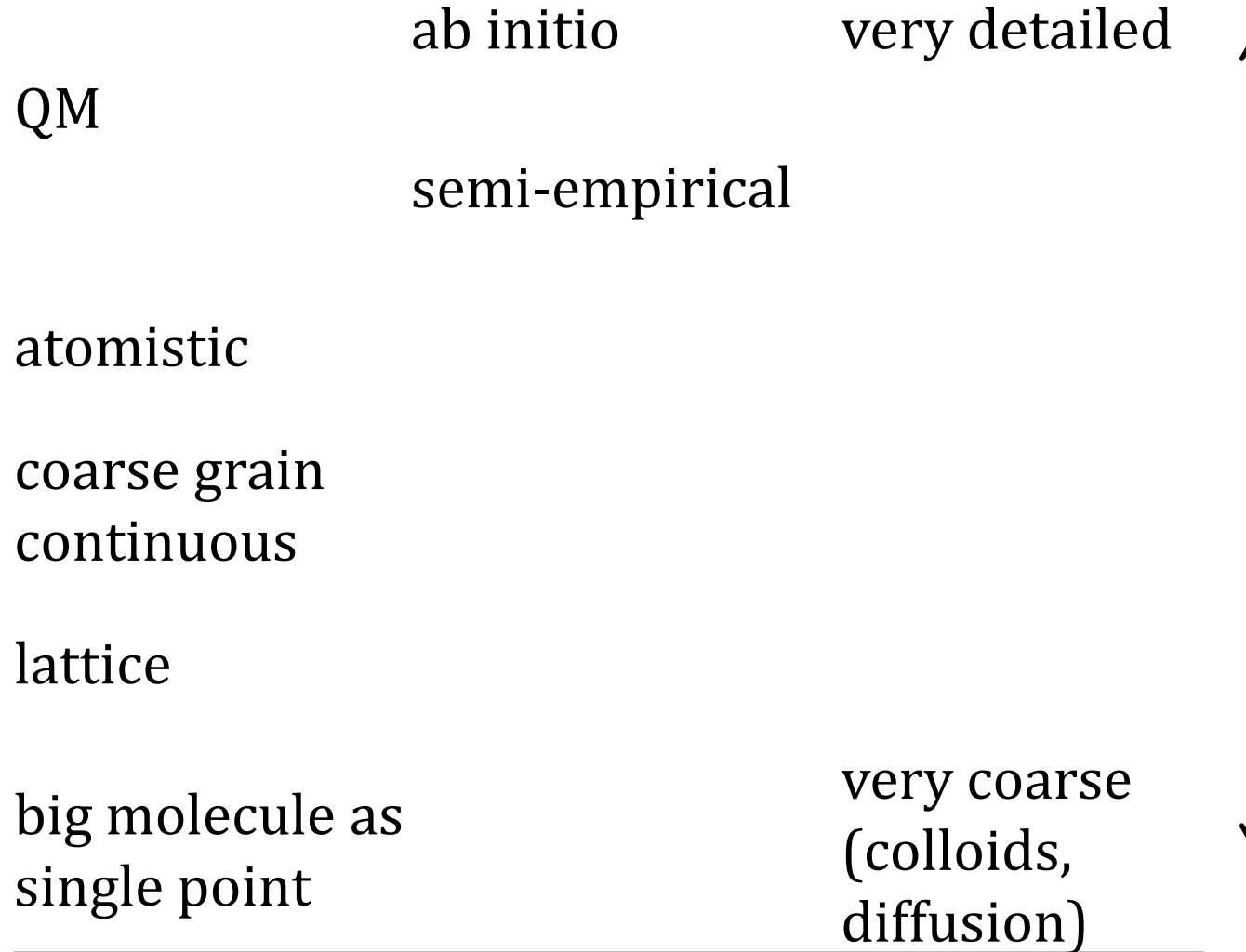
Need good models for energy

Definitions

This topic is classical / atomistic

- often referred to as “molecular mechanics”
- quantum effects are not reproduced

Different levels of models



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 is a force field ?

Set of equations / formulae that tell us about the force acting on a particle

Classic example

- I have charge
- bring another charge near, it feels a force due to the first
- what is the equation telling me about the energy ?



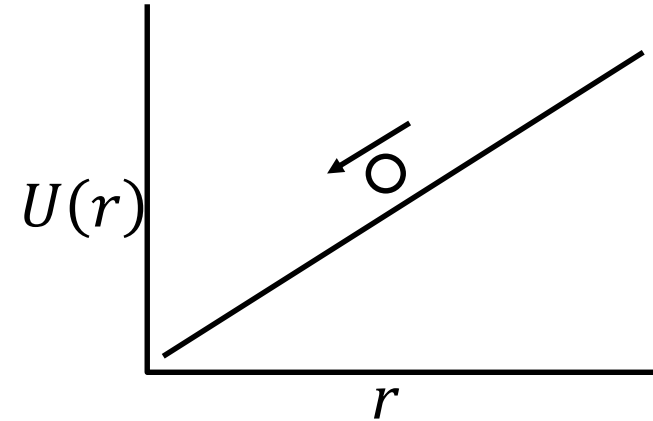
$$U(r_{ij}) = \left(\frac{1}{4\pi\epsilon_0} \right) \frac{q_i q_j}{r_{ij}}$$

and the force...

Force fields, energies, derivatives

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

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



How do we want force ?

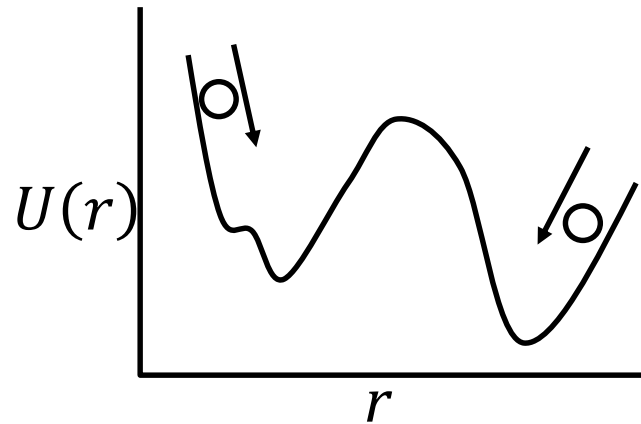
- in x, y, z terms \vec{r}

$$F_x = \frac{-\partial U}{\partial x}$$

$$F_y = \frac{-\partial U}{\partial y}$$

$$F_z = \frac{-\partial U}{\partial z}$$

really



A simple force

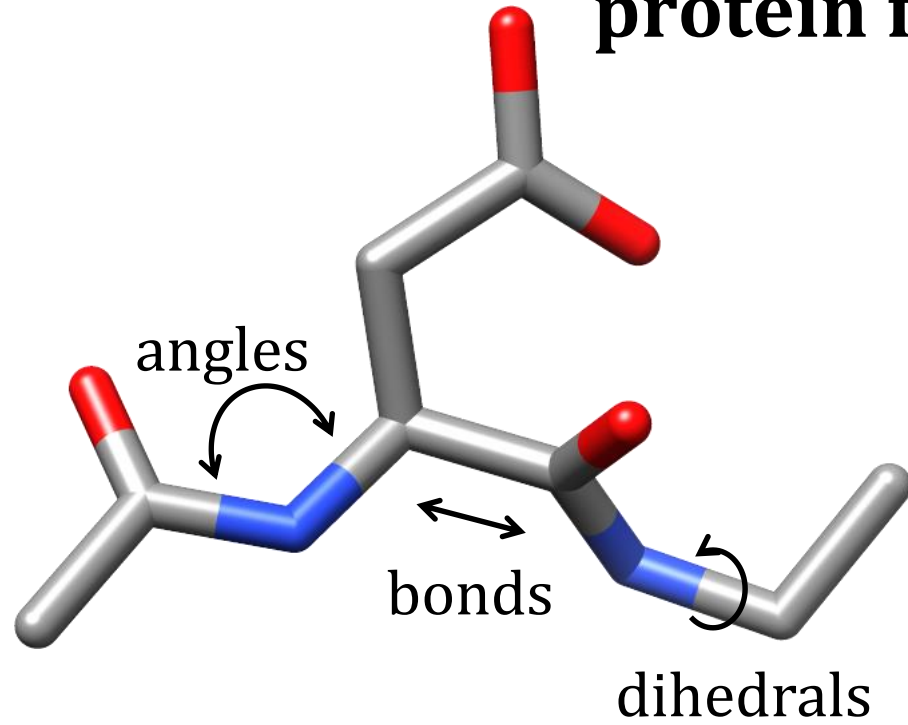
With an energy $U(r_{ij}) = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}$

force is $\vec{F} = \frac{-dU}{d\vec{r}} = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}^2} \hat{r}_{ij}$

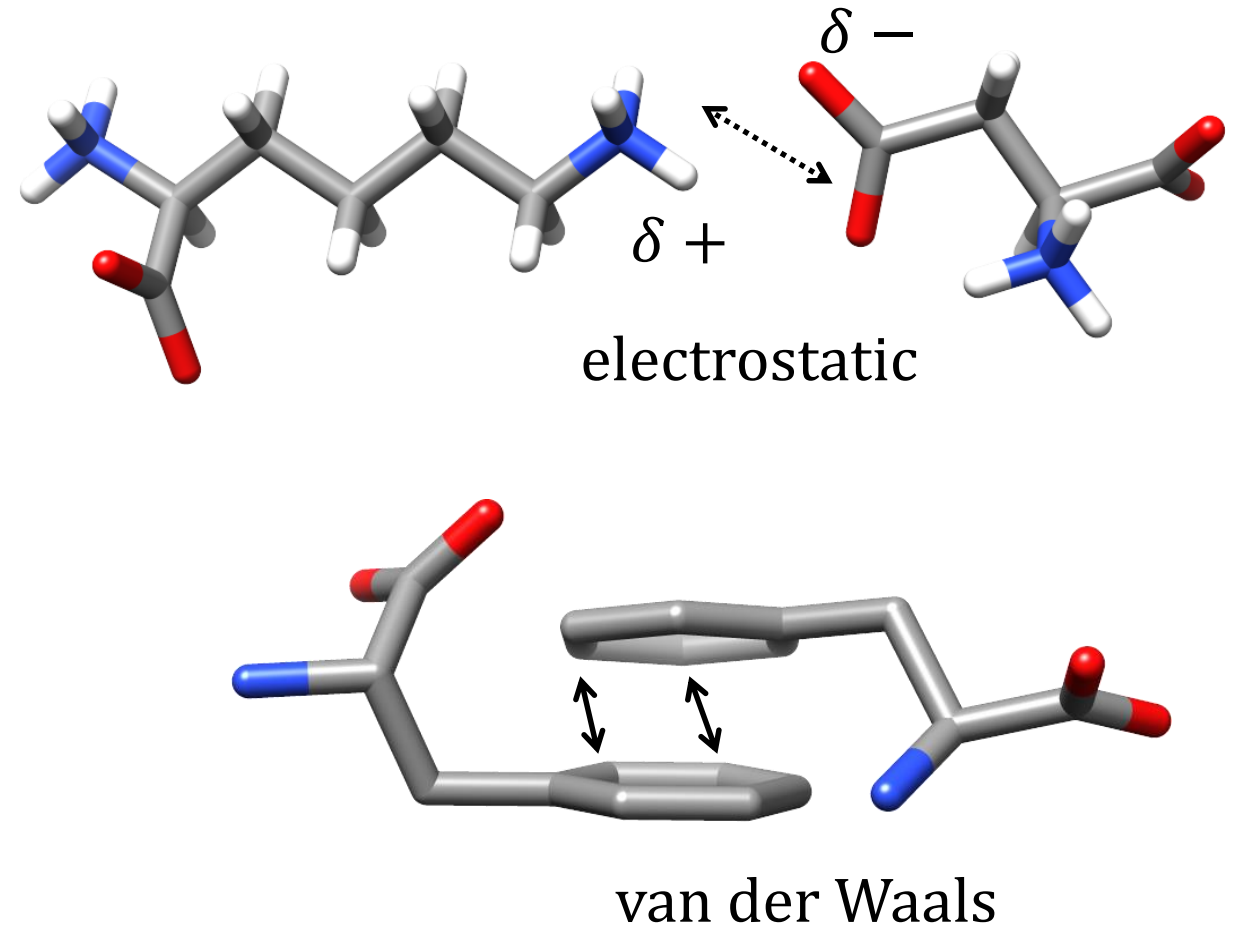
Rule

- if the derivative of an energy is non-zero
 - there is a force (remember for exams)

protein force field (pictures)



3 types of
bonded energies



2 types of
non-bonded energies

Protein force fields as equation

$$U(\{\vec{r}\}) =$$

$$\begin{aligned} & \sum_{i=1}^{N_{bond}} \frac{k_i}{2} (r_i - r_{i,0})^2 && \text{bonds} \\ & + \sum_{i=1}^{N_{angles}} \frac{k_i}{2} (\cos \theta_i - \cos \theta_{i,0})^2 && \text{angles} \\ & + \sum_{i=1}^{N_{dihedral}} k_i (1 + \cos(n\varphi_i - v_{i,0})) && \text{dihedrals} \\ & + \sum_{i=1}^{N_{atom}} \sum_{j=i+1}^{N_{atom}} \left(4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right) + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right) && \text{non-bonded} \end{aligned}$$

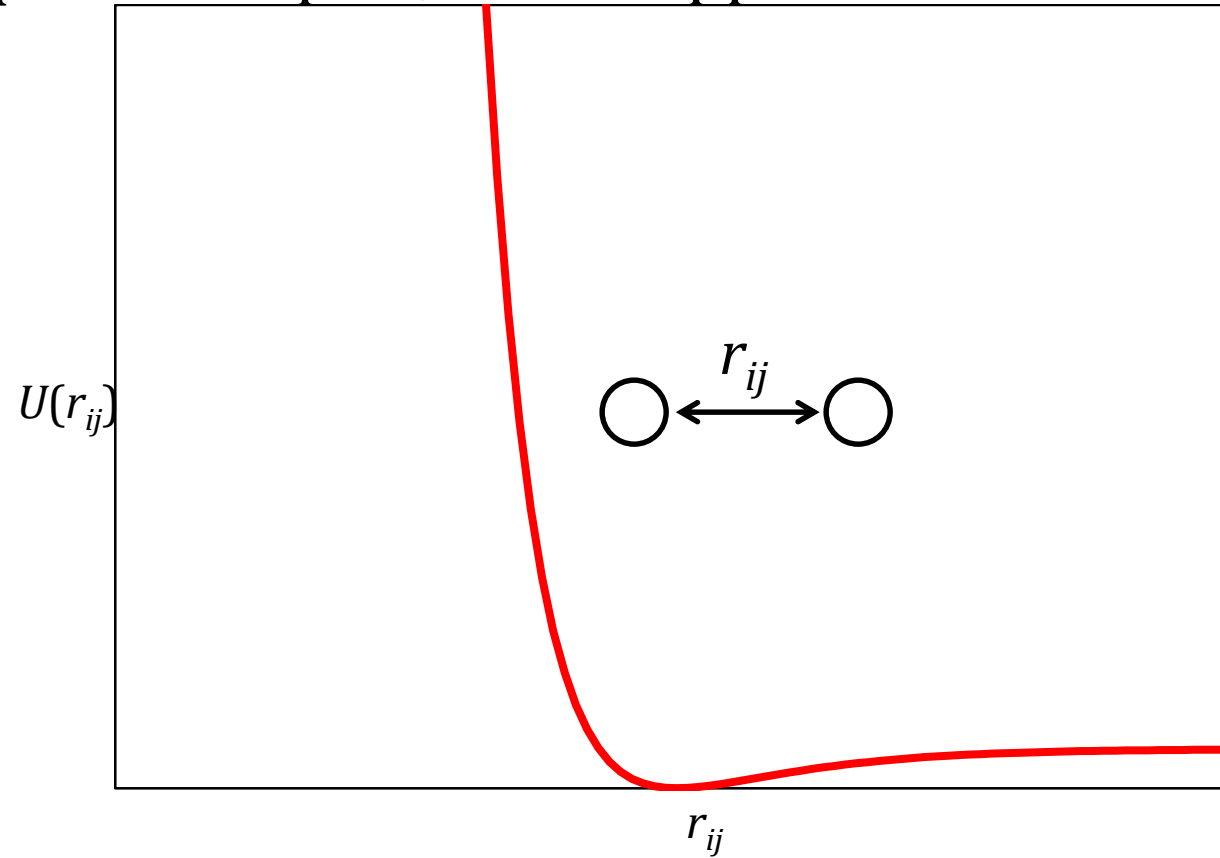
Not a bad approximation

x_i is a value in the coordinates, $x_{i,0}$ is a literature reference value

Bonds

What are bonds really ?

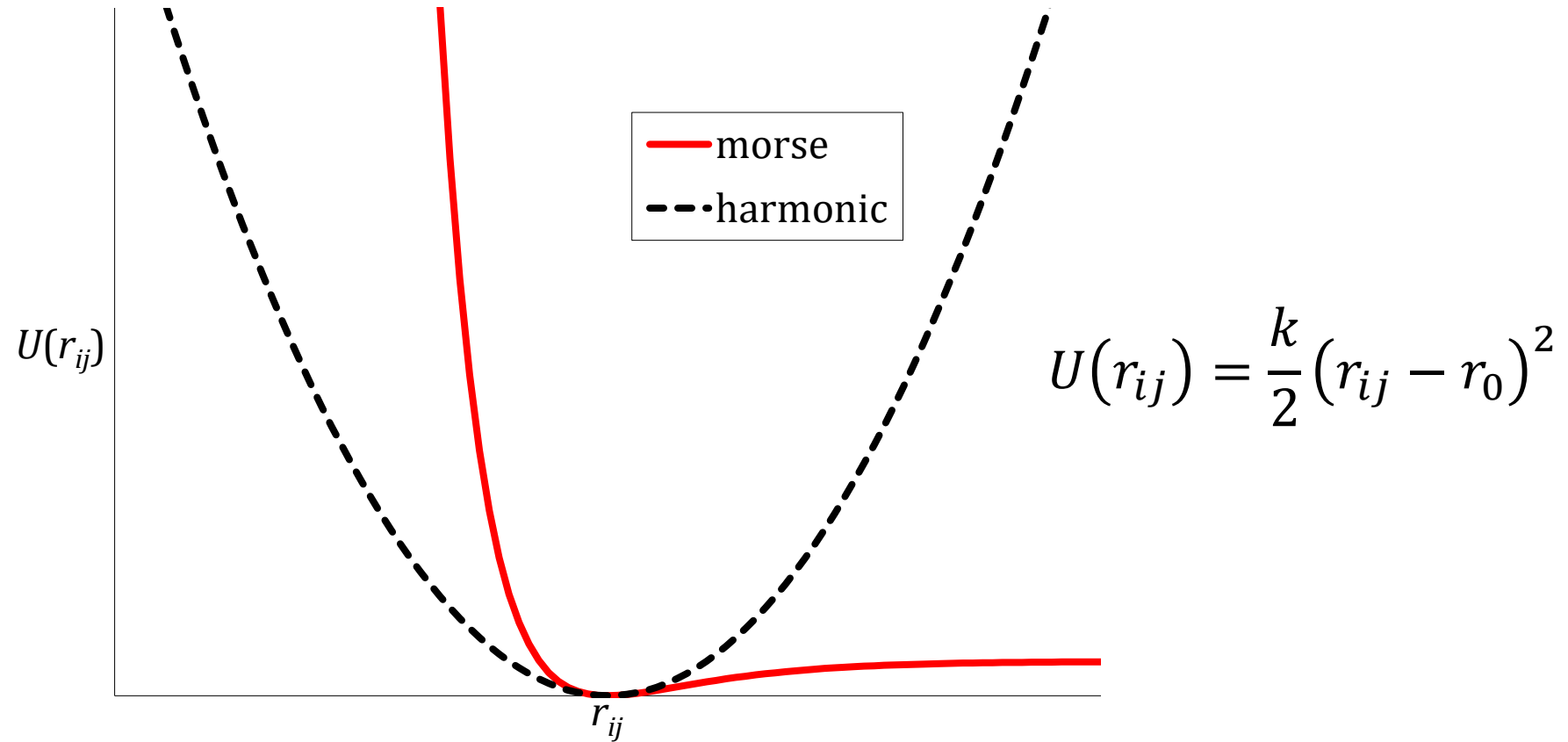
- as we pull two particles apart, what happens ?



- can we approximate ?

Bonds - approximate

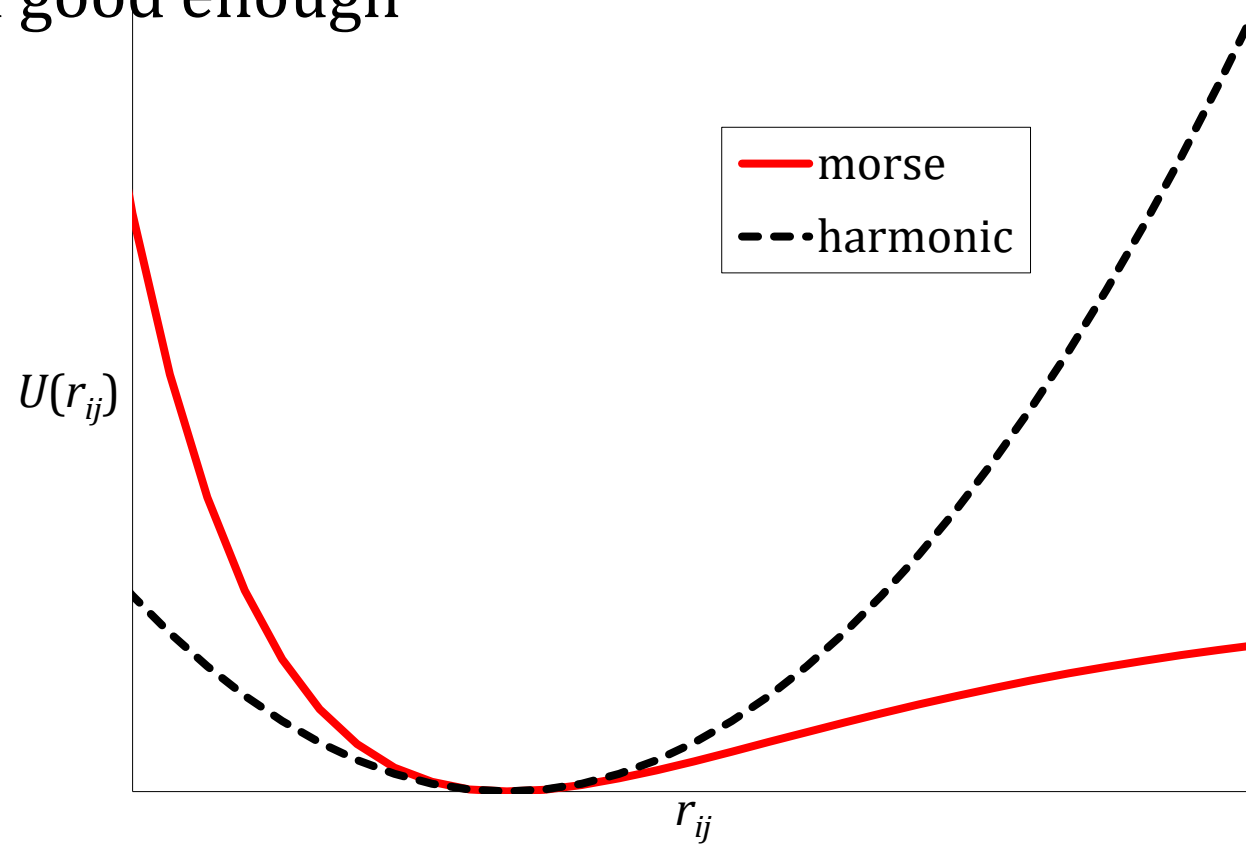
Is the black line a good approximation ?



Bonds – good enough

Bonds do not stretch much

- harmonic is often good enough



Why do we write $\frac{k}{2} (r_{ij} - r_0)^2$? (not just k)

Bonds – the end

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

as scalar.. $F = \frac{-dU}{dr} = -k(r - r_0)$

- so first form looks like Hookes law

Do bonds matter ?

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

Angles

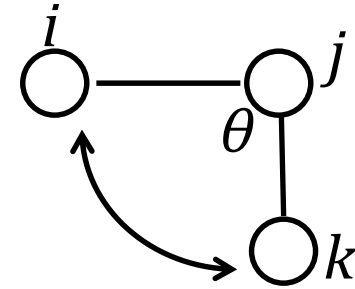
Angles are not as boring as bonds

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

or

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

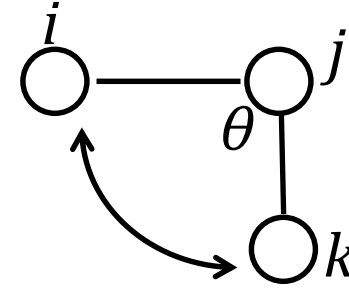
but to get forces is messy (use chain rule) ... why ?



Why are forces difficult ?

We use physical models like

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



Force is obvious

- looking at picture
- in terms of $\cos \theta$
- in terms of θ ?
- we want x, y, z

$$\vec{F}_{angle}(\vec{r}_i) = \frac{-\partial U(\vec{r}_i)}{\partial(\vec{r}_i)}$$

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

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

Angle forces

$$= \frac{-\partial U_{angle}(\vec{r}_i)}{\partial \vec{r}_i}$$

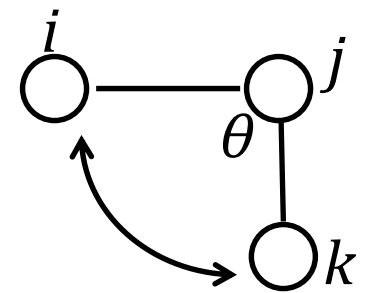
$$F_{angle_i} = \frac{-\partial U_{angle}(\vec{r}_i)}{\partial \cos \theta_{ijk}} \frac{\partial \cos \theta_{ijk}}{\partial \vec{r}_i}$$

$$= k(\cos \theta_{ijk} - \cos \theta_0) \left(\frac{\vec{r}_{kj}}{r_{jk}} - \frac{\vec{r}_{ij}}{r_{ij}} \cos \theta_{ijk} \right) \frac{1}{r_{ij}}$$

The other atoms ?

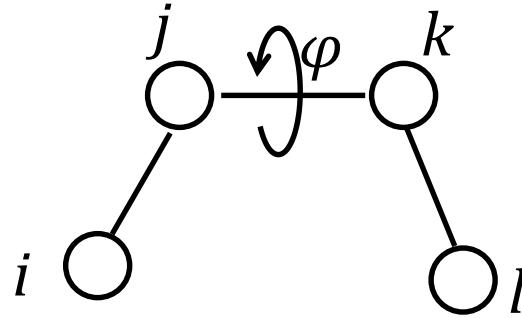
similar expression for F_k

$$F_j = -(F_i + F_k)$$

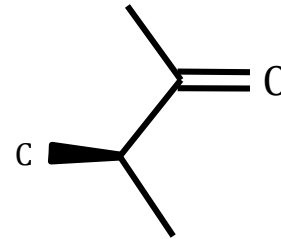
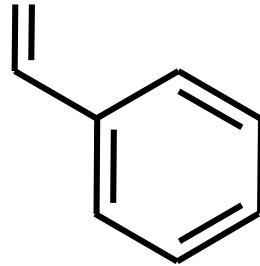


dihedral / torsion angles

basically...



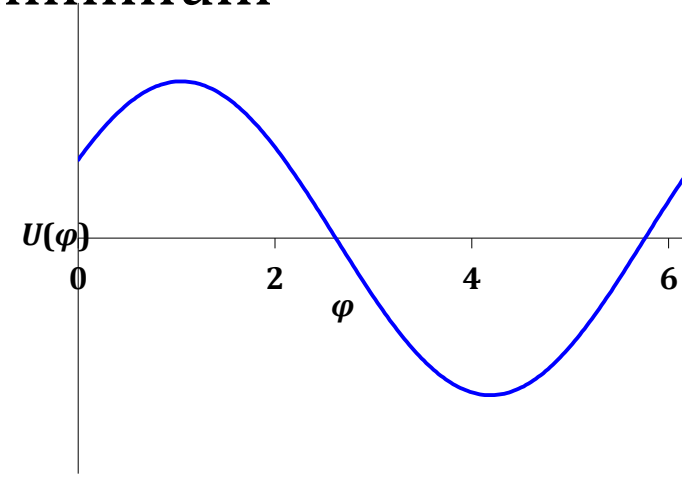
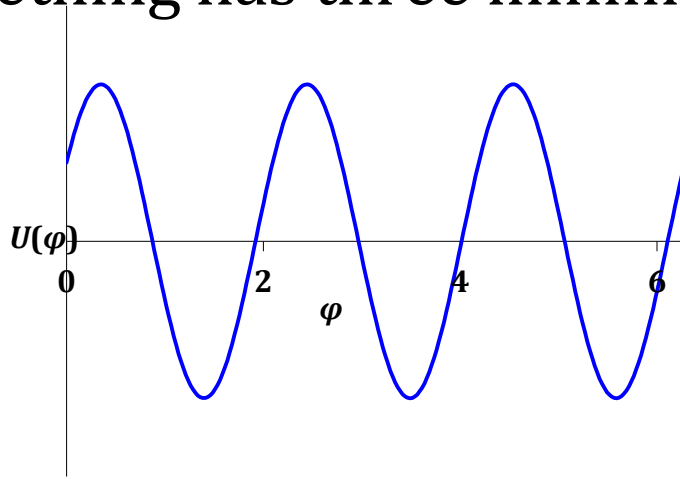
but details will vary



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

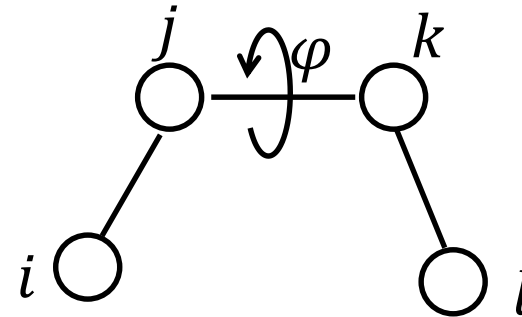
Form of dihedral term

Maybe something has three minima or one minimum



What are the causes ?

- do i and l interact ?
- electron clouds from j and k ?
- model $k_i(1 + \cos(n\varphi_i - \nu_0))$



model for dihedral angles

$$U(\vec{r}) = k_i(1 + \cos(n\varphi_i - \nu_0))$$

n controls multiplicity

- $n = 3$ butane
- $n = 2$ peptide bond

At room temperature, barriers are similar to kT so

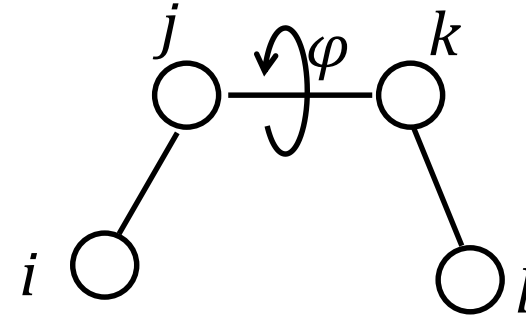
- atoms do rotate, but there are preferences (from spectroscopy)

How good is the model? Not perfect

- 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

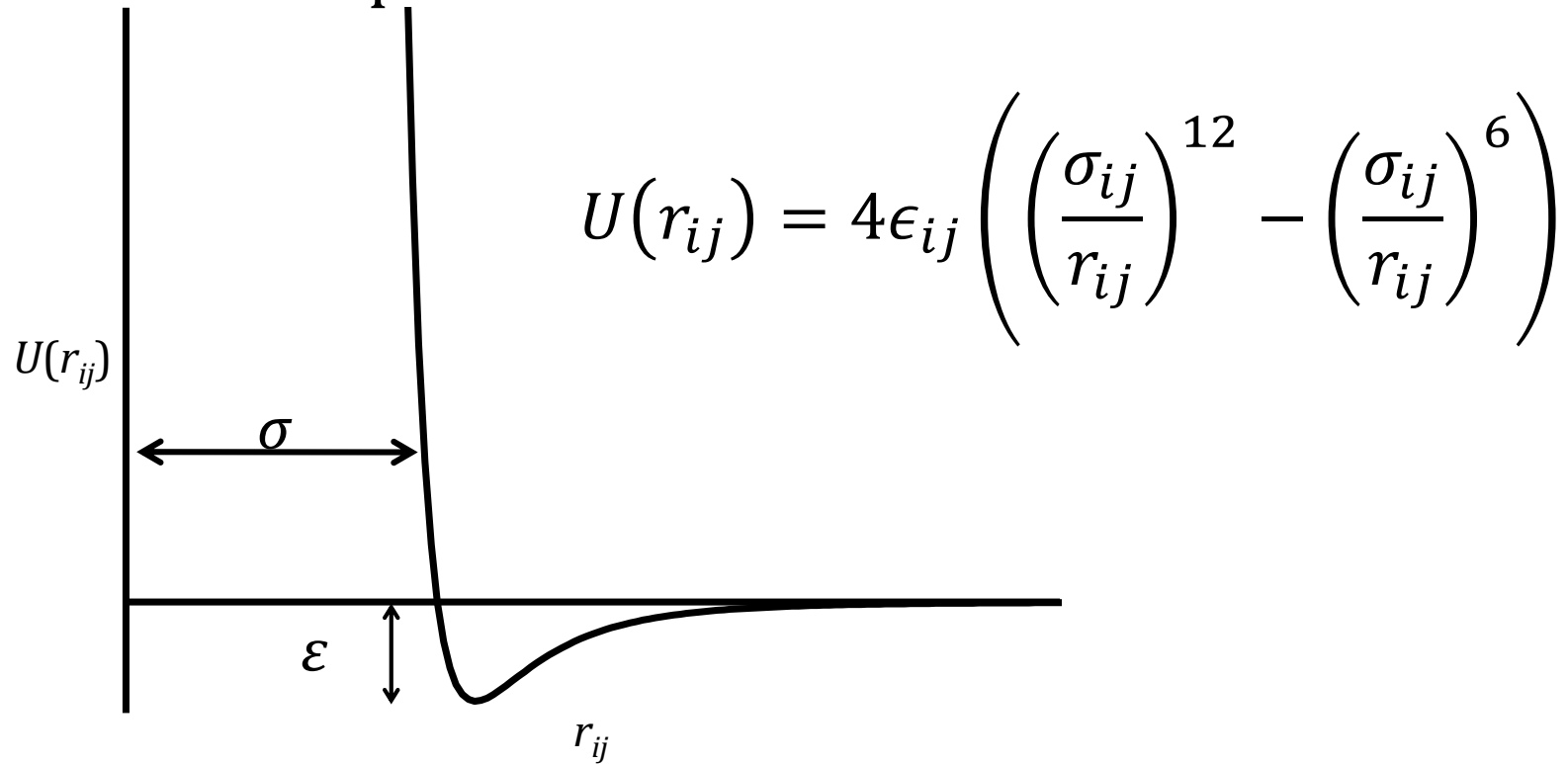
Why are they separated from others ?

- 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



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

van der Waals – how good is it ?

Repulsive and attractive look similar

- quite different

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

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 just 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

Electrostatics

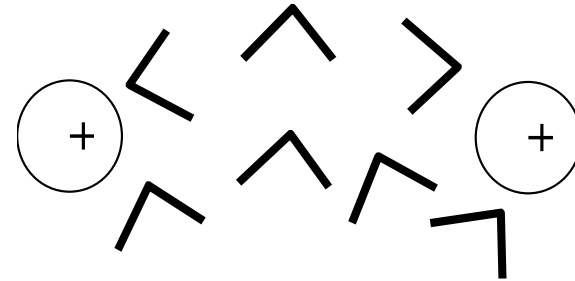
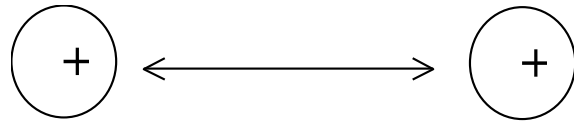
Coulombs law

Sounds easy $U(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$ or $U(r_{ij}) = \frac{q_i q_j}{D r_{ij}}$

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
- cpu intensive

Conflicting goals

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

Cost of model

Cost of model ?

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

Many simplifications

- some are popular and important

Cheaper models - United atoms / heavy atoms

When does a proton matter ?

- charge interactions, H bonds

When is a proton not interesting ?

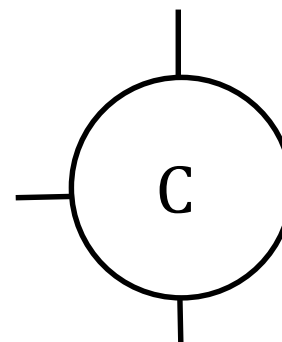
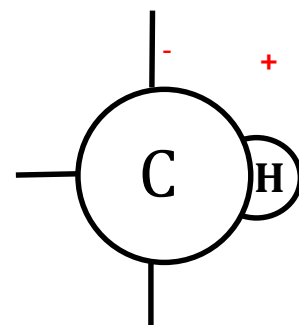
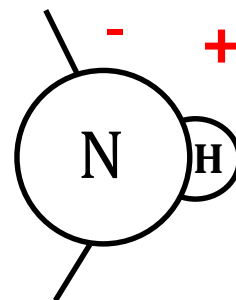
- most aliphatic and aromatic
- hardly a charge / tiny radius

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, CH₂, CH₃



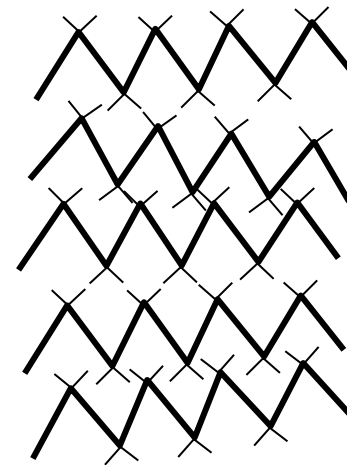
United atoms how bad ?

Gain of united atoms ?

- roughly halves number of atoms
- Costs ?
 - dynamics ? no problem
 - structure ? not too bad

Problems ?

- some small effects can be seen in certain systems
 - lipids
- maybe some effects in proteins



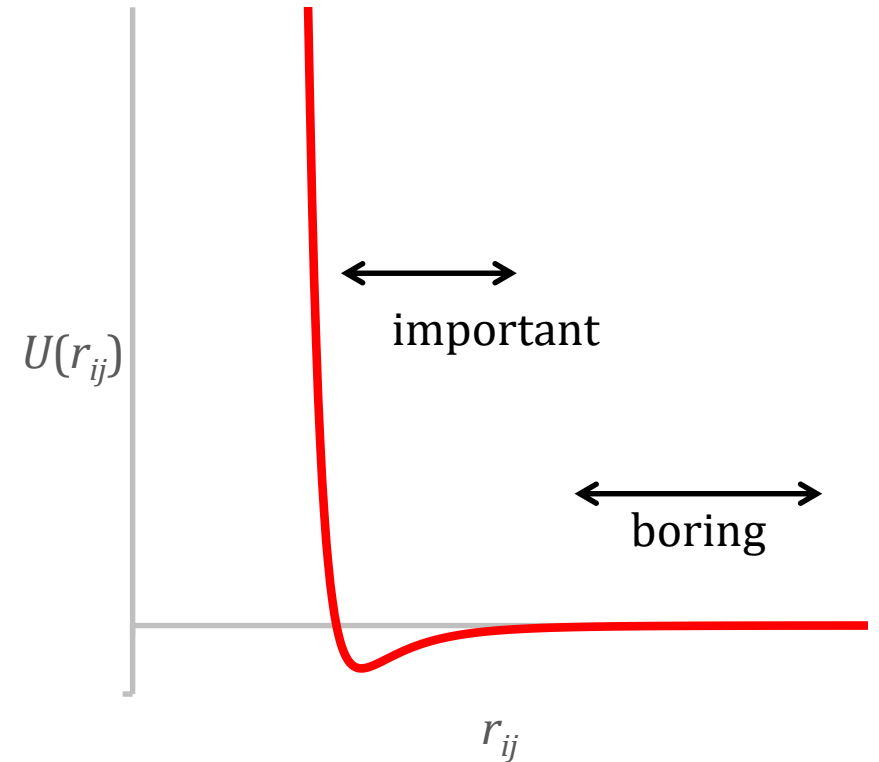
Cutoffs (cheaper models)

How important is an interaction ?

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

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

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



- more difficult.. for other terms

cutoffs - problems

Electrostatics

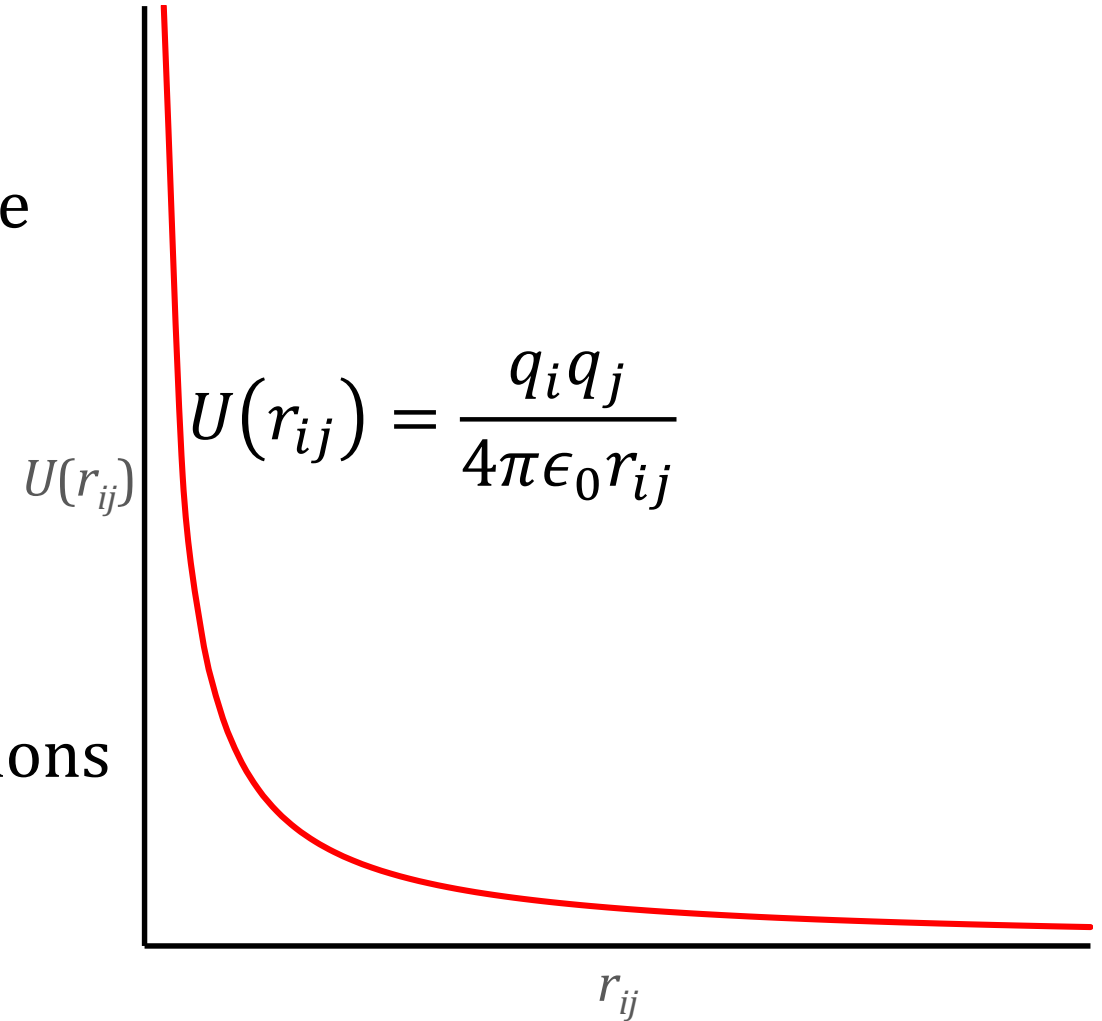
- r^{-1} shrinks slowly
- but eventually, we should be able to ignore
- bigger cutoff (10 – 15 Å)

In practice

- use one cutoff for all non-bonded calculations

Problems

- subtle – look at derivative



Possible improvements

Better bonds ?

- more sophisticated than harmonic $(r - r_0)^2$
- can do - not very interesting

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

- can do - not worth worrying about

Electrostatics

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

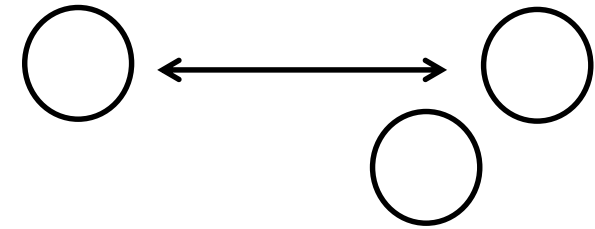
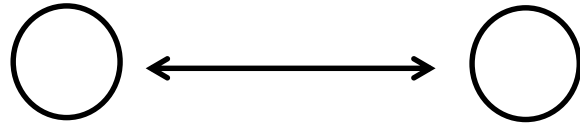
Water

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

Problems and fixes in model

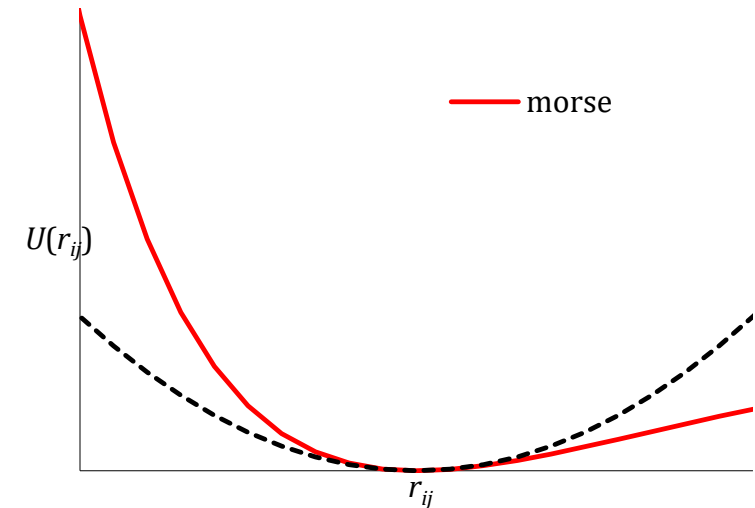
Model is fundamentally wrong

- 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

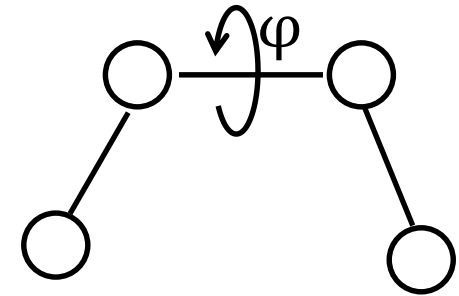
Simple example

- what controls density of a polar fluid
 - charges ?
 - Lennard-Jones terms ?
 - both

Complicated example

- rate of rotation..
 - torsional term (k in torsion term)
 - size of atoms and barriers (ϵ, σ) (less important)
 - angles ? (less important)

Errors in one part of force field compensated elsewhere



Testing force fields

What should a force field do ?

- structure
 - simulate a protein
 - it should not blow up (necessary / not sufficient)
 - density
 - anything you can measure and calculate (distances, chemical shifts..)
- reproduce energy changes (free energies)
- global minimum (free) energy should agree with experiment
- dynamics properties
 - which torsion angles rotate at room temperature ?
 - order parameters from NMR

Transferability

Perfect model of physics would work in all cases

- 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, ... for each type of bond / angle /dihedral
- Sources
 - literature
 - mass
 - partial charges ?
 - high level calculations
 - measurements on small molecules (crystallography)
 - bond lengths, geometry
 - trial and error (example)
 - simulate a liquid
 - 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

Model for potential energy

- proteins, organic molecules ..

Main model

- ignore water
- 3 kinds of bonded interactions
- 2 non-bonded

Conservative force field

- $U(r_{ij})$ energy depends on coordinates only
 - no time component

Energy continuous

- useful
- derivative always defined
 - very useful

more summary

- main model
- speed-ups – cutoffs, united atoms
- where do parameters come from ?
- what should parameters do ?
- why may you not see errors in parameters ?

Where next ?

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

- Simulation
- first some rules from statistical mechanics