Molecular Dynamics

Molecular Dynamics (MD) probes the relationship between molecular structure, motion, and function

See video of dasatinib binding to Src kinase

JACS Shan et al., 133(24):9181-9183 (2011)

http://pubs.acs.org/doi/suppl/10.1021/ja202726y

Specifically **this linked video** (AVI)

A classical force field approximates QM, allows treatment of larger systems and timescales

- Ignores motions of electrons, assumes behavior of system can be described by motions of nuclei

- Uses "force field", potential energy function describing energy as function of nuclear position

- Works well when:

  - Born-Oppenheimer approximation is valid

  - Electronic structure is not of interest

  - Temperature is moderate, not too low

  - No bond breaking/forming

  - Electrons are highly localized (no metals, pi-bonded systems, ...)

A classical system is fully described by the positions and momenta of the atoms (nuclei)

- We technically deal with nuclei, but we will talk about atoms instead

- Each atom has a specific position (**r**) and momentum (**p**); the set of these fully specifies the state of the system

$$\mathbf{r}^N = (x_1, y_1, z_1, x_2, ..., y_N, z_N)$$
$$\mathbf{p}^N = (p_{x,1}, p_{y,1}, p_{z,1}, p_{x,2}, ..., p_{y,N}, p_{z,N})$$

- A microstate specifies these 6N pieces of information

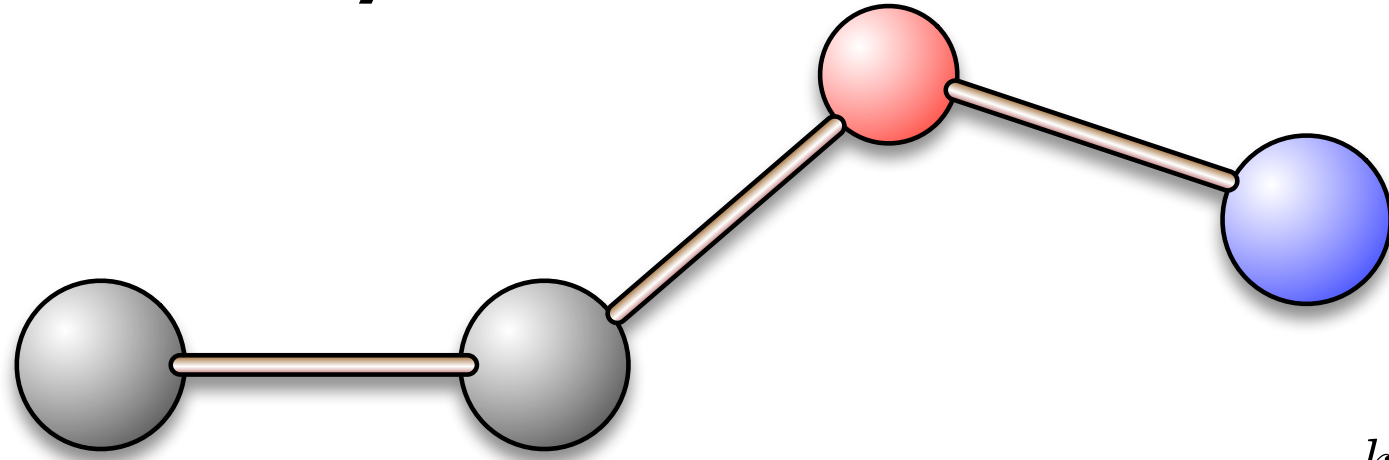# The kinetic energy is a simple function of the momentum; the potential energy depends only on the positions

$$K(\mathbf{p}^N) = \sum_i \frac{|\mathbf{p_i}|^2}{2m_i} \qquad\qquad U(\mathbf{r}^N) = U(\mathbf{r_1}, \mathbf{r_2}, ...)$$

The potential energy is an approximation to the energy that would be obtained for the ground state from QM

The time evolution of the system is described by Newton's equations:
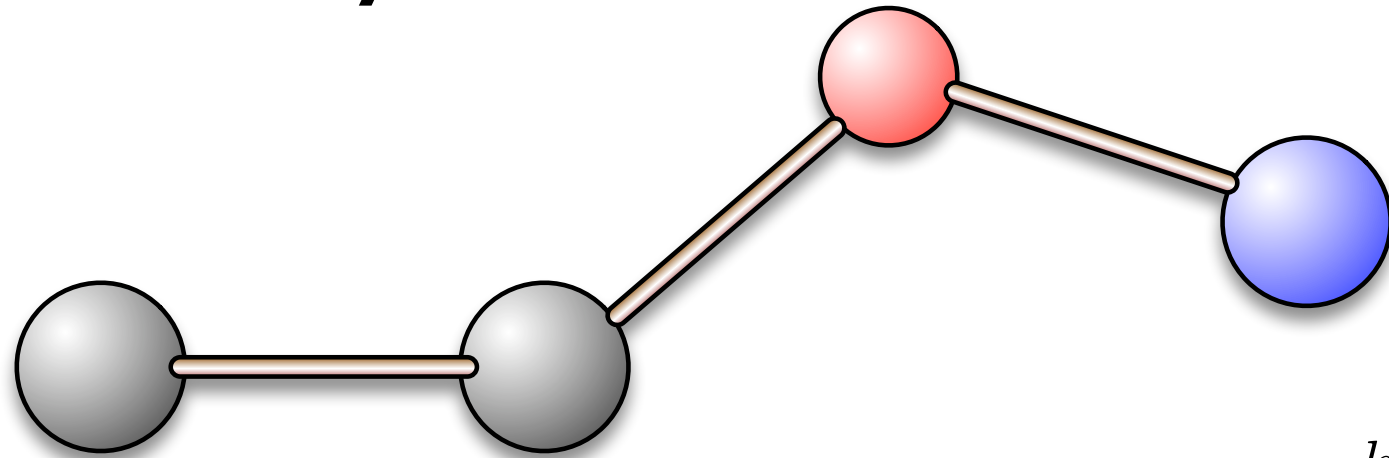
$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = \mathbf{F}_i \qquad \text{or} \qquad \frac{d\mathbf{p}_i}{dt} = -\frac{dU}{d\mathbf{r}_i}(\mathbf{r_1}, \mathbf{r_2}, ...)$$

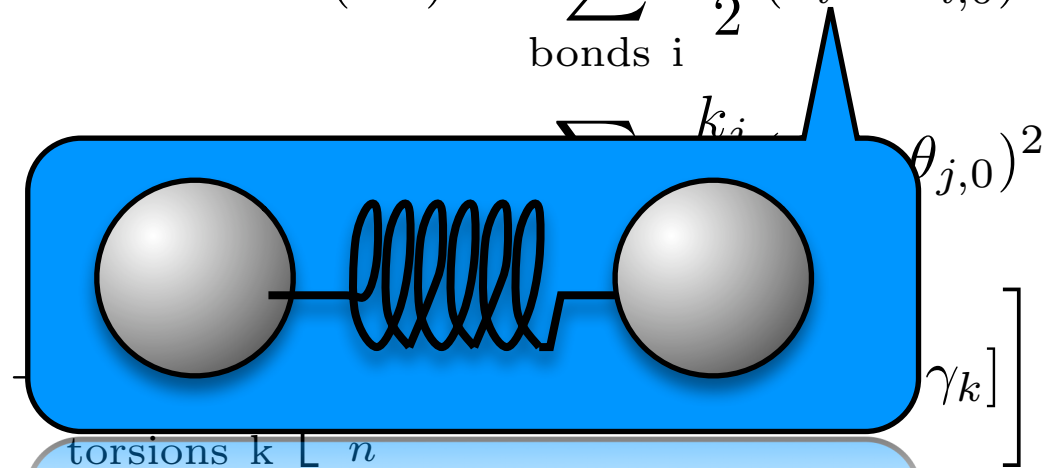# A classical force field uses a combination of terms, many of which are harmonic approximations

$$U(\mathbf{r}^N) = \sum_{\text{bonds i}} \frac{k_i}{2}(d_i - d_{i,0})^2$$

$$+ \sum_{\text{angles j}} \frac{k_j}{2}(\theta_j - \theta_{j,0})^2$$

$$+ \sum_{\text{torsions k}} \left[\sum_n u_{k,n}[1 + cos(\omega_k n + \gamma_k] \right]$$

$$+ \sum_{\text{pairs ij}} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + 4\epsilon_{ij} \left[\left(\frac{r_{ij}}{\sigma_{ij}} \right)^{-12} - \left(\frac{r_{ij}}{\sigma_{ij}} \right)^{-6} \right]$$

A classical force field uses a combination of terms, many of which are harmonic approximations

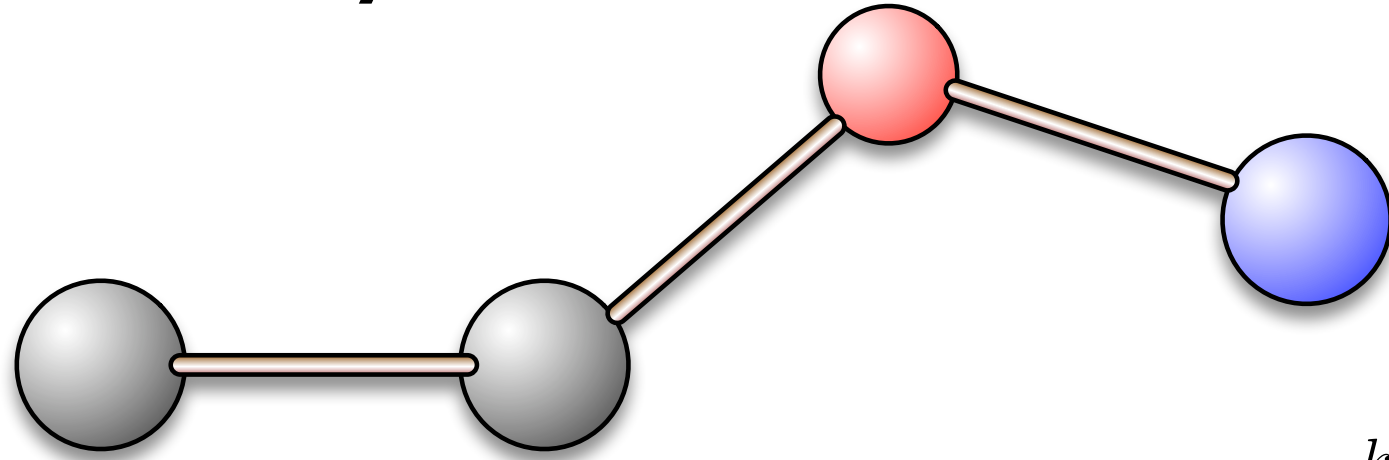$$U(\mathbf{r}^N) = \sum_{\text{bonds } i} \frac{k_i}{2}(d_i - d_{i,0})^2$$

$$\sum \frac{k_j}{2}(\theta_j - \theta_{j,0})^2$$

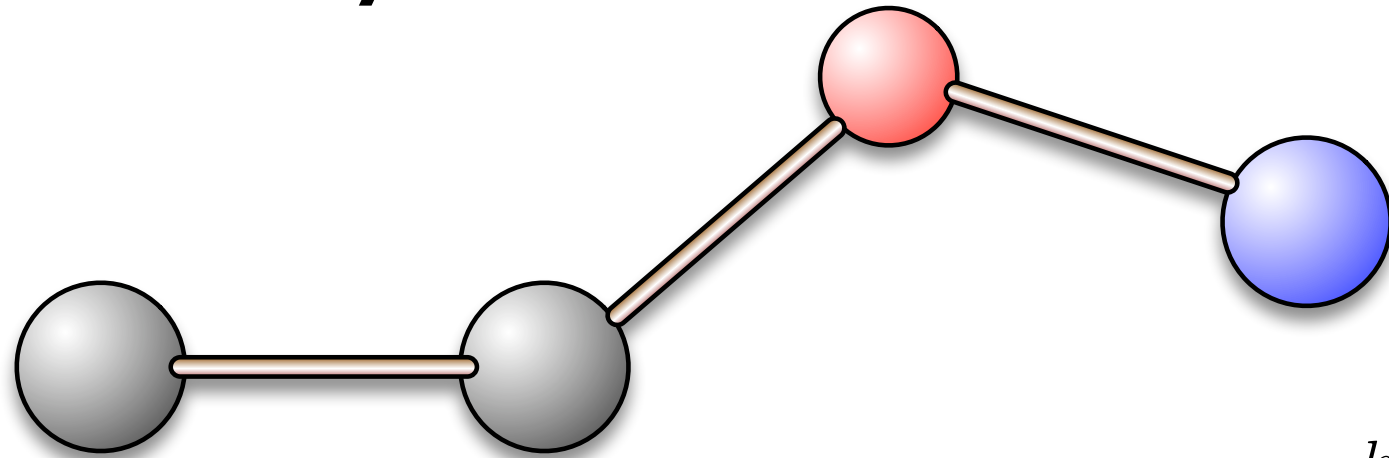$$\sum_{\text{torsions } k}\left[\cdots \frac{1}{n} \cdots \gamma_k\right]$$

$$+ \sum_{\text{pairs } ij} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + 4\epsilon_{ij}\left[\left(\frac{r_{ij}}{\sigma_{ij}}\right)^{-12} - \left(\frac{r_{ij}}{\sigma_{ij}}\right)^{-6}\right]$$

A classical force field uses a combination of terms, many of which are harmonic approximations
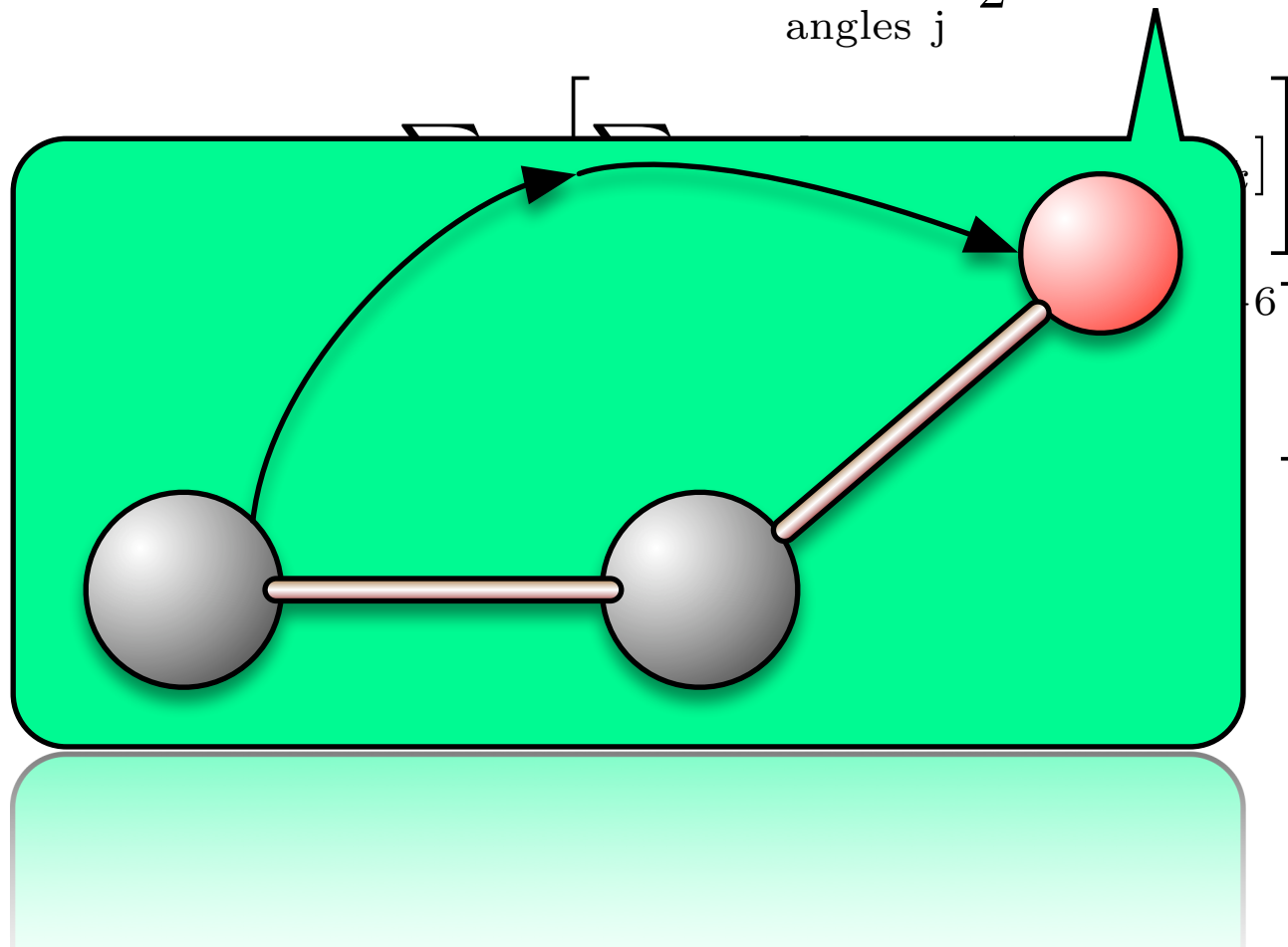

$$U(\mathbf{r}^N) = \sum_{\text{bonds i}} \frac{k_i}{2}(d_i - d_{i,0})^2$$

$$+ \sum_{\text{angles j}} \frac{k_j}{2}(\theta_j - \theta_{j,0})^2$$

$$+ \sum_{\text{torsions k}} \left[\sum_n u_{k,n}[1 + cos(\omega_k n + \gamma_k] \right]$$

$$+ \sum_{\text{pairs ij}} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + 4\epsilon_{ij} \left[\left(\frac{r_{ij}}{\sigma_{ij}} \right)^{-12} - \left(\frac{r_{ij}}{\sigma_{ij}} \right)^{-6} \right]$$

A classical force field uses a combination of terms, many of which are harmonic approximations

$$U(\mathbf{r}^N) = \sum_{\text{bonds i}} \frac{k_i}{2}(d_i - d_{i,0})^2$$

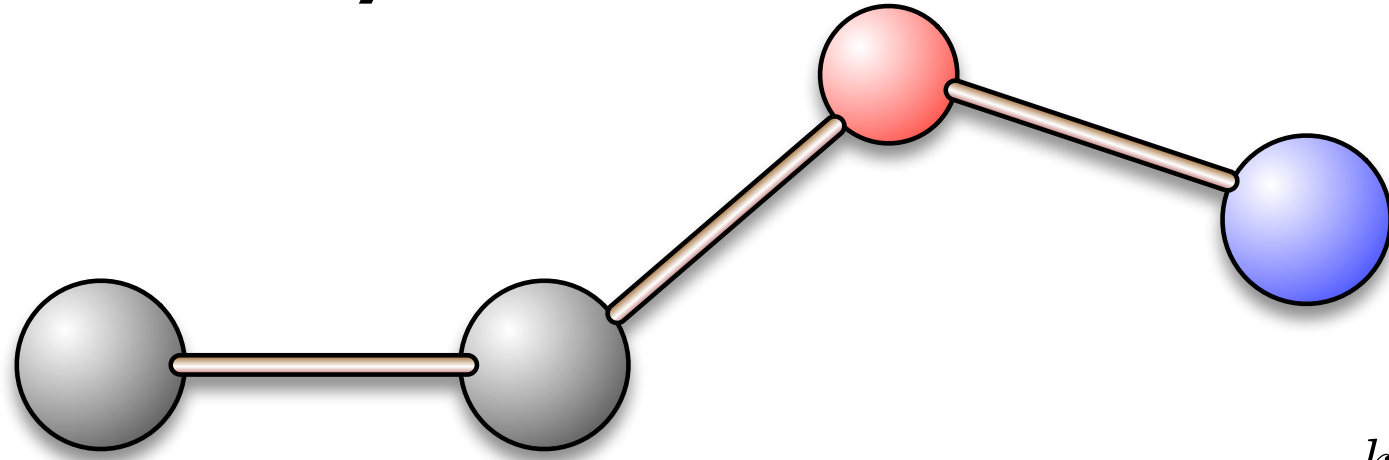$$+ \sum_{\text{angles j}} \frac{k_j}{2}(\theta_j - \theta_{j,0})^2$$

A classical force field uses a combination of terms, many of which are harmonic approximations

$$U(\mathbf{r}^N) = \sum_{\text{bonds i}} \frac{k_i}{2}(d_i - d_{i,0})^2$$

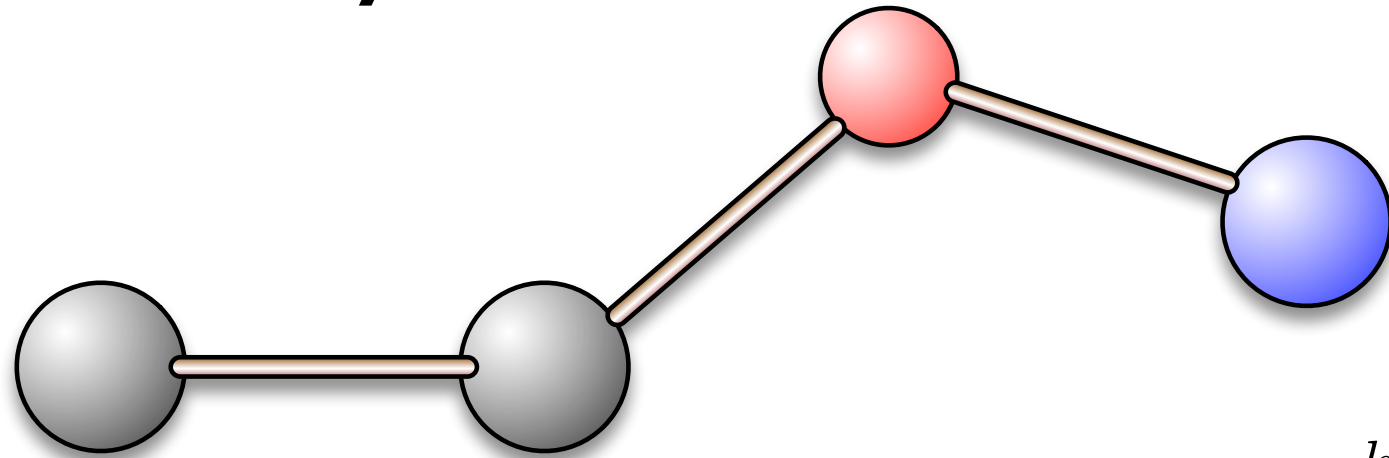$$+ \sum_{\text{angles j}} \frac{k_j}{2}(\theta_j - \theta_{j,0})^2$$

$$+ \sum_{\text{torsions k}} \left[\sum_n u_{k,n}[1 + cos(\omega_k n + \gamma_k] \right]$$

$$+ \sum_{\text{pairs ij}} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + 4\epsilon_{ij}\left[\left(\frac{r_{ij}}{\sigma_{ij}}\right)^{-12} - \left(\frac{r_{ij}}{\sigma_{ij}}\right)^{-6}\right]$$

A classical force field uses a combination of terms, many of which are harmonic approximations

$$U(\mathbf{r}^N) = \sum_{\text{bonds i}} \frac{k_i}{2}(d_i - d_{i,0})^2$$

$$+ \sum_{\text{angles j}} \frac{k_j}{2}(\theta_j - \theta_{j,0})^2$$

$$+ \sum_{\text{torsions k}} \left[\sum_n u_{k,n}[1 + cos(\omega_k n + \gamma_k] \right]$$

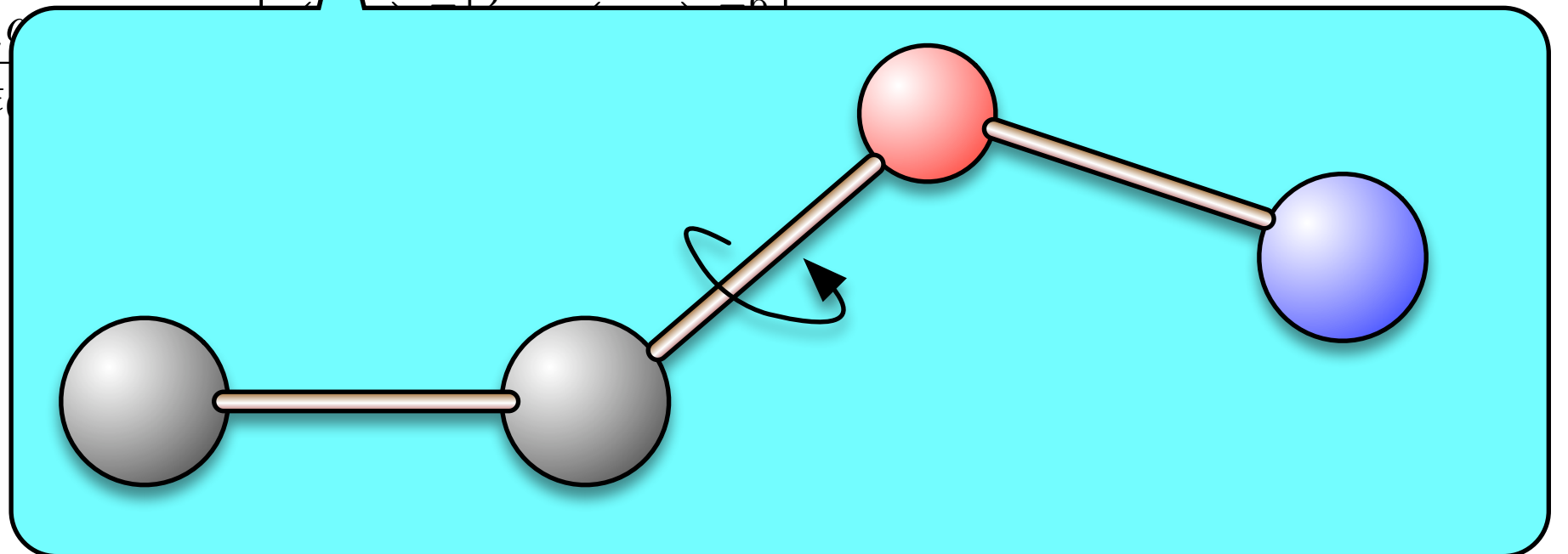$$+ \sum_{\text{pairs ij}} \frac{q_i q}{4\pi\epsilon}$$

A classical force field uses a combination of terms, many of which are harmonic approximations
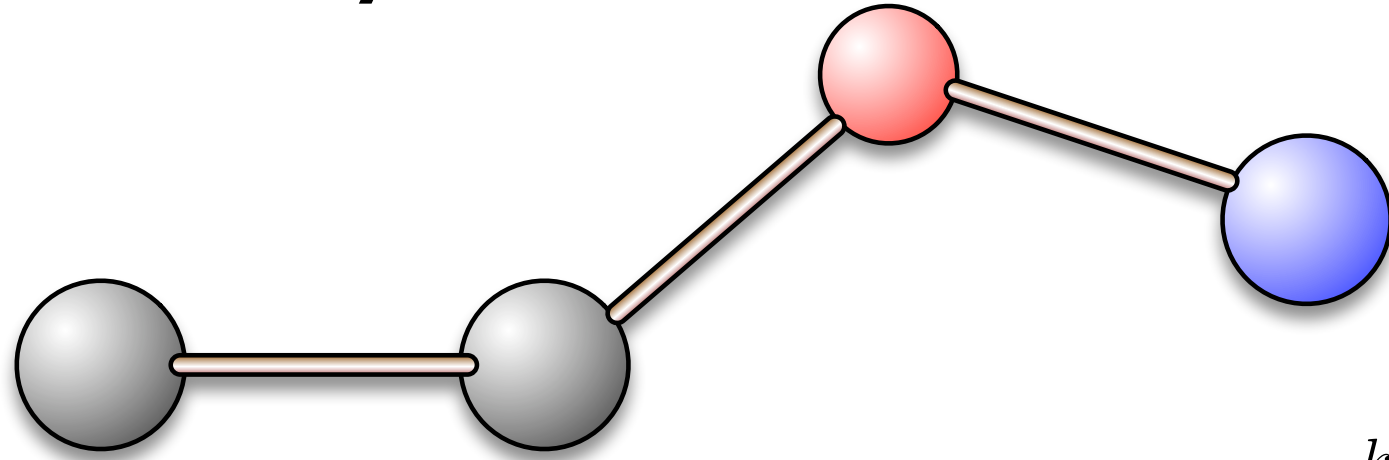

$$U(\mathbf{r}^N) = \sum_{\text{bonds i}} \frac{k_i}{2}(d_i - d_{i,0})^2$$

$$+ \sum_{\text{angles j}} \frac{k_j}{2}(\theta_j - \theta_{j,0})^2$$

$$+ \sum_{\text{torsions k}} \left[\sum_n u_{k,n}[1 + cos(\omega_k n + \gamma_k] \right]$$

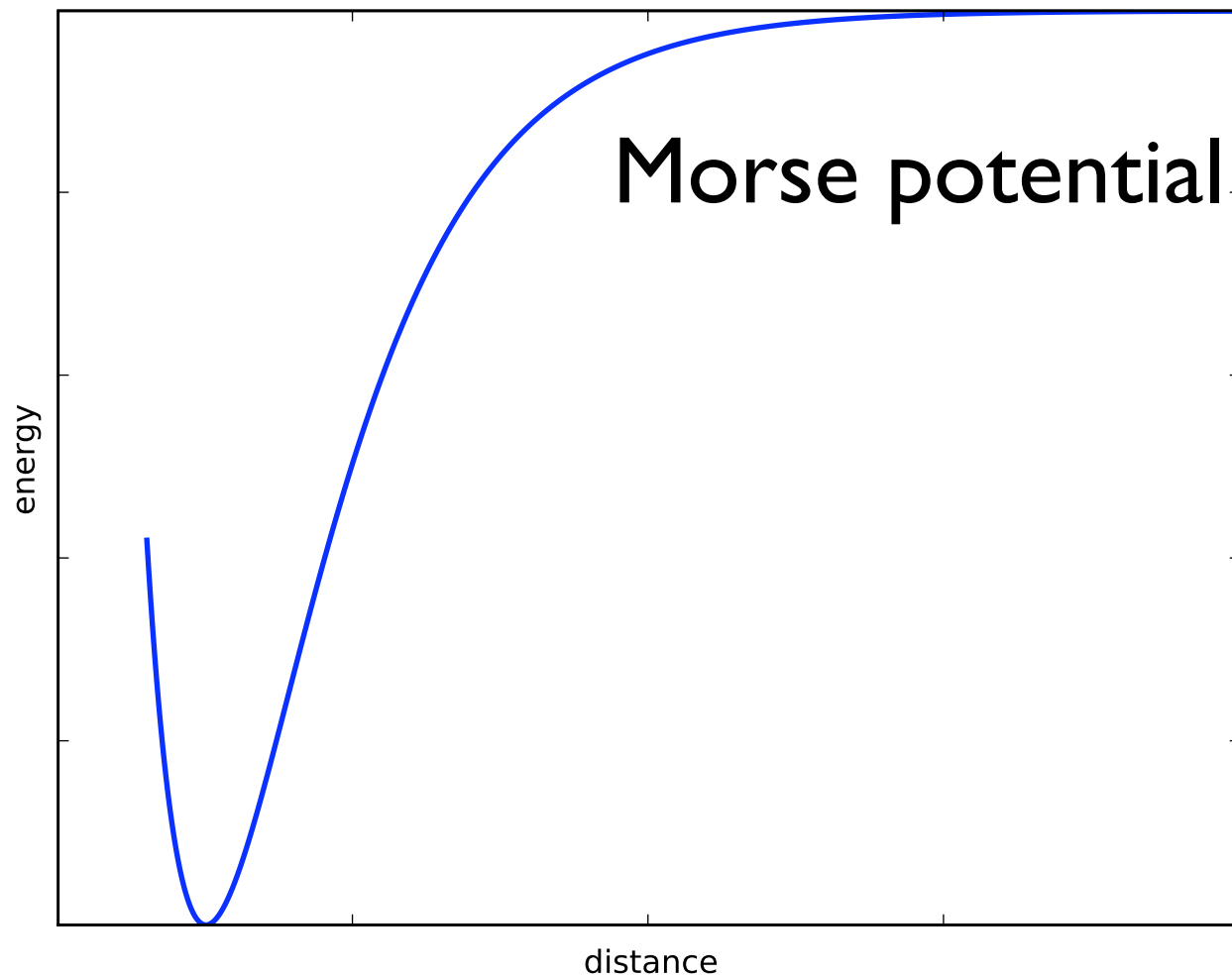$$+ \sum_{\text{pairs ij}} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + 4\epsilon_{ij} \left[\left(\frac{r_{ij}}{\sigma_{ij}} \right)^{-12} - \left(\frac{r_{ij}}{\sigma_{ij}} \right)^{-6} \right]$$

The Morse potential describes bond stretching and dissociation reasonably well; the harmonic approximation is decent

Morse potential

Morse and harmonic

Morse: $u(d) = D_e\big[1 - e^{-a(d-d_0)}\big]^2$

Harmonic: $u(d) = \dfrac{k}{2}(d - d_0)^2$
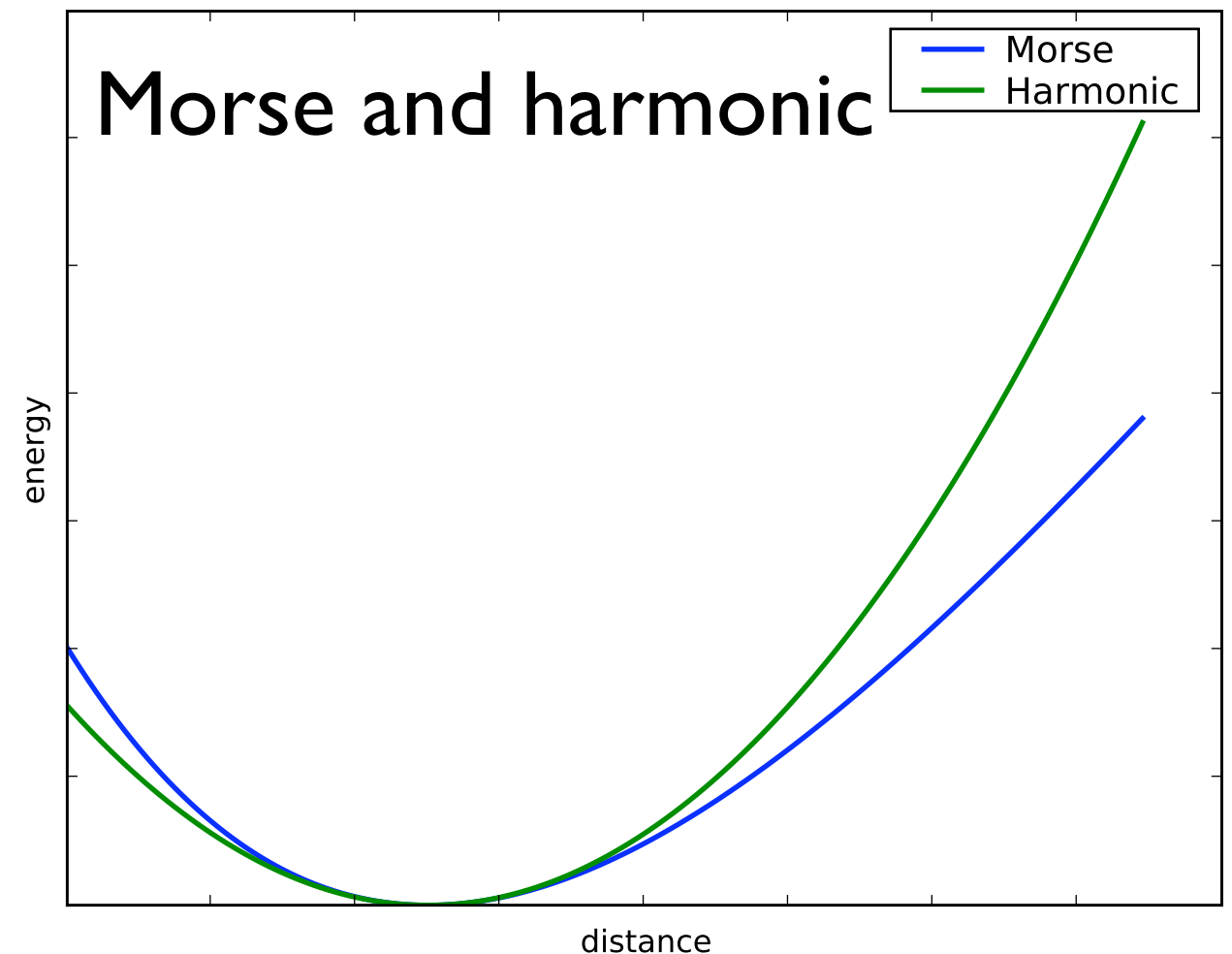
Harmonic terms come up often because of Taylor expansions

The Morse potential describes bond stretching and dissociation reasonably well; the harmonic approximation is decent

Morse potential

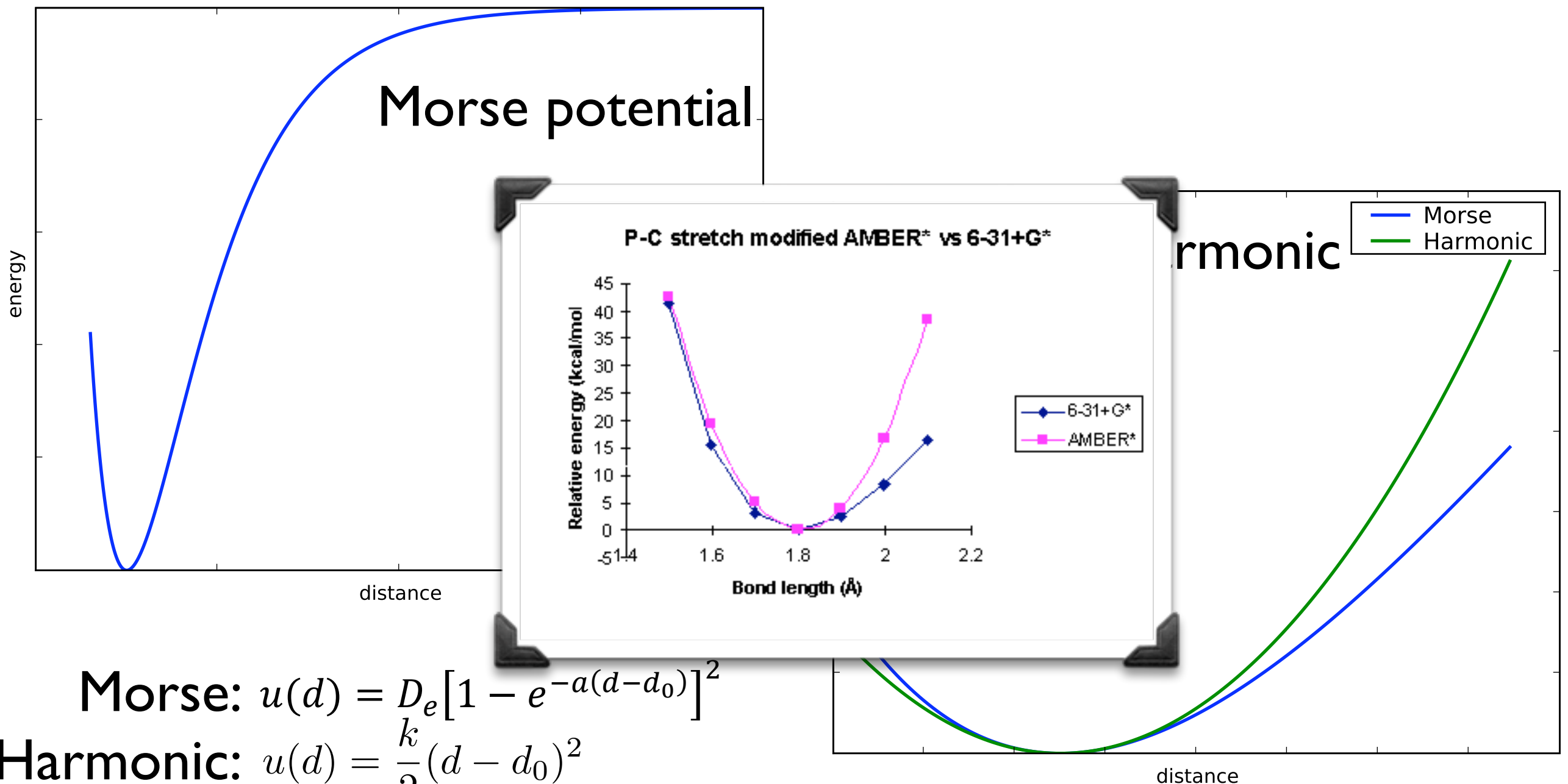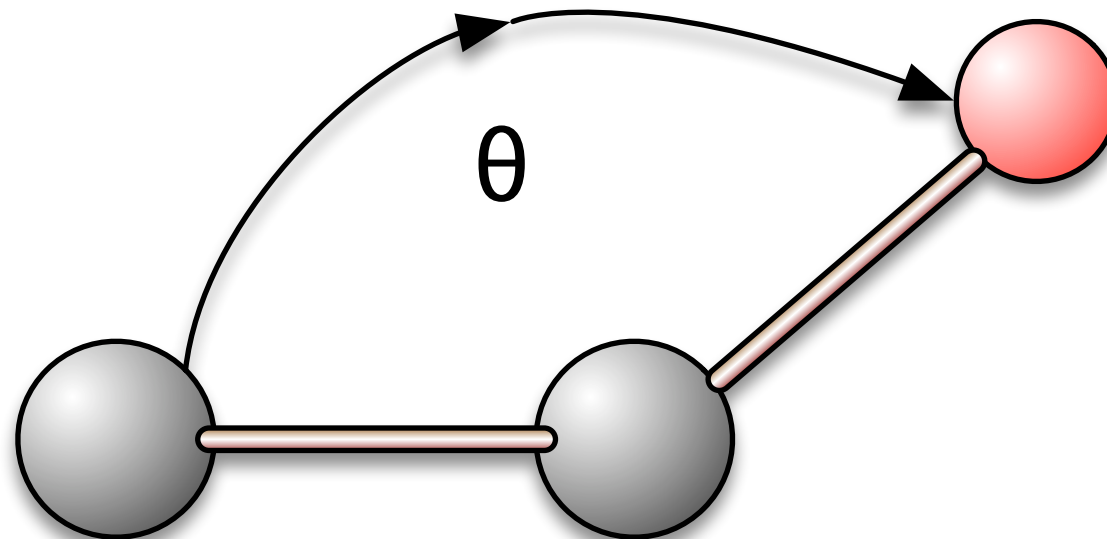P-C stretch modified AMBER* vs 6-31+G*

Morse: $u(d) = D_e\big[1 - e^{-a(d-d_0)}\big]^2$

Harmonic: $u(d) = \dfrac{k}{2}(d - d_0)^2$

Harmonic terms come up often because of Taylor expansions

Angle bending is a simple harmonic term

$$u(\theta) = \frac{k}{2}(\theta - \theta_0)^2$$

Angle bending is less stiff than bond stretching, but still relatively stiff

Torsional potentials go beyond simple harmonic approximations
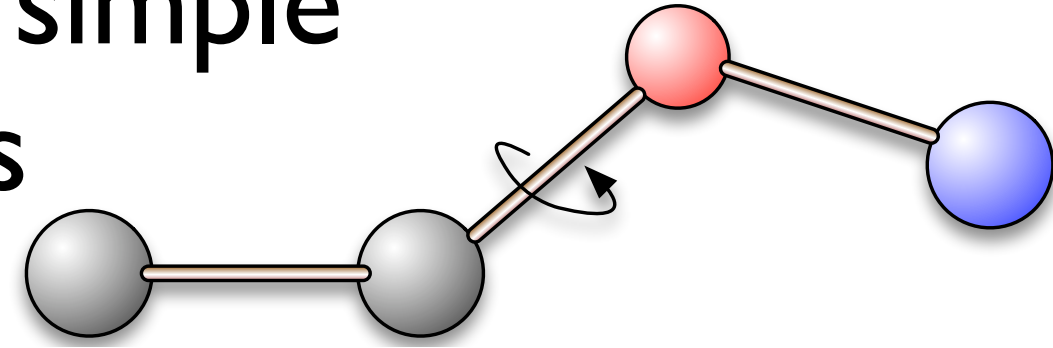
HCPO torsion modified AMBER* vs 6-31+G*

COPO torsion, modified AMBER* vs 6-31+G*
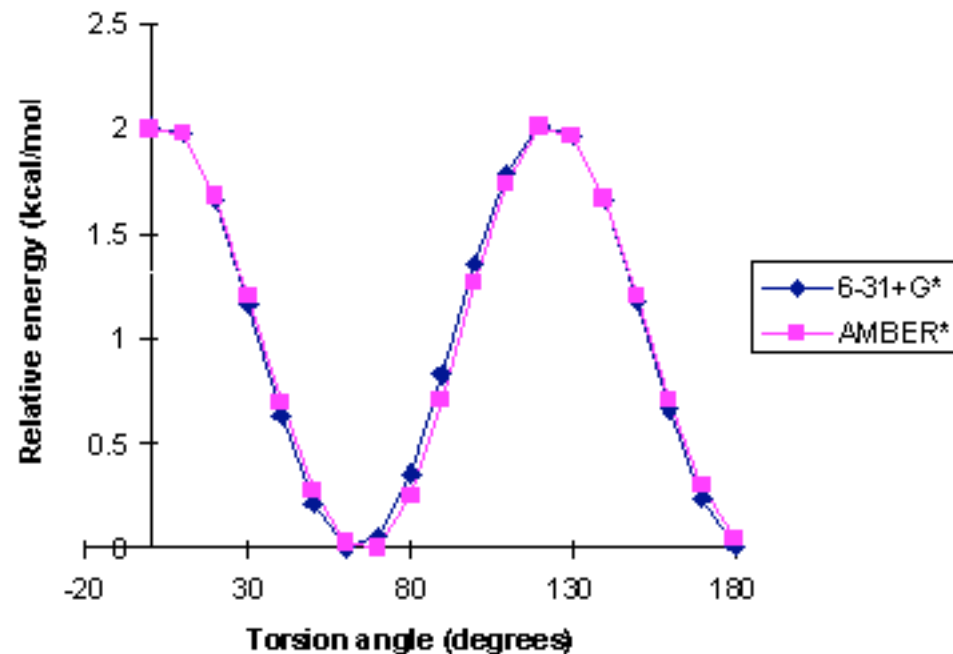
Examples from http://scholar.lib.vt.edu/theses/available/
etd-51797-145925/unrestricted/full_text/ch5.htm
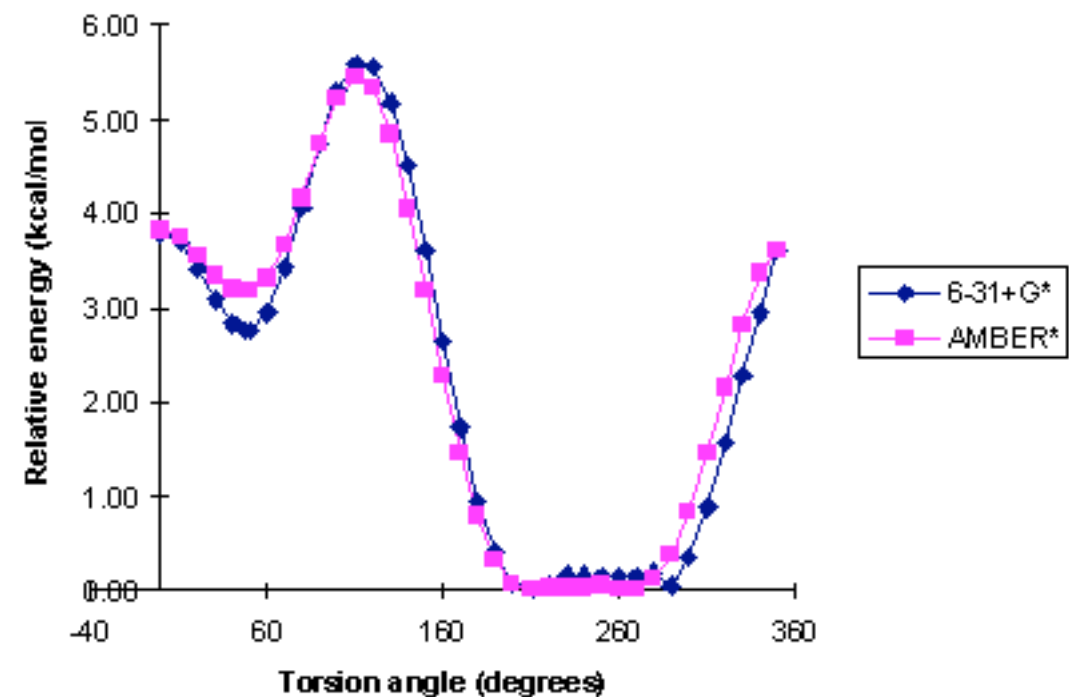
These are represented with a cosine expansion which can approximate them well (see above): $u(\omega) = \sum_n u_n[1 + cos(\omega n + \gamma]$

With enough terms, could fit perfectly, but n is kept relatively small for computational convenience, but large enough for good fit

Improper torsions are used to handle planarity and out-of-plane bending

http://www.pumma.nl/index.php/Theory/Potentials
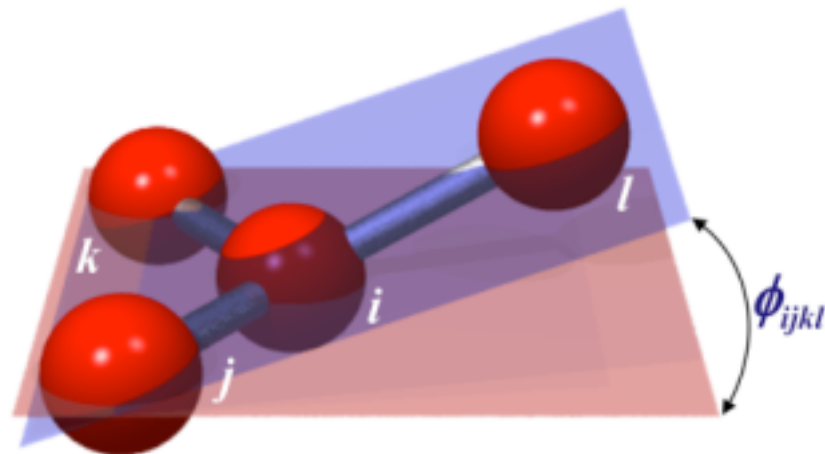
$$u(\omega) = k(1 - cos\ 2(\omega - \omega_0))$$
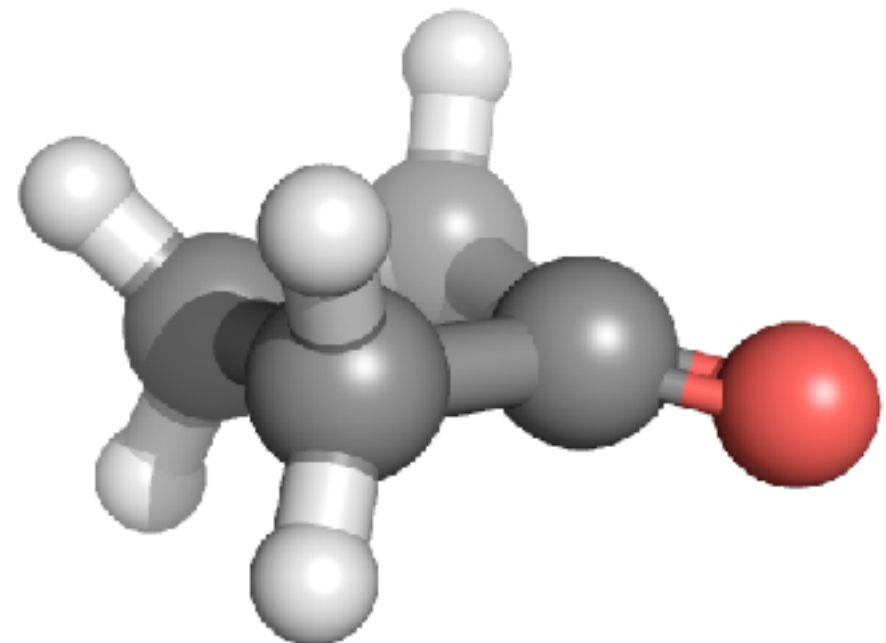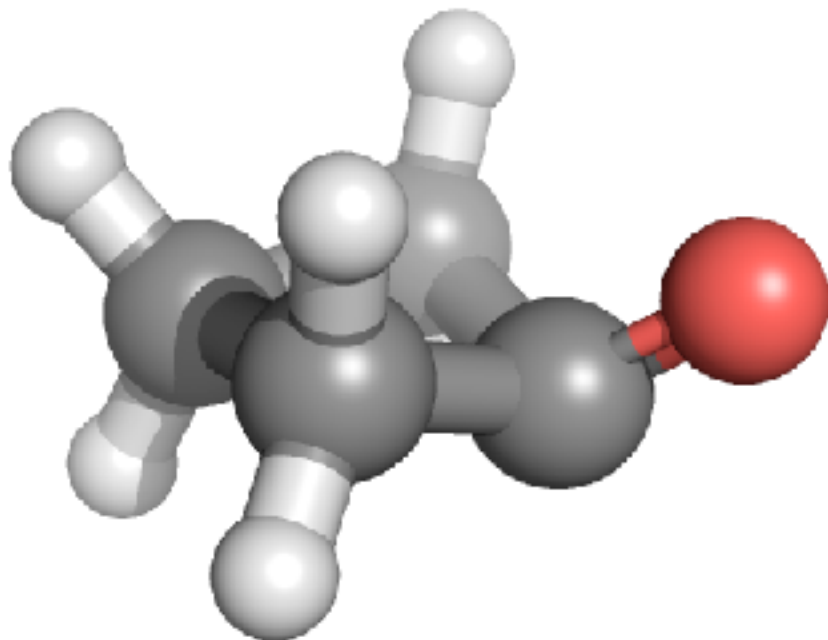
Nonbonded interactions: Between any atoms that are not bonded

- Distant internal atoms

- Atoms in other molecules (intermolecular)

- Three major types:

  - Electrostatic

  - van der Waals attraction (London dispersion)

  - Excluded volume repulsion

Electrostatic interactions are treated as interactions between fixed point charges at atomic centers

$$u(r) = \frac{q_1 q_2}{4\pi\epsilon_0 r}$$

Legend:
- ++, -- (blue)
- +- (green)

y-axis: energy
x-axis: atom-atom distance
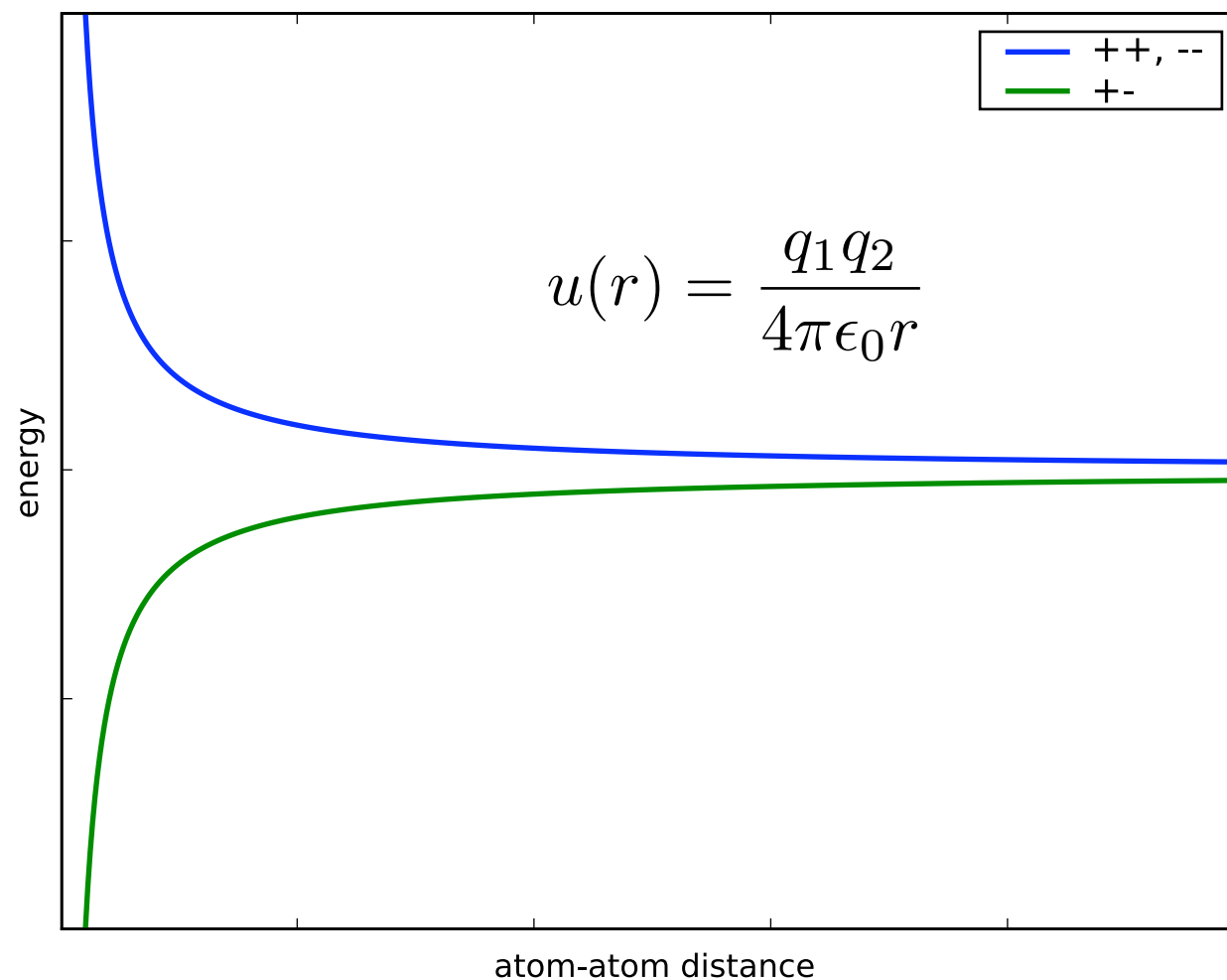
Each atom is usually assigned a *partial charge*, expressed in units of e, which is typically a fraction of e (magnitude 0 up to ~0.8 are typical in neutral molecules)

Partial charges are typically derived from electrostatic potentials from QM
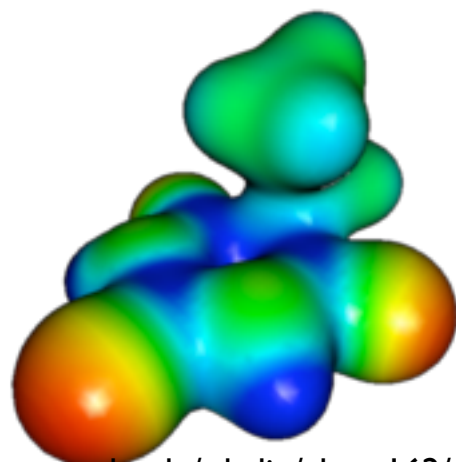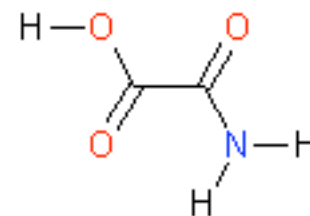
- An electrostatic potential (ESP) is derived from QM

- Charges are fit to reproduce the ESP

  - Sometimes charges are restrained to be 'physical' (RESP)

- Note: partial charges are not a physical observable; a good fit should reproduce:

  - Dipole and quadrupole moments

  - Electrostatic potential at key distances

Dispersion interactions are attractive, due to electronic polarization

- Random displacements of electrons leads to polarization of the electron cloud, resulting in attractive atomic interactions

- We call these van der Waals interactions

- A toy model for which we can solve the Schrödinger equation suggests they decay as $1/r^6$ (Leach 4.10.1)

The Lennard-Jones functional form combines dispersion with a crude estimate of repulsion

- Atoms repel when electron clouds begin to overlap due to Coulomb, Pauli: $u(r) \propto exp(-cr)$ approx.

  But any rapid increase is reasonably good: $u(r) \propto r^{-m}$ (m >= 6). Often, m = 12 for computational ease

- Hence, overall Lennard-Jones functional form:

$$u(r_{ij}) = 4\epsilon \left[\left(\frac{r_{ij}}{\sigma} \right)^{-12} - \left(\frac{r_{ij}}{\sigma} \right)^{-6} \right]$$
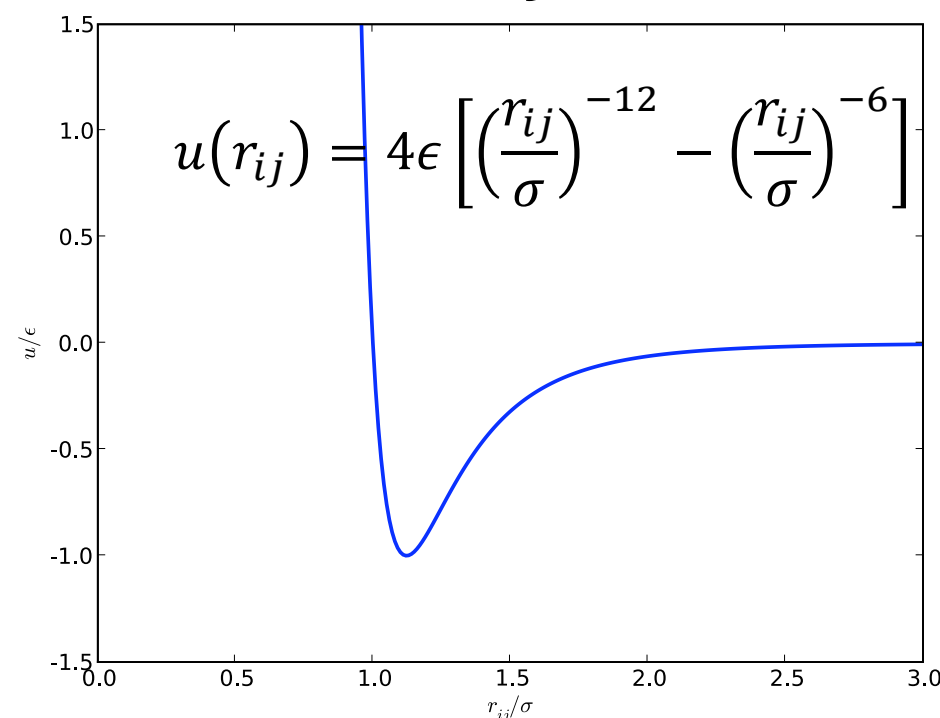
Also written:

$$= Ar_{ij}^{-12} - Cr_{ij}^{-6}$$

where:

$$A = 4\epsilon\sigma^{12} = \epsilon r_m^{12}$$

$$C = 4\epsilon\sigma^6 = 2\epsilon r_m^6$$

Note overall similarity of shape of QM potential for HF dimer we saw in psi4 example

Psi4 example
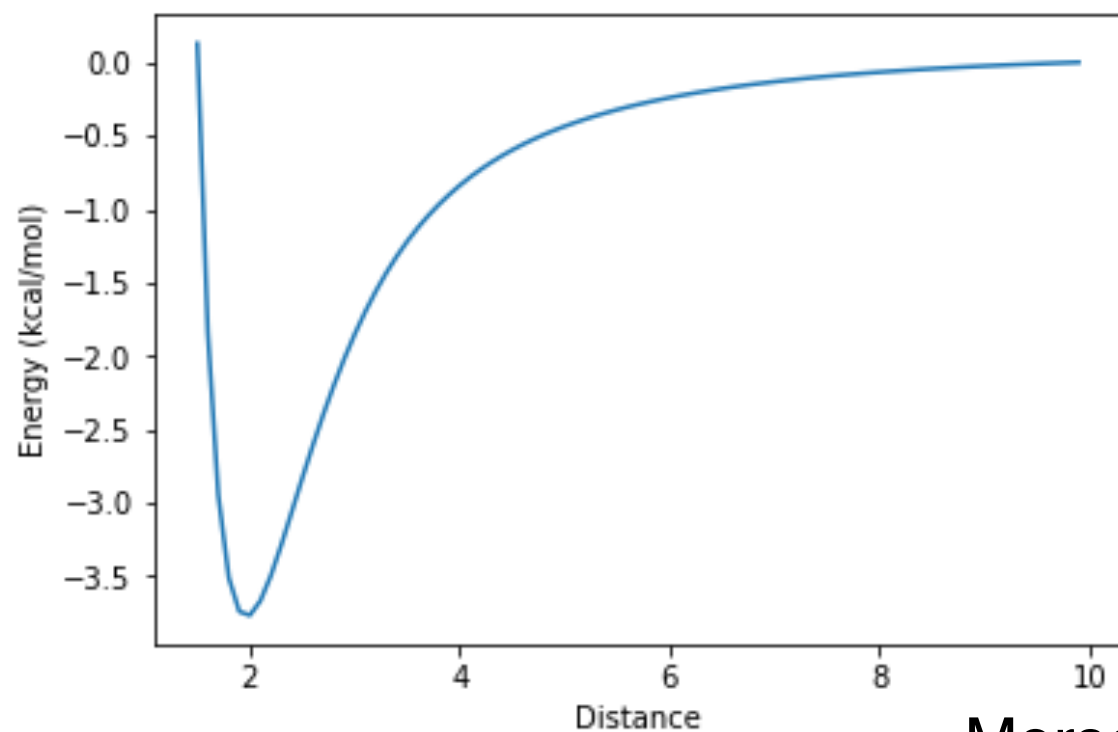
LJ
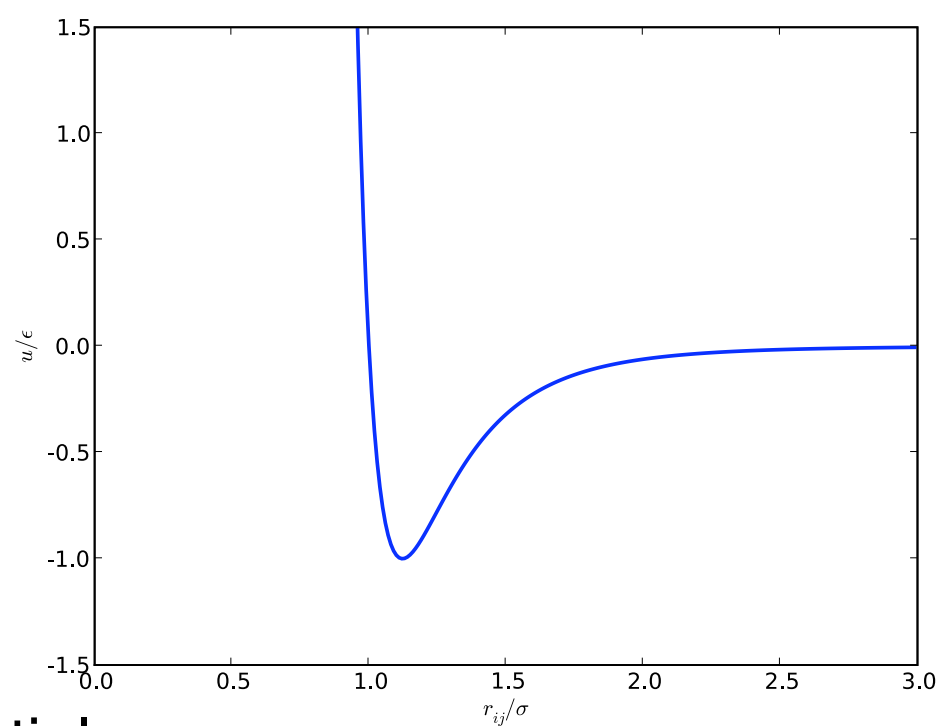
Morse potential
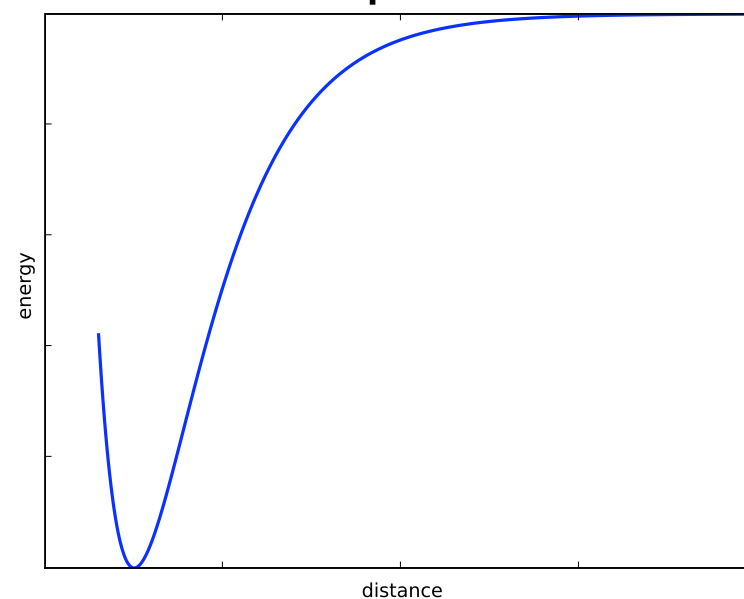
Where do the parameters in a force field come from? QM, experimental data

- Bond stretching, angle bending done first

  - Experimental vibrational and structural data

- LJ parameters from crystal packing, critical point data

  - Sometimes use simulations to ensure match experimental data on bulk properties (density, enthalpy of vaporization, etc.)

- Partial charges computed

  - Ab initio calculations, fit of elec. potential

- Torsions last

  - Affected by all the other energies

  - Match to QM torsional potential

$$U(\mathbf{r}^N) = \sum_{\text{bonds } i} \frac{k_i}{2}(d_i - d_{i,0})^2$$

$$+ \sum_{\text{angles } j} \frac{k_j}{2}(\theta_j - \theta_{j,0})^2$$

$$+ \sum_{\text{torsions } k} \left[\sum_n u_{k,n}[1 + cos(\omega_k n + \gamma_k)] \right]$$

$$+ \sum_{\text{pairs } ij} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + 4\epsilon_{ij} \left[\left(\frac{r_{ij}}{\sigma_{ij}} \right)^{-12} - \left(\frac{r_{ij}}{\sigma_{ij}} \right)^{-6} \right]$$

Force fields are (and should be) designed to be transferable as much as possible

- Goal: You shouldn't have to make a new force field for every new protein

  - Not even for every new small molecule

  - Otherwise, predictions would be extremely difficult and methods would be of limited practical value

- Often informed by chemistry

  - i.e. one set of LJ parameters for SP2-hydbridized carbons, another for SP3, etc. C(sp2) - H bonds have one set of stretching parameters, etc.

  - Torsional potentials and charges are particularly challenging

- Adequate model with few parameters preferred over great model with many

  - Fewer adjustable parameters often (but not always) means greater transferability

There are a variety of common pairwise-additive force fields

- **AMBER (Kollman, UCSF)**

  - Proteins, nucleic acids, carbohydrates

  - GAFF for general small molecules

- **CHARMM (Karplus, Harvard)**

  - Proteins, nucleic acids, lipids, organics

  - Small molecule force field coming

- **OPLS (Jorgensen, Purdue/Yale)**

  - Liquids, proteins, ... general molecules

- **GROMOS (Groningen)**

There are many popular water models which are not necessarily linked to force fields

- **Three point:**

  - TIP3P, SPC

  - SPC/E (optimized for Ewald, which we will discuss later)

- **Four point:**

  - TIP4P, BF (?)

  - TIP4P-Ew (Ewald)

- **Five point**

  - TIP5P,

  - TIP5P-E (Ewald, by Rick)

  - ...

Problem: In a typical simulation of a molecule in water, most time is spent simulating water

- Solvents can be expensive to simulate because they occupy so much of the system

- We could simulate bigger systems if we didn't have to explicitly track all of the solvent atoms and forces

- Is there a way we can avoid this?

Systems exist not just in a single conformation or state, but in a mix of states given by their relative energies

$U(\mathbf{r}^N)$

particle coordinates, $\mathbf{r}^N$
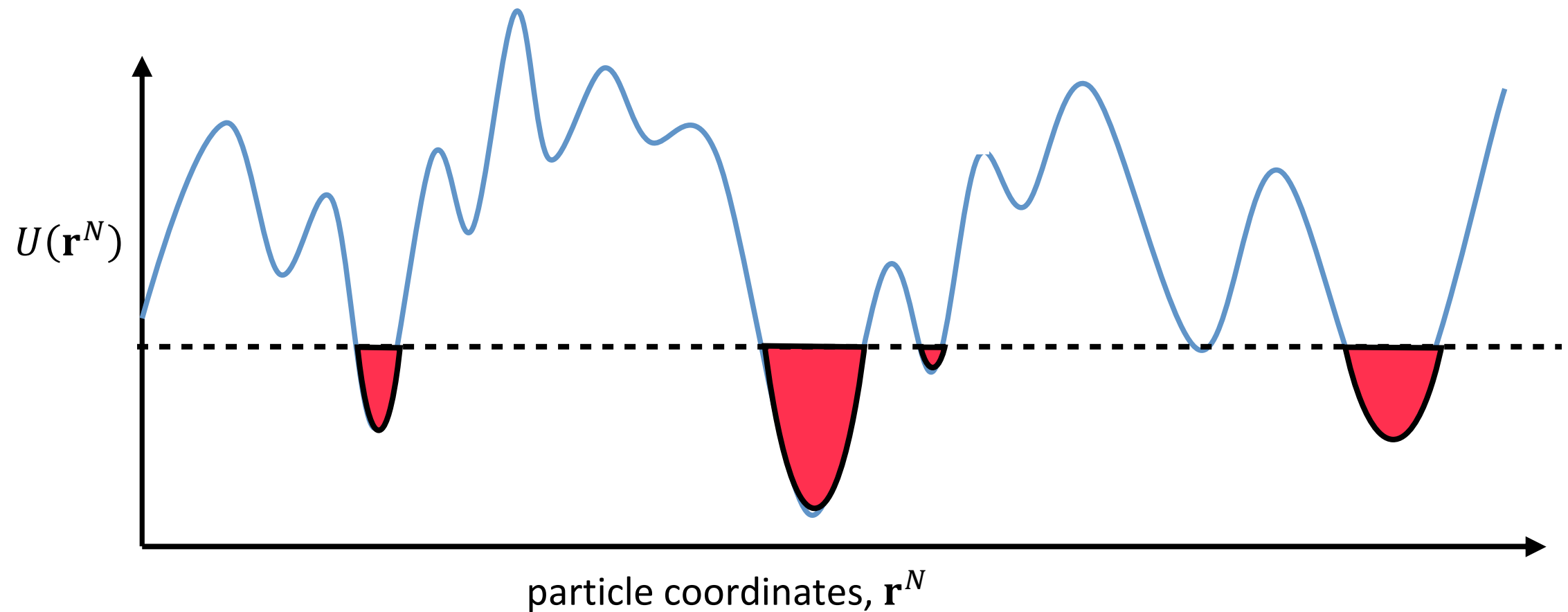
Equilibrium properties are related to energy landscapes

$$\mathcal{P}_i(\mathbf{p}_i^N, \mathbf{x}_i^N) = \frac{e^{-E(\mathbf{p}_i^N, \mathbf{x}_i^N)/k_B T}}{Q}$$

$U(\mathbf{r}^N)$

particle coordinates, $\mathbf{r}^N$
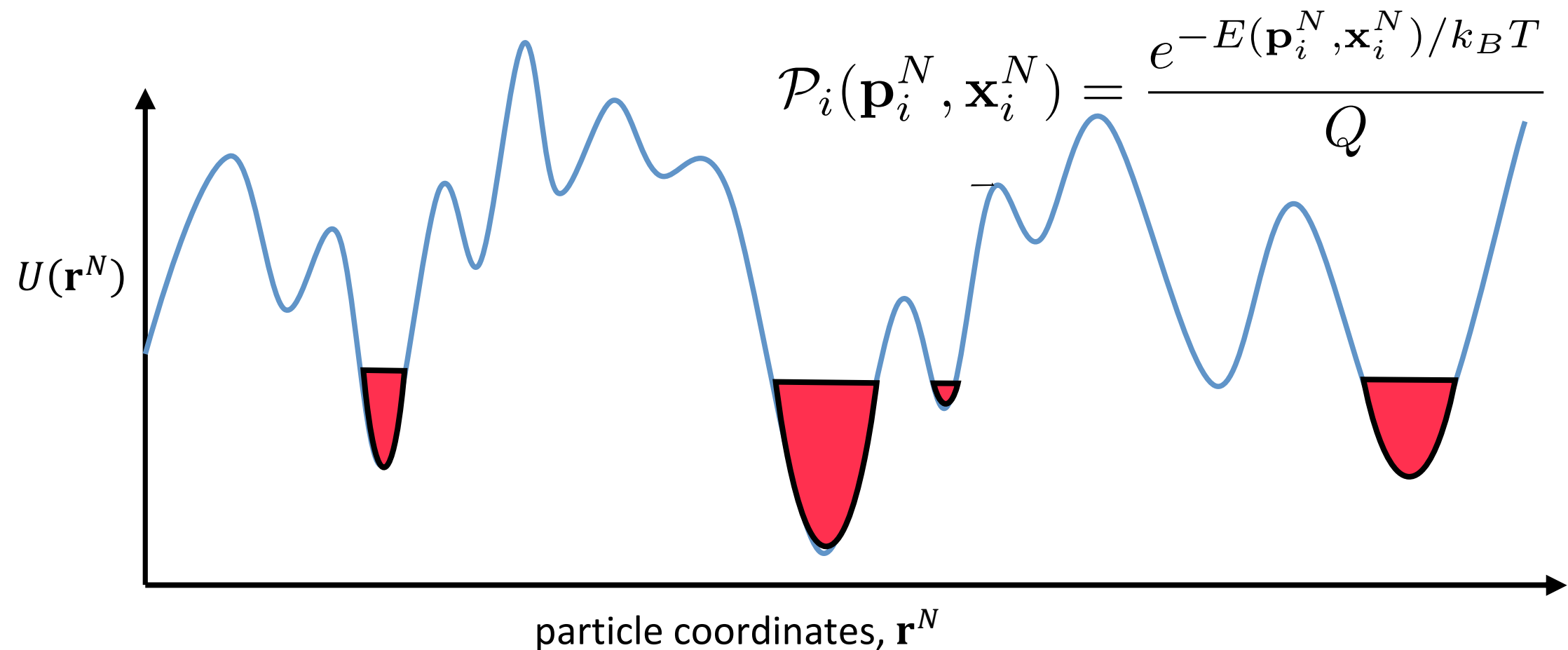
If we are interested in computing some observable W, we have to average over populated states:

$$\langle W \rangle = \int W(x)\mathcal{P}(x)dx$$

(structure, spectroscopic properties, distance, stability, entropy, enthalpy, ...)

Minimization is not an effective tool for exploring energy landscapes

- Minimization only takes us downhill
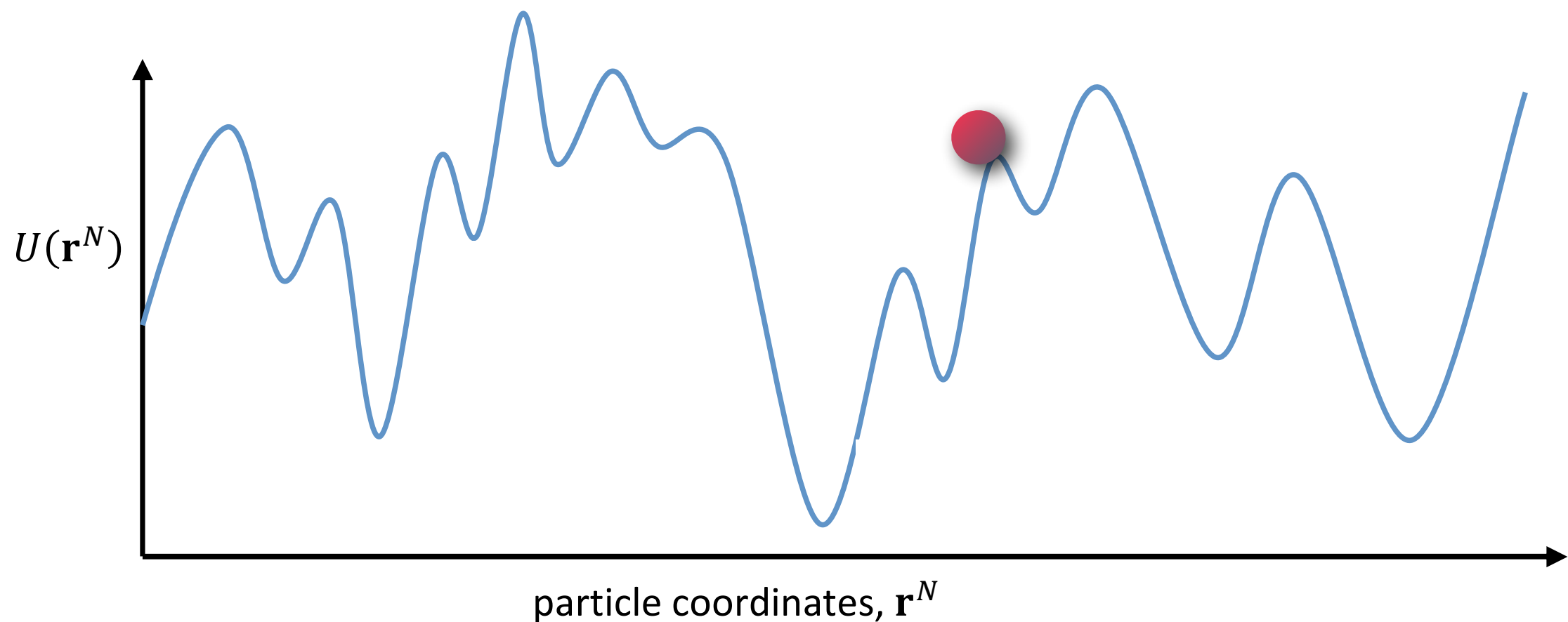
We need a better way to get equilibrium properties and dynamics

- Minimization only takes us downhill

$U(\mathbf{r}^N)$

particle coordinates, $\mathbf{r}^N$
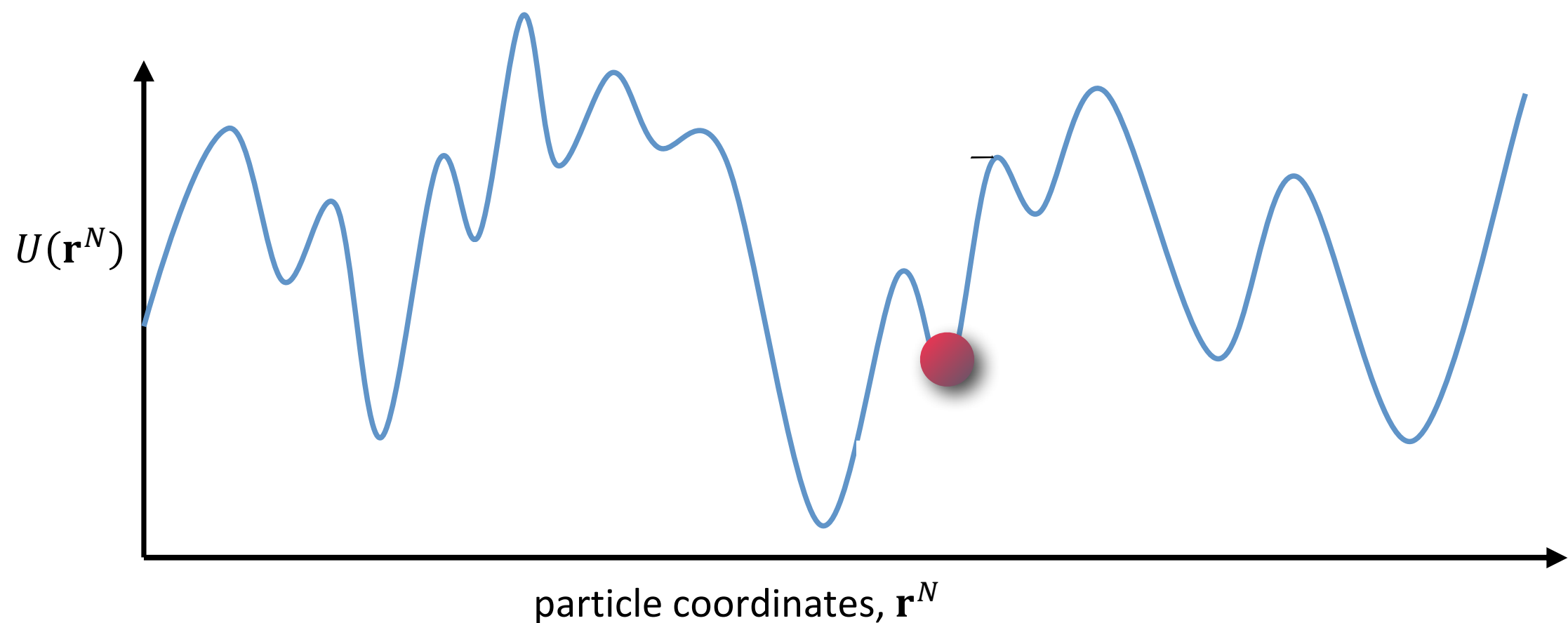
What about tracking the motion (dynamics) of the system?
If it's a classical system, use Newton's laws

- ## Newton's Laws

  - A body continues to move in straight line at constant velocity unless a force acts upon it

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

  - Force equals rate of change of momentum

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} = \frac{d(m\mathbf{v})}{dt}, \qquad \mathbf{F} = m\frac{d\mathbf{v}}{dt} = m\mathbf{a},$$

  - To every action there is an equal and opposite reaction

In MD successful configurations of the system are generated by integrating Newton's laws of motion

- Trajectory is obtained by solving the differential equation

$$\frac{\mathrm{d}^2 x_i}{\mathrm{d}\,t^2} = \frac{F_x i}{m_i}$$

- In real life models

  - Force on each particle will change whenever the particle change in position, or whenever any of the other particles with which it interacts changes position

- We need a 'integrator' to solve the equations

  - Integrators work based on the idea of taking small steps in time and predicting future coordinates and velocities from those at the present step

A Taylor expansion is a way of writing the approximate value of something in terms of its initial value and derivatives of its change

Taylor Expansion

$$f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \cdots.$$

$n=0$
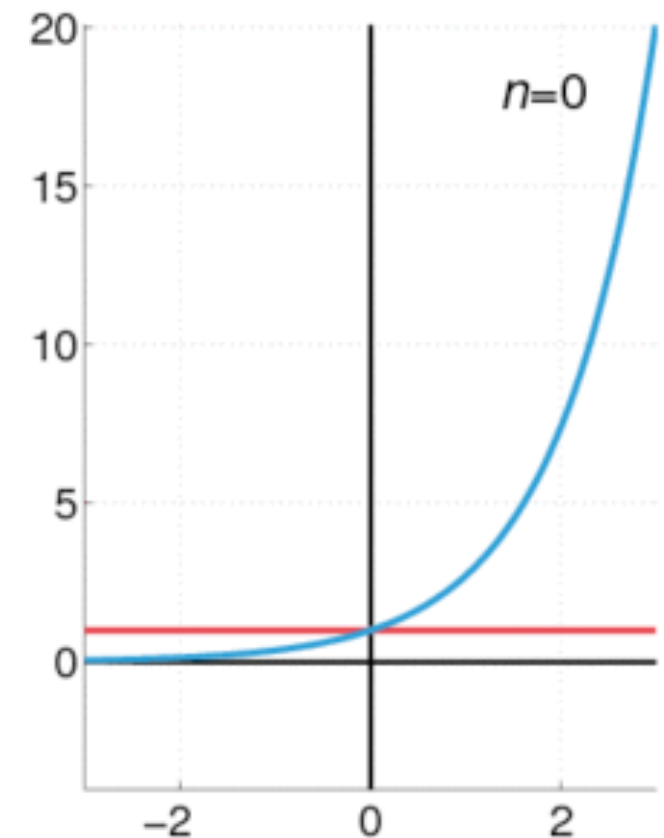
$$e^x = 1 + \frac{x^1}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \cdots = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \frac{x^5}{120} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

A Taylor expansion is a way of writing the approximate value of something in terms of its initial value and derivatives of its change

Taylor Expansion

$$f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \cdots.$$

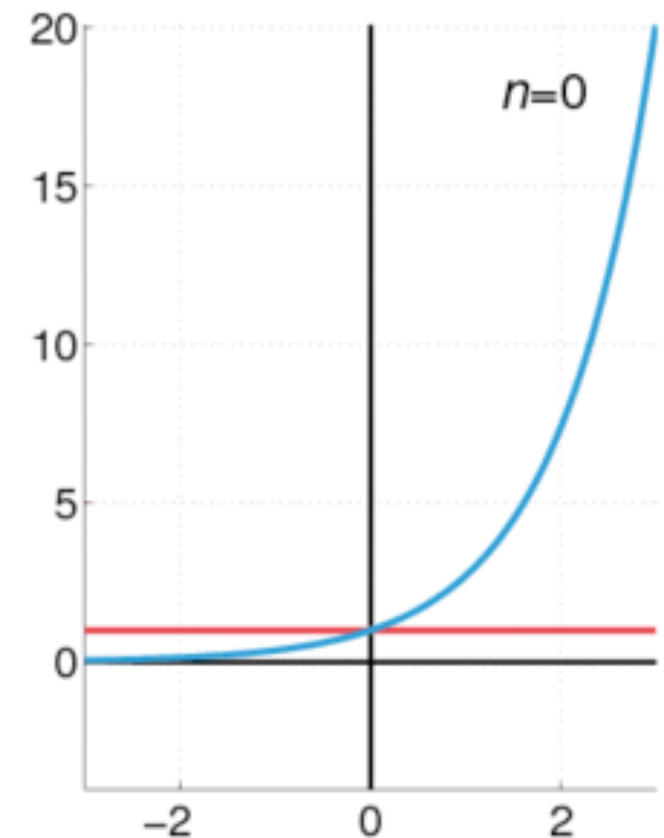$$e^x = 1 + \frac{x^1}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \cdots = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \frac{x^5}{120} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

The position Verlet algorithm is a well-known MD integrator

- Begin with a Taylor expansion of r:

$$r(t + \Delta t) = r(t) + \frac{dr(t)}{dt}\Delta t + \frac{d^2 r(t)}{dt^2}\frac{\Delta t^2}{2} + \frac{d^3 r(t)}{dt^3}\frac{\Delta t^3}{6} + \vartheta(\Delta t^4)$$

$$= r(t) + v(t)\Delta t + \frac{f(t)}{m}\frac{\Delta t^2}{2} + \frac{d^3 r(t)}{dt^3}\frac{\Delta t^3}{6} + \vartheta(\Delta t^4)$$

- Also note

$$r(t - \Delta t) = r(t) - v(t)\Delta t + \frac{f(t)}{m}\frac{\Delta t^2}{2} - \frac{d^3 r(t)}{dt^3}\frac{\Delta t^3}{6} + \vartheta(\Delta t^4)$$

- Add the equations and rearrange:

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + \frac{f(t)}{m}\Delta t^2 + \vartheta(\Delta t^4)$$

And of course $\quad f(t) = -\dfrac{dU(r(t))}{dr}$

The position Verlet algorithm does not use the velocity or track it

- The position Verlet algorithm uses only positions and forces; if we want the velocities we have to calculate them

  - Can be approximated with $v(t) = \dfrac{r(t + \Delta t) - r(t - \Delta t)}{2\Delta t} + \vartheta(\Delta t^3)$

- One disadvantage: requires storing two sets of positions -- r(t), r(t-Δt)

Velocity Verlet tracks velocities and doesn't need to track positions

- Expressions can be reformulated to give:

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2$$

$$v(t + \Delta t) = v(t) + \frac{f(t + \Delta t) + f(t)}{2m}\Delta t$$

- This provides both positions and velocities, without having to use positions from two different steps

- The position at the next time is given in terms of the current position, velocity, and force; after this is calculated, the velocity at the next time is given in terms of the current velocity, the new force, and the old force

Velocity Verlet tracks velocities and doesn't need to track positions

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2$$

$$v(t + \Delta t) = v(t) + \frac{f(t + \Delta t) + f(t)}{2m}\Delta t$$

- How it works in practice:

    1. Take current set of positions and compute force

        Now we have current velocity, position and force

    2. Insert these into above equation for position and find new position

    3. Compute new force

    4. Insert the new force with the old force into the velocity equation, find new velocity

    (Repeat)

32

There are other integrators, such as leapfrog and higher order methods

- Leapfrog uses velocities found *only* on half-step integrals

$$v(t + \Delta t/2) = v(t - \Delta t/2) + \frac{f(t)}{m}\Delta t$$

$$r(t + \Delta t) = r(t) + v(t + \Delta t/2)\Delta t$$

- Higher order integrators go out further in the Taylor expansion

- Can take bigger timesteps with same level of accuracy (to Newton's eqns)

- But going to higher order makes them slower

- Often better/easier to stay with lower order method and take smaller steps

There are several considerations in setting timesteps and choosing an integrator

- Accuracy to Newton's equations

- Energy conservation

  - Short-term: Fluctuations in energy from step to step; not so good in Verlet type algorithms, but average out over time

  - Long-term: Drift -- does the total energy go up or down over time? Verlet algorithms fairly good. Again, affected by timestep -- larger timesteps mean more drift.

- Stability: Do large forces easily result in motions that are too large and crashes?

- Velocity Verlet: Pretty stable, reasonable long-term energy drift, fast -- overall good (and common) choice

Timesteps typically must be chosen to be smaller than fastest timescales in system

- Bond vibrations often require timesteps 0.5-2.0 fs

  - (1 fs = 10^{-15} s)

  - Bonds to hydrogen can be even faster

  - Constraining bond lengths to hydrogen often allows 1-2 fs timesteps

- Bond bending, 2.0 fs

- Translational motion, 5-10 fs

- Too long a time step often causes "exploding"

Initial positions and velocities must be specified before starting a simulation

- **Initial positions**

  - Can be taken from a structure if available

  - Place consistent with constraints such as bond angles, idealized geometry

  - Can be placed on lattice

  - Or placed randomly based on sterics (utilities such as packmol do this)

  - THEN **energy minimize**

    - Failure to energy minimize often results in instability, "exploding"

- **Initial velocities -- assign randomly from Maxwell Boltzmann distribution**

$$\wp(v_{x,i}) = \left(\frac{m_i}{2\pi k_B T}\right)^{\frac{1}{2}} \exp\left(-\frac{m_i v_{x,i}^2}{2k_B T}\right) \qquad \text{(Gaussian, width related to temperature)}$$

Integrators don't follow exact motions, but this is generally OK

- Two trajectories starting from exactly the same set of configurations but very slightly different momenta will quickly become very different

  - Difference grows exponentially in time!

  - Good in that it means it's easy to generate different sets of results from one starting structure

  - It can thus be hard to exactly reproduce previous results

  - We can't claim to be tracking true Newtonian solution exactly

- That's OK:

  - We are interested in statistical properties, not single trajectories

  - Average properties of long simulations, or of many separate simulations

37

If we are interested in correct statistical properties, we need populations to be right, not necessarily dynamics

- If we want the correct populations, we need integrators to be:

  - Time reversible (Verlet is)

  - Area preserving (Verlet is):

    - The amount of energy landscape at a particular energy has a certain area which should be constant

    - If our algorithm allows the area to change, it means that the energy will not be conserved

Recap

- MD integrators work based on taking small timesteps and solving for positions, velocities at each step, based on the forces

- Velocity Verlet and related integrators are good, stable choices that yield correct distributions