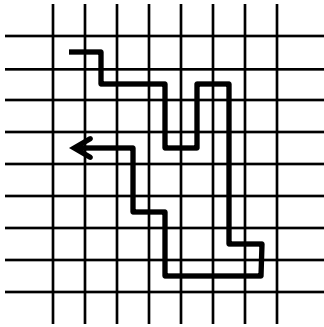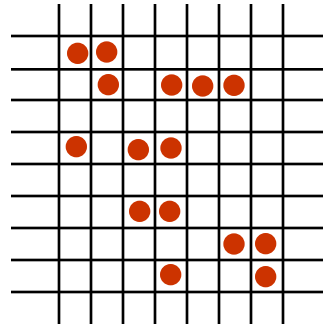# Monte Carlo and MD simulations

Andrew Torda, April 2009 strukt und sim

- What we observe in any system ?
  - averages of observables (pressure, energy, density)
- Given enough time system will visit all states

time

random
hopping

- my observable *A*

$$A_{obs} = \frac{1}{(b-a)} \int_a^b A_t dt$$

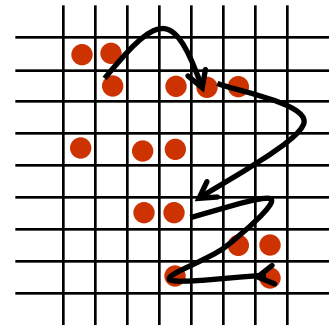$$A_{obs} = \frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} A_i$$

Time and space averages

- If we believe

$$A_{obs} = \frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} A_i$$

- then

$$A_{obs} = \sum_{j}^{states} p_j A_j$$

$$\equiv \langle A \rangle$$

- $<A>$ is ensemble average and usually $\overline{A}$ is time average
- so if sample with correct probability, we can find A_{obs}
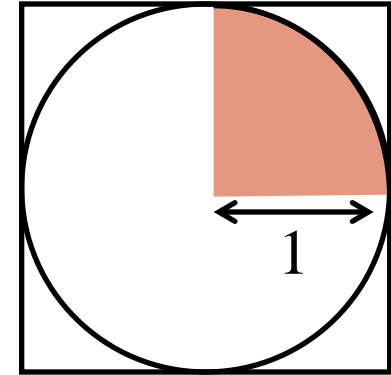- order of visiting states does not matter

Monte Carlo

- How to calculate π with random numbers

$$\frac{\text{points}_{red}}{\text{points}_{square}} = \frac{\frac{1}{4}\pi r^2}{\text{area in square}}$$

$$\pi = \frac{4\ \text{points}_{red}}{\text{points}_{square}}$$

while (not happy)

    pick random x, y

    n_{square}++

    if $((x^2+y^2) < 1)$ n_{red}++

print 4 n_{red} / n_{square}

Generating distributions / Monte Carlo

- generating points in a circle ? (a bit silly)

$$p_{in\_circle} = \begin{cases} 1 & x^2 + y^2 \leq 1 \\ 0 & x^2 + y^2 > 1 \end{cases}$$

- we could work out the area of a circle (integrate) by picking random numbers
- what does Monte Carlo simulation mean ?
  - generating points according to some distribution to find an average or integral
- what is our distribution in physical systems ?
  - Boltzmann distribution

Monte Carlo and Boltzmann distributions

- Boltzmann probability distribution

$$p_i = \frac{e^{-E_i/kT}}{\sum_j e^{-E_j/kT}}$$ often written as $$p_i = \frac{e^{-E_i/kT}}{Z}$$

- if we could generate this distribution, we could reproduce most properties of a system
- leads to a scheme (not possible)

an impossible scheme

while (not happy)

    generate configuration $\mathbf{r}_i$ (conformation of protein, …)

    calculate p_i (number between 0 and 1)

    generate random number x

    if $(x < p_i)$

        accept $\mathbf{r}_i$

    else

        reject $\mathbf{r}_i$

$$p_i = \frac{e^{-E_i/kT}}{\sum_j e^{-E_j/kT}}$$

- result ? a set of $\mathbf{r}_i$ with Boltzmann distribution
- problem ? we do not know $\sum_j e^{-E_j/kT}$

a better scheme

- we cannot generate points from $\quad p_i = \dfrac{e^{-E_i/kT}}{\sum\limits_j e^{-E_j/kT}}$

- what if we have two configurations ? $\quad \dfrac{p_i}{p_j} = \dfrac{e^{-E_i/kT}}{Z} \quad \dfrac{Z}{e^{-E_j/kT}}$

- if we have one configuration to start
  - we can work out the probability of a second $\qquad = e^{E_j - E_i/kT}$
- convenient convention $\qquad\qquad = e^{-\Delta E/kT}$
  - going from old→new $\Delta E < 0,$
    - $E_{new} - E_{old} < 0$, energy is better / more negative

Metropolis Monte Carlo

$$\frac{p_{new}}{p_{old}} = e^{-\Delta E/kT}$$

- generating a distribution
- if $\Delta E < 0$, new is likely (more than 1)
- if $\Delta E > 0$, old is p_{new} is possible

- ```
  generate starting configuration r_o
  while (not happy)
      generate r_new
      calculate E_new and ΔE
      if ΔE < 0
          set r_o to r_new
      else
          x = rand [0:1]
  ```
  $$\text{if}\left(x \le e^{-\Delta E/kT} \right)$$
  ```
              set r_o to r_new
  ```

- what if ΔE slightly > 0 ?
  - 0.0000000001
- what if $\Delta E = 10^6$?
- small uphill moves are OK
- bigger moves are less likely

Properties of Monte Carlo

- the set of $\mathbf{r}_o$ is a valid distribution (ensemble)
  - for some property A

$$A_{obs} = \langle A \rangle = \frac{1}{N_{visited}} \sum_i^{N_{visited}} A_i$$

  - A could be density, structural property, E, …
- only works for one temperature T

- does this fit with picture ?
- could I calculate entropy / free energy ?
  - only for very simple systems

better U / energy

more arrangements

$E(\mathbf{r})$

configurations ($\mathbf{r}$)
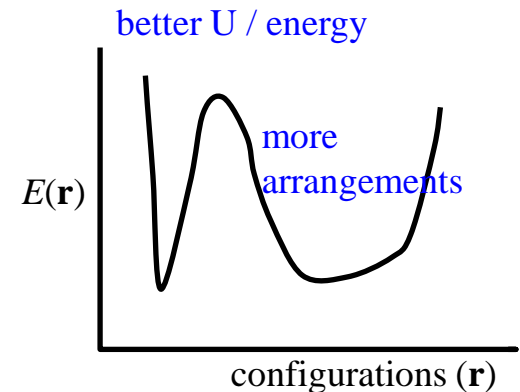
Equilibrium

- MC only for system at equilibrium / generates system at equilibrium
- Toy system with 3 states
  - for some T
  - $p_1 = \frac{5}{8}$ $\qquad$ $p_2 = \frac{1}{4}$ $\qquad$ $p_3 = \frac{1}{8}$
  - if I have 80 copies of the system, most are in state_1
  - I start a simulation with 70 copies in state 2. 5 in other states
  - all moves $2\rightarrow1$ are accepted
  - moves $1 \rightarrow 2$ are less frequent +
    - before equilibrium distribution state_1 is less populated
  - moving to equilibrium depends on
    - population
    - probability

E $\uparrow$ E_3

E_2

E_1
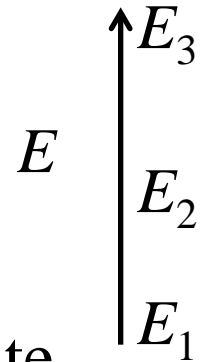
Detailed balance

- For any two states (state$_i$ and state$_j$)
- flow $i \rightarrow j$ must equal $j \rightarrow i$
  - otherwise ?
- flow $i \rightarrow j$ depends
  - population N_i
  - probability $\pi(i \rightarrow j)$

Detailed balance
$$N_i\, \pi(i \rightarrow j) = N_j\, \pi(j \rightarrow i)$$
- detailed balance must apply for any pair i, j

Ergodic

- Assumptions
  - I can do integrals because
    - I will visit every state
    - I can calculate p_i for all states
  - I will visit every state
- alternatively

- for any i, j
  - $\pi(\, i \rightarrow j) > 0$
  - may require a finite number of steps: $i \rightarrow k \rightarrow m \rightarrow j$ OK
- must be satisfied

Moves

version 1

- decide on r_{max}
- pick a particle at random
- pick random Δx, Δy, Δz
  - $0 < \Delta a < r_{max}$
- apply move
- accept / reject move
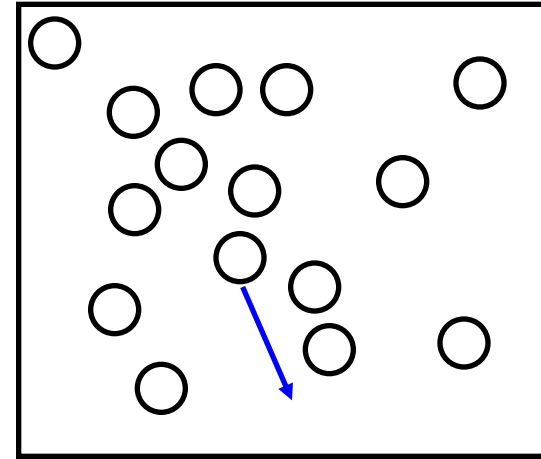
version 2

- decide on smaller r_{max}
- foreach particle
  - pick random Δx, Δy, Δz
    - $0 < \Delta a < r_{max}$
- apply move
- accept / reject
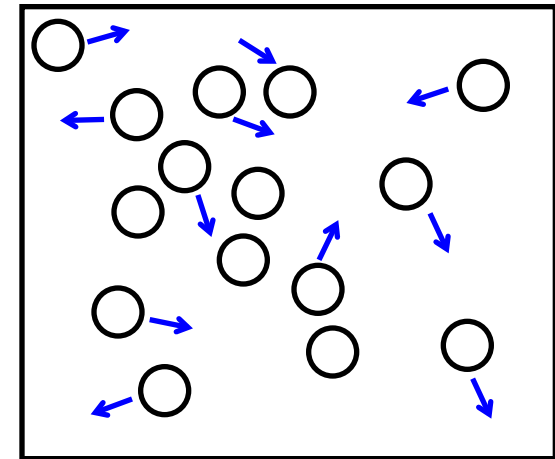
Moves

- both kinds of move OK
- note
  - "accept / reject"
- more generally,
  - how big is r_{max} ?
  - big
    - system moves faster
    - more moves rejected
- what if my particles are not spheres ?
  - rotations also necessary

- time has no meaning
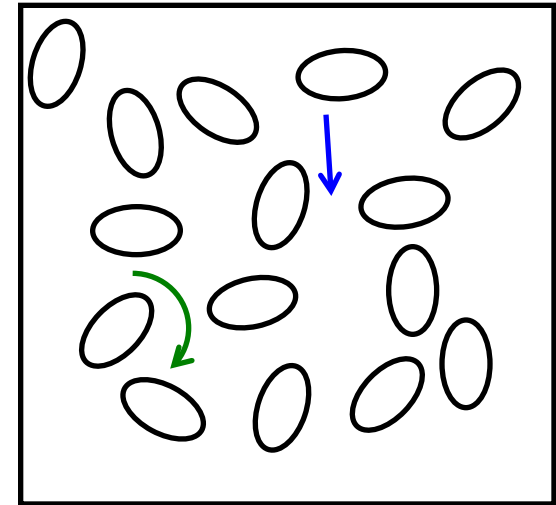
Bonded systems

- protein (lipid, polymer, ..)
- random Δx ?
  - nearly all will stretch a bond
    - high energy : rejected move
- only feasible method
  - random rotations $\Delta\theta$
- in general
  - most kinds of simple moves OK
  - must maintain detailed balance, ergodicity
  - question of efficiency
    - high rejection rate means lots of wasted calculations
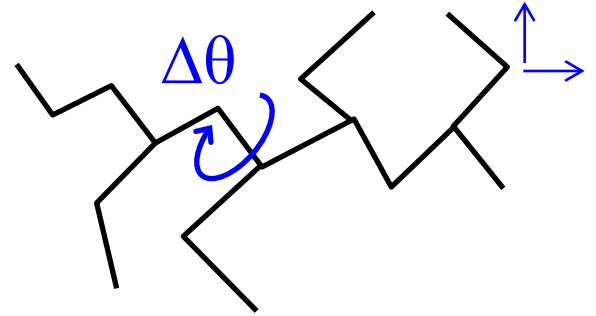
NVT

- Remember back to Gibbs / Helmholtz

$$\frac{p_{new}}{p_{old}} = e^{-\Delta E/kT}$$

- I have defined temperature
  - and $N_{particles}$ and V
- called NVT simulation
- could I have varied something else ?
  - what if I tried to put particles in / take out ?
    - sometimes energy ↑sometimes↓
  - the system will fluctuate around $<N>$
  - this would not be NVT
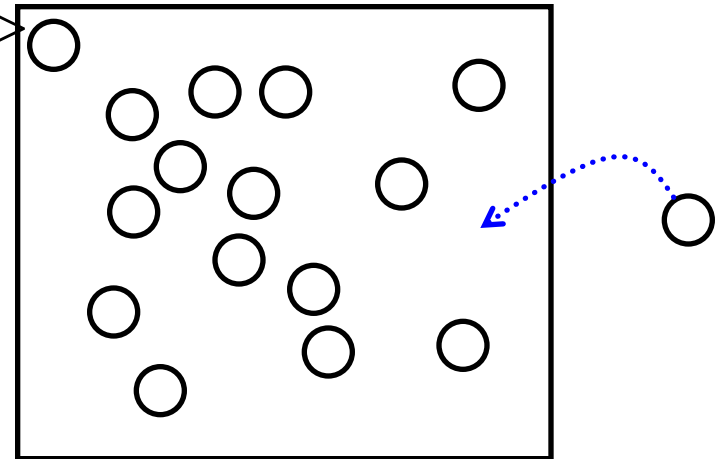
Periodic Boundary Conditions

- technical point relevant to gases, proteins in water…
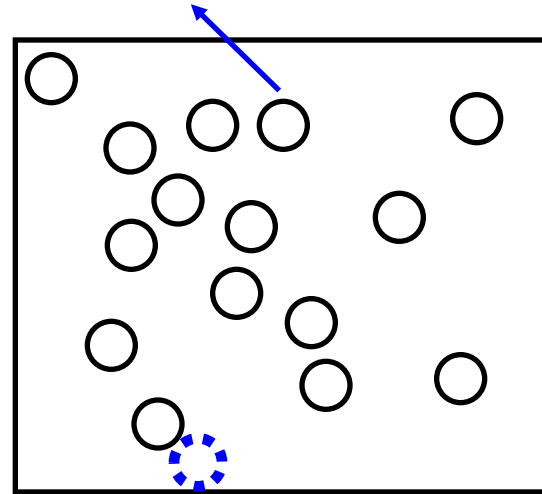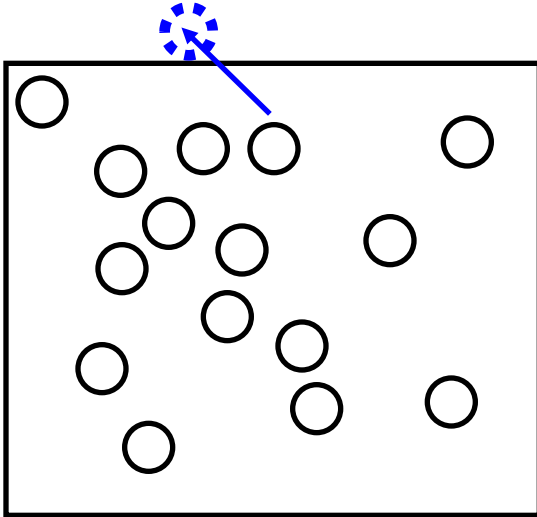
- behaves like an infinite system
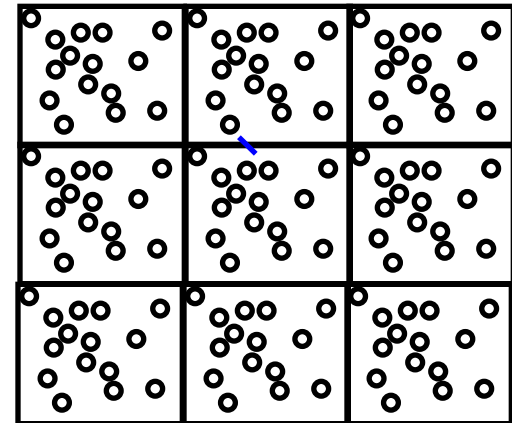
Infinite interactions ?

- neighbours of blue particle
  - only use the nearer
  - not really an infinite system

- volume defined by box
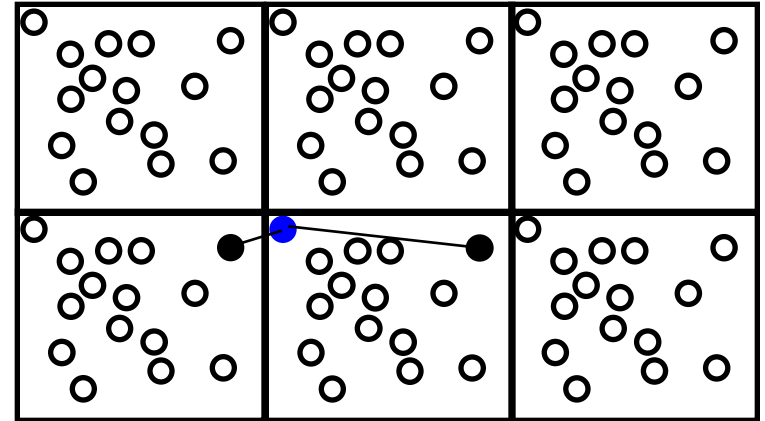
Problems with Monte Carlo

- basic scheme

  while (not happy)

      propose move

      accept / reject move

- if we use small steps
  - system moves slowly: long time to visit all states
- big steps ?
  - calculate energy
  - reject move
    - no progress, wastes time

Dense Systems and Monte Carlo

- If we do random moves ?
  - most moves rejected
- dense systems ?
  - liquids
  - proteins, polymers, …
- Solutions
  - cleverer MC moves (later)
  - MD
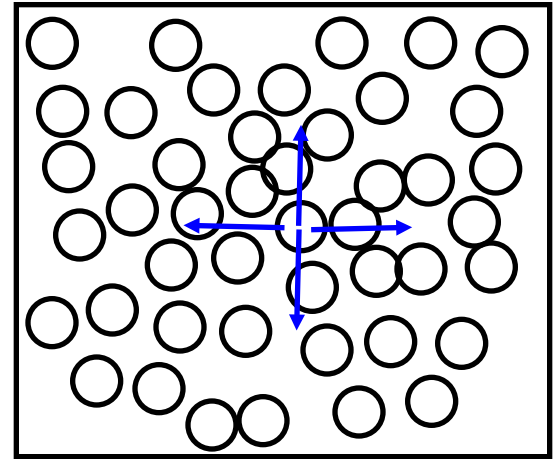
Why do molecular dynamics simulations ?

Real world

- box of gas, molecule in space, protein molecule in water
- atoms hit each other,
    - share energy, box expands/contracts, ..
    - soon reaches equilibrium
    - visits low energies (often), high energies (less often)
    - visits entropically favoured regions
- we stick in a thermometer
- measure density, …

- what have the atoms done ?
    - feel forces and move
    - an MD simulation just copies this

What do we expect ? Molecular Dynamics

$$\vec{F} = -\frac{\partial U}{\partial \vec{r}}$$
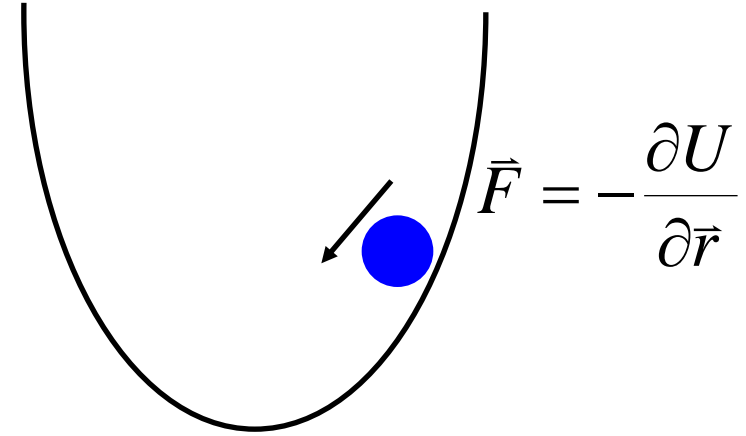
one particle in a well

- Unlike MC, particles have kinetic energy E_{kin}

Kinetic and potential energy

- Our system is isolated (no work done)
- E_{tot} never changes
  - conserves energy (no work done on system)
- $E_{tot} = E_{pot} + E_{kin}$

E_{kin} small
E_{pot} big

E_{kin} big
E_{pot} small
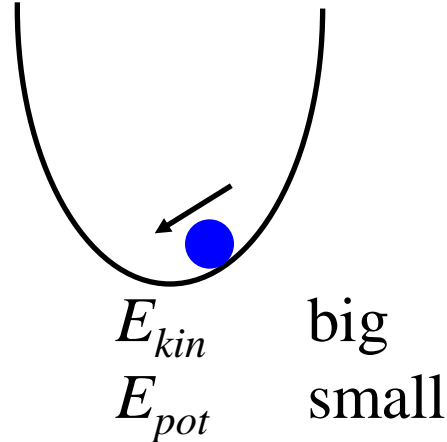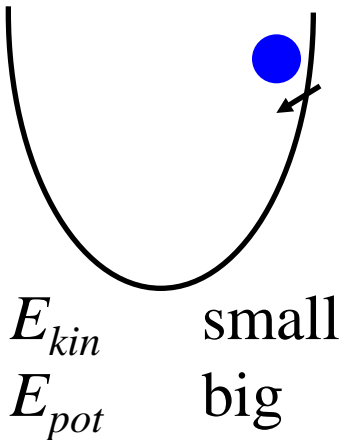
- For one particle $E_{tot} = E_{pot} + E_{kin} = $ constant

Lots of particles

- particles hitting each other
  - exchanging energy
- total system
  - conserves energy
- one particle ?
  - maybe at bottom but moving slow ($E_{kin} + E_{pot}$ small)
  - per particle energy no longer conserved (may gain or lose)
- many particles
  - distribution of velocities
  - distribution of potential energies
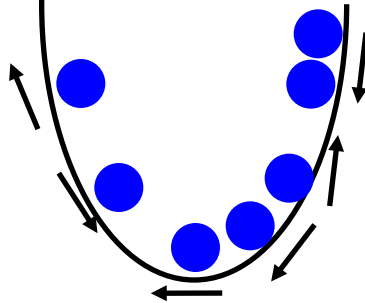
Boltzmann distribution in real world

One version of real world (N, V, T)

- constant number of particles, volume, temperature
- today $E = E_{kin} + E_{pot}$
- Z is partition function
- earlier

$$p_i = \frac{e^{-E_i/kT}}{Z}$$

$$Z = \sum_i e^{-E_i/kT}$$

- but now we have kinetic energy E_{kin} (**p**)
- where $\mathbf{p} = m\dot{\mathbf{x}}$
  - potential energy $E_{pot}(\mathbf{r})$
- if we write in continuous form …

Partition function for MD

- Usually write $H(\mathbf{p},\mathbf{r}) = E_{kin}(\mathbf{p}) + E_{pot}(\mathbf{r})$
  - "Hamiltonian"
- All the states are defined by all possible momenta and coordinates
  - sum over these: $Z(N,V,T) \propto \int d\mathbf{p} \int d\mathbf{r}\, e^{-H(\mathbf{p},\mathbf{r})/kT}$

MD Method

- For any particle we can calculate forces
- Newtons law
  - $F = ma$ often better written

  $$\ddot{\vec{x}} = \vec{F}m^{-1}$$
- if we know acceleration
  - we can get velocity
- from velocity
  - can get coordinates

$$\vec{F} = -\frac{\partial U}{\partial \vec{r}}$$

```
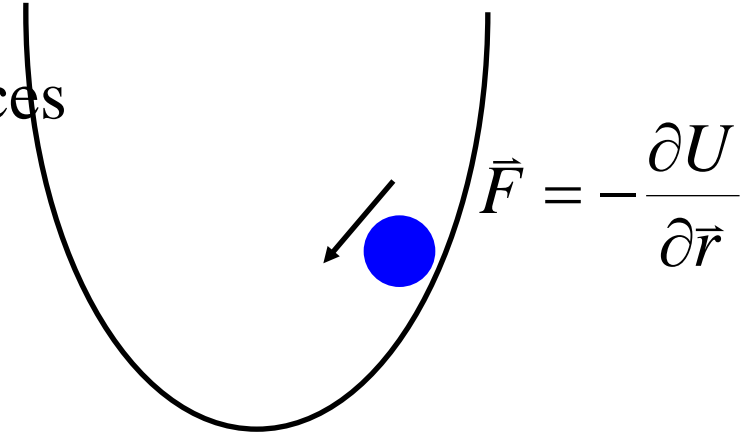while (nstep < max_step)
    calculate forces
    integrate to get new coordinates
    nstep ++
```

$\Big\}$ averaging, sampling, …

Starting system

Initial coordinates

- protein model
- protein from protein data bank (PDB)
- protein + proposed ligand
- box of liquid

Do initial coordinates matter ?

- in principle: no – infinitely long simulation visits all configurations, reaches equilibrium
- in practice: yes
  - bad examples
    - no simulation is long enough to predict protein conformation
  - take water configuration and run at ice temperature

Initial velocities

- first consider temperature – reflects kinetic energy

$$\left\langle \frac{1}{2} m v_\alpha^2 \right\rangle = \frac{1}{2} kT$$

- where v_α^2 could be v_x, v_y, v_z
- leads to definition

$$T(t) = \sum_{i=1}^{N} \frac{m_i v_i^2(t)}{kN_f}$$

- where N_f is number degrees of freedom $\approx 3N$
- we could use this to get initial velocities $\quad \left\langle v_\alpha^2 \right\rangle = kT / m$

Initial velocities

- would one $<v^2>$ be OK ?
  - not very good
    - E_{kin} correlated with E_{pot}
- either
  - use more sophisticated distribution
  - do not worry
    - ultimately system will go to equilibrium
      - velocities will reach sensible values
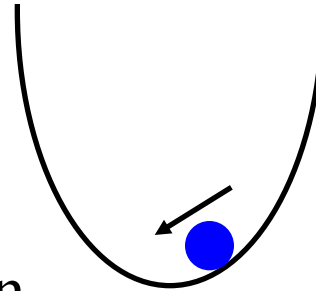
Getting new velocities / coordinates

- constant acceleration
  - $x_t = x_0 + vt + \frac{1}{2} at^2$

  - or $x_t = x_0 + \dot{x}t + \frac{1}{2}\ddot{x}t^2$
  - ok for ball falling in gravity
- try to use formula to predict future time

$\ddot{x}$ big
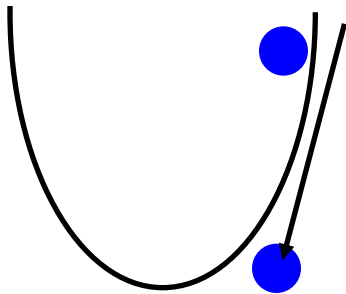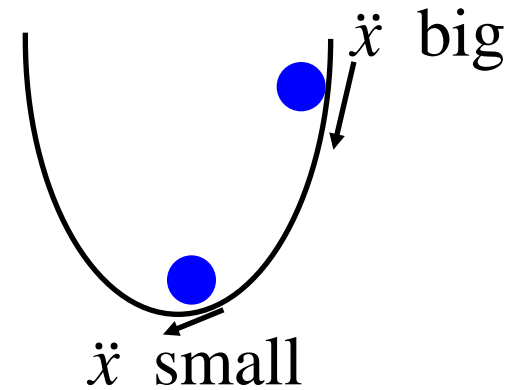
$\ddot{x}$ small

big Δt / step
big error
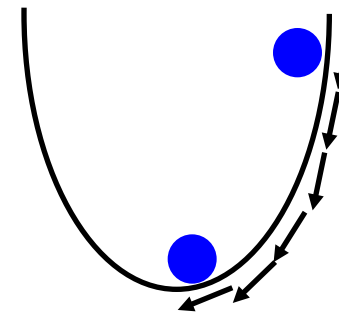
small Δt / step
small error
slow

Fundamental problem with integration

- We want to use big Δt (speed)
- We must use small Δt (accuracy)
- All Δt will give us some error
  - numerical integration is never perfect
- How small is Δt ?
  - depends on fastest frequency / steepest walls in energy
    - usually bonds
  - for proteins at room temperature
    - $\Delta t \approx 1$ fs (femtosecond 10^{-15}s)
  - high temperature Δt should be smaller
- practical integrators
  - remove velocity – slightly more sophisticated

Noise and heating

- general rule
  - noise heats the system
  - formally difficult to prove
  - $E_{kin} = \frac{1}{2}\,mv^2$

🔵 no kinetic energy   🔵 ↗↙ E_{kin} due to noise

🔵 ——————————→   🔵 ∿∿∿→ ↕extra velocity
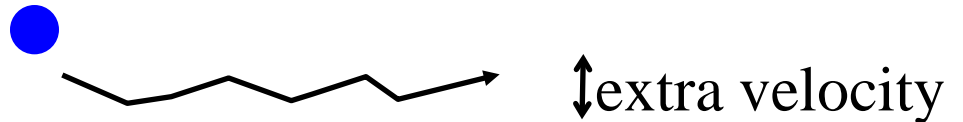
Noise-free Simulation

- Energy conservation : Absolute rule $E_{pot} = f(\mathbf{r})$
  - no time component
  - invariant under translation, rotation
- When violated ?
  - ($\mathbf{r}$) does not change, but E_{pot} changes: E_{tot} changes

Noise Sources

Integrator

- coordinates do not match velocity E_{kin} wrong: $(E_{kin} + E_{pot}) \neq$ constant
- energy not conserved
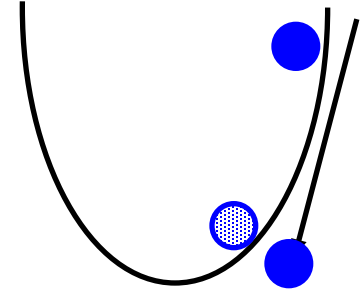
Numerical noise

- $E_{pot} = f(\mathbf{r})$
- initial coordinates ($\mathbf{r}$) quoted to 3 decimal places
- really less accurate

Cutoffs

- within cutoff rotation restricted
- outside cutoff rotation suddenly free
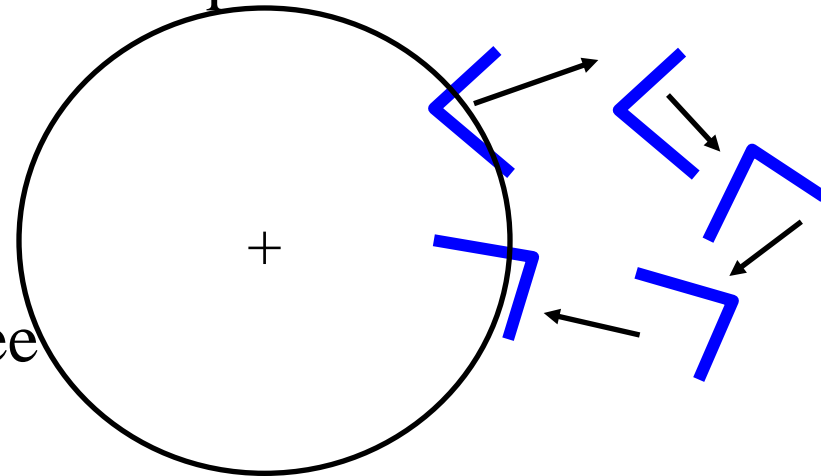
Result

- heating

Equilibrium

- Remember MC story
  - system not at equilibrium ? eventually equilibrates
- MD
  - start in high energy E_{pot}
  - E_{pot} converted to E_{kin}
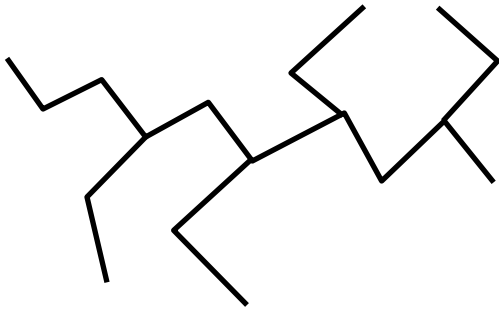
- Some high energy conformation
  - relaxes
  - E_{pot} converted to E_{kin}
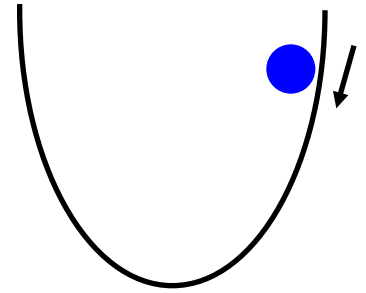
- MD system will not
  - really find low energy
  - known temperature

MD in a closed system

- An isolated molecule should not lose energy
- A repeated box will not lose energy
- Formally system is
  - NVE (constant $N_{particles}$, volume, energy)

- Problems
  - we want to set the temperature of the system
  - we may have noise / heat creating energy

- Cure
  - thermostat
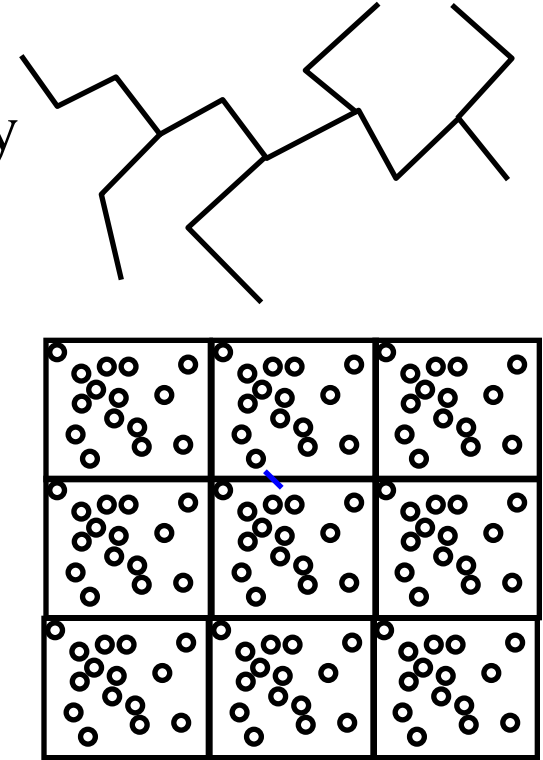
- Übung notes
  - when calculating rmsd maybe one needs the `-s` option to specify a topology file

- more elaborate explanation of coupling next week

Bath

- imagine infinite bath at desired temperature
- heat will flow in or out
- at equilibrium no flow of heat
  - maybe removal of noise/heat
- how to implement ? Many ways
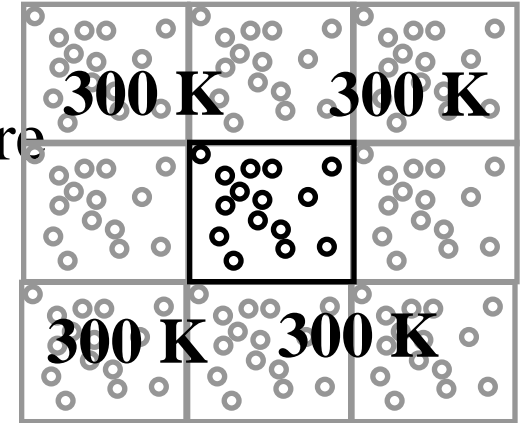
Occasionally:

1. introduce a fake particle desired temperature / collide
2. pick a particle at random / give average *v* for temperature
3. Easy method –weak coupling…

Weak Coupling

- Remember temperature* $E_{kin} = \sum_i^N \frac{1}{2} m_i v_i^2 = \frac{3}{2} NkT$

- Heat leaves system depending on how wrong temperature is

$$\frac{dT(t)}{dt} = \frac{T_0 - T(t)}{\tau_T}$$

- T_0 is reference temperature
- τ_T is a coupling / relaxation constant
  - τ_T tiny, heat moves fast. τ_T big, …
- to implement this idea ? Multiply velocities

*Slight simplification of formula

Classic reference: Berendsen, HJC, Postma, JPM, van Gunsteren, WF, DiNola, A, Haak, JR, "Molecular dynamics with coupling to an external bath", J. Chem Phys, 81, 3684, (1984)

Implementation of weak coupling

- scale velocities, $v_{new} = \lambda\ v_{old}$
$$\lambda = \left(1 + \frac{\Delta t}{\tau_T}\left(\frac{T_0}{T} - 1\right)\right)^{\frac{1}{2}}$$
- intuitively
  - Δt (time step) big ? temperature will change more
  - what if $T_0 = T$?
  - square root ?
    - wrong T reflects a difference in v^2
- can we break this ?
  - scaling is applied to whole system
    - what if part is hot ?
    - it may remain hotter than black part
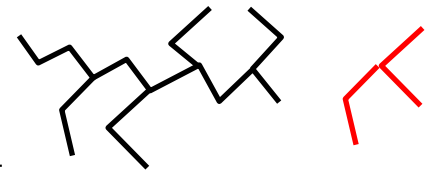
Importance of heat baths

- Does not conserve energy
- In principle
  - bring a system to equilibrium
- In practice
  - avoid damage due to numerical errors / approximations
- For a system at equilibrium
  - heat bath should do nothing

- Does allow for artificial tricks
  - gently heat a system and watch behaviour
  - gently cool a system and "anneal" it (more later)
- Extension to other properties
  - analogous reasoning for pressure bath

Summary of MD

- Philosophy
  - natural way to copy/model/simulate nature
- Lets one model processes in real time

MC	MD
• any cost/energy OK	requires continuous $E_{pot}(\mathbf{r})$
• time usually invalid	gives time scales
• most moves OK	physical trajectories
• Temperature enforced by acceptance	has explicit E_{kin}

both include entropic terms