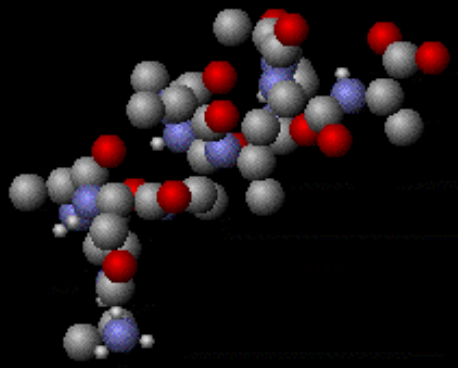
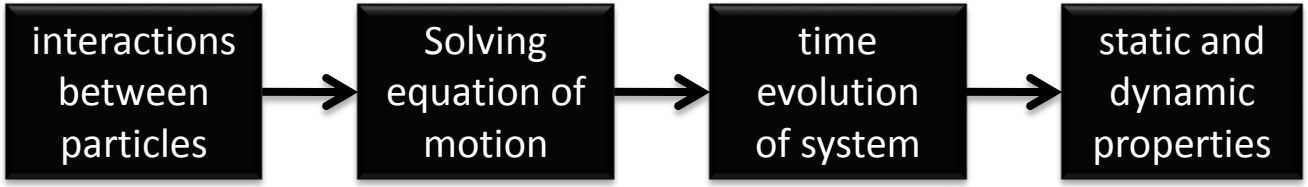


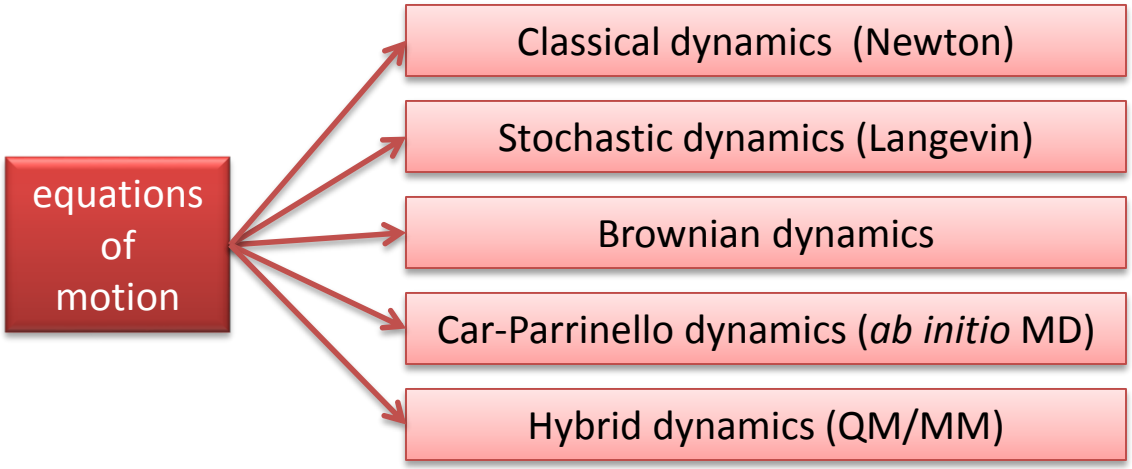
Molecular Dynamic Simulations

Yazdan Asgari
2020

Molecular Dynamic Simulation

detailed microscopic modeling on atomic scale

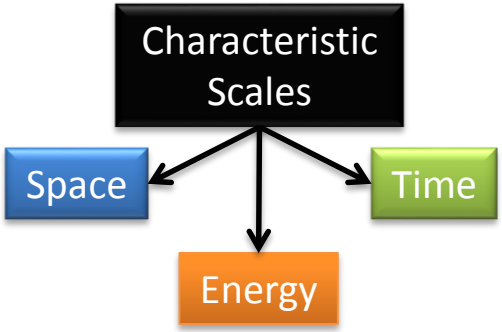




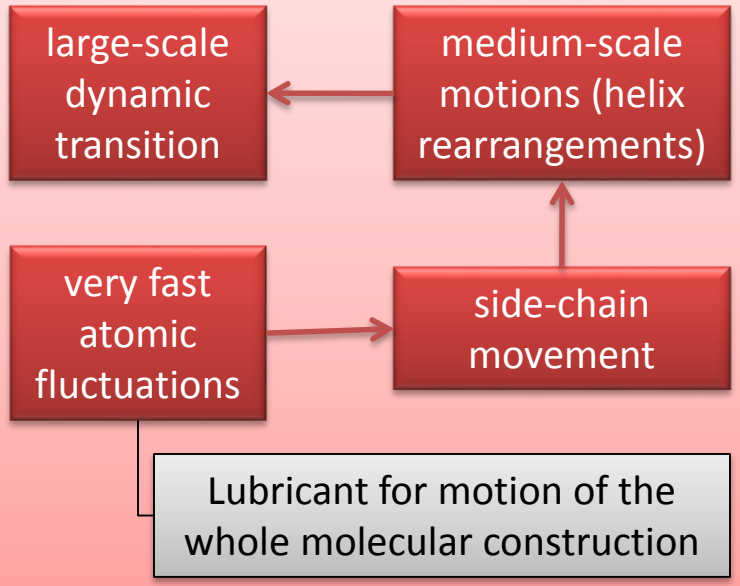
Types of Macromolecular Motions

Macromolecules (Proteins)
a broad range of characteristic motions

fast localized ⟷ slow large-scale



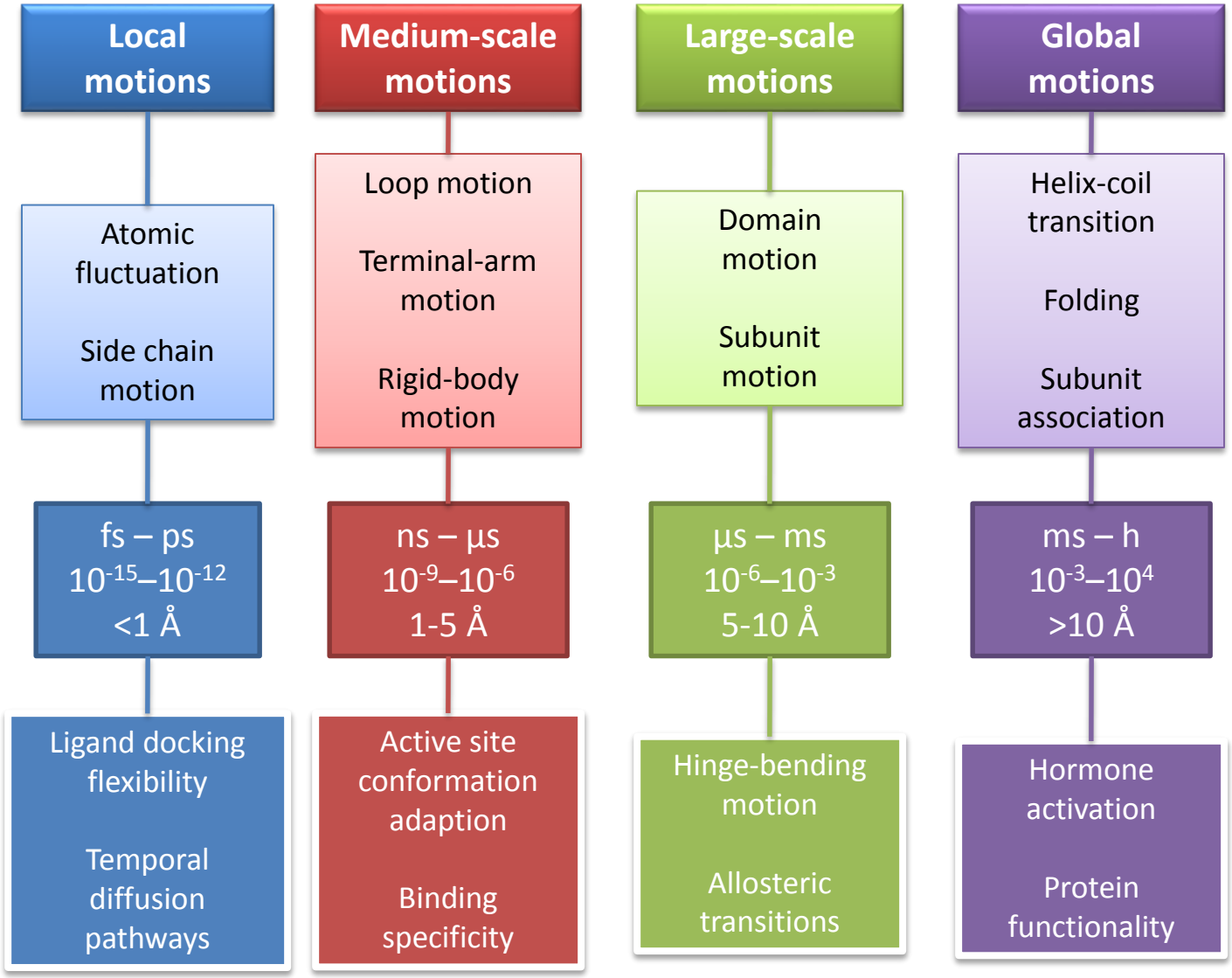
interdependency and coupling of motions



Even in the study of slow large-scale motions it is not possible to ignore fast small-scale motions.

Fast small-scale motions impose limitations on the simulation time step and length.

Characteristic Motions in Proteins



Statistical Mechanics

microscopic states

microscopic variables

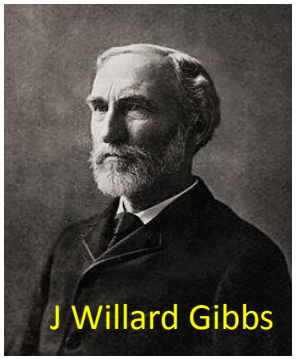
$\{\mathbf{r}_1, \mathbf{p}_1, \mathbf{r}_2, \mathbf{p}_2, \ldots, \mathbf{r}_N, \mathbf{p}_N\}$

Statistical Mechanics

macroscopic states

macroscopic variables

$\{N, V, E\}$ or $\{N, V, T\}$ or ...



For each *macrostate* very (very very) many different *microstates* are possible.

Ensemble
Collection of huge number of mental copies of a system

Ensemble Averages
macroscopic quantities are averages of the corresponding microscopic quantities over all systems in the ensemble.

Microcanonical Ensemble

N, V, E	N, V, E	N, V, E
N, V, E	N, V, E	N, V, E
N, V, E	N, V, E	N, V, E

Canonical Ensemble

N, V, T	N, V, T	N, V, T
N, V, T	N, V, T	N, V, T
N, V, T	N, V, T	N, V, T

Microscopic Motion

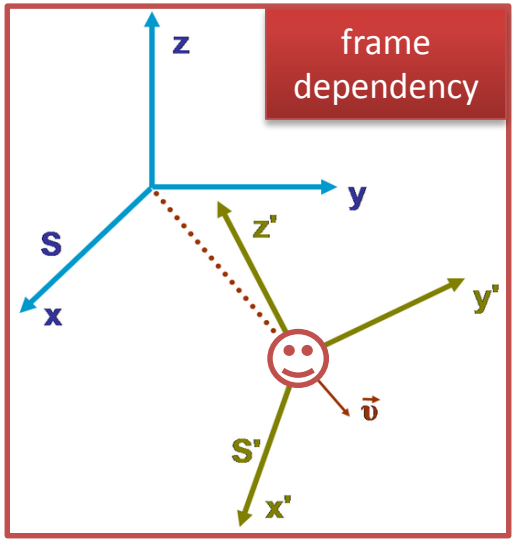
Linear Momentum

If an object is moving in any reference frame, then it has momentum in that frame.

mass

velocity

$$\mathbf{p} = m\mathbf{v} \Rightarrow \begin{cases} p_x = mv_x \\ p_y = mv_y \\ p_y = mv_y \end{cases}$$



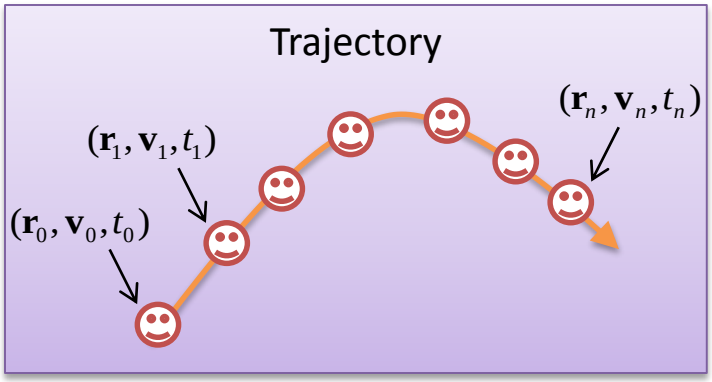
Equation of motion

motion under influence of a force and as a function of time

$$\mathbf{F} = \frac{d\,\mathbf{p}}{d\,t} = m\frac{d\,\mathbf{v}}{d\,t} = m\mathbf{a}$$

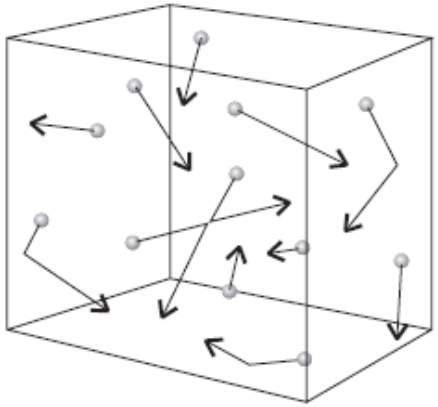
force

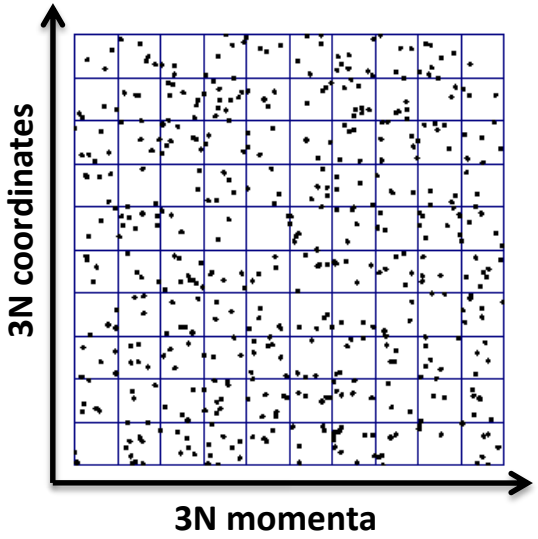
acceleration

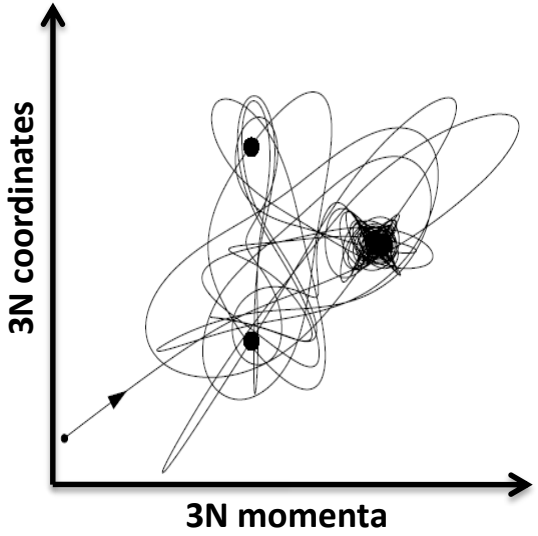


Phase space

The position and velocity variables define the *phase space* of a system.







$$\mathbf{r} = \{x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_N, y_N, z_N\}$$

$$\mathbf{p} = \{p_{1,x}, p_{1,y}, p_{1,z}, p_{2,x}, p_{2,y}, p_{2,z}, \ldots, p_{N,x}, p_{N,y}, p_{N,z}\}$$

Microscopic state of the system is a single point in phase space.

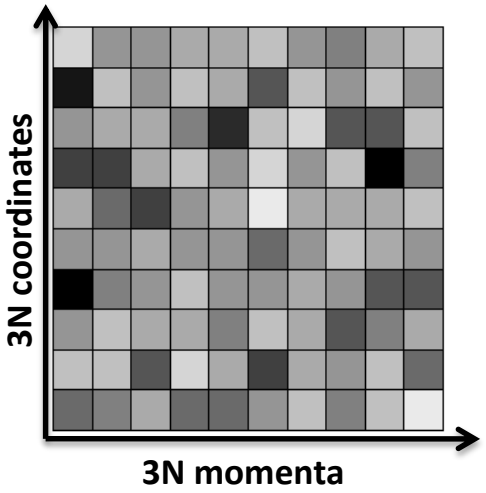
The path of the motion in phase space is called the trajectory.

The Statistical Mechanics Basis of MD

classical system
classical Hamiltonian
Total energy as a function of coordinates **r** and momenta **p**

$$H(\mathbf{r},\mathbf{p}) = K(\mathbf{p}) + U(\mathbf{r}) = \sum_i \frac{p_i^2}{2m_i} + U(\mathbf{r})$$

potential energy function independent of time and velocity.



Canonical Partition Function and Boltzmann distribution function
probability of finding the system at each and every point (microstate) in phase space.

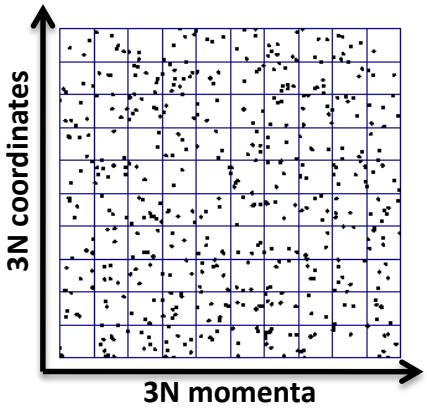
Continuous:

$$\rho(\mathbf{r},\mathbf{p}) = \frac{e^{-H(\mathbf{r},\mathbf{p})/k_B T}}{Q} \qquad Q = \int_V d\mathbf{r} \int_{-\infty}^{+\infty} d\mathbf{p}\; e^{-H(\mathbf{r},\mathbf{p})/k_B T}$$

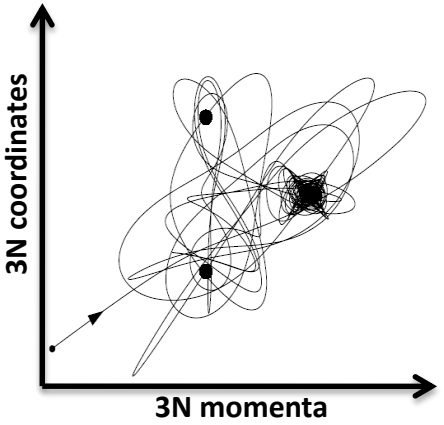
Discrete:

$$\rho_i = \frac{e^{-E_i/k_B T}}{Q} \qquad Q = \sum_i e^{-E_i/k_B T}$$

Averaging Microstates Over Phase Space



dynamic variables
any other function of **r** and/or **p**
(total energy, kinetic energy, fluctuations, …)



Thermodynamic Averages
(Ensemble Averages)

$$\langle A(\mathbf{r},\mathbf{p})\rangle_Q = \int d\mathbf{r} \int d\mathbf{p}\ \rho(\mathbf{r},\mathbf{p})\ A(\mathbf{r},\mathbf{p})$$

take into account every possible state of system.

Average over *all points* in phase space
at a *single time*.

Dynamic Averages
(Time Averages)

$$\langle A(\mathbf{r},\mathbf{p})\rangle_\tau = \frac{1}{\tau}\int_0^\tau dt\ A(\mathbf{r}(t),\mathbf{p}(t)) \qquad \langle A\rangle = \frac{1}{M}\sum_{i=1}^{M} A(\mathbf{r}_i,\mathbf{p}_i)$$

follow the motion of a single point through phase space.

Average over a *single point* in phase space
at *all times*.

Ergodic Hypothesis
for an infinitely
long trajectory

$$\lim_{\tau\to\infty}\langle A(\mathbf{r},\mathbf{p})\rangle_\tau = \langle A(\mathbf{r},\mathbf{p})\rangle_Q$$

Points of an enough long trajectory, will eventually cover all of phase space.

Calculation of Simple Thermodynamic Properties

Internal Energy

$$U = \langle E \rangle = \frac{1}{M} \sum_{i=1}^{M} E_i$$

Heat Capacity

multiple run evaluation

$$C_V = \left(\frac{\partial U}{\partial V} \right)_T$$

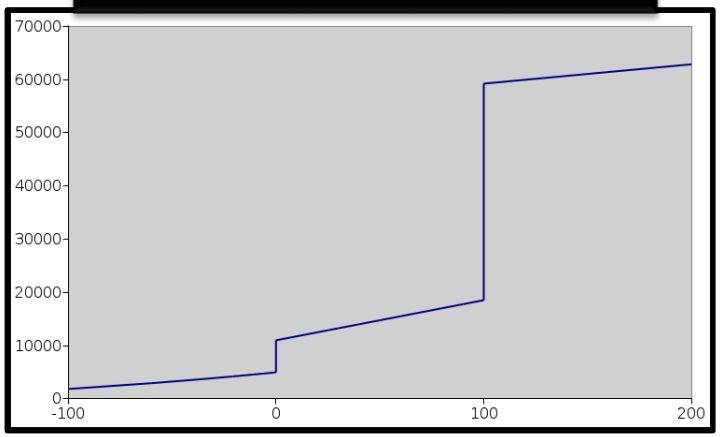
single run evaluation

$$C_V = \{ \langle E^2 \rangle - \langle E \rangle^2 \} / k_B T^2$$

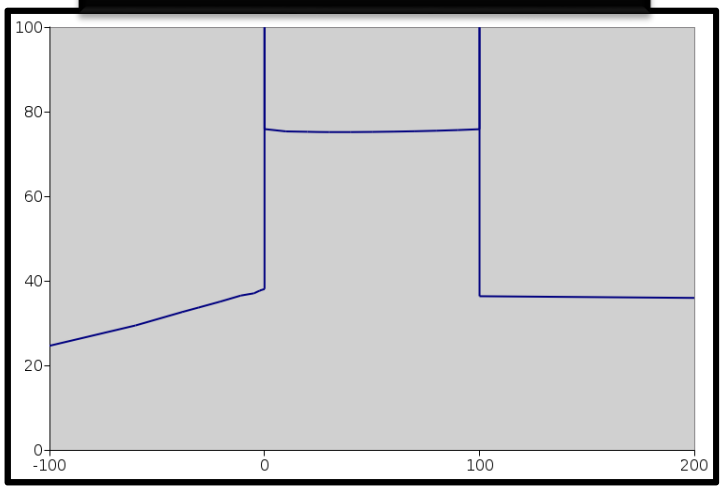
$$\langle (E - \langle E \rangle)^2 \rangle = \langle E^2 \rangle - \langle E \rangle^2$$

$$C_V = \langle (E - \langle E \rangle)^2 \rangle / k_B T^2$$

Internal Energy versus Temperature



Heat Capacity versus Temperature



Calculation of Simple Thermodynamic Properties

Pressure

virial theorem

$$W = \sum_{i=1}^{N} \mathbf{r}_i \cdot \mathbf{f}_i = -3Nk_BT$$

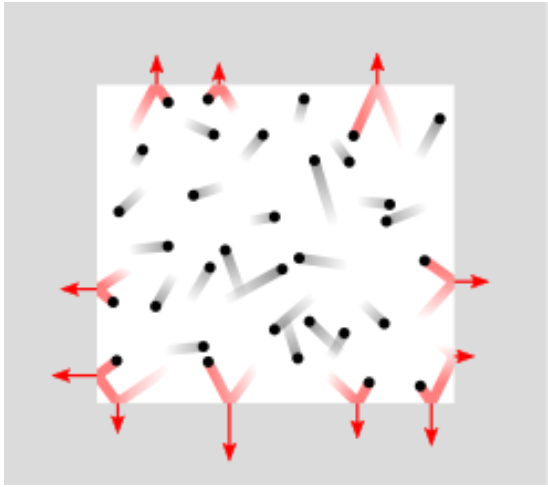
independent particles

$$W = -3PV = -3Nk_BT \qquad PV = Nk_BT$$

interacting particles

$$f_{ij} = \frac{d\,u(r_{ij})}{d\,r_{ij}}$$

$$W = \underbrace{-3PV}_{\text{interaction with container}} + \underbrace{\sum_{i=1}^{N} \sum_{j=i+1}^{N} r_{ij} f_{ij}}_{\text{Interaction with other particles}} = -3Nk_BT$$



Calculation of Simple Thermodynamic Properties

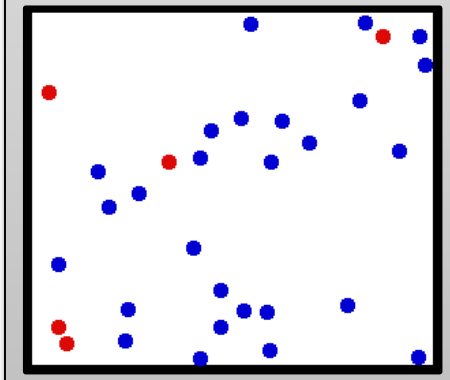
Theorem of Equipartition of Energy
In thermal equilibrium, energy is shared equally among all of its various forms.

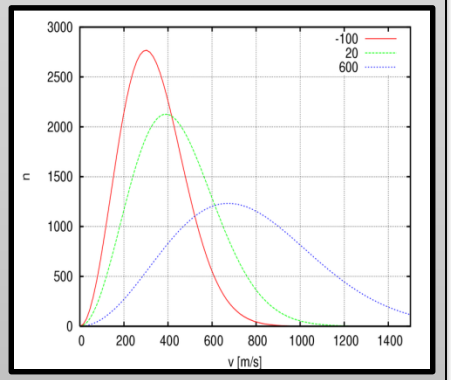
Each degree of freedom contributes $k_BT/2$.

Holds only for ergodic systems in thermal equilibrium.

All states with the same energy must be equally likely to be populated.

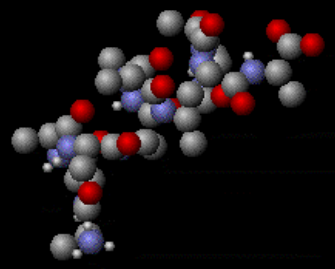
It must be possible to exchange energy among all its various forms.





Kinetic energy of a molecule can fluctuate wildly.
Average energy can be calculated at any temperature.

Thermal Motion
an α-helical peptide



Jittery motion is random and complex.
Energy of any atoms can fluctuate wildly.

Temperature

$$K(\mathbf{p}) = \sum_i \frac{p_i^2}{2m_i} = \frac{k_BT}{2}(3N - N_c)$$

N_c = number of constraints

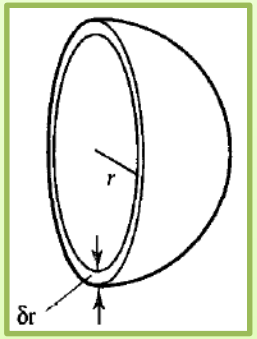
Total linear momentum of system is often constrained to zero (N_c=3).

Calculation of Simple Thermodynamic Properties

Radial Distribution Functions
(Pair Correlation Function)

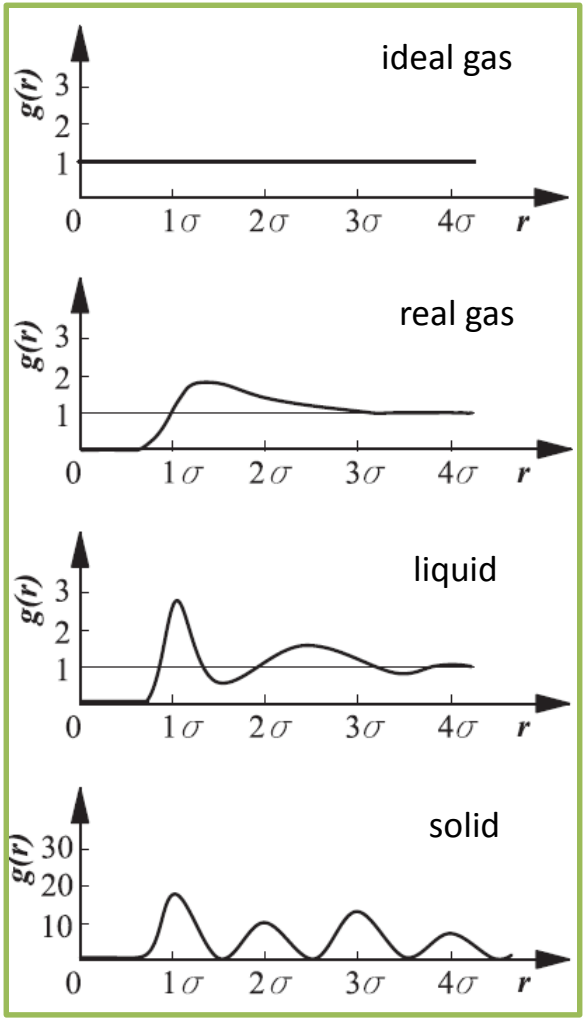
probability of finding a pair of atoms at a distance r, relative to probability expected for a completely random distribution at the same density.

Calculation Method:
neighbors around each particle are Sorted into distance 'bins'. The number of neighbors in each bin is then averaged over the entire simulation.



$$E = \tfrac{3}{2}Nk_BT + 2\pi N\rho \int_0^\infty r^2 v(r) g(r)\, dr$$

$$PV = Nk_BT - \frac{2\pi N\rho}{3k_BT} \int_0^\infty r^2 r \frac{dv(r)}{dr} g(r)\, dr$$



Newton's Equation of Motion

simplistic form

second-order differential equation

$$F_i = m_i a_i = m_i \ddot{r}_i$$

The force is determined by the gradient of the potential energy function.

$$F_i = -\nabla_i U(\mathbf{r})$$

general formulation

a pair of coupled first order differential equations

$$\dot{r}_k = \frac{\partial H(\mathbf{r}, \mathbf{p})}{\partial p_k}; \qquad \dot{p}_k = -\frac{\partial H(\mathbf{r}, \mathbf{p})}{\partial r_k}$$

A set of two first-order differential equations is often easier to solve than a single second-order differential equation.

Properties of Newton's Equation of Motion

Serve as ''handles'' to ensure that the numerical solution is correct

Conservation of energy:
Assuming that **U** and **H** do not depend explicitly on time or velocity.

Hamiltonian is a constant of motion or there is conservation of total energy

$$\partial H/\partial t = 0 \rightarrow dH/dt \text{ is zero}$$

Time reversibility

Changing the signs of all velocities will cause the molecule to retrace its trajectory.

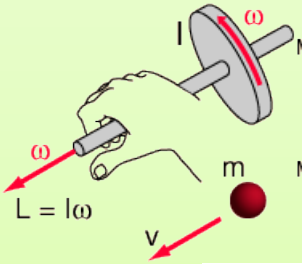
Time reversibility can be reproduced by numerical trajectories only over very short periods of time because of the chaotic nature of large molecular systems.

Conservation of linear and angular momentum

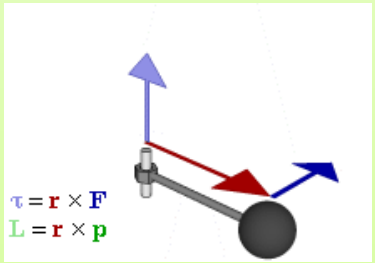
If the potential function U depends only on particle separation and there is no external field applied

$$\boldsymbol{P} = \sum_i p_i$$

$$\boldsymbol{L} = \sum_i r_i \times p_i = \sum_i m_i r_i \times \dot{r}_i$$



Angular Momentum	=	Moment of Inertia	x	Angular Velocity
L	=	I	x	ω
Linear Momentum	=	Mass	x	Velocity
p	=	m	x	v



Molecular Dynamics: Computational Algorithms

numerical procedures for integrating the differential equation

finite-difference approach
standard method for solving ordinary differential equations

Coordinates and velocities at a time **t+Δt** are obtained from coordinates and velocities at an earlier time **t**.

The equations are solved on a step-by-step basis.

choice of time interval

Δt depends on the properties of the molecular system simulated.

Δt must be significantly smaller than the characteristic time of the motion studied.

Taylor expansion of position
at time **t+Δt** about time **t**

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \dot{\mathbf{r}}(t)\Delta t + \frac{1}{2}\ddot{\mathbf{r}}(t)\Delta t^2 + \cdots$$

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{1}{2}\mathbf{a}(t)\Delta t^2 + \cdots$$

Rewrite expansion in a discrete form

$$\mathbf{r}_{n+1} = \mathbf{r}_n + \mathbf{v}_n\Delta t + \frac{1}{2}\left(\frac{\mathbf{F}_n}{\mathbf{m}}\right)\Delta t^2 + O(\Delta t^3)$$

$$\mathbf{v}_{n+1} = (\mathbf{r}_{n+1} - \mathbf{r}_n)/2$$

The formulation is highly trivial and results in a low quality integration algorithm (large errors).

Computational Algorithms: Verlet Integrator

The most common integration algorithm used in the study of biomolecules.

Two Taylor expansions:

Forward expansion (t+Δt):

$$\mathbf{r}_{n+1} = \mathbf{r}_n + \mathbf{v}_n \Delta t + \frac{1}{2}\left(\frac{\mathbf{F}_n}{\mathbf{m}}\right)\Delta t^2 + O(\Delta t^3)$$

Backward expansion (t-Δt):

$$\mathbf{r}_{n-1} = \mathbf{r}_n - \mathbf{v}_n \Delta t + \frac{1}{2}\left(\frac{\mathbf{F}_n}{\mathbf{m}}\right)\Delta t^2 - O(\Delta t^3)$$

Propagating the positions:

$$\mathbf{r}_{n+1} = 2\mathbf{r}_n - \mathbf{r}_{n-1} + \frac{\mathbf{F}_n}{\mathbf{m}}\Delta t^2 + O(\Delta t^4)$$

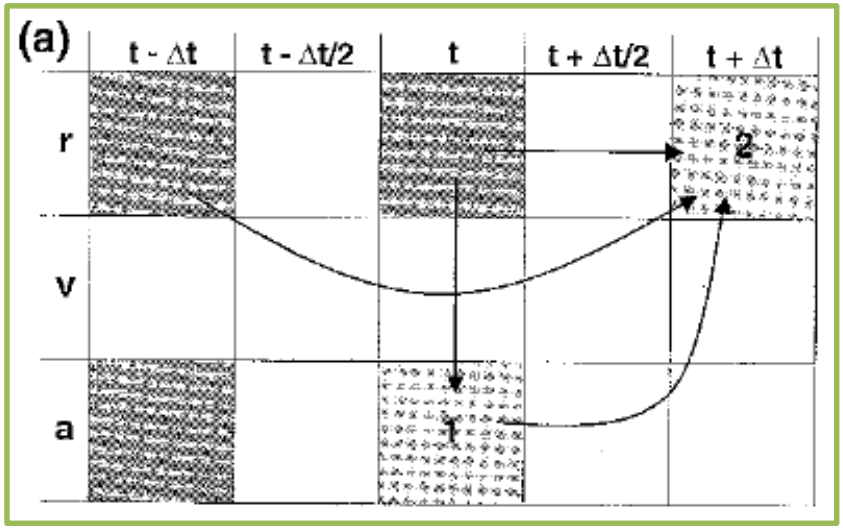
Propagating the velocities:

$$\mathbf{v}_n = \frac{\mathbf{r}_{n+1} - \mathbf{r}_{n+1}}{2\Delta t} + O(\Delta t^2)$$

execution instructions:

1) Use current position $\mathbf{r}_n$ to calculate current force $\mathbf{F}_n$.

2) Use current and previous positions $\mathbf{r}_n$ and $\mathbf{r}_{n-1}$ and current force $\mathbf{F}_n$ to calculate position in next step, $\mathbf{r}_{n+1}$.



A stable numerical method for solving Newton's equation of motion for systems ranging from simple fluids to biopolymers.

Computational Algorithms: Verlet Integrator

Advantages of the Verlet algorithm:

1) Position integration is quite accurate and independent of velocity propagation.

2) The algorithm requires only a single force evaluation per integration cycle.

3) Forward and backward expansions, guarantees time reversibility.

Disadvantages of the Verlet algorithm:

1) The velocity propagation is subject to relatively large errors.

2) $\mathbf{v}_n$ can be computed only if $\mathbf{r}_{n+1}$ is already known.

3) Algorithm is not ‘‘self-starting’’ and need a lower order expansion to initiate.

4) It must be modified to incorporate velocity-dependent forces or temperature scaling.

Computational Algorithms: Leap-Frog Integrator

Velocities are evaluated at the midpoint of the position evaluation and vice versa

$$\mathbf{r}_{n+1} = \mathbf{r}_n + \mathbf{v}_{n+1/2}\Delta t$$

$$\mathbf{v}_{n+1/2} = \mathbf{v}_{n-1/2} + \frac{\mathbf{F}_n}{\mathbf{m}}\Delta t$$

The current velocity $\mathbf{v}_n$

$$\mathbf{v}_n = (\mathbf{v}_{n+1/2} + \mathbf{v}_{n-1/2})/2$$

execution instructions:

1) Use current position $\mathbf{r}_n$ to calculate current force $\mathbf{F}_n$.

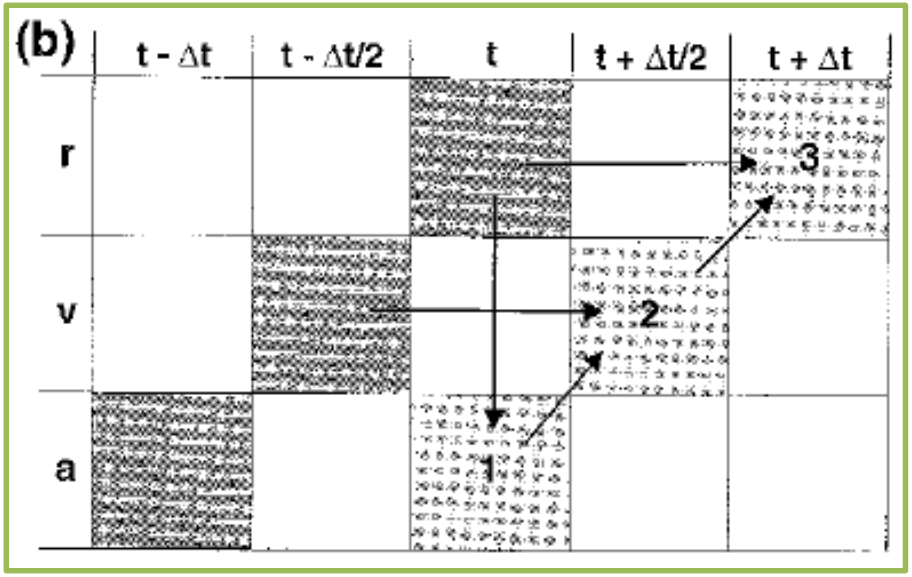
2) Use current force $\mathbf{F}_n$ and previous mid-step velocity $\mathbf{v}_{n-1/2}$ to calculate next mid-step velocity $\mathbf{v}_{n+1/2}$.

3) Use current position $\mathbf{r}_n$ and next mid-step velocity $\mathbf{v}_{n+1/2}$ to calculate position in next step, $\mathbf{r}_{n+1}$.

Advantages of the leap-frog algorithm:

1) Direct and more accurate evaluation of velocity.

2) Improved control of temperature via velocity scaling.



This algorithm is computationally a little more expensive than the Verlet algorithm.

Computational Algorithms: Velocity Verlet Integrator

Store positions, velocities, and accelerations all at the same time **t** and minimizes round-off errors

$$\mathbf{r}_{n+1} = \mathbf{r}_n + \mathbf{v}_n \Delta t + \frac{1}{2}\left(\frac{\mathbf{F}_n}{\mathbf{m}}\right)\Delta t^2$$

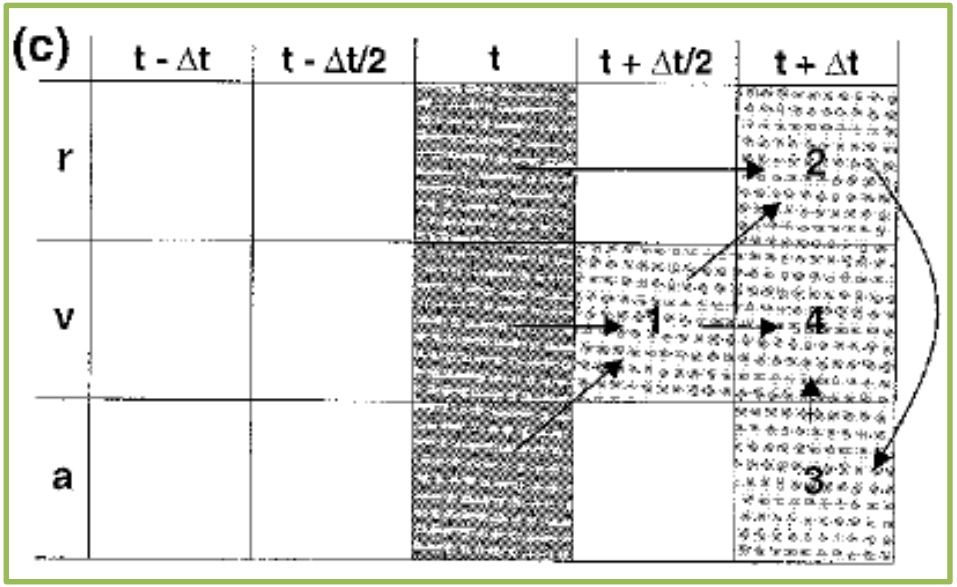
$$\mathbf{v}_{n+1} = \mathbf{v}_n + \frac{1}{2}\left[\frac{\mathbf{F}_n}{\mathbf{m}} + \frac{\mathbf{F}_{n+1}}{\mathbf{m}}\right]\Delta t$$

$$\mathbf{v}_{n+1/2} = \mathbf{v}_n + \frac{1}{2}\left(\frac{\mathbf{F}_n}{\mathbf{m}}\right)\Delta t$$

$$\mathbf{v}_{n+1} = \mathbf{v}_{n+1/2} + \frac{1}{2}\left(\frac{\mathbf{F}_{n+1}}{\mathbf{m}}\right)\Delta t$$

execution instructions:

1) Calculate position $\mathbf{r}_{n+1}$ at time t+Δt.
2) Calculate velocity at mid-step $\mathbf{v}_{n+1/2}$.
3) Calculate force $\mathbf{F}_{n+1}$ at time t+Δt.
4) Complete velocity move to $\mathbf{v}_n$ by using $\mathbf{v}_{n+1/2}$.

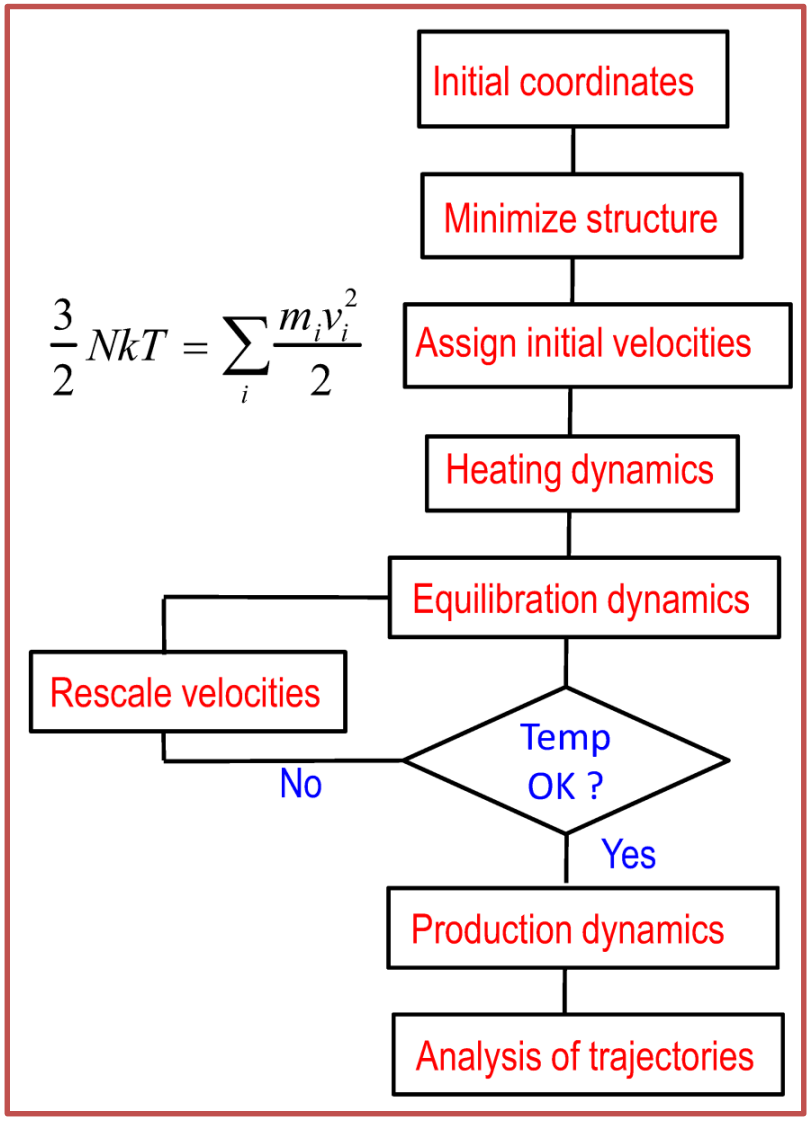


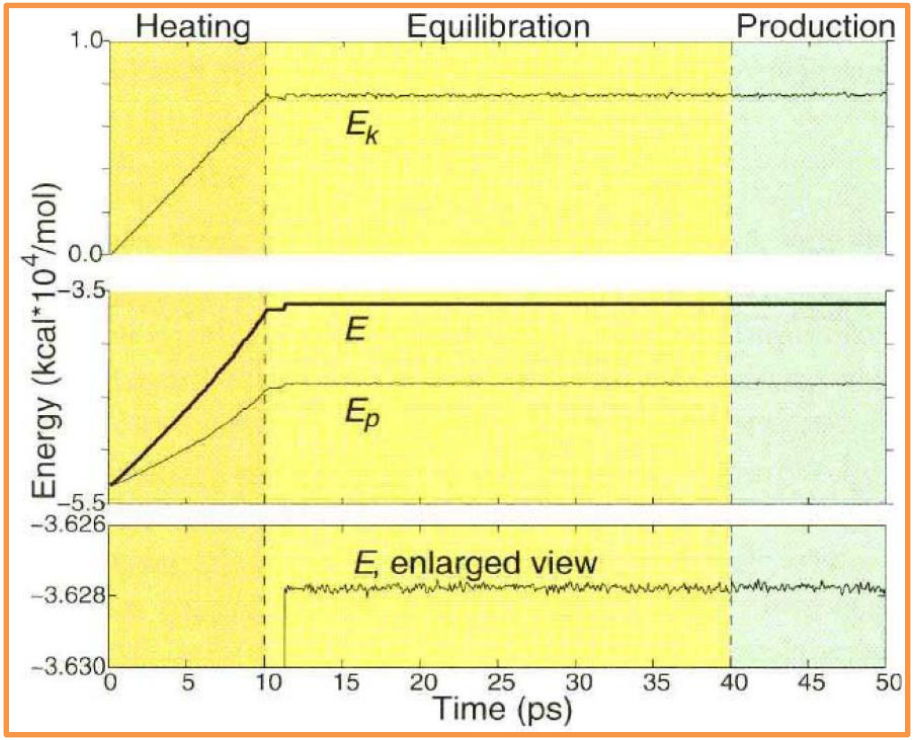
Advantages of velocity Verlet algorithm:

1) numerically very stable.
2) convenient and simple to code.
3) Accurate velocities and kinetic energy.

Computationally a little more expensive than the simpler Verlet or leap-frog algorithms

Classical MD Simulation Protocol





$$V_{\text{total}} = \sum_{\text{bonds}} K_b (r - r_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_\phi [1 + \cos(n\phi - \gamma)]$$

$$+ \sum_{\substack{\text{van der Waals} \\ i,\ j \text{ pairs}}} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + \sum_{\substack{\text{electrostatic} \\ i,\ j \text{ pairs}}} \frac{q_i q_j}{\varepsilon r_{ij}}$$

The most time demanding part

Assigning Initial Coordinates

Usual Source:
Experimentally determined structures
(X-ray crystallography and NMR)

Possible Source:
Structures based on computer models
(A variety of modeling techniques)

Experimental structures need some preparation steps before use.

Initial refinement by energy minimization:

to relieve local stresses due to non-bonded overlaps.

to relax bond length and bond angle distortions in experimental structure.

stresses are due to empirical nature of energy function and to the average nature of experimentally determined structures.

Hydrogen atom positions:

can not be determined by X-ray crystallography.

must be added separately to initial structure before MD is started.

Missing parts of protein:

flexible parts of molecule that do not have a well-defined structure (such as loops).

parts of molecule that were removed to facilitate crystallization process (such as terminal sequences).

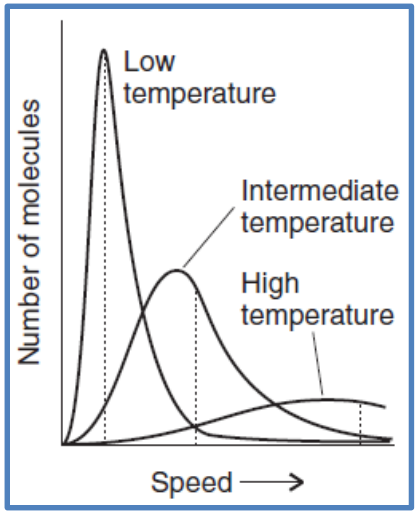
structural models may be used to fill in the gaps.

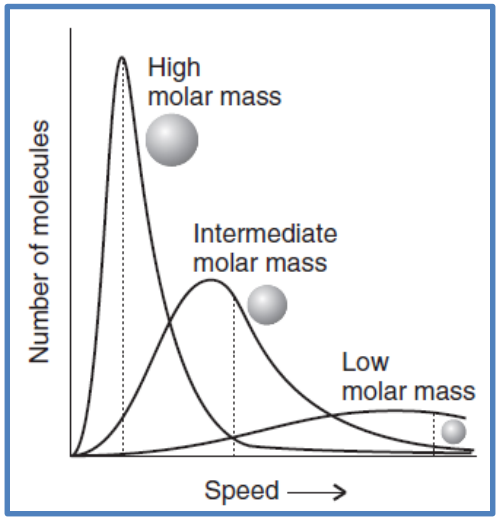
Assigning Initial Velocities

The only relevant information about atomic velocities is temperature T.

Maxwell Distribution
Initial velocities are randomly assigned from standard Maxwell velocity distribution.

$$P(v)dv = \left(\frac{m}{2\pi k_B T}\right)^{1/2} \exp\left[\frac{-mv^2}{2k_B T}\right] dv$$





"Hot Spot" Problem:
Random assignment may accidentally assign high velocities to a localized cluster of atoms that make the simulation unstable.

Equilibrium Problem:
Initial assignment is not at equilibrium.

Velocity Correlation Problem:
Expected velocity correlation between neighboring atoms is not guaranteed.

Non Zero Total Momenta Problem:
Large initial total linear momentum **P** and total angular momentum **L**.
Irrelevant translational drift and global rotation.

Assigning Initial Velocities

Gradual heat-up:
Velocities are initially assigned at a low temperature, which is then increased gradually allowing for dynamic relaxation.

$$T(t) = \frac{1}{k_B N_{\text{dof}}} \sum_{i=1}^{N_{\text{dof}}} m_i |v_i|^2$$

N_{dof} is the number of unconstrained degrees of freedom in the system

Heating by increasing atomic velocities:

1) reassigning new velocities from a Maxwell distribution at an elevated temperature.

2) scaling the velocities by a uniform factor

scaling the velocities by a factor of $[T_0/T(t)]^{1/2}$ will result in a mean kinetic energy corresponding to a desired temperature T_0

Importance of Boundary Effects and Boundary Conditions

Boundary effect on real bulk?
1 liter water at room temperature

A cube of approximately
3.3 x 10^{25} molecules.

Interactions with walls can extend
up to **10** molecular diameters.

Diameter of water molecule
is approximately **2.8 Å**.

About **2 x 10^{19}** molecules are
interacting with boundary.

About **one in 1.5 million** molecules is
influenced by interactions with
the walls of the container.

Finite size effects?

a system of 1000 water molecules

molecular dynamics simulation to
derive **'bulk'** properties.

Most molecules would be within the
influence of walls of boundary.

simulation in a vessel is not an appropriate
way to derive **'bulk'** properties.

If ignore the container?

3/4 of molecules are at the surface rather
than in bulk.

Such a situation would be relevant to
studies of liquid drops.

Periodic Boundary Conditions

Simulate a small number of particles, while they experience forces as if they were in bulk fluid.

A box of particles which is replicated in all directions in a periodic array.

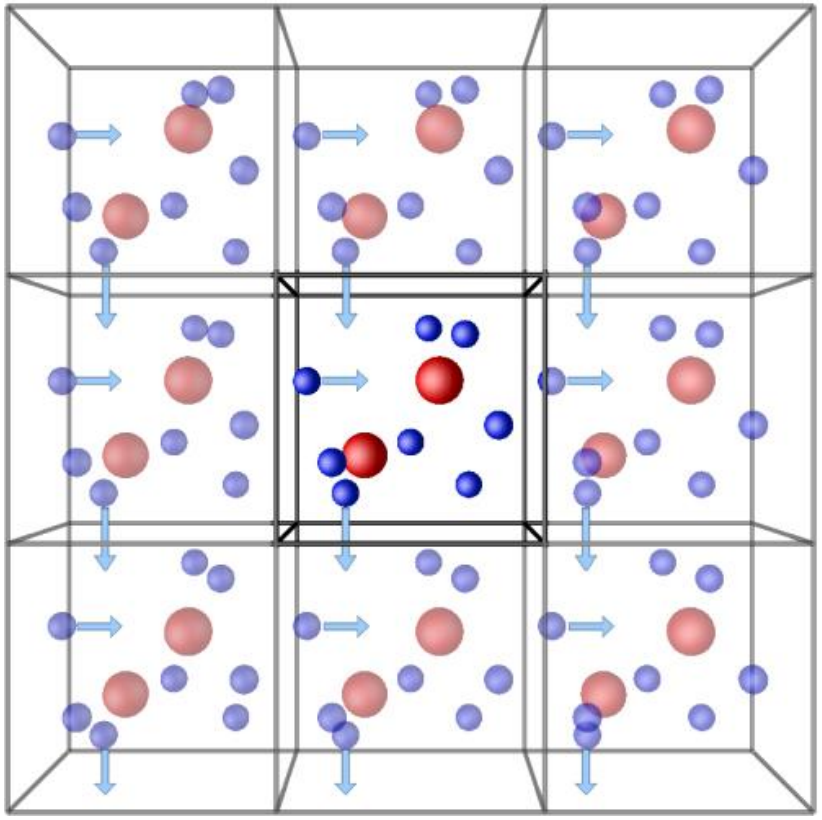
In three dimensions each cubic box would have 26 nearest neighbors.

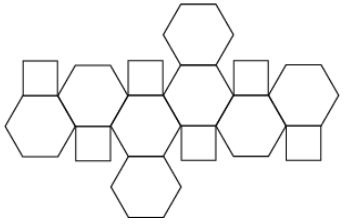
leaving particle is replaced by an image particle from the opposite side.

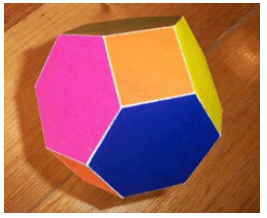
Number of particles within central box remains constant.

Any cell shape can be used if it fills all of space by translation operations.

Choose a periodic cell that reflects underlying geometry of system.



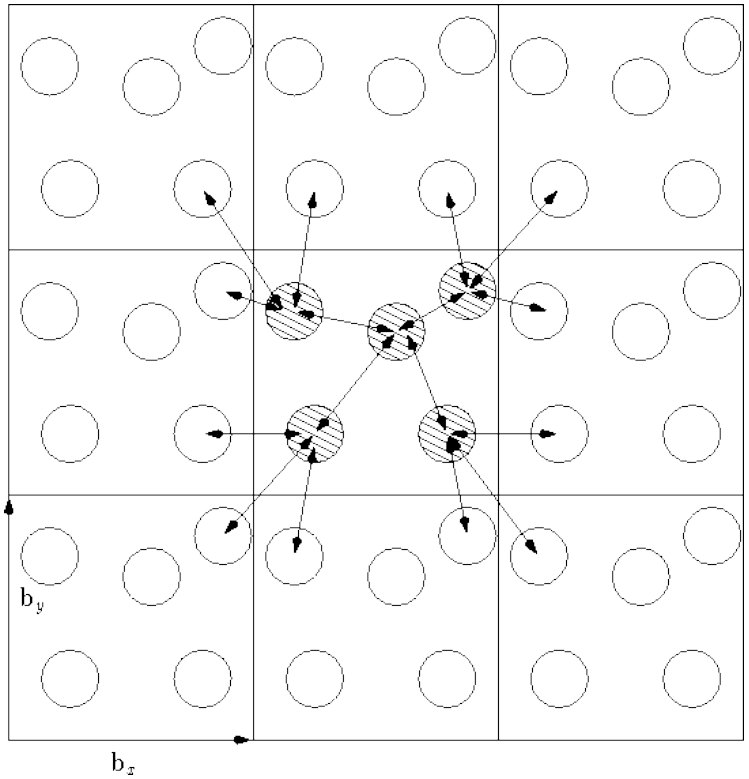




Minimum Image Convention

Bonded interactions
Bonded interactions are local.
Linear computational cost: O(N).
(N = number of atoms)

Non-Bonded interactions
Quadratic cost: $O(N^2)$.
Prohibitive for large molecules.
LJ potential falls off very rapidly
(**99%** decay from **σ** to **2.5σ**)
In principle, must be calculated between every pair of atoms in system.
In practice, use *non-bonded cutoff* and apply *minimum image convention*.



Each atom is repeated infinitely via periodic boundary method.
Each atom 'sees' at most just one image of every other atom.
Energy and/or force is calculated with closest atom or image.

$$U_{NB} = \sum_{i,j\ nonbonded} \varepsilon_{ij}\left[\left(\frac{R_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{i,j\ nonbonded} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}}$$

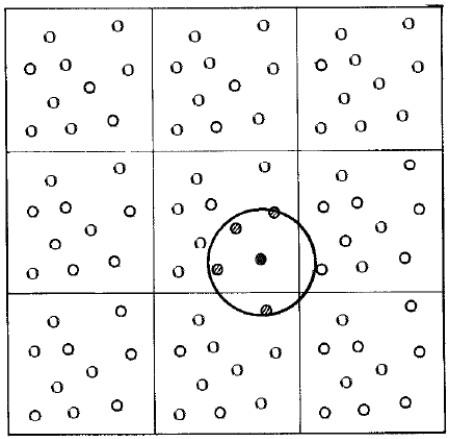
Non-Bonded Cutoff

$$U_{NB} = \sum_{i,j} \omega_{ij} S(r_{ij}) \, \varepsilon_{ij} \left[\left(\frac{R_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}}{r_{ij}} \right)^{6} \right] + \sum_{i,j} \omega_{ij} S(r_{ij}) \frac{q_i q_j}{r_{ij}}$$

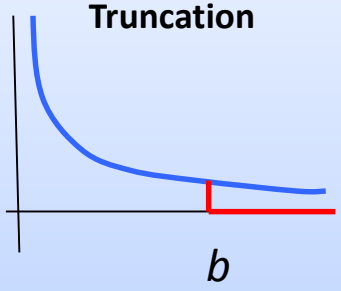
Weights: 0< ω_{ij} <1

To exclude bonded terms or to scale them (usually 1-4)

Cutoff function: *S*(*r*)
To truncate, switch or shift interaction energy

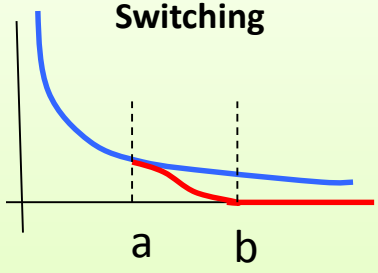


Truncation



$$S(r) = \begin{cases} 1 & r < b \\ 0 & r \geq b \end{cases}$$

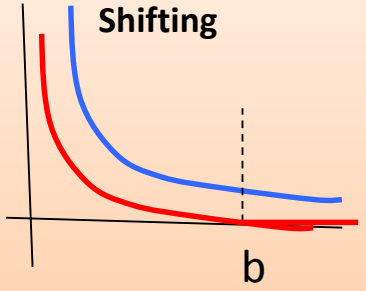
Switching



$$S(r) = \begin{cases} 1 & r < a \\ 1 + y(r)^2 [2y(r) - 3] & a \leq r \leq b \\ 0 & r > b \end{cases}$$

$$y(r) = \frac{r^2 - a^2}{b^2 - a^2}$$

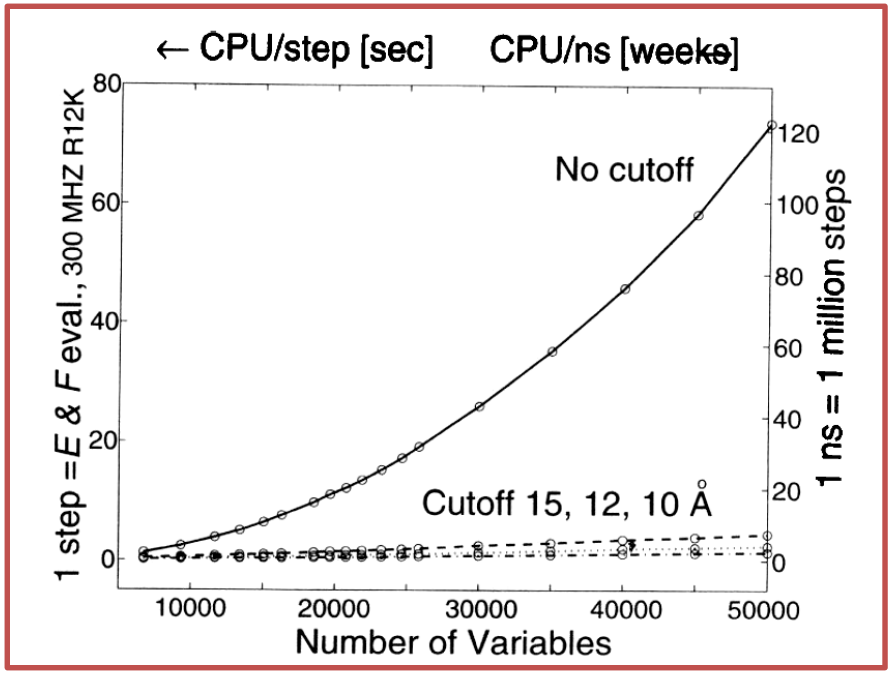
Shifting



$$S_1(r) = \left[1 - \left(\frac{r}{b} \right)^2 \right]^2 \quad r \leq b$$

$$S_2(r) = \left[1 - \frac{r}{b} \right]^2 \quad r > b$$

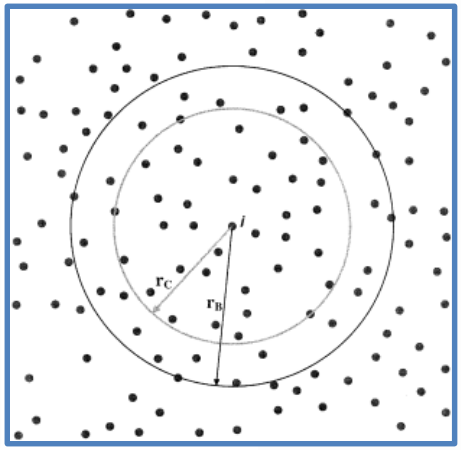
Non-Bonded Cutoff

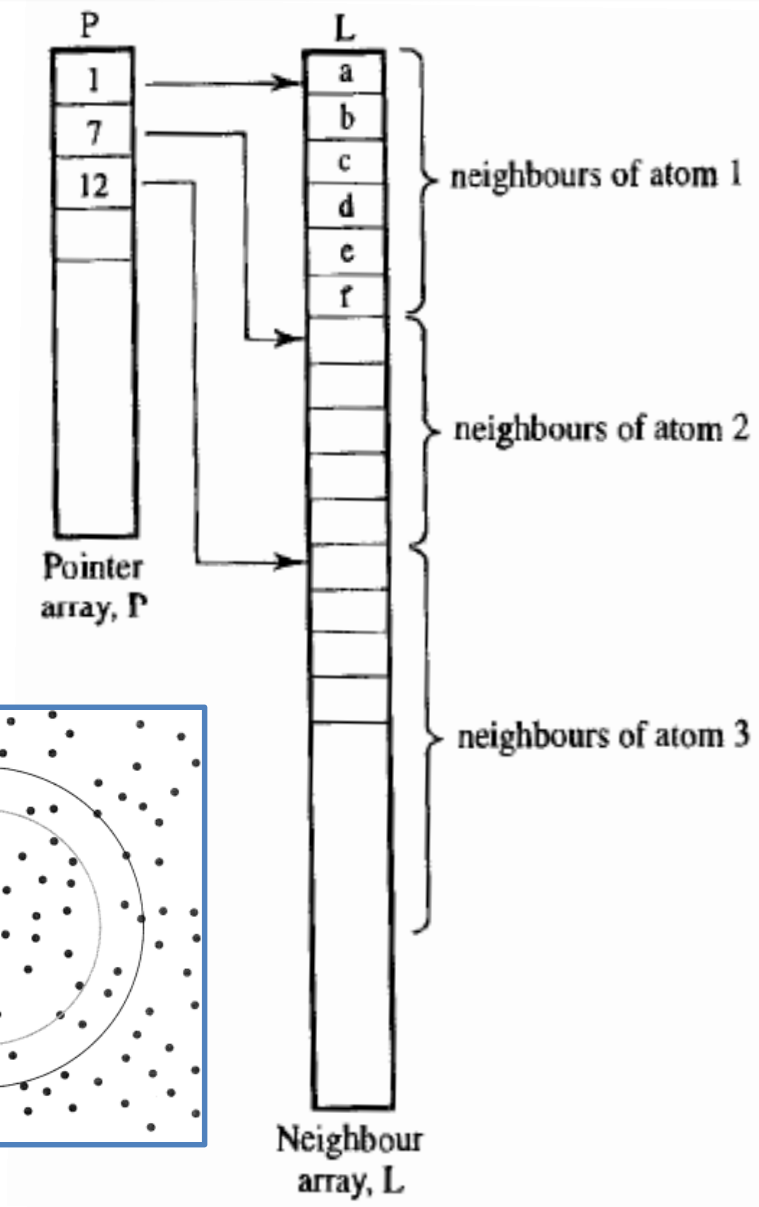


non-bonded nearest neighbor list

An atom's neighbors do not change significantly over 10 or 20 time steps.

Store atoms within cutoff distance, together with atoms slightly further away than cutoff distance.





Some issues about Periodic Boundaries

Validation
Effects of periodic boundary can be checked by comparing a variety of cell shapes and sizes.

Coverage of large-scale fluctuations

It is not possible to achieve fluctuations that have a wavelength greater than length of cell.
Example: near the liquid-gas critical point.

Wavelength of sound or shock waves and phonons in system is limited by box size.

Coverage of long-range interactions:

The range over which interactions act should be smaller than the cell size.
Lennard-Jones is relatively short-range and cell should have a side greater than **6σ**.
Argon Example: cell size around 20 Å.

Electrostatic forces act on longer ranges and should be handled in a separate manner.
In presence of ionic interactions, net charge of system must be zero.
Neutrality can be obtained by adding counterions such as sodium or chloride.

Some issues about Periodic Boundaries

Maintenance of minimum-image convention

Spherical cutoff radius for non-bonded forces should be at most half the length of box side.
A particle should not see its own image, or the same molecule twice.

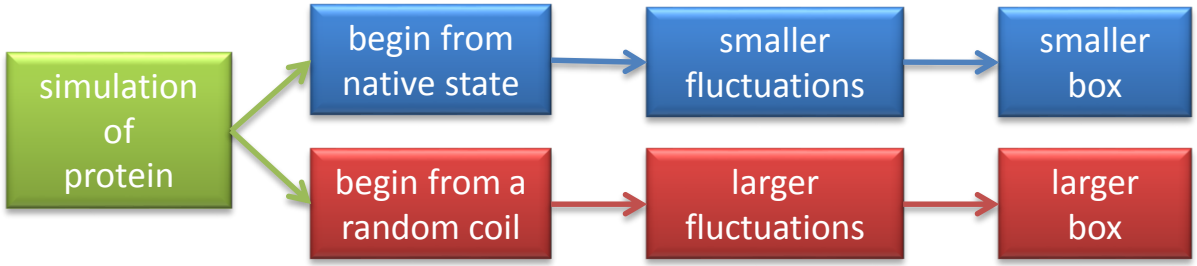
Periodic artifacts due to unphysical topology

A macromolecule should not interact with its own image in a neighboring box.
It is functionally equivalent to a molecule's "head" interacting with its own "tail".

This produces highly unphysical dynamics in most macromolecules.

Effects of solvation shells on the observed dynamics are not well understood.

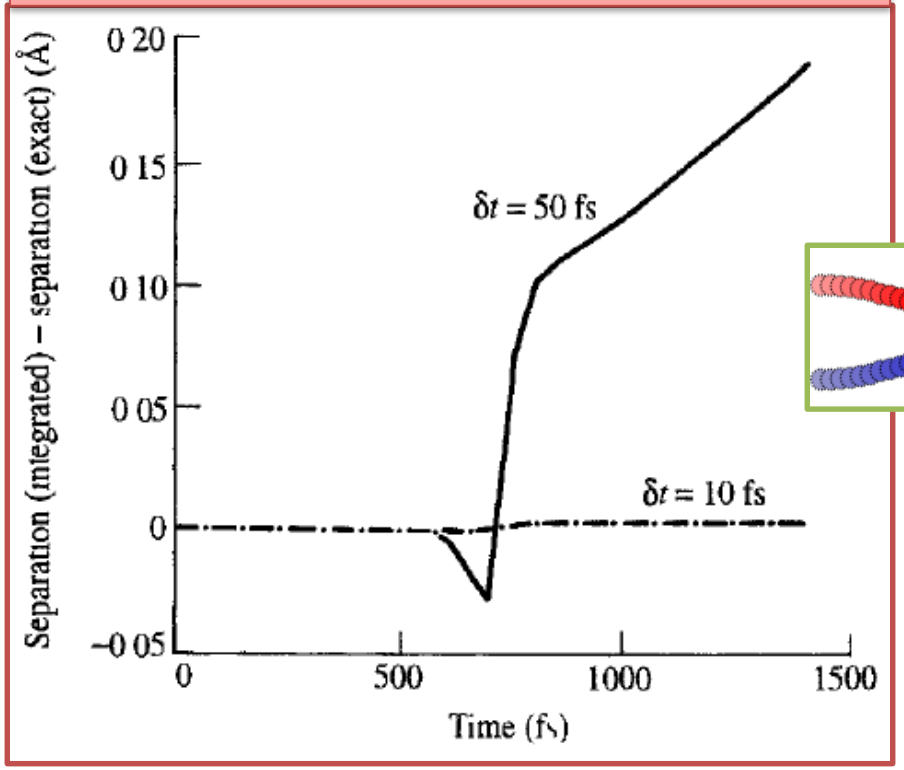
Common recommendation based on simulations of DNA: at least 1 nm of solvent around molecules of interest in every dimension.



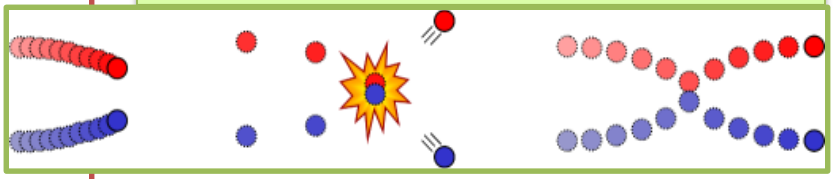
Time Step

- Not too short so that conformations are efficiently sampled
- Not too long to prevent wild fluctuations or system 'blow-up'
- An order of magnitude less than the fastest motion is ideal
  - Usually bond stretching is the fastest motion: C-H is ~10 fs so use time step of **1 fs**
  - Not interested in these motions? Constrain these bonds and double the time step

Difference between exact and numerical trajectories for approach of two argon atoms



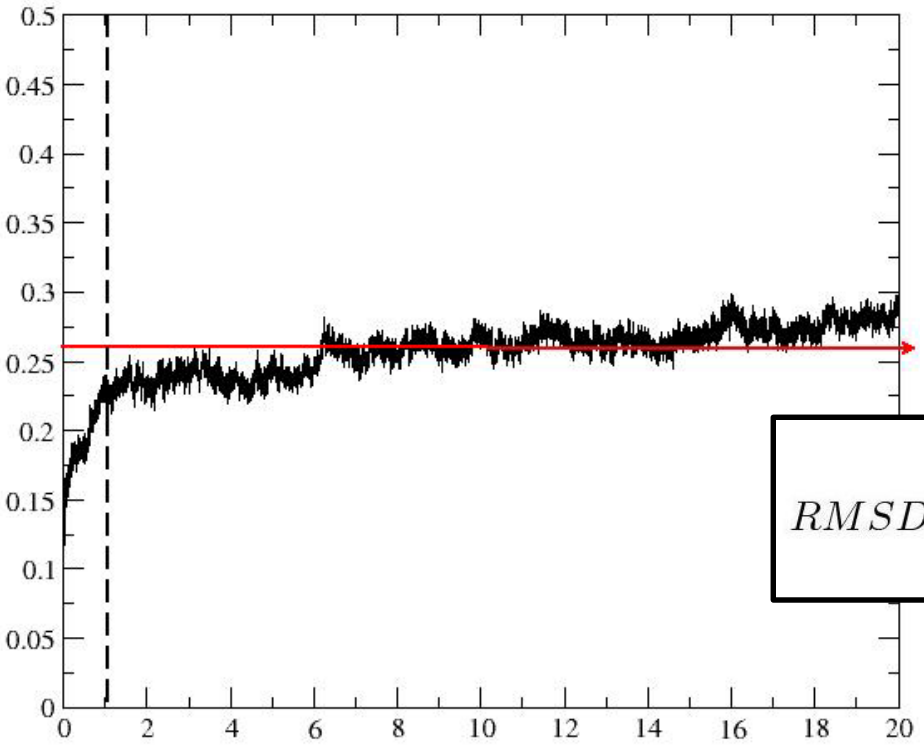
With the appropriate time step collisions occur smoothly

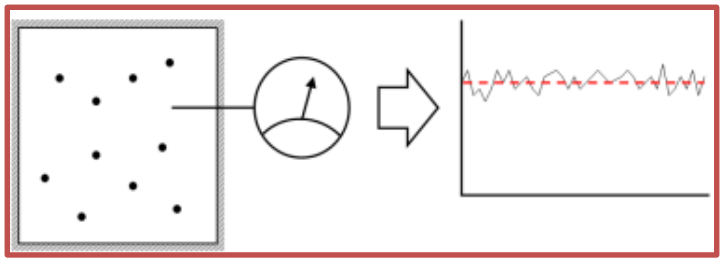


Monitoring the Equilibration

- The system is at equilibrium if:
  - Quantities fluctuate around an average value.
  - The average remains constant over time.

- Variables to monitor:
  - Structural properties (RMSD, order parameters…)
  - Thermodynamics quantities (Potential Energy…)





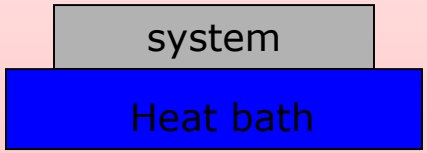
$$RMSD(t_1, t_2) = \left[\frac{1}{M} \sum_{i=1}^{N} m_i \|\mathbf{r}_i(t_1) - \mathbf{r}_i(t_2)\|^2 \right]^{\frac{1}{2}}$$

Molecular Dynamics Ensembles

Micro-Canonical ensemble:
constant N, V and E

To simulate under constant T or P:
Canonical Ensemble: NVT
Isothermal-Isobaric: NPT

Simulating at constant T:
Berendsen scheme

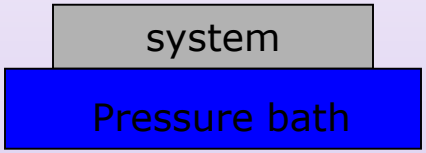


Exponentially scale the velocities at each time step by a factor λ

$$\lambda = \sqrt{1 - \frac{\Delta t}{\tau}\left(1 - \frac{T_{bath}}{T}\right)}$$

τ determines how strong the bath influences the system.

Simulating at constant T:
Berendsen scheme

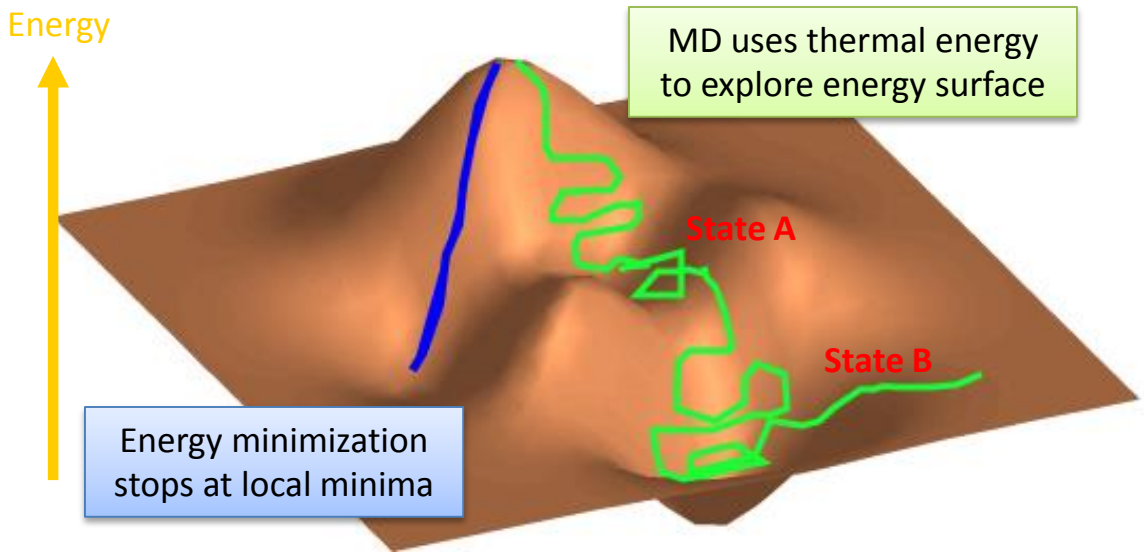


Exponentially scale volume of simulation box at each time step by a factor λ

$$\lambda = 1 - \kappa \frac{\Delta t}{\tau_P}\left(P - P_{bath}\right)$$

κ : isothermal compressibility
τ_P : coupling constant

MD As a Tool for Minimization: Crossing energy barriers

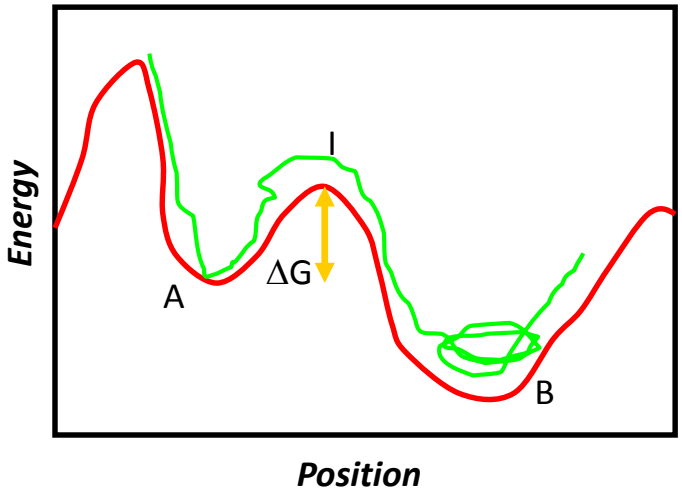


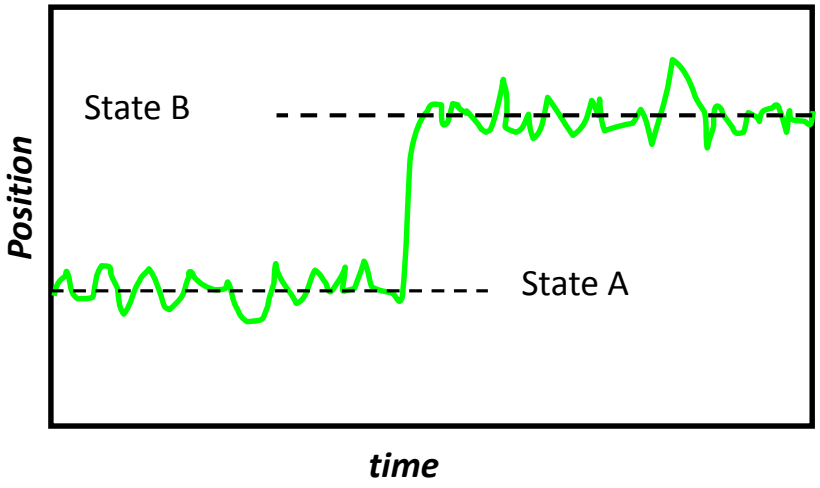
The actual transition time from A to B is very quick.

What takes time is **waiting**.

Average waiting time for going from A to B:

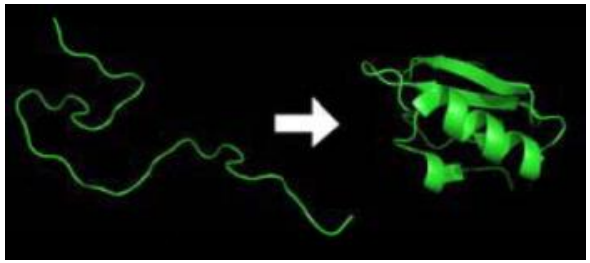
$$\tau_{A\rightarrow B} = Ce^{\frac{\Delta G}{kT}}$$



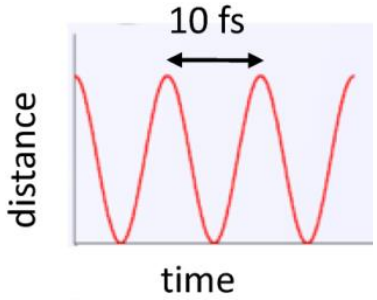


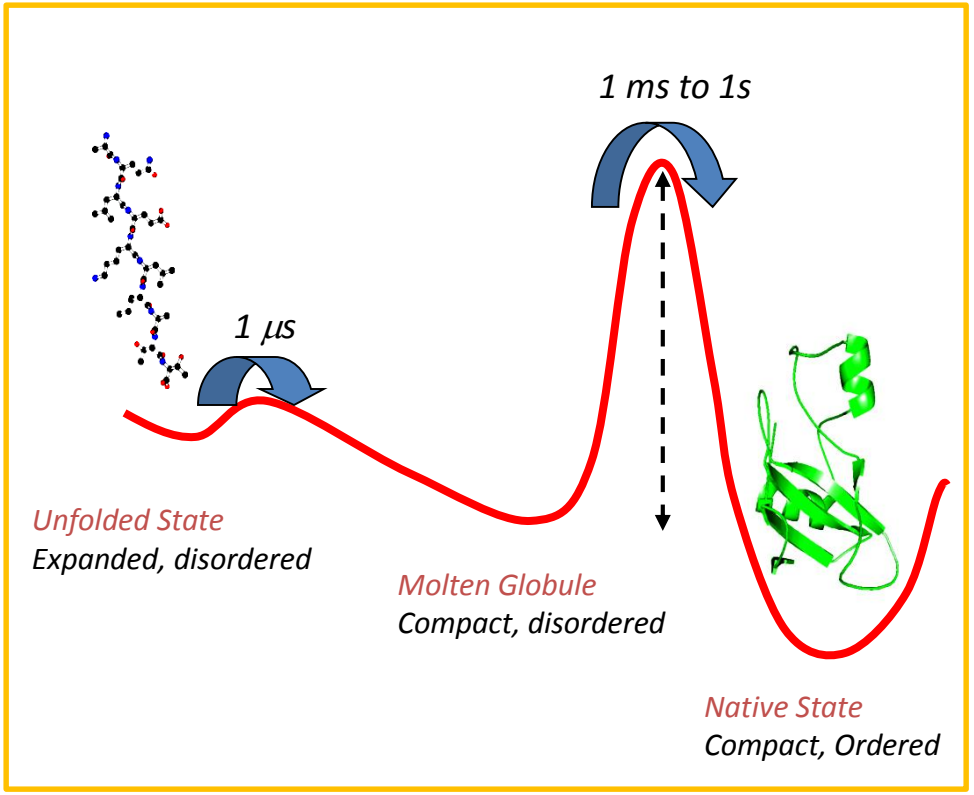
Some Practical Aspects of Biomolecular MD

(1) Slow motions often occur on timescales that not easily accessible and thus are not well characterized with simulations.



(2) The optimum time step depends on the physics of the system. In particular, it depends on the fastest motion.



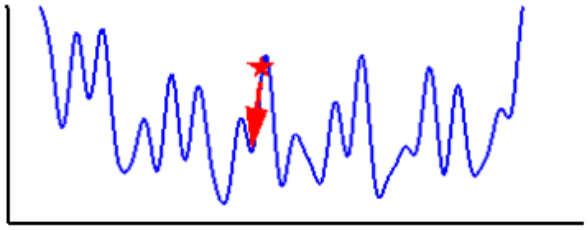


In most biological problems, the fastest motion is the C–H stretching and the slowest is protein folding.

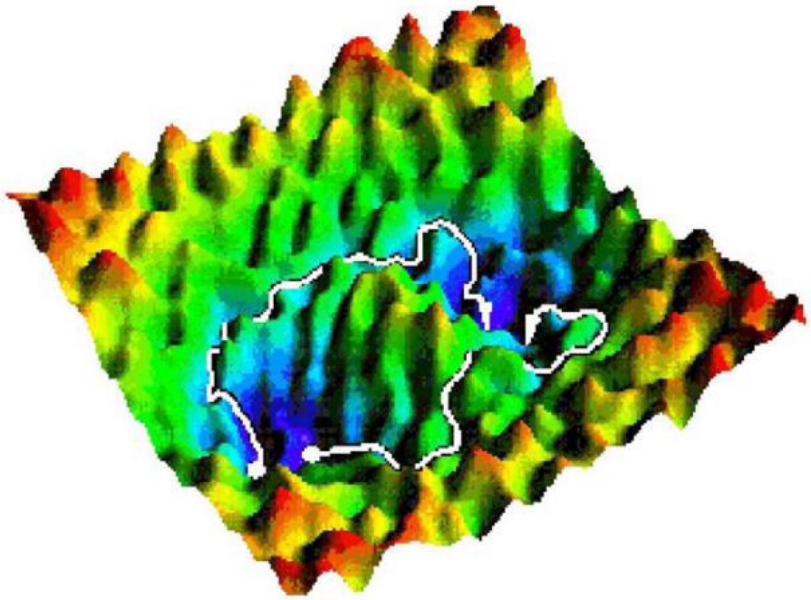
Some Practical Aspects of Biomolecular MD

(3) Convergence Problem:
How long the simulation should be to extract reliable information?

(4) Sampling of minima is essential for the conformational flexibility, and therefore, to function of biomolecules.



(5) Simulation can get stuck in one of the many minima on the energy landscape, and not allow for complete sampling of conformational space.



(6) The ergodicity should be checked:
At equilibrium, independent trajectories over an ergodic system must be self-averaging.

(7) Multiple shorter trajectories have been found to be more effective at sampling phase space than a single long one.

Some Practical Aspects of Biomolecular MD

(8) Force fields must be parameterized specifically for compounds that one is attempting to model.

Example From CHARMM:

2D dihedral energy correction map to the CHARMM 22 ϕ,ψ backbone.

$$E_{CHARMM}^{protein} = E_{bonds} + E_{UB} + E_{angle} + E_{dihe} + E_{CMAP} + E_{imp} + E_{vdw} + E_{elec}$$

$$E_{CMAP} = f(\Phi,\Psi) = \sum_{i=1}^{4}\sum_{j=1}^{4} c_{ij} \left(\frac{\Phi - \Phi_L}{\Delta_\Phi}\right)^{i-1} \left(\frac{\Psi - \Psi_L}{\Delta_\Psi}\right)^{j-1}$$

- An energy correction map based on quantum mechanical calculations.
- It improves protein backbone behavior and thus yields more accurate dynamic properties for the protein.
- CHARMM22 force field with CMAP gives the experimentally observed α-helix while without CMAP gives a π-helix for certain model peptides.