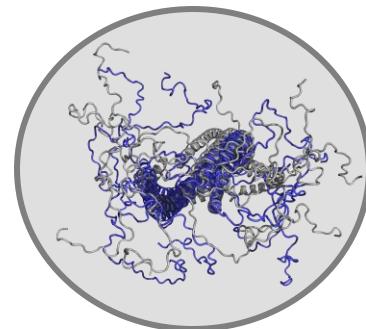


Molecular simulations for biological problems

Monika Fuxreiter

*Laboratory of Protein Interactions and Dynamics
Department of Biomedicine
University of Padova, Italy*

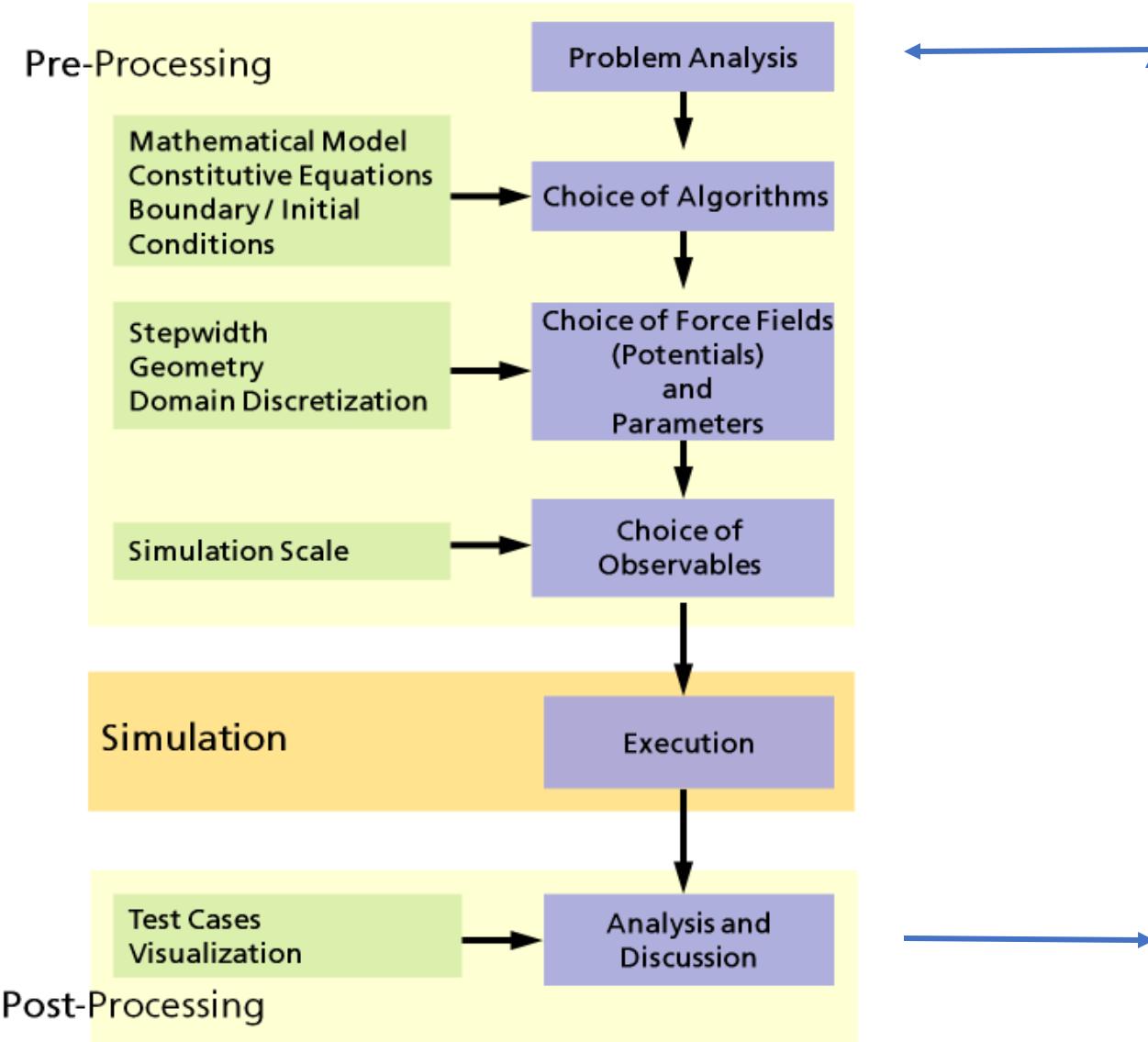


Molecular simulations for biological problems

Lecture 7

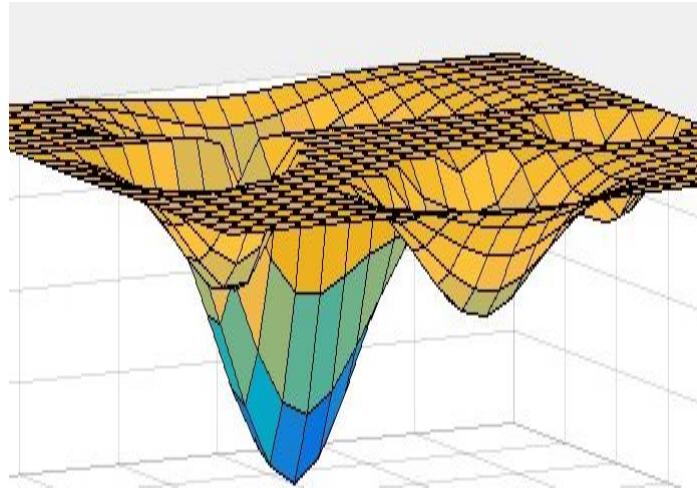
Molecular dynamics I

Computer simulations



Physics-based approaches

All-atom methods

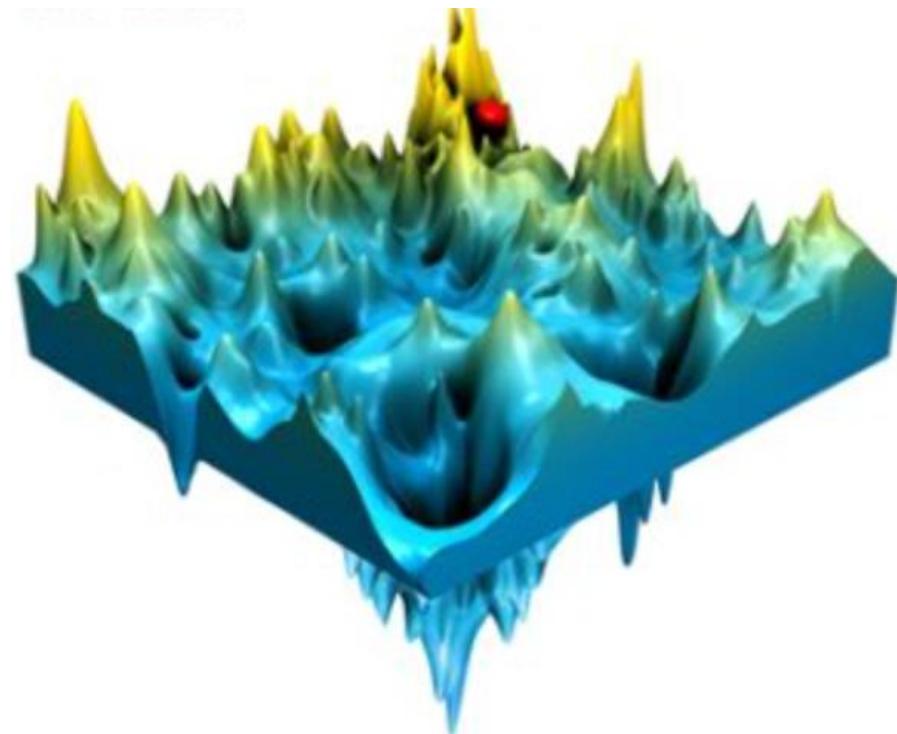


potential energy surface (PES)

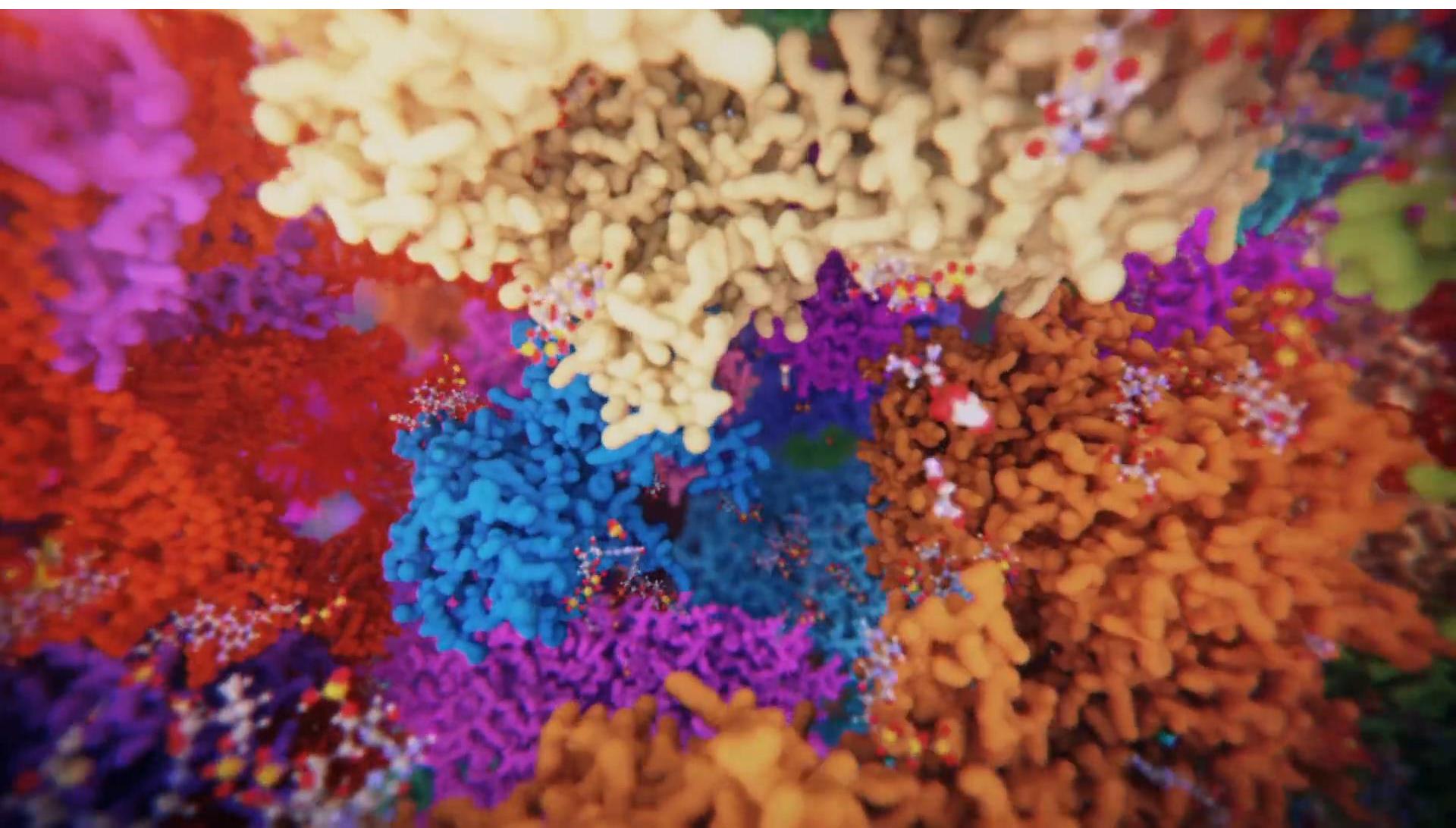
coordinates

Task 1: Define the PES (*force-field*)

Task 2: Explore the landscape (*sampling*)

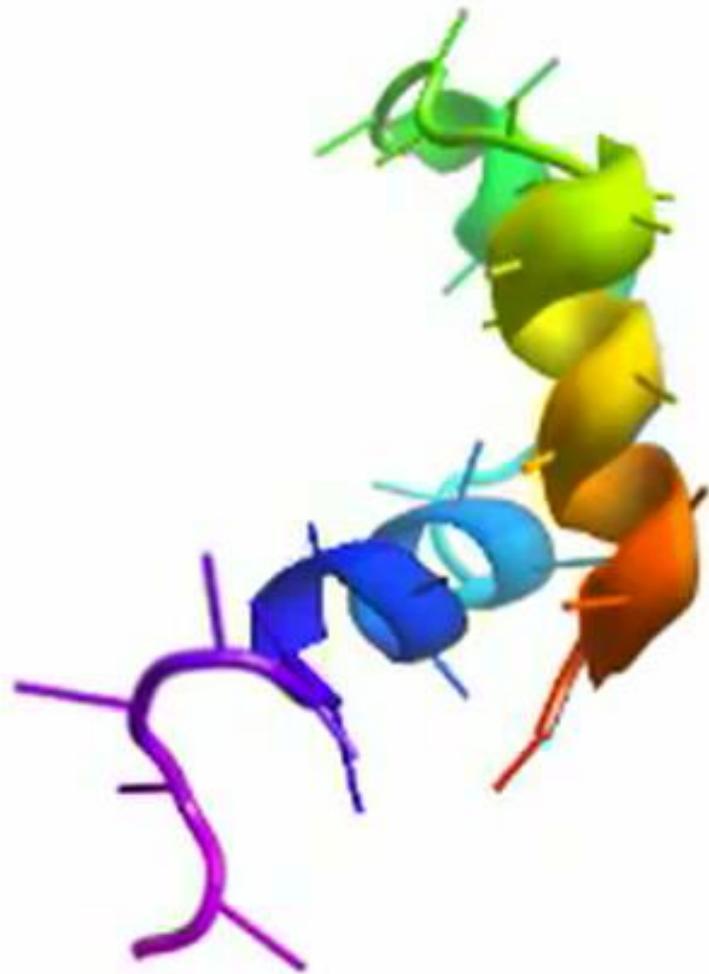


macroscopic properties

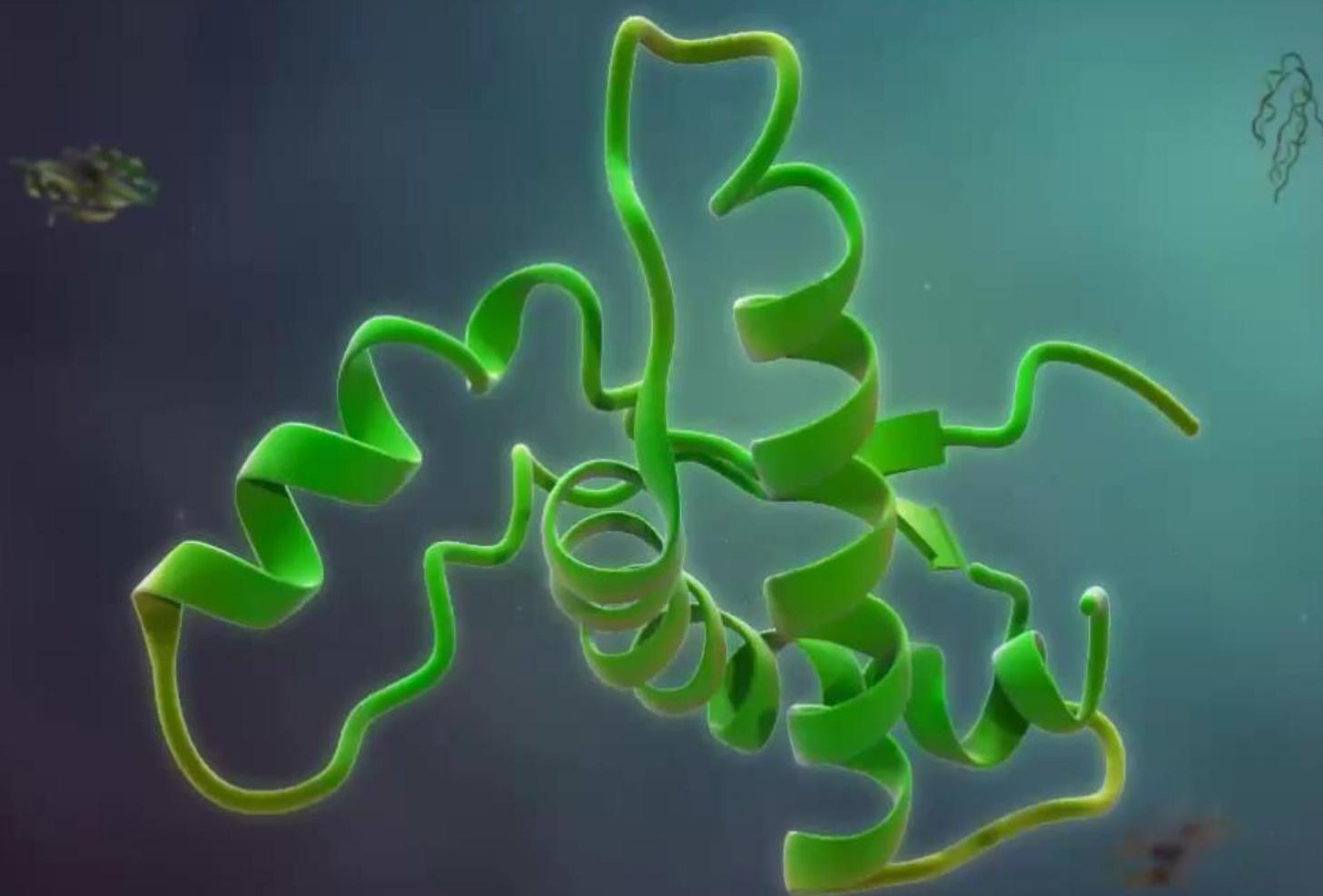


protein misfolding, aggregation

cartoon of 1res - DNA binding domain



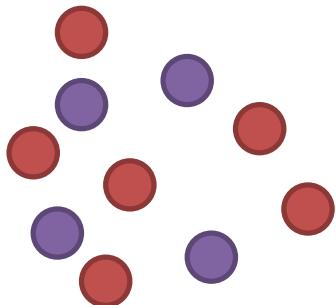
protein misfolding, aggregation



Healthy: α -helical structure



Molecular Dynamics: The Idea



Technique for computing the equilibrium and transport properties of a *classical* many-body system.

Classical – the nuclear motion of the constituent particles obeys the law of classical mechanics (*Newton's second law*).

- * for electrons and light atoms/molecules (H_2 , D_2 , He , ...)
this is not valid – the de Broglie wavelength is comparable to intermolecular spacing.

$$\Lambda = \frac{h}{\sqrt{2\pi mk_B T}}$$

$$\Lambda(Ar) = 0.16 \text{ \AA} \quad \Lambda(H_2O) = 0.24 \text{ \AA} \quad \Lambda(D_2) = 0.60 \text{ \AA} \quad \Lambda(H_2) = 0.72 \text{ \AA}$$

- * for heavier atoms and molecules, this approximation is reasonable and we should not worry about quantum effects.

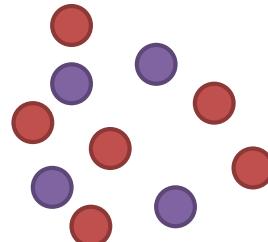
Molecular Dynamics: The Idea

Experiment:



© Alex Bannykh * www.ClipartOf.com/33054

Computational
experiment (MD):

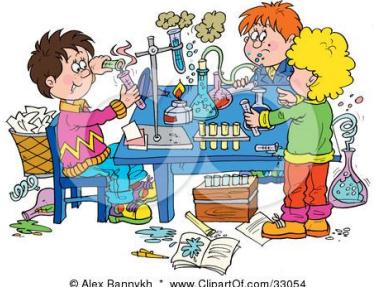


- prepare sample
- connect to an instrument
- measure over a period of time
(statistical noise)

- select a model system of N particles
- solve Newton's equation of motion
- calculate properties of the system

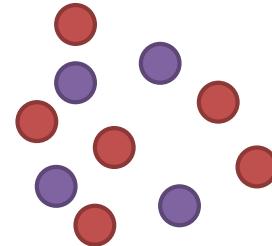
Molecular Dynamics: The Idea

Experiment:



© Alex Bannykh * www.ClipartOf.com/33054

Computational
experiment (MD):



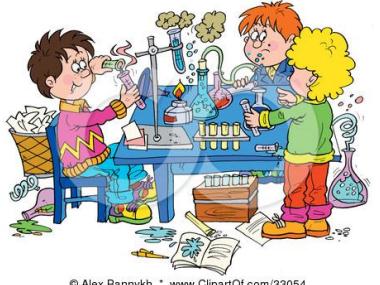
It is essential that we can compare our simulation with experiment, but many properties that we can measure in a simulation can not be measured in a real experiment.

For example, consider a MD simulation of liquid water. We can measure the instantaneous positions and velocities of every molecule in the liquid ...

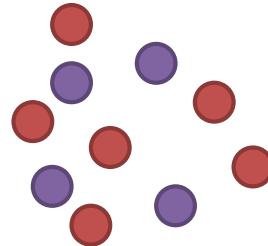
...but no real experiment can provide us with such detailed information!

Molecular Dynamics: The Idea

Experiment:



Computational
experiment (MD):



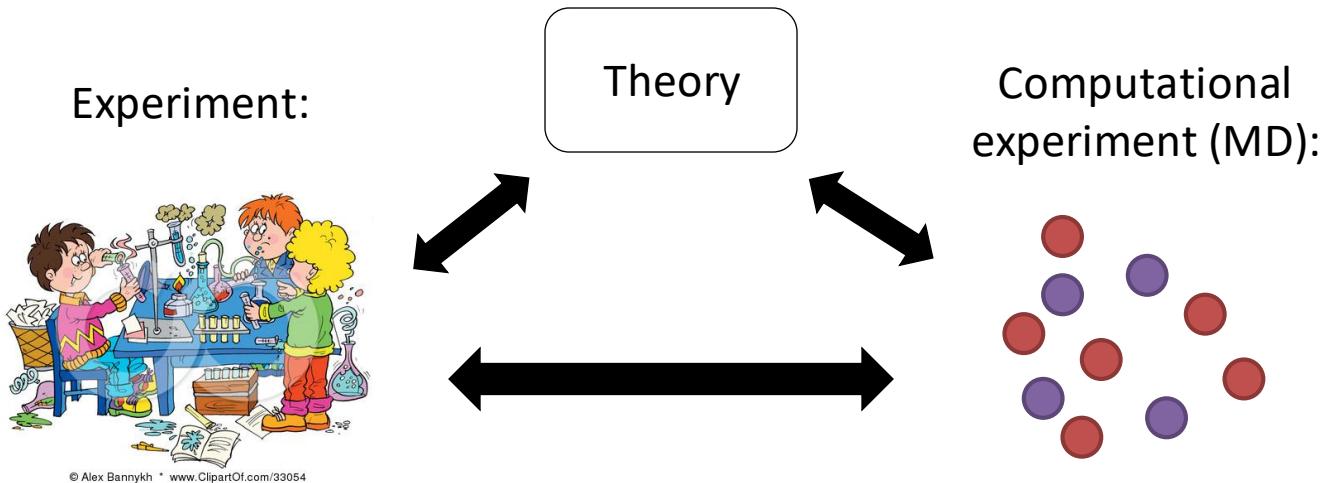
A typical experiment provides average information, averaged over a large number of particles and also over the time of measurement.

So we need to compute averages from our simulations... enter STATISTICAL MECHANICS!

MC: Ensemble averaging

MD: Time averaging

Molecular Dynamics: The Idea



A typical experiment provides average information, averaged over a large number of particles and also over the time of measurement.

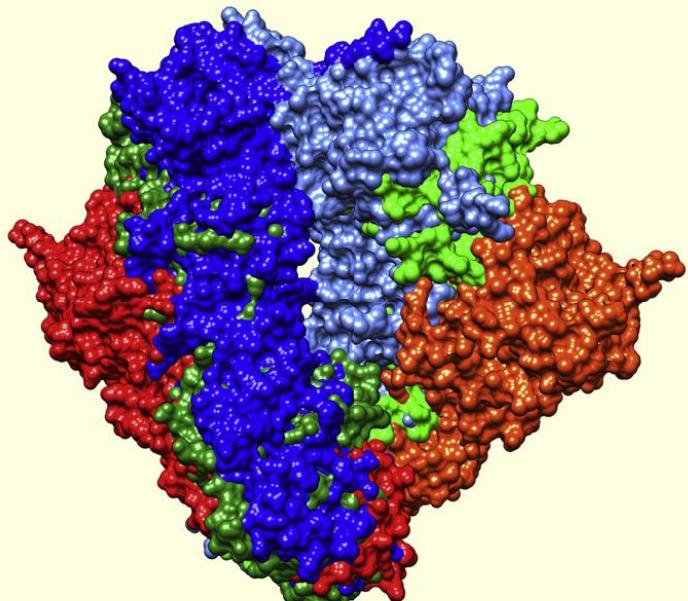
So we need to compute averages from our simulations... enter STATISTICAL MECHANICS!

MC: Ensemble averaging

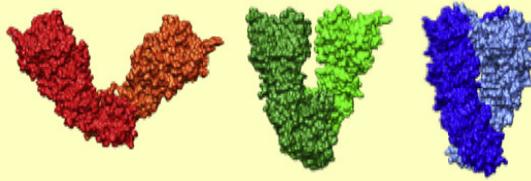
MD: Time averaging

(a)

structural ensemble



(b)



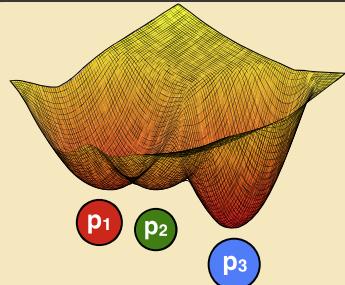
structures

state 1

state 2

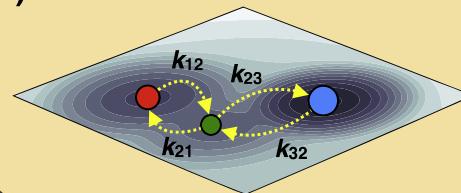
state 3

(c)



thermodynamics

(d)



dynamics

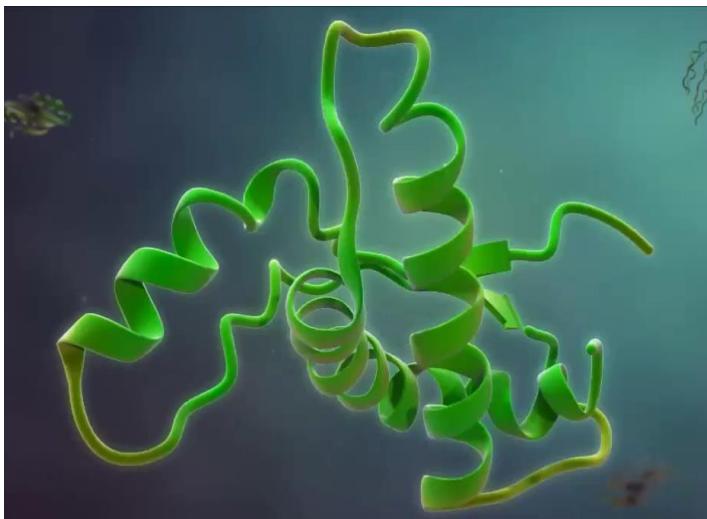
Some Applications of Molecular Dynamics

Wide range of problems in different research fields

- **Chemistry and biochemistry:** structure of biomolecules, dynamics of biomolecules, reactions, drug design, vibrational relaxation and energy transfer, ...
- **Statistical mechanics and physics:** theory of liquids, correlated many-body motion, liquid-to-glass transition, phase transitions, structure and properties of small clusters, ..
- **Materials science:** point, linear and planar defects in crystals, microscopic mechanisms of fracture, extreme conditions, melting, glass properties,...

Molecular dynamics

Some applications:



- solution conformational ensembles
- flexibility
- substrate binding
- functionally relevant movements
- dynamical aspects of function
- correlated properties
- diffusion
- catalysis
- hydration
- electrostatics
- structure determination
- force field development

Molecular dynamics

continuous energy

$$E_{kin} = \frac{1}{2}mv^2$$

$$\mathbf{r}_o, \mathbf{v}_o \Rightarrow \mathbf{r}(t), \mathbf{v}(t)$$

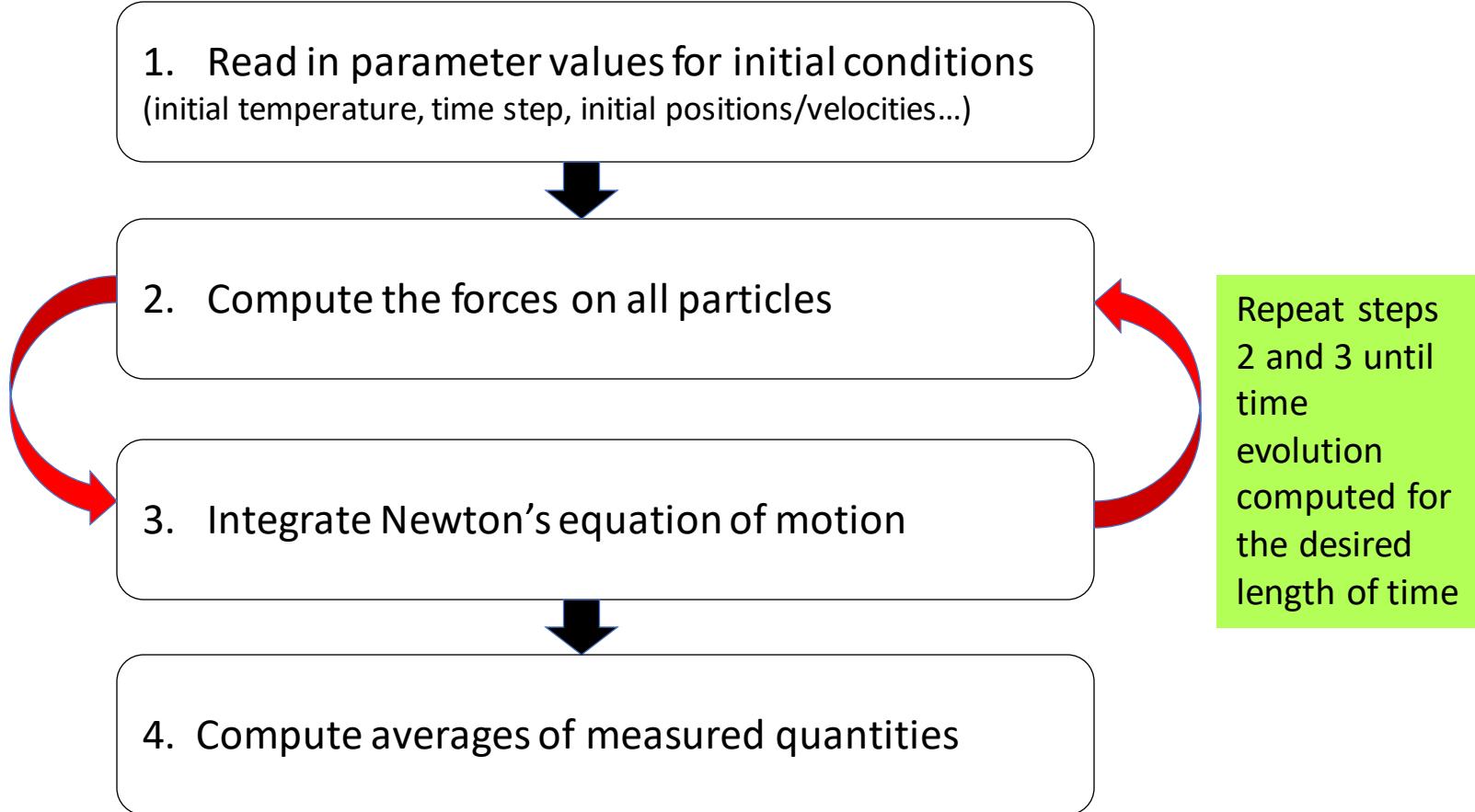
position and velocity $\vec{\mathbf{r}}, \vec{\mathbf{v}}$
to be computed at any time t

$$\mathbf{F} = m\mathbf{a}$$

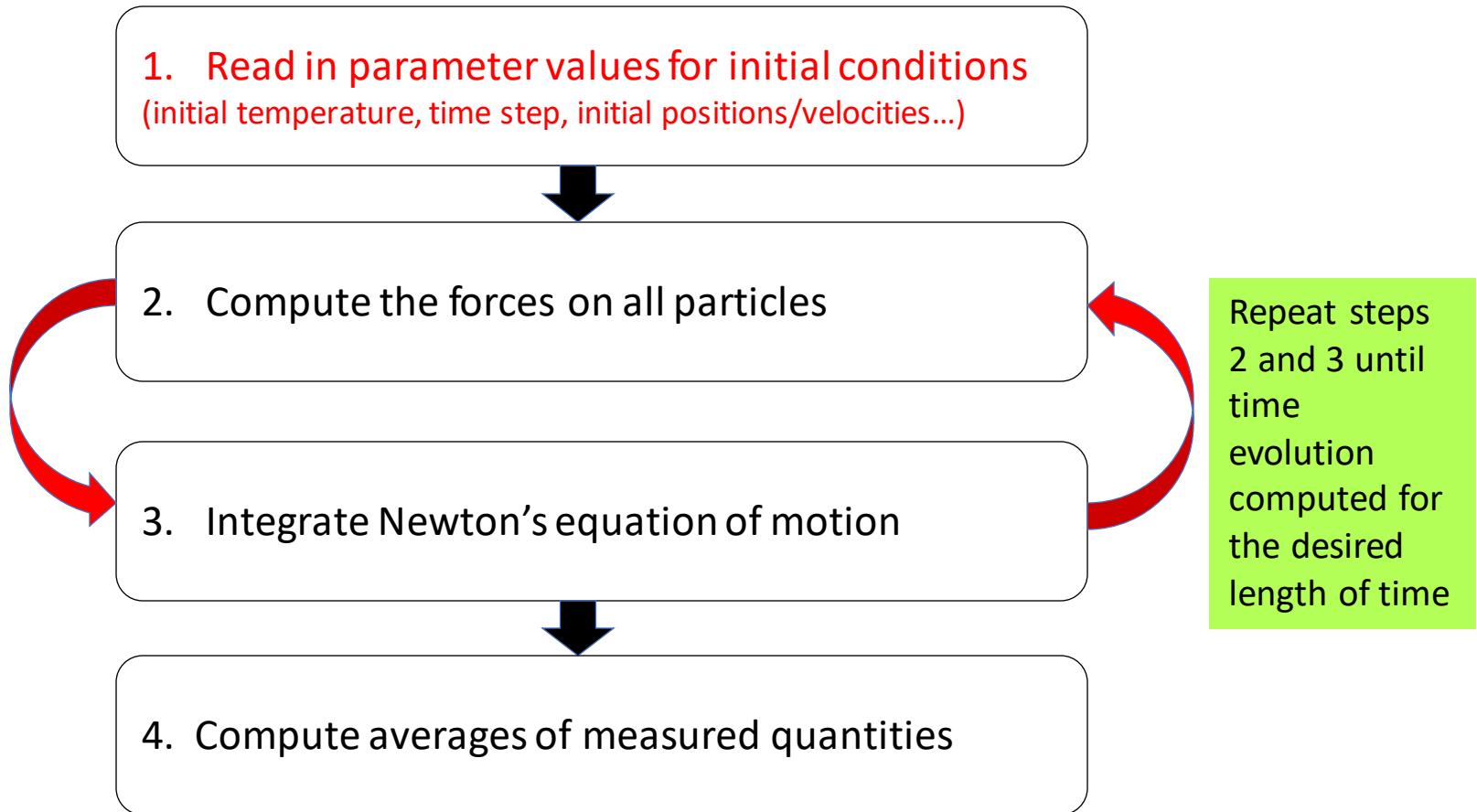
$$\vec{\mathbf{r}}(t) = \vec{\mathbf{r}}_o + \vec{\mathbf{v}}_o t + \frac{1}{2} \mathbf{a} t^2$$

$$\vec{\mathbf{v}}(t) = \vec{\mathbf{v}}_o + \mathbf{a}t$$

Molecular Dynamics: The Program

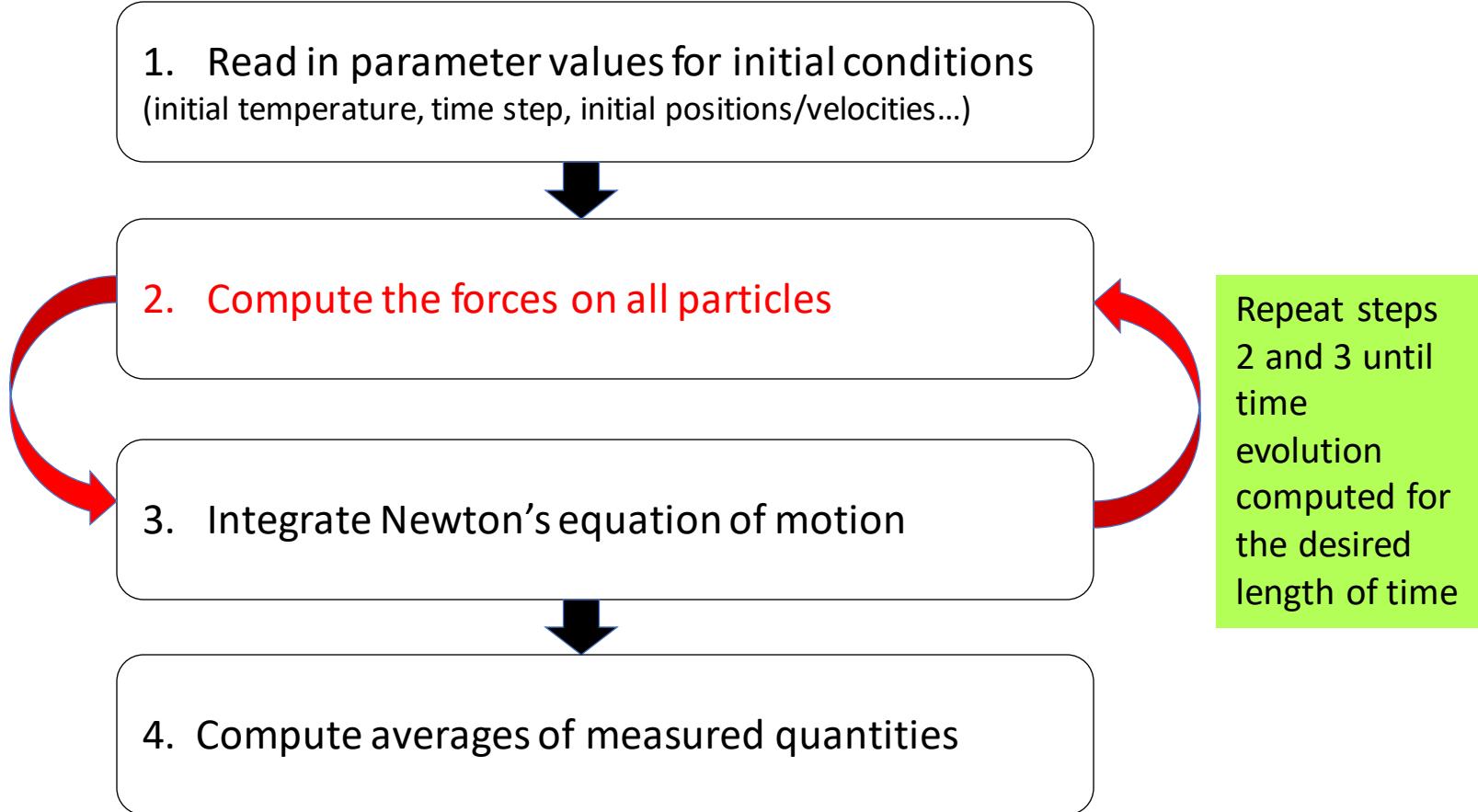


Molecular Dynamics: The Program



- Particle positions should be chosen compatible with the structure that we are aiming to simulate (avoid overlap of the atomic and molecular cores)
... will come back to this later!

Molecular Dynamics: The Program



Pair Forces

In our system, N particles interact via a potential that depends only on the position of each particle:

$$\mathcal{V} = \mathcal{V}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \equiv \mathcal{V}(\mathbf{r}^N)$$

$$\mathbf{f}_i(\mathbf{r}^N) = -\frac{\partial \mathcal{V}(\mathbf{r}^N)}{\partial \mathbf{r}_i}$$

$$m_i \ddot{\mathbf{r}}_i = \mathbf{f}_i(\mathbf{r}^N)$$

$$\mathbf{p}_i = m_i \dot{\mathbf{r}}_i$$

To start, we approximate the potential using pair-wise additive interactions:

$$\mathcal{V}(\mathbf{r}^N) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1, j \neq i}^N v(r_{ij}) \quad , \quad r_{ij} = |\mathbf{r}_i - \mathbf{r}_j| = |\mathbf{r}_j - \mathbf{r}_i| = r_{ji}$$

$$\mathbf{f}_i = -\frac{\partial \mathcal{V}}{\partial \mathbf{r}_i} = -\frac{1}{2} \sum_{j \neq i}^N \left(\frac{\partial v(r_{ij})}{\partial \mathbf{r}_i} + \frac{\partial v(r_{ji})}{\partial \mathbf{r}_i} \right) = -\sum_{j \neq i}^N \frac{\partial v(r_{ij})}{\partial \mathbf{r}_i} = \sum_{j \neq i}^N \mathbf{f}_{ij}$$

Pair forces satisfy Newton's 3rd law in detail :

$$\mathbf{f}_{ji} = -\mathbf{f}_{ij}$$

Properties of Classical Dynamics

Energy Conservation

Total Energy (Hamiltonian) defined as: $\mathcal{H} = \mathcal{K} + \mathcal{V}$

With Kinetic Energy $\mathcal{K} = \sum_i^N \frac{1}{2} m_i \dot{\mathbf{r}}_i^2$

For purely deterministic Newtonian mechanics, **total energy is a constant of the motion:** $\frac{d\mathcal{H}}{dt} = 0$

Proof:

$$\frac{d\mathcal{H}}{dt} = \frac{d}{dt} \left[\sum_i^N \frac{\mathbf{p}_i^2}{2m_i} + \mathcal{V} \right] = \sum_i^N \frac{\mathbf{p}_i \cdot \dot{\mathbf{p}}_i}{m_i} + \sum_i^N \frac{\partial \mathcal{V}}{\partial \mathbf{r}_i} \cdot \dot{\mathbf{r}}_i$$

$$\frac{d\mathcal{H}}{dt} = \sum_i^N \dot{\mathbf{r}}_i \cdot \mathbf{f}_i + \sum_i^N \frac{\partial \mathcal{V}}{\partial \mathbf{r}_i} \cdot \dot{\mathbf{r}}_i = - \sum_i^N \dot{\mathbf{r}}_i \cdot \frac{\partial \mathcal{V}}{\partial \mathbf{r}_i} + \sum_i^N \frac{\partial \mathcal{V}}{\partial \mathbf{r}_i} \cdot \dot{\mathbf{r}}_i = 0$$

Newton's 2nd law

Energy conservation

Properties of Classical Dynamics

Time Reversal Symmetry

Trajectories are completely determined by the initial conditions – including the reverse direction in time.

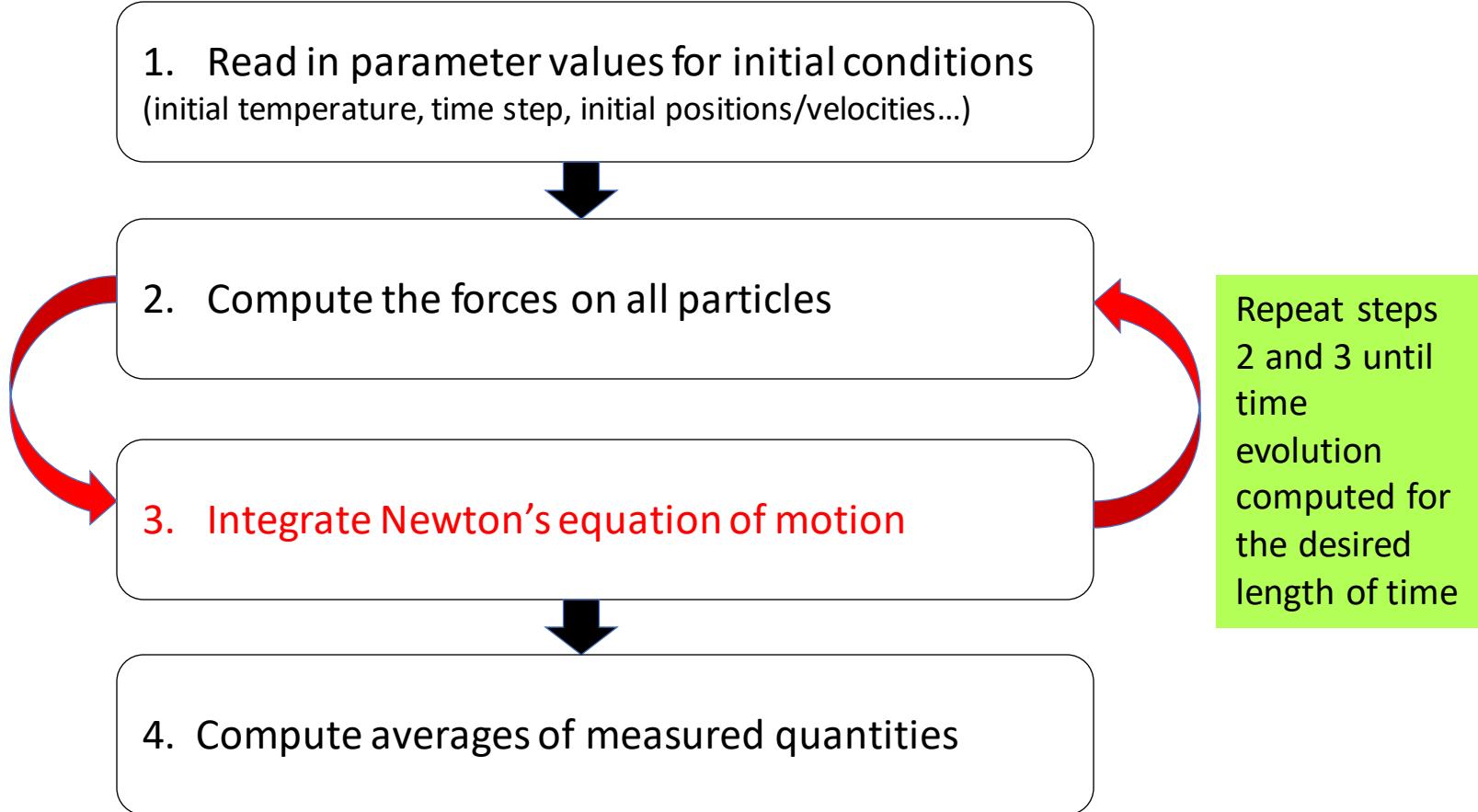
$$\mathbf{r}^N(t | \mathbf{r}_0^N, \mathbf{p}_0^N) \equiv \begin{cases} \mathbf{r}^N(t) \\ \mathbf{r}^N(0) = \mathbf{r}_0^N, \mathbf{p}^N(0) = \mathbf{p}_0^N \end{cases}$$

Consequence: **Time reversal symmetry** - if at a certain time, all velocities (or momenta) are reversed, the system retraces its trajectory backwards in time.

$$\begin{aligned} \mathbf{r}^N(t | \mathbf{r}^N(0), -\mathbf{p}^N(0)) &= \mathbf{r}^N(-t | \mathbf{r}^N(0), \mathbf{p}^N(0)) \\ \mathbf{p}^N(t | \mathbf{r}^N(0), -\mathbf{p}^N(0)) &= -\mathbf{p}^N(-t | \mathbf{r}^N(0), \mathbf{p}^N(0)) \end{aligned}$$

This is a fundamental symmetry of Newton's equation of motion.

Molecular Dynamics: The Program



Integrating the Equation of Motion

Given the current positions and velocities/momenta, we would like to *propagate* the dynamics forward or backward in time in **discrete steps** of length δt , to generate a chain of configurations (“trajectory”).

$$\mathbf{r}^N(t_0) = \mathbf{r}^N(0), \dots, \mathbf{r}^N(t_{m-1}) = \mathbf{r}^N(t_m - \delta t), \mathbf{r}^N(t_m), \mathbf{r}^N(t_{m+1}) = \mathbf{r}^N(t_m + \delta t), \dots$$

Most obvious is to use a Taylor expansion in time:

$$r_i(t + \delta t) \approx r_i(t) + \dot{r}_i(t)\delta t + \frac{1}{2}\ddot{r}_i(t)\delta t^2 + \frac{1}{6}\dddot{r}_i(t)\delta t^3 + O(\delta t^4)$$

Using what we know: $r_i(t + \delta t) = \mathbf{r}_i(t) + \delta t \mathbf{v}_i(t) + \frac{\delta t^2}{2m_i} \mathbf{f}_i(t) + O(\delta t^3)$ (1)

We can similarly Taylor expand the velocities:

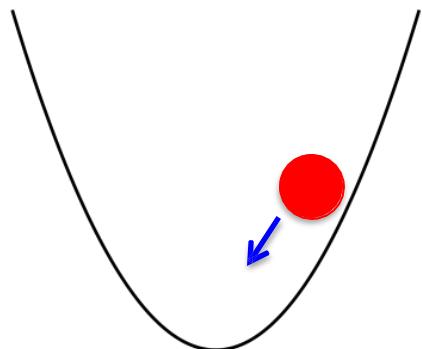
$$\mathbf{v}_i(t + \delta t) = \mathbf{v}_i(t) + \frac{\delta t}{m_i} \mathbf{f}_i(t) + O(\delta t^2) \quad (2)$$

Together, **1** and **2** can be used to simulate system, known as the “Euler Algorithm”
This turns out to be **inaccurate in a bad way**.

Example: Simple Harmonic Oscillator (SHO)

Not a multi-particle system, but... a reference model we can solve analytically

Potential: $V(x) = \frac{1}{2} kx^2$



Equations of motion easily derived:

$$f = -\frac{dV}{dx} = -kx$$

$$m\frac{d^2x}{dt^2} = -kx$$

Possible solutions are:

$$x(t) = \cos \omega t$$

$$x(t) = \sin \omega t$$

with angular frequency: $\omega = \sqrt{\frac{k}{m}}$

Quick check!

$$\dot{x} = -\omega \sin \omega t \quad \ddot{x} = -\omega^2 \cos \omega t = -\omega^2 x$$

$$\text{Then: } m\ddot{x} = -m\omega^2 x = -kx = f$$

General solution: $x = A \cos \omega t + B \sin \omega t$

General solution for SHO

Since we are solving a second order differential equation (in time), there will be two degrees of freedom in the general solution:

$$m \frac{d^2x}{dt^2} = -kx$$

The general solution can be written in the form:

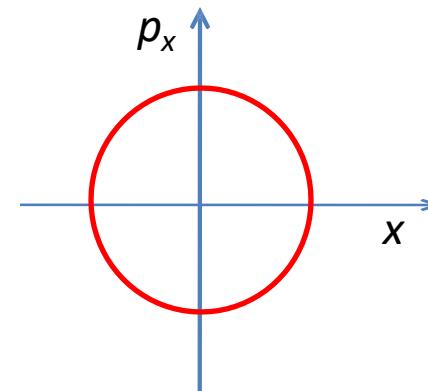
$$x_0 = x(t_0) = A \cos \omega t_0 + B \sin \omega t_0$$

$$v_0 = \dot{x}(t_0) = -A\omega \sin \omega t_0 + B\omega \cos \omega t_0$$

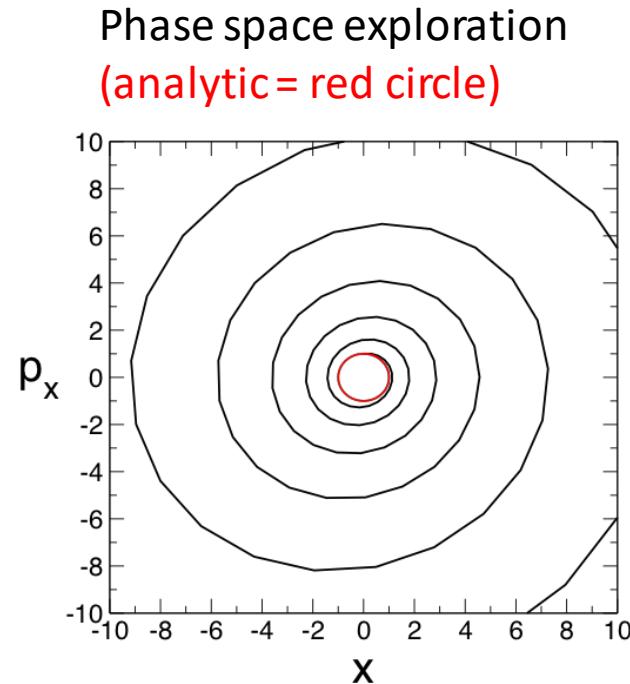
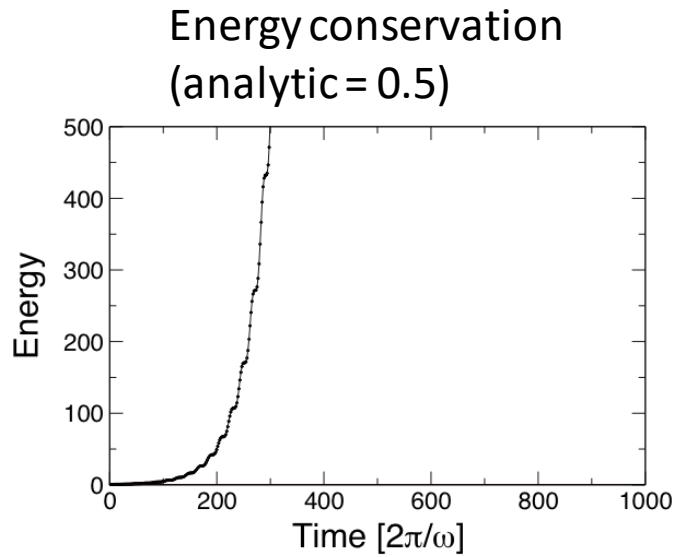
With: $A = x_0$ $B = \frac{v_0}{\omega}$

Once the initial conditions x_0 and v_0 are specified the dynamics is **completely deterministic**... for both future and past times

Phase space ($\mathbf{r}^N, \mathbf{p}^N$) is two dimensional for simple harmonic oscillator i.e. (x, p_x)



Euler results for SHO



By any measure, the Euler algorithm is not good!

Euler results for SHO

The naive “forward Euler” algorithm:

- is not time reversible
- does not conserve volume in phase space
- suffers from energy drift

Better approach: “Verlet” algorithm...

The “Position” Verlet Algorithm

As before we start with the Taylor expansion forward in time:

$$\mathbf{r}_i(t + \delta t) = \mathbf{r}_i(t) + \delta t \mathbf{v}_i(t) + \frac{\delta t^2}{2m_i} \mathbf{f}_i(t) + \frac{\delta t^3}{6} \mathbf{b}_i(t) + O(\delta t^4) \quad (1)$$

Can also expand backward in time:

$$\mathbf{r}_i(t - \delta t) = \mathbf{r}_i(t) - \delta t \mathbf{v}_i(t) + \frac{\delta t^2}{2m_i} \mathbf{f}_i(t) - \frac{\delta t^3}{6} \mathbf{b}_i(t) + O(\delta t^4) \quad (2)$$

(1) + (2) gives the position update:

$$\mathbf{r}_i(t + \delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \delta t) + \frac{\delta t^2}{m_i} \mathbf{f}_i(t) + O(\delta t^4) \quad (3)$$

(1) – (2) gives the velocities (but this is not required to propagate dynamics).

$$\mathbf{v}_i(t) = \frac{1}{2\delta t} [\mathbf{r}_i(t + \delta t) - \mathbf{r}_i(t - \delta t)] + O(\delta t^3)$$

Eq. 3 constitutes the Verlet algorithm with is accurate to $O(\delta t^4)$ compared with $O(\delta t^3)$ for Euler.

Velocity Verlet

Previously described “position Verlet” makes the increase in accuracy over Euler. Velocity Verlet is a different formulation which is convenient for many algorithms, and turns out to be equivalent to position Verlet.

(i) Advance positions $\mathbf{r}_i(t)$ using a Taylor expansion (as in Euler):

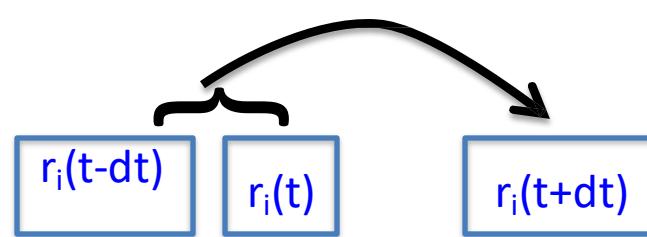
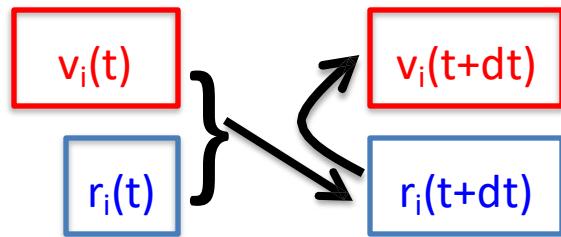
$$\mathbf{r}_i(t + \delta t) = \mathbf{r}_i(t) + \delta t \mathbf{v}_i(t) + \frac{\delta t^2}{2m_i} \mathbf{f}_i(t)$$

(ii) Compute forces $\mathbf{f}_i(t+\delta t)$ at $\mathbf{r}_i(t+\delta t)$ and store for the next step

$$\mathbf{f}_i(t + \delta t) = \mathbf{f}_i(\{\mathbf{r}_i(t + \delta t)\}) = \mathbf{f}_i\left(\left\{\mathbf{r}_i(t) + \delta t \mathbf{v}_i(t) + \frac{\delta t^2}{2m_i} \mathbf{f}_i(t)\right\}\right)$$

(iii) Use $\mathbf{f}_i(t)$ and $\mathbf{f}_i(t+\delta t)$ to update velocities (differs from Euler)

$$\mathbf{v}_i(t + \delta t) = \mathbf{v}_i(t) + \frac{\delta t}{2m_i} [\mathbf{f}_i(t) + \mathbf{f}_i(t + \delta t)]$$



Derivation of Velocity Verlet

Advance positions $r_i(t)$ by Taylor expansion from t to $t+\delta t$:

$$r_i(t + \delta t) = r_i(t) + \delta t v_i(t) + \frac{\delta t^2}{2m_i} f_i(t) + \dots$$

Substitute in (backward) Taylor expansion from $t+\delta t$ to t :

$$r_i(t) = r_i(t + \delta t) - \delta t v_i(t + \delta t) + \frac{\delta t^2}{2m_i} f_i(t + \delta t) + \dots$$

Adding this to (forward) expansion t to $t+\delta t$:

Adding this to (forward) expansion t to $t + \delta t$

gives expression for advancing velocities:

or advancing velocities

Note that δt^3 terms
cancel here as in
position Verlet



Equivalence of Velocity Verlet and Position Verlet

Subtract prediction r_i for step $t - \delta t \rightarrow t$ from $t \rightarrow t + \delta t$:

$$\begin{aligned} r_i(t + \delta t) - r_i(t) &= r_i(t) - r_i(t - \delta t) + \\ \delta t[v_i(t) - v_i(t - \delta t)] &+ \frac{\delta t^2}{2m_i}[f_i(t) - f_i(t - \delta t)] \end{aligned}$$

Eliminate velocity by inserting prediction step $t - \delta t \rightarrow t$:

$$v_i(t) = v_i(t - \delta t) + \frac{\delta t}{2m_i}[f_i(t - \delta t) + f_i(t)]$$

giving:

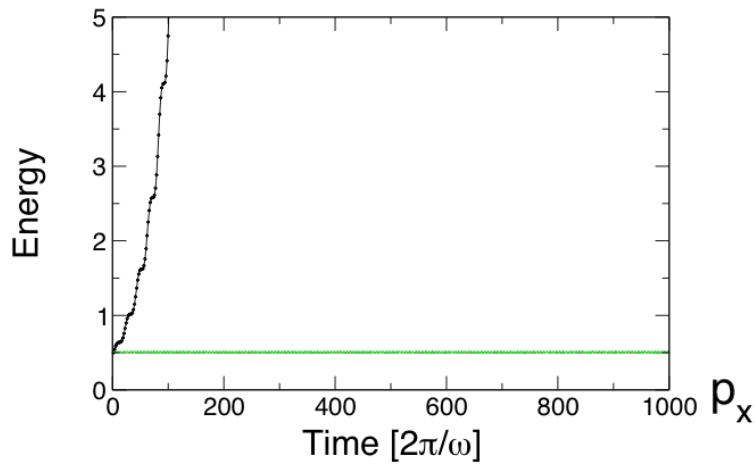
$$r_i(t + \delta t) - r_i(t) = r_i(t) - r_i(t - \delta t) + \frac{\delta t^2}{m_i}f_i(t)$$

Bringing $r_i(t)$ to the r.h.s we find the Verlet prediction for step $t \rightarrow t + \delta t$:

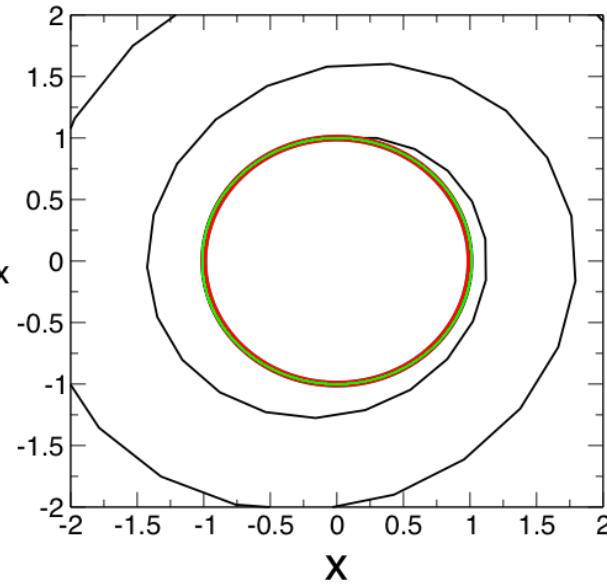
to the r.h.s. we find the Verlet prediction for st

Velocity Verlet results for SHO

Energy conservation:
Verlet = green line



Phase space exploration:
Black = Euler
Red = Analytic
Green = Verlet



Velocity Verlet seems very good, but ...

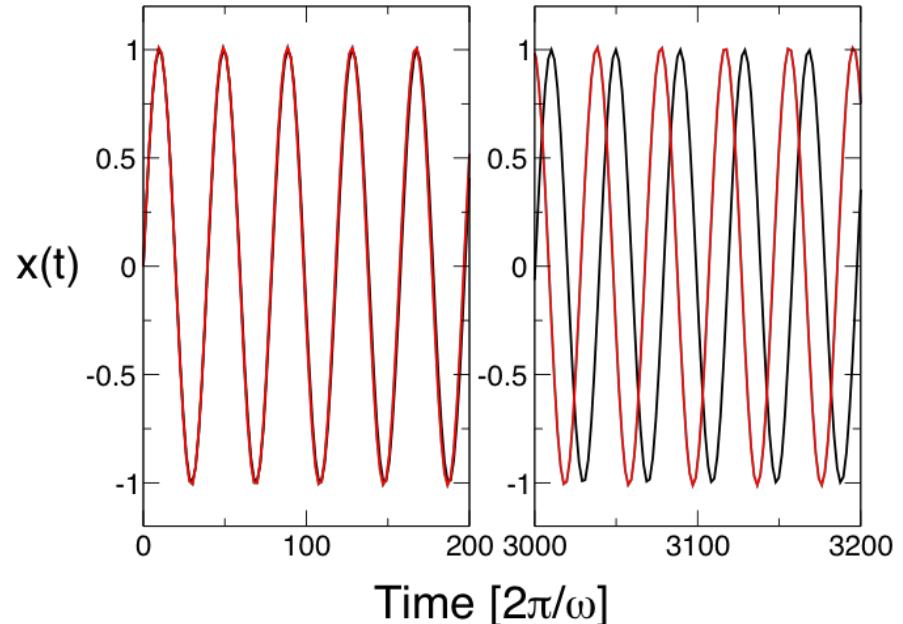
Velocity Verlet: “Lyupanov Instability”

From phase-space diagram,
Verlet integration seems
perfect...

However, $x(t)$ deviates from
exact solution quite quickly!

Would be more dramatic with
more degrees of freedom.

This so-called “Lyupanov
instability” plagues all discrete
integration algorithms



Miraculously, Verlet conserves energy despite deviating from the “true dynamics”
of the system!

Good Things About Verlet

Verlet or related schemes are used for almost all MD simulations.

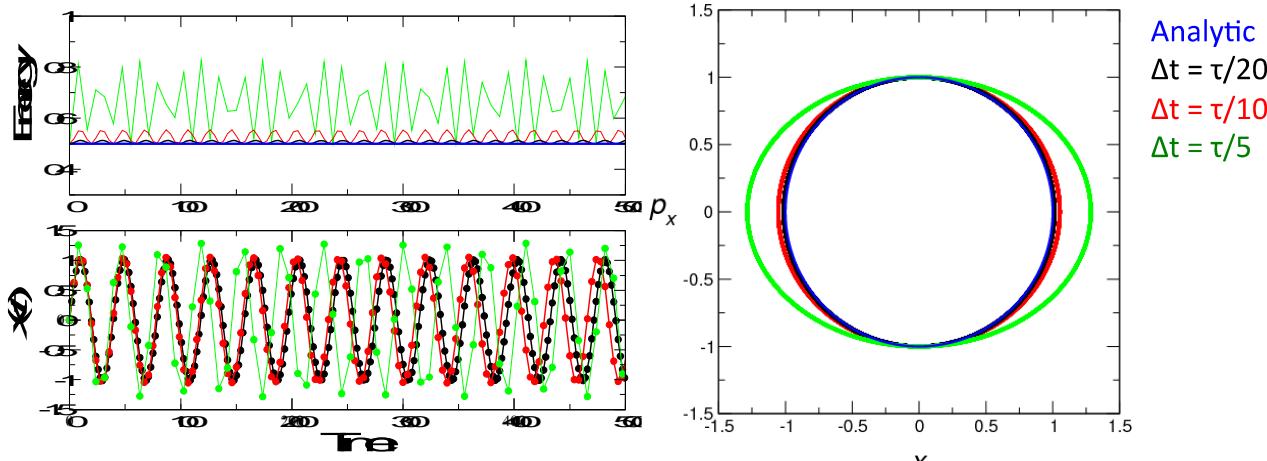
Why?

- Simple to implement and requires only positions and forces
- Computationally efficient
- Energy conservation is excellent – better than higher order integration algorithms
- Explicitly obeys time-reversal symmetry

Choosing the Time Step

From a performance point of view, we would like to choose the largest time step possible. Too large a time step will obviously lead to large errors in the numerical integration.

Velocity Verlet for SPO with different step sizes:



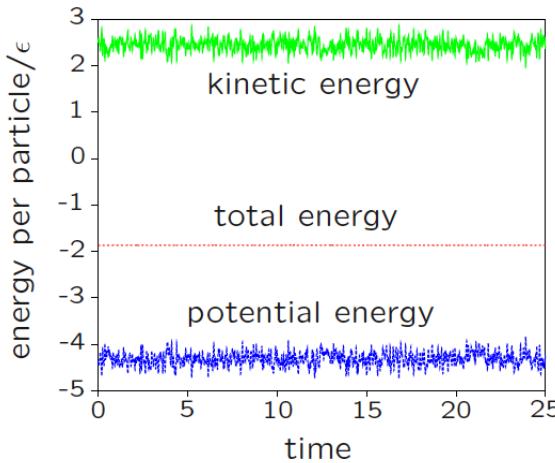
Generally the time step should be less than the period of the highest frequency oscillation in the system – a step size of $\delta t = \tau/20$ is usually “safe” for Verlet-based schemes

For a harmonic oscillator, $\tau = (2\pi/\omega)$.

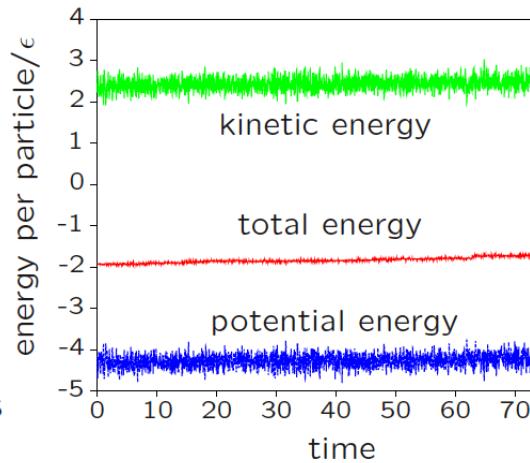
Choosing the Time Step

The best indication of this break down is drift in the total energy which should be rigorously conserved according to Newton's equation of motion.

A more complex system: 108 Lennard-Jones particles in a cubic box



Time step optimized
 $\delta t^* = 0.005$, 5000 steps
Total energy conserved



Time step too **long**
 $\delta t^* = 0.015$, 5000 steps
Small drift in total energy

$$\delta t \text{ in reduced units: } t^* = t \left(\frac{\epsilon}{m\sigma^2} \right)^{1/2}$$

Note that although the total energy is conserved, both kinetic and potential energy fluctuate!

Molecular dynamics

flowchart

- starting configuration (optimised structure)
- setting of T, assigning velocities

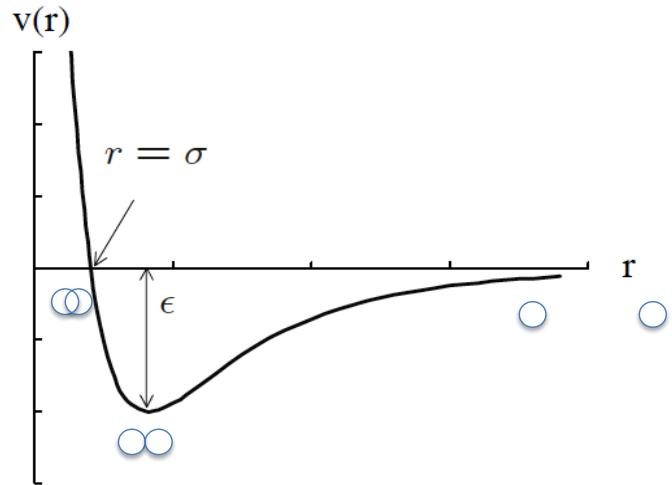
$$p(v_{ix}) = \left(\frac{m}{2\pi k_B T} \right)^{1/2} \exp \left[-\frac{1}{2} \frac{mv_{ix}^2}{k_B T} \right]$$

- 
- calculation of next position (integration algorithm, step-size)
 - calculation of forces
 - calculation of velocities

Choosing Initial Positions and Velocities

Lennard-Jones pair potential:
model for Liquid Argon.

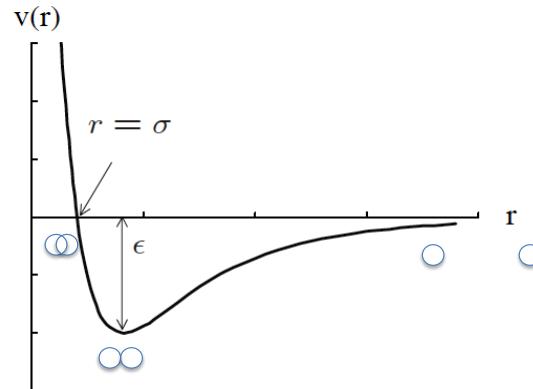
$$V_{LJ}(r) = 4\epsilon \left[\frac{(\sigma/r)^{12}}{(\sigma/r)^6} - \frac{1}{(\sigma/r)^6} \right]$$



Initial positions. All reasonable initial conditions are in principle acceptable (put positions at random can sometimes be dangerous)
e.g. For proteins – use coordinates from the X-ray/NMR structure

Choosing Initial Positions and Velocities

$$V_{\text{LJ}}(r) = 4 \left[\frac{1}{2} \frac{1}{r^6} - \frac{1}{2} \frac{1}{r^{12}} \right]$$



Initial velocities : standard strategy

- Setup random velocities
- Compute the velocity center of mass and set it to zero
- Rescale the velocities by a factor $\sqrt{T_{\text{desired}}/T(t)}$ to reach the desired temperature, as the total kinetic energy in the simulation fluctuates:

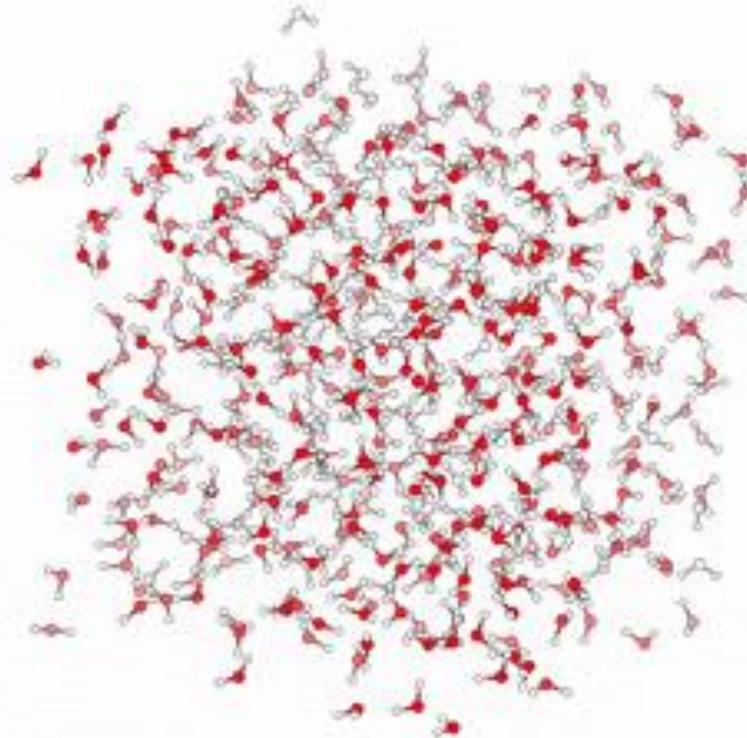
$$T(t) = \frac{1}{2} \sum_{i=1}^N \frac{m_i v_i^2(t)}{k_B N_f}$$

Recall: From the **equipartition theorem**, the kinetic energy per degree of freedom may be related to the temperature in the microcanonical ensemble *via*:

$$\left\langle \frac{1}{2} m v_a^2 \right\rangle = \frac{1}{2} k_B T$$

....we will come back to this later!

Molecular dynamics

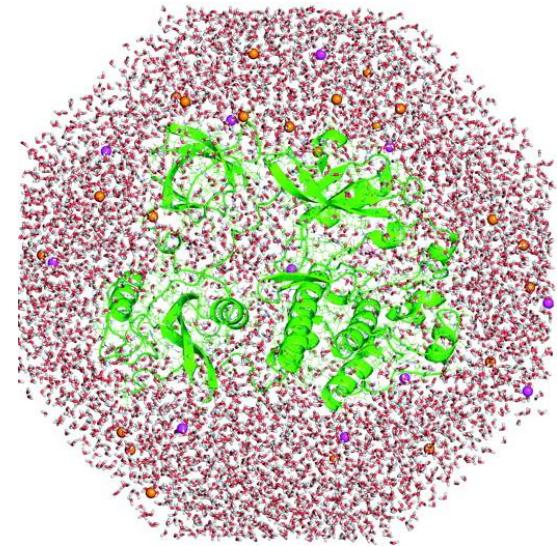


solvation and boundary conditions
optimising position and cell

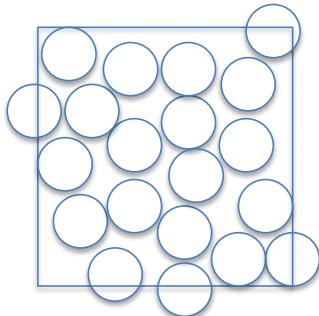
Molecular dynamics

Model generation

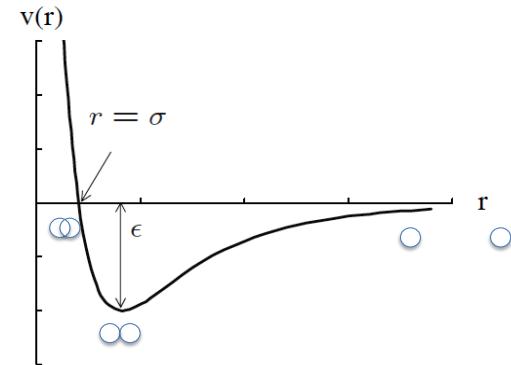
- Initial configuration
 - solvation
 - ◆ *primary hydration shell*
 - ◆ *2-3 solvation shells*
 - ◆ *cell optimisation (MC waters, pre-eq.)*
 - ions
 - periodic boundary conditions (PBC)
 - ensemble (NVT, NPT, μ VT)
 - constraints $E(x) = k(x - x_0)^2$
 - ◆ *rigid bodies*
 - ◆ *constraining to start configurations*
 - ◆ *fixed geometry (hydrogens)*



Periodic Boundary Conditions

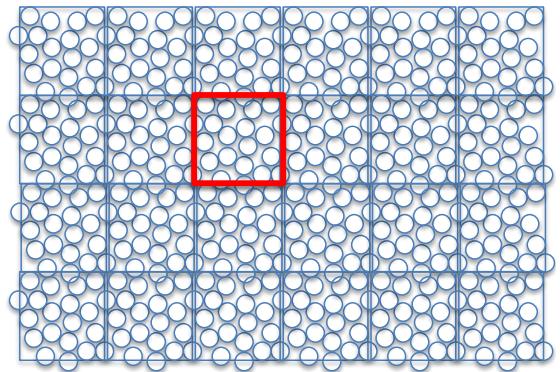


$$V_{\text{LJ}}(r) = 4 \left[\frac{1}{2} \frac{1}{r^6} - \frac{1}{2} \frac{1}{r^{12}} \right]$$



In order to describe liquids with uniform (average) density we can either take a very big cluster and hope that in the interior of the cluster surface effects can be neglected, or use periodic boundary conditions.

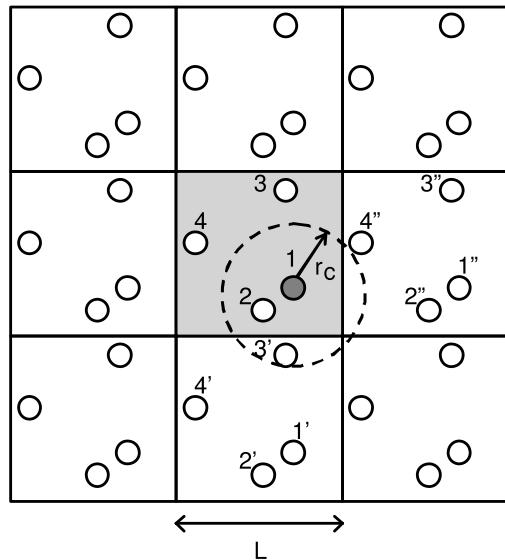
Periodic boundary conditions replicate a MD cell with the shape of a parallelepiped, cube ..., and its contents, all over space mimicking the homogeneous state of a liquid or solid. This eliminates edge/surface effects; other “finite-size” effects remain.



Periodic Boundary Conditions

The total potential can be written as:

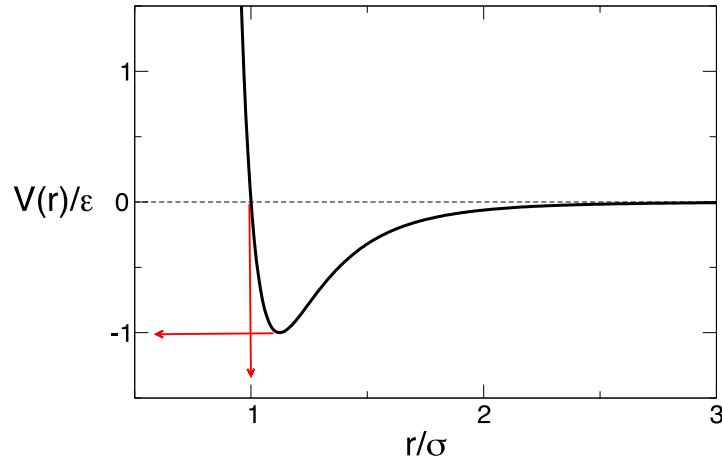
$$\mathcal{V}(\mathbf{r}^N) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1, j \neq i}^N v(r_{ij})$$



For long range interactions, this implies a sum over every cell of an infinite lattice, though at least we know where all the particles are!
Fortunately, if interactions are short-ranged, they can be truncated beyond some cut-off distance r_c . All pairs further apart than r_c are ignored.

With this assumption, provided that $r_c < L/2$, with L the length of the simulation box, each atom i is guaranteed to interact with *at most* one of the images of neighbour j .

Truncation of Potential



Spherical cutoff: neglect interactions beyond radius r_c :

$$v^{trunc}(r) = \begin{cases} v^{lj}(r) & r \leq r_c \\ 0 & r > r_c \end{cases}$$

Compute many fewer pair interactions!

$$V_{LJ}(r) = 4 \left[\frac{\left(\frac{1}{r}\right)^{12}}{2} - \frac{\left(\frac{1}{r}\right)^6}{2} \right]$$

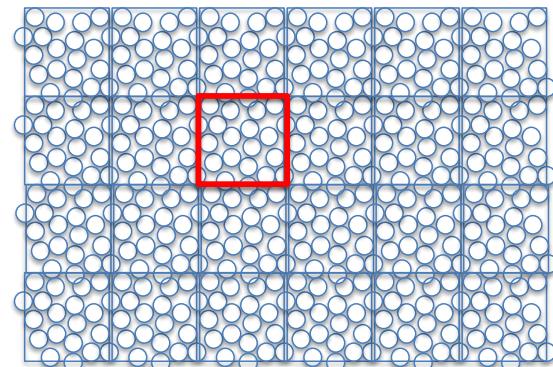
In reduced units:

$r < 1$ strongly repulsive

$r > 1$ attractive

$r \approx 1.12$ minimum $v_0 = -1$

$r = 3$ effectively zero, $v(3) \approx -0.005$

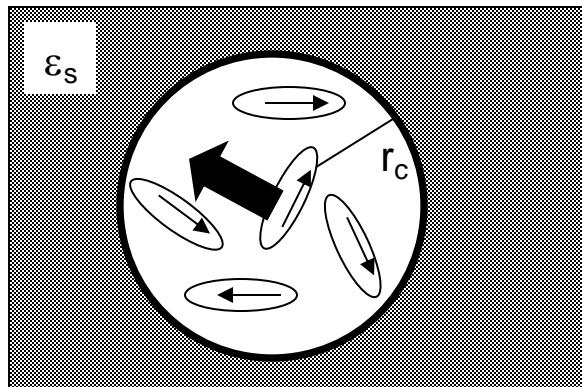


Molecular dynamics

bulk phase , environment

Problems: long-range interactions

Reaction field method (Onsager)



r_c : cutoff

ϵ_s : dielectric constant of bulk phase

$$\xi_i = \frac{2(\epsilon_s - 1)}{(\epsilon_s + 1)} \left(\frac{1}{r_c^3} \right) \sum_{j; r_{ij} \leq r_c} \mu_j$$

interaction with reaction field

$\xi_i \mu_j$

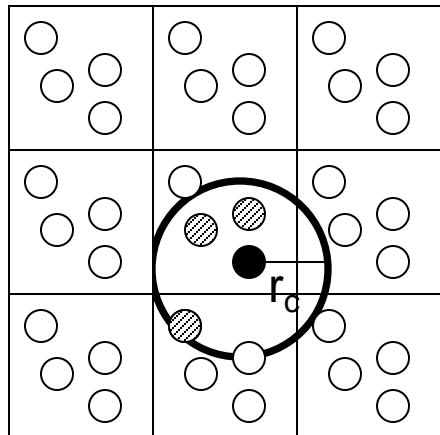
boundary issue: energy is not continuous

Molecular dynamics

bulk phase , environment

periodic boundary conditions

Minimum image:



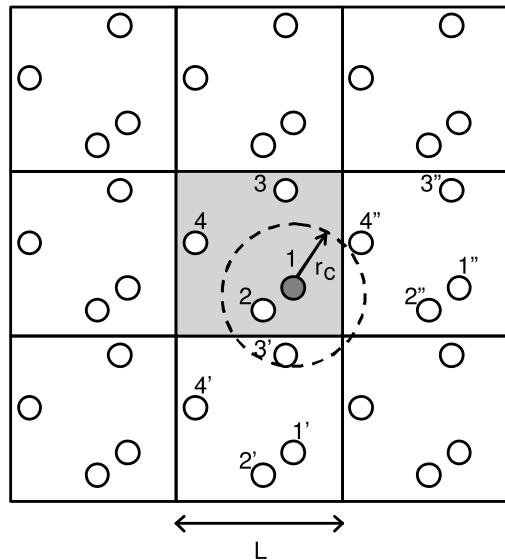
r_c : cutoff

- interactions only with the closest replica
- $r_c < a/2, b/2, c/2$
- $r_c \sim 10\text{\AA}$

Periodic Boundary Conditions

The total potential can be written as:

$$\mathcal{V}(\mathbf{r}^N) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1, j \neq i}^N v(r_{ij})$$



For long range interactions, this implies a sum over every cell of an infinite lattice, though at least we know where all the particles are!
Fortunately, if interactions are short-ranged, they can be truncated beyond some cut-off distance r_c . All pairs further apart than r_c are ignored.

With this assumption, provided that $r_c < L/2$, with L the length of the simulation box, each atom i is guaranteed to interact with *at most* one of the images of neighbour j .

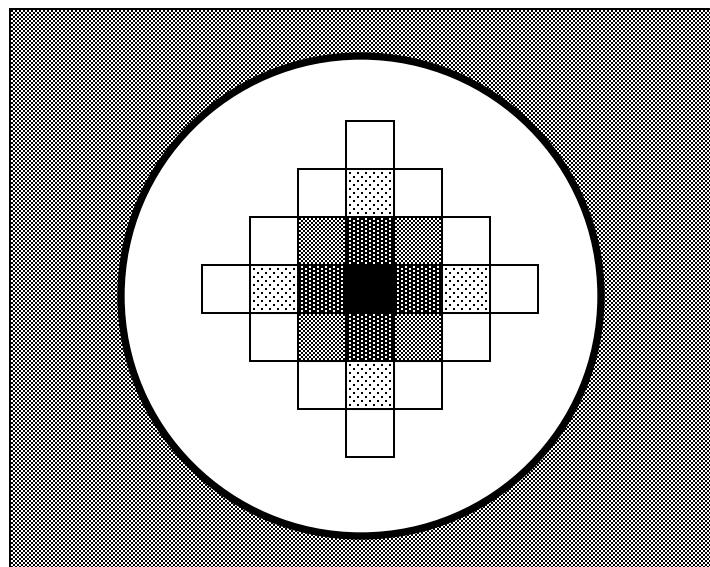
Molecular dynamics

bulk phase , environment

periodic boundary conditions

Ewald method: „crystal-like simulation”

$$\varepsilon_r=0 ; \varepsilon_r=\infty$$



$$E = \frac{1}{2} \sum_{n=0}^{\infty} \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{4\pi\varepsilon_0 |r_{ij} + n|}$$

interactions with all images

as sum is computed in a Fourier space,
slowly converges

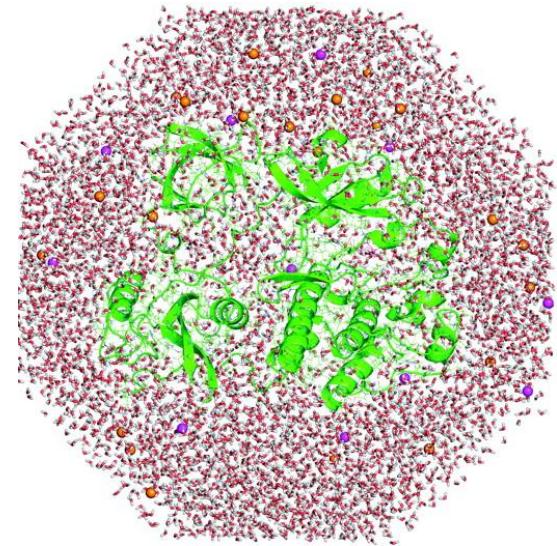
long-range contributions do not have singularity

PME: Particle Mesh Ewald (between point charges) using
B-spline interpolation

Molecular dynamics

Model generation

- Initial configuration
 - solvation
 - ◆ *primary hydration shell*
 - ◆ *2-3 solvation shells*
 - ◆ *cell optimisation (MC waters, pre-eq.)*
 - ions
 - periodic boundary conditions (PBC)
 - ensemble (NVT, NPT, μ VT)
 - constraints $E(x) = k(x - x_0)^2$
 - ◆ *rigid bodies*
 - ◆ *constraining to start configurations*
 - ◆ *fixed geometry (hydrogens)*

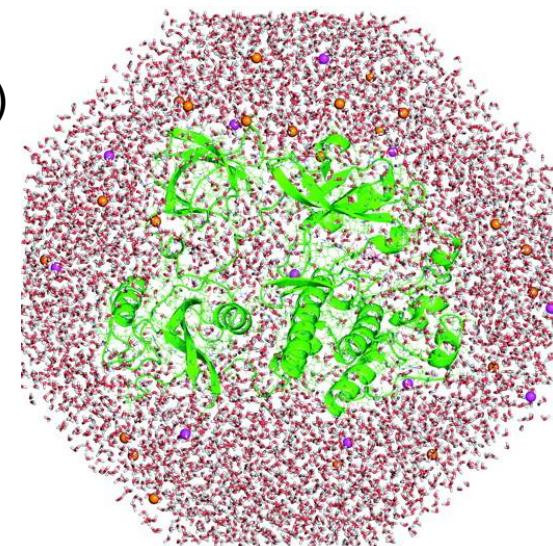


Molecular dynamics

Model preparation

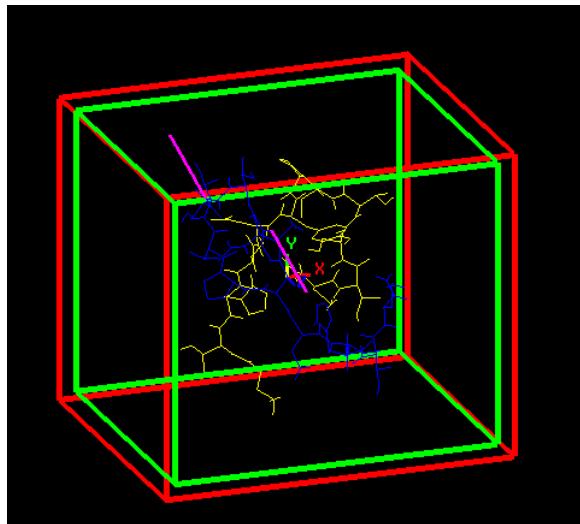
- Optimisation of the initial structure
 - solvent (*min.*)
 - solvent + ions optimisation MM
 - solvent + ions optimisation (MD or MC, low T)
(- density adjustments)
- Heating, stabilisation of temperature ($T = 0 \rightarrow T_{act}$)

} protein fixed

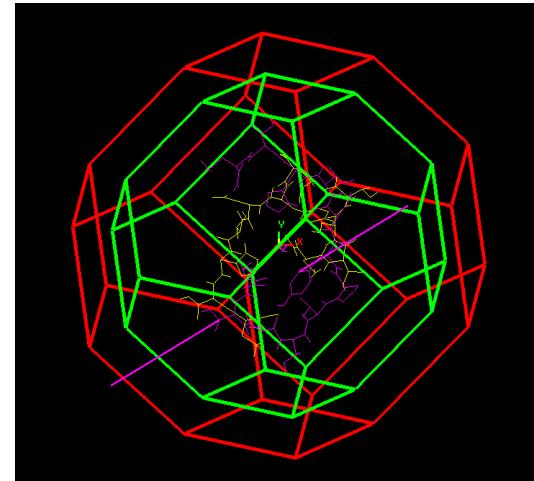


Molecular dynamics

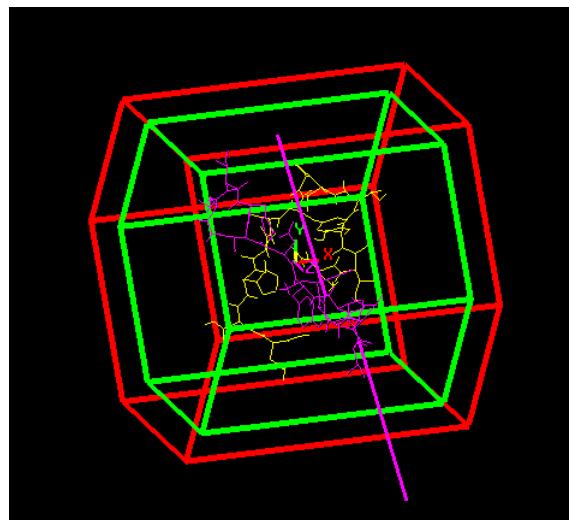
periodic boundary conditions



orthogonal



truncated octahedron



hexagonal

The Canonical Ensemble

In practice, we are rarely interested in the microcanonical ensemble, since real systems are seldom isolated from their environment.

The **canonical ensemble NVT** (*constant number of particles, volume, temperature*) and the **NPT ensemble** (*constant number of particles, pressure, temperature*) are the usual ensembles of interest.

The partition function for the canonical ensemble $Q_N(V, T)$, exponentially weights all points by their energy:

$$Q_N(V, T) = f(N) \int d\mathbf{r}^N d\mathbf{p}^N \exp \left[-\frac{\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)}{k_B T} \right]$$

The average of a property A is then:

$$\begin{aligned} \langle A \rangle_{NVT} &= \int d\mathbf{r}^N d\mathbf{p}^N \rho_{NVT}(\mathbf{r}^N, \mathbf{p}^N) \mathcal{A}(\mathbf{r}^N, \mathbf{p}^N) \\ &= \frac{f(N)}{Q_N} \int d\mathbf{r}^N d\mathbf{p}^N \mathcal{A}(\mathbf{r}^N, \mathbf{p}^N) \exp [-\beta \mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)] \end{aligned}$$

The Configuration Integral

The Hamiltonian is the sum of terms depending on the momenta \mathbf{p}^N and positions \mathbf{r}^N :

$$H(\mathbf{r}^N, \mathbf{p}^N) = K(\mathbf{p}^N) + V(\mathbf{r}^N) = \sum_i \frac{\mathbf{p}_i^2}{2m_i} + V(\mathbf{r}^N)$$

Therefore, the canonical weights can be factorized:

$$\exp[-\beta \mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)] = \exp\left[-\beta \sum_i^N \frac{\mathbf{p}_i^2}{2m_i}\right] \times \exp[-\beta \mathcal{V}(\mathbf{r}^N)]$$

Which allows the canonical partition function to be factorized:

$$Q_N = f(N) \int_{\text{phase space}} \exp(-\beta H(\mathbf{r}^N, \mathbf{p}^N)) d\mathbf{r}^N d\mathbf{p}^N = f(N) \int_{\text{momenta}} \exp(-\beta K(\mathbf{p}^N)) d\mathbf{p}^N \int_{\text{positions}} \exp(-\beta V(\mathbf{r}^N)) d\mathbf{r}^N$$

$$\exp[-A/k_B T] = Q_N = \boxed{(N! \Lambda^{3N})^{-1}} \boxed{Z_N}$$

Ideal gas part

$$\Lambda = \frac{h}{\sqrt{2\pi m k_B T}}$$

Configuration integral

$$Z_N = \int^V d\mathbf{r}^N \exp[-\beta \mathcal{V}(\mathbf{r}^N)]$$

The Configuration Integral

$$\exp[-A/k_B T] = Q_N = (N! \Lambda^{3N})^{-1} Z_N$$

Ideal gas part

$$\Lambda = \frac{h}{\sqrt{2\pi m k_B T}}$$

Configuration integral

$$Z_N = \int^V d\mathbf{r}^N \exp[-\beta \mathcal{V}(\mathbf{r}^N)]$$

The ideal gas part of the partition function is not very interesting!

All the information about the equilibrium distribution of positions is contained in the configuration integral, which is a function of *positions only*.

Although the time evolution of the positions and momenta are clearly coupled via the equations of motion, when averaging over the phase space they are statistically independent.

e.g. when choosing initial velocities from an equilibrium distribution, we do not have to consider the positions of the atoms

Constant-NVT molecular dynamics

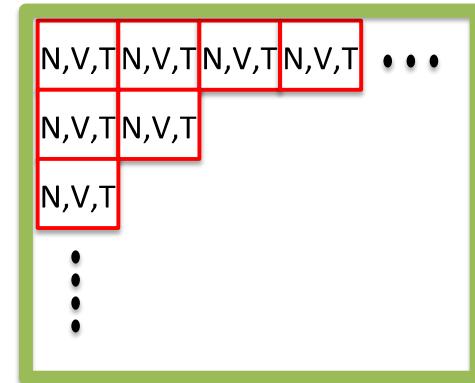
Newtonian dynamics gives us a microcanonical distribution.

How can we perform molecular dynamics simulations at constant temperature?

Formally, we could run a simulation of an ENORMOUS (i.e. macroscopic) system with Newtonian dynamics – once it has equilibrated, we know it will have approximately the canonical distribution at some temperature (*recall this is how the canonical distribution is derived*). In practice, we must simulate small, finite systems.

From a statistical mechanical point of view, we can impose a temperature on a system by bringing it into thermal contact with a large heat bath at temperature T.

This system can exchange energy with a large thermal bath. The total system is closed, so $U=E_{\text{bath}}+E_{\text{sys}}$ is fixed, and the system and bath are weakly coupled so we can ignore their interaction energy.



Constant-NVT molecular dynamics

Question: if we were to keep the average kinetic energy per particle constant (instantaneous temperature), would we simulate the true constant temperature ensemble?

For a classical system, the Maxwell-Boltzmann velocity distribution follows:

$$P(p) = \frac{1}{\sqrt{2\pi m}} e^{-p^2/(2m)}$$

$$k_B T = m \langle v_a^2 \rangle$$

As we mentioned before, this can be used to measure temperature in a constant-NVE simulation.

$$\langle p^2 \rangle = \int d\mathbf{p} p^2 P(p) = \frac{3m}{b} \quad \langle p^4 \rangle = \int d\mathbf{p} p^4 P(p) = 15 \left(\frac{m}{b} \right)^2$$

$$\frac{\langle p^2 \rangle^2}{\langle p^2 \rangle^2} \circ \frac{\langle p^4 \rangle - \langle p^2 \rangle^2}{\langle p^2 \rangle^2} = \frac{15(m/b)^2 - (3m/b)^2}{(3m/b)^2} = \frac{2}{3}$$

Constant-NVT molecular dynamics

Question: if we were to keep the average kinetic energy per particle constant (instantaneous temperature), would we simulate the true constant temperature ensemble?

For a classical system, the Maxwell-Boltzmann velocity distribution follows:

$$P(p) = \frac{1}{\sqrt{2\pi m}} e^{-p^2/(2m)}$$

$$k_B T = m \langle v_a^2 \rangle$$

As we mentioned before, this can be used to measure temperature in a constant-NVE simulation.

$$\langle p^2 \rangle = \int d\mathbf{p} p^2 P(p) = \frac{3m}{b} \quad \langle p^4 \rangle = \int d\mathbf{p} p^4 P(p) = 15 \frac{m^2}{b}$$

$$\frac{\langle p^2 \rangle^2}{\langle p^4 \rangle^2} \circ \frac{\langle p^4 \rangle - \langle p^2 \rangle^2}{\langle p^2 \rangle^2} = \frac{15(m/b)^2 - (3m/b)^2}{(3m/b)^2} = \frac{2}{3}$$

CONSTANT TEMPERATURE DOES NOT
IMPLY THAT THE KINETIC ENERGY PER
PARTICLE IS CONSTANT!

Constant-NVT molecular dynamics

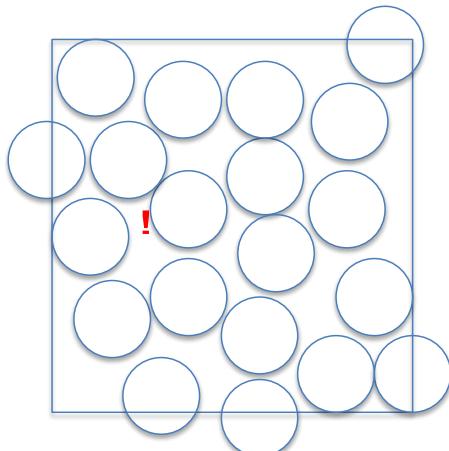
So how should we proceed?

We need to couple our system to a heat bath that imposes the desired temperature.

There is no really “correct” way of doing this; thermostats for doing so can be broadly classified as stochastic (e.g. Andersen, Langevin) or deterministic (e.g. Nose-Hoover).

stochastic algorithm idea:

Represent this ‘coupling’ by stochastic impulsive forces that act occasionally on randomly selected particles - think of this as a Monte Carlo move that changes the energy.



Between these moves, the system evolves at constant energy according to the Newtonian laws of motion.

The stochastic collisions ensure that all accessible constant-energy surfaces are each visited according to their Boltzmann weight.

Molecular dynamics

T control

$$\langle E_{kin} \rangle = \sum_{i=1}^N \frac{|p_i|^2}{2m} = \frac{k_B T}{2} (3N - N_c)$$

Scaling

$$\Delta T = \frac{1}{2} \sum_{i=1}^N \frac{2}{3} \frac{2m(\lambda v_i)^2}{Nk_B} = (\lambda^2 - 1)T(t)$$

$$\lambda = \sqrt{\frac{T_{target}}{T(t)}}$$

Molecular dynamics

T control

Berendsen thermostat

$$\frac{dT(t)}{dt} = \frac{1}{\tau} (T_{bath} - T(t))$$

τ : coupling parameter (with the bath)

$$\Delta T = \frac{\delta t}{\tau} (T_{bath} - T(t))$$

$\tau \sim 0.5 \text{ ps}$, $\delta t \sim 1 \text{ fs}$

$$\lambda^2 = 1 + \frac{\delta t}{\tau} \left(\frac{T_{bath}}{T(t)} - 1 \right)$$

distinct scaling for protein and solvent

problem: not strictly canonical ensemble

Molecular dynamics

T control

Nosé-Hoover thermostat /canonical ensemble/

new degree of freedom (s)

$$E_{pot} = (f+1)k_B T \ln s$$

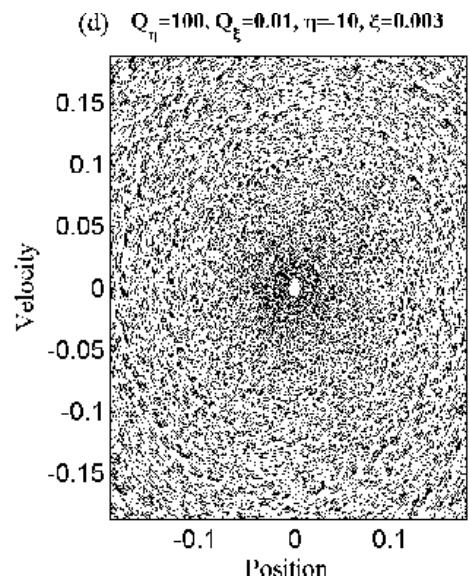
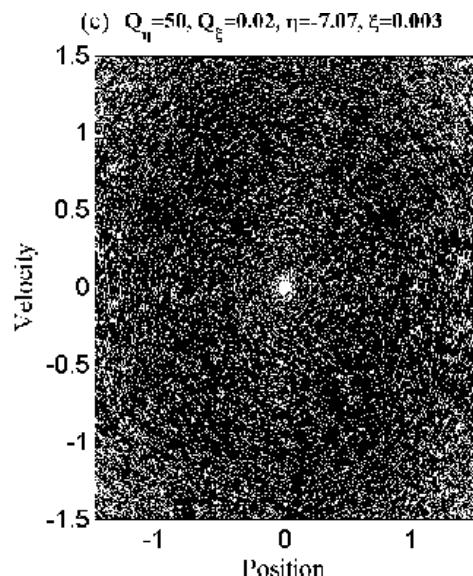
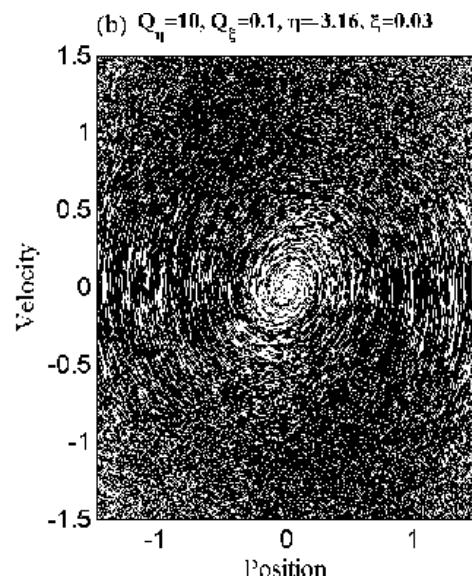
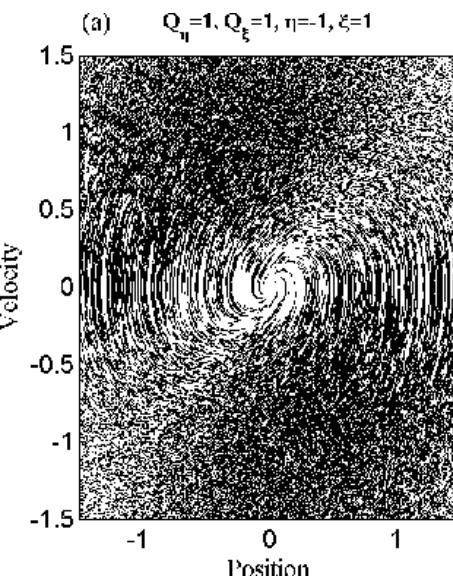
Q : fictitious mass of s

$$E_{kin} = (Q/2)(ds/dt)^2$$

$$V_i = s \frac{dr_i}{dt}$$

$$\delta t = s \delta t$$

timestep can change as s can change



The Andersen Thermostat

The Andersen Thermostat is a simple stochastic thermostat where velocities are simply reassigned from the Maxwell-Boltzmann distribution.

1. Particles undergo “collisions” with a frequency v
2. At each time step, the probability of a particle experiencing a collision is $v*dt$. Particles are chosen for collision if a random number ρ on the interval $[0;1]$ is less than $v\delta t$.
3. Any chosen particle has all components of its velocity reassigned from a Maxwell-Boltzmann distribution corresponding to the desired temperature T .

The frequency v is dictated by the coupling strength between the bath and the system. Guaranteed to give canonical distribution, but reassignment of velocities strongly affects dynamics (velocities after collisions are not correlated with those before) – not good for computation of time-dependent properties (diffusion, etc...).

Useful for sampling conformational space!!!

Molecular dynamics

p control

Volume adjustments

compressibility

scaling the volume,

isotropic

anisotropic

$$P = \frac{1}{V} \left[Nk_B T - \frac{1}{3k_B T} \sum_{i=1}^N \sum_{j=i+1}^N r_{ij} \frac{dE(r_{ij})}{dr_{ij}} \right]$$

$$\kappa = \frac{1}{V} \left(\frac{\partial V}{\partial P} \right)_T = \frac{1}{k_B T} \frac{\langle V^2 \rangle - \langle V \rangle^2}{\langle V^2 \rangle}$$

$$\frac{dP(t)}{dt} = \frac{1}{\tau_p} (P_{bath} - P(t))$$

$$\lambda = 1 - \kappa \frac{\delta t}{\tau_p} (P - P_{bath})$$

$$r'_i = \lambda^{1/3} r_i$$

Molecular dynamics

p control

Volume adjustments

compressibility

scaling the volume,

isotropic

anisotropic

$$P = \frac{1}{V} \left[Nk_B T - \frac{1}{3k_B T} \sum_{i=1}^N \sum_{j=i+1}^N r_{ij} \frac{dE(r_{ij})}{dr_{ij}} \right]$$

$$\kappa = \frac{1}{V} \left(\frac{\partial V}{\partial P} \right)_T = \frac{1}{k_B T} \frac{\langle V^2 \rangle - \langle V \rangle^2}{\langle V^2 \rangle}$$

$$\frac{dP(t)}{dt} = \frac{1}{\tau_p} (P_{bath} - P(t))$$

$$\lambda = 1 - \kappa \frac{\delta t}{\tau_p} (P - P_{bath})$$

$$r'_i = \lambda^{1/3} r_i$$