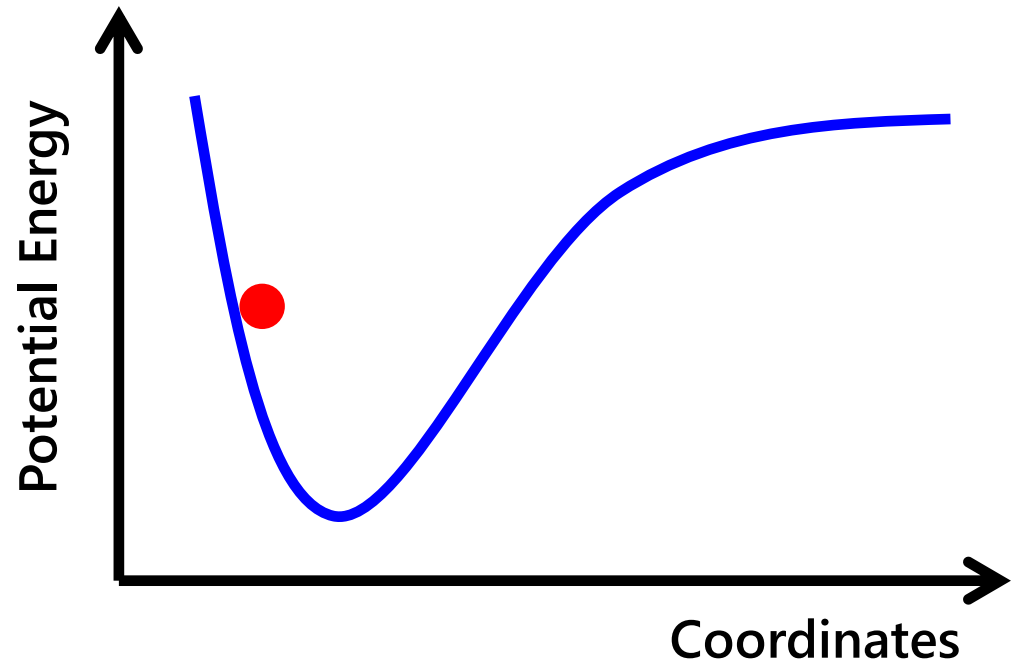
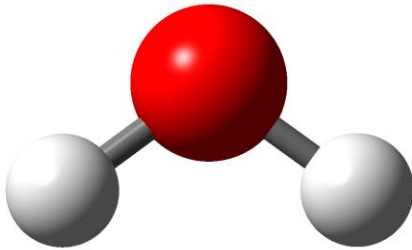




Molecular Dynamics

Macromolecular Systems

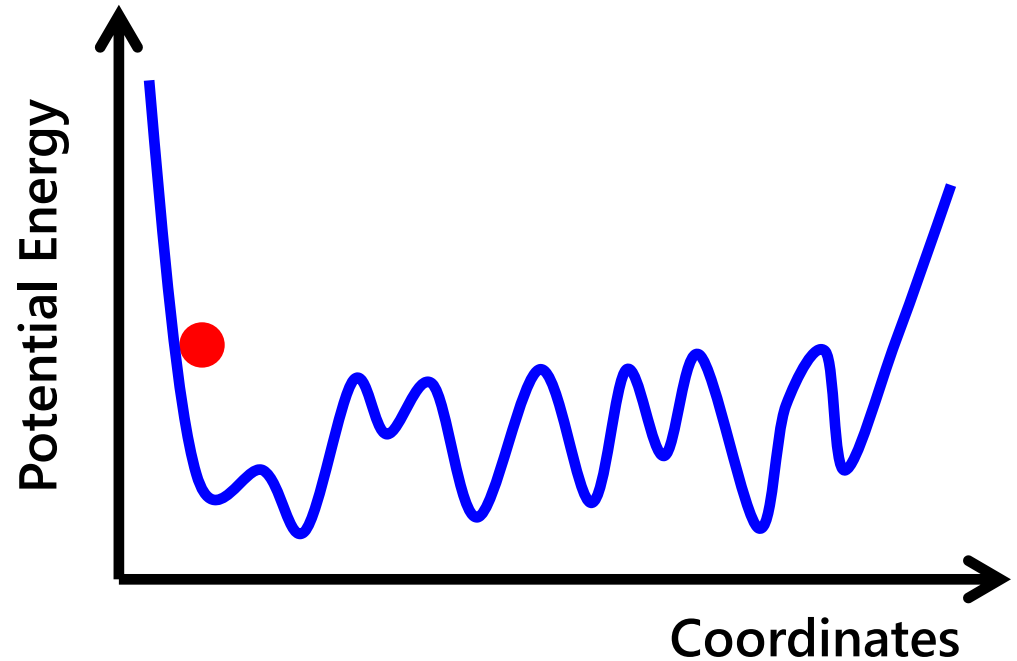
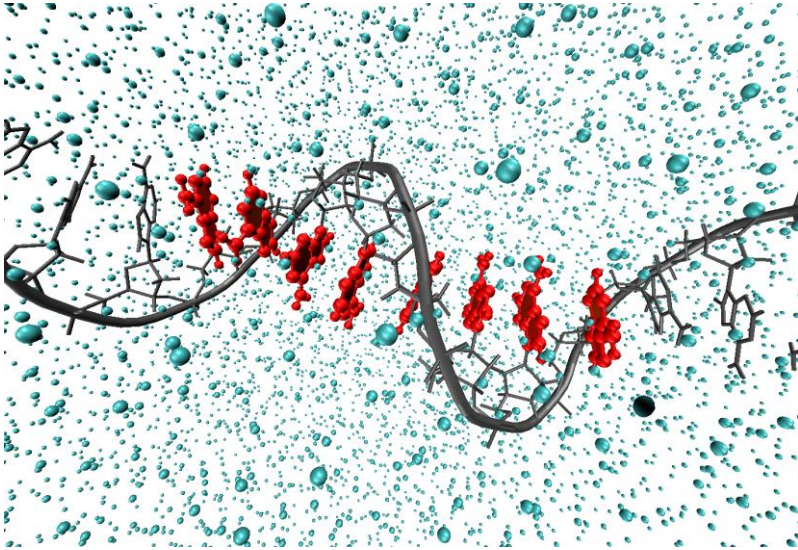
Small-Size Systems



- The potential energy surface has a well-defined minimum
- Usual approach: properties of the molecule are computed at the minimum-energy geometry

Macromolecular Systems

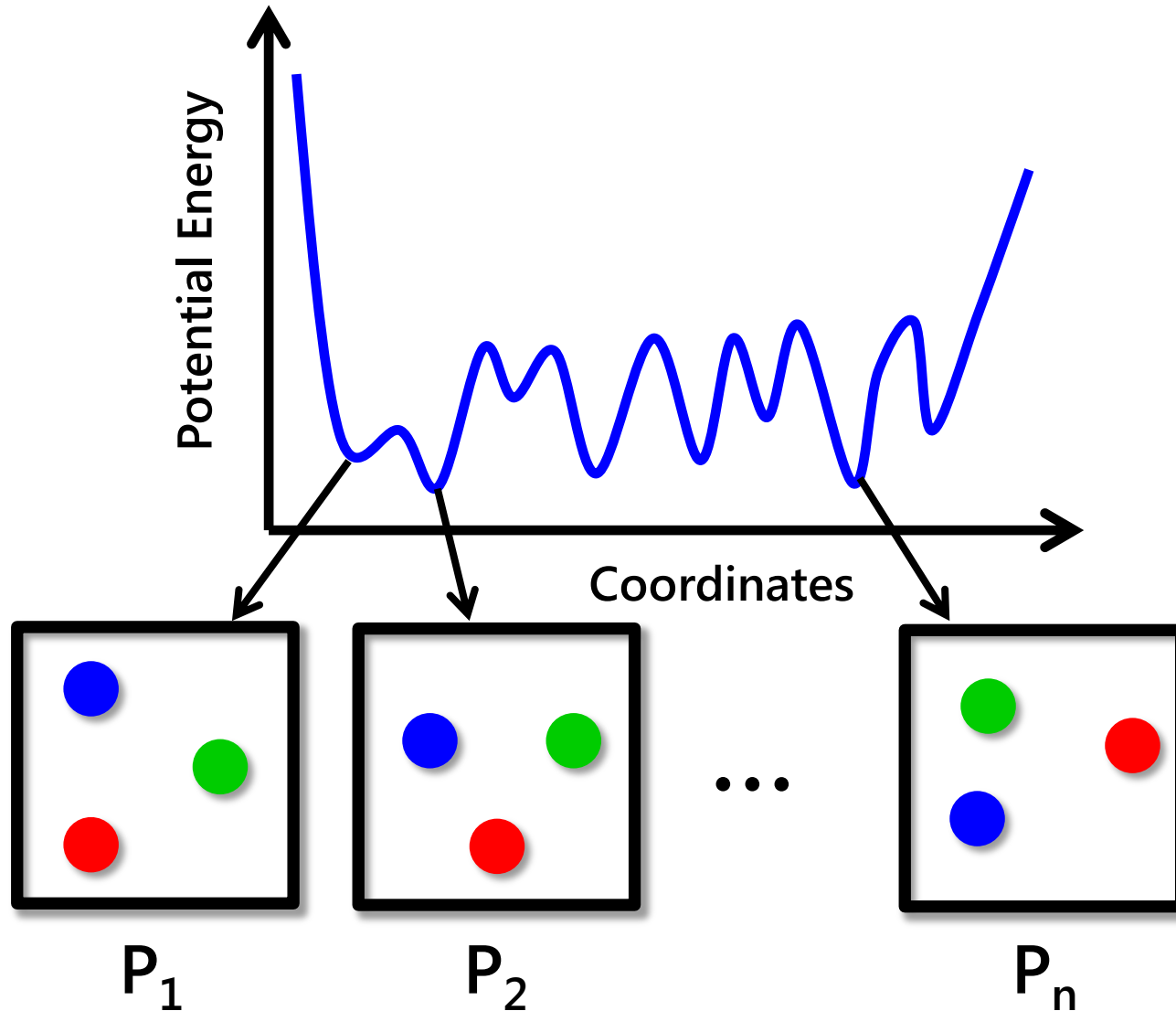
Large-Size Systems (Macromolecules)



- The potential energy surface presents several thermally accessible minima
- An efficient sampling of the potential energy surface is necessary

Macromolecular Systems

- Goal: to generate all geometry configurations of the system



Classical Molecular Dynamics

Classical Equations of Motion (Hamilton Equations)

$$F = m \cdot a = m \frac{dv}{dt} = \frac{dp}{dt} \quad (1) \quad \text{Newton equation}$$

If the force acting on the atoms of the system is conservative (it depends only on the position of the atoms), the force is related to the potential energy by:

$$F = -\frac{dV}{dq} \quad (2) \quad \text{By combining equations (1) and (2):} \quad \boxed{\frac{dp}{dt} = -\frac{dV}{dq}} \quad (3)$$

$$T = \frac{1}{2}mv^2 = \frac{1}{2}\frac{m^2}{m}v^2 = \frac{p^2}{2m} \quad (4)$$

By differentiating the kinetic energy with respect to the linear momentum:

$$\frac{dT}{dp} = \frac{2p}{2m} = \frac{p}{m} = \frac{m \cdot v}{m} = v = \frac{dq}{dt} \quad (5) \quad \boxed{\frac{dq}{dt} = \frac{dT}{dp}} \quad (6)$$

Classical Molecular Dynamics

Equations (8) and (15) form the Velocity Verlet algorithm:

$$q(t + \Delta t) = q(t) + \Delta t v(t) - \frac{\Delta t^2}{2!} \frac{dV(t)}{dq} \quad (8)$$

$$v(t + \Delta t) = v(t) - \frac{\Delta t}{2m} \left(\frac{dV(t + \Delta t)}{dq} + \frac{dV(t)}{dq} \right) \quad (15)$$

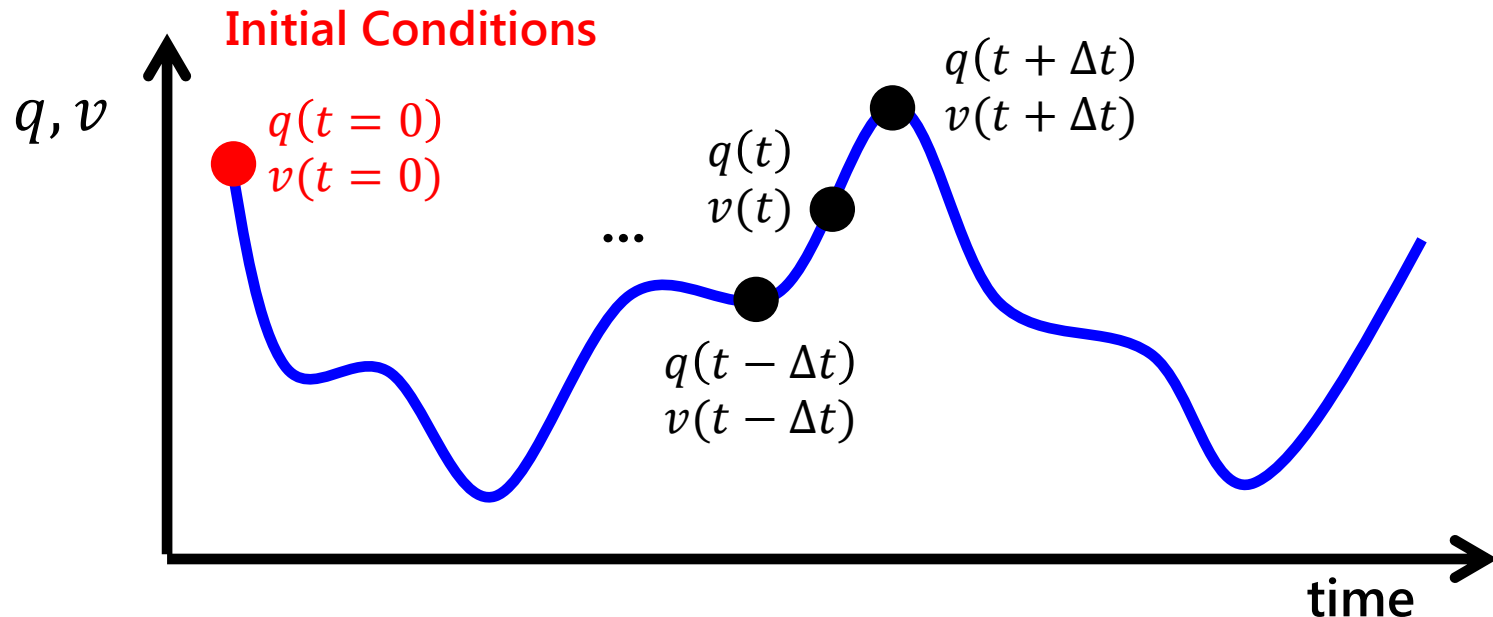
1. Calculation of $\frac{dV(t)}{dq}$
2. Calculation of $q(t + \Delta t)$ from equation (8)
3. Calculation of $\frac{dV(t+\Delta t)}{dq}$
4. Calculation of $v(t + \Delta t)$ from equation (15)
5. Repetition of steps 1 to 4 until the end of the simulation

Classical Molecular Dynamics

Equations (8) and (15) form the Velocity Verlet algorithm:

$$q(t + \Delta t) = q(t) + \Delta t v(t) - \frac{\Delta t^2}{2!} \frac{dV(t)}{dq} \quad (8)$$

$$v(t + \Delta t) = v(t) - \frac{\Delta t}{2m} \left(\frac{dV(t + \Delta t)}{dq} + \frac{dV(t)}{dq} \right) \quad (15)$$



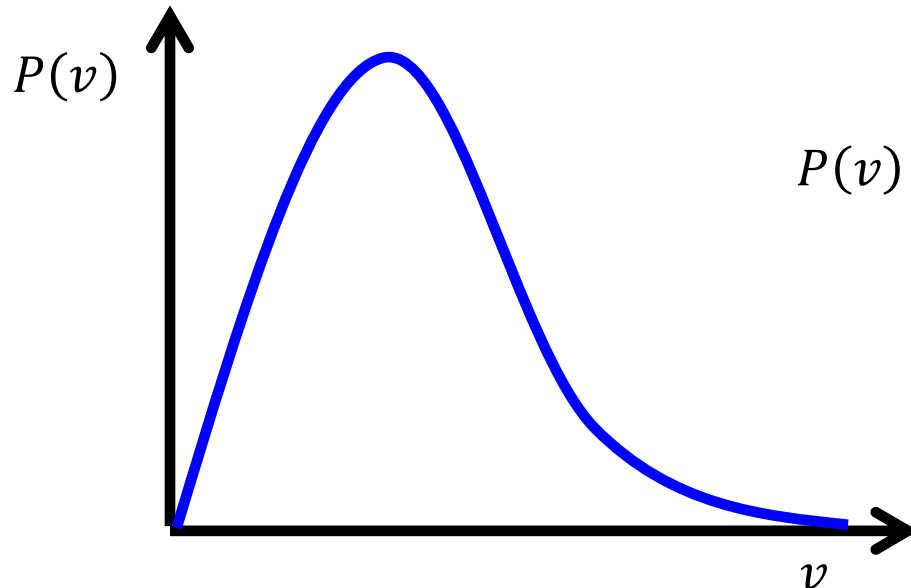
Classical Molecular Dynamics

- Initial coordinates $q(t = 0)$ are taken from experiments or are guessed from a theoretical model

Protein Data Bank: <http://www.rcsb.org>

Currently, experimental structures of 128962 systems are available (DNA strands, proteins, enzymes ...)

- Initial velocities $v(t = 0)$ are taken from a Maxwell-Boltzmann distribution at the desire temperature (assumption of thermal equilibrium)



$$P(v) = \left(\frac{m}{2\pi kT}\right)^{3/2} 4\pi v^2 e^{-\frac{mv^2}{2kT}} \quad (16)$$

Classical Molecular Dynamics

Resolution of the equations of motion requires the computation of the derivative of the potential energy with respect to the coordinates (gradient of the potential energy)

$$q(t + \Delta t) = q(t) + \Delta t v(t) - \frac{\Delta t^2}{2!} \frac{dV(t)}{dq} \quad (8)$$

$$v(t + \Delta t) = v(t) - \frac{\Delta t}{2m} \left(\frac{dV(t + \Delta t)}{dq} + \frac{dV(t)}{dq} \right) \quad (15)$$

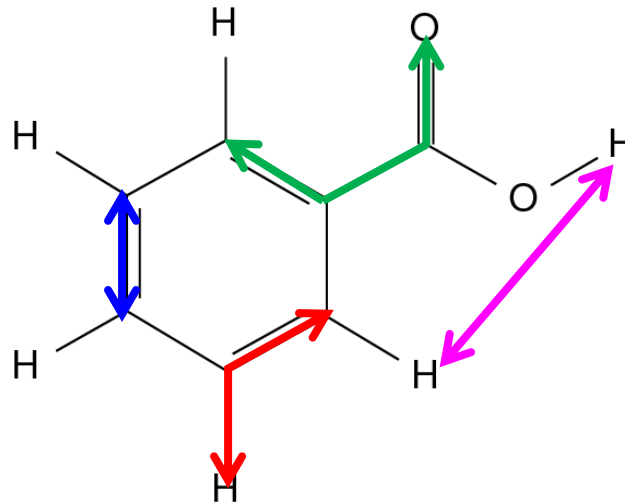
The gradient can be computed by three different ways:

- By using a force field → Classical Molecular Dynamics
- By solving the Schrödinger equation $\hat{H}\psi = E\psi$ → Ab initio Molecular Dynamics
- By a combination of both → Quantum Mechanics/Molecular Mechanics (QM/MM) Molecular Dynamics

Force Fields

A force field is a simple equation that relates the potential energy of the system with its internal coordinates (bond distances, bond angles ...)

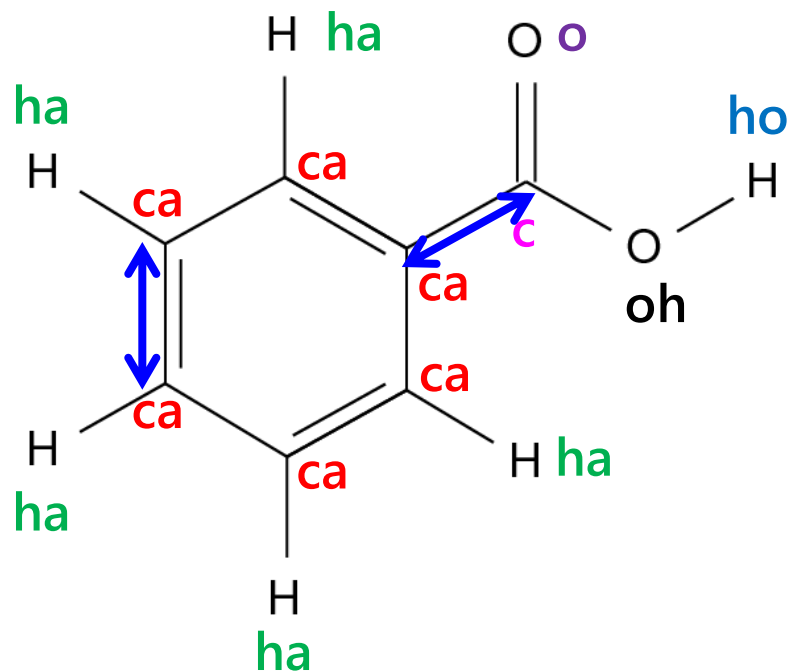
In the most used force fields, the potential energy is split into bonded and non-bonded interactions



$$V_{total} = \underbrace{\sum_{i=1}^{N_{bond}} V_{bond} + \sum_{i=1}^{N_{angle}} V_{angle} + \sum_{i=1}^{N_{dihed}} V_{dihed}}_{\text{Bonded Interactions}} + \underbrace{\sum_{i=1}^{N_{nb}} V_{nb}}_{\text{Non-Bonded Interactions}} \quad (17)$$

Force Fields

- The atoms of the molecule are classified in different atom types to distinguish interactions between the same chemical class of atoms



ca: sp^2 C in aromatic system

ha: H bonded to aromatic C

c: sp^2 C of carbonyl group

o: O with one connected atom

oh: O in hydroxyl group

ho: H in hydroxyl group

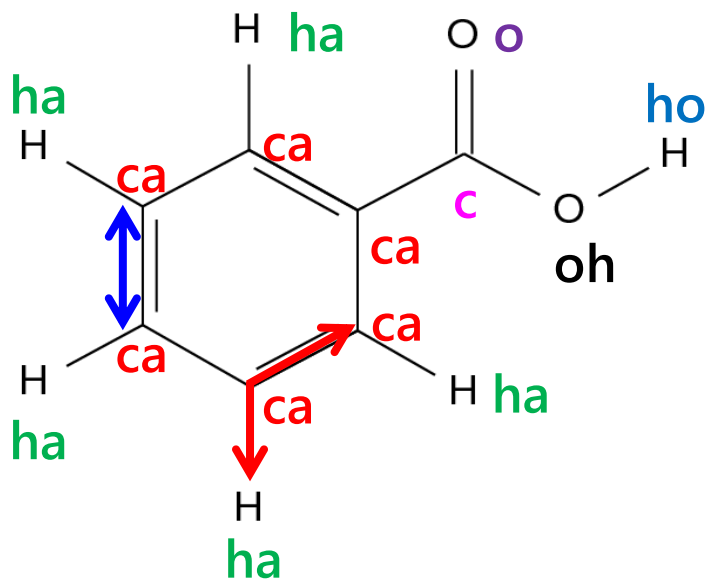
General Amber Force Field (GAFF)

J. Wang et al. *J. Comput. Chem.* **2004**, 15, 1157

- The parameters of the force field are different for each atom type.
- For example, the carbon **ca** has different parameters than the carbon **c**.

Force Fields

$$V_{total} = \sum_{i=1}^{N_{bond}} V_{bond} + \sum_{i=1}^{N_{angle}} V_{angle} + \sum_{i=1}^{N_{dihed}} V_{dihed} + \sum_{i=1}^{N_{nb}} V_{nb}$$

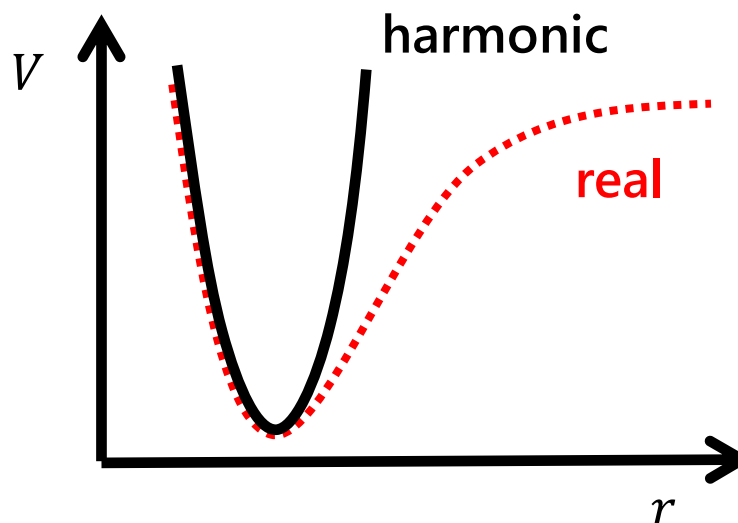


- Bond distances and angles are described by harmonic potentials:

$$V_{bond} = \frac{1}{2} k_b (r - r_0)^2 \quad (18)$$

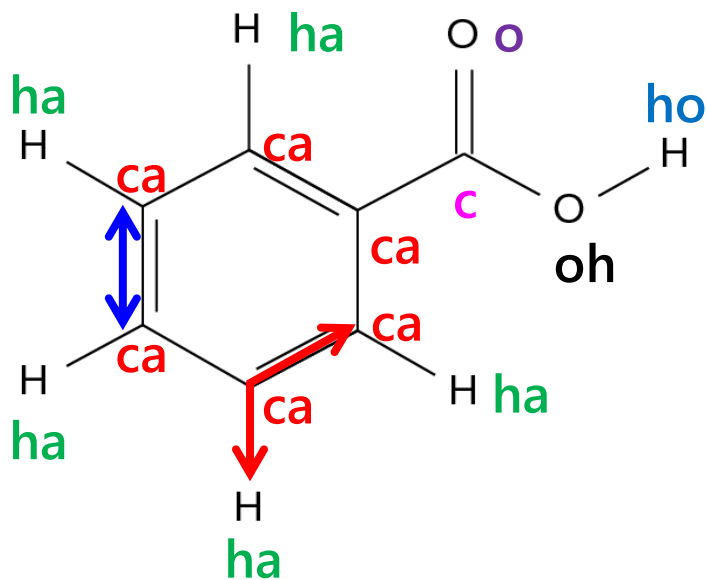
$$V_{angle} = \frac{1}{2} k_a (\theta - \theta_0)^2 \quad (19)$$

- Harmonic potentials are correct only close to the equilibrium geometry
- Harmonic potentials canNOT describe bond breaking and formation!



Force Fields

$$V_{total} = \sum_{i=1}^{N_{bond}} V_{bond} + \sum_{i=1}^{N_{angle}} V_{angle} + \sum_{i=1}^{N_{dihed}} V_{dihed} + \sum_{i=1}^{N_{nb}} V_{nb}$$



- Bond distances and angles are described by harmonic potentials:

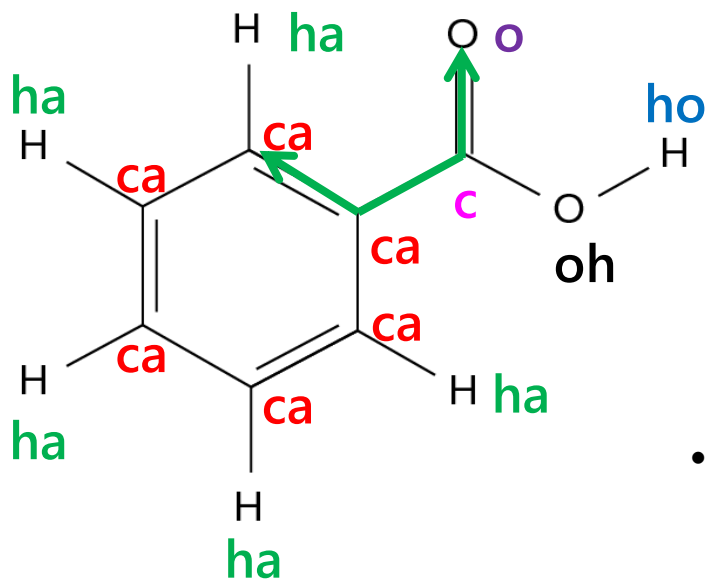
$$V_{bond} = \frac{1}{2} k_b (r - r_0)^2 \quad (18)$$

$$V_{angle} = \frac{1}{2} k_a (\theta - \theta_0)^2 \quad (19)$$

- Bond types: **ca-ca**, **ca-ha**, **ca-c**, **c-o**, **c-oh**, **oh-ho**
- Angle types: **ca-ca-ca**, **ca-ca-ha**, **ca-ca-c**, **ca-c-o**, **ca-c-oh**, **c-oh-ho**
- Each bond and angle type has different parameters k_b , r_0 , k_a and θ_0

Force Fields

$$V_{total} = \sum_{i=1}^{N_{bond}} V_{bond} + \sum_{i=1}^{N_{angle}} V_{angle} + \sum_{i=1}^{N_{dihed}} V_{dihed} + \sum_{i=1}^{N_{nb}} V_{nb}$$



- Torsion or 1-4 interactions are described by Fourier potentials:

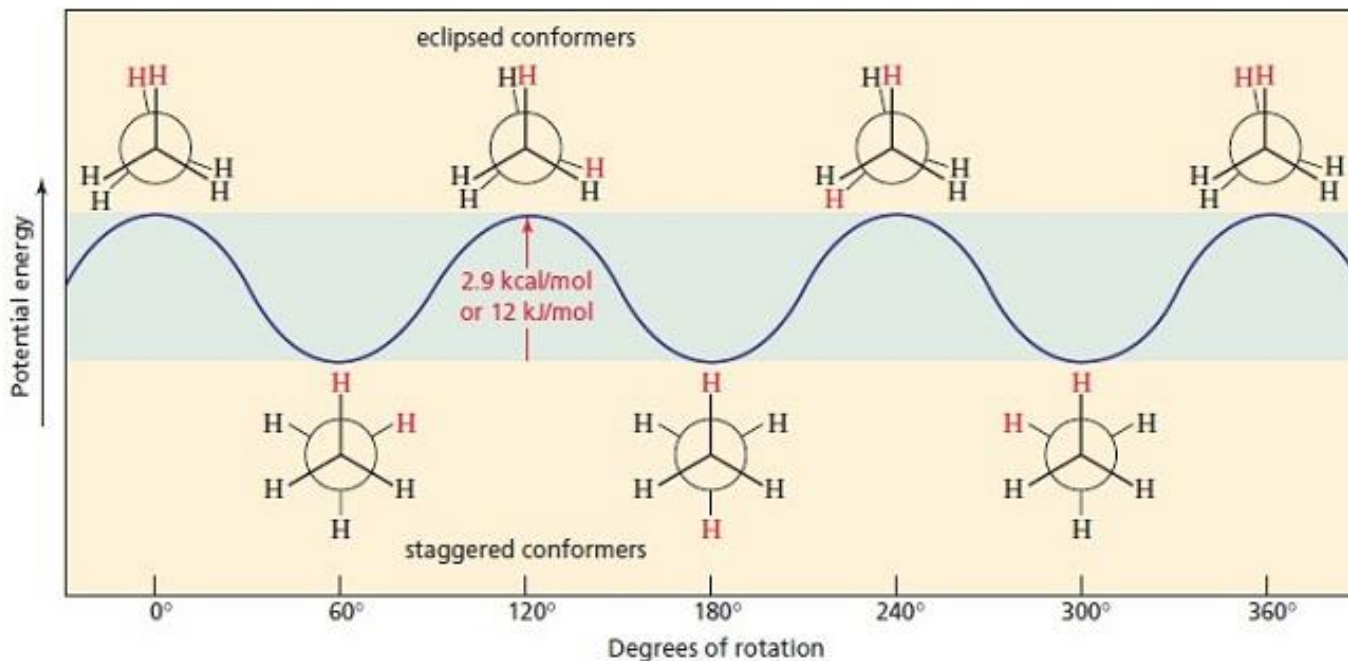
$$V_{dihed} = k_t (1 + \cos(n\omega - \gamma)) \quad (20)$$

- ω is the angle formed by the planes defined by the atoms 1-2-3 and 2-3-4
- k_t , n and γ are parameters that define the shape of the potential
- Torsion types: **ca-ca-ca-ca**, **ca-ca-ca-ha**, **ha-ca-ca-ha**, **ha-ca-ca-c**, **ca-ca-c-oh**, ...

Classical Molecular Dynamics

Torsion potential for ethane

<http://www.kshitij-iitjee.com/Conformations-of-Alkanes>



$$V_{dihed} = k_t(1 + \cos(n\omega - \gamma)) \quad (20)$$

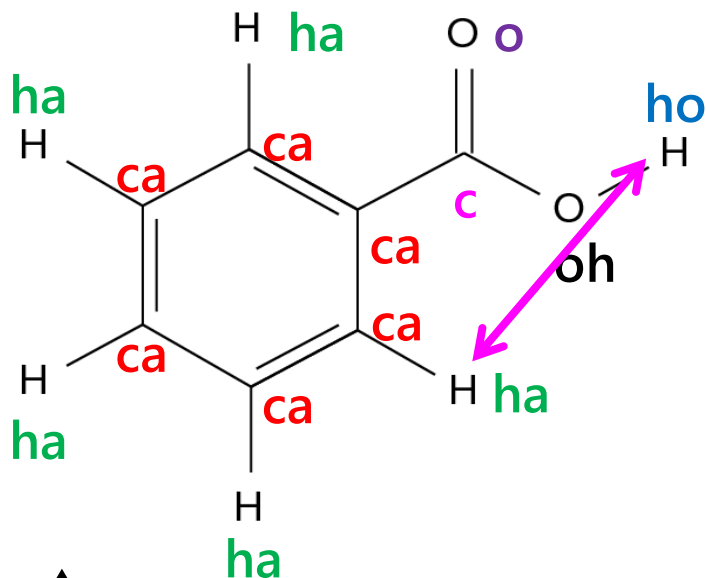
n : number of minima ($n = 3$)

k_t : height of the barrier ($k_t = 2.9$ kcal/mol)

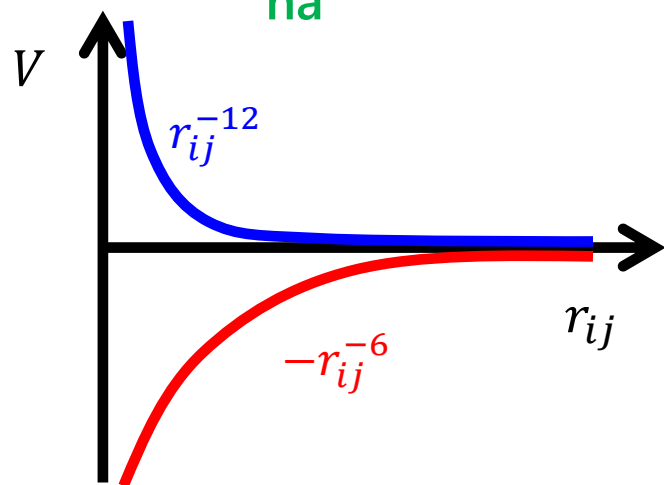
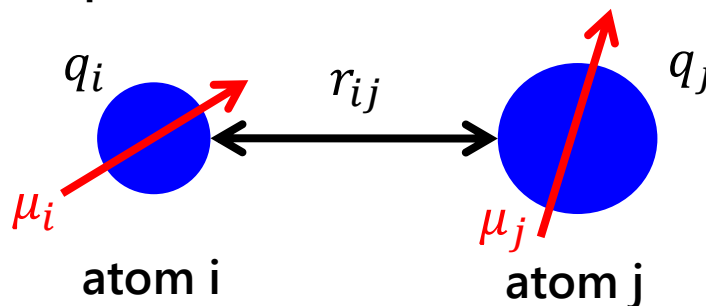
γ : angle that determines the position of the minima. For example, if n is odd and $\gamma = 0$ or if n is even and $\gamma = 180$ the minimum is located at 180° .

Force Fields

$$V_{total} = \sum_{i=1}^{N_{bond}} V_{bond} + \sum_{i=1}^{N_{angle}} V_{angle} + \sum_{i=1}^{N_{dihed}} V_{dihed} + \sum_{i=1}^{N_{nb}} V_{nb}$$



- Non-bonded interactions have three components: Coulomb, van der Waals and repulsion interactions

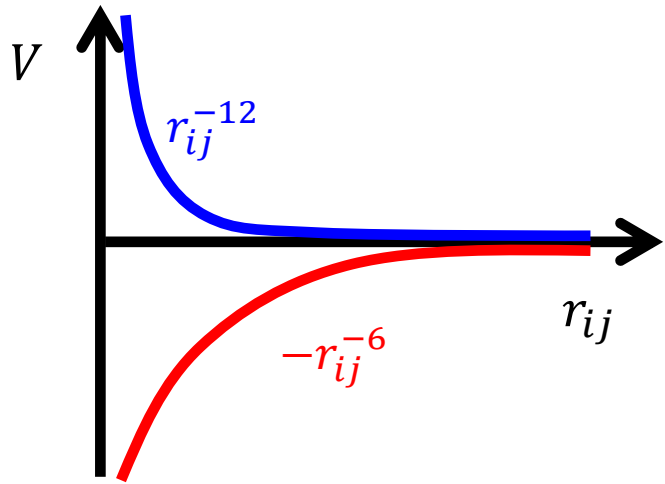


$$V_C = \frac{q_i q_j}{4\pi\epsilon r_{ij}} \quad (21)$$

$$V_{vdW} \propto -\frac{1}{r_{ij}^6} \quad (22)$$

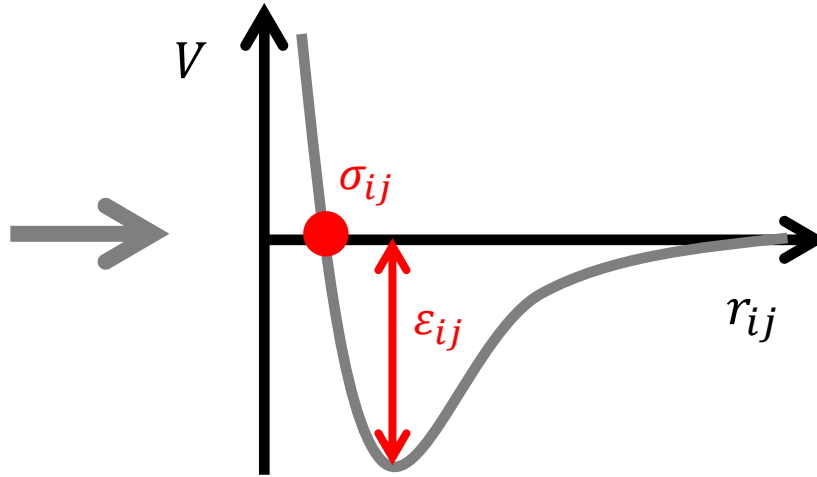
$$V_{rep} \propto e^{-r_{ij}} \approx \frac{1}{r_{ij}^{12}} \quad (23)$$

Force Fields



$$V_{vdW} \propto -\frac{1}{r_{ij}^6} \quad (22)$$

$$V_{rep} \propto e^{-r_{ij}} \approx \frac{1}{r_{ij}^{12}} \quad (23)$$



Both interactions are modelled by the Lennard-Jones potential:

$$V_{LJ} = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (24)$$

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j} \quad (25) \quad \sigma_{ij} = \frac{\sigma_i + \sigma_j}{2} \quad (26)$$

The parameters ϵ_i , ϵ_j , σ_i and σ_j are different for each atom type.

Force Fields

$$V_{total} = \sum_{i=1}^{N_{bond}} V_{bond} + \sum_{i=1}^{N_{angle}} V_{angle} + \sum_{i=1}^{N_{dihed}} V_{dihed} + \sum_{i=1}^{N_{nb}} V_{nb} \quad (17)$$

$$V_{total} = \sum_{i=1}^{N_{bond}} \frac{1}{2} k_b (r - r_0)^2 + \sum_{i=1}^{N_{angle}} \frac{1}{2} k_a (\theta - \theta_0)^2 + \sum_{i=1}^{N_{dihed}} k_t (1 + \cos(n\omega - \gamma)) + \sum_{i>j} \frac{q_i q_j}{4\pi\epsilon r_{ij}} + 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (27)$$

- Equation (27) relates the potential energy of the system with the geometry
- The parameters $k_b, r_0, k_a, \theta_0, k_t, n, \gamma, q_i, \epsilon_{ij}$ and σ_{ij} are chosen to reproduce experimental data or quantum mechanical calculations

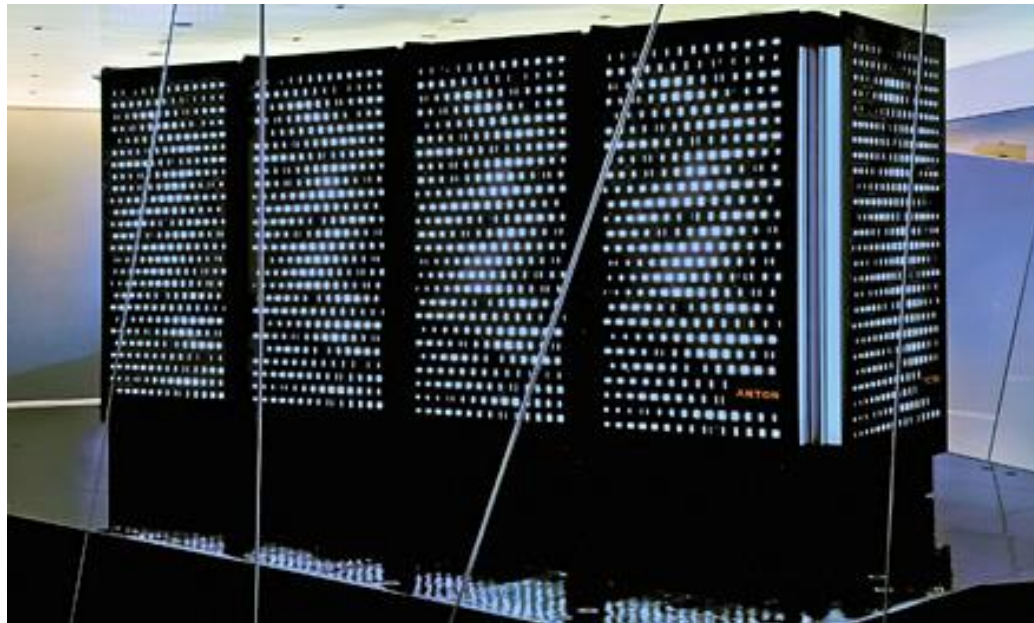
- We need the potential energy to solve the classical equations of motion (Velocity Verlet)

$$q(t + \Delta t) = q(t) + \Delta t v(t) - \frac{\Delta t^2}{2!} \frac{dV(t)}{dq} \quad (8)$$

$$v(t + \Delta t) = v(t) - \frac{\Delta t}{2m} \left(\frac{dV(t + \Delta t)}{dq} + \frac{dV(t)}{dq} \right) \quad (15)$$

Periodic Boundary Conditions

- Currently simulations of systems composed by up to 10^6 atoms can be simulated
- The longest computational time is around 1 ms (10^{-3} s)
- In general, the time step during simulations is 2 fs (10^{-15} s) $\rightarrow 10^{12}$ time steps!



<https://en.wikipedia.org/>

Anton supercomputer of David Shaw Research company in New York

Periodic Boundary Conditions

- The development of graphic processing units (GPUs) for computing (and not only for gaming) during the last years allow simulating larger systems and longer times

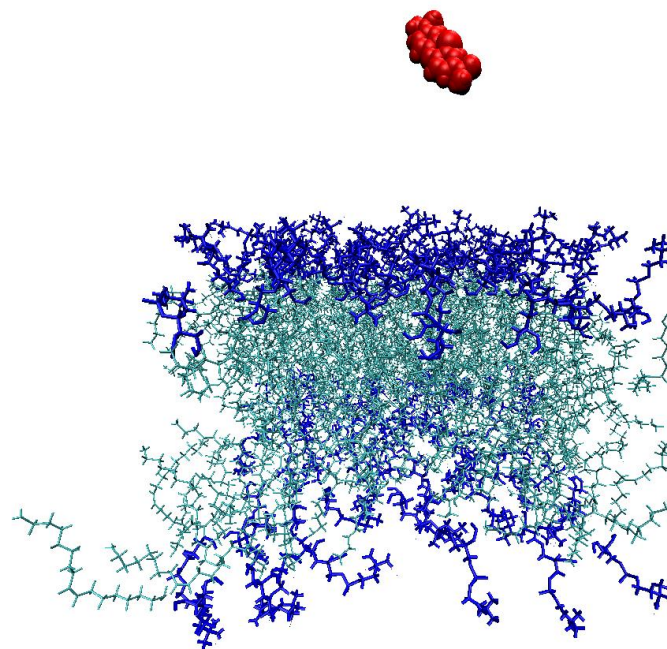


GPU

VS



CPU



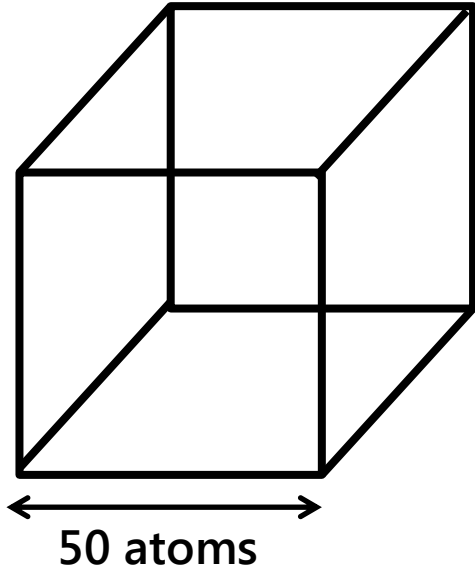
Classical MD with 97023 atoms (Amber16)

- GPU: 32 ns/day
- CPU: 0.13 ns/day

GPU is 246 times faster!

Periodic Boundary Conditions

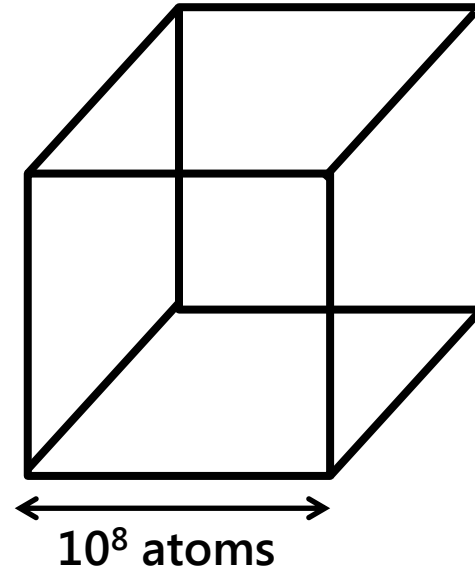
- We are still far from being able to simulate "real" macroscopic systems with $\sim 10^{23}$ atoms



$$50 \times 50 \times 50 = 125000 \text{ atoms}$$

$$50 \times 50 \times 6 = 15000 \text{ atoms on the surface of the cube}$$

12% of the atoms interact with the vacuum

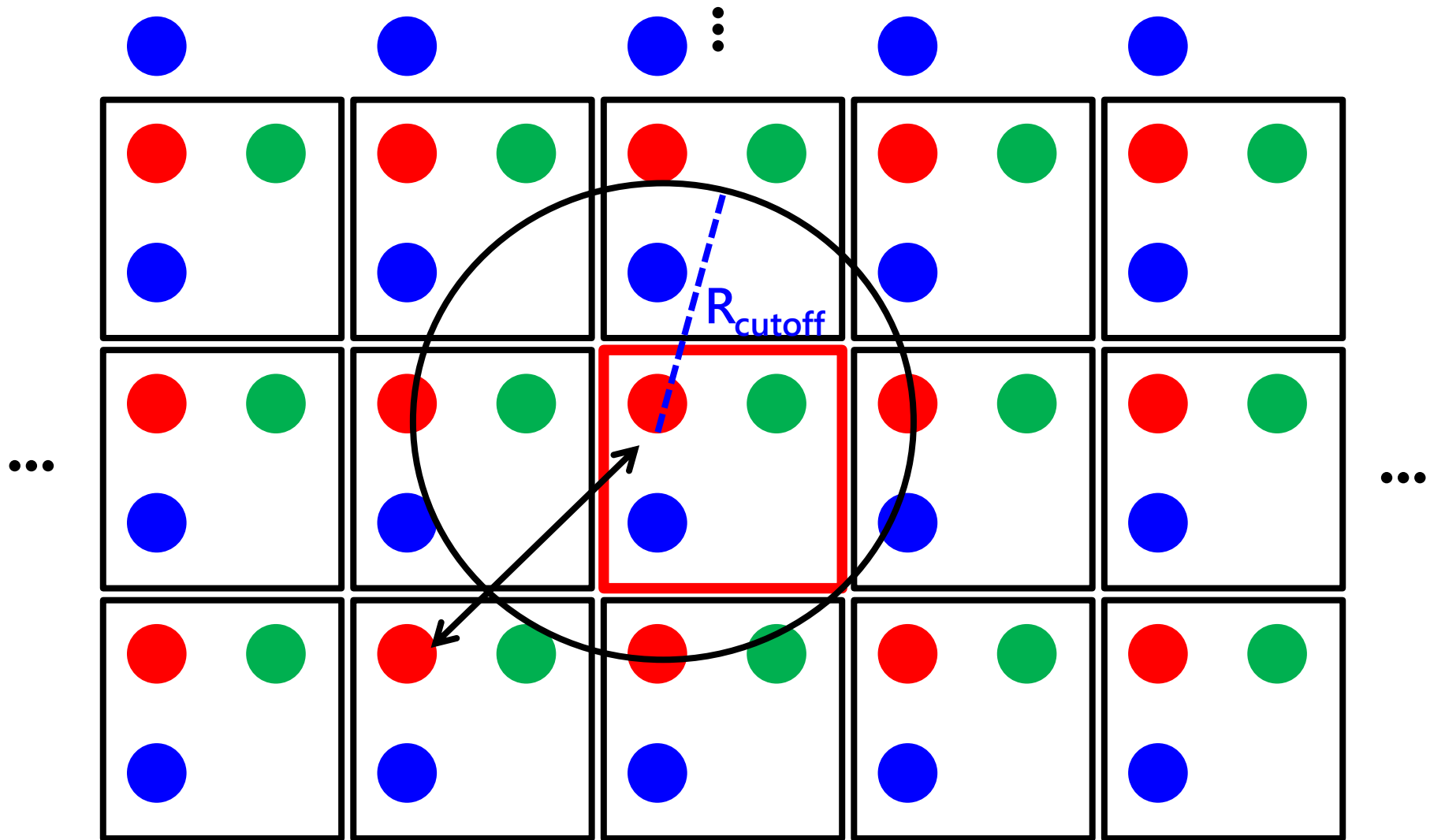


$$10^8 \times 10^8 \times 10^8 = 10^{24} \text{ atoms}$$

$$10^8 \times 10^8 \times 6 = 6 \times 10^{16} \text{ atoms on the surface of the cube}$$

Only 0.000006% of the atoms interact with the vacuum!

Periodic Boundary Conditions



Primary cell

20000 atoms
100 nearest image cells

10^{12} interactions
($\sim N_{\text{atoms}}^2$)

Experimental Conditions: P and T

Control of Temperature: Thermostats

Temperature is related with the kinetic energy (and velocities and momenta) of the atoms of our system

$$K = \frac{1}{2} \sum_i m_i v_i^2 = \frac{1}{2} \sum_i \frac{p_i^2}{m_i} \quad (33)$$

\downarrow
 $p = mv$

$$K = \frac{3}{2} kT \quad (34)$$

$$T = \frac{1}{3k} \sum_i \frac{p_i^2}{m_i} \quad (35)$$

Thermostats control the temperature by scaling in different ways the momenta (or velocities)

- Velocity Scaling

$$v_i^{new} = v_i^{old} \lambda \quad (36)$$

$$\lambda = \left(\frac{T_d}{T_{MD}(t)} \right)^{1/2} \quad (37)$$

Velocities are scaled every time step or every N steps

Experimental Conditions: P and T

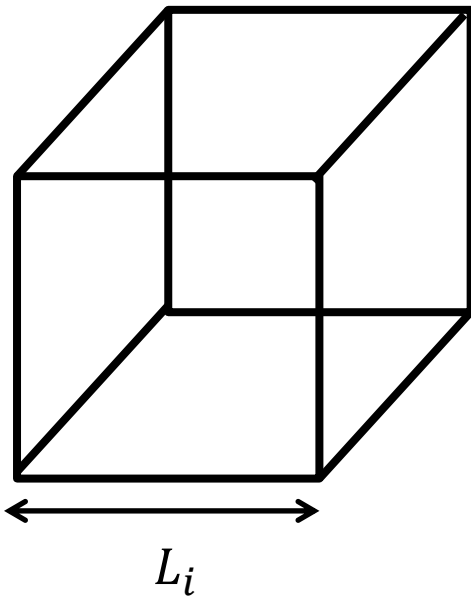
Control of Pressure: Barostats

The pressure and volume of a system are related by the virial equation of state:

$$\frac{PV_m}{RT} = 1 + \frac{B}{V_m} + \frac{C}{V_m^2} + \dots \quad (41)$$

The pressure can be controlled by modifying the volume of the simulation box.

- Berendsen Barostat



The box vectors L are scaled by a factor μ that depends on the pressure:

$$L_i^{new} = L_i^{old} \mu \quad (42)$$

$$\mu = \left(1 - \frac{\Delta t}{\tau} \gamma (P_d - P_{MD}(t)) \right)^{1/3}$$

γ is the thermal compressibility

The "rise time" τ describes the strength of the coupling to the pressure bath. The larger τ , the weaker the coupling