

Molecular dynamics and Monte Carlo methods

Agricultural Bioinformatics Research Unit,
Graduate School of Agricultural and Life
Sciences, The University of Tokyo
Tohru Terada

Today's topics

- Equation of motion
- Molecular dynamics (MD) methods
 - Exercise 1
- Comparability with experimental data
- Monte Carlo methods
 - Exercise 2
- Constant-temperature MD methods

Review of classical mechanics

- Newton's equation of motion

$\mathbf{F} = m\mathbf{a}$ \mathbf{F} : force, m : mass, \mathbf{a} : acceleration

\mathbf{F} and \mathbf{a} are 3D vectors → bold

m is a scalar number → italic

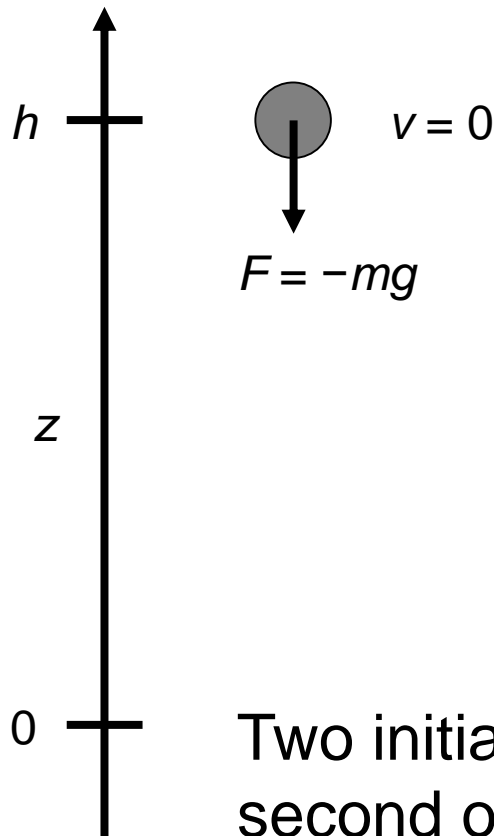
- \mathbf{a} is second derivative of \mathbf{r} with respect to t .

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \left(\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt} \right)$$

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} = \frac{d^2\mathbf{r}}{dt^2} = \left(\frac{d^2x}{dt^2}, \frac{d^2y}{dt^2}, \frac{d^2z}{dt^2} \right)$$

Solution of equation of motion (1)

- Release a ball with mass m



$$m \frac{d^2 z}{dt^2} = -mg \quad \rightarrow \quad \frac{d^2 z}{dt^2} = \frac{dv_z}{dt} = -g,$$

$$v_z(t) = -gt + C, \quad v_z(0) = 0 \quad \rightarrow \quad C = 0$$

$$\frac{dz}{dt} = v_z(t) = -gt, \quad z(t) = -\frac{1}{2}gt^2 + D$$

$$z(0) = h \quad \rightarrow \quad D = h \quad \rightarrow$$

$$z(t) = -\frac{1}{2}gt^2 + h$$

Two initial conditions are necessary to solve second order differential equations

Solution of equation of motion (2)

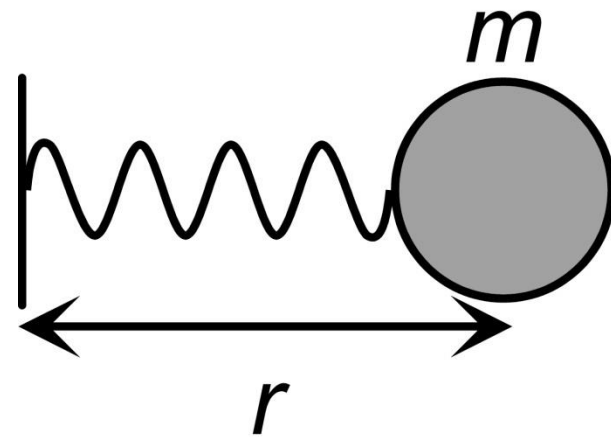
- Motion of a spring

Mass: m

Spring length: r

Equilibrium length: r_0

Spring constant: k



$$F = -k(r - r_0)$$

$$F = -kq$$

$$m \frac{d^2 r}{dt^2} = -k(r - r_0)$$

$$\xrightarrow{q = r - r_0}$$

$$m \frac{d^2 q}{dt^2} = -kq$$

Solution of equation of motion (3)

$$\frac{d^2 q}{dt^2} = -\frac{k}{m} q = -\omega^2 q, \quad \omega = \sqrt{\frac{k}{m}}$$

$$q(t) = A \cos \omega t + B \sin \omega t \quad (\text{General solution})$$

$$q(0) = q_0, \quad dq/dt|_{t=0} = v(0) = 0 \quad (\text{Initial conditions})$$

$$A = q_0, \quad B = 0$$

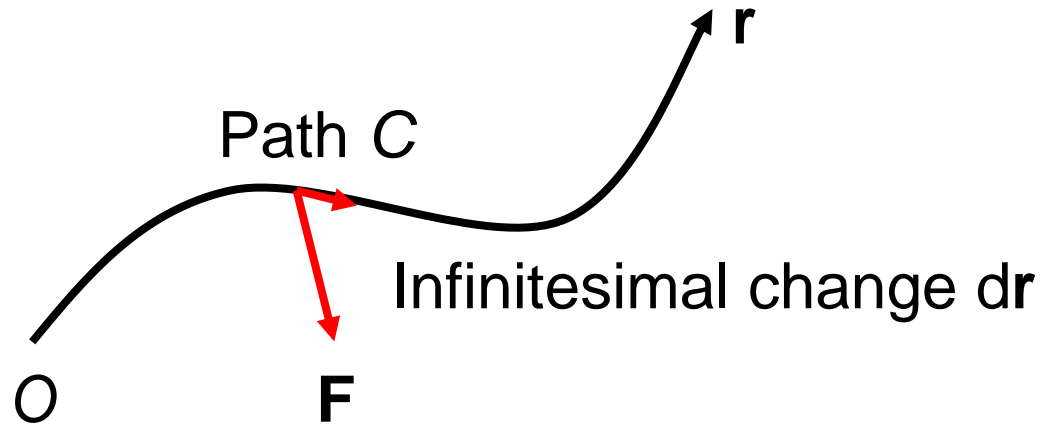
$$q(t) = q_0 \cos \omega t, \quad r = q_0 \cos \omega t + r_0$$

Potential energy and force (1)

Definition of potential energy $E(\mathbf{r})$ at position \mathbf{r}

$$E(\mathbf{r}) = -\int_C \mathbf{F} \cdot d\mathbf{r}$$

Integrate the inner product of \mathbf{F} and $d\mathbf{r}$ along path C



Potential energy and force (2)

Change of $E(\mathbf{r})$ by an infinitesimal change of \mathbf{r}

$$\begin{aligned} E(\mathbf{r} + \Delta\mathbf{r}) - E(\mathbf{r}) &= E(x + \Delta x, y + \Delta y, z + \Delta z) - E(x, y, z) \\ &= E(x + \Delta x, y + \Delta y, z + \Delta z) - E(x, y + \Delta y, z + \Delta z) \\ &\quad + E(x, y + \Delta y, z + \Delta z) - E(x, y, z + \Delta z) \\ &\quad + E(x, y, z + \Delta z) - E(x, y, z) \\ &= \frac{\partial E}{\partial x} \Delta x + \frac{\partial E}{\partial y} \Delta y + \frac{\partial E}{\partial z} \Delta z = \nabla E \cdot \Delta\mathbf{r} \end{aligned}$$

From the definition,

$$E(\mathbf{r} + \Delta\mathbf{r}) - E(\mathbf{r}) = -\mathbf{F} \cdot \Delta\mathbf{r}$$

The comparison gives the following:

$$\mathbf{F} = -\nabla E$$

Force is calculated from potential energy function.

Energy conservation law (1)

- The sum of potential energy E and kinetic energy K ($H = E + K$) is conserved in an isolated system.

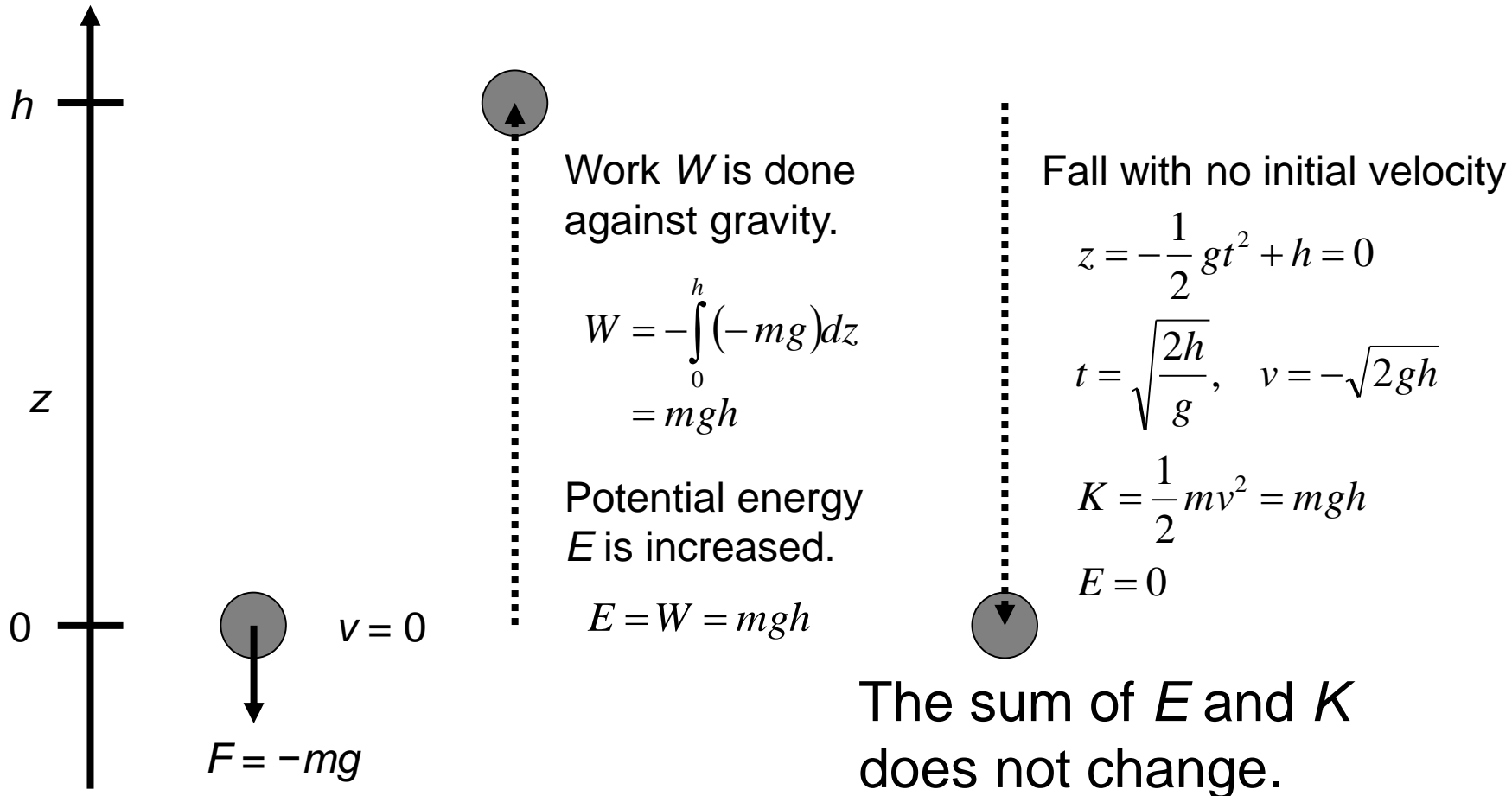
- Kinetic energy

$$K = \frac{1}{2} m |\mathbf{v}|^2$$

- Proof

$$\begin{aligned} \frac{dH}{dt} &= \frac{dK}{dt} + \frac{dE}{dt} = m\mathbf{v} \cdot \frac{d\mathbf{v}}{dt} + \frac{d\mathbf{x}}{dt} \cdot \frac{\partial E}{\partial \mathbf{x}} \\ &= m\mathbf{v} \cdot \mathbf{a} - \mathbf{v} \cdot \mathbf{F} = \mathbf{v} \cdot (m\mathbf{a} - \mathbf{F}) = 0 \end{aligned}$$

Energy conservation law (2)



Energy conservation law (3)

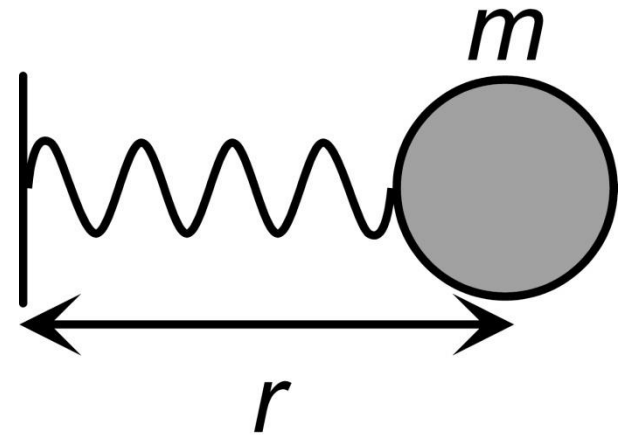
Motion of a spring

$$r(t) = q_0 \cos \omega t + r_0$$

$$v(t) = dr(t)/dt = -q_0 \omega \sin \omega t$$

$$E(r) = -\int_{r_0}^r -k(r' - r_0) dr' = \frac{k}{2} (r - r_0)^2$$

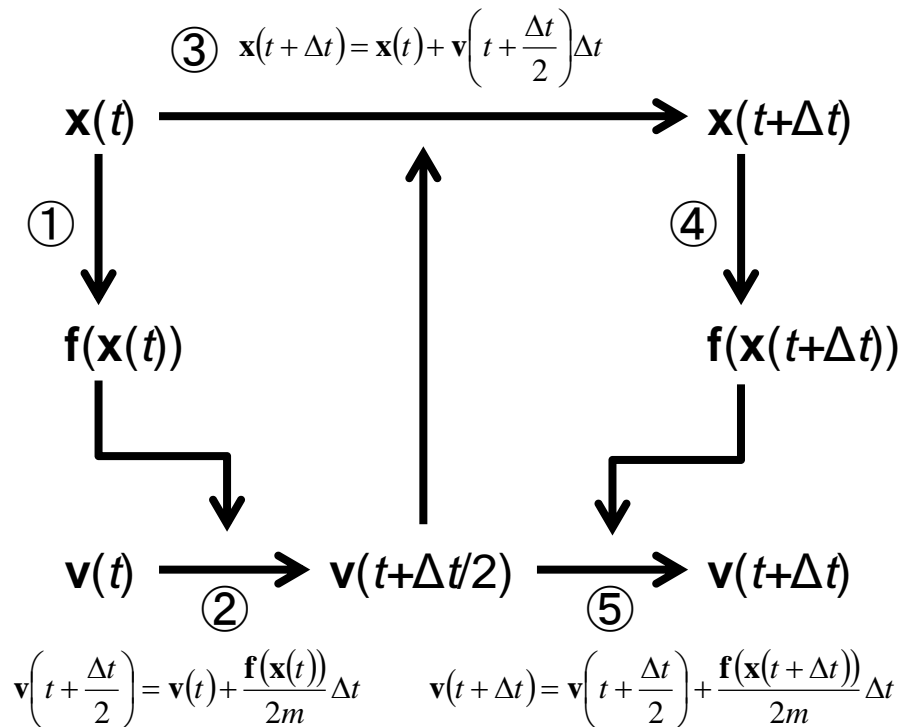
$$\begin{aligned} H &= \frac{m}{2} [v(t)]^2 + \frac{k}{2} [r(t) - r_0]^2 \\ &= \frac{mq_0^2 \omega^2}{2} \sin^2 \omega t + \frac{kq_0^2}{2} \cos^2 \omega t = \frac{1}{2} kq_0^2 \end{aligned}$$



Molecular dynamics methods

- In the molecular dynamics (MD) methods, the force exerted on each atom is calculated from the potential energy function and the time evolutions of its position and velocity are calculated according Newton's equation of motion.
- The equation of motion cannot be solved analytically for the system composed of three or more atoms. → Use of numerical methods
- In the numerical methods, the position and the velocity of each atom are successively calculated at small intervals of time (time step: Δt).
- Accuracy of the method is measured by the deviation of the energy from the initial value.

The velocity Verlet method



- ① Calculate the force for the position at t .
- ② Calculate the velocity at $t + \Delta t/2$.
- ③ Calculate the position at $t + \Delta t$.
- ④ Calculate the force for the position at $t + \Delta t$.
- ⑤ Calculate the force for the position at $t + \Delta t$.
- ⑥ Return to ② and repeat.

A harmonic oscillator (1)

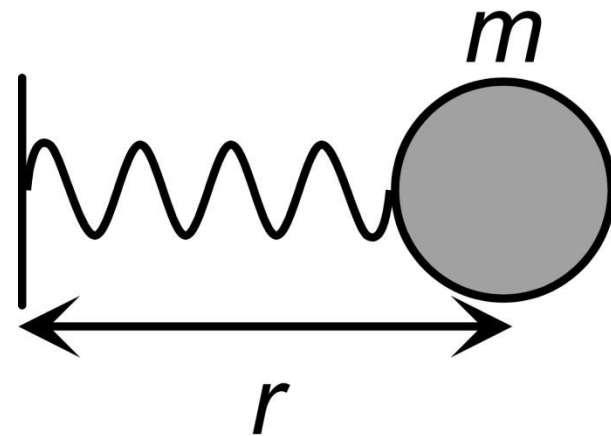
- Motion of a spring

Mass: m

Spring length: r

Equilibrium length: r_0

Spring constant: k



$$F = -k(r - r_0)$$

$$F = -kq$$

$$m \frac{d^2 r}{dt^2} = -k(r - r_0)$$

$$\xrightarrow{q = r - r_0}$$

$$m \frac{d^2 q}{dt^2} = -kq$$

A harmonic oscillator (2)

- Analytical solution

Initial position: q_0 , initial velocity: 0

$$q(t) = q_0 \cos \omega t, \quad v(t) = -q_0 \omega \sin \omega t$$

- A perl script implementing velocity Verlet (osc.pl)

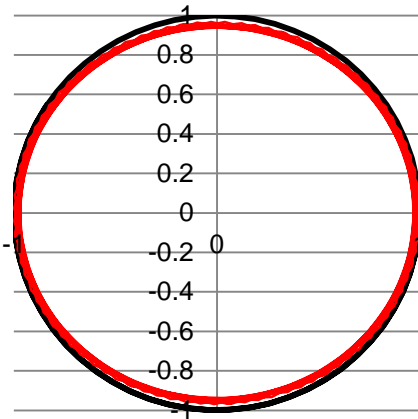
```
$q=1.0;$v=0.0;
$m=1.0;$k=1.0;
$dt=0.01;$nstep=100;
sub calc_force {
    my $q=$_[0];
    my $f=-$k*$q;
    my $e=0.5*$k*$q**2;
    return ($e,$f);
}
open(OUT,">osc$dt.csv");
```

```
($e,$f)=calc_force($q);
for($i=1;$i<=$nstep;$i++) {
    $v+=0.5*$f/$m*$dt;
    $q+=$v*$dt;
    ($e,$f)=calc_force($q);
    $v+=0.5*$f/$m*$dt;
    $H=0.5*$m*$v**2+$e;
    print OUT $i*$dt,",",,$q,
            ",",$v,",",,$H,"¥n";
}
```

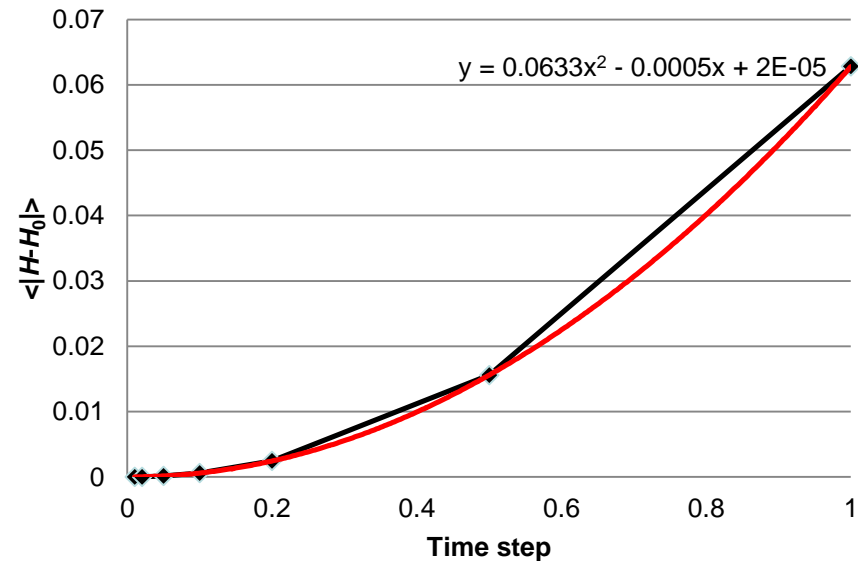
Exercise 1

1. Download `osc.pl` and `osc.xlsx` from the web site of this lecture and run `osc.pl` with different $\$dt$ values of 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, and 1.0.
2. For each $\$dt$ value, calculate the average of absolute difference, $\langle |H - H_0| \rangle$, between the total energy H and its initial value $H_0 = 0.5$.
3. Plot $\langle |H - H_0| \rangle$ as a function of $\$dt$ using `osc.xlsx`.
4. Discuss how $\langle |H - H_0| \rangle$ changes as $\$dt$ increases.

Time step vs. error



Trajectories of (q, v) in the simulations with Δt of 0.1 (black) and 0.5 (red).



Plot of $\langle |H - H_0| \rangle$ as a function of Δt

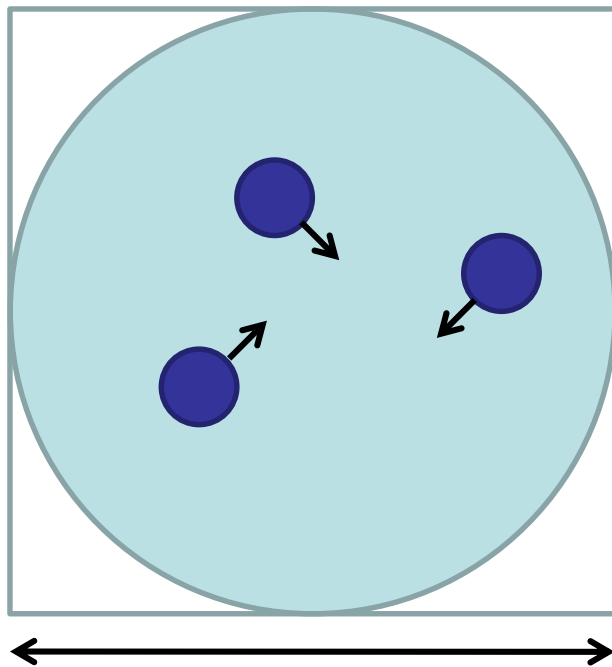
$$H = \frac{mv^2}{2} + \frac{kq^2}{2} = 0.5 \quad \Rightarrow \quad v^2 + q^2 = 1 \quad |H - H_0| \text{ corresponds to the deviation from the circle.}$$

Appropriate choice of time step

- The smaller deviation of the total energy is obtained with the smaller time step.
- In general, the time step should be 1/10–1/20 of the cycle of the fastest motion.
- As for the case of proteins, the fastest motion is the X–H bond stretching motion (X = C, N, O, or S). The frequency is about 3000 cm^{-1} and the cycle is about 10 fs.
- Therefore, 0.5–1.0 fs is used for Δt .

A multi-particle system (1)

- A multi-particle system with van der Waals interaction (vdw.pl)



\$width

```
$natom=3;           # Number of particles
$width=10.0;        # Width of initial
                    # particle distribution
$scale=1.0;         # Scaling factor for
                    # initial velocity
$fcap=1.0;          # Force constant for
                    # spherical boundary
$sigma=1.0;         # Atom radius
$epsilon=1.0;       # Well depth
$mass=1.0;          # Atomic mass
$nstep=100000;     # Number of MD steps
$nsave=100;         # Frequency of saving
                    # trajectory
$dt=0.001;          # Time step
$seed=150525;       # Random seed
```

A multi-particle system (2)

- Initial placement
 - Particles are randomly placed within a cube of side length \$width\$ centered on the origin.
- Initial velocities
 - Randomly assigned (parameter: \$scale\$)
- Potential energy function

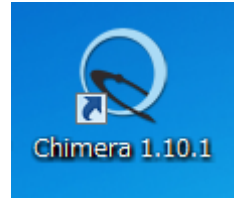
$$E = \sum_{i=1}^N \sum_{j=i+1}^N 4\epsilon \left(\frac{\sigma^{12}}{r_{ij}^{12}} - \frac{\sigma^6}{r_{ij}^6} \right) + \sum_{i=1}^N E_{\text{cap}}(r_i)$$

$$E_{\text{cap}}(r_i) = \begin{cases} 0 & r_i < r_{\text{cut}} \\ f_{\text{cap}}(r_i - r_{\text{cut}})^2 & r_i \geq r_{\text{cut}} \end{cases}$$

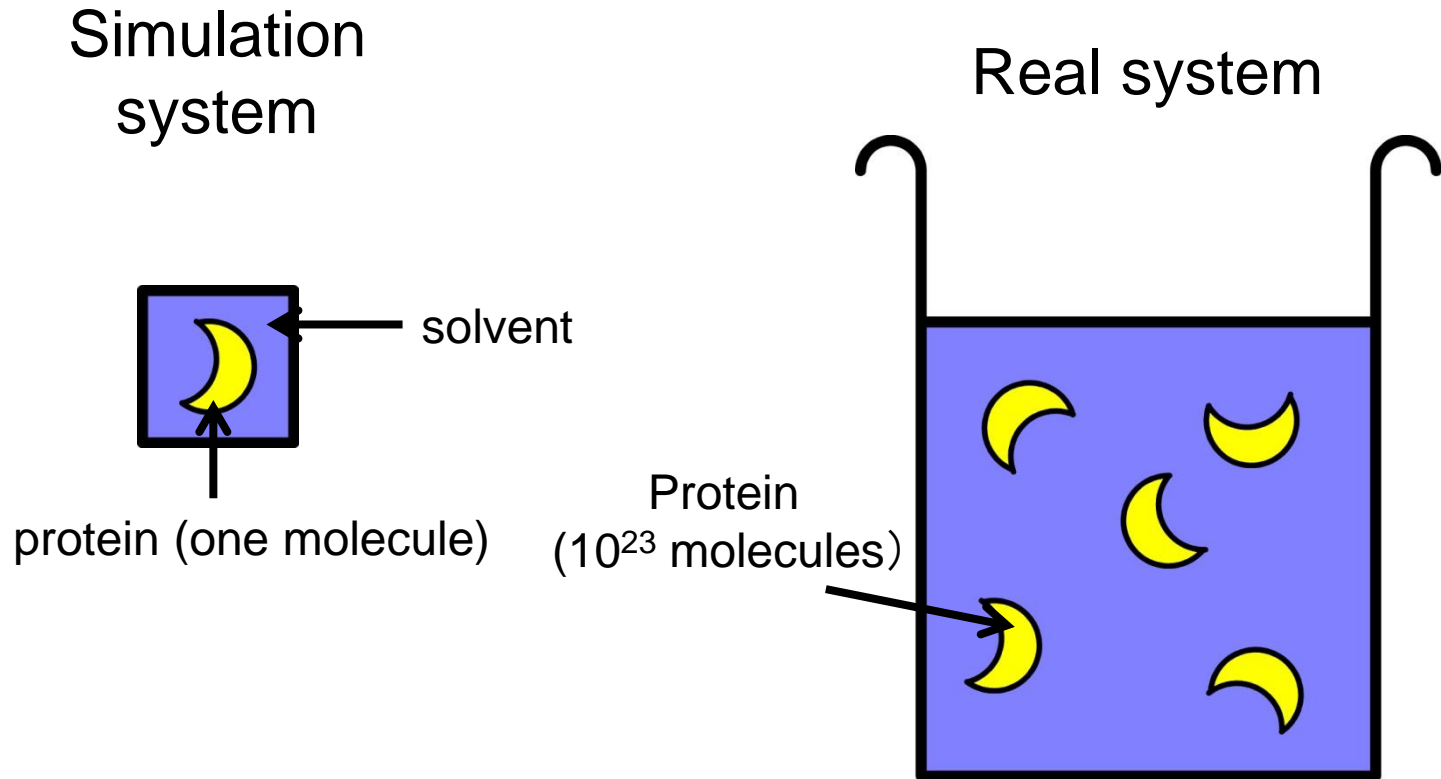
r_i : distance from the origin
 r_{cut} : half of \$width\$

A multi-particle system (3)

- Visualize the result with UCSF Chimera.
- Start Chimera 1.10.1.
- Choose “Tools”→“MD/Ensemble Analysis”→“MD Movie”, change Trajectory format to “PDB”, change PDB frames contained in to “Single file”, set “vdw.pdb” to the file, and “OK.”
- Click the play button to start animation.
- Examine how the motion changes by changing the parameters.

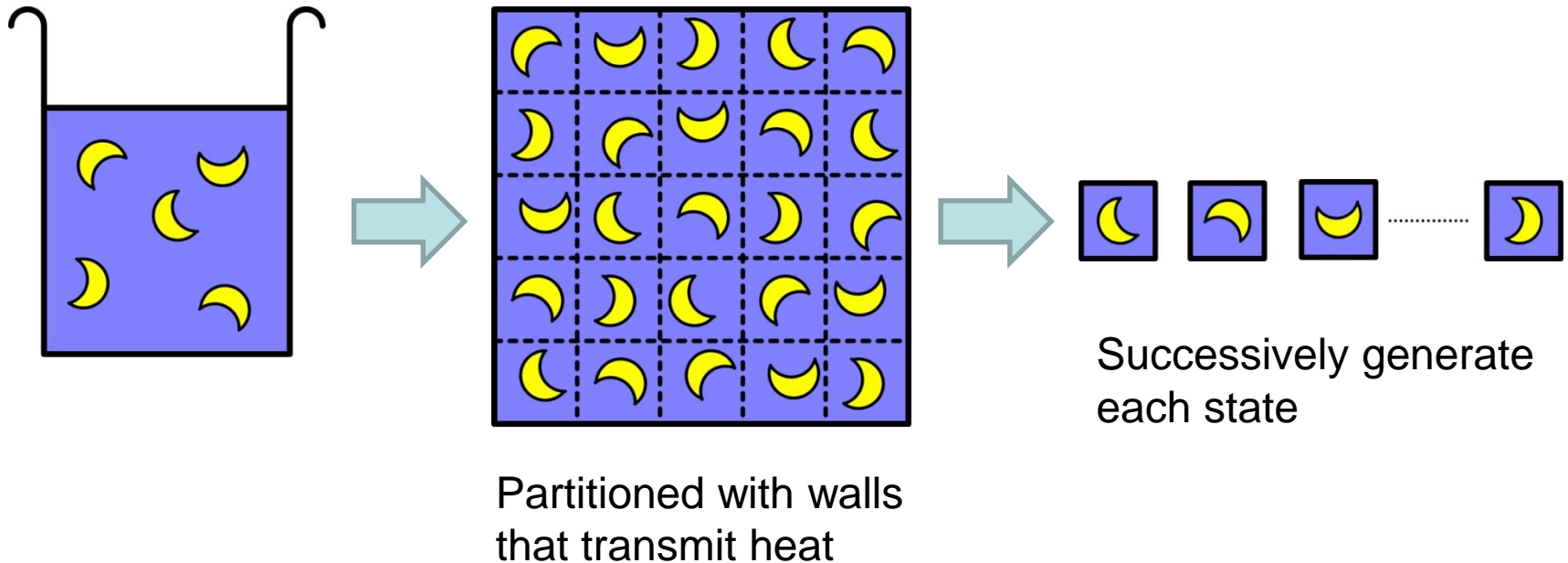


Comparability with experimental data (1)



- Is the result of a molecular simulation comparable with the experimental data?

Comparability with experimental data (2)



- The real system can be reproduced by successively generating each state in a molecular simulation, but how to generate it?

Comparability with experimental data (3)

- Experimental data is the weighted average of the observable dependent on the state.

$$\langle A \rangle = A_i \rho_i, \quad \sum_i \rho_i = 1 \quad \rho_i: \text{probability of state } i$$

- Probability of each state follows the canonical distribution.

$$\rho_i = Z^{-1} \exp(-e_i/k_B T) \quad e_i: \text{energy of state } i$$

$$Z = \sum_i \exp(-e_i/k_B T) \quad Z: \text{partition function}$$

→Generate each state following the canonical distribution.

A harmonic oscillator (1)

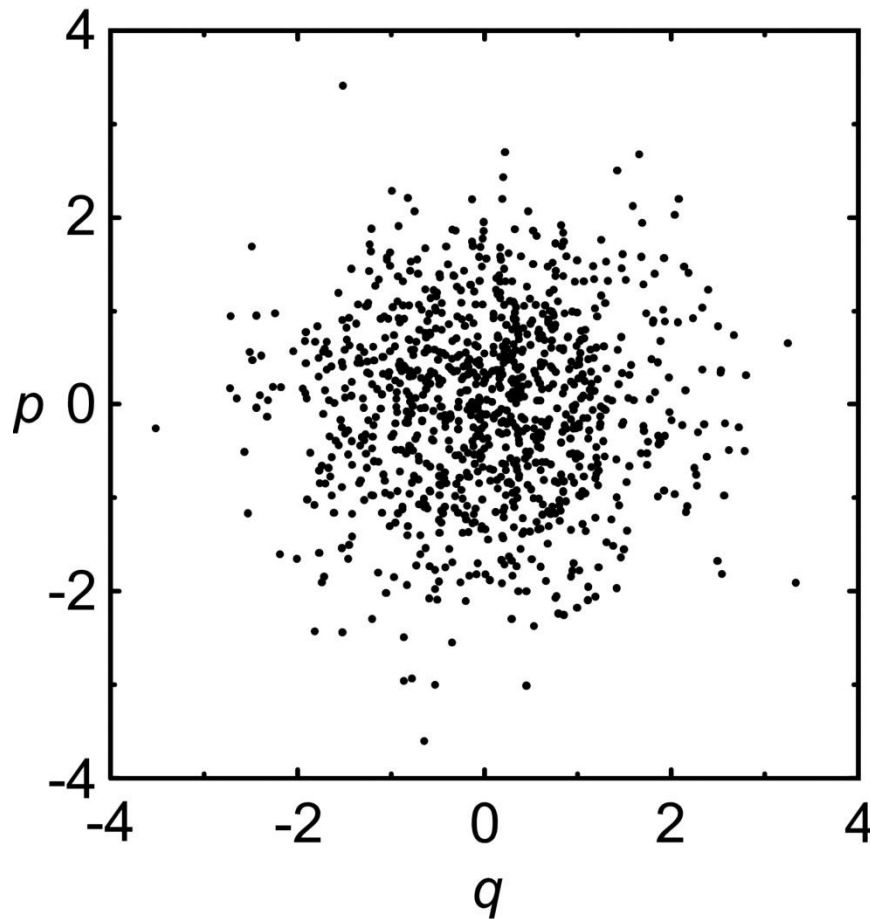
- The canonical distribution of (q, p)

$$H(q, p) = \frac{p^2}{2m} + \frac{k}{2}q^2, \quad \rho(q, p) = \frac{\exp\left[-\frac{H(q, p)}{k_B T}\right]}{\int \exp\left[-\frac{H(q, p)}{k_B T}\right] dq dp}$$

- Generate (q, p) following this distribution
 - Evaluate the accuracy using the average of the total energy

$$\langle H \rangle = \int H(q, p) \rho(q, p) dq dp = k_B T$$

A harmonic oscillator (2)



- When $m = k = k_B T = 1$, this corresponds with the normal distribution.

$$\rho(q, p) = \frac{1}{2\pi} \exp\left(-\frac{q^2 + p^2}{2}\right)$$

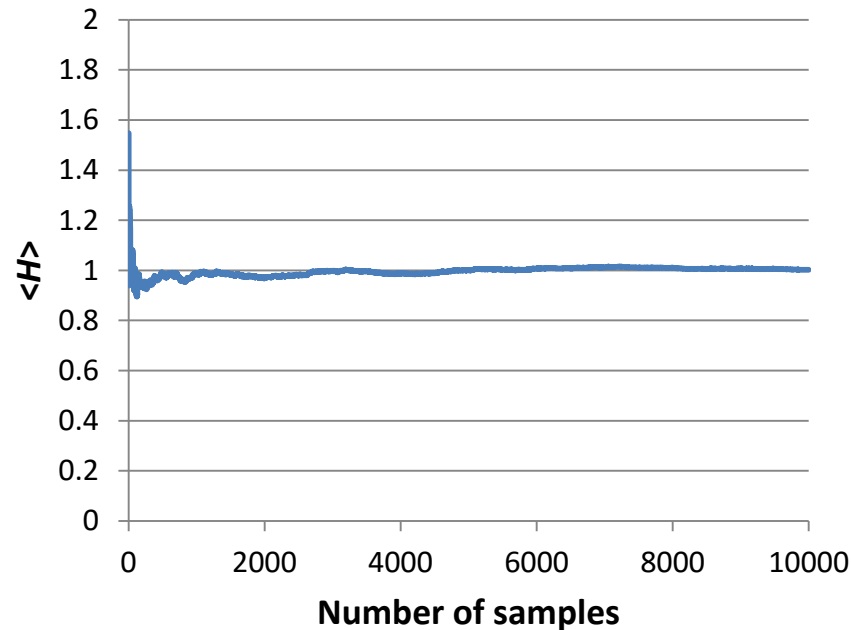
- Check the convergence of $\langle H \rangle$ by increasing the number of samples.

An example program

```
$kT=1.0;
$pi=atan2(1.0,1.0)*4.0;
$npt=10000;

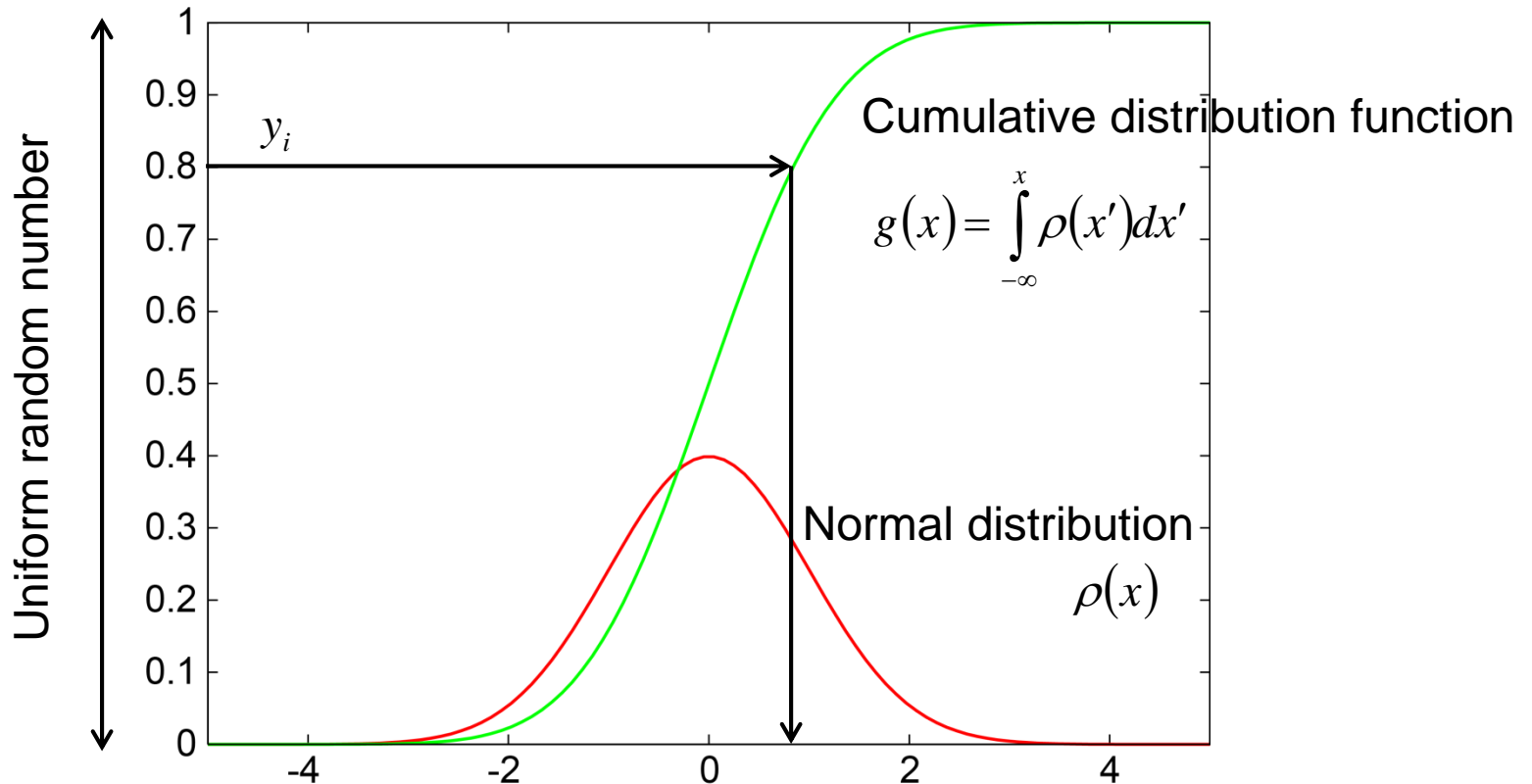
open(OUT,">normal.csv");
$val1=0.0;
for($i=1;$i<=$npt;$i++) {
    $x1=rand;
    $x2=rand;
    #Convert uniform dist. into normal dist.
    $q=sqrt(-2.0*$kT*log($x1))*cos(2.0*$pi*$x2);
    $p=sqrt(-2.0*$kT*log($x1))*sin(2.0*$pi*$x2);
    $H=0.5*$q**2+0.5*$p**2;
    $val1+=$H;           #Average of total energy
    if($i % 100 == 0) {
        printf(OUT "%d,%f¥n",$i,$val1/$i);
    }
}
```

A harmonic oscillator (3)



- $\langle H \rangle$ quickly converges to the theoretical value, 1.

Sample generation (1)



Convert the number into the one following the normal distribution

$$g(x_i) = y_i, \quad x_i = g^{-1}(y_i)$$

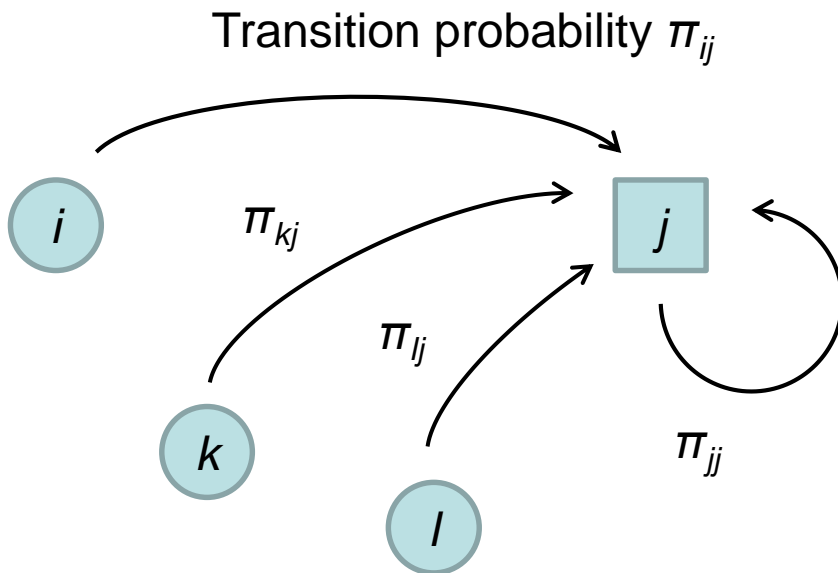
Sample generation (2)

- The above-mentioned method is only applicable to the cases where the cumulative probability function is computable.
- As for the biomacromolecular systems, the relation between bonded and non-bonded interactions is so complex that it is impossible to analytically compute the cumulative probability function.
- Due to the large internal degree of freedom, numerical calculation is also impossible.



Use of Markov chain

Markov chain



- Let π_{ij} be the transition probability from state i to state j .
- Let ρ_i^0 be the probability of state i before transition. Then, the probability of state j after transition is:

$$\rho_j^1 = \sum_i \rho_i^0 \pi_{ij}$$

- The transition probability matrix satisfies:

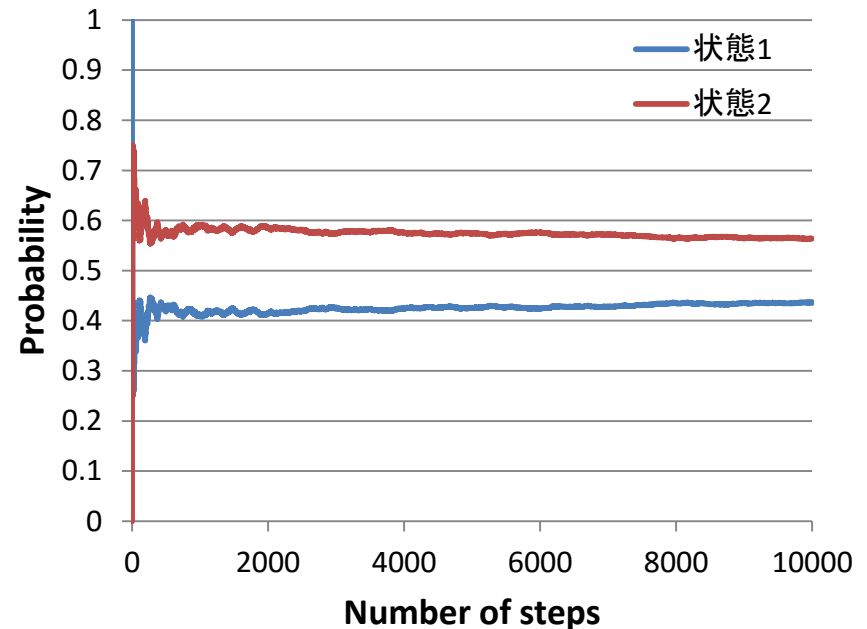
$$\sum_j \pi_{ij} = 1$$

An example (1)

- There are two states
- The transition probabilities are:

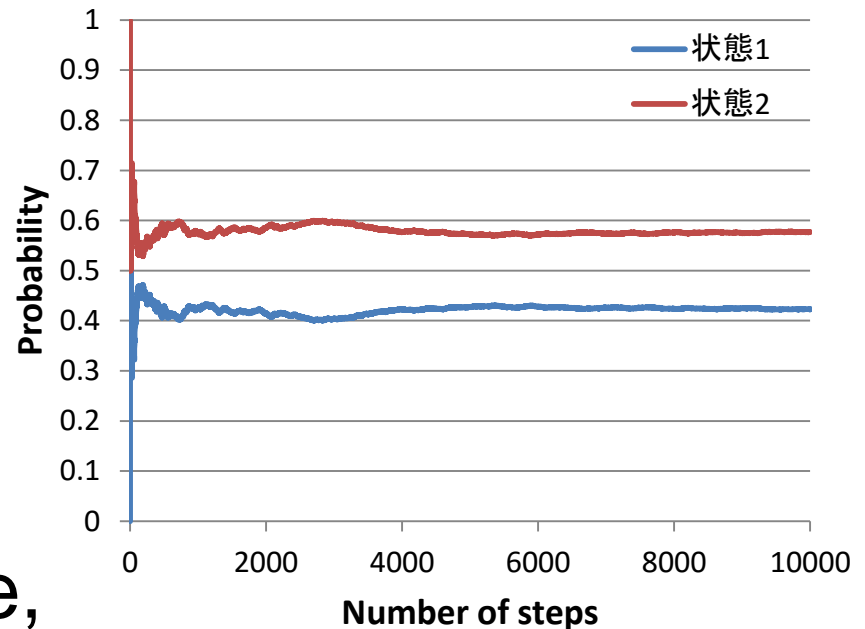
$\pi_{11} = 0.6$	$\pi_{12} = 0.4$
$\pi_{21} = 0.3$	$\pi_{22} = 0.7$

- The plot shows the result of the simulation starting with state 1.



An example (2)

- The probabilities converge to the same values if the simulation starts with state 2.



- Upon the convergence, the equation, $\boldsymbol{\rho} = \boldsymbol{\rho}\boldsymbol{\pi}$, holds.
- $\boldsymbol{\rho}$ is an eigenvector of $\boldsymbol{\pi}$.
→ Uniquely determined for a given $\boldsymbol{\pi}$.

The Metropolis method (1)

- We want to find $\boldsymbol{\pi}$ from $\boldsymbol{\rho}$.
- If the detailed balance condition holds, $\boldsymbol{\rho} = \boldsymbol{\rho}\boldsymbol{\pi}$ holds.

$$\rho_i \pi_{ij} = \rho_j \pi_{ji} \Rightarrow \sum_i \rho_i \pi_{ij} = \sum_i \rho_j \pi_{ji} = \rho_j \sum_i \pi_{ji} = \rho_j$$

- Metropolis *et al.* proposed the following:

$$\begin{cases} \pi_{ij} = \alpha_{ij} & \text{if } \rho_j \geq \rho_i \text{ and } i \neq j \\ \pi_{ij} = \alpha_{ij} (\rho_j / \rho_i) & \text{if } \rho_j < \rho_i \text{ and } i \neq j \end{cases}$$

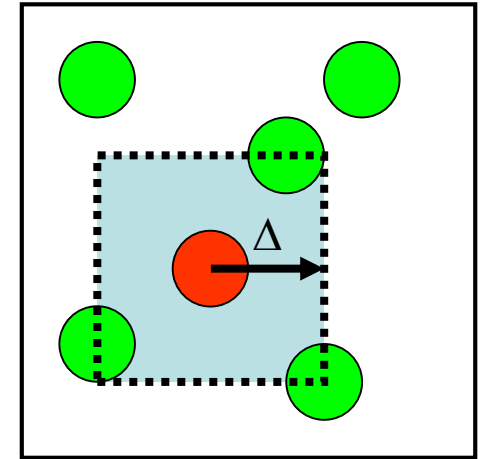
$$\alpha_{ij} = \alpha_{ji}, \quad \sum_{j \neq i} \alpha_{ij} = 1, \quad \pi_{ii} = 1 - \sum_{j \neq i} \pi_{ij}$$

The Metropolis method (2)

1. Move an atom of interest randomly within a cube with side length 2Δ centered on the atom.

$$\alpha_{ij} = \frac{1}{N_{\Delta}} \quad \text{Within the cube}$$

$$\alpha_{ij} = 0 \quad \text{Outside the cube}$$



2. If the energy of the destination state, e_j , is lower than that of the original state, e_i , accept the movement. If not, accept it with the following probability:

$$\rho_j / \rho_i = \exp[-(e_j - e_i) / k_B T] = \exp(-\Delta e_{ji} / k_B T)$$

3. If not accepted, use the original state as the new state.

$$\pi_{ii} = 1 - \sum_{j \neq i} \pi_{ij} = \sum_{j \neq i} (\alpha_{ij} - \pi_{ij}) = \sum_{\substack{j \neq i \text{ and} \\ \rho_j < \rho_i}} \alpha_{ij} (1 - \rho_j / \rho_i)$$

Exercise 2

1. Download metropolis.pl from the web site of this lecture and run it.
2. Plot $\langle H \rangle$ as a function of the sample number and ascertain that $\langle H \rangle$ approaches to 1 as the sample number increases.
3. Run metropolis.pl with $\$delta = 0.1$ and plot the result similarly.
4. How does the plot change by using a smaller $\$delta$ value? Discuss the reason for the change.

metropolis.pl

```
$nstep=10000; #Number of steps
($q,$p)=(0.0,0.0); #Initial conditions
$delta=1.0; #Maximum displacement
$kT=1.0;$m=1.0;$k=1.0; #kT, mass, force constant
$delta_q=$delta/sqrt($k);
$delta_p=$delta*sqrt($m);

open(OUT,">metropolis$delta.csv"); #Output file

$save=0.0;
$H=&calc_H($q,$p); #Initial energy
for($i=1;$i<=$nstep;$i++) {
    $q_new=$q+2.0*$delta_q*(rand()-0.5); #Trial
    $p_new=$p+2.0*$delta_p*(rand()-0.5); #Trial
    $H_new=&calc_H($q_new,$p_new);
    $probability=exp(($H-$H_new)/$kT);
    if($probability >= 1.0 || $probability >= rand()) { #Judgement
        $q=$q_new;$p=$p_new;$H=$H_new;
    }
    $save+=$H;
    printf(OUT "%d,%f\n",$i,$save/$i) if($i % 100 == 0);
}
close(OUT);

sub calc_H { #Energy function
    my ($q,$p)=@_;
    return 0.5*$p*$p/$m+0.5*$k*$q*$q;
}
```

Application to biomacromolecules

- “State” corresponds to “structure.”
- The Metropolis method can be performed by moving an arbitrary chosen atom.
- However, moving an atom changes the bond length and the energy increases in most cases.
 - The trial is rarely accepted.
- To solve this problem, only dihedral angles are changed without changing the bond lengths and angles. But, there are still difficulties.
 - Application to a multiple molecular system is difficult.
 - In the protein core where atoms are densely packed, changing a dihedral angle tends to cause overlapping of atoms.

Constant-temperature MD (1)

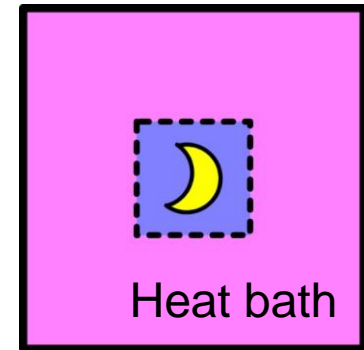
- Sampling from the canonical distribution is enabled by keeping the temperature constant in the MD simulation.

$$\frac{3}{2} Nk_{\text{B}}T = \sum_{i=1}^N \frac{m_i |\mathbf{v}_i|^2}{2}$$

- Since MD simulations are conducted in the Cartesian coordinate system, they are more easily applied to biomacromolecules.
- The average is taken over time.

Constant-temperature MD (2)

- Nosé's method
 - Nosé-Hoover chain method
 - An improved method
 - Constraint method
 - Only the distribution of the coordinates is canonical.
 - Langevin dynamics
 - Temperature is controlled by random force and friction.
 - Berendsen's weak coupling method
 - Distribution is not canonical.
 - But frequently used because of its easiness and stability
- } Considers degree of freedom of heat bath



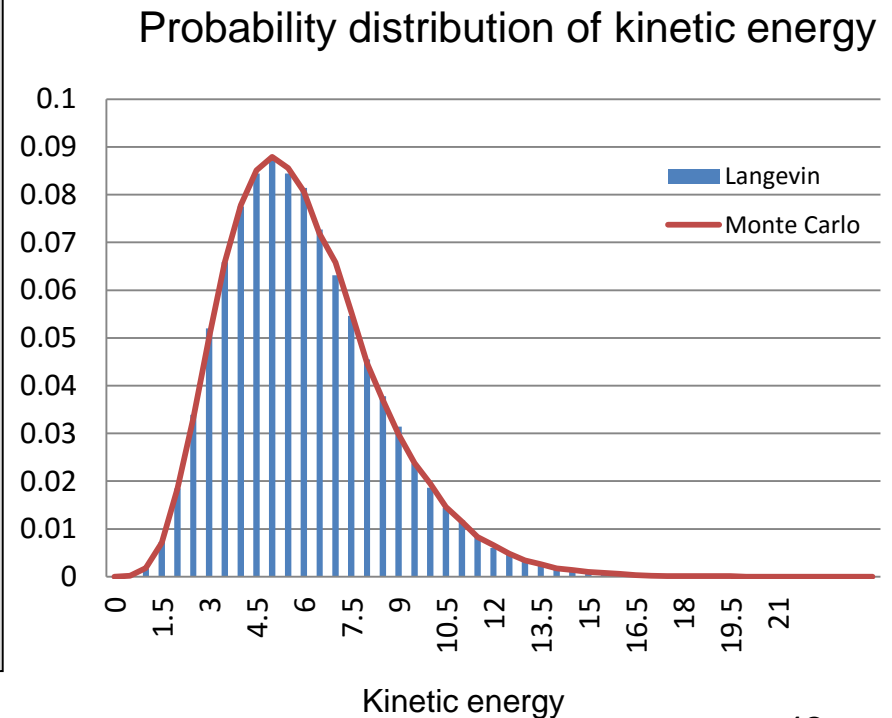
Langevin dynamics

- Exchange of energy between the real system and the head bath occurs by the collision between the atoms of the real system and randomly moving fictitious particles of the heat bath.
 - If the real system is hot, the energy is transmitted toward the heat bath and vice versa.
- Equations of motion
$$m\mathbf{a} = \mathbf{F}(\mathbf{x}) - \underbrace{\gamma \mathbf{v}}_{\text{friction}} + \underbrace{\mathbf{R}(t)}_{\text{force by collision with fictitious particles}}$$
- \mathbf{R} satisfies the following conditions:
$$\langle \underbrace{\mathbf{R}(t)}_{\text{average}} \rangle = 0, \quad \langle \underbrace{\mathbf{R}(t) \cdot \mathbf{R}(t')}_{\text{variance-covariance}} \rangle = 6k_{\text{B}}T\gamma\delta(t-t')$$

An example

- A multi-particle system with van der Waals interaction (vdw_langevin.pl)

```
$natom=4;           # Number of particles
$width=10.0;        # Width of initial
                    # particle distribution
$fcap=1.0;          # Force constant for
                    # spherical boundary
$sigma=1.0;         # Atom radius
$epsilon=1.0;       # Well depth
$mass=1.0;          # Atomic mass
$nstep=100000;     # Number of MD steps
$nsave=100;         # Frequency of saving
                    # trajectory
$dt=0.001;         # Time step
$seed=150525;       # Random seed
$gamma=10.0;        # friction coefficient
$kT=1.0;
```



Berendsen's weak coupling

1. Calculate instantaneous temperature T' after each step of velocity Verlet

$$\frac{3}{2} N k_B T' = \sum_{i=1}^N \frac{m_i |\mathbf{v}_i|^2}{2}$$

2. Scale the velocity by a factor of χ . Parameter τ controls the speed of adaptation.

$$\chi = \left[1 + \frac{\Delta t}{\tau} \left(\frac{T}{T'} - 1 \right) \right]^{1/2}$$

Assignment submission instructions

- Put the results and discussion of the exercises in the slides of a PowerPoint file.
- Send the PowerPoint file as an attachment to email.
- Put “Molecular modeling assignments” in the Subject field of the email.
- Be sure to put your name and ID card number (if you are a student) in the body of the email.
- Send the email to Prof. Tohru Terada (tterada@iu.a.u-tokyo.ac.jp).