

# Molecular Dynamics: Integrators for eqs. of motion

Molecular Dynamics (MD) simulations are carried out in the *Microcanonical ensemble*

The equations of motion can be obtained from Hamiltonian formalism:

$$\frac{\partial \mathbf{p}_i}{\partial t} = - \frac{\partial H(\mathbf{p}, \mathbf{q})}{\partial \mathbf{q}_i} \quad , \quad \frac{\partial \mathbf{q}_i}{\partial t} = \frac{\partial H(\mathbf{p}, \mathbf{q})}{\partial \mathbf{p}_i}$$

In cartesian coordinates:

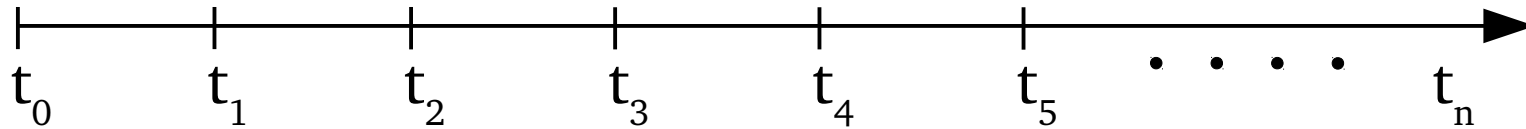
$$\frac{\partial \mathbf{p}_i}{\partial t} = - \frac{\partial U(\mathbf{r})}{\partial \mathbf{r}_i} \quad , \quad \frac{\partial \mathbf{r}_i}{\partial t} = \frac{\partial K(\mathbf{p})}{\partial \mathbf{p}_i} = \frac{\mathbf{p}_i}{m_i}$$

Reasons for using Lagrangean/Hamiltonian formalisms:

- not always adequate to perform simulations in cartesian coordinates, e.g., systems with rigid bonds require special treatment;
- the rigorous introduction of thermostats and barostats to the eqs. of motion;
- the natural formalism to associate with Statistical Mechanics and Quantum Mechanics;
- one nice example of the power of Lagrangean formalisms is the Car-Parrinello method (CPMD), where classical eqs. of motion are set up for the electronic wavefunctions.

# Molecular Dynamics: Integrators for eqs. of motion

They are discrete time approximations for Newton's eqs. of motion.



$t_n = ndt$  , where  $dt$  is the time step

$\mathbf{r}(t)$  is substituted by the series  $\mathbf{r}(0), \mathbf{r}(dt), \mathbf{r}(2dt), \mathbf{r}(3dt), \dots$

## Verlet and Velocity-Verlet (VV) algorithms

The forward Taylor expansion yields:

$$\begin{aligned}\mathbf{r}(t + dt) &= \mathbf{r}(t) + \frac{d\mathbf{r}}{dt} dt + \frac{d^2\mathbf{r}}{dt^2} \frac{dt^2}{2} + \frac{d^3\mathbf{r}}{dt^3} \frac{dt^3}{6} + \mathcal{O}(dt^4) \\ &= \mathbf{r}(t) + \mathbf{v}(t)dt + \frac{\mathbf{f}(t)}{m} \frac{dt^2}{2} + \frac{d^3\mathbf{r}}{dt^3} \frac{dt^3}{6} + \mathcal{O}(dt^4) \quad , \quad \mathbf{f}(t) = -\frac{\partial U(\mathbf{r}(t))}{\partial \mathbf{r}}\end{aligned}$$

The backward expansion yields:

$$\mathbf{r}(t - dt) = \mathbf{r}(t) - \mathbf{v}(t)dt + \frac{\mathbf{f}(t)}{m} \frac{dt^2}{2} - \frac{d^3\mathbf{r}}{dt^3} \frac{dt^3}{6} + \mathcal{O}(dt^4)$$

Then, solving for  $\mathbf{r}(t + dt)$ , we get the **Verlet** formula

$$\mathbf{r}(t + dt) = 2\mathbf{r}(t) - \mathbf{r}(t - dt) + \frac{\mathbf{f}(t)}{m} \frac{dt^2}{2} + \mathcal{O}(dt^4)$$

# Molecular Dynamics: Integrators for eqs. of motion

The simple **Verlet** formula

$$\mathbf{r}(t + dt) = 2\mathbf{r}(t) - \mathbf{r}(t - dt) + \frac{\mathbf{f}(t)}{m} \frac{dt^2}{2} + \mathcal{O}(dt^4)$$

has some **drawbacks**. The velocity must be calculated aside by

$$\mathbf{v}(t) = \frac{\mathbf{r}(t + dt) - \mathbf{r}(t - dt)}{2dt} + \mathcal{O}(dt^3) .$$

The **Velocity-Verlet** is an alternative, in this case we have two sets of eqs.

$$\mathbf{r}(t + dt) = \mathbf{r}(t) + \mathbf{v}(t)dt + \frac{\mathbf{f}(t)}{2m} dt^2$$

$$\mathbf{v}(t + dt) = \mathbf{v}(t) + \frac{\mathbf{f}(t + dt) + \mathbf{f}(t)}{2m} dt = \left( \mathbf{v}(t) + \frac{\mathbf{f}(t)}{m} \frac{dt}{2} \right) + \frac{\mathbf{f}(t + dt)}{m} \frac{dt}{2}$$

The integration over  $dt$  is achieved in two steps.

***This method is the one most frequently used.***

# Molecular Dynamics: Integrators for eqs. of motion

## The Leap-frog algorithm

Equivalent to the Verlet algorithms, but yields  $\mathbf{r}$  and  $\mathbf{v}$  at intercalated times

$$\mathbf{r}(n \cdot dt) \text{ and } \mathbf{v}((n + 1/2)dt).$$

The equations are:

$$\mathbf{v}(t + dt/2) = \mathbf{v}(t - dt/2) + \frac{\mathbf{f}(t)}{m} dt$$

$$\mathbf{r}(t + dt) = \mathbf{r}(t) + \mathbf{v}(t + dt/2) dt$$

The **drawback** is for calculating the total energy. For that sake, the average velocity can be calculated as

$$\mathbf{v}(t) = \frac{\mathbf{v}(t + dt) - \mathbf{v}(t - dt)}{2}.$$

## Errors for Verlet-based algorithms

**Local** error in position is  $\mathcal{O}(dt^4)$  , **local** error in velocity is  $\mathcal{O}(dt^2)$ .

**Global** error in position and velocities is  $\mathcal{O}(dt^2)$  .

In molecular dynamics simulations, the global error is typically far more important than the local error, and the *Verlet integrator* is therefore known as a **second-order integrator**.

# Molecular Dynamics: Integrators for eqs. of motion

## *Higher Order Methods*

A number of efficient higher order techniques exists in the literature, for instance:

- Predictor-Corrector Methods
- Runge-Kutta Methods
- Adams Methods

In principle, these methods have a better rate of convergence scaling linearly with  $dt$ , however, they also have a computational overhead that makes them slower.

Therefore, for general MD applications, it is better to use a lower order method with a smaller time-step ( $dt$ ) to compensate for the lower accuracy.

The use of higher order methods might be necessary, nonetheless, to investigate unstable situations, like systems close to a phase transition or a critical point.

# Molecular Dynamics: Integrators for eqs. of motion

## General Considerations

Energy conservation: the integrators (any of them) should conserve energy. This is a good test to verify code correctness.

small oscillations of  $E$  around  $\langle E \rangle$  are acceptable to one part in  $10^4 - 10^5$  ;

the energy  $E$  should not drift, this is a sign of something going wrong ;

by decreasing the time-step  $dt$  the energy conservation should improve .

Criteria for time-step:

| <i>system</i>                           | <i>time - step</i> |
|---|--------------------|
| non-adiabatic dynamics                  | $< 0.1 fs$         |
| molecules (bond vibrations)             | $0.5 - 1.0 fs$     |
| molecules (rigid bonds / angle bending) | $\sim 2 fs$        |
| atomic translations                     | $5 - 10 fs$        |

big time steps can lead to particles to overlap and give rise to strong forces that send atoms apart.

# Molecular Dynamics

direct comparison with experiment requires that the boundary conditions (BC) imposed on the simulated system are adequate with the experimental conditions:

- **hard** (*must be satisfied at every timepoint during the simulation*)
- **soft** (*average must be equal to the macroscopic value during the simulation*)

## 1) Spatial BC (**hard** constraints):

- vacuum
- fixed (e.g., solute + solvent in a droplet )
- periodic

## 2) Thermodynamic BC (characterize the macrostate of the system):

- extensive quantities, like E, N or V, are **hard** constraints
- intensive quantities, like T, P,  $\mu$ , are **soft** constraints

## 3) Experimental BC

- used to enforce agreement between simulation and experiment (**soft**)

## 4) Geometrical BC

- such as fixing specific bond lengths or molecular geometry (**hard**).

# Molecular Dynamics

| Independent | Dependent | Ensemble                    |
|-------------|-----------|-----------------------------|
| $NVE$       | $\mu PT$  | Microcanonical              |
| $NVT$       | $\mu PE$  | Canonical                   |
| $NPH$       | $\mu VT$  | Isoenthalpic-isobaric       |
| $NPT$       | $\mu VH$  | Isothermal-isobaric (Gibbs) |
| $\mu VL$    | $NPT$     | Grand-microcanonical        |
| $\mu VT$    | $NPL$     | Grand-canonical             |
| $\mu PR$    | $NVT$     | Grand-isothermal-isobaric   |
| $\mu PT$    | $NVR$     | Generalized                 |

canonical ensemble:

$$\sigma_E^2 = \langle E^2 \rangle_{NVT} - \langle E \rangle_{NVT}^2 = k_B T c_V$$

$$\sigma_T^2 = \langle T^2 \rangle_{NVT} - \langle T \rangle_{NVT}^2 = 2T^2 / N_{DOF}$$

isothermal-isobaric  
(Gibbs) ensemble:

$$H = E + PV$$

$$\sigma_V^2 = \langle V^2 \rangle_{NPT} - \langle V \rangle_{NPT}^2 = V k_B T \beta_T$$

$$\sigma_T^2 = \langle H^2 \rangle_{NPT} - \langle H \rangle_{NPT}^2 = k_B T^2 c_P$$



# Molecular Dynamics: Thermostats and Barostats

Modification of the Newtonian MD equations with the purpose of generating a thermodynamical ensemble at constant temperature (pressure) is called a thermostat (barostat) algorithm.

- to match experimental conditions
- to study temperature-dependent processes
- to dissipate the heat in dissipative non-equilibrium MD simulations
- to enhance the efficiency of a conformational search (e.g., simulated annealing)
- to avoid steady energy drifts caused by the accumulation of numerical errors

# Molecular Dynamics: Thermostats

Typical methods for applying soft boundary conditions:

- the penalty-function,
- weak-coupling (Berendsen),
- extended-system (Nosé, Nosé-Hoover)
- stochastic-coupling methods (Anderson)

**easy:** there are many ways to ensure that the average of an instantaneous quantity takes a specified value

**difficult:** to ensure that the simulation actually samples the correct ensemble and, in particular, provides the correct fluctuations for the specific instantaneous observable in the given ensemble

# Berendsen Thermostat

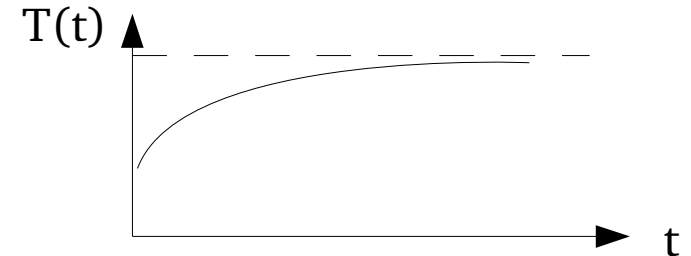
this thermostat is based on the idea of a first-order relaxation equation for T

$$\frac{dT}{dt} = -\frac{1}{\tau} [T(t) - T_B] \quad \Longrightarrow \quad T(t) = T_B e^{-t/\tau} + T_B \left(1 + e^{-t/\tau}\right)$$

$T(t)$  = instantaneous temperature

$\tau$  = Berendsen relaxation time

$T_B$  = thermal-bath temperature



$\tau$  must be short compared to the experimental timescale,  
but long compared to the time separating atomic collisions

A definition of temperature for the atomic system:

$$\langle K \rangle \equiv K = \frac{1}{2} k_B N_{Dof} T \quad \Longrightarrow \quad T = \frac{2}{k_B N_{Dof}} K$$

velocity rescale factor:  $\lambda$

$$T(t + \delta t) = T(t) - \frac{\delta t}{\tau} (T(t) - T_B)$$

$$\lambda \equiv \frac{T(t + \delta t)}{T(t)} = 1 + \frac{\delta t}{\tau} \left( \frac{T_B}{T(t)} - 1 \right)$$

- The Berendsen equations of motion are smooth and deterministic, but time-irreversible.
- The ensemble generated by its eqs. of motion is not a canonical ensemble.
- It neglects stochastic fluctuations on the microscopic timescale.

# Nosé-Hoover Thermostat

The idea of a thermostat based on an extended-system method is due to Nosé: extend the real system by addition of an artificial  $(N_{df} + 1)^{\text{th}}$  dynamical variable  $s'$  (associated with a fictitious “mass”  $Q > 0$  and satisfying  $s' > 0$ ) that plays the role of a **time-scaling parameter**.

The atomic coordinates are identical in both systems (extended and real), but the extended-system velocities are amplified

$$dt' = s' dt \implies \mathbf{r}' = \mathbf{r} \quad , \quad \mathbf{v}' = \frac{\mathbf{v}}{s'} \quad , \quad s' = s \quad , \quad \dot{s}' = \frac{\dot{s}}{s'}$$

$$s' \frac{d}{dt'} \mathbf{r}' = \frac{d\mathbf{r}}{dt} \implies \mathbf{v}' = \frac{\mathbf{v}}{s'}$$

The Lagrangean for the extended-system:

$$g = (N_{DoF} + 1)$$

$$\mathcal{L} = \underbrace{\frac{1}{2} \sum_i^N m_i s'^2 \dot{\mathbf{r}}_i'^2 - V(\mathbf{r})}_{\text{real system}} + \underbrace{\frac{1}{2} Q \dot{s}'^2 - g k_B T \ln s'}_{\text{bath}}$$

the extended-system momenta are amplified by a factor  $s'$  compared to the real-system momenta

$$\mathbf{p}'_i = \frac{\partial \mathcal{L}}{\partial \dot{\mathbf{r}}'_i} = m_i s'^2 \mathbf{v}'_i = m_i s'^2 \frac{\mathbf{v}_i}{s'} \longrightarrow \mathbf{p}'_i = s' \mathbf{p}_i$$

# Nosé-Hoover Thermostat

The Nosé hamiltonian for the extended-system:

$$\mathcal{H} = \underbrace{\sum_i^N \frac{\mathbf{p}'_i}{2m_i s'^2} + V(\mathbf{r})}_{\text{real system}} + \underbrace{\frac{p'_s{}^2}{2Q} + gk_B T \ln s'}_{\text{bath}} = \mathcal{H}_0\left(\mathbf{r}, \frac{\mathbf{p}'}{s'}\right) + \frac{p'_s{}^2}{2Q} + gk_B T \ln s'$$

This function is a constant of the motion and evaluates to  $E'$ , the total energy of the extended-system.

The corresponding Hamiltonian equations of motion:

$$\begin{aligned} \dot{\mathbf{p}}'_i &= -\frac{\partial \mathcal{H}}{\partial \mathbf{r}'_i} & \text{and} & & \dot{\mathbf{r}}'_i &= \frac{\partial \mathcal{H}}{\partial \mathbf{p}'_i} = \frac{\mathbf{p}_i}{m_i s'^2} \\ \dot{p}'_s &= -\frac{\partial \mathcal{H}}{\partial s'} = \frac{1}{s'} \left( \sum_i \frac{\mathbf{p}'_i{}^2}{m_i s'^2} - gk_B T \right) & \text{and} & & \dot{s}' &= \frac{\partial \mathcal{H}}{\partial p'_s} = \frac{p'_s}{s'} \end{aligned}$$

The Nosé equations of motion sample a *microcanonical ensemble in the extended-system, with a constant energy  $E'$*

$$\mathcal{H}\left(\mathbf{r}', \frac{\mathbf{p}'}{s'}, p'_s, s'\right)$$

***But the energy of the real system is not constant***

$$\mathcal{H}_0\left(\mathbf{r}', \frac{\mathbf{p}'}{s'}\right)$$

# Nosé-Hoover Thermostat

The Nosé hamiltonian for the extended-system:

$$\mathcal{H} = \underbrace{\sum_i^N \frac{\mathbf{p}'_i}{2m_i s'^2} + V(\mathbf{r})}_{\text{real system}} + \underbrace{\frac{p'_s{}^2}{2Q} + gk_B T \ln s'}_{\text{bath}} = \mathcal{H}_0(\mathbf{r}, \frac{\mathbf{p}'}{s'}) + \frac{p'_s{}^2}{2Q} + gk_B T \ln s'$$

Nosé showed that the phase-space distribution resulting from the equations of motion is canonical in the variables  $(r, p=p'/s)$

$$\begin{aligned} \Omega(E', V, N) &\propto e^{\beta E'} \int_{-\infty}^{\infty} e^{-\beta p'_s{}^2 / (2Q)} dp'_s \int d\mathbf{r}^N d\mathbf{p}^N e^{-\beta \mathcal{H}(\mathbf{r}, \mathbf{p})} \\ &\propto e^{\beta E'} \sqrt{\frac{2Q\pi}{\beta}} \underbrace{Z(T, N, V)}_{\text{real system}} \end{aligned}$$

Thus, canonical distributions can be generated with smooth, deterministic and time-reversible trajectories.

The use of an extended system with a stretched timescale ( $dt' = s' dt$ ) is not very intuitive, and the sampling of a trajectory at uneven time intervals is rather impractical for the investigation of the dynamical properties of a system.

# Nosé-Hoover Thermostat

The Nosé equations of motion are:

$$\begin{aligned}\dot{\mathbf{r}}'_i &= \frac{\mathbf{p}'_i}{m s'^2} & \text{and} & & \dot{\mathbf{p}}'_i &= \mathbf{F}_i \\ \dot{s}'_i &= \frac{p'_s}{Q} & \text{and} & & \dot{\mathbf{p}}'_s &= \sum \frac{\mathbf{p}'_i{}^2}{m s'^3} - \frac{g k_B T}{s'}\end{aligned}$$

Hoover showed that these equations of motion can be cast in terms of real-system variables: (I) by transforming to the scaled momenta ( $\mathbf{p} = \mathbf{p}'/s'$ ) and (II) to the scaled time  $dt = dt'/s'$ .

Define the friction coefficient:  $\xi = \frac{\dot{s}}{s} = \frac{d \ln s}{dt} = \frac{p_s}{Q}$ .

In scaled momenta and scaled time, the *Lagrangian equations of motion* read:

$$\begin{aligned}\dot{\mathbf{r}}_i &= \frac{\mathbf{p}_i}{m} & \text{and} & & \dot{\mathbf{p}}_i &= \mathbf{F}_i - \xi \mathbf{p}_i \\ \xi &= \dot{s}/s & \text{and} & & \dot{\xi} &= \frac{1}{Q} \left[ \sum \frac{\mathbf{p}_i^2}{m} - g k_B T \right]\end{aligned}$$

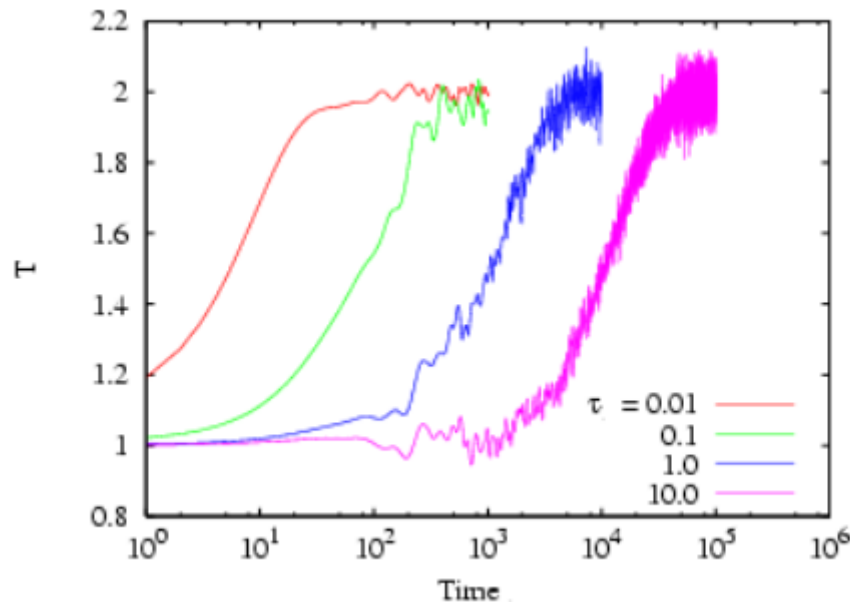
The NH hamiltonian is a constant of the motion (evaluating to E')

$$\mathcal{H}(\mathbf{r}, \mathbf{p}, s, p_s) = \sum_i^N \frac{\mathbf{p}_i^2}{2m_i} + V(\mathbf{r}) + \frac{p_s^2 s^2}{2Q} + g k_B T \ln s$$

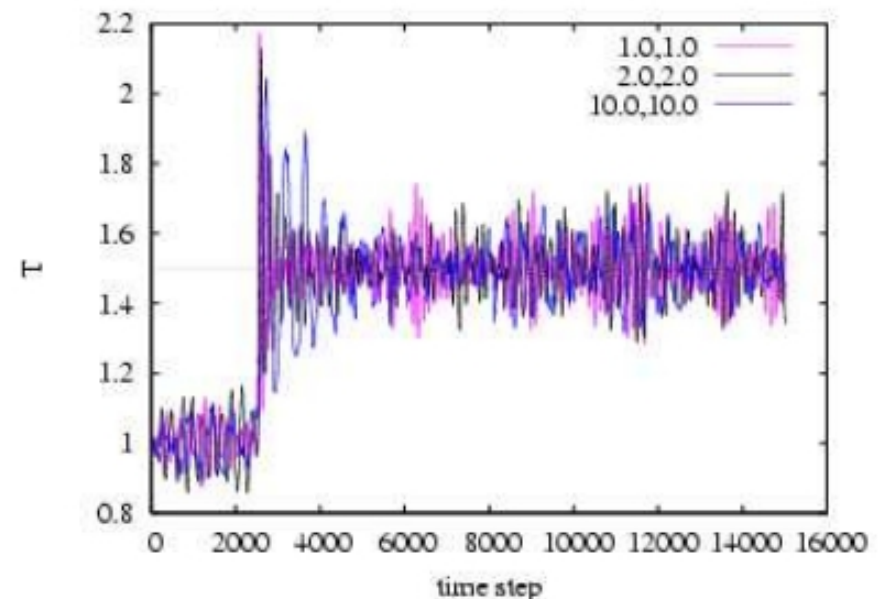
# Nosé-Hoover Thermostat

- The N-H equations of motion are smooth, deterministic and time-reversible. However, just as the Nosé algorithm, Nosé-Hoover dynamics may lead to temperature oscillations.
- too large values of  $Q$  (loose coupling) may cause a poor temperature control, and the canonical distribution will only be obtained after very long simulation times.
- too small values of  $Q$  (tight coupling) generates a microcanonical ensemble, and may cause high-frequency temperature oscillations.

Berendsen thermostat



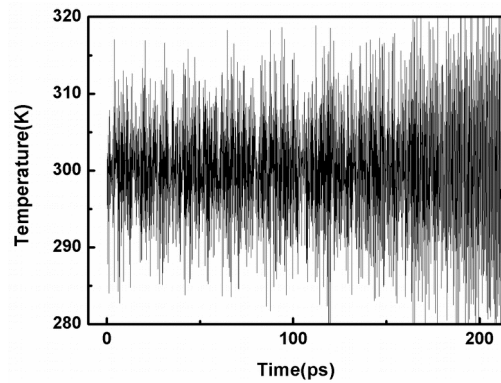
Nosé-Hoover thermostat



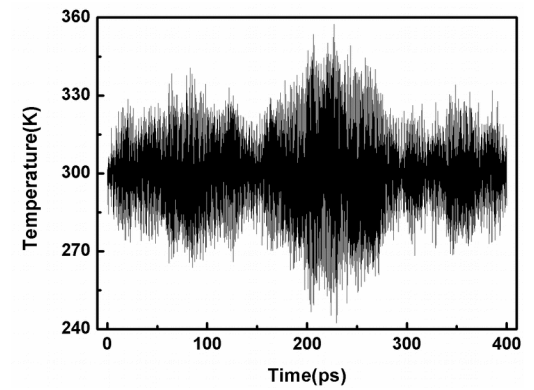


# Nosé-Hoover Thermostat

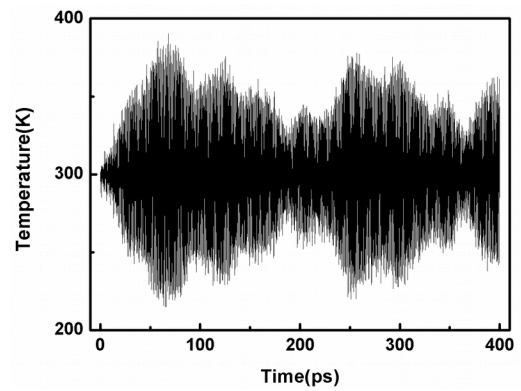
$$\tau = 0.02$$



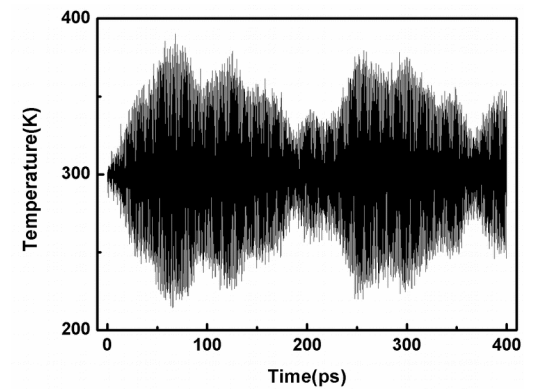
$$\tau = 0.03$$



$$\tau = 0.5$$



$$\tau = 1$$



# Molecular Dynamics: Thermostats

deterministic thermostats include friction coefficients in the atomic equations of motion.

$$\ddot{\mathbf{r}}_i(t) = m_i^{-1} \mathbf{F}_i(t) - \gamma(t) \dot{\mathbf{r}}_i(t)$$