# Simulating thermodynamic ensembles

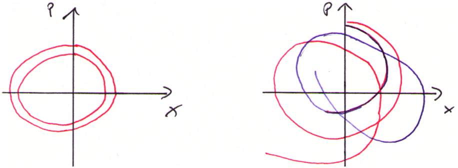
## what you simulate is what you would measure

Marcus Elstner and Tomáš Kubař

2019, May 16

# Intro

- system of classical particles interacting with potential $V$
  – deterministic system
- given initial conditions, $\vec{r}_0$ and $\vec{v}_0$,
  trajectory of the system, $\vec{r}(t)$ and $\vec{v}(t)$,
  is determined for all of the future $t \to \infty$



- so-called chaotic systems – strictly deterministic, too
- chaos – two trajectories that are close in phase space initially
  will depart exponentially from each other
  (solution of the eqns of motion is unstable)

# Intro

stochastic process – whenever we do not have sufficient information
about all of the degrees of freedom of the system
then, we have to describe the system with statistical mechanics

what we need – techniques to control basic simulation parameters
– temperature, optionally pressure
important – do this in such a way that
correct phase space density at given temperature is obtained

phase space – different sampling at high and at low temperatures
– different ensembles will be generated

$T$ – determines if a certain region of phase space shall be visited
= a certain molecular structure shall be observed

# Isolated system

- exchanges with surroundings neither energy (heat / work)
  nor matter (atoms / molecules)
- total energy of system: $E = E_{\text{kin}} + E_{\text{pot}} = \text{const}$
- individually, $E_{\text{kin}}$ and $E_{\text{pot}}$ fluctuate in the course of time
  as they are being transformed into each other
- is what we get when using the Verlet method for a molecule
- trajectory in the microcanonical ensemble

# Isolated system

kinetic theory of gases → relation of $E_{\text{kin}}$ and temperature:

$$\begin{aligned} \langle E_{\text{kin}} \rangle &= \frac{3}{2} NkT \\ \text{comparewith } \langle E_{\text{kin}} \rangle &= \frac{1}{2} \sum_i m_i \left\langle v_i^2 \right\rangle \\ \text{so } T &= \frac{\sum_i m_i \left\langle v_i^2 \right\rangle}{3Nk} \end{aligned}$$

‘local’ / ‘instantaneous’ temperature

$$T = \frac{\sum_i m_i v_i^2}{3Nk}$$

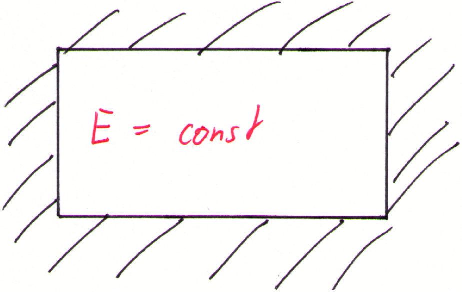
- fluctuates in time
- may differ between different parts of system

# Isolated and closed system

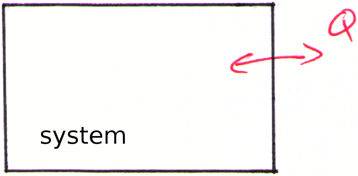
experimental setup (a test tube with a sample)

- usually in thermodynamic equilibrium with the surroundings
- temperature (and opt. pressure) equals that of surroundings

isolated system



closed system



system

environment (bath)

# Closed system

- thermal contact of system with surroundings
- exchange of energy in the form of heat
  until the temperature of surroundings is reached
- strictly speaking: $T$ only defined with such thermal contact
  ($\rightarrow$ N/A in case of isolated system)
- trajectory from MD simulation of such a system
  will provide a canonical ensemble

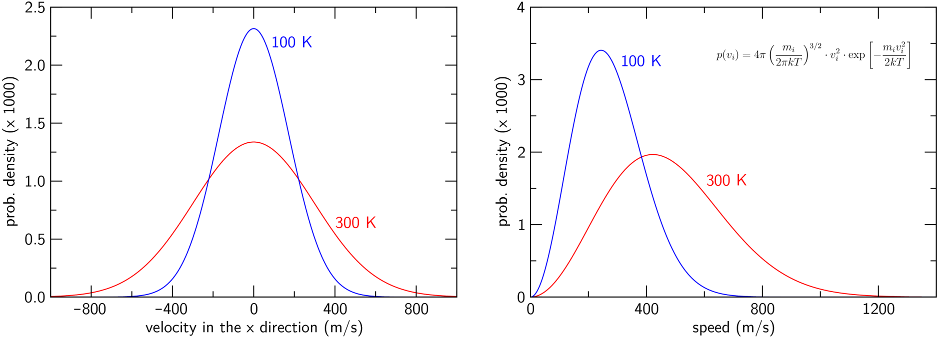
# Canonical ensemble

velocity / speed of atoms – Maxwell–Boltzmann distribution

$$\begin{aligned}
\mathcal{P}(v_{x,i}) &= \sqrt{\frac{m_i}{2\pi kT}} \cdot \exp\left[-\frac{m_i v_{x,i}^2}{2kT}\right] \\
\mathcal{P}(v_i) &= 4\pi \left(\frac{m_i}{2\pi kT}\right)^{3/2} \cdot v_i^2 \cdot \exp\left[-\frac{m_i v_i^2}{2kT}\right]
\end{aligned}$$

# Canonical ensemble

velocity / speed of atoms – Maxwell–Boltzmann distribution



$$p(v_i) = 4\pi \left(\frac{m_i}{2\pi kT}\right)^{3/2} \cdot v_i^2 \cdot \exp\left[-\frac{m_i v_i^2}{2kT}\right]$$

(for an $N_2$ molecule considered as IG)

# Equipartition theorem

Every DoF contains the same average amount of kinetic energy of

$$\left\langle \frac{1}{2} m_i v_{x,i}^2 \right\rangle = \frac{1}{2} kT$$

Each atom $i$ has 3 DoF $x_i$, $y_i$ and $z_i$ (and $v_i^2 = v_{x,i}^2 + v_{y,i}^2 + v_{z,i}^2$) $\rightarrow$

$$\langle E_{\text{kin}} \rangle = \left\langle \sum_i \frac{1}{2} m_i v_i^2 \right\rangle = \left\langle \sum_i \frac{1}{2} m_i v_{x,i}^2 + \frac{1}{2} m_i v_{y,i}^2 + \frac{1}{2} m_i v_{z,i}^2 \right\rangle = \frac{3}{2} NkT$$

Such a distribution of velocity and kinetic energy is a property of systems in contact with heat bath (not of isolated system)

Attention if constraints are being applied:

– reduced number of DoF $3N - N_c$ here and in all what follows!

(Gleichverteilungssatz) (DoF = degree of freedom)

# Equipartition theorem

Virial theorem:

$$\left\langle \sum_k q_k \frac{\partial H}{\partial q_k} \right\rangle = \left\langle \sum_k p_k \frac{\partial H}{\partial p_k} \right\rangle = \left\langle \sum_k p_k \frac{dq_k}{dt} \right\rangle = -\left\langle \sum_k q_k \frac{dp_k}{dt} \right\rangle$$

(DoFs: generalized coordinates $q_k$, generalized momenta $p_k$)

Generalized equipartition theorem – extension of virial theorem:

$$\begin{aligned} \left\langle p_k \frac{\partial H}{\partial p_k} \right\rangle = kT \rightarrow \left\langle \frac{1}{m} \cdot p_k^2 \right\rangle &= kT \rightarrow \quad \left\langle \frac{1}{2} m \cdot v_k^2 \right\rangle = \frac{1}{2} kT \\ \left\langle q_k \frac{\partial H}{\partial q_k} \right\rangle = kT \rightarrow \quad \langle q_k \cdot F_k \rangle &= -kT \\ \rightarrow \quad \left\langle \vec{r}_i \cdot \vec{F}_i \right\rangle &= -3kT \end{aligned}$$

How can we control the temperature in simulation?

An additional algorithm – several options:

- a naïve thermostat
- Berendsen thermostat
- Bussi thermostat (v-rescale)
- Nosé–Hoover thermostat
- Andersen thermostat
- stochastic dynamics

# A naïve thermostat – scaling of velocities

in a Verlet MD simulation – ‘instantaneous temperature’ $T$ deviates from the target $T_{\text{ref}}$ (of bath = the surroundings)

$$T(t) = \frac{2}{3}\frac{E_{\text{kin}}(t)}{Nk} \neq T_{\text{ref}}$$

$T(t)$ – another designation for $E_{\text{kin}}$ determined by velocities

a simple idea – scale the velocities by a certain factor $\lambda$:

$$\begin{aligned} T_{\text{ref}} &= \frac{1}{\frac{3}{2}Nk} \cdot \frac{1}{2}\sum_i m_i \left(\lambda \cdot v_i\right)^2 = \\ &= \lambda^2 \cdot \frac{1}{\frac{3}{2}Nk} \cdot \frac{1}{2}\sum_i m_i v_i^2 \; = \; \lambda^2 \cdot T \end{aligned}$$

scaling of all velocities by $\lambda = \sqrt{T_{\text{ref}}/T} \;\rightarrow\; T_{\text{ref}}$ reached exactly

# A naïve thermostat – scaling of velocities

- very crude way of controlling the temperature
- rescaling of velocities affects the 'natural' way of evolution of the system
- velocities – not sure if the distribution is correct (M–B)
- importantly, system does not sample any canonical ensemble
  – phase space density is not that of a canonical ensemble
  – very important because everything is calculated as averages:

$$\langle A\rangle = \frac{1}{Z}\int \rho\cdot A\,\mathrm{d}\vec{r}\,\mathrm{d}\vec{p}$$

- possibly: wrong sampling → wrong averages

# Berendsen thermostat

How to avoid the drastic changes to the dynamics?

> adjust the velocities more smoothly, in the direction of $T_{\text{ref}}$, resigning on $T_{\text{ref}}$ to be recovered in every step immediately

After every step of MD simulation:

1. measure $T$
2. determine the new temperature to be reached

   $$T > T_{\text{new}} > T_{\text{ref}} \text{ or } T < T_{\text{new}} < T_{\text{ref}}$$

3. perform scaling of velocities to reach $T_{\text{ref}}$

# Berendsen thermostat

- system coupled to infinite bath with temperature $T_{\text{ref}}$
- imagine that $T$ changes between two time steps according to

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{1}{\tau}\left(T_{\text{ref}} - T\right)$$

  – exponential decay of temperature towards $T_{\text{ref}}$:

- rate of change of $T$ (due to the change of velocities)
  is proportional to the deviation of actual $T$ from $T_{\text{ref}}$;
  constant of proportionality – relaxation time $\tau$
- change of $T$ due to thermostat, during 1 finite time step:

$$\Delta T = \frac{\Delta t}{\tau}\left(T_{\text{ref}} - T\right)$$

# Berendsen thermostat

$$\Delta T = \frac{\Delta t}{\tau}\left(T_{\text{ref}} - T\right)$$

- to achieve $\Delta T$, velocities are scaled by $\lambda$ as above:

$$\begin{aligned} T_{\text{new}} &= T + \Delta T = T + \frac{\Delta t}{\tau}\left(T_{\text{ref}} - T\right) \\ \lambda &= \sqrt{\frac{T_{\text{new}}}{T}} = \sqrt{1 + \frac{\Delta t}{\tau}\left(\frac{T_{\text{ref}}}{T} - 1\right)} \end{aligned}$$

- usually: $\tau = 0.1 - 10$ ps
- $T$ is still fluctuating – however around the desired value $T_{\text{ref}}$

# Note: fluctuations

fluctuation of temperature – desired property
for canonical ensemble – variance of 'inst. temperature' $T$:

$$\sigma_T^2 = \left\langle (T - \langle T \rangle)^2 \right\rangle = \left\langle T^2 \right\rangle - \langle T \rangle^2$$

and relative variance

$$\frac{\sigma_T^2}{\langle T \rangle^2} = \frac{2}{3N}$$

large number of atoms $N$: fluctuations → 0
finite-sized systems: visible fluctuation of temperature

- feature of the canonical ensemble
- we would not obtain this with the simple velocity scaling

# Berendsen thermostat

drawbacks:

- correct canonical ensemble is still not generated!
- various parts of the system (different individual molecules,
  or solute × solvent) may exhibit different temperatures,
  while the temperature of the entire system is 'correct'
  – may remain like that for extended periods of time
- gradually moves the energy from the fastest modes of motion
  to the slowest/weakest ones, violating the equipartition
  the fastest – bond stretching and angle bending
  loss of energy → 'freezing' of the molecules
  the slowest – 3 transl'ns (+ 3 rot'ns) of the entire system
  energy gain → 'flying (+ spinning) ice cube'

# Bussi thermostat (v-rescale)

The workflow of an MD simulation:

1. make 1 step of MD with the Verlet integrator
2. calculate the temperature
3. make 1 step of an auxiliary continuous stochastic dynamics
   to propagate the temperature
   – a complicated way to say:
   predict the instantaneous temperature in the right way
4. rescale the velocities to enforce that value of temperature

# Bussi thermostat (v-rescale)

$$\Delta T = \frac{\Delta t}{\tau}\left(T_{\text{ref}} - T\right) + \frac{\Delta W}{\sqrt{\tau}} \cdot 2\sqrt{\frac{T_{\text{ref}}\,T}{N_{\text{DoF}}}}$$

- additional “correction” to Berendsen thermostat
- $\Delta W$ – Wiener process – sequence of random numbers
- no additional parameters (except for a random seed)
- produces a correct canonical ensemble

# Nosé–Hoover thermostat

- conceptionally and mathematically $>$ difficult to understand
- heat bath is treated not as an external element
  rather as an integral part of the system
- the bath – an additional DoF $s$ with fictitious mass $Q$
  may be understood as time-scaling parameter:

$$\mathrm{d}t' = s \cdot \mathrm{d}t$$

- eqns of motion will be propagated for this extended system,
  for which an energy-like quantity will be conserved
- generates canonical NVT ensemble of the molecular system!

# Nosé–Hoover thermostat

expression for the energy of the system involves the bath:

$$E_{\mathrm{pot}} = U(r) + g \cdot kT_{\mathrm{ref}} \cdot \log s$$

- $g$ – number of DoF of the system $= 3N + 1$
- $T_{\mathrm{ref}}$ – reference temperature

$$E_{\mathrm{kin}} = \sum_i \frac{1}{2} m_i s^2 \dot{r}_i'^2 + \frac{1}{2} Q s'^2$$

- attention needed – derivatives w.r.t. the modified time:

$$\mathrm{d}t' \neq \mathrm{d}t \rightarrow \dot{r}_i' \neq \dot{r}_i$$

# Nosé–Hoover thermostat

- eqns of motion for the extended system ($3N + 1$ DoF) derived with the Hamiltonian formalism:

$$H(r', \dot{r}', s, \dot{s}') \quad = \quad E_{\text{pot}} + E_{\text{kin}}$$

- eqns for the molecular DoF:

$$\frac{\mathrm{d}r_i}{\mathrm{d}t'} = \frac{1}{m_i} \cdot \frac{\partial H}{\partial \dot{r}_i'} \qquad \frac{\mathrm{d}\dot{r}_i'}{\mathrm{d}t'} = -\frac{1}{m_i} \cdot \frac{\partial H}{\partial r_i'}$$

- eqns for the additional DoF:

$$\frac{\mathrm{d}s}{\mathrm{d}t'} = \frac{1}{Q} \cdot \frac{\partial H}{\partial \dot{s}'} \qquad \frac{\mathrm{d}\dot{s}'}{\mathrm{d}t'} = -\frac{1}{Q} \cdot \frac{\partial H}{\partial s'}$$

# Nosé–Hoover thermostat

- we obtain these eqns of motion:

$$\ddot{r}_i' = \frac{F_i}{m_i} \cdot \frac{1}{s^2} - \frac{2\dot{s}'}{s} \cdot \dot{r}_i'$$

black – usual Newtonian eqns of motion
red – bath $s$ incorporated into the propagation,
rather than correcting velocities a posteriori

- one more eqn of motion – for the bath coordinate $s$:

$$\ddot{s}' = \frac{1}{Qs} \left( \sum_i m_i s^2 \dot{r}_i^2 - g \cdot kT_{\mathrm{ref}} \right)$$

# Nosé–Hoover thermostat

such eqns are impractical because

- they work with transformed velocities $\dot{r}'$ and accelerations $\ddot{r}'$
- time steps are not equally long ($\Delta t' = s \cdot \Delta t$)

to make things clearer, we

- return from the modified time scale $t'$ to the usual $t$
- transform the eqns to the usual vaiables $\dot{r}$, $\ddot{r}$
- for the 'velocity of bath': pass from $\dot{s}$ to $\gamma = \frac{\dot{s}}{s}$

# Nosé–Hoover thermostat

final form of the eqns of motion:

$$\ddot{r}_i = \frac{F_i}{m_i} - \gamma \cdot \dot{r}_i$$

2nd term: formally – a kind of 'friction' (bath)

$$\dot{\gamma} = \frac{1}{Q}\left(T - T_{\text{ref}}\right)$$

note: $\sum_i \frac{1}{2} m_i \dot{r}_i^2 = 3N \cdot \frac{1}{2} kT$

# Nosé–Hoover thermostat

- strength of coupling controlled by $Q$ – more intuitively time $\tau$:

$$Q = \frac{\tau^2 \cdot T_{\mathrm{ref}}}{4\pi^2}$$

  – meaning of period of oscillation of the kinetic energy
- difference between $\tau$ in Berendsen and in Nosé–Hoover:
  Berendsen – exponential damping of $\Delta T$ with $\tau$
  Nosé–Hoover – oscillatory relaxation of $T$ with period $\tau$
- N–H thermostat – incorporated in eqns of motion
  – inseparable part of the integrator,
  rather than an a posteriori correction
- generates canonical phase-space density, used frequently

# Andersen thermostat

from time to time, some particles (atoms) are selected randomly to undergo a 'collision' with the particles of a heat bath, which changes their velocities suddenly

this algorithm has a stochastic character:

- start MD with a standard integrator (Verlet. . . )
- select randomly the atoms that shall be hit by the bath
- for these atoms, draw new velocities from Maxwell–Boltzmann distribution, and keep all of the other atoms untouched

# Andersen thermostat

advantage: generates canonical ensemble (if implemented right)
– rate of collisions must be neither too low (inefficient)
nor too high (collisions would dominate over eqns of motion)

disadvantage: no continuity of momentum
- dynamic properties incorrect (diffusion, viscosity. . . )

MD simulations of a molecule / molecular complex in vacuo
- Verlet integrator: no energy flow between modes of motion
  - equipartition of energy possibly violated
- Andersen thermostat is a viable solution
  (other stochastic algorithms applicable, see further on)

# Langevin dynamics

a.k.a. stochastic dynamics
additional terms in the eqns of motion:

$$m_i \ddot{\vec{r}}_i = \vec{F}_i - m_i \gamma \dot{\vec{r}}_i + \sqrt{\frac{2 m_i \gamma k T_{\mathrm{ref}}}{\Delta t}} \vec{R}_i(t)$$

- velocity-dependent friction $m_i \gamma \dot{\vec{r}}_i$ (removes energy)
  - shall be $\gamma \Delta t \ll 1$, related to diffusivity: $D = kT/m\gamma$
  - frictional drag (like, of solvent imposed on a solute)
- random force or noise $\vec{R}_i$ (adds energy)
  - from a standard normal distribution, uncorrelated
  - if higher $T_{\mathrm{ref}}$ desired → larger additional force required
  - random collisions (like, with solvent molecules)

modified eqns of motion → the integrator has to be modified

# Brownian dynamics

overdamped Langevin dynamics, or dynamics without inertia

– no acceleration takes place, $\ddot{\vec{r}}_i = \vec{o}$:

$$\cancel{m_i \ddot{\vec{r}}_i} = \vec{F}_i - m_i \gamma \dot{\vec{r}}_i + \sqrt{\frac{2 m_i \gamma k T_{\text{ref}}}{\Delta t}} \vec{R}_i(t)$$

$$\dot{\vec{r}}_i = \frac{\vec{F}_i}{m_i \gamma} + \sqrt{\frac{2 k T_{\text{ref}}}{m_i \gamma \Delta t}} \vec{R}_i(t)$$

- 1st order ODE – integration: $\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \dot{\vec{r}}_i(t) \Delta t$
- over-damped → long time step $\Delta t$ possible
- very different from explicit all-atom MD simulation
- applications: large molecules without explicit solvent
  polymers, proteins, nucleic acid molecules, colloid systems

# Controlling temperature

Summary

- scaling of velocities (Berendsen, v-rescale)
- extended Lagrangian (Nosé–Hoover)
- stochastic algorithms (Andersen, Langevin)

# Introducing pressure

Chemical reality – constant pressure rather than constant volume

How to calculate pressure?

From the (generalized) equipartition theorem:

$$\frac{1}{3}\left\langle \sum_i \vec{r}_i \cdot \vec{F}_i^{\text{tot}} \right\rangle = \frac{1}{3}\left\langle \sum_i \vec{r}_i \left( \vec{F}_i + \vec{F}_i^{\text{ext}} \right) \right\rangle = -NkT$$

$\vec{F}_i^{\text{ext}}$ – force due to external pressure:

$$\frac{1}{3}\left\langle \sum_i \vec{r}_i \cdot \vec{F}_i^{\text{ext}} \right\rangle = -PV$$

$\vec{F}_i$ – force from interactions between atoms

introduce the virial of force:

$$\Xi = -\frac{1}{3}\sum_i \vec{r}_i \cdot \vec{F}_i$$

# Introducing pressure

Adding the components of forces:

$$-PV - \langle \Xi \rangle = -NkT$$

Consider 'instantaneous' temperature and pressure from now on:

$$\begin{aligned} -PV - \Xi &= -NkT \\ P &= \frac{NkT}{V} - \frac{1}{V} \cdot \Xi = \frac{2}{3} E_{\text{kin}} - \frac{1}{V} \cdot \Xi \end{aligned}$$

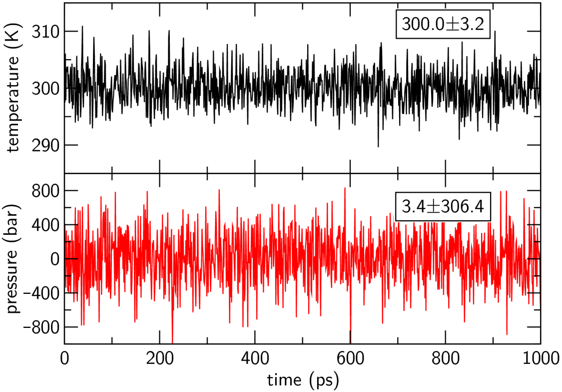
Pass to pairwise forces $\vec{F}_{ij}$ between atoms on distances $\vec{r}_{ij}$:

$$\Xi = -\frac{1}{3} \sum_i \vec{r}_i \vec{F}_i = -\frac{1}{3} \sum_{i<j} \vec{r}_{ij} \vec{F}_{ij}$$

$$P = \frac{1}{3V} \left( \sum_i m_i |\vec{v}_i|^2 + \sum_{i<j} \vec{r}_{ij} \vec{F}_{ij} \right)$$

# Measuring pressure

- no mention of the idea of particles colliding with the wall
- virial pressure fluctuates greatly and may even be negative:



(DNA oligomer in water at $T_{\text{ref}} = 300$ K, $P_{\text{ref}} = 1.0$ bar)

# Controlling pressure

we can calculate the pressure
    – so how do we maintain it at a constant value?
barostat – algorithm that is equivalent of a thermostat,
    and it varies volume of the simulation box instead of velocities

the scaling of the volume is usually isotropic,
    except for special systems (e.g. membranes)
        it shall be semi-isotropic (xy+z) for such geometries

several options are available:

# Berendsen barostat

- equivalent to the Berendsen thermostat
- molecular system coupled to a 'force / pressure bath' – piston
- direct rescaling of box lengths and atom coordinates:

$$\begin{aligned} \mu &= 1 - \frac{\beta}{3}\frac{\Delta t}{\tau}\left(P_{\text{ref}} - P\right) \\ \vec{r}_i^{\,\prime} &= \mu \cdot \vec{r}_i \end{aligned}$$

- $\beta$ – compressibility; $\beta = 0.000045\ \text{bar}^{-1}$ for $H_2O$

# Parrinello–Rahman barostat

- extended-ensemble barostat – much like Nosé–Hoover algo.
- eqns of motion contain box lengths $b$ as additional DoFs:

$$\ddot{r}_i = \frac{F_i}{m_i} - \frac{\dot{b}}{b} \cdot \dot{r}_i$$

- additional eqn of motion for the dimensions of the box:

$$\ddot{b} = \frac{V}{b} \cdot W^{-1} \cdot (P - P_{\text{ref}})$$

- strength of coupling – due to mass parameter $W^{-1}$:

$$W^{-1} = \frac{4\pi^2}{3} \frac{\beta}{\tau^2} \frac{1}{L}$$

$\tau$ – relaxation time (parameter)