Advanced Molecular Dynamics

Lyopanov instability and the shadow theorem (4.3.4)
Lagrangian and Hamiltonians (Appendix A)
MD at constant temperature (6.1)
Constraints (15.1)
Computer experiments (4.4)

Contents

- Recap of the Verlet algorithms
- Lyopunov instability and the shadow theorem
- Lagrangian and Hamiltonian approach
- MD at constant temperature: Thermostats
- Constraints
- Ensemble averaging and computing static and dynamic properties
- Example of biomolecular simulation

Recap: The Verlet algorithms

Original Verlet algorithm

$$\mathbf{r}(t+\Delta t) \approx 2\mathbf{r}(t) - \mathbf{r}(t-\Delta t) + \frac{\Delta t^2}{m}\mathbf{f}(t)$$

Downside regular verlet algorithm: velocity is not known, worse accuracy.

Velocity verlet (Andersen 1983):

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

$$\mathbf{v}(t+\Delta t) \approx \mathbf{v}(t) + \frac{\Delta t}{2m}\left[\mathbf{f}(t+\Delta t) + \mathbf{f}(t)\right]$$

(Is based on Trotter decomposition of Liouville operator formulation,
also basis of Multiple time steps).

Velocity Verlet has the advantage of allowing multiple time stepping

Is Verlet a good algorithm?

Verlet algorithm

- is time reversible
- does conserve volume in phase space
- (is “symplectic”)
- does not suffer from energy drift

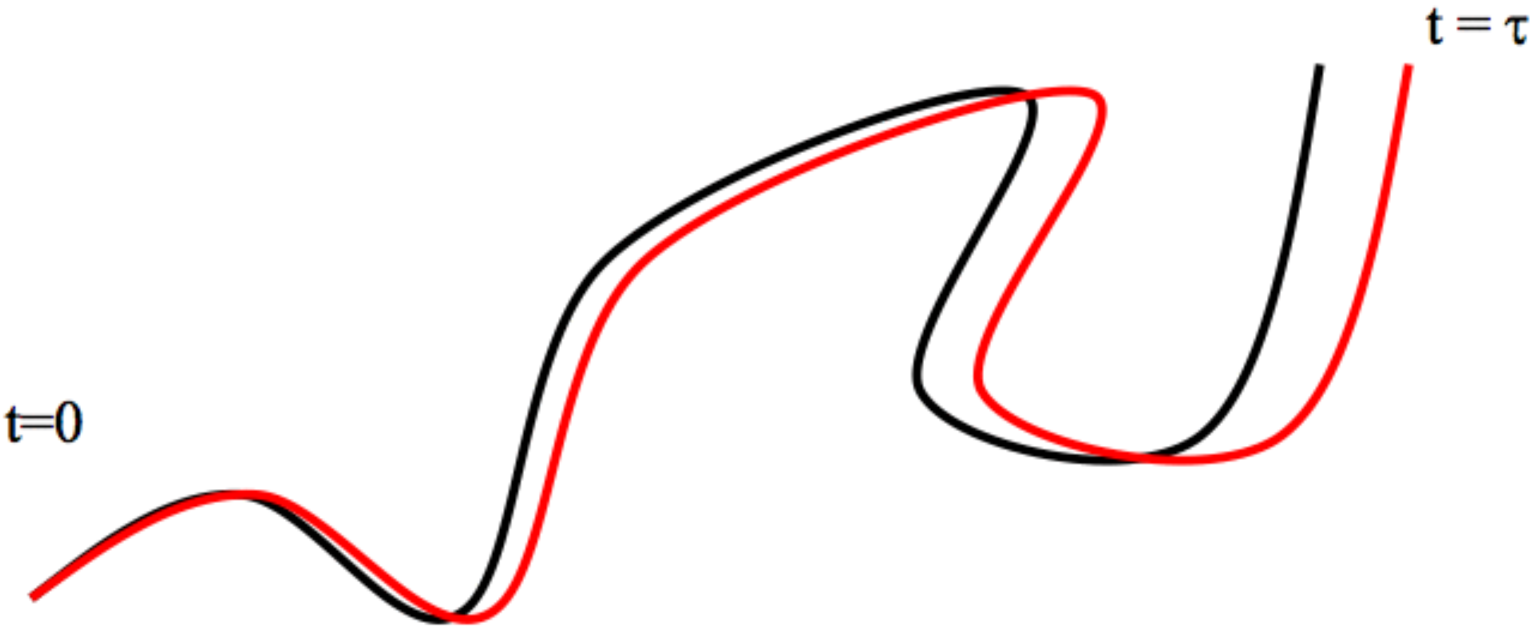
...but is it a good algorithm?

i.e. does it predict the time evolution of the system correctly???

Molecular chaos

Dynamics of “well-behaved” classical many-body system is chaotic.

Consequence: Trajectories that differ very slightly in their initial conditions diverge exponentially (“Lyapunov instability”)



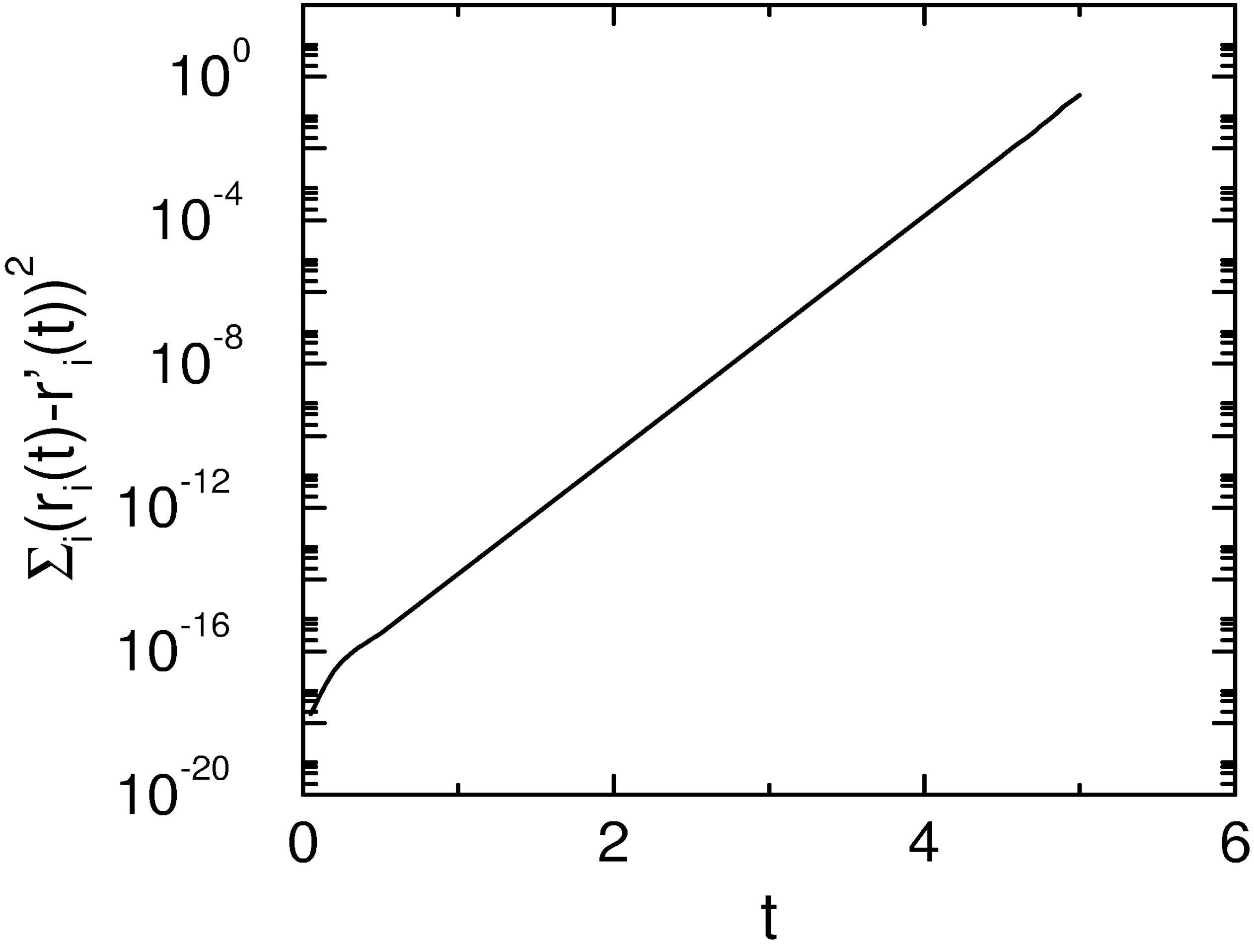
Lyapunov instability

The Lyapunov disaster in action...

$$\left(\mathbf{r}^N(0),\mathbf{p}^N(0)\right)$$

$$\left(\mathbf{r}^N(0),\mathbf{p}_1(0),\cdots,\mathbf{p}_j(0)+\varepsilon,\mathbf{p}_i(0)-\varepsilon\cdots,\mathbf{p}_N(0)\right)$$

$$\varepsilon = 10^{-10}$$



Any small error in the numerical integration of the equations of motion, will blow up exponentially....

always...

...and for any algorithm!!

SO:

Why should anyone believe Molecular Dynamics simulations ???

Answers:

1. In fact, one should not…

2. Good MD algorithms (e.g. Verlet) can also be considered as good Monte Carlo algorithms –they therefore yield reliable STATIC properties ("Hybrid Monte Carlo")

3. What is the point of simulating dynamics, if we cannot trust the resulting time-evolution???

4. All is well (probably), because of...

   The Shadow Theorem.

Shadow theorem

- For any realistic many-body system, the shadow theorem is merely a hypothesis.
- It basically states that Good algorithms generate numerical trajectories that are “close to” a REAL trajectory of the many-body system.
- Question: Does the Verlet algorithm indeed generate “shadow” trajectories?
- Take a different look at the problem.
  - Do not discretize NEWTON’s equation of motion…
  - ...but discretize the ACTION

Lagrangian Classical mechanics

- Newton: $$F_i = m_i \frac{d^2 x_i(t)}{dt^2}$$

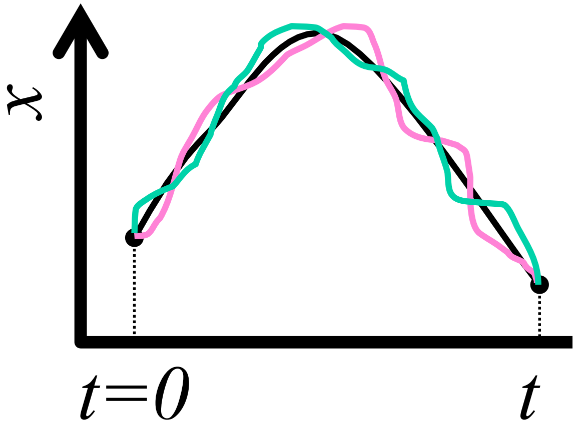
- Lagrange:
  - Consider a system that is at a point r_0 at time t=0 and at point r_t at time t=t, then the system follows a trajectory r(t) such that:

$$S \equiv \int_0^t dt' \mathcal{L}(r(t'))$$

is an extremum. The Lagrangian L is defined as:

$$\mathcal{L}(r(t)) = K - U(r)$$

kinetic energy



$S<S$

$S<S$

Langrangian

For example, if we use cartesian coordinates:

$$\mathcal{L}(r(t)) = \sum_{i=1}^{N} \frac{1}{2} m_i \dot{r}_i^2 - U(r_1, r_2, \ldots r_N)$$

What does this lead to?

Consider the “true” path R(t), with R(0)=r_0 and R(t)=r_t.

Now, consider a path close to the true path:

$$r(t') = R(t') + \delta r(t')$$

Then the action S is an extremum if

$$\frac{\partial S}{\partial r(t')} = 0 \text{ for all } t$$

What does this lead to?

Discretized action

$$S_{cont} = \int_{t_0}^{t_1} dt \mathcal{L}(t)$$

$$S_{disc} = \Delta t \sum_{i=0}^{i_{max}} \mathcal{L}(t_i) \qquad \mathcal{L}(t_i) = K(t_i) - U(t_i)$$

For a one dimensional system this becomes

$$\mathcal{L}(t_i)\Delta t = \frac{1}{2} m \Delta t \frac{(x_{i+1} - x_i)^2}{\Delta t^2} - U(x_i)\Delta t$$

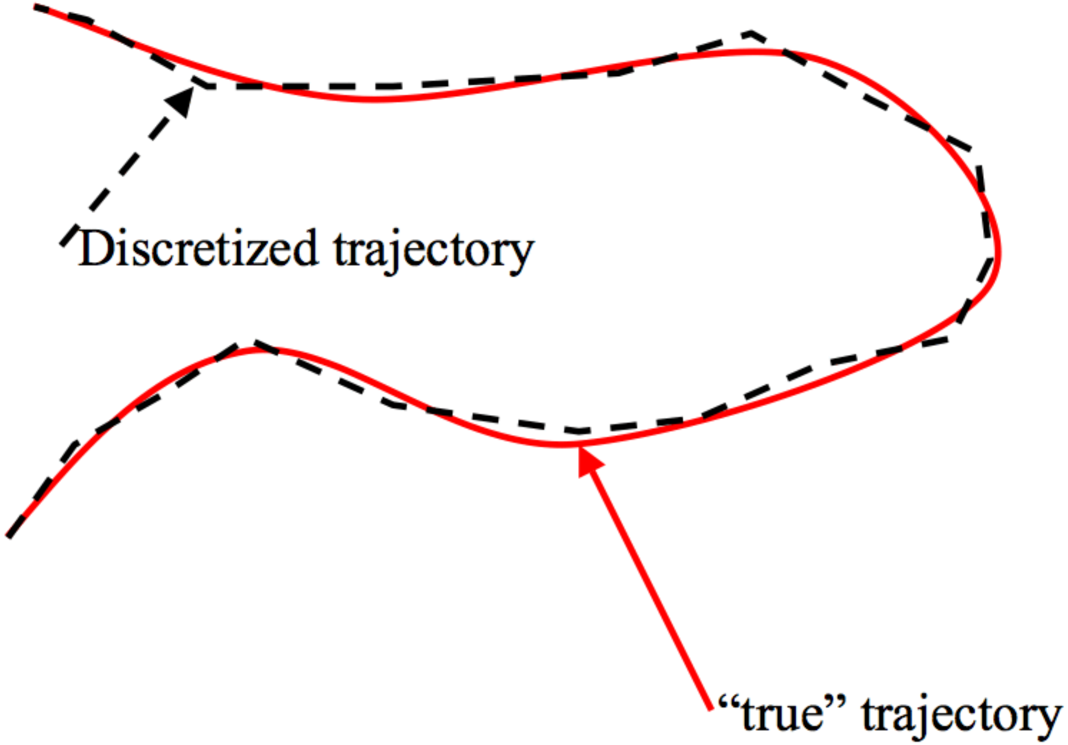
$$S_{disc} = \sum_{i=1}^{i_{max}} \left[\frac{m(x_{i+1} - x_i)^2}{2\Delta t} - U(x_i)\Delta t \right]$$

Minimize the action

Now do the standard thing: Find the extremum for small variations in the path, i.e. for small variations in all x_i.

$$\frac{\partial S_{disc}}{\partial x_i} = 0 \ \text{ for all } i$$

This will generate a discretized trajectory that starts at time t_0 at X, and ends at time t at X_t.



Minimizing the action

$$\frac{\partial S_{disc}}{\partial x_i} = \frac{\partial}{\partial x_i} \sum_{i=1}^{i_{max}} \left[\frac{m(x_{i+1} - x_i)^2}{2\Delta t} - U(x_i)\Delta t \right]$$

$$\frac{\partial S_{disc}}{\partial x_i} = \frac{-m(x_{i+1} - x_i) + m(x_i - x_{i-1})}{\Delta t} - \Delta t \frac{\partial U(x_i)}{\partial x_i}$$

$$0 = \frac{m}{\Delta t} \left(2x_i - x_{i+1} - x_{i-1} - \frac{\Delta t^2}{m} \frac{\partial U(x_i)}{\partial x_i} \right)$$

$$0 = 2x_i - x_{i+1} - x_{i-1} - \frac{\Delta t^2}{m} \frac{\partial U(x_i)}{\partial x_i}$$

$$x_{i+1} = 2x_i - x_{i-1} + \frac{\Delta t^2}{m} F(x_i)$$

- which is the Verlet algorithm!

- The Verlet algorithm generates a trajectory that satisfies the boundary conditions of a REAL trajectory –both at the beginning and at the endpoint.

- Hence, if we are interested in statistical information about the dynamics (e.g. time-correlation functions, transport coefficients, power spectra...) ...then a “good” MD algorithm (e.g. Verlet) is fine.

Contents

- Recap of the Verlet algorithms
- Lyopunov instability and the shadow theorem
- Lagrangian and Hamiltonian approach
- MD at constant temperature: Thermostats
- Constraints
- Ensemble averaging and computing static and dynamic properties
- Example of biomolecular simulation

Lagrangian approach

Lagrangian is sum of two terms

$$\mathcal{L}(\dot{r}, r) = K(\dot{r}) - U(r) = \frac{m\dot{r}^2}{2} + U(r)$$

$$\frac{\partial \mathcal{L}}{\partial \dot{r}} = \frac{\partial K}{\partial \dot{r}} = p \qquad \frac{\partial \mathcal{L}}{\partial r} = -\frac{\partial U}{\partial r} = F$$

$$p = \frac{\partial \mathcal{L}(\dot{r}, r)}{\partial \dot{r}} \qquad \dot{p} = \frac{\partial \mathcal{L}(\dot{r}, r)}{\partial r}$$

Newton : F=ma

Hamiltonian approach

$$L(r,\dot{r}) \qquad p = \frac{\partial L(r,\dot{r})}{\partial \dot{r}} \qquad \dot{p} = \frac{\partial L(r,\dot{r})}{\partial r}$$

$$H(r,p) = \dot{r}p - L(r,\dot{r})$$

$$\mathrm{dH}(r,p) = \mathrm{d}(\dot{r}p) - \mathrm{dL}(r,\dot{r})$$

$$= \dot{r}\mathrm{d}(p) \cancel{+ \dot{p}\mathrm{d}(\dot{r})} - \left[\frac{\partial L(r,\dot{r})}{\partial r}\mathrm{dr} \cancel{+ \frac{\partial L(r,\dot{r})}{\partial \dot{r}}\mathrm{d}\dot{r}}\right]$$

$$= -\dot{p}\mathrm{dr} + \dot{r}(\mathrm{d}p)$$

$$\mathrm{dH}(r,p) = \left(\frac{\partial H}{\partial r}\right)\mathrm{dr} + \left(\frac{\partial H}{\partial p}\right)\mathrm{d}p$$

$$\begin{cases} \dot{r} = \dfrac{\partial H}{\partial p} \\[2ex] \dot{p} = -\dfrac{\partial H}{\partial p} \end{cases}$$

Hamilton’s equations of motion

Hamiltonian approach

The Hamiltonian is defined as

$$H(p,r) = p\dot{r} - \mathcal{L}(\dot{r},r)$$

$$H(p^N,r^N) = U(r^N) + \sum_i \frac{p_i^2}{2m_i}$$

Hamilton’s equations are then

$$\dot{r} = \frac{\partial\mathcal{H}(r,p)}{\partial p} = \frac{p}{m}$$

$$\dot{p} = -\frac{\partial\mathcal{H}(r,p)}{\partial r} = -\frac{\partial U(r^N)}{\partial r}$$

Integrating equations of motion (by Verlet) conserves the Hamiltonian

Conservation of Hamiltonian

$$dH(p,r) = \frac{\partial H}{\partial p} dp + \frac{\partial H}{\partial r} dr$$

$$\frac{\partial H}{\partial p} = \dot{r} \qquad \frac{\partial H}{\partial r} = -\dot{p}$$

$$\frac{dH(p,r)}{dt} = \frac{\partial H}{\partial p} \dot{p} + \frac{\partial H}{\partial r} \dot{r} = \dot{r}\dot{p} - \dot{p}\dot{r} = 0$$

So a solution to the Hamiltonians equation conserves the TOTAL energy

$$E = K + U$$

MD generates the NVE ensemble

How do we sample the canonical ensemble with MD?

Contents

- Recap of the Verlet algorithms
- Lyopunov instability and the shadow theorem
- Lagrangian and Hamiltonian approach
- MD at constant temperature: Thermostats
- Constraints
- Ensemble averaging and computing static and dynamic properties
- Example of biomolecular simulation

Constant Temperature: a naïve approach

Velocity scaling

$$\frac{3}{2}k_B T = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}mv_i^2$$

$$v_i \rightarrow v_i\sqrt{\frac{T_{\mathrm{req}}}{\mathrm{T}}}$$

This is called the **isokinetic** thermostat

Do we sample the canonical ensemble?

Partition function

$$Q_{NVT} = \frac{1}{h^{3N} N!} \int \mathrm{dp}^N \exp\left[-\beta \sum p_i^2 / 2m\right] \int \mathrm{dr}^N \exp\left[-\beta U\left(\mathrm{r}^N\right)\right]$$

Maxwell-Boltzmann velocity distribution

$$P(p) = \left(\frac{\beta}{2\pi m}\right)^{3/2} \exp\left[-\beta p^2 / 2m\right]$$

$$\left\langle p^2 \right\rangle = \int \mathrm{dp} P(p) p^2$$

$$= \left(\frac{\beta}{2\pi m}\right)^{3/2} \int \mathrm{d}p 4\pi p^4 \exp\left[-\beta p^2 / 2m\right]$$

$$= \frac{3m}{\beta}$$

$$P(p)=\left(\frac{\beta}{2\pi\right)^{3/2}$$

$$\left\langle p^2\right\rangle=\int \mathrm{dp}\qquad m\right] \quad =\frac{3m}{\beta}$$

$$\left\langle p^4\right\rangle=\int \mathrm{dp}$$

Fluctuations in t

$$\frac{\sigma_{p^2}^2}{\left\langle p^2\right\rangle^2}=\frac{\left\langle p\right.}{}$$

Fluctuations in t

$$\frac{\sigma_{k_BT}^2}{\left\langle k_BT\right\rangle^2}=\frac{\left\langle\left(\right.\right.}{}$$

$$\left\langle k_BT\right\rangle=\left\langle\frac{1}{3N}\sum_{i=1}^{N}p_i^2/m\right\rangle=\frac{1}{3Nm}N\left\langle p^2\right\rangle$$

$$\left\langle\left(k_BT\right)^2\right\rangle=\left\langle\frac{1}{\left(3mN\right)^2}\left(\sum_{i=1}^{N}p_i^2\right)^2\right\rangle$$

$$=\left\langle\frac{1}{\left(3mN\right)^2}\left(\sum_{i=1}^{N}p_i^4+\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}p_i^2p_j^2\right)\right\rangle$$

$$=\frac{1}{\left(3mN\right)^2}\left(N\left\langle p^4\right\rangle-N\left(N-1\right)\left\langle p^2\right\rangle^2\right)$$

$$\frac{\sigma_{k_BT}^2}{\left\langle k_BT\right\rangle^2}=\frac{N\left\langle p^4\right\rangle-N\left(N-1\right)\left\langle p^2\right\rangle^2-\left(N\left\langle p^2\right\rangle\right)^2}{\left(N\left\langle p^2\right\rangle\right)^2}$$

$$=\frac{1}{N}\frac{\left\langle p^4\right\rangle-\left\langle p^2\right\rangle^2}{\left\langle p^2\right\rangle^2}=\frac{2}{3N}$$

So, no the canonical ensemble is not isokinetic

Thermostat: From NVE to NVT

- Introduce thermostat in MD trajectory:

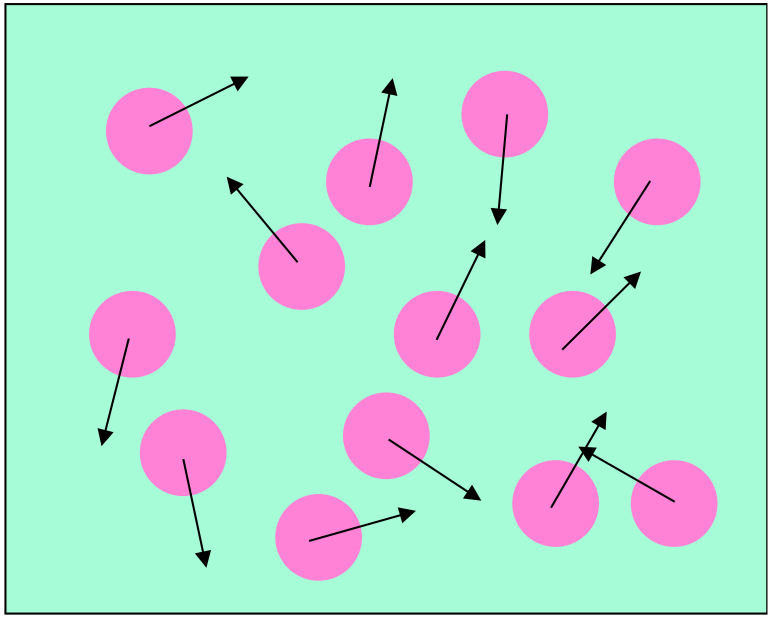
- deterministic thermostat:
  - **Nose-Hoover**
- stochastic thermostats:
  - **Andersen**
  - Langevin
  - **Bussi**
  - **Nose Hoover- Langevin**

All of these alter the velocities such that the trajectory samples the canonical NVT ensemble, and the partition function becomes

$$Q = \frac{1}{N!\Lambda^{3N}} \int e^{-\beta \mathcal{U}(r)} dr$$

These thermostats differ in how they achieve this

Andersen Thermostat



- Every particle has a fixed probability to collide with the Andersen demon
- After collision the particle is give a new velocity

$$P(v) = \left(\frac{\beta}{2\pi m} \right)^{3/2} \exp\left[-\beta m v^2 / 2 \right]$$

- The probabilities to collide are uncorrelated (Poisson distribution)

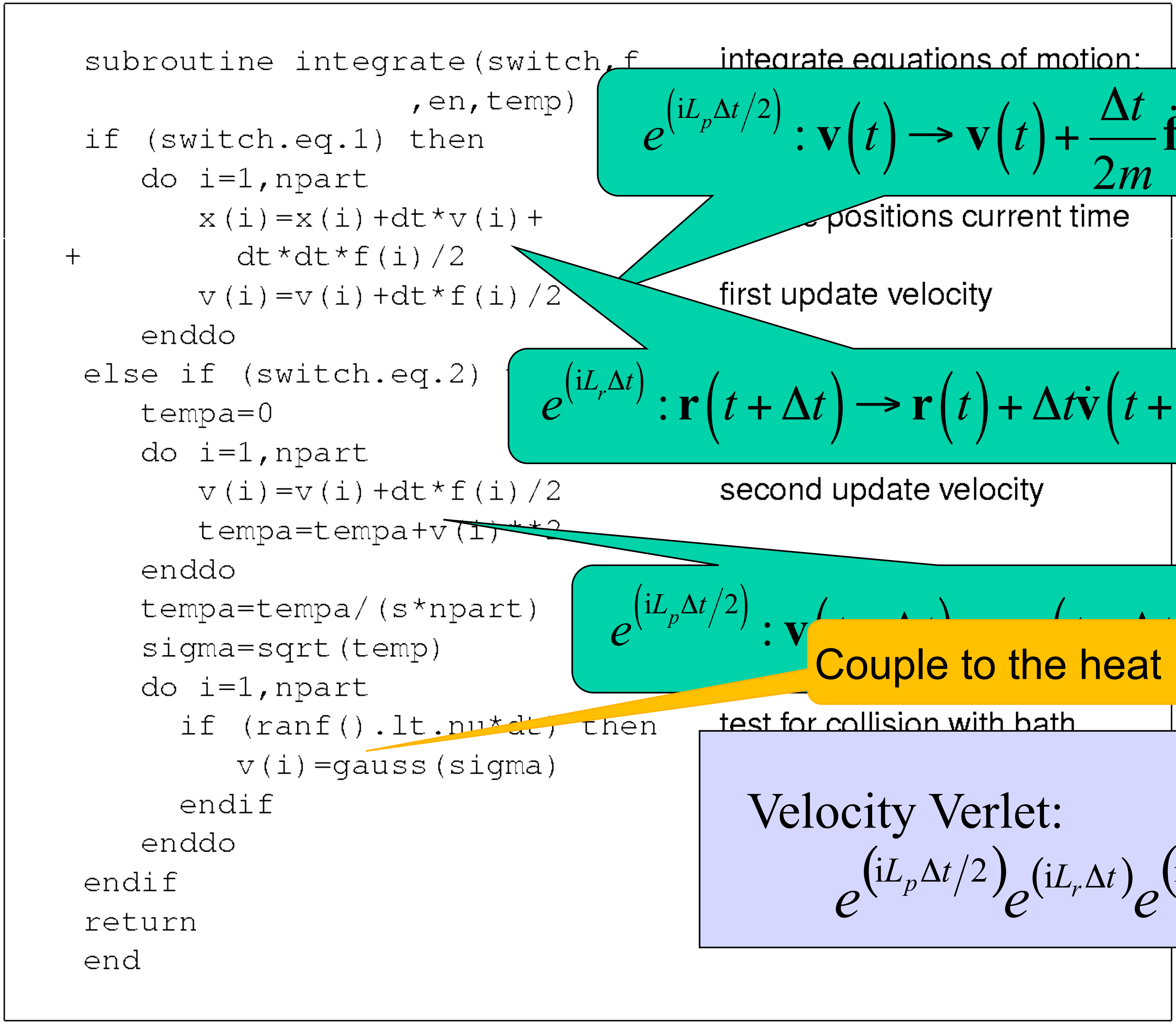
$$P(t;\nu) = \nu \exp\left[-\nu t \right]$$

Algorithm 14 (Molecular Dynamics: Andersen Thermostat)

```
program md_Andersen                          MD at constant temperature
call init(temp)                              initialization
call force(f,en)                             determine the forces
t=0
do while (t.lt.tmax)                         MD loop
   call integrate(1,f,en,temp)               first part of the eqs. of motion
   call force(f,en)                          determine the forces
   call integrate(2,f,en,temp)               second part of eqs. of motion
   t=t+dt
   call sample                               sample averages
enddo
stop
end
```

$$
\begin{aligned}
r(t+\Delta t) &= r(t) + v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 \\
v(t+\Delta t) &= v(t) + \frac{f(t+\Delta t) + f(t)}{2m}\Delta t.
\end{aligned}
$$

Algorithm 15 (Equations of Motion: Andersen Thermostat)



```
   subroutine integrate(switch,f
                          ,en,temp)
   if (switch.eq.1) then
      do i=1,npart
         x(i)=x(i)+dt*v(i)+
+           dt*dt*f(i)/2
         v(i)=v(i)+dt*f(i)/2
      enddo
   else if (switch.eq.2) t
      tempa=0
      do i=1,npart
         v(i)=v(i)+dt*f(i)/2
         tempa=tempa+v(i)**2
      enddo
      tempa=tempa/(s*npart)
      sigma=sqrt(temp)
      do i=1,npart
        if (ranf().lt.nu*dt) then
           v(i)=gauss(sigma)
        endif
      enddo
   endif
   return
   end
```

integrate equations of motion:

positions current time

first update velocity

second update velocity

test for collision with bath

$$e^{(iL_p\Delta t/2)}:\mathbf{v}(t)\rightarrow\mathbf{v}(t)+\frac{\Delta t}{2m}\dot{\mathbf{f}}(0)$$

$$e^{(iL_r\Delta t)}:\mathbf{r}(t+\Delta t)\rightarrow\mathbf{r}(t)+\Delta t\dot{\mathbf{v}}(t+\Delta t/2)$$

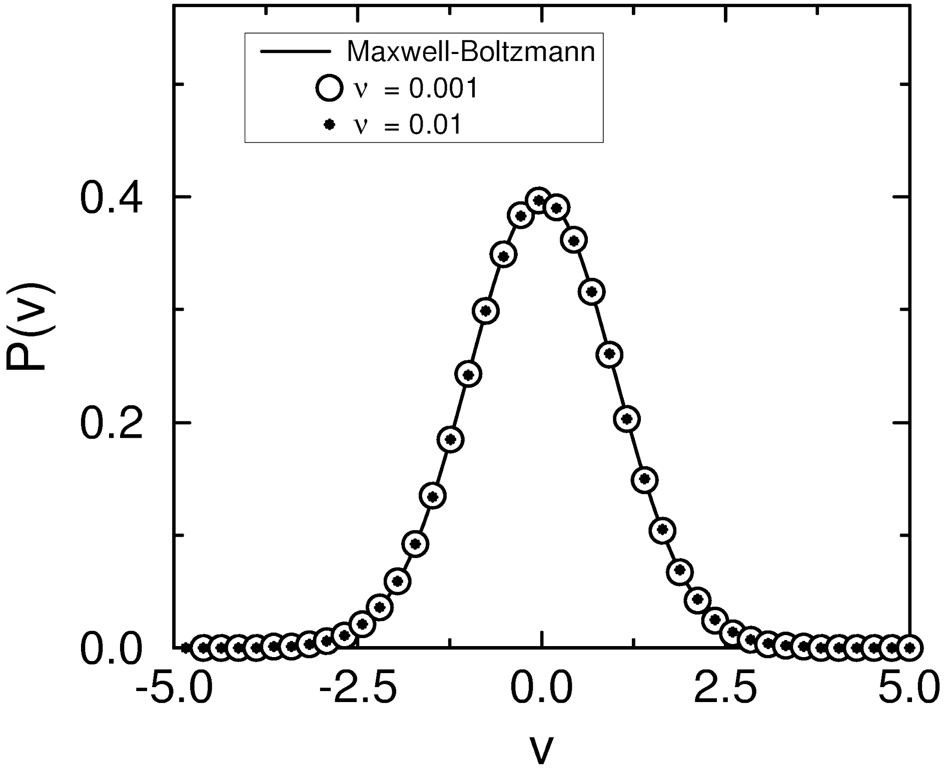
$$e^{(iL_p\Delta t/2)}:\mathbf{v}\quad\frac{\Delta t}{2m}\dot{\mathbf{f}}(t)$$

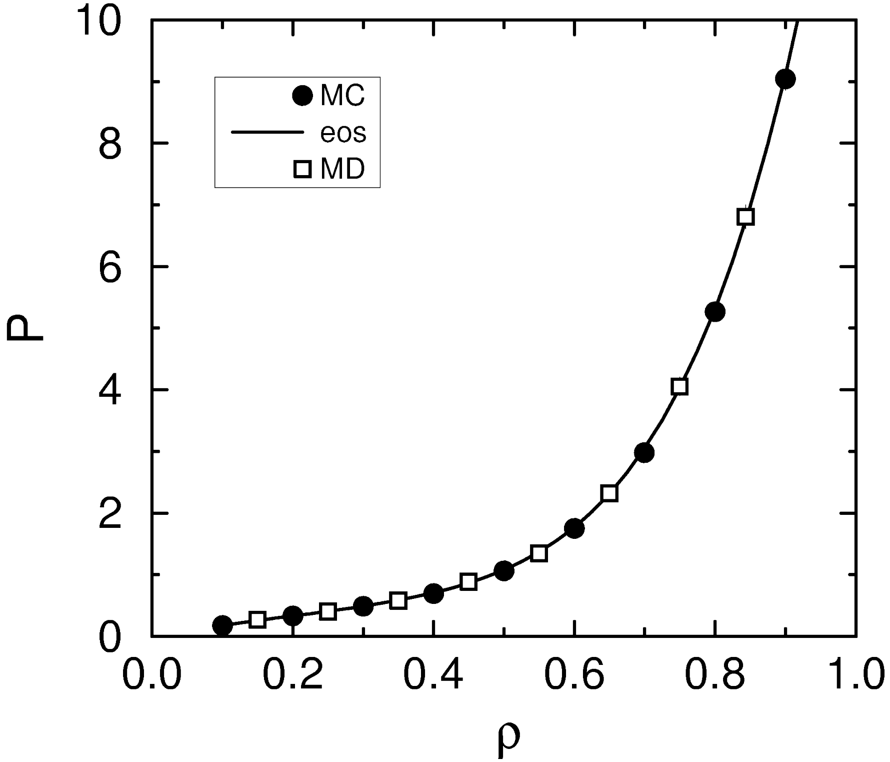
Couple to the heat bath

Velocity Verlet:

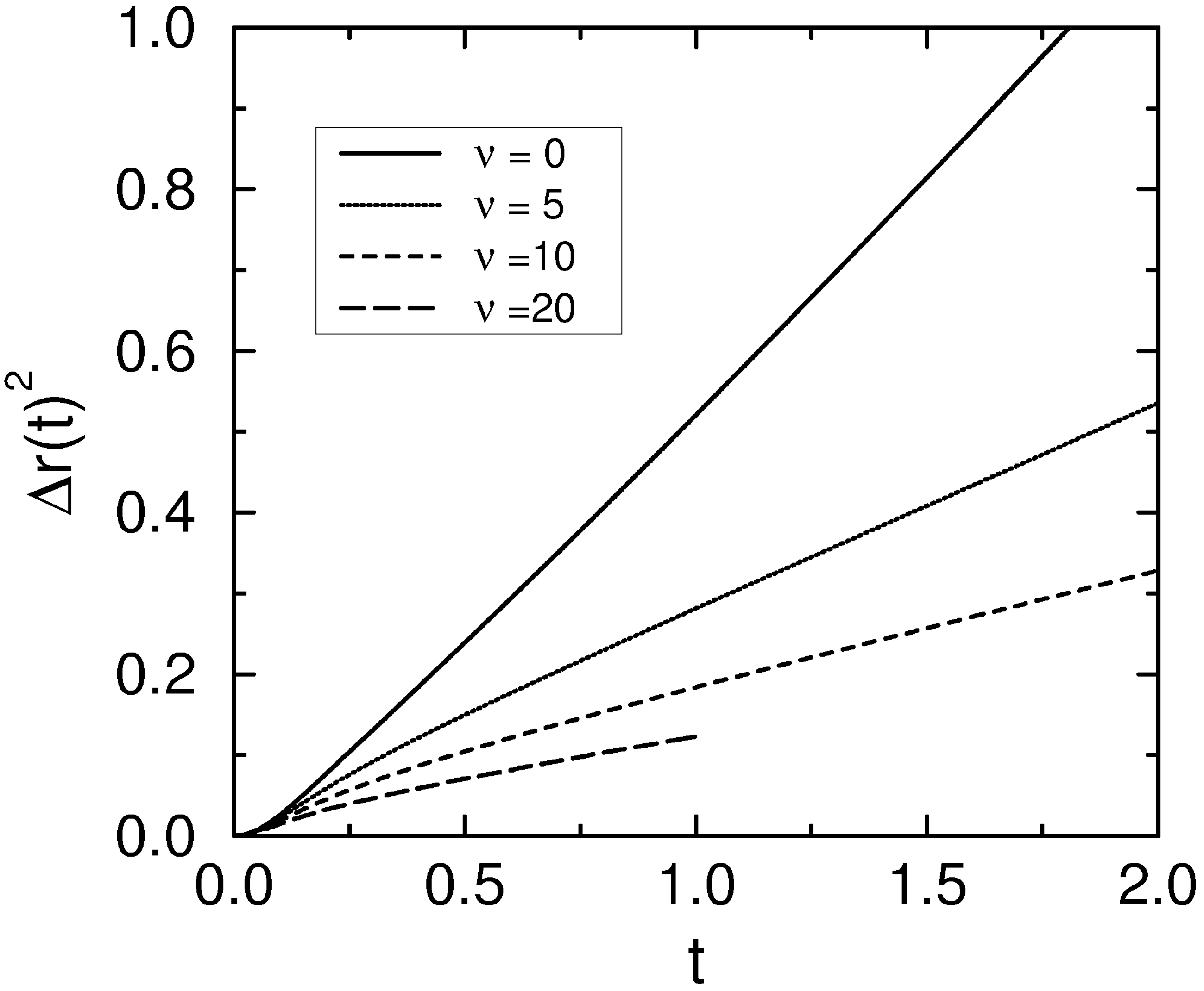
$$e^{(iL_p\Delta t/2)}e^{(iL_r\Delta t)}e^{(iL_p\Delta t/2)}$$

Andersen thermostat: static properties





Andersen thermostat: dynamic properties



Velocity rescaling revisited

- The Andersen thermostat samples the correct ensemble, but is rather violent to the dynamics.

- The **Bussi** thermostat combines rescaling with a stochastic approach:
  - Evolve a single time step with a symplectic integrator e.g. velocity Verlet
  - Calculate the kinetic energy K_t .
  - Evolve the kinetic energy $K_t = K_t + dK$ for a t
    single time step using an auxiliary continuo

Wiener noise term

$$dK = (\bar{K} - K)\frac{dt}{\tau} + 2\sqrt{\frac{K\bar{K}}{N_f}}\frac{dW}{\sqrt{\tau}}.$$

relaxation time parameter

  - Rescale the velocities so as to enforce thi
    energy K_t.

$$\alpha = \sqrt{\frac{K_t}{K}}$$

Bussi thermostat

Solving the differential equation

$$dK = (\bar{K} - K)\frac{dt}{\tau} + 2\sqrt{\frac{K\bar{K}}{N_f}}\frac{dW}{\sqrt{\tau}}.$$

gives

$$\alpha^2 = e^{-\Delta t/\tau} + \frac{\bar{K}}{N_f K}(1 - e^{-\Delta t/\tau})\left(R_1^2 + \sum_{i=2}^{N_f} R_i^2\right)$$

$$+ 2e^{-\Delta t/2\tau}\sqrt{\frac{\bar{K}}{N_f K}(1 - e^{-\Delta t/\tau})}R_1.$$

(Bussi et al 2007)

Where the Ri are independent Gaussian random numbers

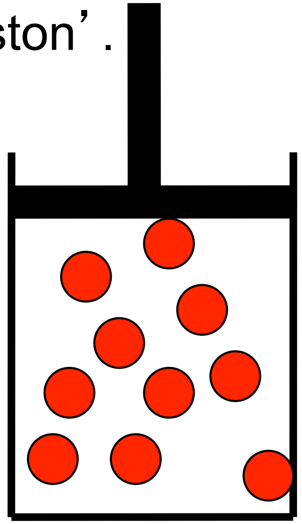
Nose Hoover thermostat

goal: compute MD trajectory sampling NVT ensemble.

Take kinetic energy out of the system and put it back in via a ‘piston’.

piston can be seen as additional variable *s* storing kinetic energy

Approach: extended Lagrangian



$$\mathcal{L}_{\text{Nose}} = \sum_{i=1}^{N} \frac{1}{2} m s^2 \dot{r}_i^2 - U\left(r^N\right) + \frac{1}{2} Q \dot{s}^2 - \frac{g}{\beta} \ln s$$

extended variable

effective mass

constant to be set

$$p_i = \frac{\partial \mathcal{L}}{\partial \dot{r}_i} = m s^2 \dot{r}_i \qquad p_s = \frac{\partial \mathcal{L}}{\partial \dot{s}} = Q \dot{s}$$

$$\mathcal{H}_{\text{Nose}} = \sum_{i=1}^{N} \frac{p_i^2}{2ms^2} + \frac{p_s}{2Q} + U\left(r^N\right) + \frac{g}{\beta} \ln s$$

Nose-Hoover Thermostat

$$H_{\text{Nose}} = \sum_{i=1}^{N} \frac{p_i^2}{2ms^2} + \frac{p_s}{2Q} + U\left(r^N\right) + \frac{g}{\beta} \ln s$$

now define

$$\mathcal{H}(p', r) = \sum_{i=1}^{N} \frac{p_i'^2}{2m_i} + U(r^N) \qquad\qquad p' = p/s$$

then it is possible to show that the partition function Z_{nose} is

$$Z_{nose} \propto \frac{1}{N!} \int dp'^N dr^N \exp\left[-\beta \frac{3N+1}{g} \mathcal{H}(p', r)\right]$$

for g=3N+1 the system samples the canonical distribution if p’ is interpreted as the real momentum

$$\delta\left(h(s)\right) = \frac{\delta\left(s - s_0\right)}{h'(s)}$$

$$h(s) = H\left(p',r\right) + \frac{p_s^2}{2Q} + \frac{g}{\beta}\ln s - E$$

$$h'(s) = \frac{g}{\beta s}$$

$$s_0 = \exp\left\{\frac{\beta}{g}\left[E - H\left(p',r\right) - \frac{p_s^2}{2Q}\right]\right\}$$

Recall

MD $Q_{NVE} = \frac{1}{N!}\int \mathrm{d}p^N \int \mathrm{d}r^N \delta\left(E - H\left(r^N,p^N\right)\right)$

MC $Q_{NVT} = \frac{1}{N!}\int \mathrm{d}p^N \int \mathrm{d}r^N \exp\left[-\beta H\left(r^N,p^N\right)\right]$

$$Q_{\mathrm{No}} = \frac{1}{N!}\int \mathrm{d}p_s \int \mathrm{d}p'^N \int \mathrm{d}r^N \int \mathrm{d}s\, \delta\left(H\left(p',r\right) + \frac{p_s^2}{2Q} + \frac{g}{\beta}\ln s - E\right)$$

$$= \frac{1}{N!}\int \mathrm{d}p_s \int \mathrm{d}p'^N \int \mathrm{d}r^N \int \mathrm{d}s\, \frac{\beta s^{3N+1}}{g}\, \delta\left(s - e^{\frac{\beta}{g}\left[E - H\left(p',r\right) - \frac{p_s^2}{2Q}\right]}\right)$$

$$= \frac{1}{N!}\int \mathrm{d}p_s \int \mathrm{d}p'^N \int \mathrm{d}r^N \frac{\beta}{g} e^{\frac{\beta(3N+1)}{g}\left[E - H\left(p',r\right) - \frac{p_s^2}{2Q}\right]}$$

Gaussian integral

$$= \frac{C}{N!}\int \mathrm{d}p'^N \int \mathrm{d}r^N e^{-\frac{\beta(3N+1)}{g}H\left(p',r\right)} \qquad g = 3N+1$$

Constant plays no role in thermodynamics

$$= \frac{C}{N!}\int \mathrm{d}p'^N \int \mathrm{d}r^N e^{-\beta H\left(p',r\right)}$$

Nose-Hoover Thermostat

$$H_{\text{Nose}} = \sum_{i=1}^{N} \frac{p_i^2}{2ms^2} + \frac{p_s}{2Q} + U\left(r^N\right) + \frac{g}{\beta} \ln s$$

equations of motion follow from Hamilton's equations.

$$\frac{\mathrm{d}r_i}{\mathrm{d}t} = \frac{\partial H_{\text{Nose}}}{\partial p_i} = \frac{p_i}{ms^2} \qquad \frac{\mathrm{d}p_i}{\mathrm{d}t} = -\frac{\partial H_{\text{Nose}}}{\partial r_i} = -\frac{\partial U\left(r^N\right)}{\partial r_i}$$

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \frac{\partial H_{\text{Nose}}}{\partial p_s} = \frac{p_s}{Q} \qquad \frac{\mathrm{d}p_s}{\mathrm{d}t} = -\frac{\partial H_{\text{Nose}}}{\partial s} = \frac{1}{s}\left(\sum \frac{p_i^2}{ms^2} - \frac{g}{\beta}\right)$$

Nose Hoover implementation

NH equation of motion can be rewritten as (Hoover 1984)

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

$$\dot{\mathbf{p}}_i = \mathbf{f}_i - \xi \mathbf{p}_i$$

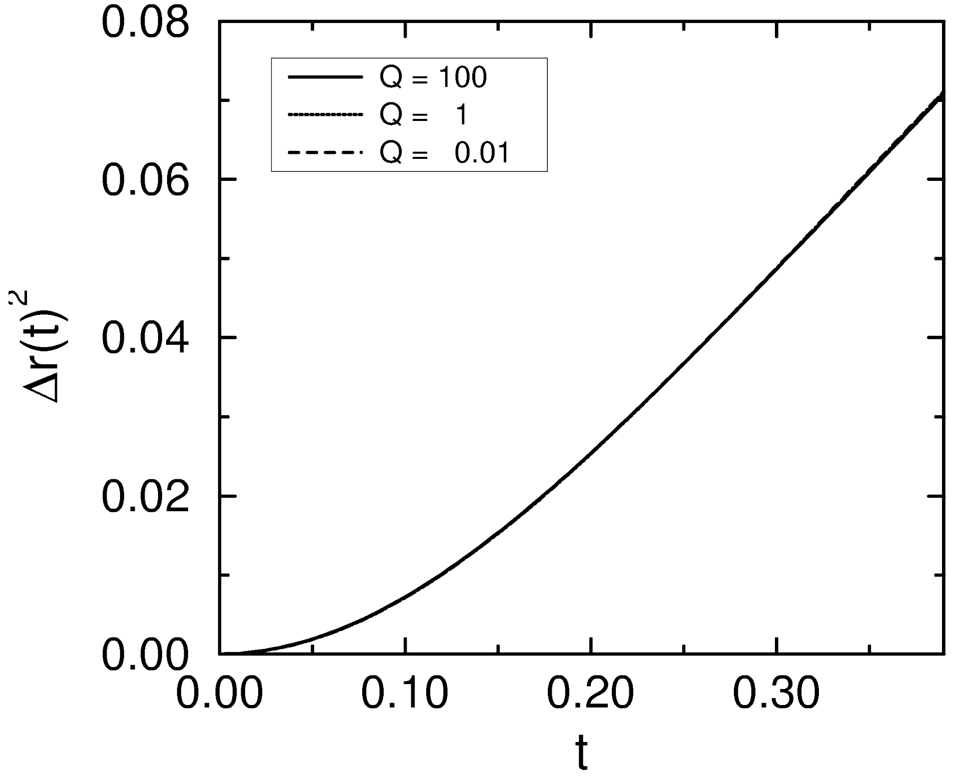
$$\dot{\xi} = \left(\sum_i p_i^2 / m_i - 3N/\beta \right) /Q$$

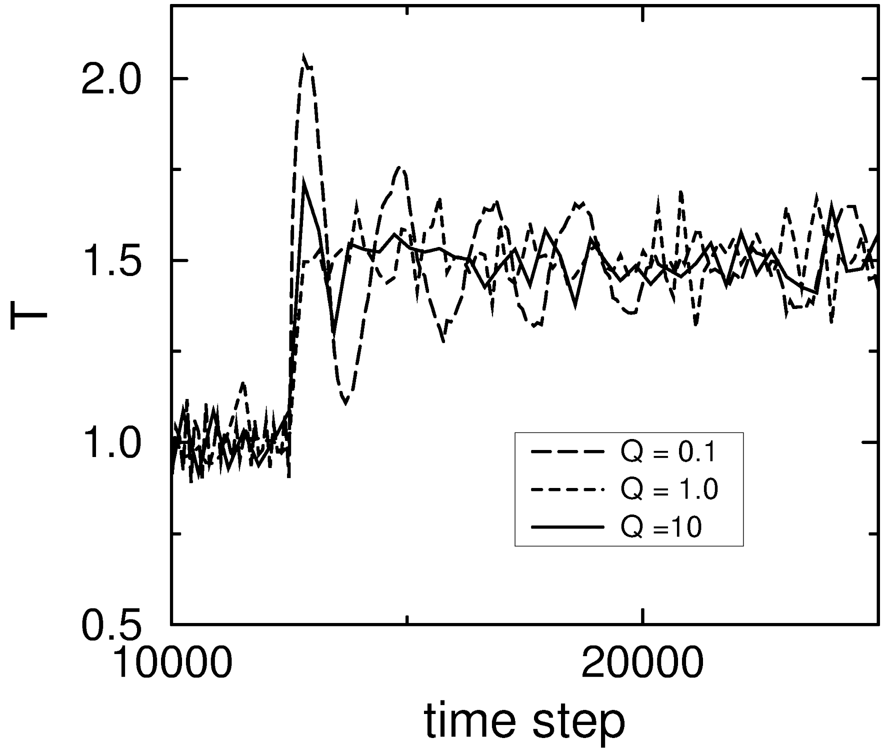
Where $\xi = p_s/Q$ now denotes a kind of ‘friction’ term

Mass Q determines the damping

Effect of mass Q

Lennard-Jones fluid





mean square displacement

temperature relaxation

Nose Hoover Langevin thermostat

- Nose Hoover is not ergodic (NH Chains can alleviate this)
- Better option might be Nose-Hoover-Langevin (Leimkuhler 2012)

$$\begin{aligned}\dot{q} &= p \\ \dot{p} &= f(q) - \xi p \\ \dot{s} &= (p^2 - k_{\mathrm{B}} T)/\mu - \gamma \xi + \sigma \dot{w}\end{aligned}$$

μ *is* mass

Friction and noise term

- Newton, Nosé-Hoover, Ornstein-Uhlenbeck
- Trotter formula, e.g.:

$$\begin{aligned}e^{-\Delta t \hat{L}} \approx\; & e^{-(\Delta t/2)\hat{L}_{\mathrm{NH}}} + e^{-(\Delta t/2)\hat{L}_p} + e^{-(\Delta t/2)\hat{L}_q} + e^{-\Delta t \hat{L}_{\mathrm{OU}}} \\ & + e^{-(\Delta t/2)\hat{L}_q} + e^{-(\Delta t/2)\hat{L}_p} + e^{-(\Delta t/2)\hat{L}_{\mathrm{NH}}}\end{aligned}$$

- with: $\hat{L} = \hat{L}_q + \hat{L}_p + \hat{L}_{\mathrm{NH}} + \hat{L}_{\mathrm{OU}}$
  and $P(p, q; t + \Delta t) = e^{-\Delta t \hat{L}} P(p, q; t)$

Implementation of NHL thermostat

(Leimkuhler 2012) $c_1 = \exp(-\gamma \frac{\Delta t}{2}) \quad c_2 = \sqrt{\frac{2 k_B T}{\mu}(1 - c_1^2)}$

$$
\begin{aligned}
\xi &:= c_1 \xi + c_2 R \\
p &:= e^{-\xi \Delta t/2} p \\
p &:= p + \frac{\Delta t}{2} f(q) \\
q &:= q + \frac{\Delta t}{2} p \\
\xi &:= \xi + \Delta t (2 E_{\text{kin}} - N_f k_B T)/\mu \\
q &:= q + \frac{\Delta t}{2} p
\end{aligned}
$$

force call

$$
\begin{aligned}
p &:= p + \frac{\Delta t}{2} f(q) \\
p &:= e^{-\xi \Delta t/2} p \\
\xi &:= c_1 \xi + c_2 R
\end{aligned}
$$

Note the similarity to Velocity rescaling (Bussi)

Constant Pressure

- The pressure can be kept constant using a similar extended Lagrangian formalism as the Nose Hoover thermostat.

- rectangular boxes
  - Andersen barostat,
  - Martyna-Tuckerman-Klein barostat,
  - Nosé-Hoover Langevin barostat
- variable box shape
  - Parrinello Rahman barostat

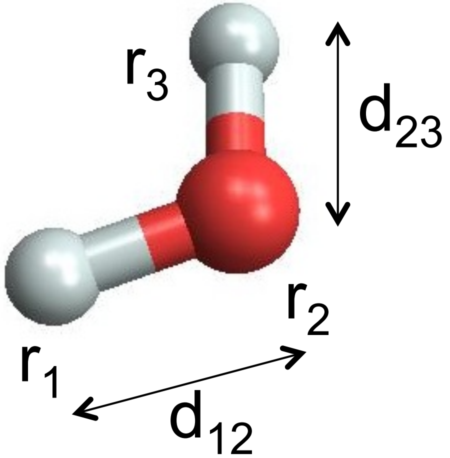
Contents

- Recap of the Verlet algorithms
- Lyopunov instability and the shadow theorem
- Lagrangian and Hamiltonian approach
- MD at constant temperature: Thermostats
- Constraints
- Ensemble averaging and computing static and dynamic properties
- Example of biomolecular simulation

Constraints

- Constraints are used to eliminate high frequency bond vibration in e.g. CH bonds by fixing the the bond length.

$$m_1\ddot{r}_1 = f_1 + g_1$$
$$m_2\ddot{r}_2 = f_2 + g_2$$
$$m_3\ddot{r}_3 = f_3 + g_3$$



- The g's are the constraint forces needed to keep the distances constant

$$\chi_{12} = r_{12}^2(t) - d_{12} = 0$$
$$\chi_{23} = r_{23}^2(t) - d_{23} = 0$$

- The constraint force is then

$$\mathbf{g}_a = \frac{1}{2}\lambda_{12}\nabla_{\mathbf{r_a}}\chi_{12} + \frac{1}{2}\lambda_{23}\nabla_{\mathbf{r_a}}\chi_{23}$$

Lagrange multiplier

SHAKE algorithm

- SHAKE algorithm approximates **g** by g$^{(r)}$

$$m_a \ddot{\mathbf{r}}_a = \mathbf{f}_a + \mathbf{g}_a \approx \mathbf{f}_a + \mathbf{g}_a^{(r)}$$

- This force **g**$^{(r)}$ enters into the Verlet algorithm as

$$r_a(t+\Delta t) = r_a'(t+\Delta t) + (\Delta t^2 / m_a)_a \mathbf{g}_a^{(r)}(t)$$

- This constraint force must be directed along the bond, and obey Newton's 3rd law (action = - reaction)

$$g_1^{(r)} = \lambda_{12} r_{12}$$

$$g_2^{(r)} = \lambda_{23} r_{23} - \lambda_{12} r_{12}$$

$$g_3^{(r)} = -\lambda_{23} r_{23}$$

- The λ's are solved iteratively.

- The similar RATTLE algorithm handles constraints in Velocity Verlet
- Other methods exist, such as LINCS

Contents

- Recap of the Verlet algorithms
- Lyopunov instability and the shadow theorem
- Lagrangian and Hamiltonian approach
- MD at constant temperature: Thermostats
- Constraints
- Ensemble averaging and computing static and dynamic properties
- Example of biomolecular simulation

Ensemble averages by ergodicity

time averages over a NVT MD trajectory

$$\bar{A} = \frac{1}{\mathcal{T}} \int_0^{\mathcal{T}} A(t) dt$$

ensemble average

$$\langle A \rangle = \frac{\int dr^N A(r^N) \exp(-\beta \mathcal{U}(r^N))}{\int dr^N \exp(-\beta \mathcal{U}(r^N))}$$

Ergodicity theorem states that for an ‘ergodic system’

$$\bar{A} = \langle A \rangle$$

Computing equilibrium properties

Ensemble averages follow from time averages

$$\langle A \rangle = \bar{A} = \frac{1}{\mathcal{T}} \int_0^{\mathcal{T}} A(t)dt$$

Temperature follows from equipartition: ½ $k_B T$ per d.o.f.

$$T = \frac{\langle 2K \rangle}{k_B f} \qquad K = \frac{1}{2}\sum_{i=1}^{N} m v_i^2$$

Where f is number of degrees of freedom

Pressure follows from virial expression

$$P = \frac{N k_B T}{V} + \frac{1}{3V}\left\langle \sum_{i<j}^{N} \mathbf{f}_{ij}\mathbf{r}_{ij} \right\rangle$$

Computing free energy landscape

Statistical mechanics gives us

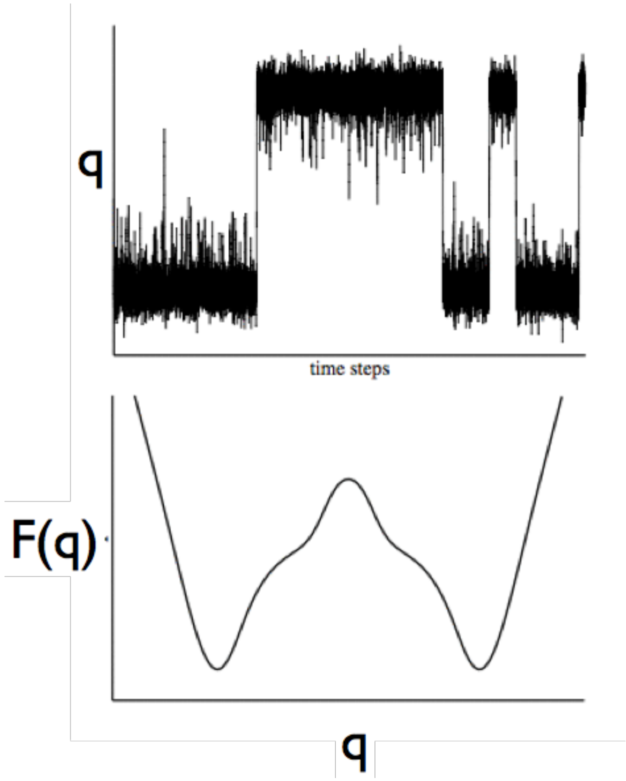
$$F = -k_B T \ln Q = -k_B T \ln \int dr^N e^{-\beta U(r^N)}$$

Project free energy on collective variable q

$$F(q) = -k_B T \ln \int dr^N \delta(q - q(r^N)) e^{-\beta U(r^N)}$$

$$F(q) = -k_B T \ln \langle \delta(q - q(r^N)) \rangle$$

$$= -k_B T \ln P(q)$$



The result is a Landau free energy

This can be generalized to multiple dimensions

$$\beta F(x, y) = -\ln P(x, y)$$

Transport coefficients: Diffusion

Diffusion equation (Fick's second law)

$$\frac{\partial c(x,t)}{\partial t} = D\frac{\partial^2 c(x,t)}{\partial x^2}$$

Solution for an initial c(x,0)=δ(0): all molecules at origin

$$c(x,t) = \frac{1}{\sqrt{4\pi Dt}} \exp\left(-\frac{x^2}{4Dt}\right)$$

Mean square displacement of the molecules

$$\langle x^2(t)\rangle = \frac{\int dx x^2 c(x,t)}{\int dx c(x,t)} = \int dx \frac{x^2 e^{-\frac{x^2}{4Dt}}}{\sqrt{4\pi Dt}} = 2Dt$$

Time derivative gives

$$2D = \lim_{t\to\infty} \frac{d\left\langle x^2(t)\right\rangle}{dt}$$

General c(x,t)

$$\frac{\partial}{\partial t}\int dx\ x^2 c(x,t) = D\int dx\ x^2 \frac{\partial^2 c(x,t)}{\partial\, x^2}.$$

$$2D = \lim_{t\to\infty} \frac{d\left\langle x^2(t)\right\rangle}{dt}$$

Diffusion in 3 dimensions

$$\boxed{D = \frac{1}{6}\lim_{t\to\infty}\frac{d\langle \mathbf{r}^2(t)\rangle}{dt}}$$

Relation to velocity

Relation to velocity

$$\Delta x(t) = \int_0^t dt' \; v_x(t').$$

$$\left\langle x^2(t) \right\rangle = \left\langle \left(\int_0^t dt' \; v_x(t') \right)^2 \right\rangle$$

$$= \int_0^t \int_0^t dt' dt'' \left\langle v_x(t') v_x(t'') \right\rangle$$

$$= 2 \int_0^t \int_0^{t'} dt' dt'' \left\langle v_x(t') v_x(t'') \right\rangle .$$

$$\left\langle v_x(t') v_x(t'') \right\rangle = \left\langle v_x(t' - t'') v_x(0) \right\rangle .$$

$$\frac{\partial\langle x^2(t)\rangle}{\partial t} = 2\frac{\partial}{\partial t}\int_0^t dt' \int_0^{t'} dt'' \langle v_x(t')v_x(t'')\rangle$$

$$\Big\langle v_x(t')v_x(t'')\Big\rangle = \Big\langle v_x(t'-t'')v_x(0)\Big\rangle\,.$$

$$2D = \lim_{t\to\infty} 2\int_0^t dt'' \Big\langle v_x(t-t'')v_x(0)\Big\rangle$$

Define $\tau = t - t''$

$$\boxed{D = \int_0^\infty d\tau\ \langle v_x(\tau)v_x(0)\rangle}$$

Green –Kubo relation

Also exists for other transport coefficients, such as viscosity and conductivity

Contents

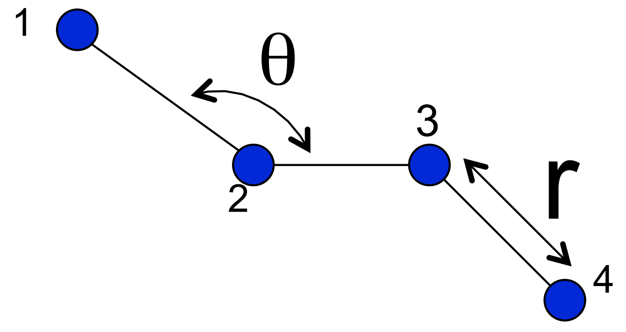
- Recap of the Verlet algorithms
- Lyopunov instability and the shadow theorem
- Lagrangian and Hamiltonian approach
- MD at constant temperature: Thermostats
- Constraints
- Ensemble averaging and computing static and dynamic properties
- Example of biomolecular simulation

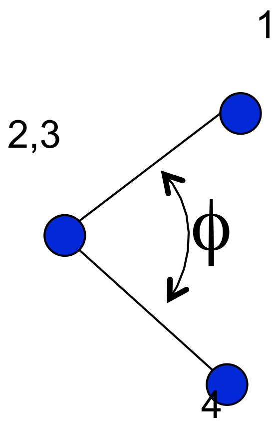
All-atom force fields for biomolecules

- Potential energy for protein

$$V(\mathbf{r}) = \sum_{bonds} k_r (r - r_{eq})^2 + \sum_{angles} k_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{1}{2} v_n (1 + \cos(n\phi - \phi_0))$$

$$+ \sum_{i<j} \left(\frac{a_{ij}}{r_{ij}^{12}} - \frac{b_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon r_{ij}} \right)$$



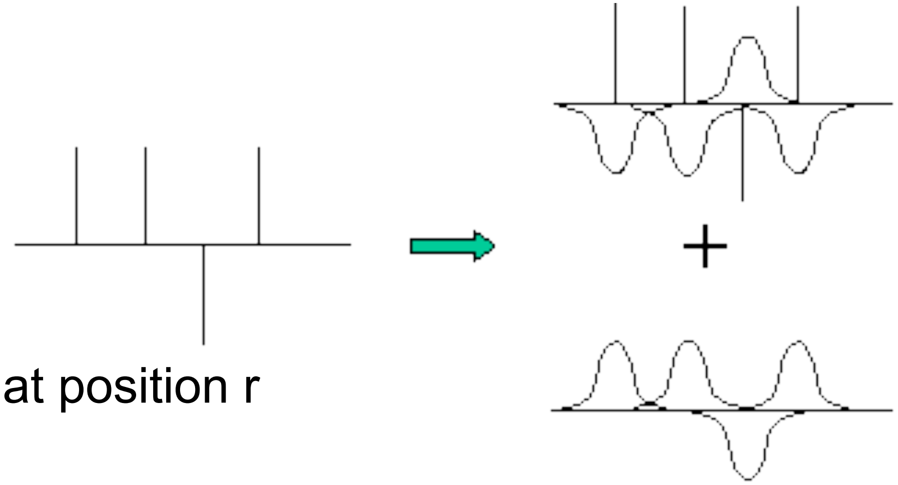


vdW interactions only between non-bonded |i-j|>4

Ewald sums

- Coloumb interaction

$$U_{Coul} = \frac{1}{2}\sum_{i=1}^{N} q_i \phi(r_i)$$



- With φ(r) the electrostatic potential at position r

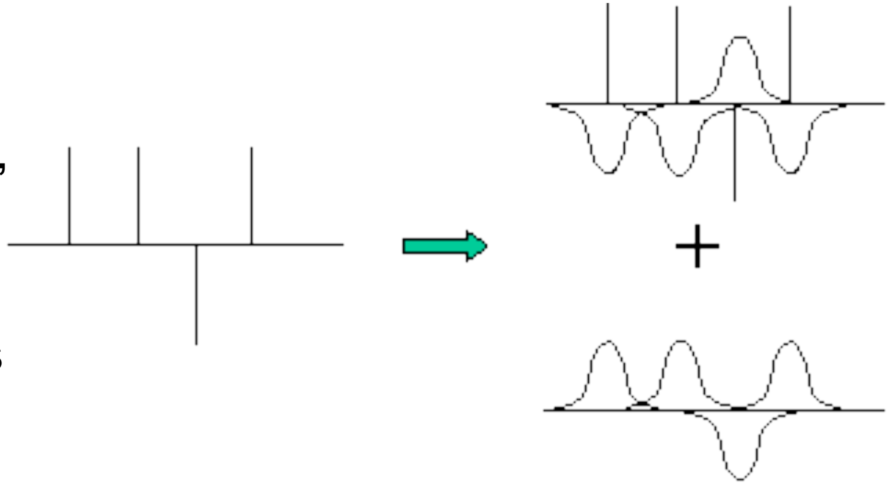
$$\phi(r_i) = \sum_{j,\mathbf{n}} \frac{q_j}{\left| r_{ij} + \mathbf{n}L \right|}$$

- The sum runs over all periodic images n

- This equation does converge poorly.

- The trick of the Ewald sum is to add a screened potential of the opposite sign, such each charge q is canceled

- A direct sum of the screened potentials converges much quicker.



- Thus 3 contributions
  - Direct sum of point charges q with Gaussian screening with charge (converges quickly)
  - Compensating screening with charge q (can be represented by a Fourier series)
  - Self energy correction

$$U_{coul} = \frac{1}{2V}\sum_{k\neq 0}\frac{4\pi}{k^2}\left|\rho(k)\right|^2 \exp\left(-k^2/4\alpha\right) - \left(\frac{\alpha}{\pi}\right)^{1/2}\sum_{1=0}^{N} q_i^2 + \frac{1}{2}\sum_{i\neq j}^{N}\frac{q_i q_j erfc\left(\sqrt{\alpha}\, r_{ij}\right)}{r_{ij}}$$

Currently available empirical force fields

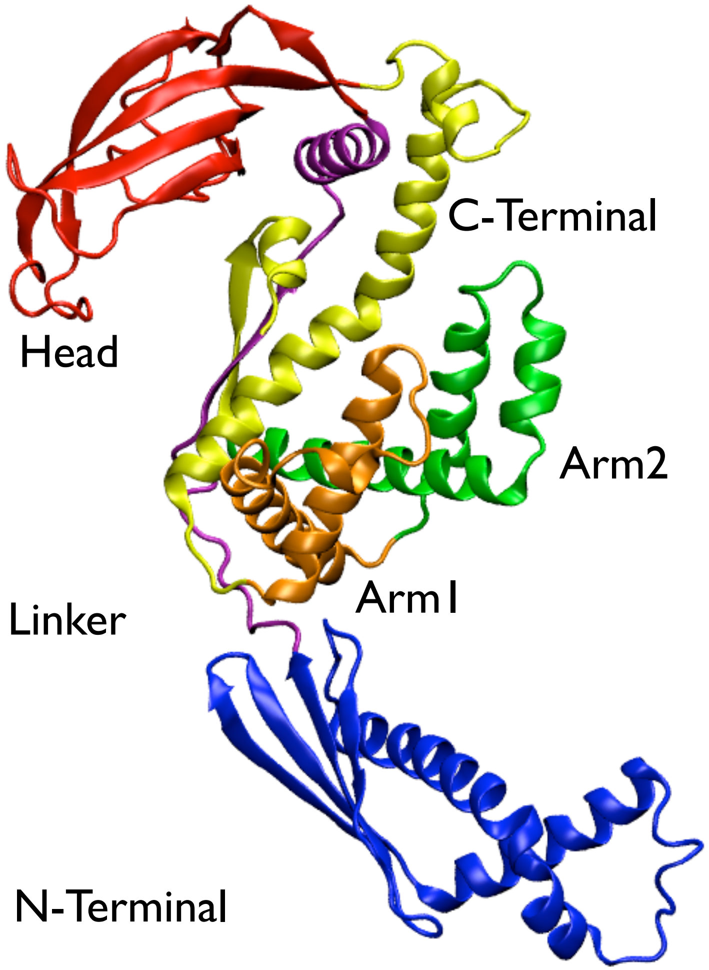
CHARMm	(MacKerrel et 96)
AMBER	(Cornell et al. 95)
GROMOS	(Berendsen et al 87)
OPLS-AA	(Jorgensen et al 95)
ENCAD	(Levitt et al 83)

- Subtle differences in improper torsions, scale factors 1-4 bonds, united atom rep.
- Partial charges based on empirical fits to small molecular systems
- Amber & Charmm also include ab-initio calculations
- Not clear which FF is best : top 4 mostly used

- Water models also included in description
  - TIP3P, TIP4P
  - SPC/E
- Current limit: 10^6 atoms, microseconds (with Anton ms)

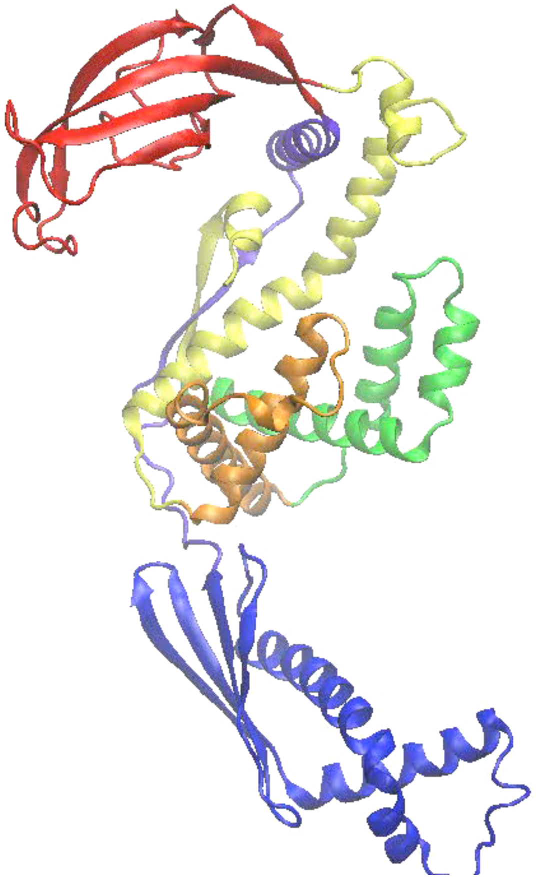
X-tal structure of Trigger Factor (TF)

- Chaperone protein characterized in E. coli
- Dual role:
  - Ribosome exit tunnel
  - Downstream in cytosol
- Flexible dragon-shaped structure
- 432 amino acid residues
- 3 domains:
  - N-terminal: Tail
  - C-terminal: Body + Arms
  - PPIase domain: Head



A. Hoffmann, B. Bukau, and G. Kramer, Biochimica et Biophysica Acta 1803 (2010), 650

Solution structure very different from x-tal

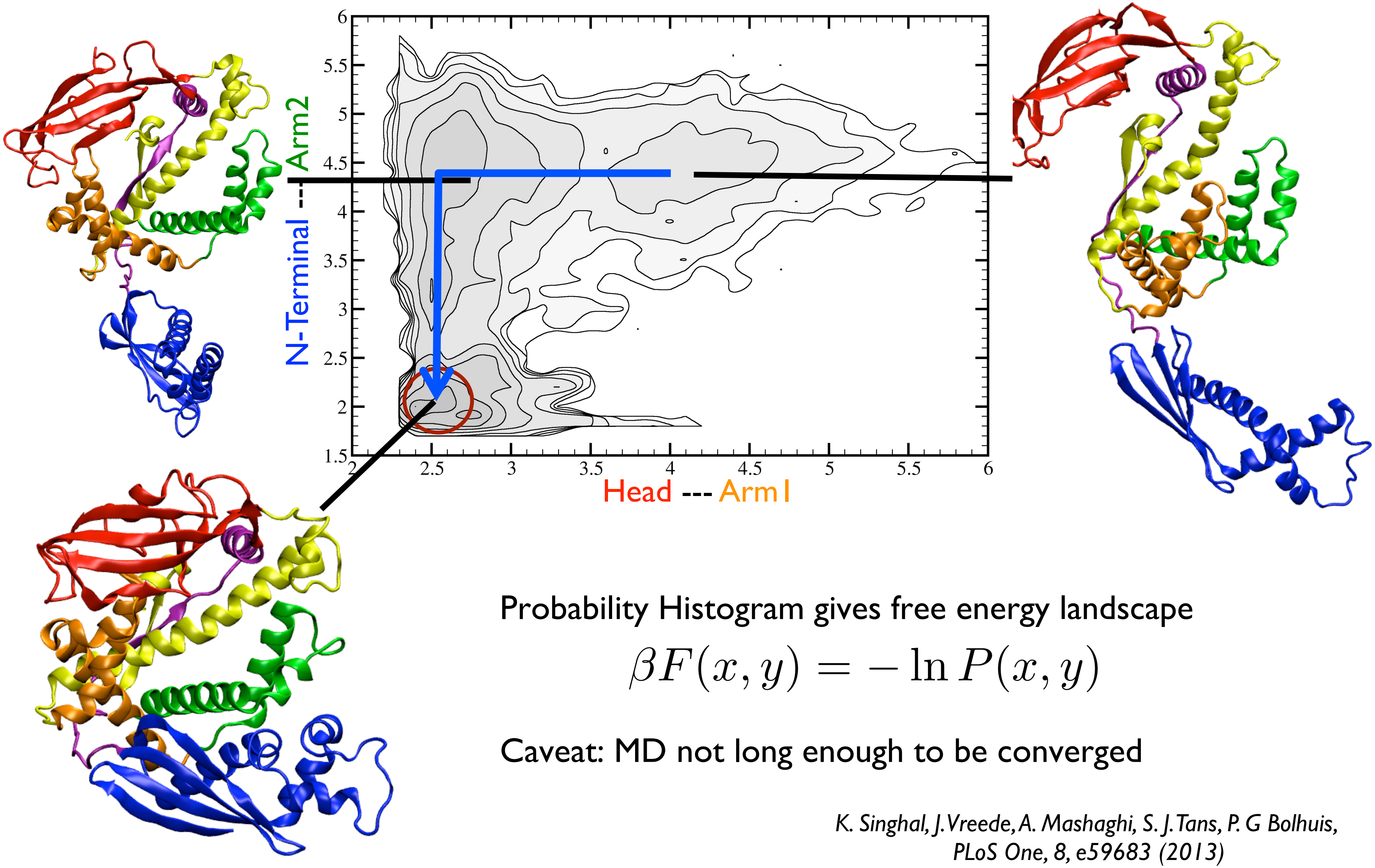


MD 250 ns trajectories in 50 mM salt solution

Amber FF

System with explicit water: 200000 atoms

Collapse Mechanism



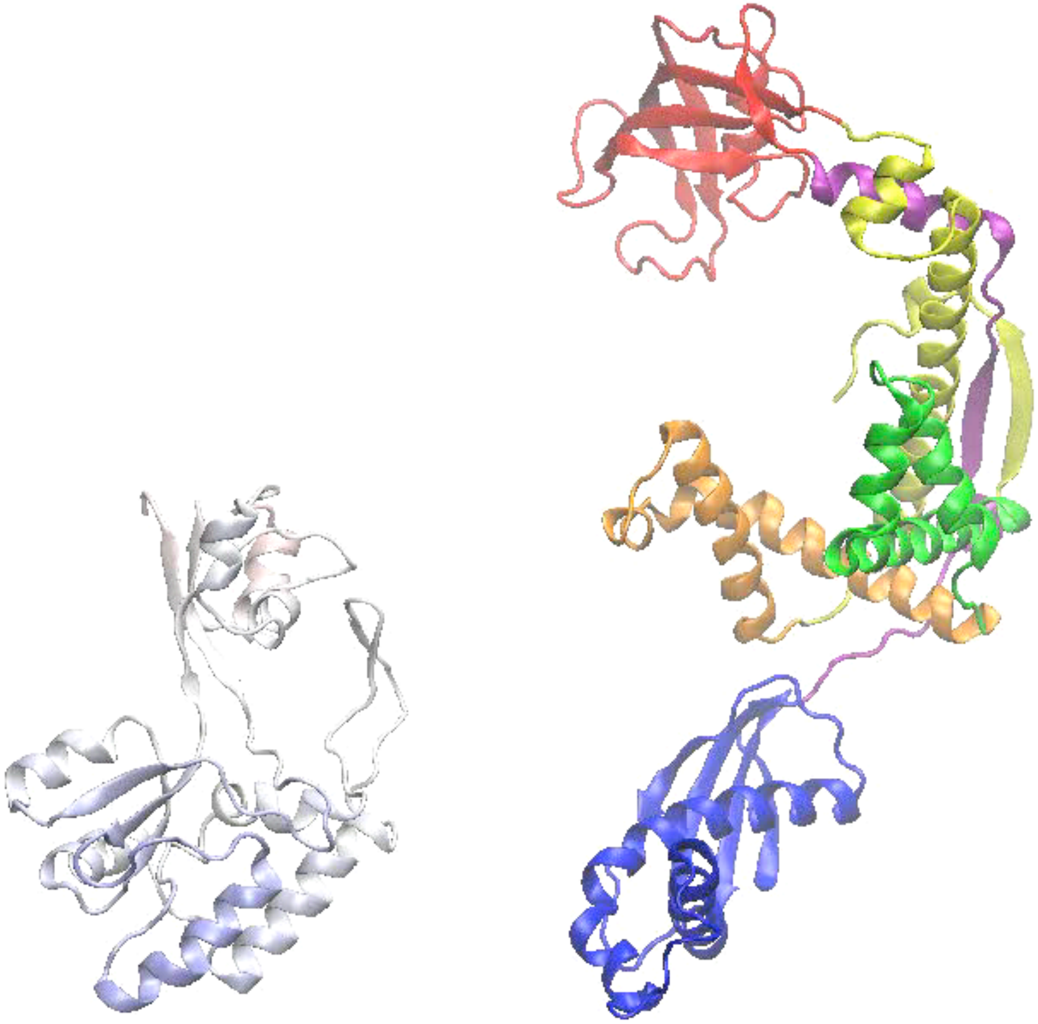
Probability Histogram gives free energy landscape

$$\beta F(x,y) = -\ln P(x,y)$$

Caveat: MD not long enough to be converged

K. Singhal, J. Vreede, A. Mashaghi, S. J. Tans, P. G Bolhuis, PLoS One, 8, e59683 (2013)

Trigger factor interacting with substrate

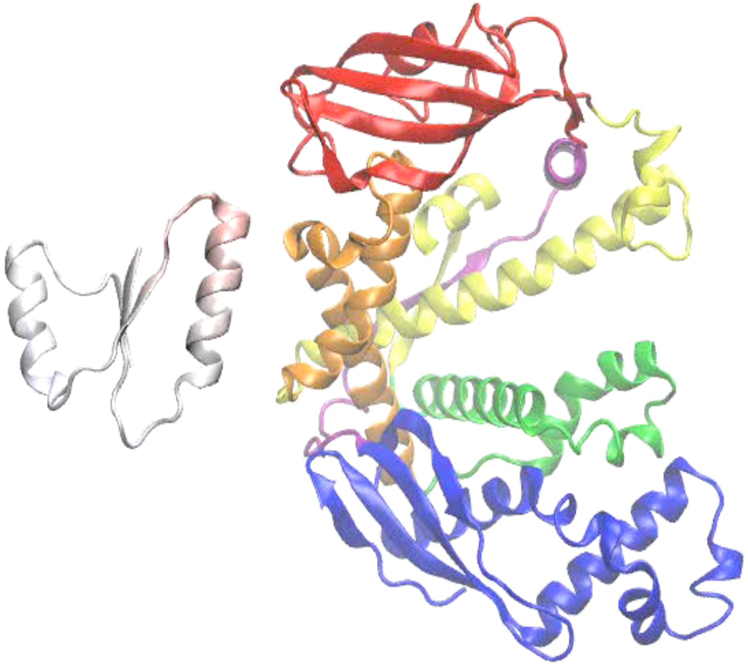


Conclusions Trigger factor

- Trigger factor very flexible in solution
  - crystal structure collapses
  - first hydrophobic collapse of Head-Arm1
  - followed by hydrophilic interaction of Nterm-Arm2

- Structural motions in TF stabilized in presence of substrate protein
- N-terminal crucial: mostly hydrophilic interactions with MBP subfold
- C-terminal crucial: also hydrophobic interactions with unfolded chain

- Clearly MD is not long enough
- Dynamics is dominated by rare events caused by high barriers



Next Thursday: rare events!

R. McKEE..
©2012 the AUGUSTA CHRONICLE 12/21
CAGLECARTOONS.COM