MSc in Theoretical Chemistry

Applications of Statistical Mechanics: Molecular Simulation Methods in Chemistry

Thomas Ouldridge

t.ouldridge1@physics.ox.ac.uk

`http://www-thphys.physics.ox.ac.uk/people/ThomasOuldridge/Site/Teaching.html`

Course Outline

Lecture 1 – Introduction to simulation.

Lecture 2 – Constant energy Molecular Dynamics.

Lecture 3 – Dynamical simulations in other ensembles.

- Thermostats, Langevin Dynamics, Brownian Dynamics.

Lecture 4 – Introduction to Monte Carlo sampling.

Lecture 5 – More Monte Carlo sampling.

- Other ensembles, Biased move generation.

Lecture 6 – Sampling phase diagrams and accelerating equilibration.

- Thermodynamic integration, Gibbs Ensemble, Umbrella Sampling, Parallel Tempering.

Lecture 7 – Dynamical sampling of rare events.

- Forward Flux Sampling, Transition Path Sampling.

Lecture 8 – Issues with simulation.

- Improving efficiency, finite system sizes, common mistakes and how to avoid them.

Aims of this course

- To introduce you to the principles behind computer simulation (not all the algorithmic details).
- To put you in a position so that you know what you are doing when running someone else's code/a simulation package.
- To put you in a position to start writing your own simulation codes.
- To highlight common issues and pitfalls with simulations.

This course won't cover:

- The complete statistical mechanical basis.
- How to program.
- Where $\mathcal{V}(\mathbf{r}^N)$ comes from.
- Every last detail of algorithms.

Recommended texts

Frenkel and Smit: Understanding Molecular Simulation: From Algorithms to Applications (Academic Press, 2002).

Allen and Tildesley: Computer Simulation of Liquids (Oxford, 1987).

Tuckerman: Statistical Mechanics: Theory and Molecular Simulation (Oxford, 2010).

Huang Statistical Mechanics (2nd Ed.) (John Wiley and sons, 1987).

Van Kampen Stochastic Processes in Physics and Chemistry (3rd Ed.) (Elsevier, 2007).

Press *et al.*: Numerical Recipes in ?: The Art of Scientific Computing, (Cambridge).

Krauth: Algorithms and Computations (Oxford, 2006).

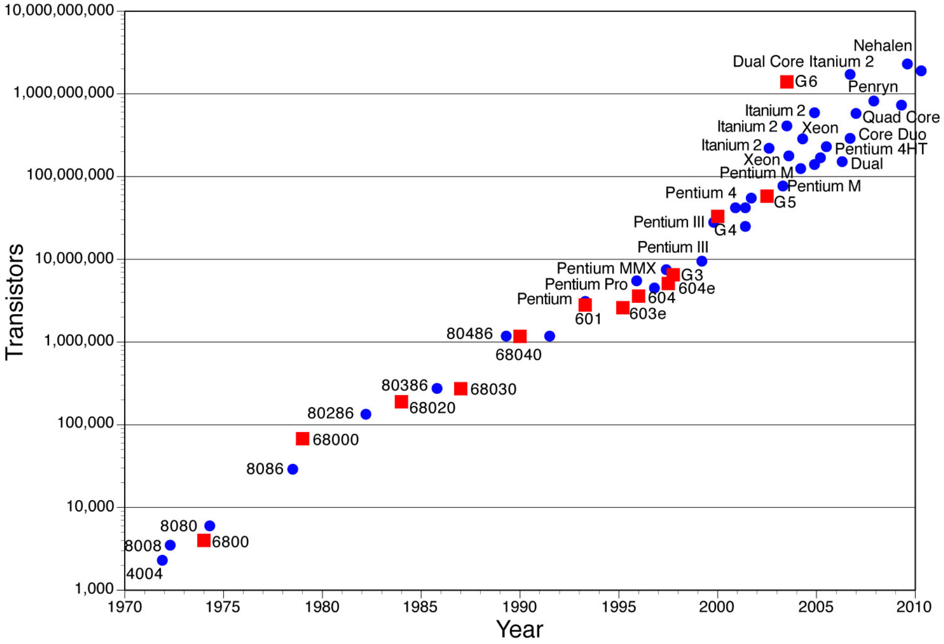
Newman and Barkema: Monte Carlo Methods in Statistical Physics (Oxford,1999).

Rapaport: The Art of Molecular Dynamics (Cambridge, 1995).

Introduction

Why computer simulations?

- Emergent properties of complex systems are a major part of modern science.
- The inevitability of approximations in analytical work
  e.g. the motion of three interacting bodies.
- Moore's law and the increasing power of computers



- The increasing sophistication of computer simulation techniques.
- Wide-spread availability of user-friendly simulation packages.

What can molecular simulations achieve given a classical potential $\mathcal{V}(\mathbf{r}^N)$ of a system of particles with $\mathbf{r}^N = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$?

- Dynamics: $\mathbf{r}^N(t), \mathbf{p}^N(t)$
  i.e. movies and trajectories.
- Thermodynamics and equilibrium properties: $\langle Q(\mathbf{r}^N(t), \mathbf{p}^N(t))\rangle$
  i.e. ensembles of configurations and averages.

Roles of computer simulations

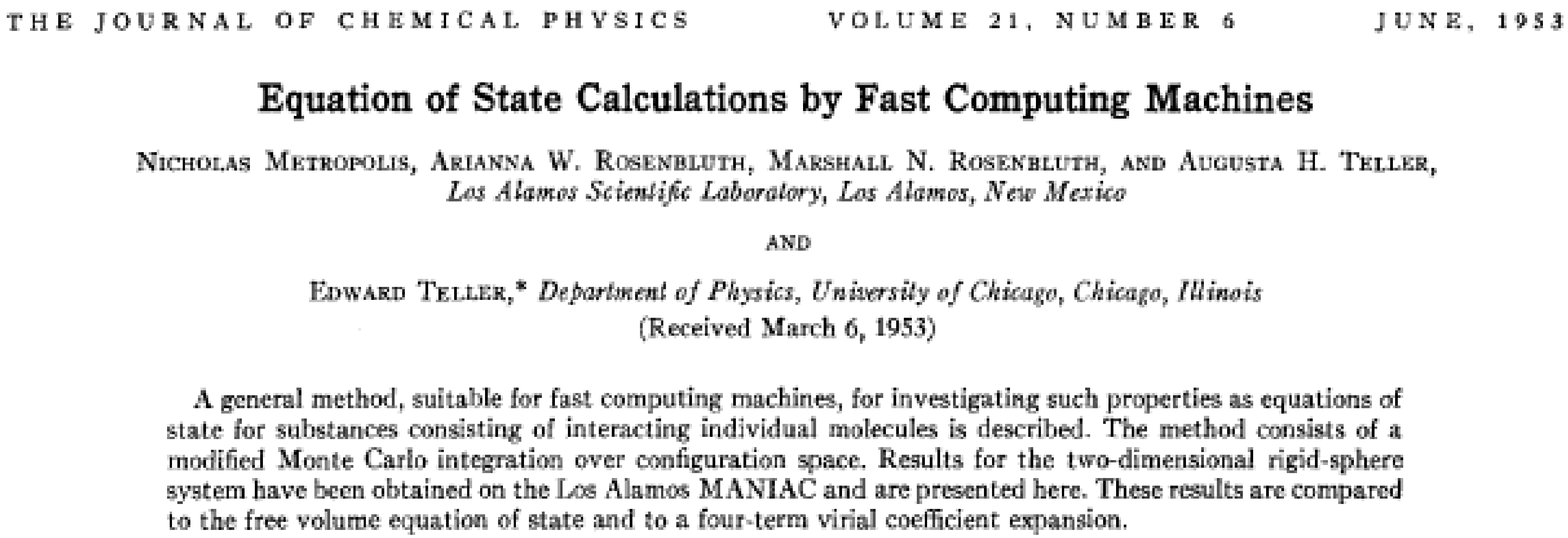
- Model systems (qualitative or semi-quantitative description of reality)
  - ▷ As a test for theory.
  - ▷ Exploring complex systems.
- Realistic models:
  - ▷ Prediction of experimental properties.
  - ▷ Direct visualization of molecular configurations.

Limits of computer simulation

- Need the potential energy function $\mathcal{V}(\mathbf{r}^N)$ (GIGO principle)
- Size of system limited (max $O(10^8)$)
- Time scales relatively short (max $O(\mu s)$)

Computer simulation: a new science

1953: Metroplis *et al.*: First Monte Carlo simulation of a molecular system



THE JOURNAL OF CHEMICAL PHYSICS VOLUME 21, NUMBER 6 JUNE, 1953

Equation of State Calculations by Fast Computing Machines

NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AND AUGUSTA H. TELLER,
Los Alamos Scientific Laboratory, Los Alamos, New Mexico

AND

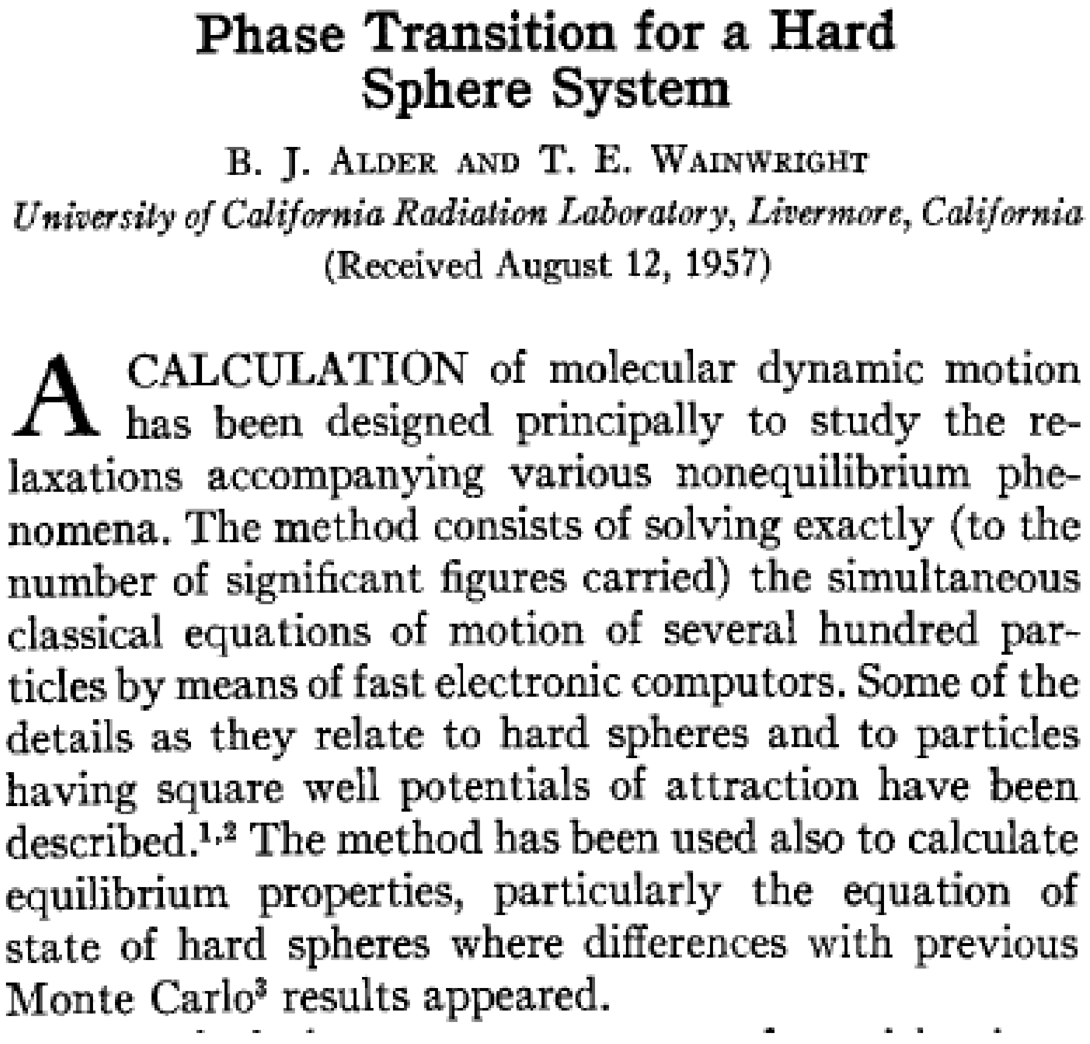
EDWARD TELLER,* *Department of Physics, University of Chicago, Chicago, Illinois*

(Received March 6, 1953)

A general method, suitable for fast computing machines, for investigating such properties as equations of state for substances consisting of interacting individual molecules is described. The method consists of a modified Monte Carlo integration over configuration space. Results for the two-dimensional rigid-sphere system have been obtained on the Los Alamos MANIAC and are presented here. These results are compared to the free volume equation of state and to a four-term virial coefficient expansion.

1956 Alder and Wainwright: First molecular dynamics (MD) simulations.

1957 Alder and Wainwright: Hard-sphere debate



Phase Transition for a Hard Sphere System

B. J. Alder and T. E. Wainwright
University of California Radiation Laboratory, Livermore, California

(Received August 12, 1957)

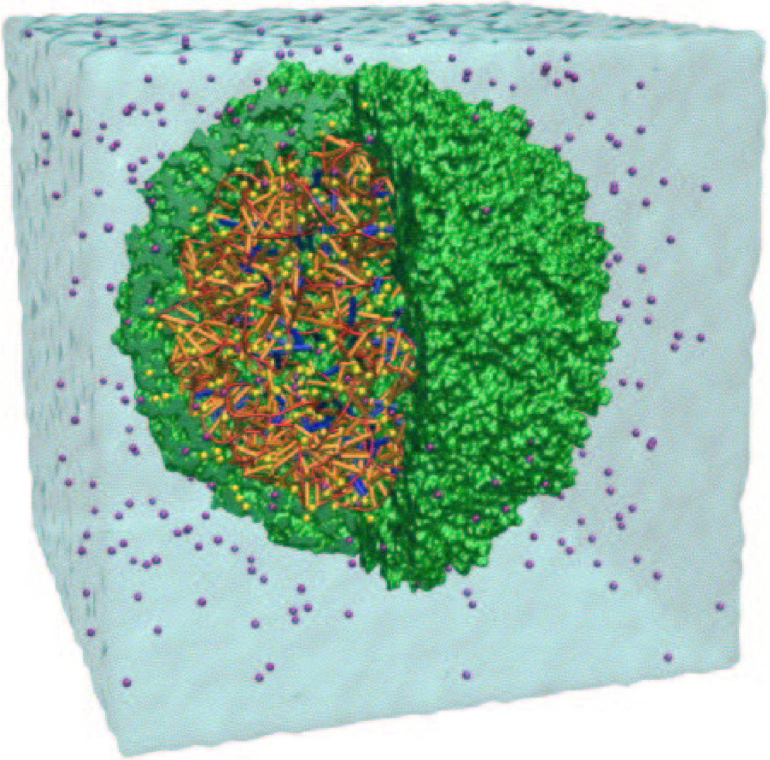
A CALCULATION of molecular dynamic motion has been designed principally to study the relaxations accompanying various nonequilibrium phenomena. The method consists of solving exactly (to the number of significant figures carried) the simultaneous classical equations of motion of several hundred particles by means of fast electronic computors. Some of the details as they relate to hard spheres and to particles having square well potentials of attraction have been described.[1,2] The method has been used also to calculate equilibrium properties, particularly the equation of state of hard spheres where differences with previous Monte Carlo[3] results appeared.

1964 Rahman: First MD with realistic interatomic potential (Ar)

1974 Rahman and Stillinger: First MD with realistic intermolecular potential (H_2O)

1977 McCammon *et al.*: First MD of a protein (BPTI) (500 atoms, 9.2 ps)

2006 Schulten group: MD of complete virus (STMV) (10^6 atoms, 50 ns)



Theoretical preliminaries: Newton's equations, Lagrangian and Hamiltonian mechanics [1]

Given a potential $\mathcal{V}(\mathbf{r}^N)$:

$$\frac{\partial^2 \mathbf{r}_i}{\partial t^2} = -\frac{1}{m_i}\frac{\partial \mathcal{V}(\mathbf{r}_i)}{\partial \mathbf{r}^N}. \tag{1}$$

In an alternative formulation due to Lagrange, we define the Lagrangian

$$\mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}, t) = \mathcal{K}(\mathbf{q}, \dot{\mathbf{q}}) - \mathcal{V}(\mathbf{q}, t). \tag{2}$$

In our simple case, $\mathbf{q} = \mathbf{r}^N$, $\mathcal{K} = \frac{1}{2}\sum_i m_i \mathbf{r}_i^2$ and $\mathcal{V}(\mathbf{q}, t) = \mathcal{V}(\mathbf{r}^N)$. The trajectory followed by a system is that which minimizes the action

$$S = \int_{t_i}^{t_f} \mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}, t) dt. \tag{3}$$

It can be shown that this trajectory is given by

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}, t)}{\partial \dot{\mathbf{q}}}\right) = \frac{\partial \mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}, t)}{\partial \mathbf{q}} \tag{4}$$

We can also define a ‘generalized’ momentum

$$\mathbf{p} = \frac{\partial \mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}, t)}{\partial \dot{\mathbf{q}}} \tag{5}$$

Hence, due to Eq. 4

$$\dot{\mathbf{p}} = \frac{\partial \mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}, t)}{\partial \mathbf{q}} \tag{6}$$

As a check, for motion in a 1-dimensional Cartesian coordinate

$$\mathcal{L}(x, \dot{x}) = \frac{1}{2} m \dot{x}^2 - \mathcal{V}(x), \tag{7}$$

$$p = \frac{\partial \mathcal{L}(x, \dot{x})}{\partial \dot{x}} = m\dot{x}, \tag{8}$$

and the equation of motion is

$$m\ddot{x} = -\frac{\partial \mathcal{V}(x)}{\partial x}. \tag{9}$$

Yet another alternative is the Hamiltonian formulation. The Hamiltonian is related to the Lagrangian by

$$\mathcal{H}(\mathbf{q},\mathbf{p},t) = \mathbf{p}\dot{\mathbf{q}} - \mathcal{L}(\mathbf{q},\dot{\mathbf{q}},t). \tag{10}$$

Here it is understood that we use $\mathbf{p}$ to eliminate $\dot{\mathbf{q}}$ from the RHS.

By comparing the Hamiltonian to the Lagrangian, we can show

$$\frac{\partial \mathcal{H}}{\partial \mathbf{p}} = \dot{\mathbf{q}} \tag{11}$$

$$\frac{\partial \mathcal{H}}{\partial \mathbf{q}} = -\dot{\mathbf{p}} \tag{12}$$

Note that this formulation gives the equations of motion in terms of two sets of first-order differential equations, rather than a single set of second-order differential equations.

When the Lagrangian does not explicitly depend on time, the Hamiltonian is conserved

$$\frac{d\mathcal{H}(\mathbf{q},\mathbf{p})}{dt} = \frac{\partial \mathcal{H}}{\partial \mathbf{p}}\dot{\mathbf{p}} + \frac{\partial \mathcal{H}}{\partial \mathbf{q}}\dot{\mathbf{q}} = -\frac{\partial \mathcal{H}}{\partial \mathbf{p}}\frac{\partial \mathcal{H}}{\partial \mathbf{q}} + \frac{\partial \mathcal{H}}{\partial \mathbf{q}}\frac{\partial \mathcal{H}}{\partial \mathbf{p}} = 0, \tag{13}$$

and expresses the conservation of energy.

As a sanity check, return to our 1-dimensional Cartesian system

$$\mathcal{H}(x, p_x) = p_x \dot{x} - \mathcal{L}(x, \dot{x}) = \frac{p_x^2}{2m} + \mathcal{V}(x), \tag{14}$$

and the equations of motion are

$$\dot{x} = \frac{\partial \mathcal{H}}{\partial p_x} = \frac{p_x}{m} \tag{15}$$

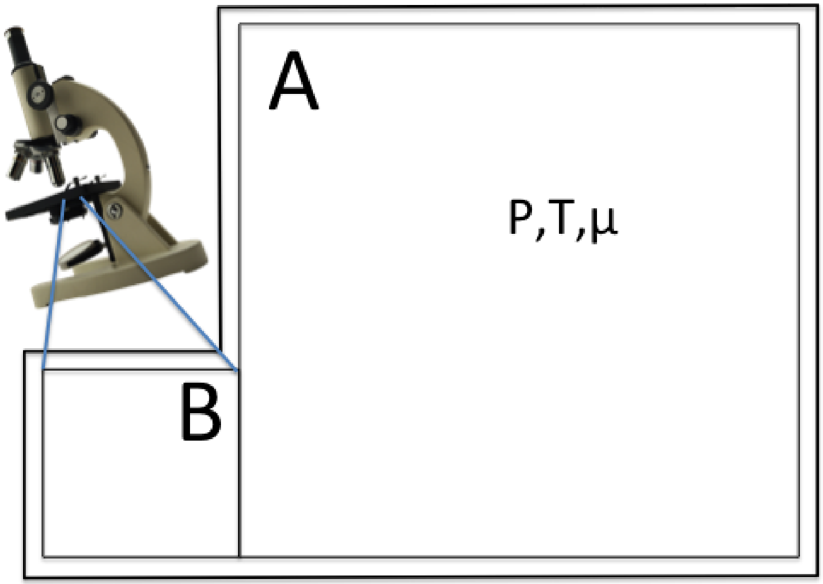
$$\dot{p}_x = -\frac{\partial \mathcal{H}}{\partial x} = -\frac{\partial \mathcal{V}(x)}{\partial x}. \tag{16}$$

Theoretical preliminaries: Statistical mechanics and Ensembles [2]

- We consider systems that can exist in a (very large) number of 'states', defined by $\mathbf{r}^N(t), \mathbf{p}^N(t)$. These states are equally distributed in *phase space*.

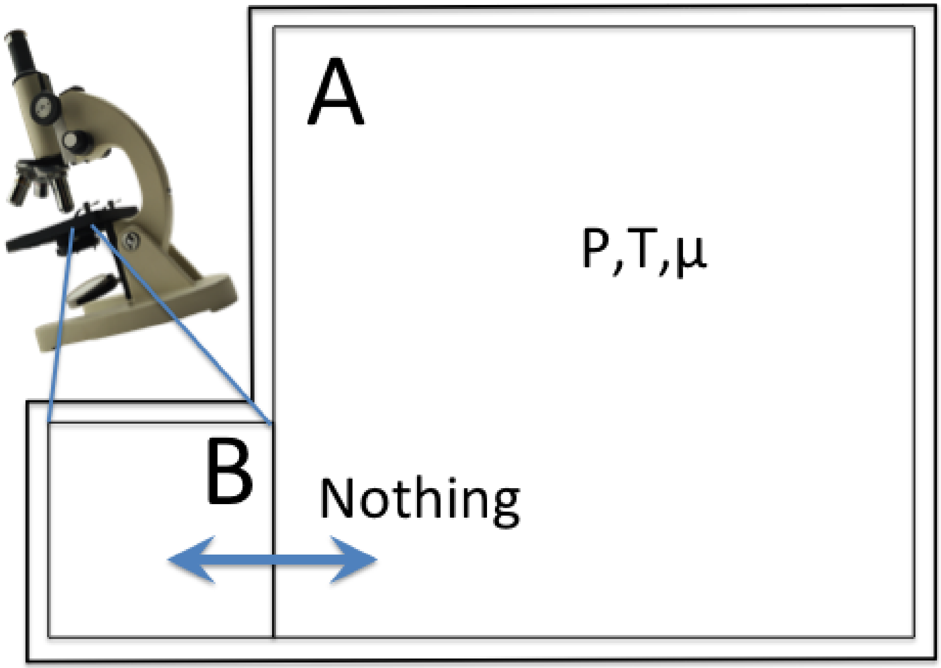
- The assumption of equal a-priori probability states that, for an isolated system with total energy E, all states of energy E are equally probable.

- Now let us construct an isolated system by combining a large system A and a small one B. If A is thermodynamically large, we can meaningfully talk about its temperature T, pressure P and chemical potential μ, regardless of B.



If A and B are isolated from each other, B is in the microcanonical ensemble.

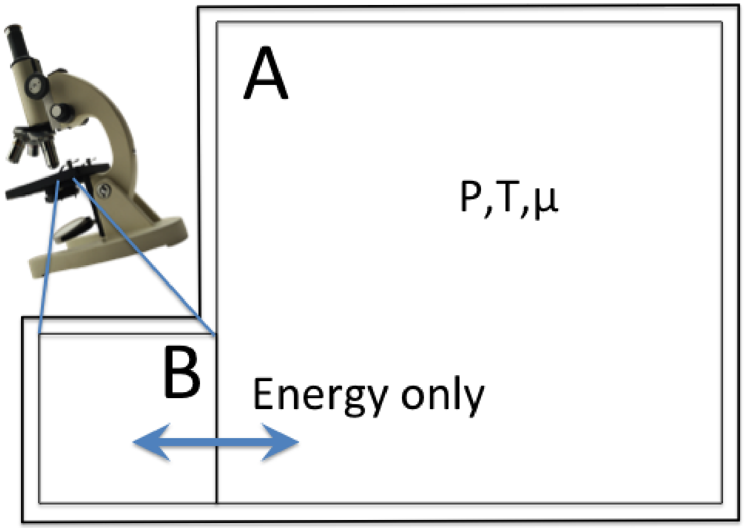
- B has well-defined volume V_B, energy E_B and particle number N_B.
- All states of B which conform to these values are equally likely.



If A and B can exchange energy with each other, B is in the canonical ensemble.

- B has well-defined volume V_B, temperature T and particle number N_B.
- A state of B that has an energy E occurs with probability

$$P(E) \propto \exp\left(-E/k_B T\right). \tag{17}$$

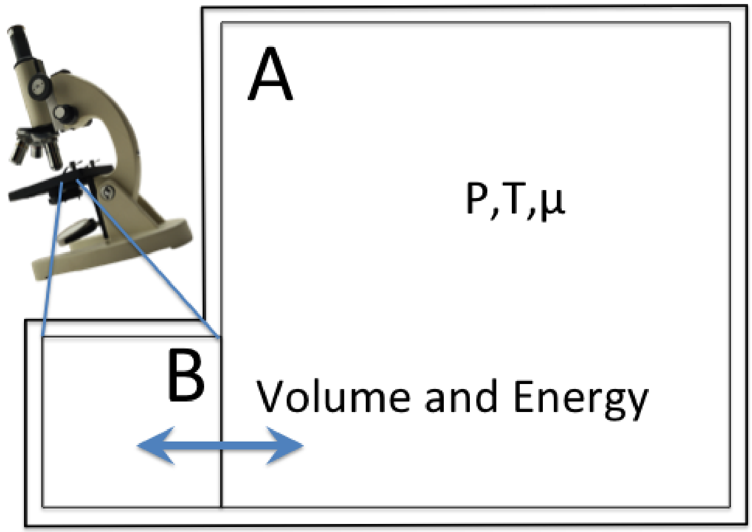


$$\mathcal{Q}(\mathcal{N},\mathcal{V},\mathcal{T}) = \sum_i \exp\left(-E_i/k_B T\right). \tag{18}$$

If A and B can exchange energy and volume each other, B is in the isobaric-isothermal ensemble

- B has well-defined pressure P, temperature T and particle number N_B.
- A state of B that has an energy E and a volume V occurs with probability

$$P(E,V) \propto \exp\left(-E/k_BT\right)\exp\left(-PV/k_BT\right). \tag{19}$$

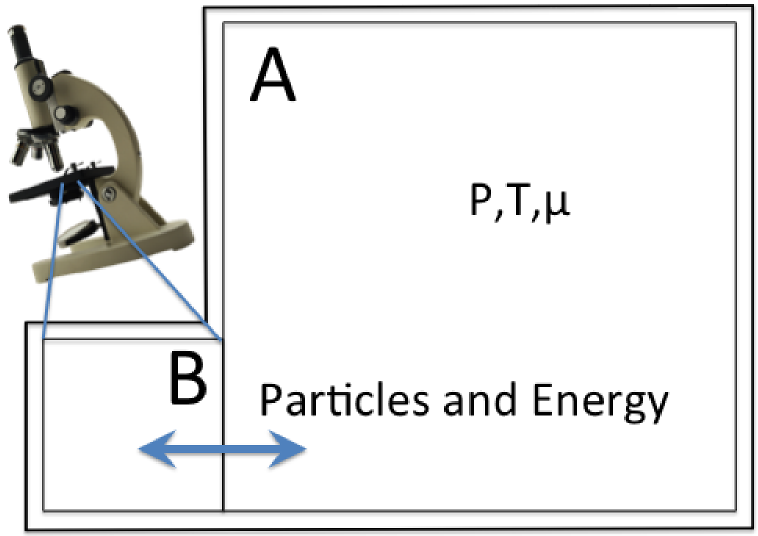


$$\Theta(N,P,T) = \frac{1}{V_0}\int dV\, Q(N,V,T)\exp(-PV/k_BT) \tag{20}$$

If A and B can exchange energy and particles with each other, B is in the grand canonical ensemble

- B has well-defined volume V_B, temperature T and chemical potential μ.
- A state of B that has an energy E and contains N particles occurs with probability

$$P(E, N) \propto \exp\left(-E/k_B T\right) \exp\left(\mu N/k_B T\right). \tag{21}$$



$$\Xi(\mu, V, T) = \sum_N Q(N, V, T) \exp(\mu N/k_B T) \tag{22}$$

Quite often, people say that the ensembles are “equivalent”. What does this mean [3]?

- Fluctuations are relatively small in a large system (away from phase transitions).
- Quantities that are not fixed by the ensemble have well-defined average values that are representative of the thermodynamic state.
- If we studied a canonical system at temperature T and measured energy E, then a microcanonical system with energy E would have a temperature T.

▷ If system B has many interacting degrees of freedom (although still small compared to our hypothetical reservoir), then we can use any ensemble that is convenient to obtain physically meaningful results.

▷ In this case, we can study an NVE (microcanonical) system and *measure* the temperature and pressure.

▷ For small, coarse-grained or discrete systems, ensembles must be considered more carefully.

▷ The use of the microcanoical ensemble can be particularly problematic.

Molecular Simulation Methods in Chemistry

Lecture 2:
Constant energy Molecular Dynamics

Thomas Ouldridge

t.ouldridge1@physics.ox.ac.uk

`http://www-thphys.physics.ox.ac.uk/people/ThomasOuldridge/Site/Teaching.html`

Molecular Dynamics

Molecular dynamics simulations are a method for computing the equilibrium and transport properties of a classical many-body system.

Classical: the nuclear motion of the constituent particles obeys the laws of classical mechanics, i.e. Newton's laws.

Classical mechanics is a good approximation for many materials. Fundamentally quantum effects most relevant for light atoms and low temperatures.

We need a potential $\mathcal{V}(\mathbf{r}^N)$ which describes the system as well as possible.

We solve Newton's (equivalently Lagrange's or Hamilton's) equations by integrating them forward in time to generate a trajectory.

Basic scheme

1. Initialize the system: choose $\mathbf{r}^N(t = 0)$ and $\mathbf{p}^N(t = 0)$
2. Compute forces.
3. Integrate equations of motion: from $t \to t + \delta t$.
4. Repeat 3 and 4 until trajectory of desired length.

Which ensemble?

For an isolated mechanical system the total energy is conserved during the motion. Proof:

$$\begin{aligned}\frac{dE}{dt} &= \frac{d}{dt}\left[\sum_i^N \frac{1}{2} m_i \dot{\mathbf{r}}_i^2 + \mathcal{V}(\mathbf{r}^N)\right] = \sum_i^N m_i \dot{\mathbf{r}}_i \cdot \ddot{\mathbf{r}}_i + \sum_i^N \frac{\partial \mathcal{V}}{\partial \mathbf{r}_i} \cdot \dot{\mathbf{r}}_i && (23) \\ &= \sum_i^N \dot{\mathbf{r}}_i \cdot \mathbf{f}_i - \sum_i^N \dot{\mathbf{f}}_i \cdot \mathbf{r}_i = 0 && (24)\end{aligned}$$

Therefore, it seems that standard molecular dynamics simulates the microcanonical ensemble (more on this later).

Newtonian dynamics also obeys time reversal symmetry. *i.e.* if we reverse the velocities of all the particles, keeping the positions the same, the system will retrace its trajectory back into the past. More formally,

$$\begin{aligned}\mathbf{r}^N\left(t, \mathbf{r}^N(0), -\mathbf{p}^N(0)\right) &= \mathbf{r}^N\left(-t, \mathbf{r}^N(0), \mathbf{p}^N(0)\right) && (25) \\ \mathbf{p}^N\left(t, \mathbf{r}^N(0), -\mathbf{p}^N(0)\right) &= -\mathbf{p}^N\left(-t, \mathbf{r}^N(0), \mathbf{p}^N(0)\right) && (26)\end{aligned}$$

Integration Schemes

All based on Taylor expansions of position in time.

A simple example. Expand position to third order in time.

$$\mathbf{r}_i(t+\delta t) = \mathbf{r}_i(t) + \dot{\mathbf{r}}_i(t)\delta t + \frac{\ddot{\mathbf{r}}_i(t)}{2}\delta t^2 + \frac{\dddot{\mathbf{r}}_i(t)}{6}\delta t^3 + \mathcal{O}(\delta t^4) \tag{27}$$

As $\dot{\mathbf{r}}_i = \mathbf{v}_i$ and using Newton's 2nd law ($\mathbf{f}_i = m_i\ddot{\mathbf{r}}_i$), rewrite as

$$\mathbf{r}_i(t+\delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t)\delta t + \frac{\mathbf{f}_i(t)}{2m_i}\delta t^2 + \frac{\dddot{\mathbf{r}}_i(t)}{6}\delta t^3 + \mathcal{O}(\delta t^4) \tag{28}$$

Similarly,

$$\mathbf{r}_i(t-\delta t) = \mathbf{r}_i(t) - \mathbf{v}_i(t)\delta t + \frac{\mathbf{f}_i(t)}{2m_i}\delta t^2 - \frac{\dddot{\mathbf{r}}_i(t)}{6}\delta t^3 + \mathcal{O}(\delta t^4) \tag{29}$$

Adding these two equations gives

$$\mathbf{r}_i(t+\delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t-\delta t) + \frac{\mathbf{f}_i(t)}{m_i}\delta t^2 + \mathcal{O}(\delta t^4) \tag{30}$$

Note that due to the cancellation of odd time derivatives of $\mathbf{r}_i$:

- this expression is accurate to third order in time, even though only forces are used.
- velocities do not appear in the expression.

Get velocity, by subtracting Eq. 29 from Eq. 28.

$$\mathbf{v}_i(t) = \frac{1}{2\delta t}\left[\mathbf{r}_i(t+\delta t) - \mathbf{r}_i(t-\delta t)\right] + \mathcal{O}(\delta t^2) \tag{31}$$

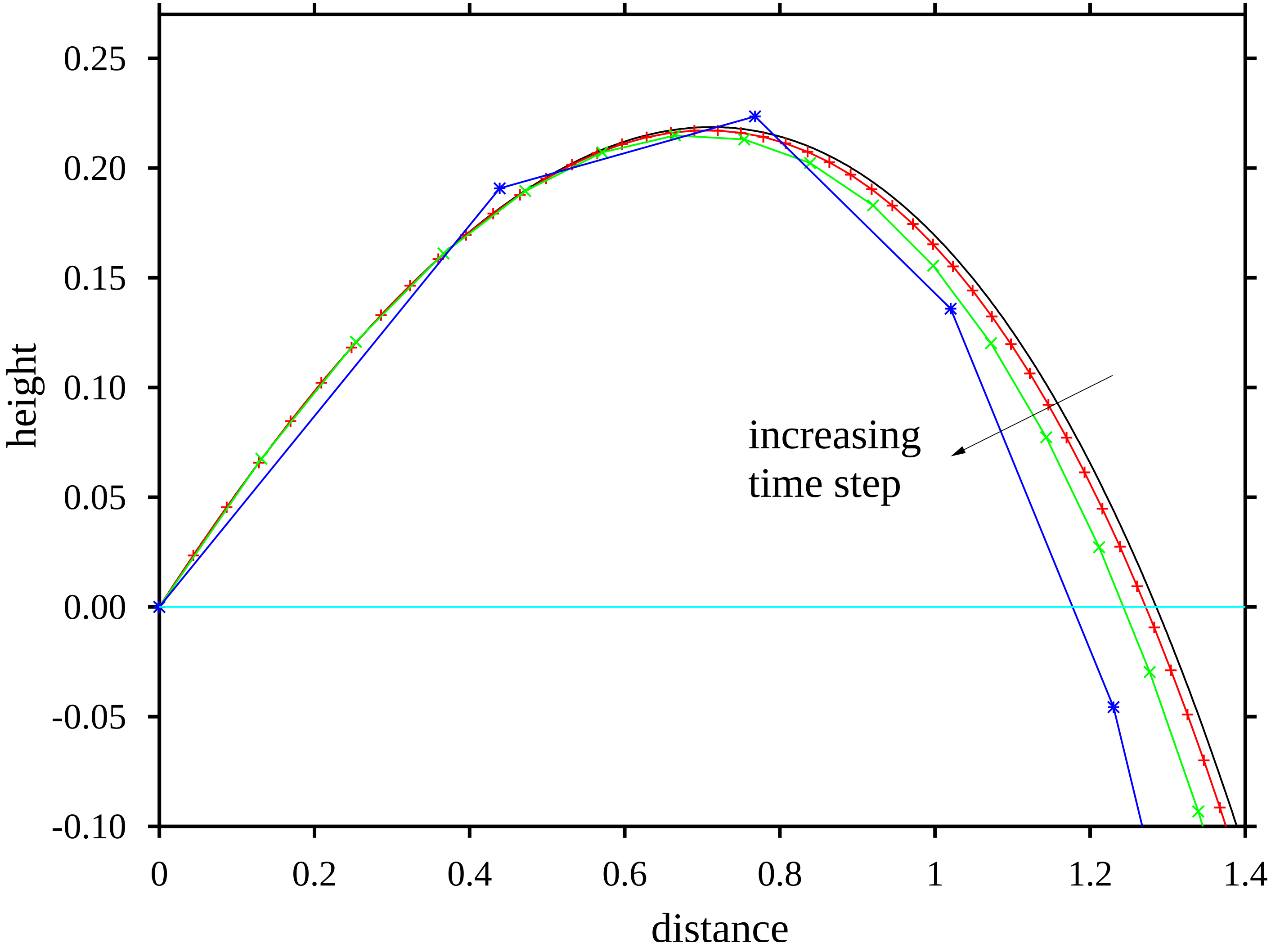
This is the Verlet scheme.

An alternative scheme that has identical trajectories to the Verlet algorithm, but a more convenient expression for the velocity is velocity Verlet.

$$\mathbf{r}_i(t+\delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t)\delta t + \frac{\mathbf{f}_i(t)}{2m_i}\delta t^2 \tag{32}$$

$$\mathbf{v}_i(t+\delta t) = \mathbf{v}_i(t) + \frac{1}{2m_i}\left[\mathbf{f}_i(t) + \mathbf{f}_i(t+\delta t)\right]\delta t \tag{33}$$

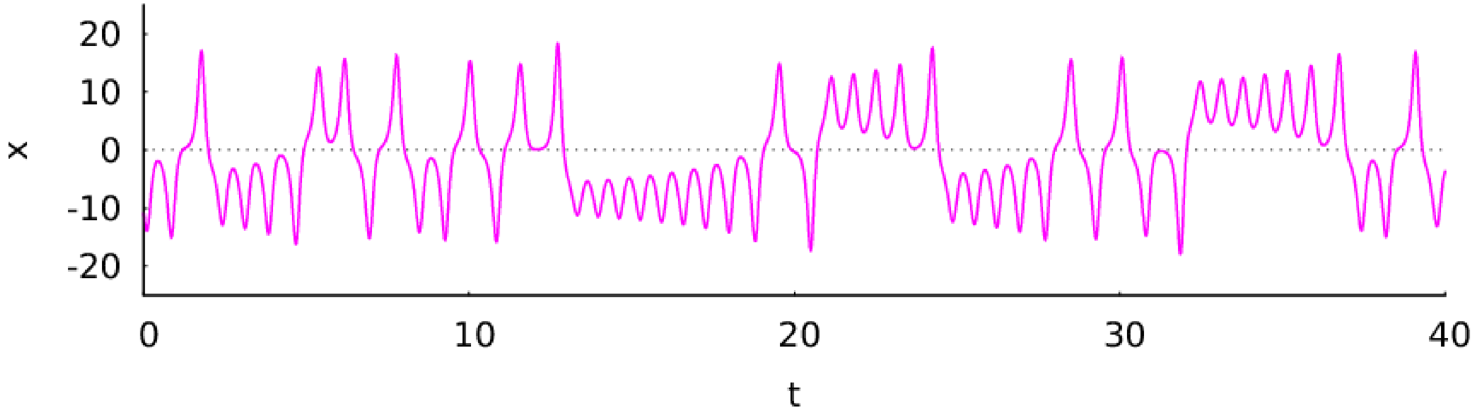
Example: Motion of a projectile with air resistance



Chaotic behavior

Systems of many particles with non-linear interactions tend to show chaotic behaviour.

- Trajectories continue without converging to steady states or predictable cycles.
- Without knowledge of the microscopic laws, motion appears to be a random process.
- Trajectories show sensitive dependence on initial conditions.



Chaotic behaviour helps make the Ergodic Hypothesis reasonable.

- Time averages of ergodic systems in the limit of trajectories of infinite length are equivalent to ensemble averages.
- Equivalently, an MD simulation will sample all states of a system with a certain energy.

i.e. $\bar{A} = \langle A \rangle_{NVE}$, where

$$\bar{A} = \lim_{\Delta t \to \infty} \frac{1}{\Delta t} \int_0^{\Delta t} A\left(\mathbf{r}^N(t), \mathbf{p}^N(t)\right) dt \tag{34}$$

$$\langle A \rangle_{NVE} = \frac{\int A\left(\mathbf{r}^N, \mathbf{p}^N\right) \delta\left[\mathcal{H}\left(\mathbf{r}^N, \mathbf{p}^N\right) - E\right] d\mathbf{r}^N d\mathbf{p}^N}{\int \delta\left[\mathcal{H}\left(\mathbf{r}^N, \mathbf{p}^N\right) - E\right] d\mathbf{r}^N d\mathbf{p}^N} \tag{35}$$

For this to work, we need anharmonic terms in the potential that couple different degrees of freedom.

- Motion in harmonic potentials can be decoupled into normal modes.
- Energy and entropy barriers can inhibit ergodicity on simulation time scales.

What do we want from an integrator?

Accuracy? *i.e.* an integrator that most closely follows the true trajectory.

Important for *satellite* dynamics, but not for *molecular* dynamics.

Chaotic effects mean that simulated trajectories soon diverge from “true” trajectories.

$\Rightarrow$ Trying to achieve accuracy in MD is pointless.

New aim: representative trajectory and good statistical properties.

- Trajectories should obey time reversal symmetry.
- Energy should be conserved over long times.
- The integrator should conserve phase space volume (it should be *symplectic*).

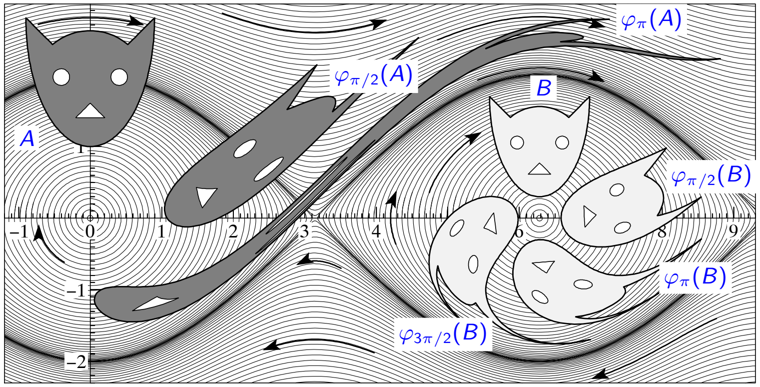
Symplectic Integrators

We have already discussed time-reversibility and energy conservation. What does conserving phase space volume mean?

- Plot trajectories on a graph where the axes are $\mathbf{r}^N$, $\mathbf{p}^N$.
- We consider a blob of states occupying some volume, and ask how this blob changes over time.
- For a Hamiltonian system, the blob gets distorted but retains its volume [1].

If the blob tended to shrink or grow in size, the system would tend to deviate strongly from the microcanonical ensemble, and the deviation would get worse with time.

▷ Conservation of phase space volume is consistent with equal a-priori probability.

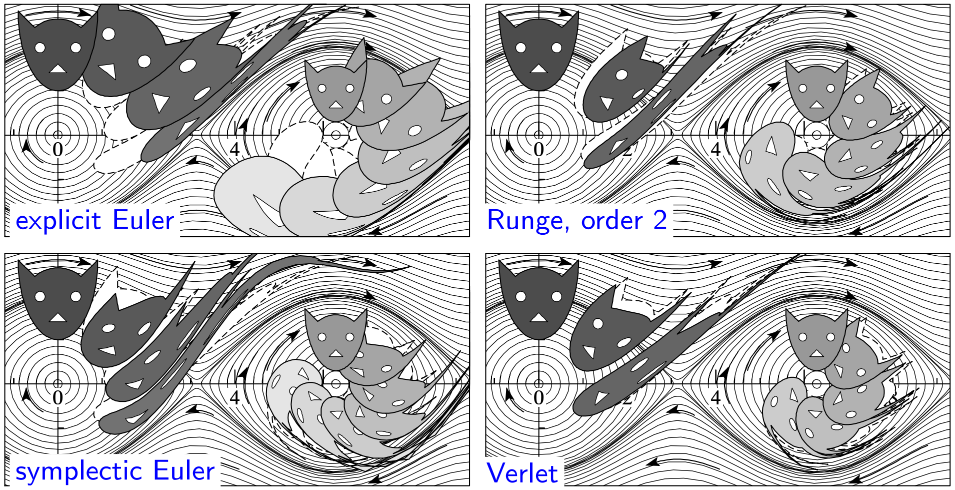


It can be shown that symplectic integrators conserve a perturbed Hamiltonian [4]

$$\mathcal{H}'(\mathbf{r}^N, \mathbf{p}^N, \Delta t) = \mathcal{H}(\mathbf{r}^N, \mathbf{p}^N) + \Delta\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N, \Delta t), \tag{36}$$

where $\Delta\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N, \Delta t) \to 0$ as $\Delta t \to 0$.

- For short Δt, deviations in the energy don't tend to grow systematically with simulation time.
- Verlet is an example of a symplectic integrator (there are others).
- It is more useful for constant energy MD than other integrators (such as RK4) that are typical "workhorses" of numerical integration and reproduce "true" trajectories more accurately.



Initialization

Need to provide initial $\mathbf{r}^N$ and $\mathbf{p}^N$.

$\mathbf{r}^N(0)$ easy if simulating an ordered state with known structure, ***e.g.*** crystal, native state of protein.

If simulating a disordered state such as a liquid, ***e.g.***

- Create random configuration and then minimise the potential energy (so that the potential energy is less than the total energy that you want).
- Or perform an initial simulation, where a crystal is rapidly heated to melt it, and then cooled to the desired total energy.

The velocities are assigned from a random distribution (ideally a Gaussian one, so that they satisfy the Maxwell-Boltzmann distribution).

The velocities are then scaled so that the system has the desired total energy E.

Measuring temperature and pressure

If the simulated system is large enough, the temperature and the pressure are well defined.

Temperature can be calculated through the *equipartition theorem* [5]

$$\left\langle \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} \right\rangle = \frac{3}{2} NkT. \tag{37}$$

Pressure is usually calculated using the *virial theorem*

$$P = \frac{NkT}{V} + \frac{1}{3V} \left\langle \sum_{i=1}^{N} \mathbf{r}_i . \mathbf{f}_i \right\rangle , \tag{38}$$

where $\mathbf{f}_i$ is the force on i due to the other particles. Note that this doesn't always work when periodic boundary conditions are involved.

Molecular Simulation Methods in Chemistry

Lecture 3:
Dynamical simulations in other ensembles

Thomas Ouldridge

t.ouldridge1@physics.ox.ac.uk

http://www-thphys.physics.ox.ac.uk/people/ThomasOuldridge/Site/Teaching.html

Constant temperature Molecular Dynamics

Why do we want to do this?

- Practicality of performing the simulation.
- Practicality of comparing to experiment.
- To capture the fluctuations that are important in small systems.

We have two basic methods:

- Introduce stochastic collisions with a “heat bath” to thermalise the system.
- Introduce a deterministic “non-Newtonian” dynamics that generates the correct ensemble.

Andersen Thermostat

Coupling to the heat bath is represented by stochastic impulsive forces that occasionally act on randomly selected particles.

These stochastic collisions move the system between different constant energy shells.

Strength of coupling controlled by ν, the frequency of stochastic collisions.

Algorithm:

- Integrate the deterministic equations of motion normally: $\mathbf{r}^N(t) \rightarrow \mathbf{r}^N(t + \Delta t)$.
- Check whether any particles are selected to undergo a collision with the heat bath. The probability that a particle is selected is $\nu \Delta t$.
- If a particles has been selected to undergo a collision, its new momentum is drawn from a Maxwell-Boltzmann distribution.

Proof: the steady state distribution of the canonical ensemble is

$$\rho_0(\mathbf{r}^N, \mathbf{p}^N) = \frac{1}{Z(N, V, T)} \exp(-\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)/kT) \tag{39}$$

What happens if we start an infinite number of simulations, distributed according to ρ_0?

- Deterministic dynamics will preserve the distribution.
- Stochastic changes of momentum will also preserve the distribution, as:
  - All momenta are equally likely to be selected for change.
  - The new momentum selected is drawn from the distribution itself.

We now evoke the *ergodic hypothesis* to claim that a single simulation will tend to sample from such a distribution in the limit of infinite simulation length.

Pros: Simple.

Cons: The dynamics is not realistic for an atomistic simulation.

Nosé-Hoover thermostat

We use an extended Lagrangian, where the extra degree of freedom acts as a reservoir and allows energy to flow back and forth between the system and the reservoir.

$$\mathcal{L}_{Nose} = \sum_{i=1}^{N} \frac{m_i}{2} s^2 \dot{\mathbf{r}}_i^2 - \mathcal{V}\left(\mathbf{r}^N\right) + \frac{Q}{2}\dot{s}^2 - \frac{L}{\beta}\ln(s). \tag{40}$$

L is a parameter and Q is an effective mass associated with s that controls the rate of temperature fluctuations. The momenta are

$$\begin{aligned} \mathbf{p}_i &= \frac{\partial \mathcal{L}}{\partial \mathbf{r}_i} = m_i s^2 \dot{\mathbf{r}}_i, & (41) \\ p_s &= \frac{\partial \mathcal{L}}{\partial \dot{s}} = Q\dot{s}. & (42) \end{aligned}$$

The Hamiltonian of this extended system is

$$\mathcal{H}_{Nose} = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i s^2} + \mathcal{V}\left(\mathbf{r}^N\right) + \frac{p_s^2}{2Q} + \frac{L}{\beta}\log(s). \tag{43}$$

The partition function of this system of $6N+2$ degrees of freedom in the microcanonical ensemble is

$$\mathcal{Q}_{Nose} = \frac{1}{N!} \int \delta \left(E - \mathcal{H}_{Nose} \right) dp_s \, ds \, d\mathbf{r}^N d\mathbf{p}^N \tag{44}$$

It can be shown [6] that, if $L = 3N + 1$,

$$\langle A(\mathbf{r}^N, \mathbf{p}^N/s) \rangle_{Nose} = \langle A \rangle_{NVT}. \tag{45}$$

The equations of motion are

$$\frac{d\mathbf{r}_i}{dt} = \frac{\partial \mathcal{H}_{Nose}}{\partial \mathbf{p}_i} = \frac{\mathbf{p}_i}{m_i s^2} \tag{46}$$

$$\frac{d\mathbf{p}_i}{dt} = -\frac{\partial \mathcal{H}_{Nose}}{\partial \mathbf{r}_i} = -\frac{\partial \mathcal{V}\left(\mathbf{r}^N\right)}{\partial \mathbf{r}_i} \tag{47}$$

$$\frac{ds}{dt} = \frac{\partial \mathcal{H}_{Nose}}{\partial p_s} = \frac{p_s}{Q} \tag{48}$$

$$\frac{dp_s}{dt} = -\frac{\partial \mathcal{H}_{Nose}}{\partial s} = \frac{1}{s} \left(\sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{m_i s^2} - \frac{L}{\beta} \right) \tag{49}$$

s in fact acts a bit like a scaling of the time, meaning that the physical time associated with each time step in simulation is actually changing. These equations can be reformulated in more physical terms, by defining

- ▷ $\mathbf{p}'_i = \mathbf{p}_i/s$.
- ▷ $dt'/dt = 1/s$.
- ▷ $\xi = p_s/Q$.

The result is

$$\frac{d\mathbf{r}_i}{dt'} = \frac{\mathbf{p}'_i}{m_i}, \tag{50}$$

$$\frac{d\mathbf{p}'_i}{dt'} = -\frac{\partial \mathcal{V}\left(\mathbf{r}^N\right)}{\partial \mathbf{r}_i} - \xi \mathbf{p}'_i, \tag{51}$$

$$\frac{d\ln s}{dt'} = \xi, \tag{52}$$

$$\frac{d\xi}{dt'} = \left(\sum_{i=1}^{N} \frac{\mathbf{p}'^2_i}{m_i} - \frac{L}{\beta}\right). \tag{53}$$

In this case, it can be shown that the canonical ensemble is sampled if $L = 3N$ [7].

Constant pressure Molecular Dynamics

An extended Lagrangian method can also be introduced to generate trajectories in the isobaric (NPT) ensemble in an analogous way.

Those interested should refer to [8].

The Langevin approach

In fully atomistic simulations of dilute solutions, solvent molecules have two particularly important effects on the solute.

- Solute molecules do not move ballistically.
- Solvent molecules act as a thermal reservoir.

Many simulations are performed using coarse-grained models and implicit solvents. We can add drag and random noise forces on each momenta of each particle to mimic these solvent effects

$$\dot{p}_i = F_i(\mathbf{r}) - \gamma_i p_i + \eta_i(t). \tag{54}$$

We conventionally use an independent gaussian random variable for each $\eta_i(t)$.

▷ $\langle \eta_i(t) \rangle = 0; \quad \langle \eta_i(t) \eta_j(t') \rangle = \Gamma_i^2 \delta(t - t') \delta_{i,j}.$

If we choose $\Gamma_i^2 = 2 m_i \gamma_i kT$ [9], then it can be shown that the particles will sample the canonical ensemble.

The Andersen thermostat can also be used with implicit solvent to generate diffusive motion.

Additional degrees of freedom

So far we have imagined that every simulation particle has only translational degrees of freedom. However, we may wish:

- To consider molecules.
- To consider electrons within atoms.

If we work with a fully atomistic representation, then can simply treat the atoms within a molecule in the same way as any other atom.

- ▷ Bonds within a molecule are much stiffer than the interactions between non-bonded atoms.
- ▷ Stiff interactions will determine the time step, and constrain it to be much smaller than is necessary for the intermolecular interactions.

Rigid and stiff molecules

There are several common approaches:

- Treat molecules as absolutely rigid, and describe them using *quaternions* [10] as well as linear coordinates.
  - ▷ Only appropriate in simple cases.
- Treat the molecules as rigid, but have separate forces acting on each atom. Constrain structure using algorithms such as SHAKE [11].
  - ▷ These algorithms are fast (approximate) methods to solve for the value of Lagrange multipliers necessary to preserve distances at each time step.
- Use a multiple time step integration algorithm, *e.g.* RESPA (reference system propagator algorithm) [12].

‘Car-Parrinello’ Approaches

For certain systems, evaluation of the energy requires an optimization of internal degrees of freedom (e.g. electrons, induced dipoles of polarizable atoms) of the atoms.

As the time scales associated with these internal degrees of freedom are much faster than those associated with the nuclear motion, it is assumed that they respond adiabatically to changes in the positions of the atoms. ***I.e.*** the electrons are always in their ground state.

Thus, in a standard molecular dynamics one would need to perform an iterative optimization at each time step.

Problems:

- Computationally expensive
- Incomplete convergence leads to the exertion of a drag on the nuclei.

Car-Parrinello molecular dynamics treats the electrons using density functional theory in the local density approximation.

It circumvents these problems using an extended Lagrangian that includes the electronic density. *I.e.* there are equations of motion for the electron density that are integrated in an analogous way to those for the nuclear coordinates.

Although the electron density is not always optimal - instead it fluctuates around this optimal value - it does not exert any systematic drag on the nuclei.

Optimization of the electron density occurs ‘on-the-fly’, rather than by iteration at each step.

Note that although CPMD computes the potential energy at each point using electronic density functional theory (rather than using a potential function) the nuclear motion is still treated classically.

Application to polarizable atoms

For a system of polarizable atoms (say with permanent dipoles), the total potential energy can be written as $\mathcal{V} = \mathcal{V}_0 + \mathcal{V}_{ind}$ the induction energy is given by

$$\mathcal{V}_{ind} = -\sum_i \mathbf{E}_i \cdot \mu_i + \frac{1}{2\alpha}\sum_i \mu_i^2 \tag{55}$$

where μ_i is the induced dipole on atom i, $\mathbf{E}_i$ is the electric field at atom i due to the other atoms, and α is the polarizability.

Minimization of $\mathcal{V}_{ind}$ with respect to the μ_i implies that the induced dipoles must satisfy

$$\mu_i = \alpha \mathbf{E}_i \tag{56}$$

Calculation of these induced dipoles, and hence the induction energy, requires the solution of these 3N linear equations, typically by iteration to self-consistency.

To avoid this iteration in an MD simulation, one can use the extended Lagrangian

$$\mathcal{L}\left(\mathbf{r}^N, \mu^N\right) = \frac{1}{2}\sum_i m\dot{\mathbf{r}}_i^2 + \frac{1}{2}\sum_i M\dot{\mu}_i^2 - \mathcal{V}\left(\mathbf{r}^N, \mu^N\right) \tag{57}$$

where M is the inertial ‘mass’ associated with the dipoles.

The equations of motion for the dipoles is then

$$M\ddot{\mu}_i = \frac{\partial \mathcal{L}}{\partial \mu_i} = \mathbf{E}_i - \frac{\mu_i}{\alpha} \tag{58}$$

The RHS of this equation represents the forces on the dipoles and is zero when they take their optimal value (Eq. 56).

Writing $\mu_i = \mu_i^{opt} + \Delta\mu_i$, gives $M\ddot{\mu}_i = -\Delta\mu_i/\alpha$.

Therefore, the induced dipoles will oscillate around their optimal values.

To ensure that the dipoles respond rapidly to changes in the atomic positions: $M \ll m$

To ensure that the dipoles only exhibit small oscillations about their optimal value: $T_\mu \ll T_r$ (temperatures defined through kinetic energies)

Latter achieved by two Nosé-Hoover thermostats.

Molecular Simulation Methods in Chemistry

Lecture 4:
Introduction to Monte Carlo sampling

Thomas Ouldridge

t.ouldridge1@physics.ox.ac.uk

http://www-thphys.physics.ox.ac.uk/people/ThomasOuldridge/Site/Teaching.html

Monte Carlo

Monte Carlo is a stochastic method to calculate integrals.

Why integrals?

Common in classical statistical mechanics

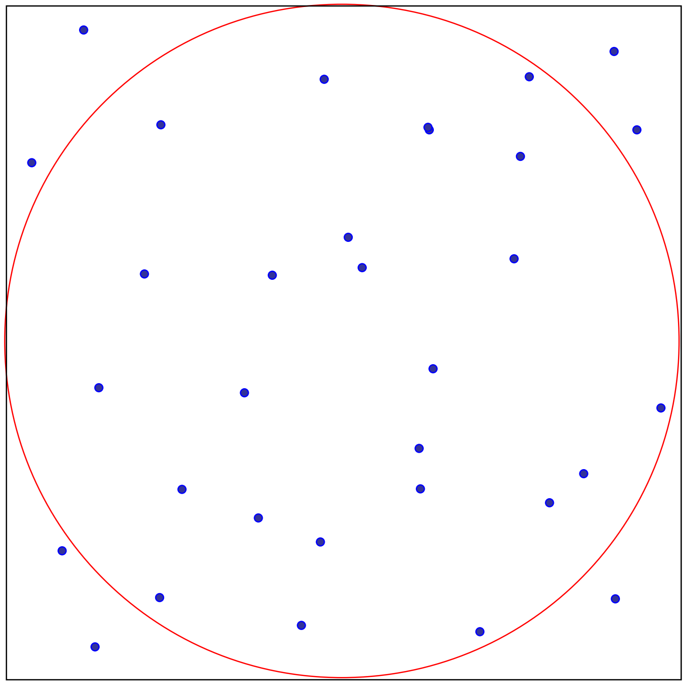
$$\langle A \rangle_{NVT} = \frac{\int A\left(\mathbf{r}^N, \mathbf{p}^N\right) \exp(-\beta E) d\mathbf{r}^N d\mathbf{p}^N}{\int \exp(-\beta E) d\mathbf{r}^N d\mathbf{p}^N} \tag{59}$$

1947: von Neumann, Metropolis and Ulam: First computational Monte Carlo application. Diffusion of neutrons in fissile material.

1953: Metropolis *et al.*: First Monte Carlo simulation of a molecular system.

Called Monte Carlo because of its use of random numbers.

Calculating π by MC



Throwing darts randomly at the target

$$\lim_{n_{throws}\to\infty} \frac{n_{hits}}{n_{throws}} = \frac{\pi}{4} \tag{60}$$

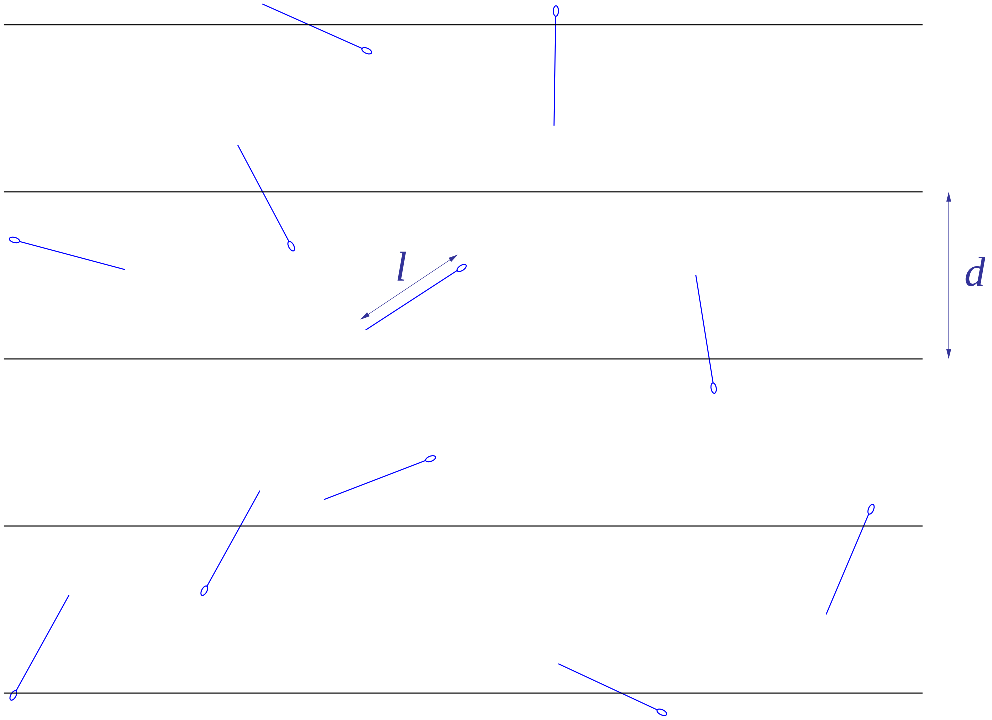
As an integral

$$\frac{1}{n_{throws}} \sum_i \mathcal{O}_i = \frac{\int_{-1}^{1} \mathcal{O}(x,y)dxdy}{\int_{-1}^{1} dxdy} \tag{61}$$

where

$$\mathcal{O}(x,y) = \begin{cases} 1 & r \leq 1 \\ 0 & r > 1 \end{cases} \tag{62}$$

Buffon's Needle

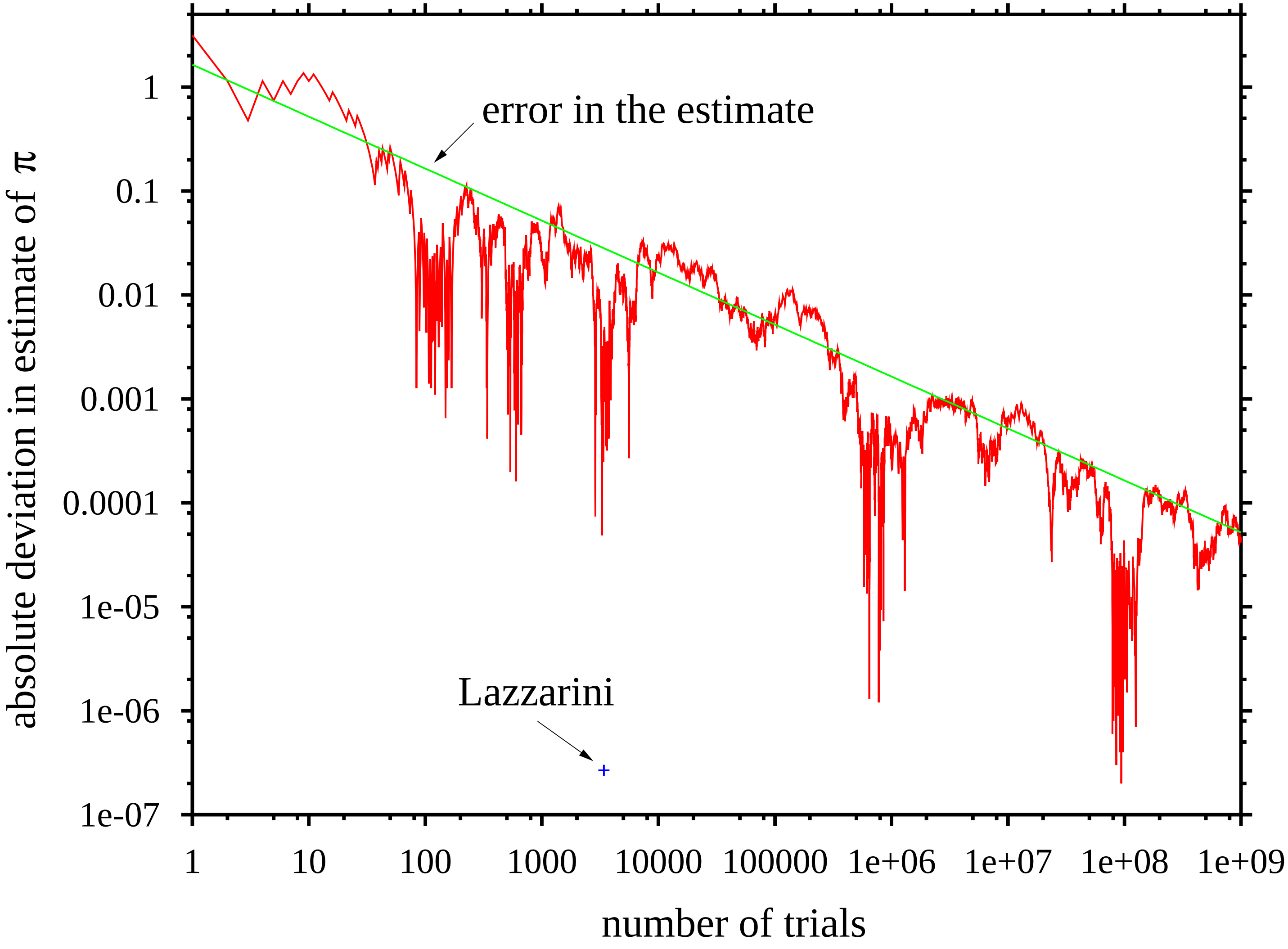


The probability of a needle overlapping a crack between floorboards can also be used to obtain π.

$$p_{overlap} = \frac{2l}{\pi d} \tag{63}$$

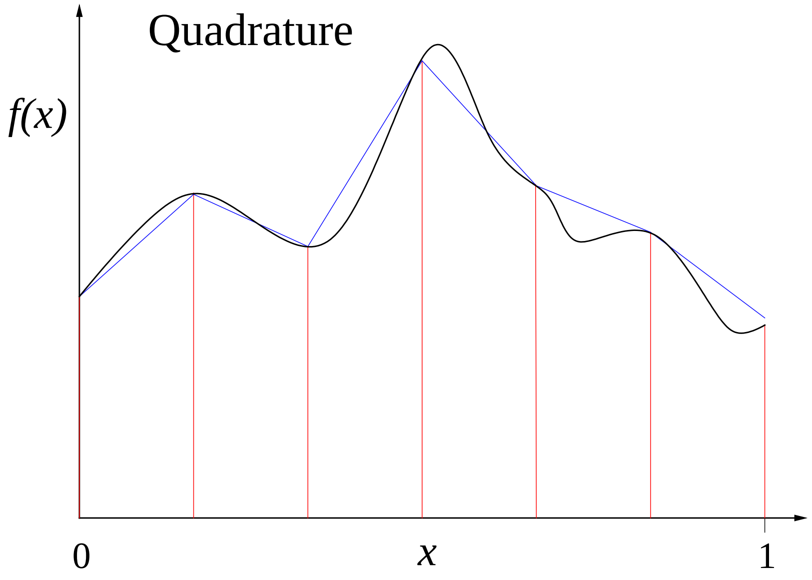
First suggested by Comte de Buffon in 1777.

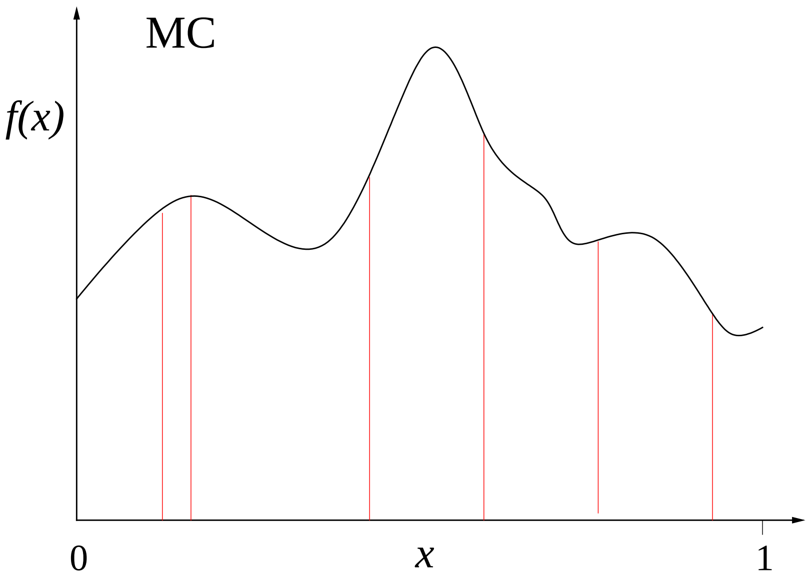
In 1901 Lazzarini claimed an estimate of $\pi = 3.1415929$ (355/113), i.e. an accuracy of 7SF, on the basis of 3408 trials.



Results of MC estimation of π.

Estimating 1D integrals





MC estimate for the integral

$$I = \int_0^1 f(x)dx \approx \frac{1}{M} \sum_{i=1}^{M} f(x_i) \tag{64}$$

where M random points x_i in the interval $[0, 1]$ have been selected.

Error analysis: MC

The error in the estimate of I, σ_I, obeys

$$\begin{aligned}
\sigma_I^2 &= \left\langle (I - \langle I \rangle)^2 \right\rangle && (65)\\
&= \left\langle \left(\frac{1}{M} \sum_{i=1}^{M} f(x_i) - \left\langle \frac{1}{M} \sum_{i=1}^{M} f(x_i) \right\rangle \right)^2 \right\rangle && (66)\\
&= \frac{1}{M^2} \left\langle \left(\sum_{i=1}^{M} (f(x_i) - \langle f(x) \rangle) \right) \left(\sum_{j=1}^{M} (f(x_j) - \langle f(x) \rangle) \right) \right\rangle && (67)\\
&= \frac{1}{M^2} \left\langle \sum_{i=1}^{M} (f(x_i) - \langle f(x) \rangle)^2 \right\rangle = \frac{1}{M} \sigma_f^2 && (68)
\end{aligned}$$

as the deviation of $f(x_i)$ and $f(x_j)$ from the mean are uncorrelated unless $i = j$.

Error analysis: Quadrature

Integration using the simple trapezium rule gives

$$I = \frac{1}{M}\left(\frac{1}{2}f(0) + \sum_{i}^{N-2} f(x_i) + \frac{1}{2}f(1)\right) + \mathcal{O}\left(\frac{1}{M^3}\right) \tag{69}$$

where the $x_i = i/N$ are equally spaced.

Thus, for 1D integrals quadrature methods converge much faster than the MC approach.

For higher dimensional integrals using the trapezium rule, the error has the same dependence on the number of points m in each dimension, i.e. $\mathcal{O}(1/m^3)$. However, as the total number of points in a d-dimension grid is $M = m^d$ the error is $\mathcal{O}(1/M^{3/d})$.

The scaling for the MC error is independent of dimension, i.e. $\sigma_I \sim 1/\sqrt{M}$.

$\Rightarrow$ MC scales better than trapezium rule for $d > 6$.

Importance sampling

As the Boltzmann weights in the statistical mechanics integrals that we want to evaluate (Eq. 59) are non-negligible for only a small fraction of configuration space, random uniform sampling will be of no use.

- If you wanted to estimate the total number of wild baby pandas in the world, a uniform sampling in space would be a waste of time.
- For 100 hard spheres at the freezing transition, only one in 10^{260} configurations have a non-zero Boltzmann weight.

We would like to preferentially sample regions where an integral has the most weight.

Back to 1D: Sample points from a probability distribution $w(x)$. Estimate of integral now:

$$I \approx \frac{1}{M} \sum_{i=1}^{M} \frac{f(x_{i|w})}{w(x_{i|w})} \tag{70}$$

where $x_{i|w}$ means that the x_i have been drawn from the distribution $w(x)$.

Even though this procedure does not change the scaling of accuracy with M, it's effect on the prefactor can be vast, particularly for thermodynamic systems.

A good choice of $w(x)$ can lead to enhanced accuracy.

- ▷ For example, we could sample the number of wild baby pandas by biasing according to a previous measurement of the total density of wild pandas.

The ideal weight function for us would be $\exp(-\beta E)/Q$. However, we don't know the partition function Q.

Metropolis resolved this impasse by using a Markov chain method.

He used a biased random walking which one only requires the relative probabilities of different configurations.

Metropolis Monte Carlo

Basic approach

- Start with a configuration o.
- Move to a new configuration n with a transition probability $\pi(o \to n)$.
- Calculate properties of interest for the current configuration and add to average.
- Repeat to create a trajectory through configuration space.

Note:

- The configuration does not have to change at each step, *i.e.*

$$\pi(o \to o) = 1 - \sum_{n \neq o} \pi(o \to n) \neq 0 \tag{71}$$

- The trajectories generated are stochastic.

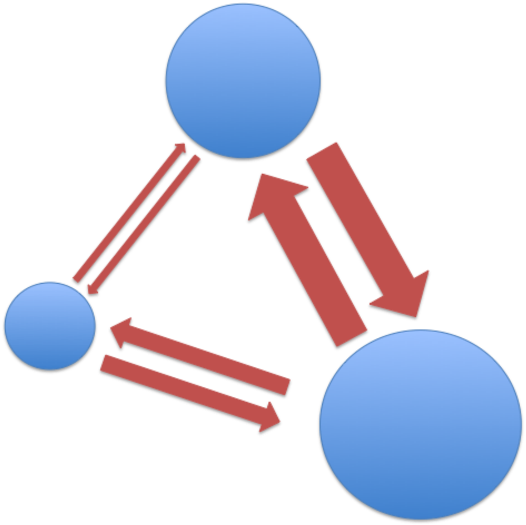
Detailed balance

Aim: sample configurations from the Boltzmann distribution $P(o) = \exp(-\beta E_o)/Q$.

- As with the Andersen thermostat, we imagine having a large number of simulations, with initial states distributed according to the Boltzmann distribution.
- We then ask whether the dynamics of the system will tend to preserve this distribution.

i.e., the net probability flow into and out of every state must balance.

$$P(o) \sum_i \pi(o \rightarrow i) = \sum_j P(j)\pi(j \rightarrow o) \tag{72}$$



This equation can be satisfied, if *detailed balance* is obeyed:

$$P(o)\pi(o \rightarrow n) = P(n)\pi(n \rightarrow o) \tag{73}$$

Obeying this condition guarantees equilibrium is maintained. It does not say if or how equilibrium will be reached.

Moves are broken into two parts

- Selecting a given destination state n, with probability $p_c(o \rightarrow n)$.
- Accepting the trial move $o \rightarrow n$, with probability $p_a(o \rightarrow n)$.

$$\pi(o \rightarrow n) = p_c(o \rightarrow n) p_a(o \rightarrow n) \tag{74}$$

As in the original Metropolis scheme, we commonly chose p_c to be symmetrical, i.e. $p_c(o \rightarrow n) = p_c(n \rightarrow o)$.

Applying *detailed balance* implies:

$$\frac{\pi(o \rightarrow n)}{\pi(n \rightarrow o)} = \frac{p_a(o \rightarrow n)}{p_a(n \rightarrow o)} = \frac{P(n)}{P(o)} = \exp\left[-\beta(E_n - E_o)\right] \tag{75}$$

Metropolis acceptance criterion

Many possible choices of $p_a(o \to n)$ that satisfy Eq. 75. The form proposed by Metropolis *et al.* (and most commonly used) is

$$p_a(o \to n) = \begin{cases} P(n)/P(o) & P(n) < P(o) \\ 1 & P(n) \geq P(o) \end{cases} \tag{76}$$

For the Boltzmann distribution, this is

$$p_a(o \to n) = \begin{cases} \exp\left[-\beta(E_n - E_o)\right] & E_n > E_o \\ 1 & E_n \leq E_o \end{cases} \tag{77}$$

Downhill moves are always accepted, whereas uphill moves are accepted with a probability that decreases with the energy difference.

Summary:

- ▷ Metropolis Monte Carlo samples from the Boltzmann distribution.
- ▷ We only know relative weights with which states are sampled, not absolute $w(x)$.

$\therefore$ We can only calculate ratios of integrals – usually ok (*e.g.* Eq. 59).

Integrating out the momenta

You may have noticed that I haven't been worrying about momenta. This is because integrals over the momentum coordinates can be easily performed [13].

$$\mathcal{Q} = \frac{1}{N!\Lambda^{3N}} Z \tag{78}$$

where Λ is the thermal wavelength, and Z is the configurational partition function:

$$Z = \int \exp\left[-\beta \mathcal{V}\left(\mathbf{r}^N\right)\right] d\mathbf{r}^N. \tag{79}$$

The $N!$ term accounts for indistinguishability [14].

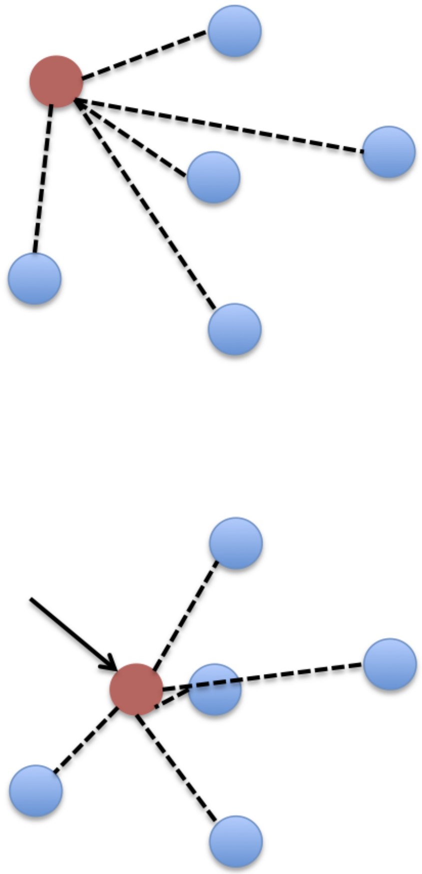
Similarly, for quantities that only depend on $\mathbf{r}^N$, no point using MC to perform the integral over the momenta in the average $\langle A \rangle_{NVT}$ (Eq. 59).

Therefore, in Metropolis MC only $\mathbf{r}^N$ is considered and configurations are sampled with probability

$$P\left(\mathbf{r}^N\right) = \frac{1}{Z} \exp\left[-\beta \mathcal{V}\left(\mathbf{r}^N\right)\right] \tag{80}$$

Basic step in more detail

To move from step k to $k+1$



- Select a particle i at random from $\mathbf{r}_k^N$
- Add a random displacement to this atom: $\mathbf{r}_i' = \mathbf{r}_i + \mathbf{\Delta}$
- Calculate the potential energy $\mathcal{V}(\mathbf{r}_{trial}^N)$ for new trial configuration.
- Accept the move $\mathbf{r}_k^N \to \mathbf{r}_{trial}^N$ with probability

$$p_a(o \to n) = \min\left\{1, \exp\left[-\beta\left(\mathcal{V}\left(\mathbf{r}_{trial}^N\right) - \mathcal{V}\left(\mathbf{r}_k^N\right)\right)\right]\right\} \tag{81}$$

On acceptance

$$\mathbf{r}_{k+1}^N = \mathbf{r}_{trial}^N \tag{82}$$

On rejection

$$\mathbf{r}_{k+1}^N = \mathbf{r}_k^N \tag{83}$$

- Add $A(\mathbf{r}_{k+1}^N)$ to average.

Notes on algorithm

Random displacement $\boldsymbol{\Delta}$ can be chosen uniformly from a cube centred on the origin:

$$\Delta_\alpha = \Delta_{max}\left(\mathrm{ran}[0,1] - 0.5\right) \qquad \alpha = x, y, z \tag{84}$$

Choice of step size, Δ_{max}:

- If Δ_{max} is too large, most steps will be rejected.
- If Δ_{max} is too small, although most steps will be accepted, the system will only move very slowly through configuration space.

Optimal step size: Lowest statistical error for given CPU time.

Rule of thumb: approximately 50% acceptance rate reasonable. Don't take this too literally!

Optimal step size will vary with conditions. *E.g.* smaller in dense systems.

How to choose Δ_{max}? One way is to adjust the step size dynamically *during the equilibration period*.

For move selection matrix to be symmetric, *i.e.* $p_c(o \rightarrow n) = p_c(n \rightarrow o)$:

- Choice of atom needs to be random.
- Step choice algorithm (including the step size) must be the same for each step.

Why move only one atom at a time?

Cost of N single-particle moves is similar to that for a single N-particle move. For pair potential,

$$\mathcal{V}\left(\mathbf{r}^N_{trial}\right) = \mathcal{V}\left(\mathbf{r}^N_k\right) + \sum_{j \neq i} v\left(r^{trial}_{ij}\right) - v\left(r^k_{ij}\right) \tag{85}$$

- In a typical thermodynamic simulation, attempted moves are more likely to cause an energy increase.
- Moving multiple particles simultaneously and independently raises the chance of rejection.

$\therefore$ Single-particle move algorithms move more rapidly through configuration space.

Jargon: an MC cycle or sweep corresponds to N single-particle moves.

Molecular Simulation Methods in Chemistry

Lecture 5:
More Monte Carlo sampling

Thomas Ouldridge

t.ouldridge1@physics.ox.ac.uk

`http://www-thphys.physics.ox.ac.uk/people/ThomasOuldridge/Site/Teaching.html`

Inferring thermodynamic properties

In a large canonical simulation, the pressure is a well-defined quantity. We can calculate using the virial theorem as given in Eq. 38.

The heat capacity can be calculated from fluctuations in the total energy [15]:

$$C_v(T) = \frac{1}{kT^2}\left(\langle E^2\rangle - \langle E\rangle^2\right) \tag{86}$$

In principle, the information collected at temperature T_0 can be used to infer properties at any temperature T'.

- Let us assume we have "binned" our states according to some quantity of interest A and the energy E, giving $p(A, E, T_0)$.
- We can calculate (up to a constant) the density of states $\Omega(A, E)$ from the histogram obtained at T_0,

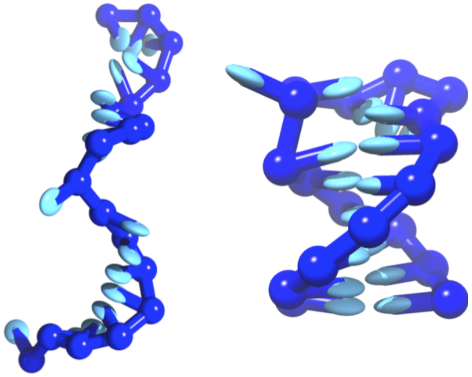
$$p(A, E, T_0) = \frac{1}{\mathcal{Q}(T_0)}\Omega(A, E)\exp\left(-E/kT_0\right). \tag{87}$$

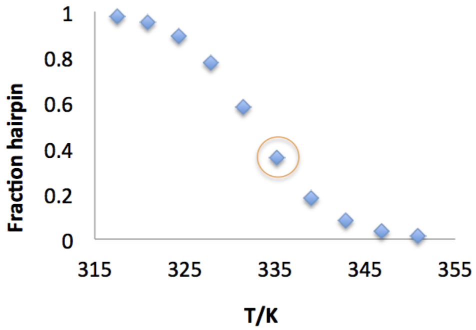
- As usual, the value of $\mathcal{Q}(T_0)$ is unknown.

$\Omega(A,E)$ can be used to calculate thermodynamic properties at any temperature, *i.e.*,

$$\langle A(T')\rangle = \frac{\int\int A\,\Omega(A,E)\exp\left(-E/kT'\right)dEdA}{\int\int \Omega(A,E)\exp\left(-E/kT'\right)dEdA} \quad (88)$$

$$= \frac{\int\int Ap(A,E,T_0)\exp\left((1/kT_0 - 1/kT')E\right)dEdA}{\int\int p(A,E,T_0)\exp\left((1/kT_0 - 1/kT')E\right)dEdA} \quad (89)$$





This extrapolation is known as *single-histogram reweighting*.

- ▷ Simplest application is for the energy itself.
- ▷ Can be easiest to do it on the fly, rather than saving $p(A,E,T_0)$.
- ▷ Only works well for small temperature gaps.
- ▷ Multi-histogram approaches use $p(A,E,T)$ from several T to obtain a 'best' estimate of $\Omega(E,T)$ [16].

Constant pressure Monte Carlo

For the isobaric-isothermal ensemble, averages are found as

$$\langle A \rangle_{NPT} = \frac{1}{\Theta_{NPT}} \int_0^\infty V^N \exp(-\beta PV) dV \int A\left(\mathbf{s}^N; V\right) \exp\left[-\beta \mathcal{V}\left(\mathbf{s}^N; V\right)\right] d\mathbf{s}^N \tag{90}$$

Scaled coordinates ($\mathbf{s}_i = \mathbf{r}_i / V^{1/3}$) have been used to separate out the volume dependence of the integral over $\mathbf{r}^N$.

A NPT MC scheme must sample configurations with the probability distribution:

$$P_{NPT}\left(\mathbf{s}^N; V\right) \propto \exp\left[-\beta\left(\mathcal{V}\left(\mathbf{s}^N\right) + PV - N\ln(V)/\beta\right)\right] \tag{91}$$

and must involve two kinds of moves

- Moves that randomly displace a particle.
- Moves that randomly change the volume V.

For both of these, $p_c(o \rightarrow n)$ must be symmetric.

The choice of move type must be random to maintain detailed balance.

For particle moves the acceptance criterion reduces to that for the canonical ensemble.

For volume moves the acceptance criterion is

$$p_a(o \to n) = \min\left\{1, \exp\left[-\beta\left(\mathcal{V}\left(\mathbf{s}^N; V_n\right) - \mathcal{V}\left(\mathbf{s}^N; V_o\right) + P\Delta V - \frac{N}{\beta}\ln\left[\frac{V_n}{V_o}\right]\right)\right]\right\} \tag{92}$$

where $\Delta V = V_n - V_o$.

The $P\Delta V$ term which favours a shrinkage of the volume is opposed by the $\ln(V_n/V_o)$ term and the tendency of particles to repel.

- The relative frequency with which the two move types are chosen should reflect their relative expense. Generally, a volume move is roughly as expensive as N single-particle moves.
- The average size of the volume move should be optimized to give reasonable acceptance ratios.

Pressure: virtual volume moves

It is possible to measure the pressure in a large canonical simulation using fake trial moves that change the volume of the simulation box.

$$P = -\left(\frac{\partial F}{\partial V}\right)_{N,T} = kT\left(\frac{\partial \ln Q}{\partial V}\right)_{N,T} \approx \frac{kT}{\Delta V}\ln\frac{Q_{V'}}{Q_V}, \tag{93}$$

with F the Helmholtz free energy and $\Delta V = V' - V$ a small volume change.

For a constant ΔV

$$\frac{Q_{V'}}{Q_V} = \frac{\frac{V'^N}{\Lambda^{3N}N!}\int \exp\left[-\beta\mathcal{V}\left(\mathbf{s}^N;V'\right)\right]d\mathbf{s}^N}{\frac{V^N}{\Lambda^{3N}N!}\int \exp\left[-\beta\mathcal{V}\left(\mathbf{s}^N;V\right)\right]d\mathbf{s}^N} \tag{94}$$

$$= \left(\frac{V'}{V}\right)^N \frac{\int \exp\left(-\beta\Delta\mathcal{V}\right)\exp\left[-\beta\mathcal{V}\left(\mathbf{s}^N;V\right)\right]d\mathbf{s}^N}{\int \exp\left[-\beta\mathcal{V}\left(\mathbf{s}^N;V\right)\right]d\mathbf{s}^N} \tag{95}$$

$$= \left(\frac{V'}{V}\right)^N \left\langle \exp\left(-\beta\Delta\mathcal{V}\right)\right\rangle. \tag{96}$$

Grand Canonical Monte Carlo

After integrating over the momenta, the probability distribution that the MC needs to sample is obtained:

$$P\left(\mathbf{s}^N; N\right) \propto \frac{V^N}{\Lambda^{3N} N!} \exp\left[-\beta\left(\mathcal{V}\left(\mathbf{s}^N\right) - \mu N\right)\right]. \tag{97}$$

- ▷ As the integral over the momentum depends on N it has to be explicitly considered.
- ▷ We also have to worry about indistinguishability.

In addition to particle displacement moves, there are two new move types

- Removal of a randomly chosen particle
- Addition of a particle at a random position

These moves should be attempted with equal probability.

The acceptance criteria are given by

$$p_a(N \to N-1) = \min\left\{1, \frac{\Lambda^3 N}{V} \exp\left[-\beta\left(\mu + \mathcal{V}\left(\mathbf{s}^{N-1}\right) - \mathcal{V}\left(\mathbf{s}^{N}\right)\right)\right]\right\} \tag{98}$$

$$p_a(N \to N+1) = \min\left\{1, \frac{V}{\Lambda^3(N+1)} \exp\left[-\beta\left(-\mu + \mathcal{V}\left(\mathbf{s}^{N+1}\right) - \mathcal{V}\left(\mathbf{s}^{N}\right)\right)\right]\right\} \tag{99}$$

Pros of grand-canonical ensemble:

- Direct access to μ
- Natural ensemble for processes, such as adsorption.

Cons

- Insertion probabilities can be very low for dense systems. Biasing techniques need to be used.

Widom Insertion

We can measure the chemical potential of a large system in a canonical simulation using *Widom insertion*.

$$\mu = \left(\frac{\partial F}{\partial N}\right)_{V,T} \approx -kT \ln\left[\frac{Q_{N+1}}{Q_N}\right]. \tag{100}$$

Rewriting this in terms of scaled coordinates

$$\begin{aligned} \mu &= -kT\ln\left[\frac{V}{\Lambda^3(N+1)}\right] - kT\ln\left[\frac{\int \exp\left[-\beta\mathcal{V}\left(\mathbf{s}^{N+1}\right)\right] d\mathbf{s}^{N+1}}{\int \exp\left[-\beta\mathcal{V}\left(\mathbf{s}^{N}\right)\right] d\mathbf{s}^{N}}\right] && (101) \\ &= \mu_{id} + \mu_{ex}. && (102) \end{aligned}$$

$\mu_{id} = -kT\ln\left[\rho\Lambda^{-3}\right]$ is the chemical potential of an ideal gas.

$$\mu_{ex} = -kT\ln\left[\int\left(\frac{\int \exp\left[-\beta\Delta\mathcal{V}\right]\exp\left[-\beta\mathcal{V}\left(\mathbf{s}^N\right)\right]d\mathbf{s}^N}{\int \exp\left[-\beta\mathcal{V}\left(\mathbf{s}^N\right)\right]d\mathbf{s}^N}\right)d\mathbf{s}_{N+1}\right] \quad (103)$$

$$= -kT\ln\left[\int \left\langle \exp\left[-\beta\Delta\mathcal{V}\right]\right\rangle_{NVT} d\mathbf{s}_{N+1}\right], \quad (104)$$

where $\Delta\mathcal{V} = \mathcal{V}\left(\mathbf{s}^{N+1}\right) - \mathcal{V}\left(\mathbf{s}^N\right)$.

Therefore, the excess chemical potential can be related to the average Boltzmann factor for inserting an extra particle into an N-particle system. The extra particle is a ‘ghost’ particle whose only purpose is to measure the chemical potential. It is never actually added to the system.

As with grand-canonical MC the insertion probability can be very low for dense systems leading to bad statistics.

Biased moves

In some systems, simple move-generating algorithms can be highly inefficient.

- Fairly dense polymer solutions.
- Systems of strongly interacting clusters

Need more sophisticated means of generating configurations.

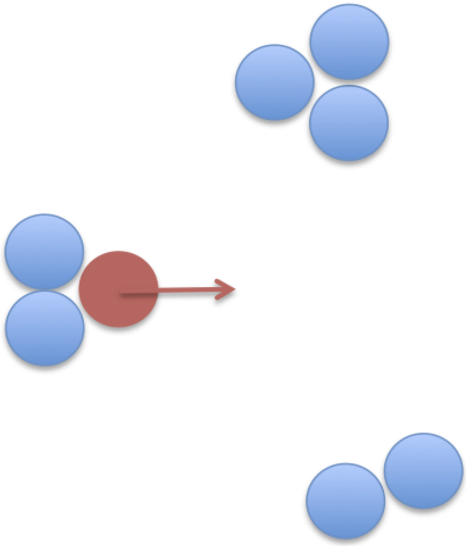
Use general Metropolis recipe, but do not choose the generation of trial moves to be symmetric.

I.e. $p_a(o \to n) = \min\{1, \chi\}$ and $p_a(n \to o) = \min\{1, 1/\chi\}$, where

$$\chi = \frac{P_n \, p_c(n \to o)}{P_o \, p_c(o \to n)} \tag{105}$$

E.g. For polymers "configurational bias Monte Carlo" [17] involves regrowing a section of polymer at each MC step. At each step in the regrowth process a number of directions for the new chain are explored, and the direction selected is biased towards those with lower energy.

Cluster moves



Single particle moves can lead to slow equilibration for strongly interacting systems.

- Large moves lead to rejection.
- Small moves don't change the configuration much.

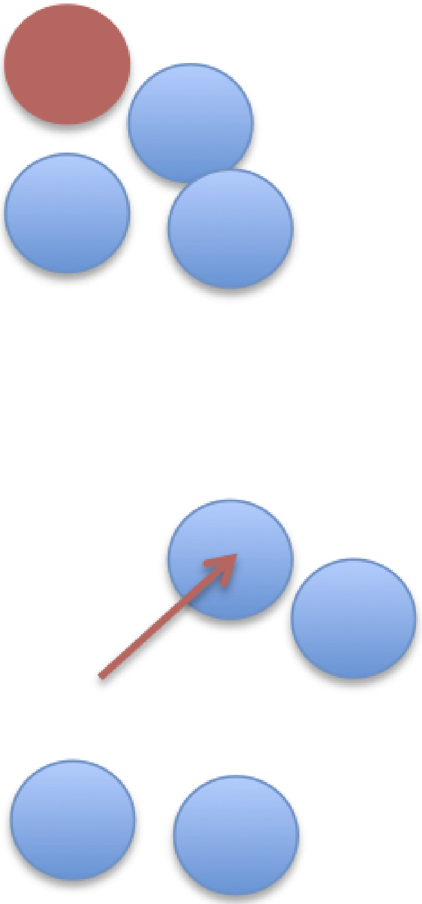
This is a particular problem in self-assembling systems.

We can propose *cluster moves* that involve simultaneous motion of a number of particles [18]. Two basic approaches to choosing a cluster:

- ▷ Choose a cluster based on the initial state of the system, then choose a trial move independently.
- ▷ Choose a cluster based on the initial state and the attempted move.

Both approaches generate trial moves in a biased fashion – acceptance factors are not the same as in single-particle Monte Carlo.

Troisi algorithm



An example (simple) cluster algorithm, applicable to systems with pairwise interactions.

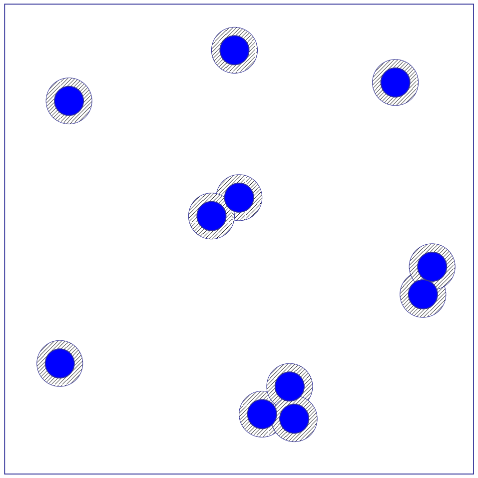
- Randomly pick a particle $\boldsymbol{i}$ as the cluster seed.
- Test all particles $\boldsymbol{j}$ with which $\boldsymbol{i}$ interacts in the initial state $\boldsymbol{o}$.
- Add $\boldsymbol{j}$ to cluster with probability $\boldsymbol{P_{ij} = 1 - \exp(\beta \min\{0, E_{ij}\})}$.
- Assign a random translation or rotation to the cluster.
- Calculate new energies.
- Let $\boldsymbol{\Delta E^{+} = E_{n}^{+} - E_{o}^{+}}$, where the subscript indicates that only positive contributions to the energy are included.
- Accept move with $\boldsymbol{p_a = \min\{1, \exp(-\beta \Delta E^{+})\}}$ [19].

The acceptance probability is justified in the notes.

Association Bias Monte Carlo

Adding explicit cluster association moves and cluster breakup moves could accelerate convergence.

Cluster association moves:



- Choose a random particle (probability $= 1/N$)
- Move this particle from $\mathbf{r}_o$ to a random trial position $\mathbf{r}_n$ in V_b, where V_b is the union of all the bonding regions around each particle. (probability $= 1/V_b$)

$$p_c(o \rightarrow n) = \frac{1}{NV_b} p_{\mathrm{assoc}} \tag{106}$$

Cluster breakup moves:

- Choose a random particle from those that are involved in clusters (probability $= 1/N_b$)
- Move this particle from $\mathbf{r}_o$ to a random trial position $\mathbf{r}_n$ (probability $= 1/V$)

$$p_c(n \rightarrow o) = \frac{1}{N_b V} p_{\mathrm{breakup}} \tag{107}$$

If the probability of performing association and breakup moves are equal, then the acceptance criteria for cluster association follows from

$$p_a(o \to n) = \min\left\{1, \frac{NV_b^o}{N_b^n V} \exp\left[-\beta\left(\mathcal{V}\left(\mathbf{r_n}^N\right) - \mathcal{V}\left(\mathbf{r_o}^N\right)\right)\right]\right\}. \tag{108}$$

Similarly, the acceptance factor for breakup moves is

$$p_a(o \to n) = \min\left\{1, \frac{N_b^o V}{NV_b^n} \exp\left[-\beta\left(\mathcal{V}\left(\mathbf{r_n}^N\right) - \mathcal{V}\left(\mathbf{r_o}^N\right)\right)\right]\right\}. \tag{109}$$

▷ Note, the V_b used is always for the state in which the trial particle is separated from the cluster, and the N_b is always for the state in which the trial particle is part of the cluster.

In practice determining V_b is not straightforward, but modifications make it simpler [20].

Molecular Simulation Methods in Chemistry

Lecture 6:
Sampling phase diagrams and accelerating equilibration

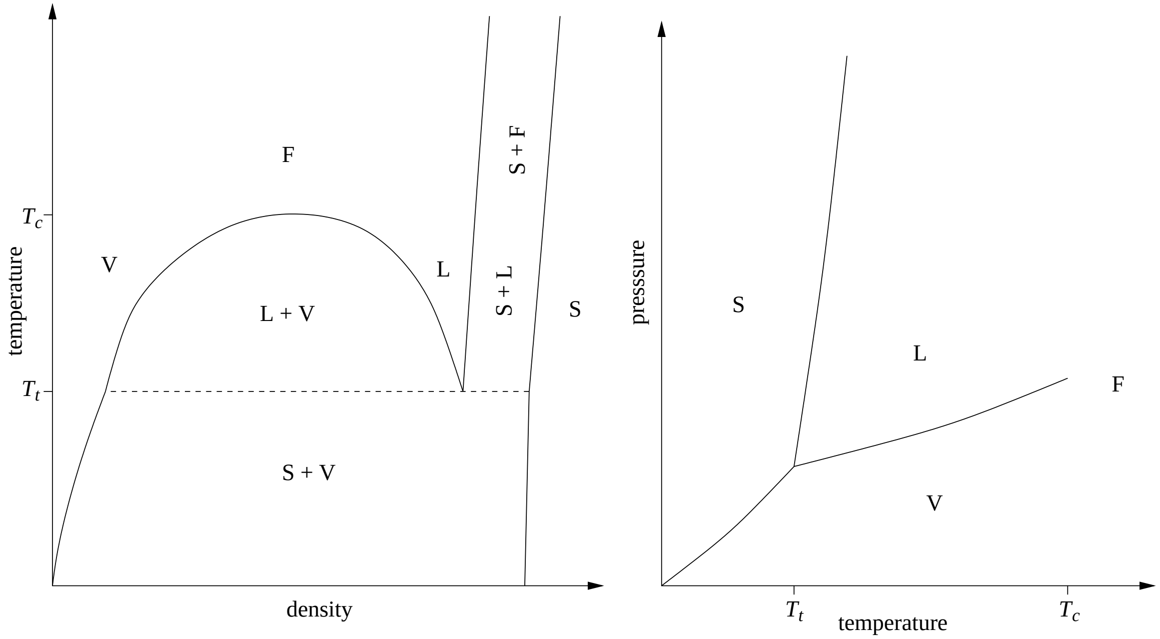
Thomas Ouldridge

t.ouldridge1@physics.ox.ac.uk

`http://www-thphys.physics.ox.ac.uk/people/ThomasOuldridge/Site/Teaching.html`

Computing Phase Diagrams

One important use of simulation techniques is to calculate phase diagrams.



Direct simulation approaches hindered by:

- dynamics of phase transition may be slow
- the effect of the interface between the two phases

The chemical potentials of the two phases are equal on coexistence lines.

Thermodynamic Integration

Trick: Compute the difference in free energy between the system of interest and a reference system for which the free energy is known (e.g. ideal gas, harmonic crystal).

- ▷ As free energies are extensive for bulk phases, specific free energies are also equal at coexistence points.

A path between these states can be defined.

$$\mathcal{V}(\lambda) = \mathcal{V}_{ref} + \lambda \left(\mathcal{V}_{sys} - \mathcal{V}_{ref}\right) \tag{110}$$

The partition function for a given λ is

$$\mathcal{Q}(N, V, T; \lambda) = \frac{1}{\Lambda^{3N} N!} \int \exp\left[-\beta \mathcal{V}(\lambda)\right] d\mathbf{r}^N \tag{111}$$

Hence,

$$\left(\frac{\partial F(\lambda)}{\partial \lambda}\right) = \frac{\int \left(\partial \mathcal{V}(\lambda)/\partial \lambda\right) \exp\left[-\beta \mathcal{V}(\lambda)\right] d\mathbf{r}^N}{\int \exp\left[-\beta \mathcal{V}(\lambda)\right] d\mathbf{r}^N} = \left\langle \frac{\partial \mathcal{V}(\lambda)}{\partial \lambda} \right\rangle_\lambda \tag{112}$$

Integrating gives

$$F(\lambda = 1) = F(\lambda = 0) + \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial \mathcal{V}(\lambda)}{\partial \lambda)} \right\rangle_{\lambda} d\lambda \tag{113}$$

In practice, simulations at a number of different values of λ are performed and the above equation integrated numerically.

- ▷ The path must be reversible, i.e. not cross a phase transition line.

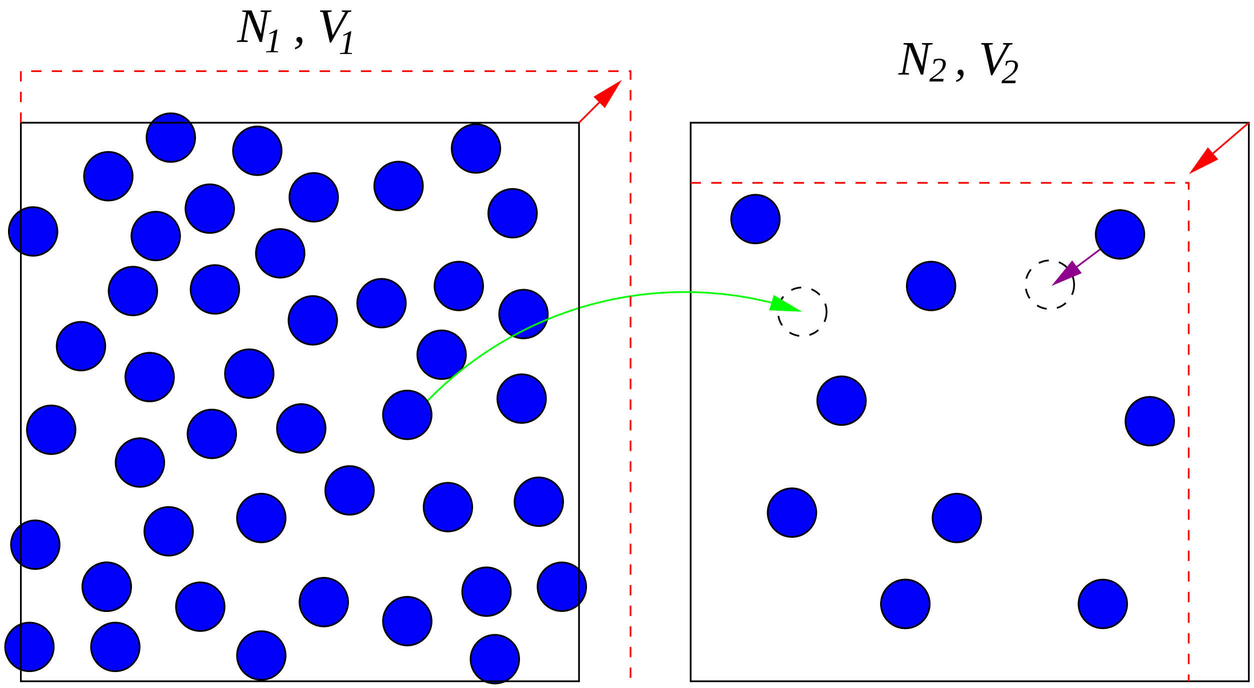
To locate a single point on the phase boundary, need to compute the free energies of both phase at a number of different state points, and then locate where they cross. *i.e.* lots of work!

Gibbs Ensemble

Direct simulation of phase coexistence is hindered by the problems of nucleation and the role of the interface in a small system.

One way round this is the Gibbs Ensemble introduced by Panagiotopoulos, where separate boxes contain the two phases and particles and volume can be exchanged between the two boxes.

Gives thermodynamic contact without physical contact.



Standard method for liquid-vapour coexistence.

Partition function for this ‘Gibbs ensemble’

$$\mathcal{Q}_G(N,V,T) = \sum_{N_1=0}^{N} \int_0^V Q_1(N_1,V_1,T)\, Q_2(N-N_1, V-V_1, T) dV_1 \tag{114}$$

$$= \sum_{N_1=0}^{N} \frac{1}{\Lambda^{3N} N_1!(N-N_1)!} \int_0^V V_1^{N_1} (V-V_1)^{N-N_1} dV_1$$

$$\times \int \exp\left[-\beta \mathcal{V}\left(\mathbf{s}^{N_1}\right)\right] d\mathbf{s}^{N_1} \int \exp\left[-\beta \mathcal{V}\left(\mathbf{s}^{N-N_1}\right)\right] d\mathbf{s}^{N-N_1} \tag{115}$$

Hence the distribution we wish to sample is:

$$P(N_1, V_1, \mathbf{s}_1^N, \mathbf{s}^{N-N_1}) \propto \frac{V_1^{N_1} (V-V_1)^{N-N_1}}{N_1!(N-N_1)!} \exp\left[-\beta\left(\mathcal{V}\left(\mathbf{s}^{N_1}\right) + \mathcal{V}\left(\mathbf{s}^{N-N_1}\right)\right)\right] \tag{116}$$

Types of move:

- Random particle displacement.
- Transfer of a randomly selected particle.
- Coupled volume change of the boxes (total volume constant).

Standard Metropolis acceptance criterion for particle displacements. Volume move:

$$p_a(o \to n) = \min\left\{1, \frac{\left(V_1^n\right)^{N_1}\left(V - V_1^n\right)^{N-1}}{\left(V_1^o\right)^{N_1}\left(V - V_1^o\right)^{N-1}} \exp\left[-\beta\left(\mathcal{V}\left(\mathbf{s}_n^N\right) - \mathcal{V}\left(\mathbf{s}_o^N\right)\right)\right]\right\} \tag{117}$$

Particle move from 1 to 2:

$$p_a(o \to n) = \min\left\{1, \frac{N_1(V - V_1)}{(N - N_1 + 1)V_1} \exp\left[-\beta\left(\mathcal{V}\left(\mathbf{s}_n^N\right) - \mathcal{V}\left(\mathbf{s}_o^N\right)\right)\right]\right\} \tag{118}$$

▷ A box is selected at random, then a particle from that box randomly selected.

Problems:

- Particle insertion slow when one system dense (worse for liquid as T is decreased).
- Identity of the boxes can change close to the critical point,
- Magnitude of fluctuations and correlation lengths increase close to the critical point.

Tracing coexistence curves

Idea:

- First find one point on the coexistence curve.
- Then trace the curve by integrating the Clausius-Clapeyron equation [21]:

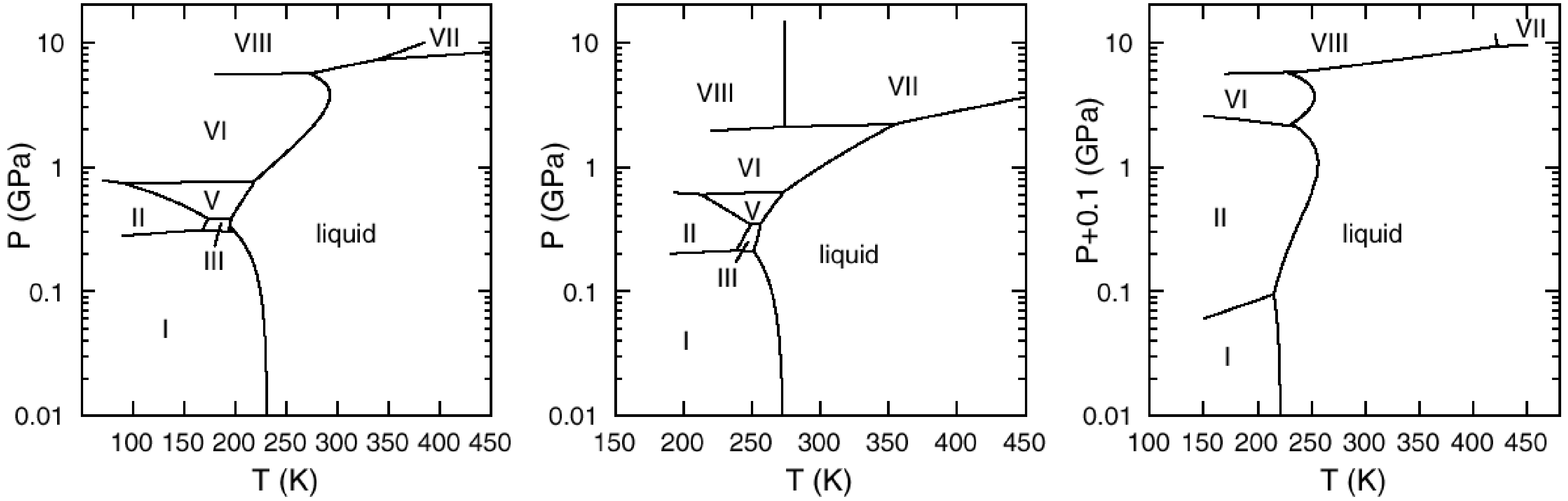
$$\frac{dP}{dT} = \frac{\Delta H}{T\Delta V} \tag{119}$$

▷ All these quantities can be calculated direct from simulation without further free energy calculations.

Method often called Gibbs-Duhem integration.

Analogous equations can be derived for tracing the phase boundaries as a function of other variables, e.g. the parameters of the intermolecular potential.

Water phase diagrams



Experimental water phase diagram (middle) compared to that for TIP4P (left) and SPC/E (right) water models.

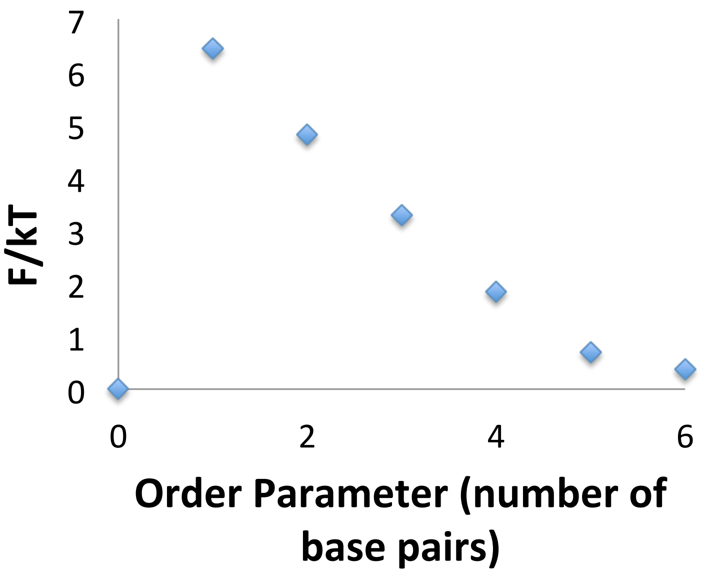
Clever tricks to speed up simulations

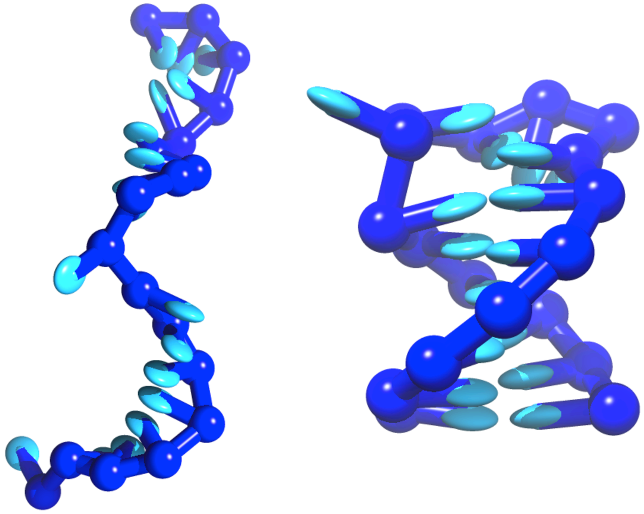
It is often useful to define a free energy as a function of some internal reaction coordinate $\lambda(\mathbf{r}^N)$.

$$F(\lambda_i) = -kT \ln\left(\int \mathrm{d}\mathbf{r}^N \exp\left(-\mathcal{V}\left(\mathbf{r}^N\right)\right)\delta_{\lambda,\lambda_i}\right) + F_c \tag{120}$$

Free energy barriers that are significantly larger than kT make sampling an equilibrium distribution with standard simulation techniques tricky.

- ▷ Here we discuss techniques for increasing barrier-crossing frequency.





Parallel Tempering

Barriers that are prohibitively high at low temperature can be easily crossed at sufficiently high temperature.

- ▷ Provided barriers are primarily *energetic* rather than *entropic*.

Basic idea: maintain equilibrium at low temperature by coupling runs to higher temperature through the occasional swapping of configurations between parallel simulations at different temperatures.

The partition function for this extended system of M coupled simulations is

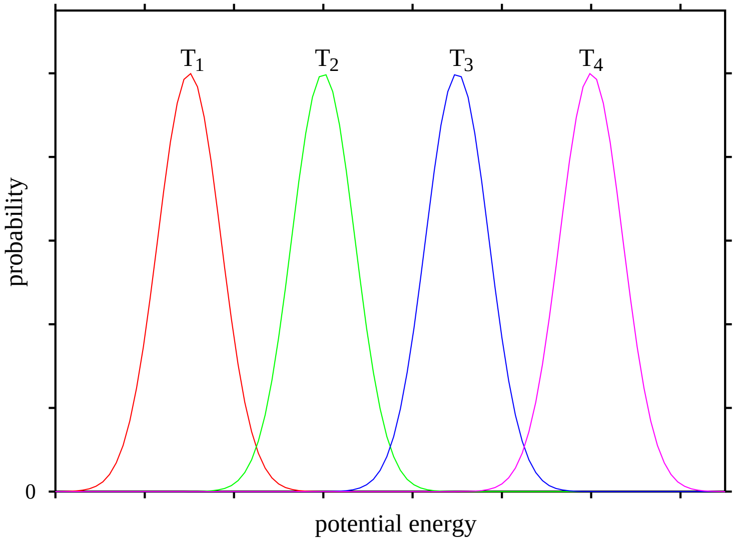
$$\mathcal{Q}_{extended}\left(N, V, \{T_k\}\right) = \prod_{k=1}^{M} \mathcal{Q}(N, V, T_k) = \prod_{k=1}^{M} \frac{1}{\Lambda_k^{3N} N!} \int \exp\left[-\beta_k \mathcal{V}\left(\mathbf{r}_k^N\right)\right] d\mathbf{r}_k^N \tag{121}$$

For the swapping moves, chose $p_c(o \to n)$ to be symmetric. *E.g.* by randomly choosing two of the M systems, and then switching their temperatures.

The acceptance criterion must satisfy:

$$\frac{p_a\left(\left\{\left(\mathbf{r}_a^N, T_a\right), \left(\mathbf{r}_b^N, T_b\right)\right\} \rightarrow \left\{\left(\mathbf{r}_b^N, T_a\right), \left(\mathbf{r}_a^N, T_b\right)\right\}\right)}{p_a\left(\left\{\left(\mathbf{r}_b^N, T_a\right), \left(\mathbf{r}_a^N, T_b\right)\right\} \rightarrow \left\{\left(\mathbf{r}_a^N, T_a\right), \left(\mathbf{r}_b^N, T_b\right)\right\}\right)} = \frac{P(\left(\mathbf{r}_b^N, T_a\right) P(\left(\mathbf{r}_a^N, T_b\right)}{P(\left(\mathbf{r}_a^N, T_a\right) P(\left(\mathbf{r}_b^N, T_b\right)}$$

$$= \frac{\exp\left[-\beta_a \mathcal{V}\left(\mathbf{r}_b^N\right) - \beta_b \mathcal{V}\left(\mathbf{r}_a^N\right)\right]}{\exp\left[-\beta_a \mathcal{V}\left(\mathbf{r}_a^N\right) - \beta_b \mathcal{V}\left(\mathbf{r}_b^N\right)\right]} = \exp\left[\left(\beta_a - \beta_b\right)\left(\mathcal{V}\left(\mathbf{r}_a^N\right) - \mathcal{V}\left(\mathbf{r}_b^N\right)\right)\right] \quad (122)$$



The probability of a switching move will be very low unless the potential energy distributions for the two temperatures being switched have sufficient overlap. ∴ Choice of temperatures is very important.

Umbrella Sampling

Basic idea: Bias the system to sample configurations near the top of the barrier by modifying the Hamiltonian.

- Let us assume that we have found a good order parameter, λ, that characterises the free-energy profile.
- Apply a biasing weight to the simulation, $W(\lambda(\mathbf{r}^N))$.

Configurations are sampled with a frequency

$$P_W(\mathbf{r}^N) \propto W(\lambda(\mathbf{r}^N)) \exp\left(-\beta \mathcal{V}\left(\mathbf{r}^N\right)\right) \tag{123}$$

Equilibrium averages can be found by unbiasing:

- ▷ Each sampled configuration contributes to a weighted average with a weight $1/W(\lambda(\mathbf{r}^N))$.

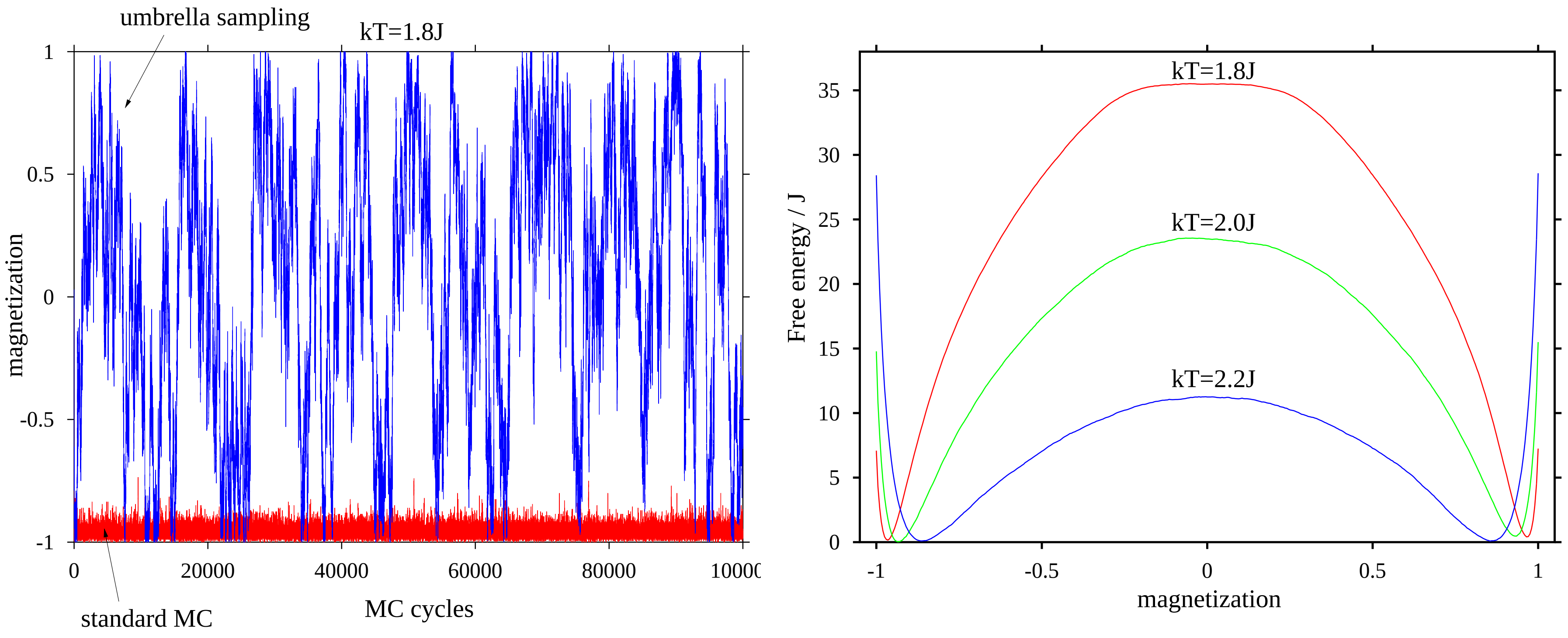
Another way to think of Umbrella Sampling is that we are introducing a fictitious potential $-kT \ln W(\lambda(\mathbf{r}^N))$.

Proof:

$$\langle A\rangle_{NVT}=\frac{\int A\left(\mathbf{r}^N\right)\exp(-\beta\mathcal{V}\left(\mathbf{r}^N\right))d\mathbf{r}^N}{\int\exp(-\beta\mathcal{V}\left(\mathbf{r}^N\right))d\mathbf{r}^N} \tag{124}$$

$$\Longrightarrow\ \langle A\rangle_{NVT}=\frac{\int\frac{A\left(\mathbf{r}^N\right)}{W(\lambda(\mathbf{r}^N))}W(\lambda(\mathbf{r}^N))\exp(-\beta\mathcal{V}\left(\mathbf{r}^N\right))d\mathbf{r}^N}{\int\frac{1}{W(\lambda(\mathbf{r}^N))}W(\lambda(\mathbf{r}^N))\exp(-\beta\mathcal{V}\left(\mathbf{r}^N\right))d\mathbf{r}^N}=\frac{\langle A/W\rangle_W}{\langle 1/W\rangle_W} \tag{125}$$

- ▷ As a rule of thumb, the ideal choice for the biasing distribution is $W(\lambda(\mathbf{r}^N))=-F(\lambda)$ since then $P_W(\lambda(\mathbf{r}^N))$ is constant.
- ▷ In practice, only need a reasonable estimate to sufficiently facilitate passage between the free energy minima. It can be obtained iteratively.
- ▷ In principle, Umbrella sampling allows exquisite control, but this depends on how good your order parameter is.
- ▷ Trivially extended to multiple dimensions.
- ▷ Can be used in a “windowed” fashion [22].

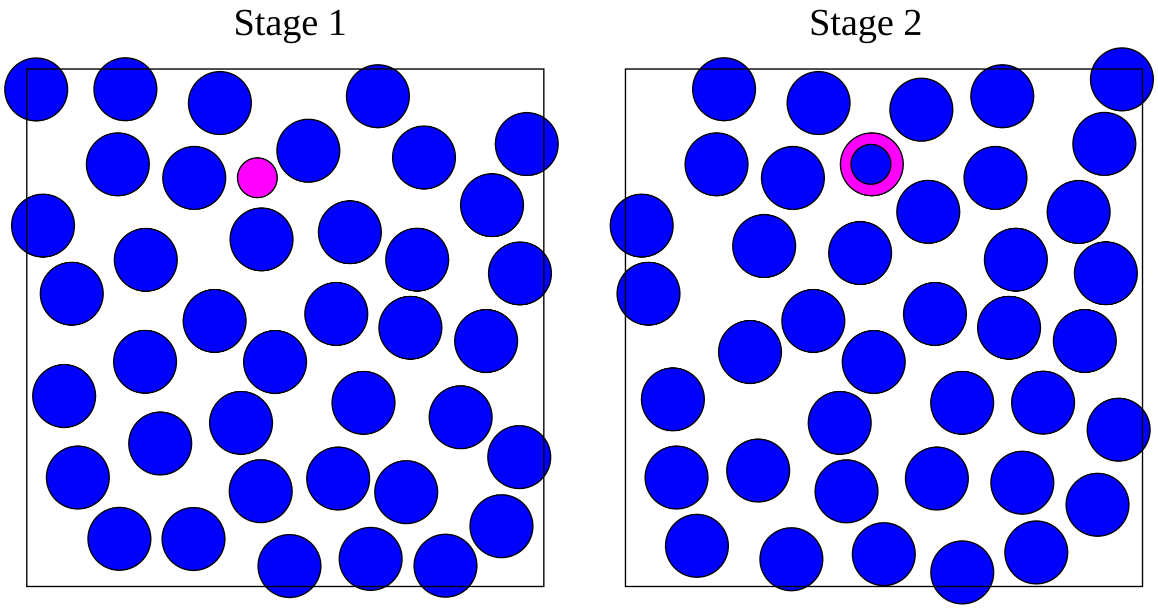


2D Ising model: 20×20 lattice.

Widom insertion and multiple staging

In the Widom insertion method, the probability of insertion is very low for dense configurations.

E.g. For hard spheres: One way to improve is to use insertion of a small sphere as an intermediate.



Stage 1: Measure fraction of time that small sphere finds no overlap.

Stage 2: Measure fraction of time that there is no overlap when small sphere is grown to full size.

Molecular Simulation Methods in Chemistry

Lecture 7:
Dynamical sampling of rare events

Thomas Ouldridge

t.ouldridge1@physics.ox.ac.uk

`http://www-thphys.physics.ox.ac.uk/people/ThomasOuldridge/Site/Teaching.html`

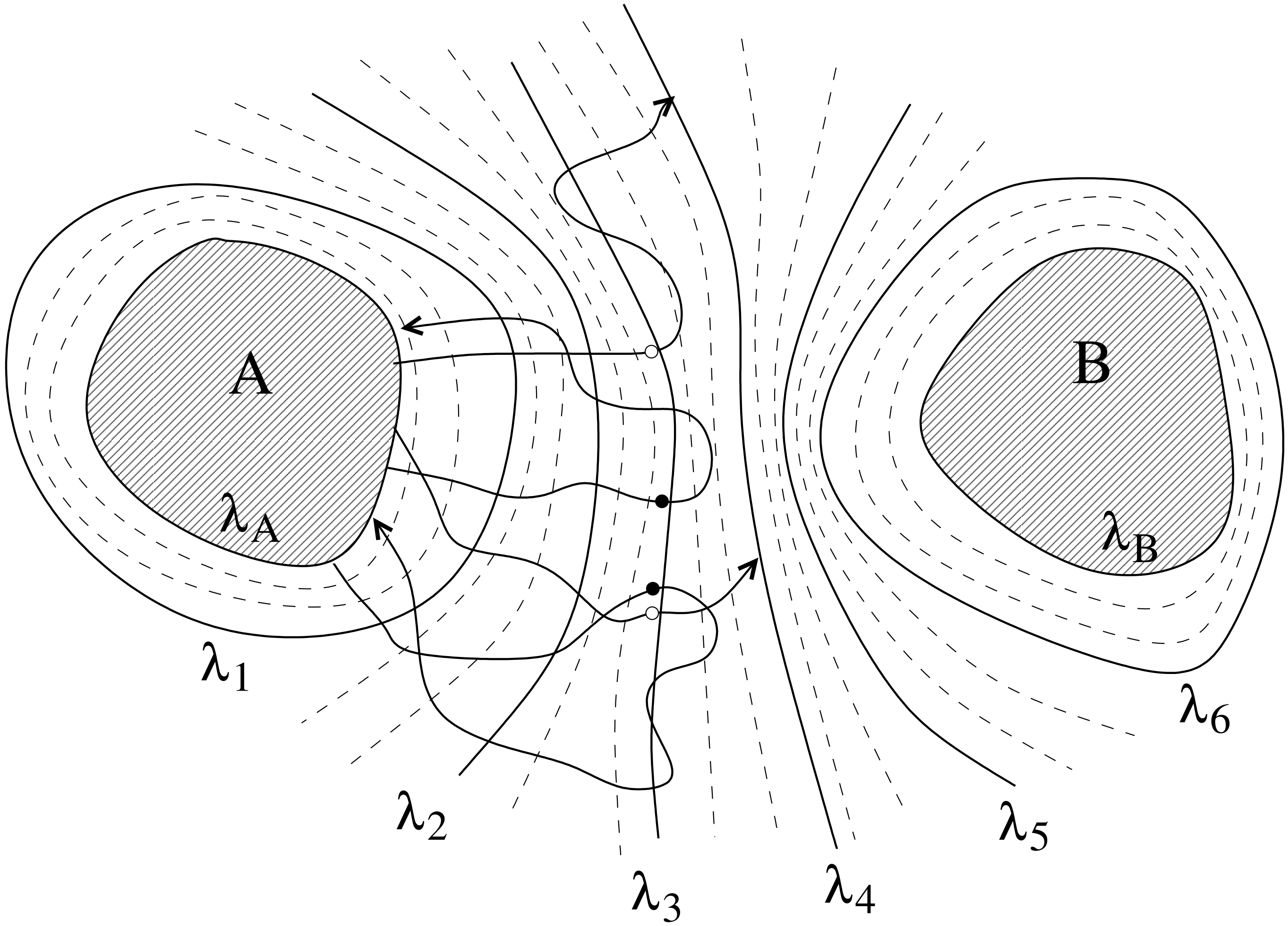
Why is wrong with previous enhanced sampling techniques?

Umbrella sampling and parallel tempering are excellent for equilibrium properties. They give limited direct information on dynamics.

- Umbrella sampling involves an artificial potential, so the dynamics of simulations will be different from the unbiased case.
- Trajectories from parallel tempering are unbiased, but split into sections at different temperature. Transitions tend to occur at temperatures that we're not interested in.

In this lecture we look at two methods which enhance sampling of true reaction pathways and allow the calculation of rate constants.

Formulation of the problem

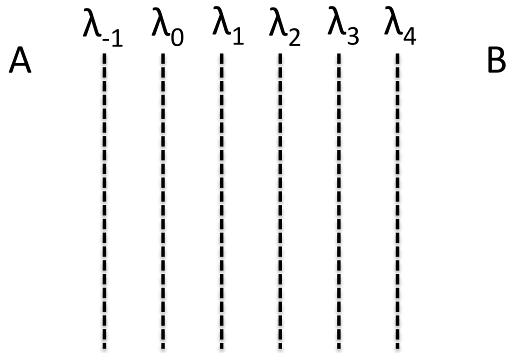


Forward Flux Sampling

Calculates the flux of trajectories leaving basin A that reach basin B before returning to A.

- Given a long simulation in which many transitions are observed, the flux of trajectories from A to B is $\Phi_{AB} = N_{AB}/(\tau f_A)$.
- N_{AB} is the number of times the simulation leaves A and then reaches B.
- τ is the total time simulated.
- f_A is the fraction of simulated time for which state A has been more recently visited than B.

Calculate by splitting into stages.



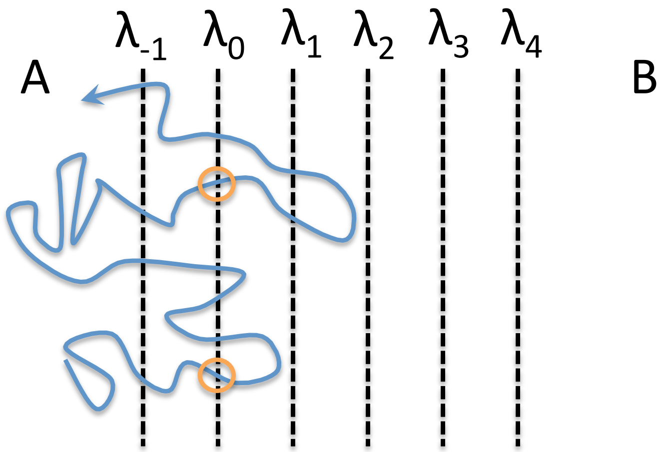
The use of the additional interface λ_{-1} is optional.

$$\Phi_{AB} = \phi(\lambda_0) p(\lambda_{\max}|\lambda_0). \tag{126}$$

- ▷ $\phi(\lambda_0)$ is the flux of trajectories crossing λ_0 for the first some since leaving A (*i.e.*, since crossing λ_{-1}).
- ▷ $p(\lambda_{\max}|\lambda_0)$ is the probability that such a trajectory will reach B (*i.e.*, cross $\lambda_{\max}$) before returning to A (crossing λ_{-1} in the other direction).

We first run a simulation to estimate $\phi(\lambda_0)$.

- This generates a set of points on the interface λ_0.

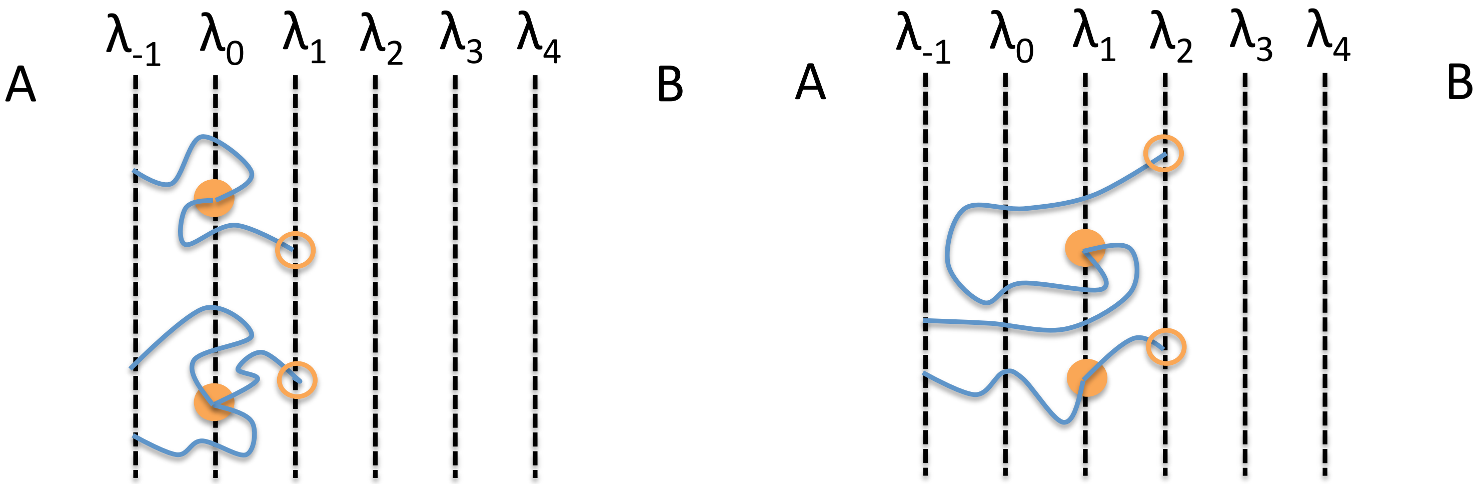


We then split up $p(\lambda_{\max}|\lambda_0)$ into stages.

$$p(\lambda_{\max}|\lambda_0) = \prod_{i=1}^{\max} p(\lambda_i|\lambda_{i-1}) \tag{127}$$

- $p(\lambda_i|\lambda_{i-1})$ is the probability that a trajectory will reach λ_i before returning to A given that it has reached λ_{i-1} for the first time since reaching A.

We use the states obtained from the previous stage to successively estimate $p(\lambda_i|\lambda_{i-1})$.

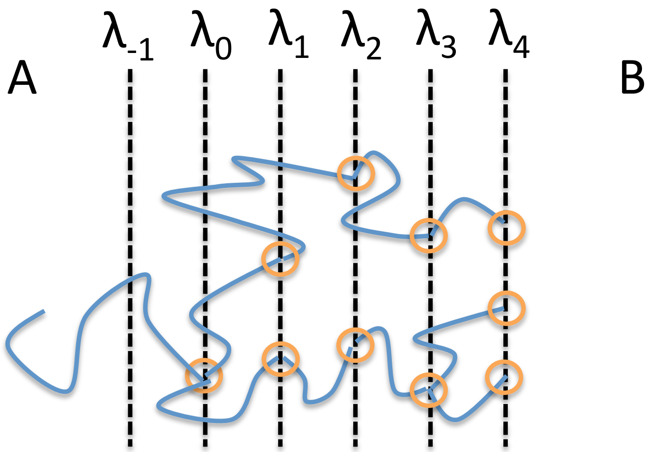


FFS results

The algorithm generates:

- A value for the overall flux.
- A value for the success rate from each interface.
- A set of (branched) trajectories sampled from the true set of transition paths.

By attempting to make progress from each point several times, trajectories are generated efficiently.



Reasons to be careful:

- Technically, you're interfaces don't need to be a particularly good choice for the reaction coordinate.
  - ▷ Interfaces must not intersect (this means you cannot assume a pathway).
  - ▷ A poor choice will tend to be inefficient.
  - ▷ It is possible to sample one type of pathway well without sampling another good alternative.
- Branched trajectories mean that sampling of later stages is more detailed than earlier stages.
  - ▷ 'Rosenbluth' methods exist to generate unbranched trajectories, involving a slightly more complicated analysis [23].
- FFS struggles when there are long-lived metastable intermediates.
  - ▷ In simple cases, one can do FFS to reach the metastable intermediate, then FFS to determine the flux out of it.
- FFS assumes the system is initially in equilibrium.
  - ▷ Non-stationary FFS allows this assumption to be lifted [24].
- FFS requires a stochastic component to the dynamics.

Transition Path Sampling

Generates an unbiased sample of trajectories that start in A at time $t = 0$ and end in B at time $t = \tau$.

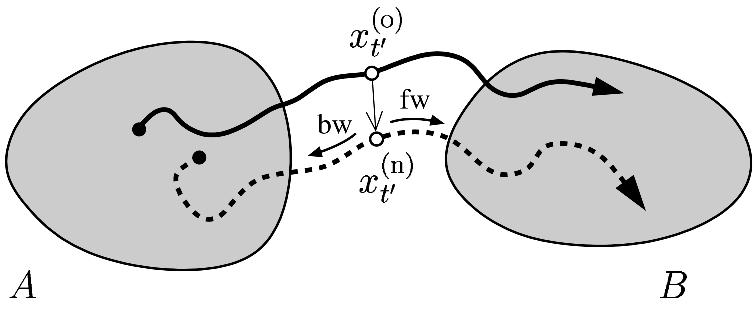
Algorithm performs a Monte Carlo sampling of these pathways.

- We initially generate a path of length τ between the two basins in whatever way we can.
    - ▷ For example, push the system half way and let it go.

We perform *shooting* and *shifting* moves to sample the trajectories that link the two basins.

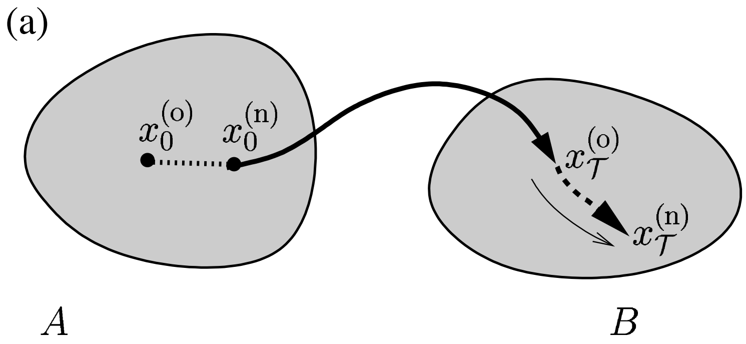
Shooting moves

- We select a random time along this trajectory from which to generate a new path.
- Generate new forward and backwards trajectories based on this point.
  - ▷ With deterministic dynamics, we must perturb the state a little.
  - ▷ With stochastic dynamics, we simply use a different random number seed (perturbing the state is possible, but introduces more acceptance factors).
- If the forward trajectory is in B at time τ, and the backward trajectory in A at time 0, we have a transition path.
- Accept this path with a factor that ensures the right statistical weight [25].
  - ▷ For Newtonian dynamics in the micro canonical ensemble (and symmetric trial perturbations), or stochastic dynamics in the canonical ensemble, if the system is initially assumed to equilibrated in A, the acceptance probability is 1. Other cases can be more complex.



Shifting moves

- We select a random (short) time δt.
- Define a new start point for a trajectory (t = 0) at δt along the old one.
    - ▷ Moves which shift in the opposite direction must also be included.
- Integrate the end point forward in time by δt to obtain the new end of the trajectory ($t = \tau$).
- If the forward trajectory is in B at time τ, and the backward trajectory in A at time 0, we have a transition path.
- Accept this path with a factor that ensures the right statistical weight [25].
    - ▷ In cases where the initial distribution is the steady-state, this factor is 1.



Results of TPS

The algorithm generates:

- A set of trajectories sampled from the true set of transition paths of length τ.

We didn't need an order parameter, only the need to distinguish state A and B.

However:

- ▷ We need to generate an initial trajectory.
- ▷ Calculate of reaction rates is tedious and requires an order parameter [23].
- ▷ The algorithm can suffer from ergodicity problems if you have very distinct pathways.
- ▷ TPS struggles with metastable intermediates just like FFS.

Transition Interface Sampling

TIS is a cross between FFS and TPS for evaluating reaction rates.

As with FFS, we wish to evaluate

$$\Phi_{AB} = \phi(\lambda_0) p(\lambda_{\max}|\lambda_0). \tag{128}$$

- ▷ $\phi(\lambda_0)$ is estimated in the same way as in FFS (through direct simulation).
- ▷ $p(\lambda_{\max}|\lambda_0)$ is split into stages, as in FFS

$$p(\lambda_{\max}|\lambda_0) = \prod_{i=1}^{\max} p(\lambda_i|\lambda_{i-1}). \tag{129}$$

- ▷ $p(\lambda_i|\lambda_{i-1})$ is calculated using a TPS approach.

To calculate $p(\lambda_i|\lambda_{i-1})$:

- For each interface $\boldsymbol{i}$, we begin with a path of length $\boldsymbol{\tau_o}$ that starts at $\boldsymbol{\lambda_0}$ and ends at $\boldsymbol{\lambda_i}$ or $\boldsymbol{\lambda_{-1}}$, having first gone through $\boldsymbol{\lambda_{i-1}}$.

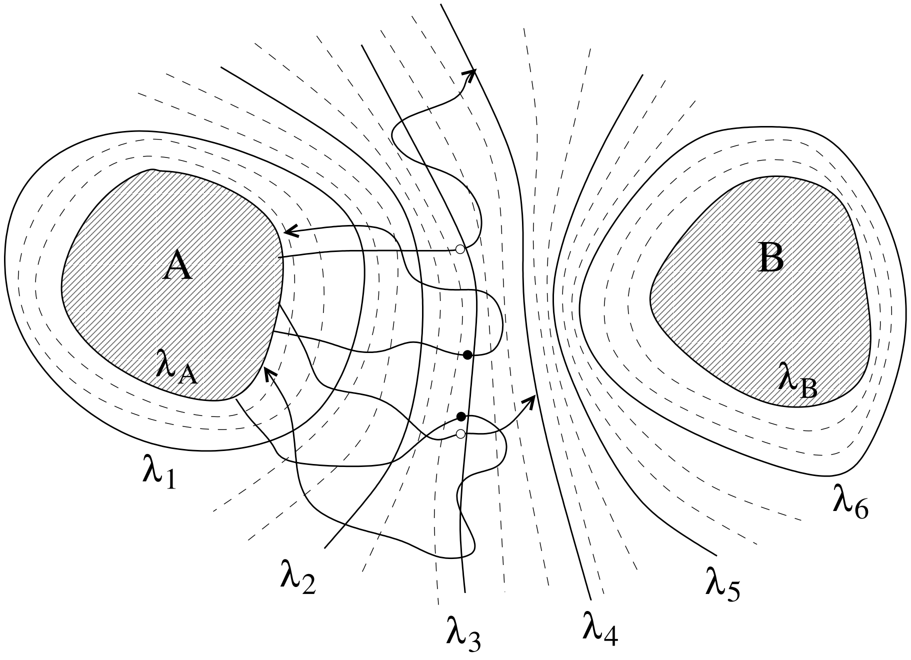
- Perform a shooting move – obtain a trajectory.

▷ Generate the new initial point (for stochastic systems, choose new seed).

▷ Generate new forward and backward trajectories. Backward trajectory ended when $\boldsymbol{\lambda_0}$ or $\boldsymbol{\lambda_i}$ reached, forwards when $\boldsymbol{\lambda_{-1}}$ or $\boldsymbol{\lambda_i}$ reached. Trajectory has length $\boldsymbol{\tau^n}$.

▷ Reject any trajectory that doesn't pass through $\boldsymbol{\lambda_{i-1}}$, and any for which the backward trajectory ends at $\boldsymbol{\lambda_i}$.

▷ For trajectories that pass through $\boldsymbol{\lambda_{i-1}}$, accept with a probability $\boldsymbol{\min\{1, \tau^o/\tau^n\}}$ [24]. Otherwise reject.



Results of TIS

We have ended up with:

- A representative set of trajectories for each interface i that start at λ_0, reach λ_{i-1} and then either proceed to λ_i or return to A.
- This allows direct evaluation of $p(\lambda_i|\lambda_{i-1})$. When combined with the initial flux, we have a total flux measurement.
- The successful trajectories at the final interface are representative reaction pathways.

However,

- ▷ We need to generate an initial trajectory for each interface.
- ▷ The algorithm can suffer from ergodicity problems if you have very distinct pathways.
- ▷ TIS struggles with metastable intermediates just like FFS and TPS.

Molecular Simulation Methods in Chemistry

Lecture 8:
Issues with simulation

Thomas Ouldridge

t.ouldridge1@physics.ox.ac.uk

http://www-thphys.physics.ox.ac.uk/people/ThomasOuldridge/Site/Teaching.html

Truncating the interactions

The most expensive part of any calculation is evaluating the energy and its derivatives.

For a pair potential, the cost of evaluating the potential is $\mathcal{O}(N^2)$, as all pairs of atoms should be formally considered.

However, the interactions between distant atoms for short-ranged interactions are often very small, and so the potential can be truncated.

If $v(r_c) \neq 0$ this leads to an error, but this can be approximately corrected for [25]:

$$\mathcal{V}\left(\mathbf{r}^N\right) = \sum_{r_{ij}<r_c} v(r_{ij}) + \frac{N\rho}{2}\int_{r_c}^{\infty} 4\pi r^2 v(r) \tag{130}$$

Jargon: this extra term is called a tail correction.

Tail correction does not converge for $v(r) \propto r^{-n}$ if $(n \leq 3)$. Need to use *Ewald summation*, which involves computing the long-range interactions in Fourier Space (where they converge more quickly), or other techniques [28].

Common ways to truncate the potential

- Simple truncation

$$v'(r) = \begin{cases} v(r) & r \leq r_c \\ 0 & r > rc \end{cases} \tag{131}$$

  MC: okay, but needs an additional impulsive correction for the pressure.

  MD: problems with discontinuity in the potential.

- Truncate and shift

$$v'(r) = \begin{cases} v(r) - v(r_c) & r \leq r_c \\ 0 & r > rc \end{cases} \tag{132}$$

  MD: still a discontinuity in the derivatives.

- Potential can also be ‘tapered’ to zero within a range of the cutoff.

Code scaling

Truncation reduces the number of interactions that need to be calculated from $\mathcal{O}(N^2)$ to $\mathcal{O}(mN)$, where m is the average number of neighbours of an atom within the cutoff distance.

However, $\mathcal{O}(N^2)$ distances still need to be checked at each step.

To reduce the overall scaling of the algorithm to $\mathcal{O}(N)$, two common methods [29]:

- Neighbour list
  Store a list of the particles that are within r_v of each atom, where $r_v > r_c$. Only calculate the interactions between particles on the neighbour lists. List needs to be updated when atoms have moved by $r_v - r_c$.

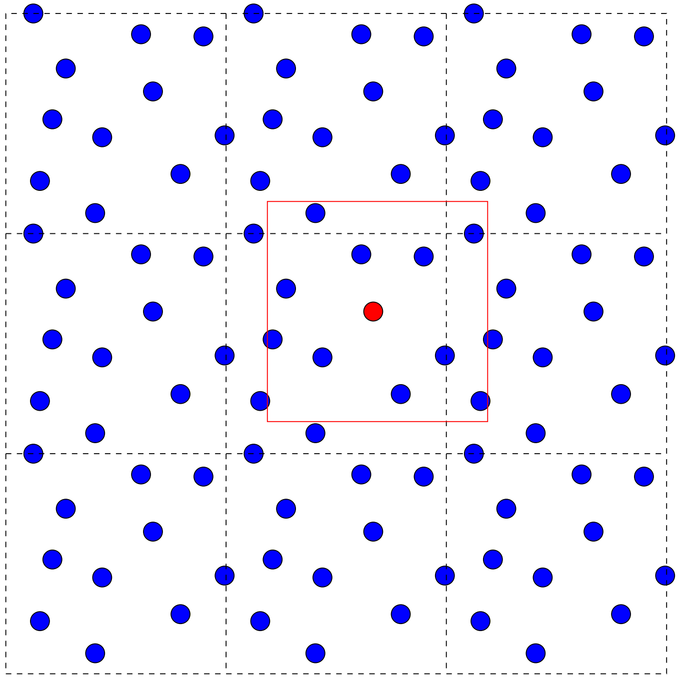
- Cell list
  Divide the simulation box up into cells. Only calculate the interactions between particles in the same or neighbouring cells.

Periodic boundary conditions

Normally, one wants to simulate a bulk system using a simulation of a finite number of particles.

For small systems, the fraction of particles at any surface is high, *e.g.* open or hard boundary conditions.

Therefore, to minimize such effects, periodic boundary conditions are used.



The simulation cell is treated as the primitive cell of an infinite periodic lattice of identical cells.

The simulation box is often chosen to be cubic, but other choices are possible.

$$\mathcal{V}_{tot} = \frac{1}{2} \sum_{i,j,\mathbf{n}} v\left(|\mathbf{r}_{ij} + \mathbf{n}L|\right), \tag{133}$$

where L is the edge length of the (cubic) box, and $\mathbf{n}$ is any vector of three integers.

Minimum image convention: an atom only interacts with the nearest image of any particle.

- ▷ If we truncate our interactions at a distance less than $L/2$, this is natural.

PBC comes with its own set of problems.

- Momentum is conserved in MD (and something similar in Nosé/Nosé-Hoover [30]) simulations.
  - ▷ This is a particular problem for Nosé/Nosé-Hoover simulations, which must be performed with zero total momentum [30]. Nosé-Hoover chains can improve this and other ergodic issues [31].
- Long-range interactions are still problematic.
- Calculation of the virial pressure can be invalid [32].
  - ▷ Still works fine for 2-body interactions with a cutoff-range smaller than half the simulation cell.
- ...

Such finite-size errors often scale as $\sim \frac{1}{\sqrt{N}}$ or $1/N$.

Finite-size scaling (i.e. examining how your property depends on N) can often be used to extrapolate to the bulk limit.

Error analysis

The variance in a quantity measured in an M step simulation is

$$\sigma_M^2(A) = \frac{1}{M}\sum_{i=1}^{M}\left(A_k - \langle A \rangle_M\right)^2 = \left\langle A^2 \right\rangle_M - \langle A \rangle_M^2 \tag{134}$$

If these measurements were independent, the variance in the average would be

$$\sigma^2\left(\langle A \rangle_M\right) \approx \frac{1}{M}\sigma_M^2 \tag{135}$$

However, the configurations generated by trajectories are *correlated*.

Auto-correlation functions can be used to measure how long it takes a system to lose "memory" of the state it was in.

$$C_{AA}(k) = \frac{1}{M}\sum_{k'}\left(A_{k'} - \langle A \rangle\right)\left(A_{k+k'} - \langle A \rangle\right) \tag{136}$$

Correlation functions typically show an exponential decay, i.e.

$$C_{AA}(k) \sim \exp\left(-k/n_\tau\right) \tag{137}$$

where $2n_\tau$ is a reasonable estimate of the number of steps between independent measurements.

$\therefore n_m = M/2n_\tau$ is a measure of the number of statistically independent measurements, and

$$\sigma^2\left(\langle A \rangle_M\right) = \frac{1}{n_M - 1}\sigma_M^2 \tag{138}$$

Block averages

An alternative approach is to use block averages.

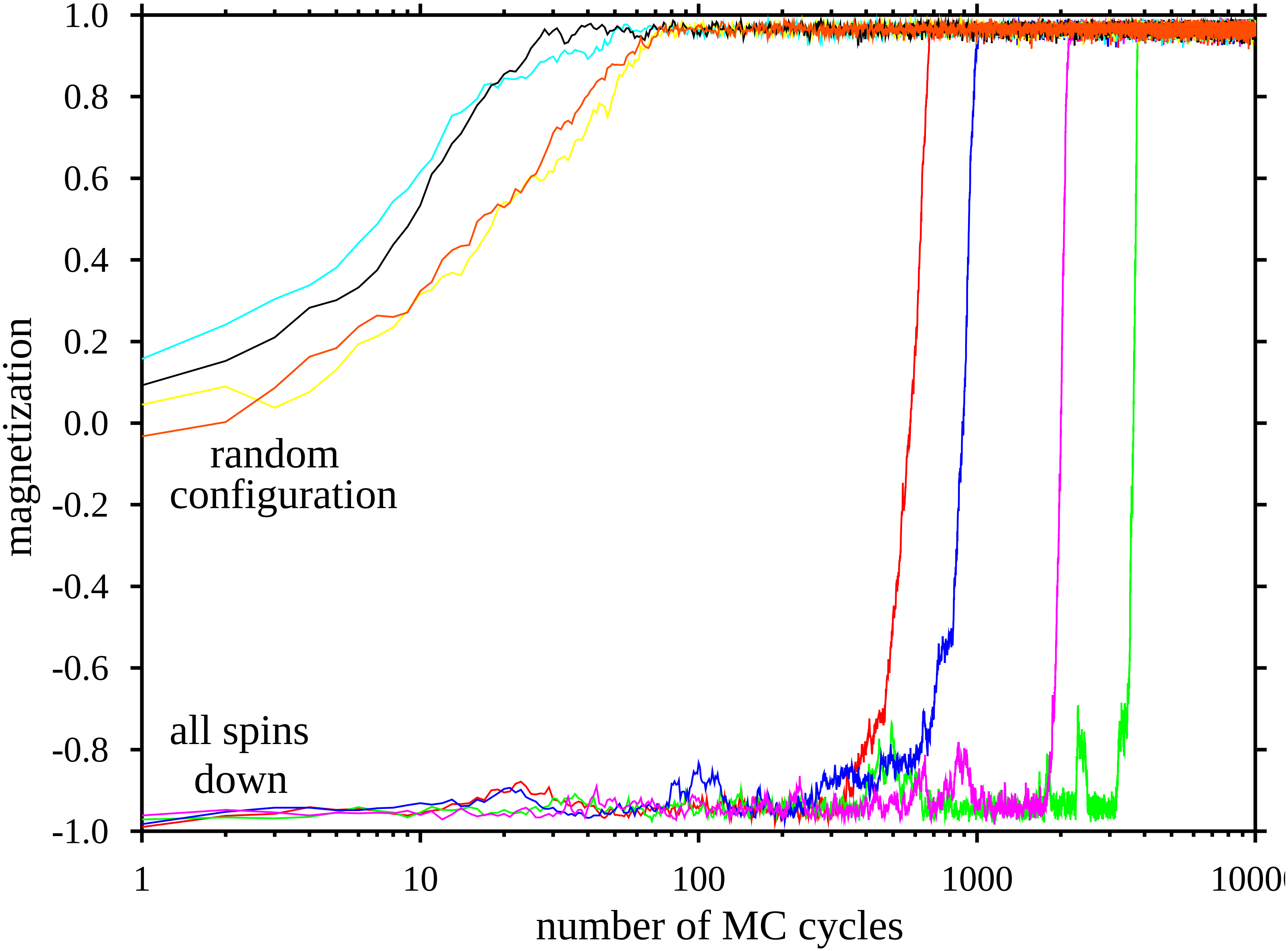
Take L partial averages $\langle A \rangle_l$ over blocks of length $l = M/L$ steps.

$$\sigma_L^2(\langle A \rangle_l) = \frac{1}{L} \sum_{i=1}^{L} (\langle A \rangle_l - \langle A \rangle_M)^2 \tag{139}$$

As the block size increases, eventually $l > 2n_\tau$ and the $\langle A \rangle_l$ become independent. $\therefore$

$$\frac{\sigma_L^2(\langle A \rangle_l)}{L - 1} \approx \sigma^2(\langle A \rangle_M) \tag{140}$$

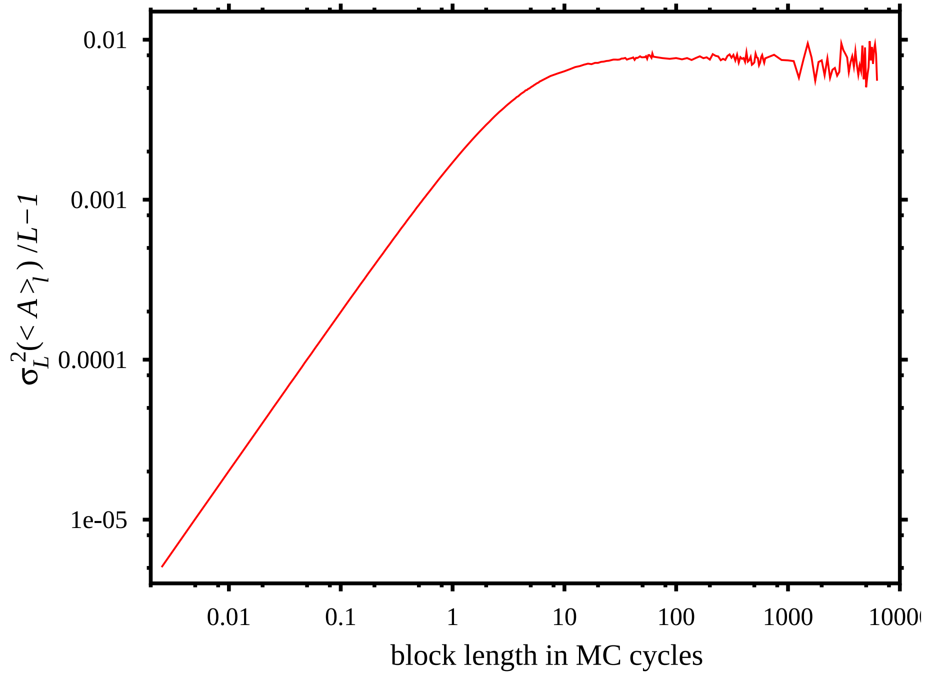
Plotting $\sigma_L^2(\langle A \rangle_l)/(L-1)$ as a function of L, this quantity will increase with decreasing L and plateau at the correct variance for $L \leq n_M$.

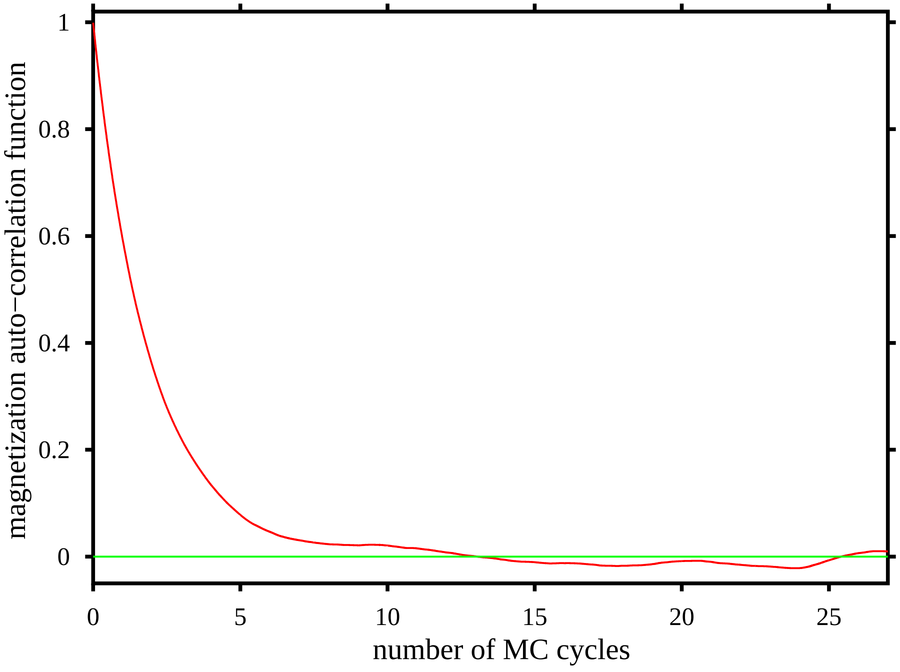

1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
magnetization
random
configuration
all spins
down
1
10
100
1000
1000
number of MC cycles

2D Ising model: 40×40 lattice, $kT/J = 1.8$.

Another alternative is to actually run N truly independent simulations, and calculate the variance between these results σ^2_{sim}. The error on your mean is then $\sigma_{\mathrm{sim}}/\sqrt{N-1}$.

- ▷ Systematic errors can arise from failure to reach equilibrium.





2D Ising model: 40×40 lattice, $kT/J = 1.8$, $H = 0.1$.

$$\mathcal{H} = -J \sum_{i,j} S_i S_j + H \sum_i S_i \tag{141}$$

A word on dynamics

Real liquid systems exhibit long-range *hydrodynamic* effects.

- ▷ In Molecular Dynamics simulations, objects will end up feeling their own flow field (either through reflection or PBC).

When Langevin and Andersen thermostats are used, they destroy momentum transport within the system.

- ▷ Cannot be used to calculate properties like viscosity and diffusion coefficients in systems.

Thermostats such as Dissipative Particle Dynamics [33] allow coarse-grained simulations to capture some long-range hydrodynamic effects.

Variable step size in Monte Carlo

The efficiency of a Monte Carlo algorithm is strongly dependent on the distribution of trial moves used.

It can be tempting to allow the maximum step size to vary

- ▷ increase move size if success rate is above 50% during the last n moves.
- ▷ decrease move size if success rate is below 50% during the last n moves.

This is OK for equilibration (when you'r not collecting data), but *the move generation algorithm must be constant during sampling*.

- ▷ Detailed balance is violated at the point at which the change is made.
- ▷ This violation can systematically differentiate between some states (particularly during self-assembly).

Random numbers

The library random number generator tends to be a bit rubbish.

- ▷ Lots of correlations, particularly for last few digits.

∴ Safest to use a more sophisticated one (*e.g.* from 'Numerical Recipes in ?).

Random number generators need seeding.

- ▷ Simple tutorials often suggest that you use 'time'. This can be problematic on clusters.

Use something else as a seed. Clusters often give each simulation a unique Job ID.

Random number generators can generate zero.

- ▷ This happens more often with worse algorithms.
- ▷ Simple implementations will generate an integer n between 0 and RAND_MAX.
- ▷ RAND_MAX can be as low as 32767 on C++ generators.

Any Monte Carlo move, no matter how crazy, will be accepted if you test the Boltzmann factor against $n/\mathrm{RAND\_MAX}$ with $n = 0$.

General tips

- Calculate things twice (in two different ways) if possible.
- Compare new code to old code, and original code to analytics.
- Don't take rules of thumb as gospel.
    - Success rates in MC should be 50%.
    - The ideal umbrella potential produces a flat landscape.
- Look at pictures and follow trajectories of your system.
- Get a feel for how a number in your code relates to a real physical system.
- Remember that free-energy landscapes $F(Q)$ depend on your definition of Q.
- Have function that outputs the state of your system in full precision, including the state of the random number generator.