Statistical methods in Atomistic Computer Simulations

1. Constant-temperature sampling and ergodicity

Michele Ceriotti

MSE 639 - EPFL - EDMX

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Outline

1. Sampling the canonical ensemble in atomistic simulations: Monte Carlo, correlation functions, ergodicity
2. Thermostatting MD efficiently: white and colored-noise Langevin dynamics
3. Rare events: transition state theory and the transition state ensemble
4. Statistical re-weighting, biased sampling, metadynamics
5. Linear and non-linear dimensionality reduction: determining automatically order parameters

Checklist of approximations

- Classical dynamics of a system of atoms with positions $\mathbf{q}$ and momenta $\mathbf{p}$, described by the Hamiltonian

$$H(\mathbf{p},\mathbf{q}) = K(\mathbf{p}) + V(\mathbf{q}) = \sum_i \frac{\mathbf{p}_i^2}{2m_i} + V(\mathbf{q})$$

  - We assume V is given (force fields, electronic structure calculation)
  - In all cases, strictly Born-Oppenheimer
  - Nuclei are treated as classical, distinguishable particles

- Want to generate $(\mathbf{p},\mathbf{q})$ configurations consistent with given thermodynamic conditions
  - We focus on the canonical (NVT) ensemble

$$P(\mathbf{p},\mathbf{q}) = e^{-\beta H(\mathbf{p},\mathbf{q})} \Big/ \int \mathrm{d}\mathbf{q}\mathrm{d}\mathbf{p}\, e^{-\beta H(\mathbf{p},\mathbf{q})}$$

Averages in the canonical ensemble

- The canonical ensemble describes a system in contact with a thermal reservoir at temperature $T = 1/k_B\beta$
- Averages of observables can be computed as integrals over the canonical Boltzmann distribution

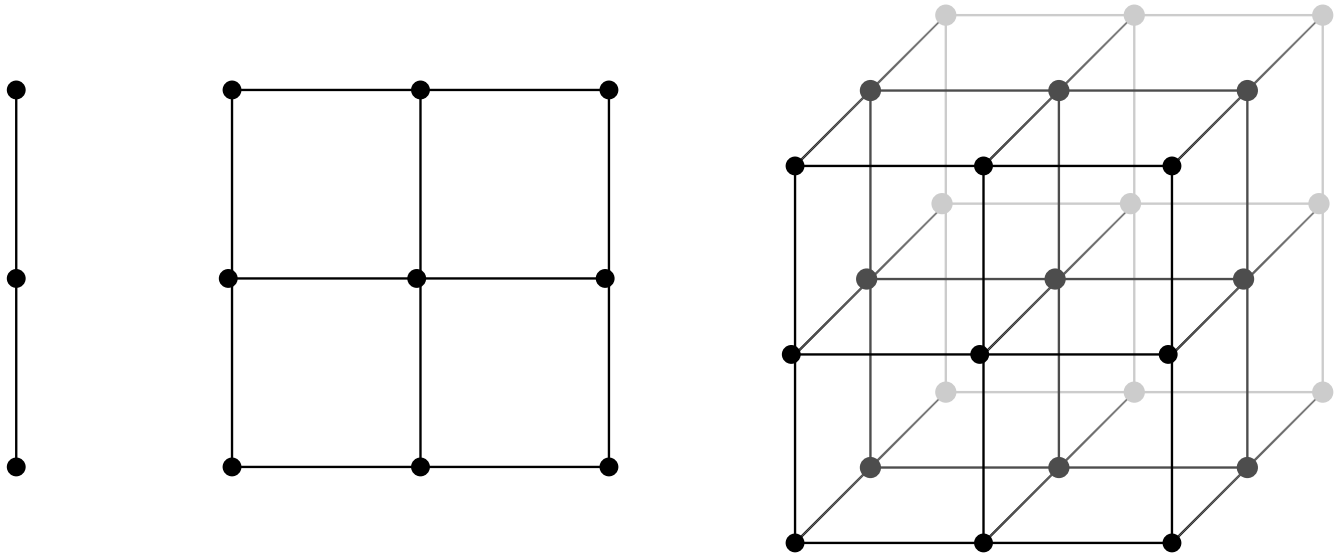
$$\langle A \rangle = \int \mathrm{d}\mathbf{q}\mathrm{d}\mathbf{p}\, P(\mathbf{p},\mathbf{q})\, A(\mathbf{p},\mathbf{q})$$

- The configuration and momentum parts of $P(\mathbf{p},\mathbf{q})$ factor out, and the momentum $P(\mathbf{p})$ is trivial

$$P(\mathbf{p},\mathbf{q}) = P(\mathbf{p}) \cdot P(\mathbf{q}) = \frac{e^{-\beta \sum_i \frac{\mathbf{p}_i^2}{2m_i}}}{\int \mathrm{d}^{3N}\mathbf{p} e^{-\beta \sum_i \frac{\mathbf{p}_i^2}{2m_i}}} \cdot \frac{e^{-\beta V(\mathbf{q})}}{\int \mathrm{d}^{3N}\mathbf{q} e^{-\beta V(\mathbf{q})}}$$

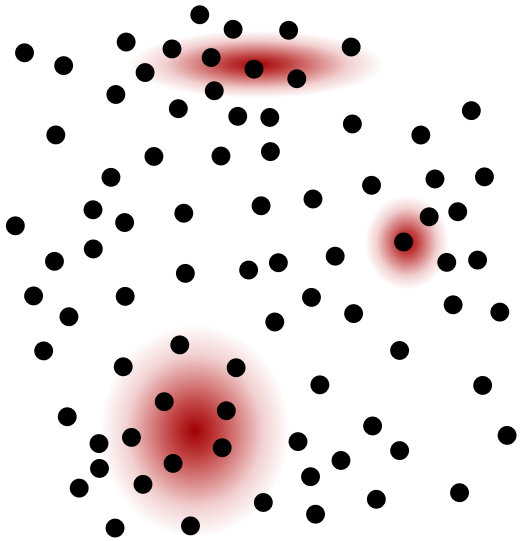
Integration by importance sampling

- Performing the integral by quadrature is impractical, because the number of grid points grows exponentially with dimensionality
- One could think of using randomly-distributed points $\mathbf{q}_i$, and computing $\langle A \rangle \approx \sum_i P(\mathbf{q}_i) A(\mathbf{q}_i) / \sum_i P(\mathbf{q}_i)$. But $P(\mathbf{q})$ is often very irregular and most points do not contribute to the integral
- Ideally, one would want to generate $\mathbf{q}_i$'s distributed as $P(\mathbf{q})$, and compute $\langle A \rangle \approx \frac{1}{M} \sum_i A(\mathbf{q}_i)$. This is **importance sampling**.



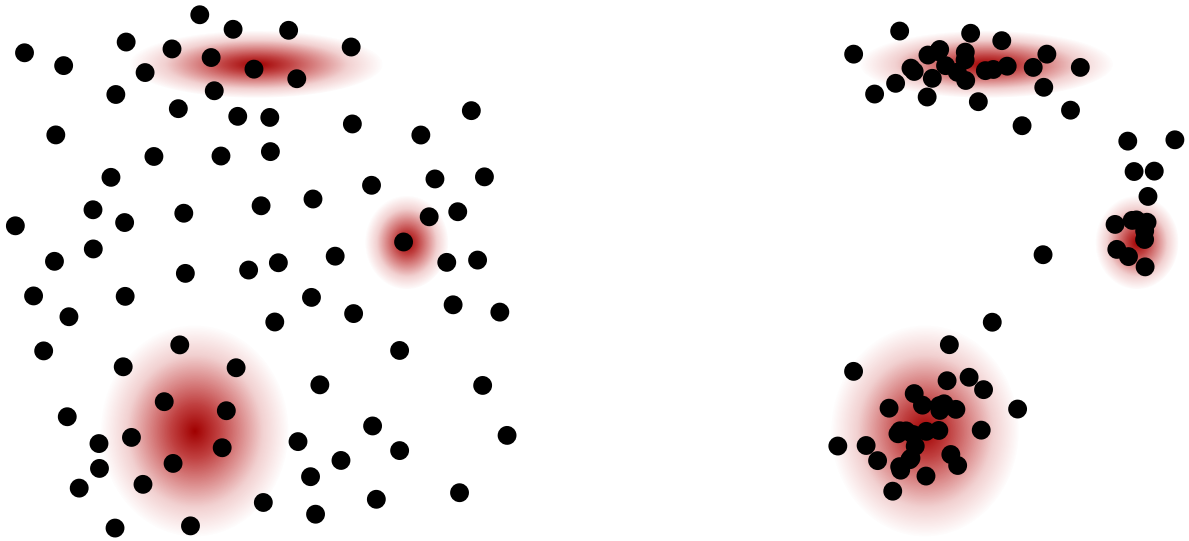
Integration by importance sampling

- Performing the integral by quadrature is impractical, because the number of grid points grows exponentially with dimensionality
- One could think of using randomly-distributed points $\mathbf{q}_i$, and computing $\langle A \rangle \approx \sum_i P(\mathbf{q}_i) A(\mathbf{q}_i) / \sum_i P(\mathbf{q}_i)$. But $P(\mathbf{q})$ is often very irregular and most points do not contribute to the integral
- Ideally, one would want to generate $\mathbf{q}_i$'s distributed as $P(\mathbf{q})$, and compute $\langle A \rangle \approx \frac{1}{M} \sum_i A(\mathbf{q}_i)$. This is **importance sampling**.



Integration by importance sampling

- Performing the integral by quadrature is impractical, because the number of grid points grows exponentially with dimensionality
- One could think of using randomly-distributed points $\mathbf{q}_i$, and computing $\langle A \rangle \approx \sum_i P(\mathbf{q}_i) A(\mathbf{q}_i) / \sum_i P(\mathbf{q}_i)$. But $P(\mathbf{q})$ is often very irregular and most points do not contribute to the integral
- Ideally, one would want to generate $\mathbf{q}_i$'s distributed as $P(\mathbf{q})$, and compute $\langle A \rangle \approx \frac{1}{M} \sum_i A(\mathbf{q}_i)$. This is **importance sampling**.



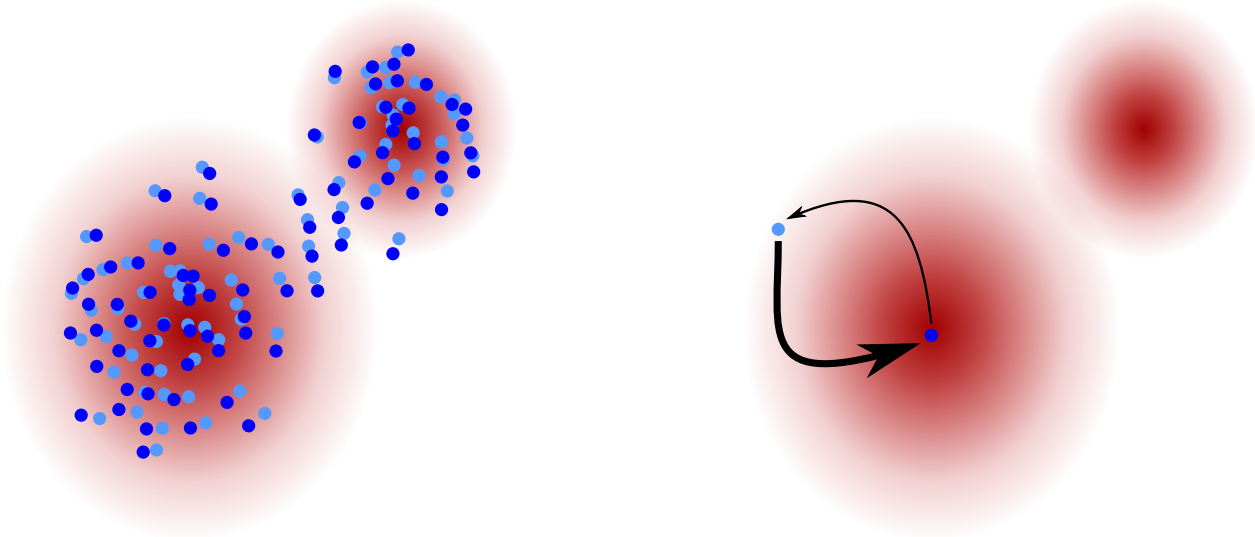
Using Monte Carlo for importance sampling

- How to generate a sequence of points distributed as $P(\mathbf{q})$?
  - Monte Carlo moves $\mathbf{q} \to \mathbf{q}'$ that generate a new point given a previous point consistent with P.
  - The move (typically) contains some randomness: probability $p(\mathbf{q} \to \mathbf{q}')$
- **Necessary** (& "sufficient") condition: P invariant under the move

$$\int \mathrm{d}\mathbf{q}\, P(\mathbf{q})\, p(\mathbf{q} \to \mathbf{q}') = P(\mathbf{q}')$$

- **Sufficient** condition: **detailed balance**

$$P(\mathbf{q})\, p(\mathbf{q} \to \mathbf{q}') = P(\mathbf{q}')\, p(\mathbf{q}' \to \mathbf{q})$$



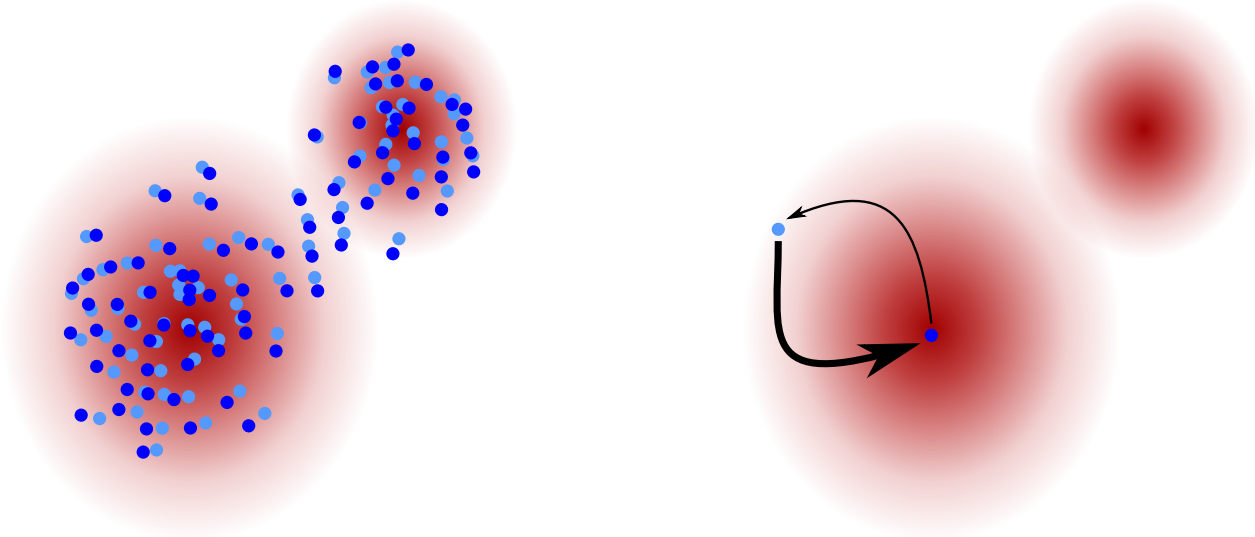
Using Monte Carlo for importance sampling

- How to generate a sequence of points distributed as $P(\mathbf{q})$?
  - Monte Carlo moves $\mathbf{q} \to \mathbf{q}'$ that generate a new point given a previous point consistent with P.
  - The move (typically) contains some randomness: probability $p(\mathbf{q} \to \mathbf{q}')$
- **Necessary** (& "sufficient") condition: P invariant under the move

$$\int \mathrm{d}\mathbf{q}\, P(\mathbf{q})\, p(\mathbf{q} \to \mathbf{q}') = P(\mathbf{q}')$$

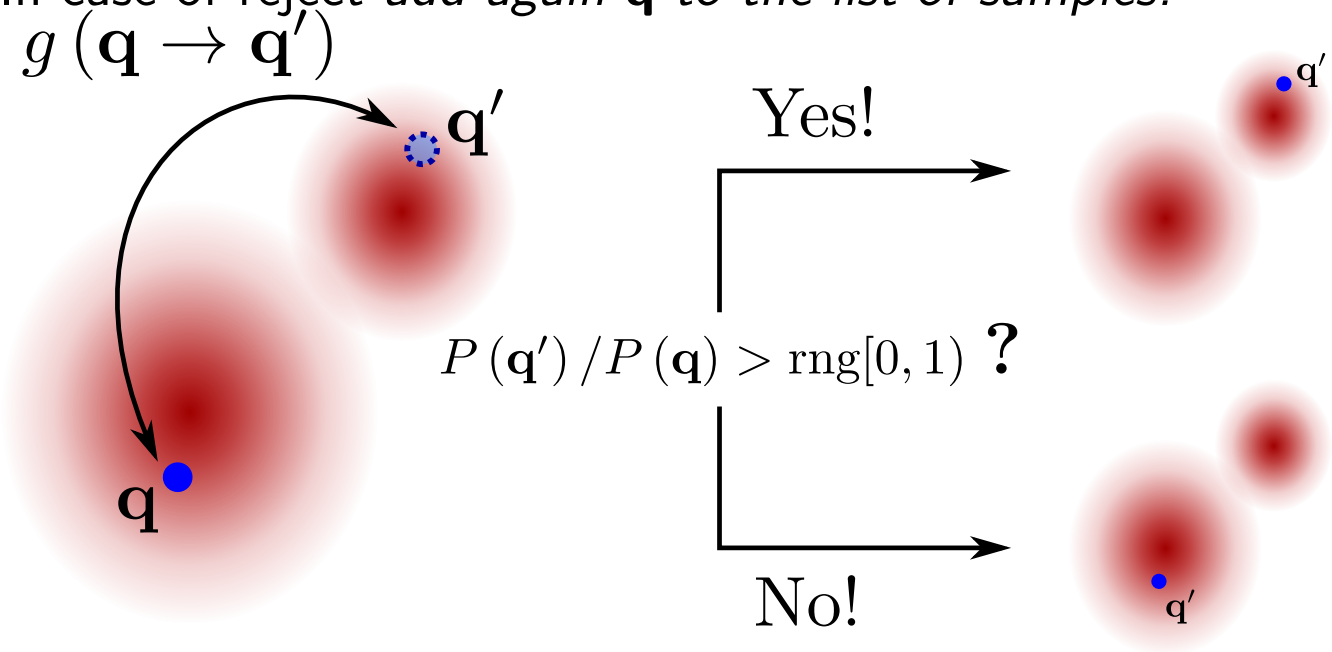
- **Sufficient** condition: **detailed balance**

$$p(\mathbf{q} \to \mathbf{q}') / p(\mathbf{q}' \to \mathbf{q}) = P(\mathbf{q}') / P(\mathbf{q}) = e^{-\beta[V(\mathbf{q}')-V(\mathbf{q})]}$$



Accept/reject: Metropolis moves

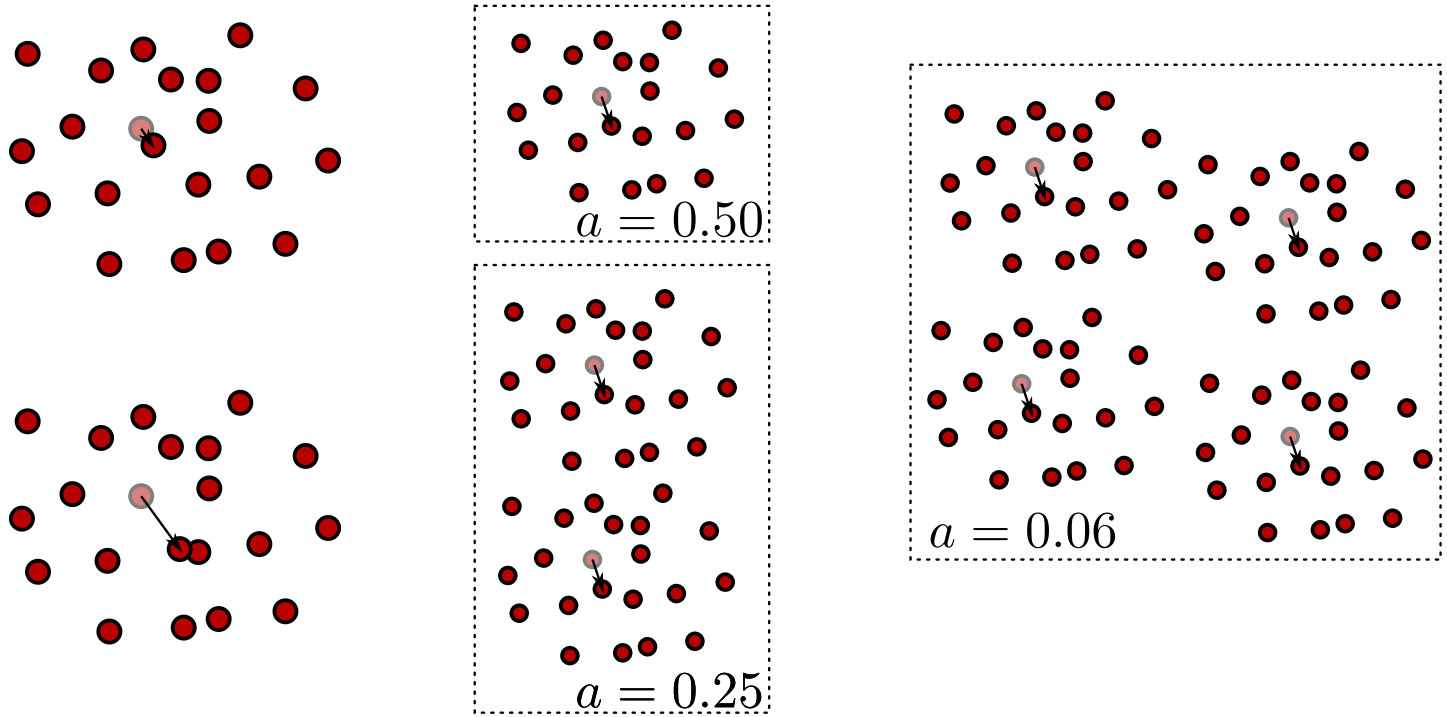
- A simple way to generate moves consistent with detailed balance. Split the move in two steps: generation/acceptance
  $p(\mathbf{q} \to \mathbf{q}') = g(\mathbf{q} \to \mathbf{q}')\, a(\mathbf{q} \to \mathbf{q}')$
  - Generate a new configuration with a criterion that is symmetric:
    $g(\mathbf{q} \to \mathbf{q}') = g(\mathbf{q}' \to \mathbf{q})$
  - Accept the new configuration with probability
    $a(\mathbf{q} \to \mathbf{q}') = \min\left[1, P(\mathbf{q}')/P(\mathbf{q})\right] = \min\left[1, e^{-\beta\left[V(\mathbf{q}')-V(\mathbf{q})\right]}\right].$
  - In case of reject *add again* $\mathbf{q}$ *to the list of samples!*



Metropolis et al., J. Chem. Phys 1953

Problems with Monte Carlo

- Must find a balance between having high acceptance, and making things happen!
- Fundamental problem: size-extensive changes in energy upon "global" moves reduce dramatically the acceptance
- One should develop "smart" moves, hard to generalize to arbitrary problems



Sampling by molecular dynamics

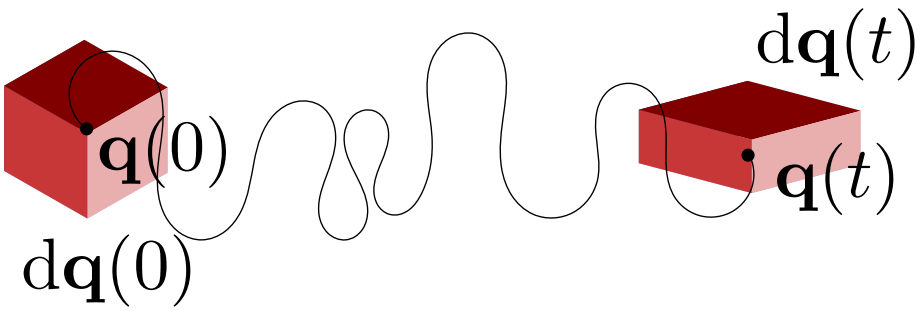
- An alternative: use Hamilton's equations to generate a trajectory

$$\dot{\mathbf{p}} = -\partial V/\partial \mathbf{q}, \qquad \dot{\mathbf{q}} = \mathbf{p}/m$$

- Does not obey detailed balance (but a generalization of it).
- Does fulfill the *necessary*, "global" condition: could prove cleanly with a Liouville operator formalism. The ingredients are:
  - Conservation of probability density

$$\frac{dP}{dt} \propto e^{-\beta H}\frac{dH}{dt}, \qquad \frac{dH}{dt} = \frac{\partial H}{\partial \mathbf{p}}\cdot\dot{\mathbf{p}} + \frac{\partial H}{\partial \mathbf{q}}\cdot\dot{\mathbf{q}} = -\frac{\partial H}{\partial \mathbf{p}}\cdot\frac{\partial H}{\partial \mathbf{q}} + \frac{\partial H}{\partial \mathbf{q}}\cdot\frac{\partial H}{\partial \mathbf{p}} = 0$$

  - Conservation of phase-space volume (symplectic property)



Integrating the equations of motion

- MD requires discrete time step dt integration of Hamilton's equations
- Many integrators exist. My favourite: symmetric-split velocity Verlet
  - Time-reversible, **symplectic**, simple implementation, easy to extend
  - Monitor total energy fluctuations and drift to get a grip of the integration error

$$\mathbf{p} \leftarrow \mathbf{p} - \frac{\partial V}{\partial \mathbf{q}} \frac{dt}{2}$$

$$\mathbf{q} \leftarrow \mathbf{q} + \frac{\mathbf{p}}{m} dt$$

$$\mathbf{p} \leftarrow \mathbf{p} - \frac{\partial V}{\partial \mathbf{q}} \frac{dt}{2}$$

- Integration error leads to (small) departure from the (micro)canonical ensemble, but no problem with system size: MD assumes the error is split evenly across the system

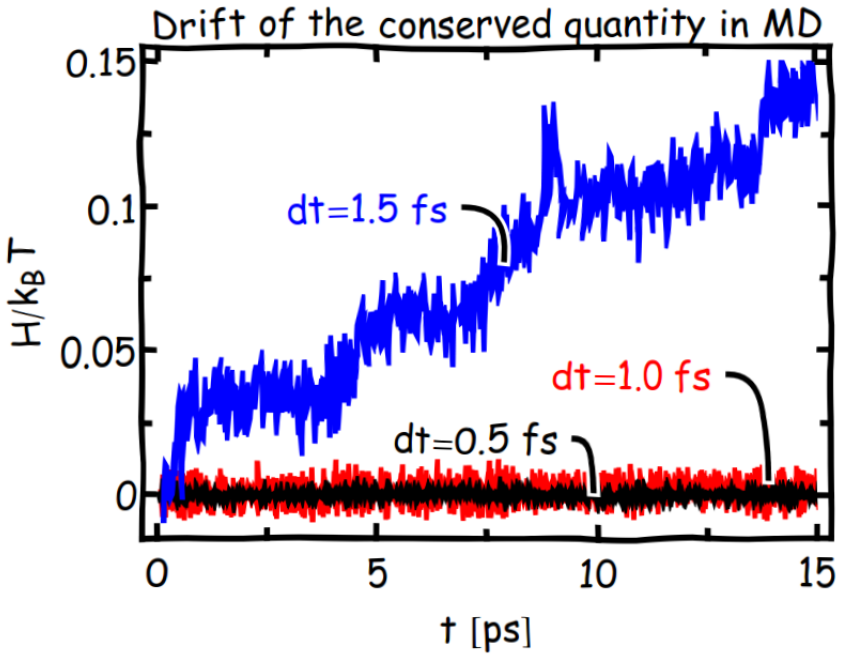
Integrating the equations of motion

- MD requires discrete time step dt integration of Hamilton's equations
- Many integrators exist. My favourite: symmetric-split velocity Verlet
  - Time-reversible, **symplectic**, simple implementation, easy to extend
  - Monitor total energy fluctuations and drift to get a grip of the integration error

$$\mathbf{p} \leftarrow \mathbf{p} - \frac{\partial V}{\partial \mathbf{q}} \frac{dt}{2}$$

$$\mathbf{q} \leftarrow \mathbf{q} + \frac{\mathbf{p}}{m} dt$$

$$\mathbf{p} \leftarrow \mathbf{p} - \frac{\partial V}{\partial \mathbf{q}} \frac{dt}{2}$$



- Integration error leads to (small) departure from the (micro)canonical ensemble, but no problem with system size: MD assumes the error is split evenly across the system

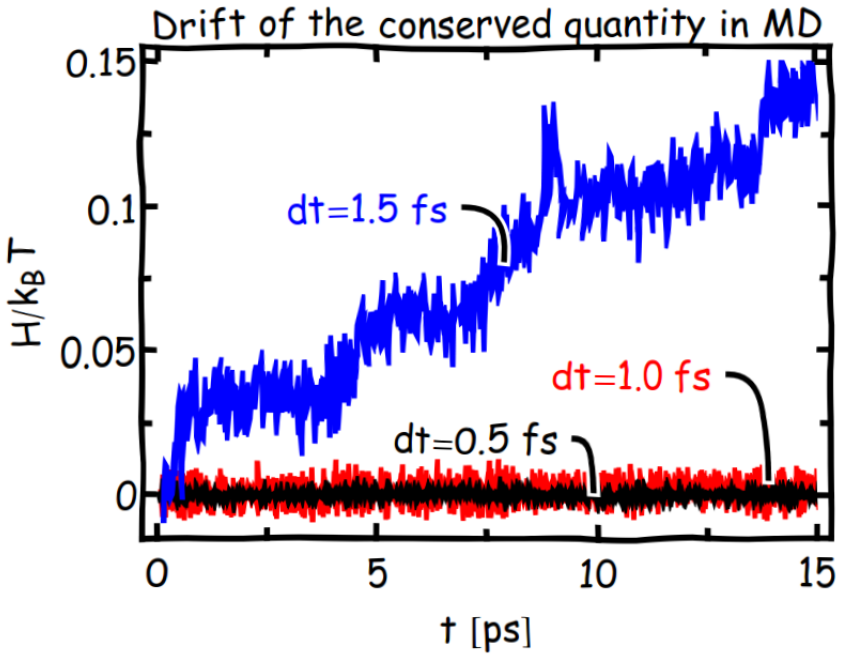
Integrating the equations of motion

- MD requires discrete time step dt integration of Hamilton's equations
- Many integrators exist. My favourite: symmetric-split velocity Verlet
  - Time-reversible, **symplectic**, simple implementation, easy to extend
  - Monitor total energy fluctuations and drift to get a grip of the integration error

$$\mathbf{p} \leftarrow \mathbf{p} - \frac{\partial V}{\partial \mathbf{q}} \frac{dt}{2}$$

$$\mathbf{q} \leftarrow \mathbf{q} + \frac{\mathbf{p}}{m} dt$$

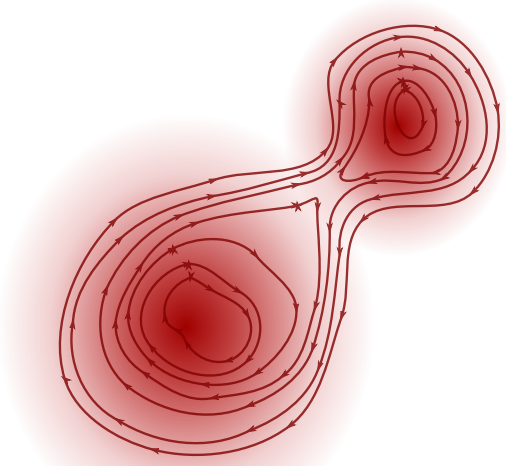
$$\mathbf{p} \leftarrow \mathbf{p} - \frac{\partial V}{\partial \mathbf{q}} \frac{dt}{2}$$



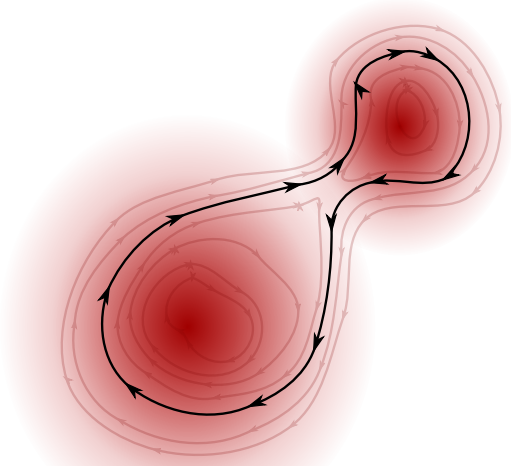
- Integration error leads to (small) departure from the (micro)canonical ensemble, but no problem with system size: MD assumes the error is split evenly across the system

Conserved energy in Hamiltonian MD

- The probability distribution is conserved *globally*, i.e. if we started many trajectories from the canonical ensemble, we would sample the whole of it
- However, individual trajectories conserve the total energy $H(\mathbf{p}, \mathbf{q})$, and so only sample partially the canonical distribution
- We have an **ergodicity** problem with MD (solution: next lecture!)



$dP/dt = 0$



$dH/dt = 0$

Ergodicity of sampling

- Underlying assumption behind all of this is **ergodicity**:

$$\lim_{M\to\infty} \frac{1}{M} \sum_i A(\mathbf{q}_i) = \int \mathrm{d}\mathbf{q}\, P(\mathbf{q})\, A(\mathbf{q})$$

  - This is an additional condition, that is very hard to prove in practice.
  - We always have a finite M, so the question is: *how much* ergodic is the trajectory? How can we "measure" ergodicity?
  - As we will see, this is closely related to the statistical error on $\langle A \rangle$

- The key tool to assess the ergodicity of a trajectory for computing $\langle A \rangle$ is the autocorrelation function

$$c_{AA}(t) = \langle A(0)\, A(t) \rangle$$

Introducing autocorrelation functions

- One can also define a version for discrete series,

$$c_{AA}(i) = \langle A(\mathbf{q}_0) A(\mathbf{q}_i) \rangle$$

- It is written $\langle A(0) A(t) \rangle$, but it really means

$$c_{AA}(t) = \frac{\langle (A(0) - \langle A \rangle)(A(t) - \langle A \rangle) \rangle}{\langle A^2 \rangle - \langle A \rangle^2}$$

. . . and it *really* means

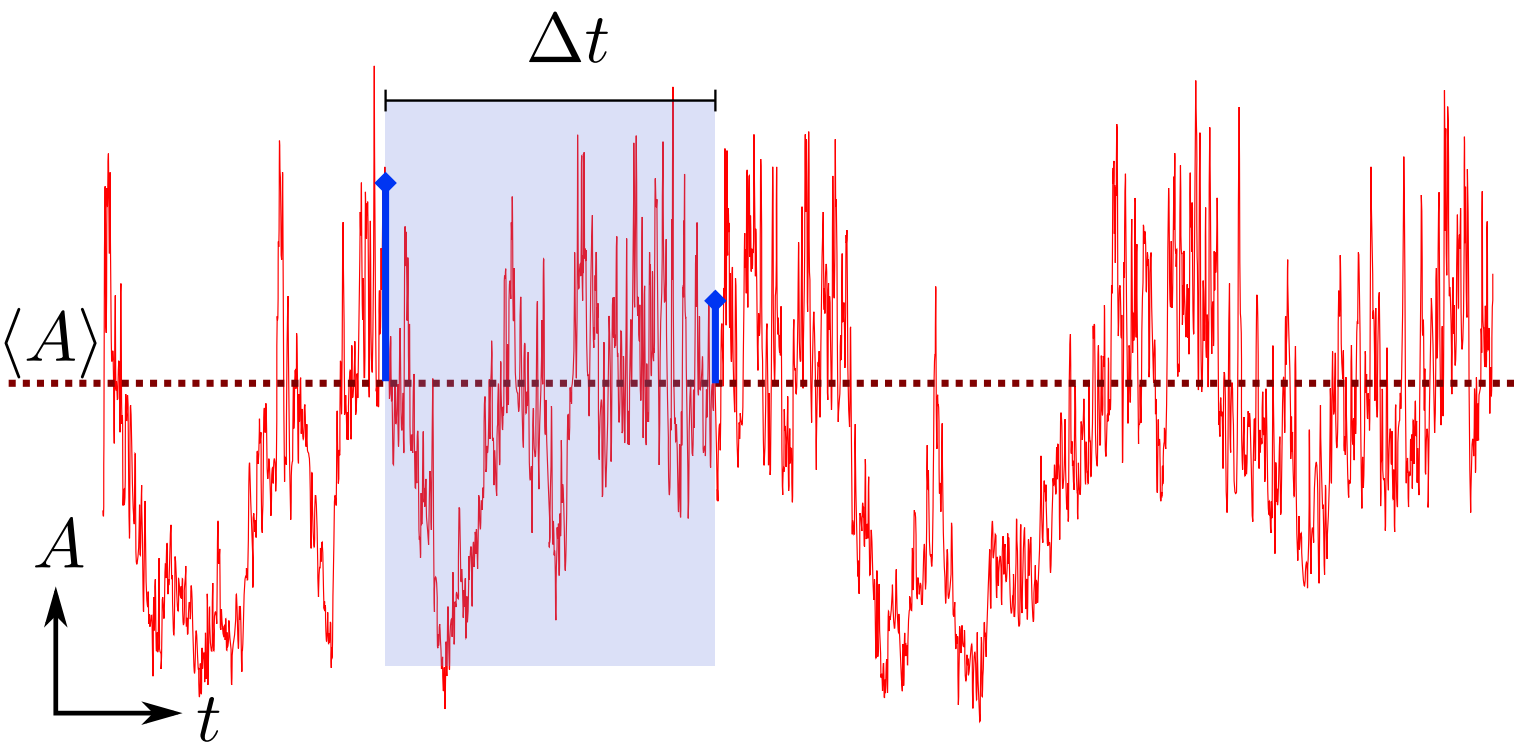
$$c_{AA}(t) = \frac{\int \mathrm{d}\mathbf{q}\mathrm{d}\mathbf{q}' P(\mathbf{q}) P(\mathbf{q}'; t|\mathbf{q}; 0) [A(\mathbf{q}) - \langle A \rangle][A(\mathbf{q}') - \langle A \rangle]}{\int \mathrm{d}\mathbf{q} P(\mathbf{q}) [A(\mathbf{q}) - \langle A \rangle]^2}$$

. . . and in practice it is computed from a trajectory $A(t)$ as

$$c_{AA}(t) = \frac{1}{\sigma_A^2} \lim_{T\to\infty} \frac{1}{T-t} \int_0^{T-t} \mathrm{d}t' \left[A(t') - \langle A \rangle\right] \left[A(t'+t) - \langle A \rangle\right].$$

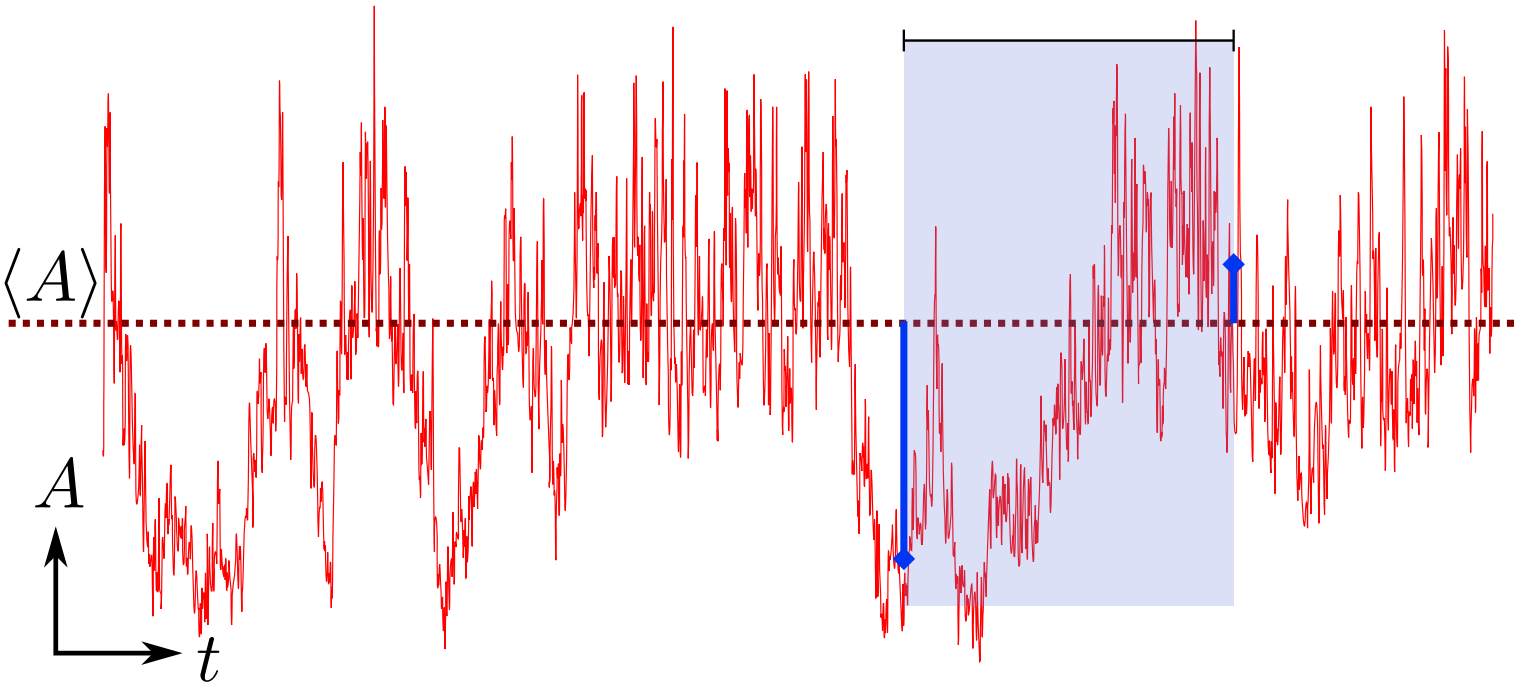
Autocorrelation function from a trajectory

$$c_{AA}(t) = \frac{1}{\sigma_A^2} \lim_{T\to\infty} \frac{1}{T-t} \int_0^{T-t} \mathrm{d}t' \left[A(t') - \langle A \rangle\right] \left[A(t'+t) - \langle A \rangle\right]$$



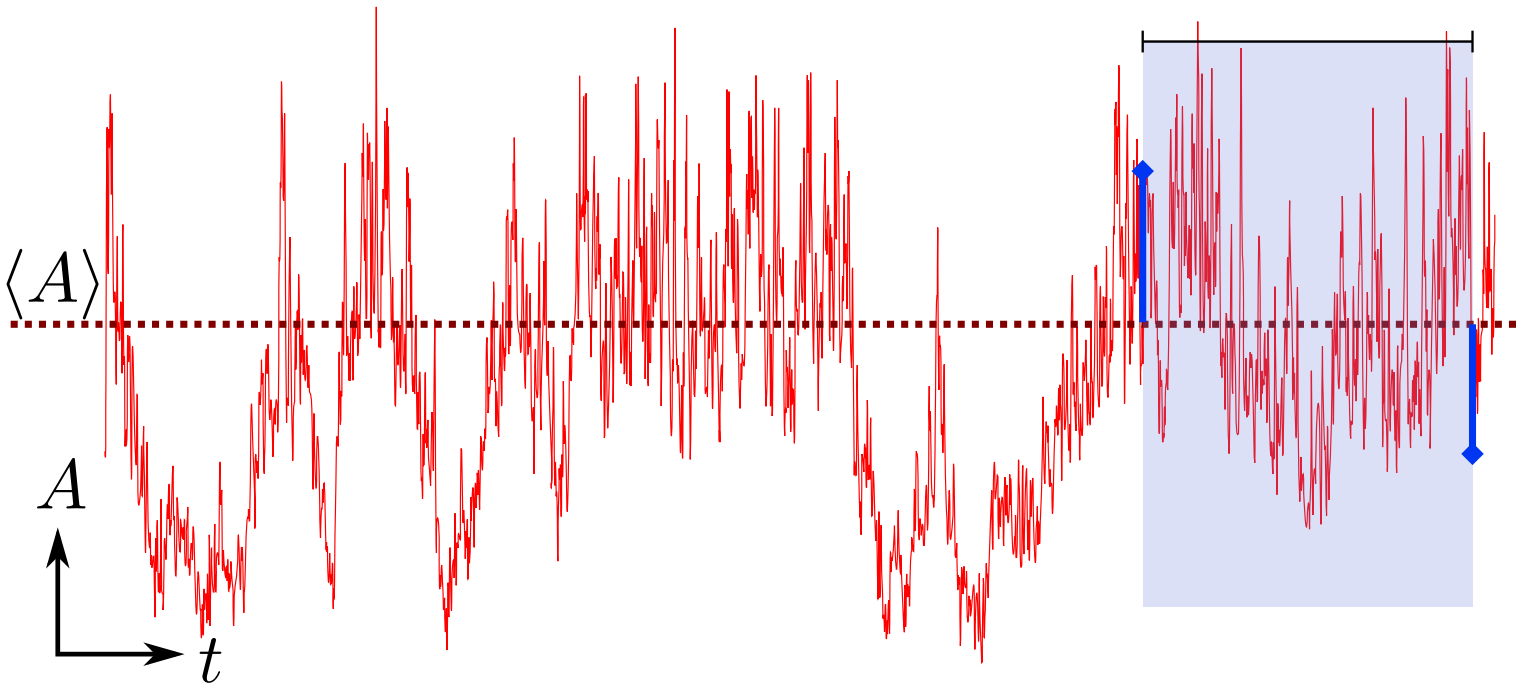
Autocorrelation function from a trajectory

$$c_{AA}(t) = \frac{1}{\sigma_A^2} \lim_{T\to\infty} \frac{1}{T-t} \int_0^{T-t} \mathrm{d}t' \left[A(t') - \langle A \rangle\right] \left[A(t'+t) - \langle A \rangle\right]$$



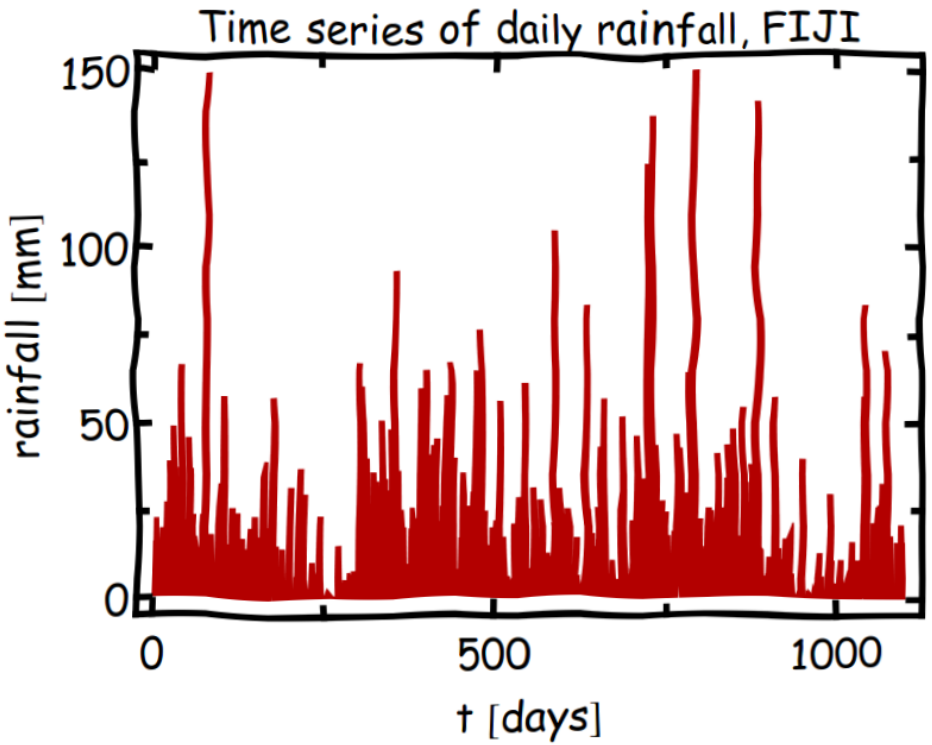
Autocorrelation function from a trajectory

$$c_{AA}(t) = \frac{1}{\sigma_A^2} \lim_{T\to\infty} \frac{1}{T-t} \int_0^{T-t} \mathrm{d}t' \left[A(t') - \langle A \rangle\right] \left[A(t'+t) - \langle A \rangle\right]$$



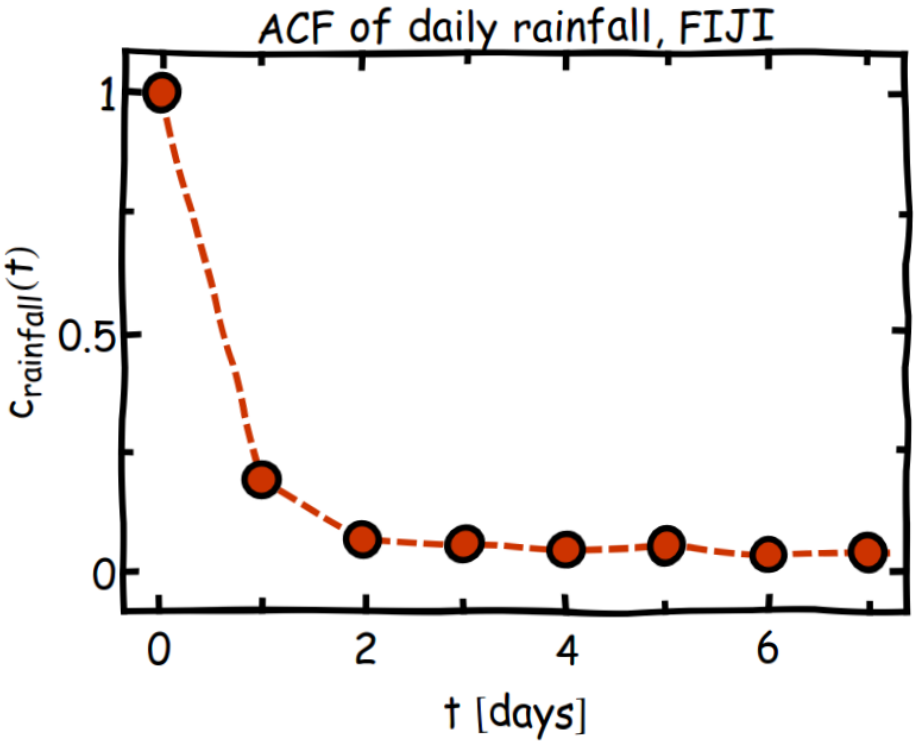
If it rained yesterday...

- $c_{AA}(t)$ tells you how much one can infer about the deviation of A from its mean at time t from its deviation at time 0.
- Example: daily rain fall on a small island in the Pacific: will it rain tomorrow? Multiple time scales can be identified.



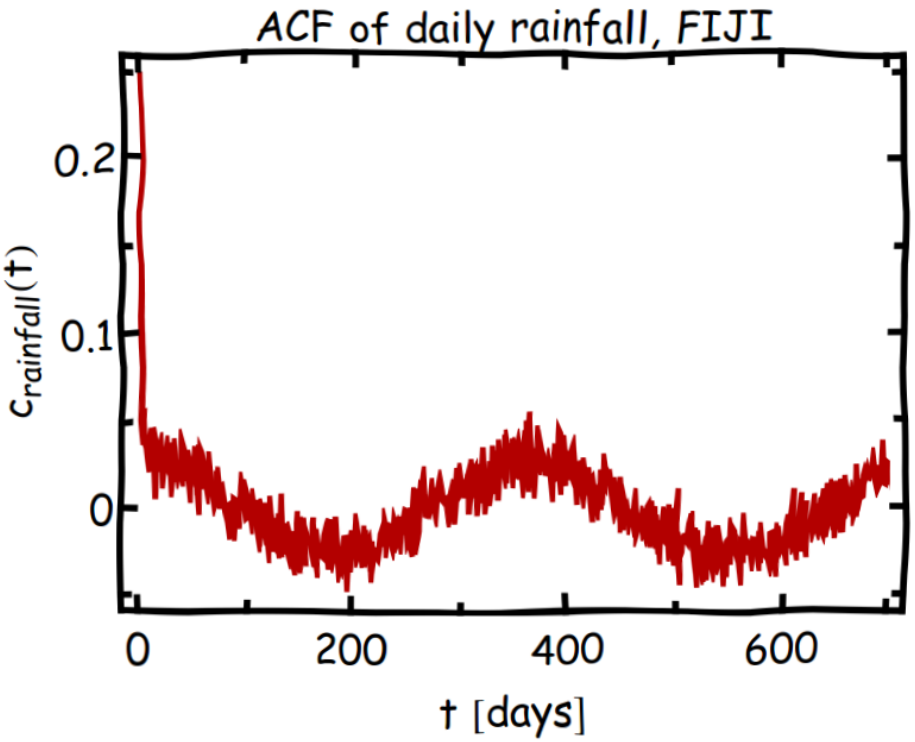
If it rained yesterday...

- $c_{AA}(t)$ tells you how much one can infer about the deviation of A from its mean at time t from its deviation at time 0.
- Example: daily rain fall on a small island in the Pacific: will it rain tomorrow? Multiple time scales can be identified.



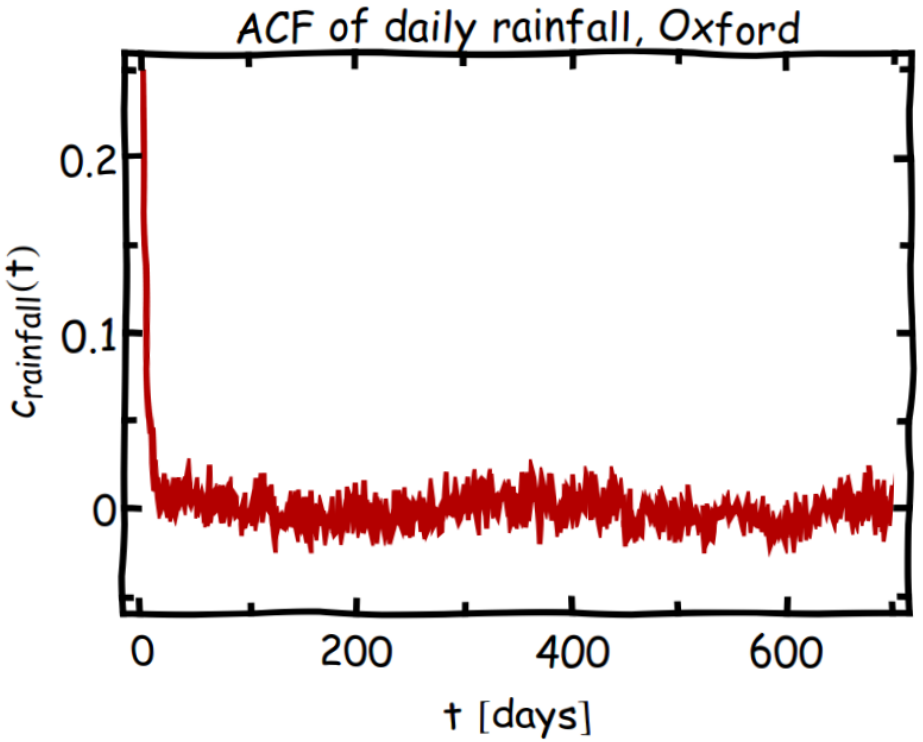
If it rained yesterday...

- $c_{AA}(t)$ tells you how much one can infer about the deviation of A from its mean at time t from its deviation at time 0.
- Example: daily rain fall on a small island in the Pacific: will it rain tomorrow? Multiple time scales can be identified.



If it rained yesterday...

- $c_{AA}(t)$ tells you how much one can infer about the deviation of A from its mean at time t from its deviation at time 0.
- Example: daily rain fall on a small island in the Pacific: will it rain tomorrow? Multiple time scales can be identified.



ACF and error in the mean

- Autocorrelation functions are closely related to the sampling efficiency
- The mean A computed from a run of length T is

$$\langle A \rangle_T = \frac{1}{T} \int_0^T A(t)\,\mathrm{d}t$$

- The mean square error in the mean $\epsilon_A^2(T)$ is then

$$\left\langle \left[\langle A \rangle_T - \langle A \rangle \right]^2 \right\rangle = \frac{1}{T^2} \int_0^T \mathrm{d}t \int_0^T \mathrm{d}t' \left\langle \left[A(t) - \langle A \rangle \right] \left[A(t') - \langle A \rangle \right] \right\rangle .$$

Noting that

$$\left\langle \left[A(t) - \langle A \rangle \right] \left[A(t') - \langle A \rangle \right] \right\rangle = \sigma_A^2 c_{AA}(t' - t) ,$$

one can do a change of variables $t' - t = \Delta$ and get

$$\epsilon_A^2(T) = \frac{\sigma_A^2}{T} \int_{-T}^T \mathrm{d}\Delta\, c_{AA}(\Delta) \left(1 - \frac{|\Delta|}{T} \right)$$

Correlation time and statistical efficiency

- For T much longer than the time scale on which $c_{AA}(t)$ decays to zero, $\epsilon_A^2(T) = \frac{\sigma_A^2}{T/\tau_A}$, where we introduced the **correlation time**

$$\tau_A = \int_{-\infty}^{\infty} c_{AA}(t)\,\mathrm{d}t$$

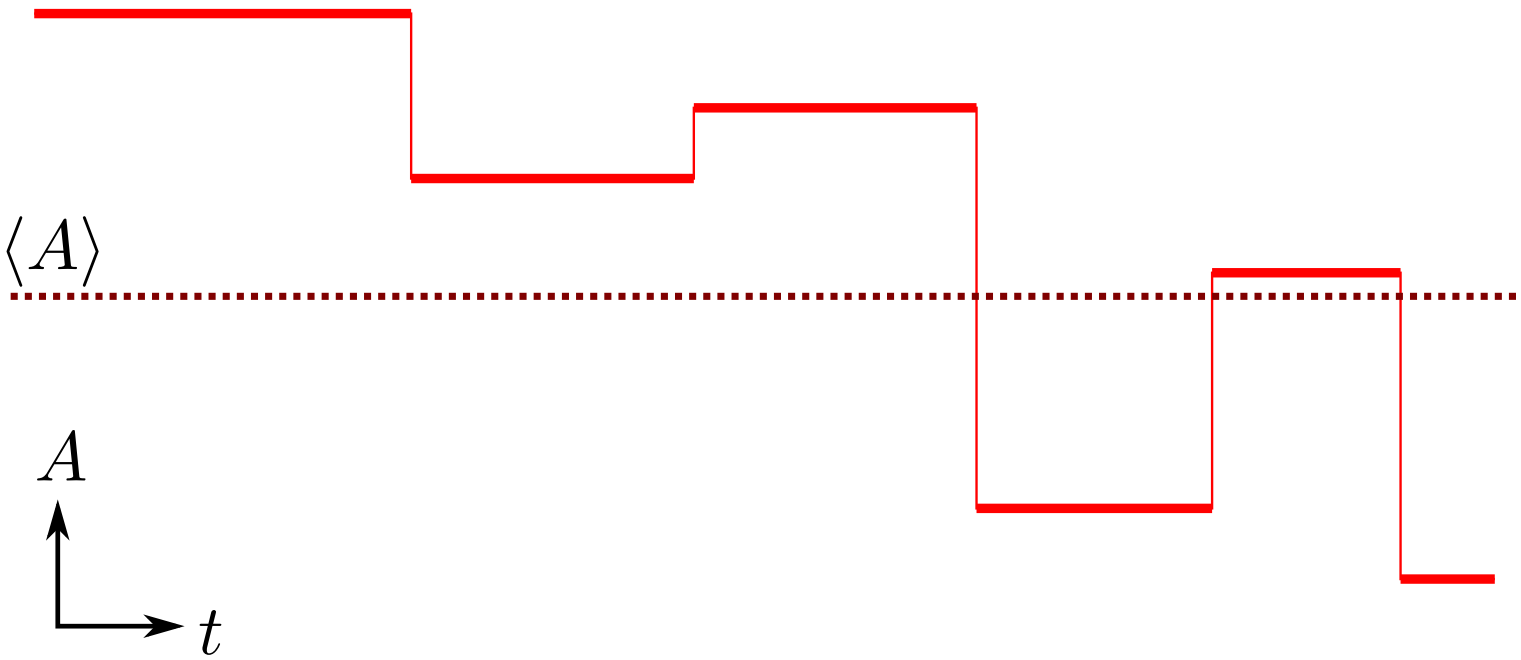
- Samples spaced by less than τ_A do not contribute to statistics

Correlation time and statistical efficiency

- For T much longer than the time scale on which $c_{AA}(t)$ decays to zero, $\epsilon_A^2(T) = \frac{\sigma_A^2}{T/\tau_A}$, where we introduced the **correlation time**

$$\tau_A = \int_{-\infty}^{\infty} c_{AA}(t)\,\mathrm{d}t$$

- Samples spaced by less than τ_A do not contribute to statistics

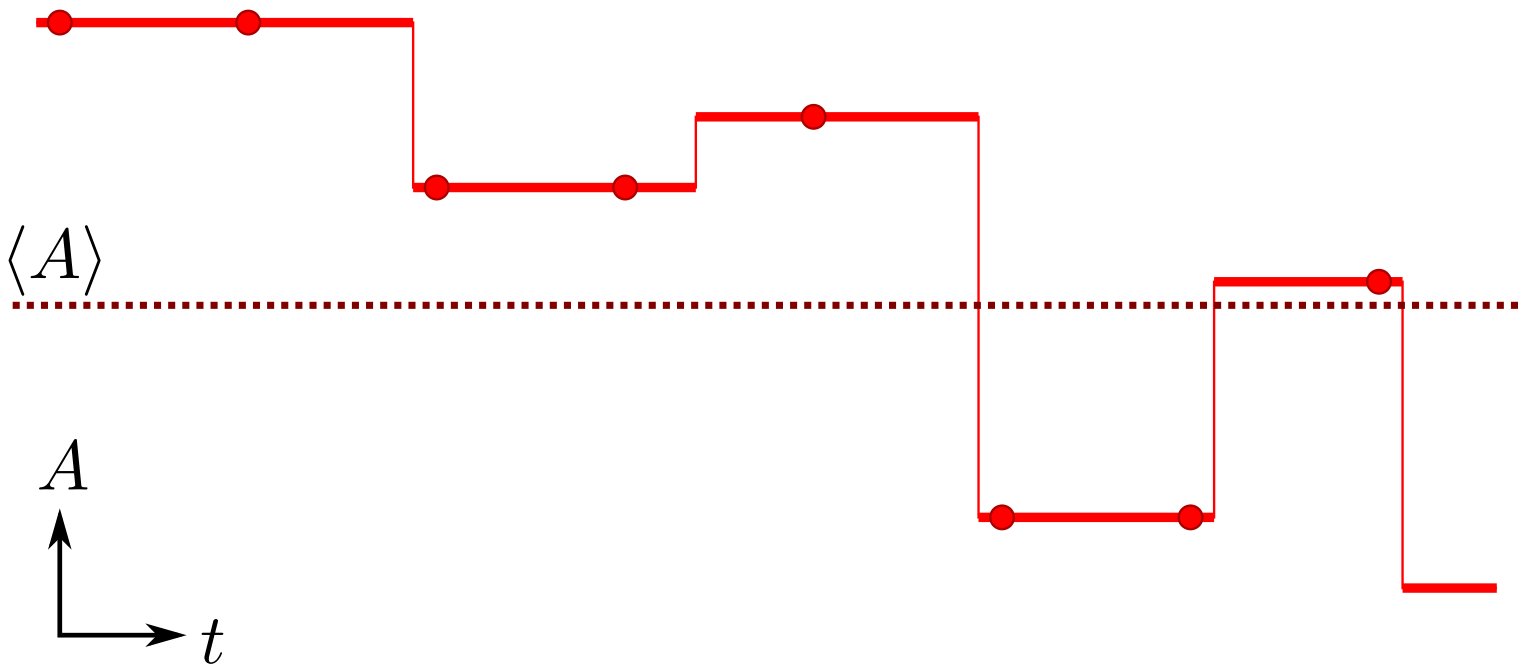


Correlation time and statistical efficiency

- For T much longer than the time scale on which $c_{AA}(t)$ decays to zero, $\epsilon_A^2(T) = \frac{\sigma_A^2}{T/\tau_A}$, where we introduced the **correlation time**

$$\tau_A = \int_{-\infty}^{\infty} c_{AA}(t)\,\mathrm{d}t$$

- Samples spaced by less than τ_A do not contribute to statistics

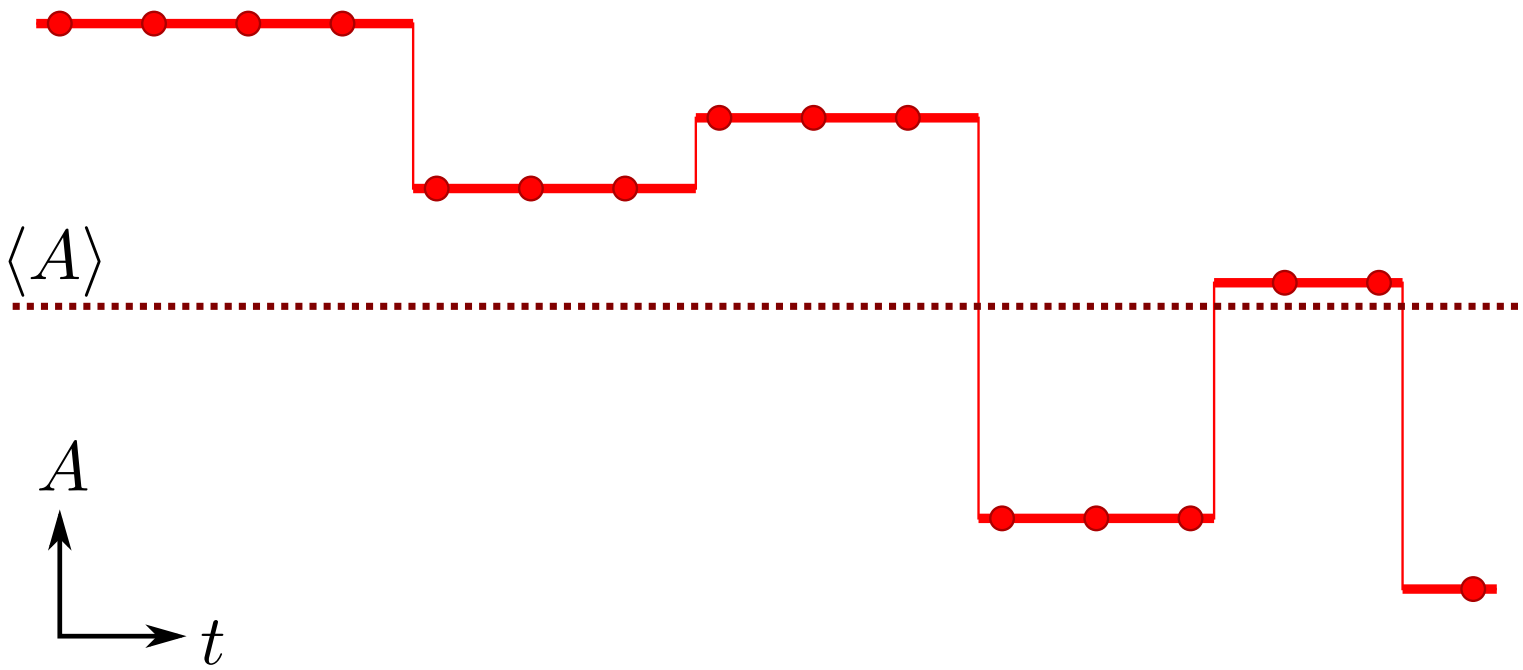


Correlation time and statistical efficiency

- For T much longer than the time scale on which $c_{AA}(t)$ decays to zero, $\epsilon_A^2(T) = \frac{\sigma_A^2}{T/\tau_A}$, where we introduced the **correlation time**

$$\tau_A = \int_{-\infty}^{\infty} c_{AA}(t)\,\mathrm{d}t$$

- Samples spaced by less than τ_A do not contribute to statistics

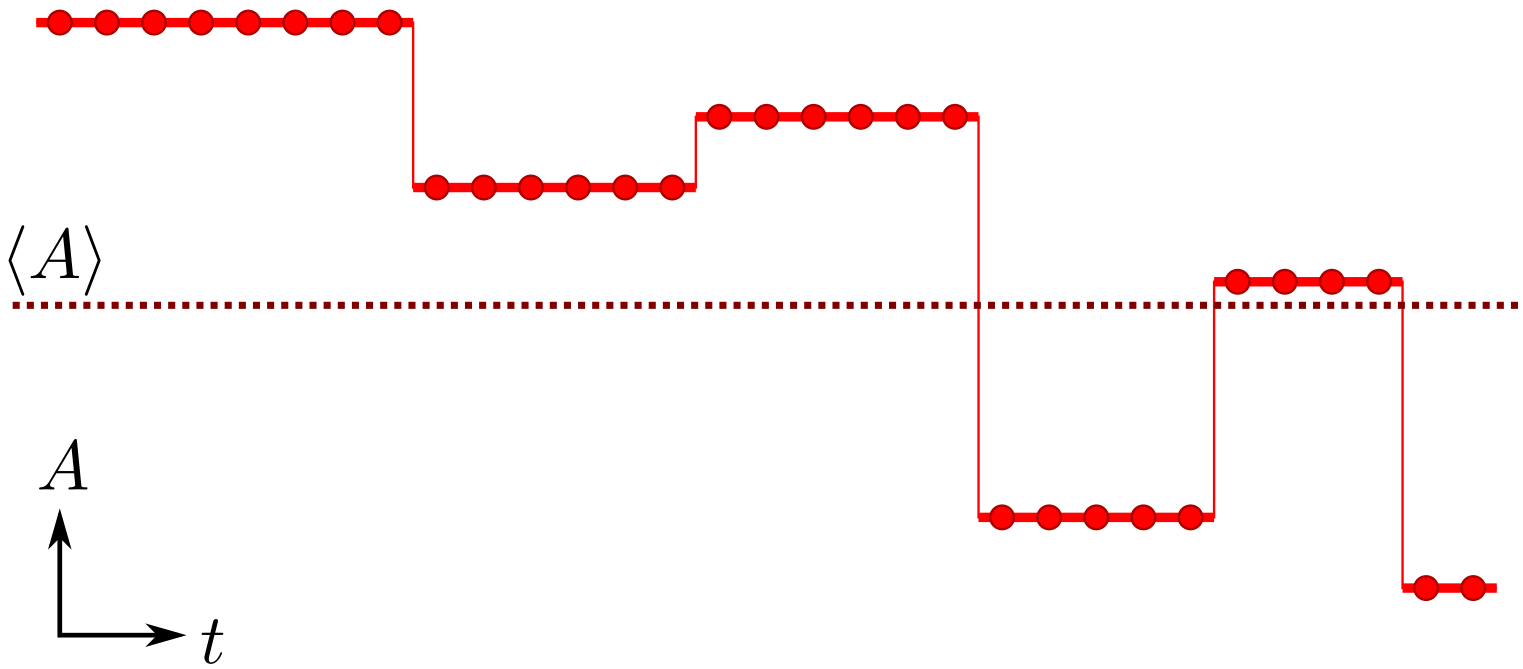


Correlation time and statistical efficiency

- For T much longer than the time scale on which $c_{AA}(t)$ decays to zero, $\epsilon_A^2(T) = \frac{\sigma_A^2}{T/\tau_A}$, where we introduced the **correlation time**

$$\tau_A = \int_{-\infty}^{\infty} c_{AA}(t)\,\mathrm{d}t$$

- Samples spaced by less than τ_A do not contribute to statistics

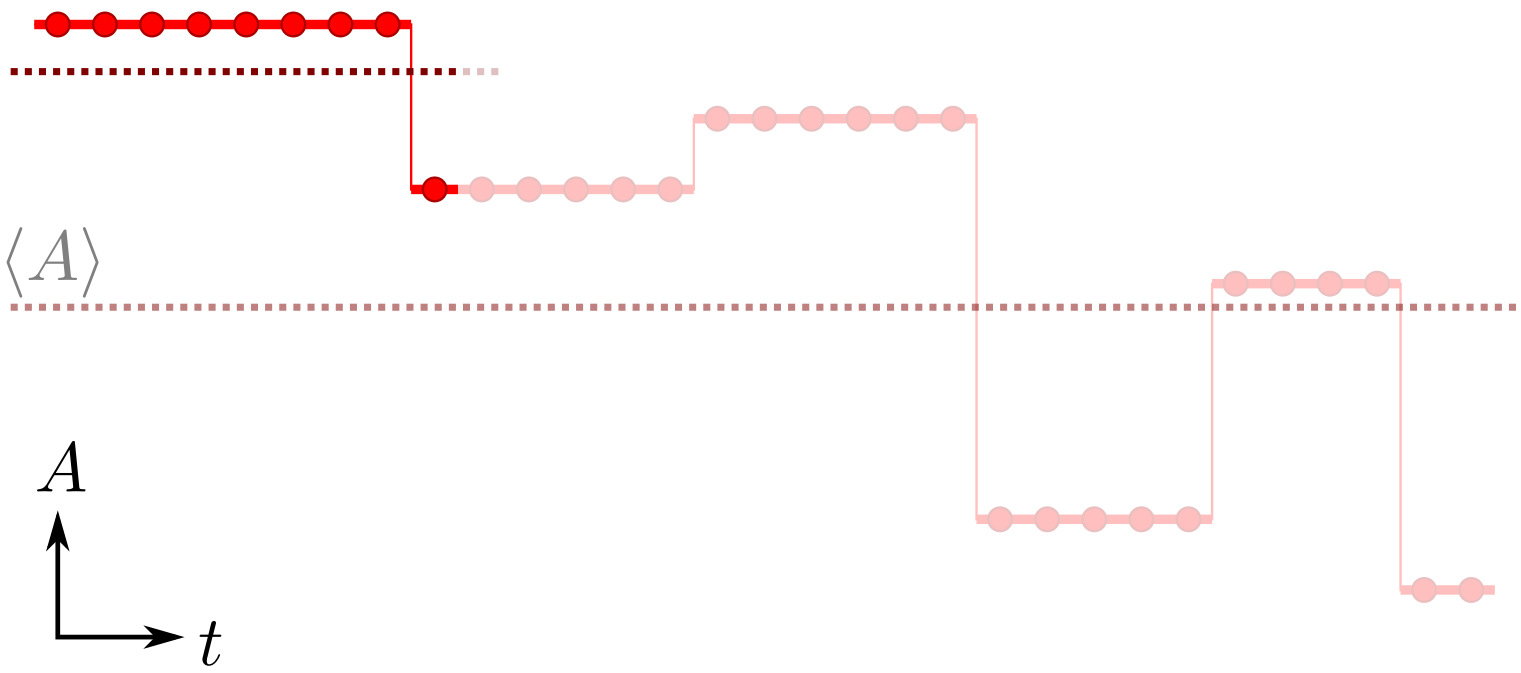


Practicalities of computing ACFs

- ACFs can be computed very efficiently with FFT
- There is a problem with the average entering the definition of c_{AA}
  - For $T \sim \tau_A$, $\langle A \rangle$ estimated from the trajectory is correlated with $A(t)$
  - The estimator for c_{AA} is biased (sort of $N \to N-1$ evaluating σ_A^2)
  - Use the exact (or most accurate) $\langle A \rangle$ whenever possible
  - In practice τ_A tends to always be underestimated

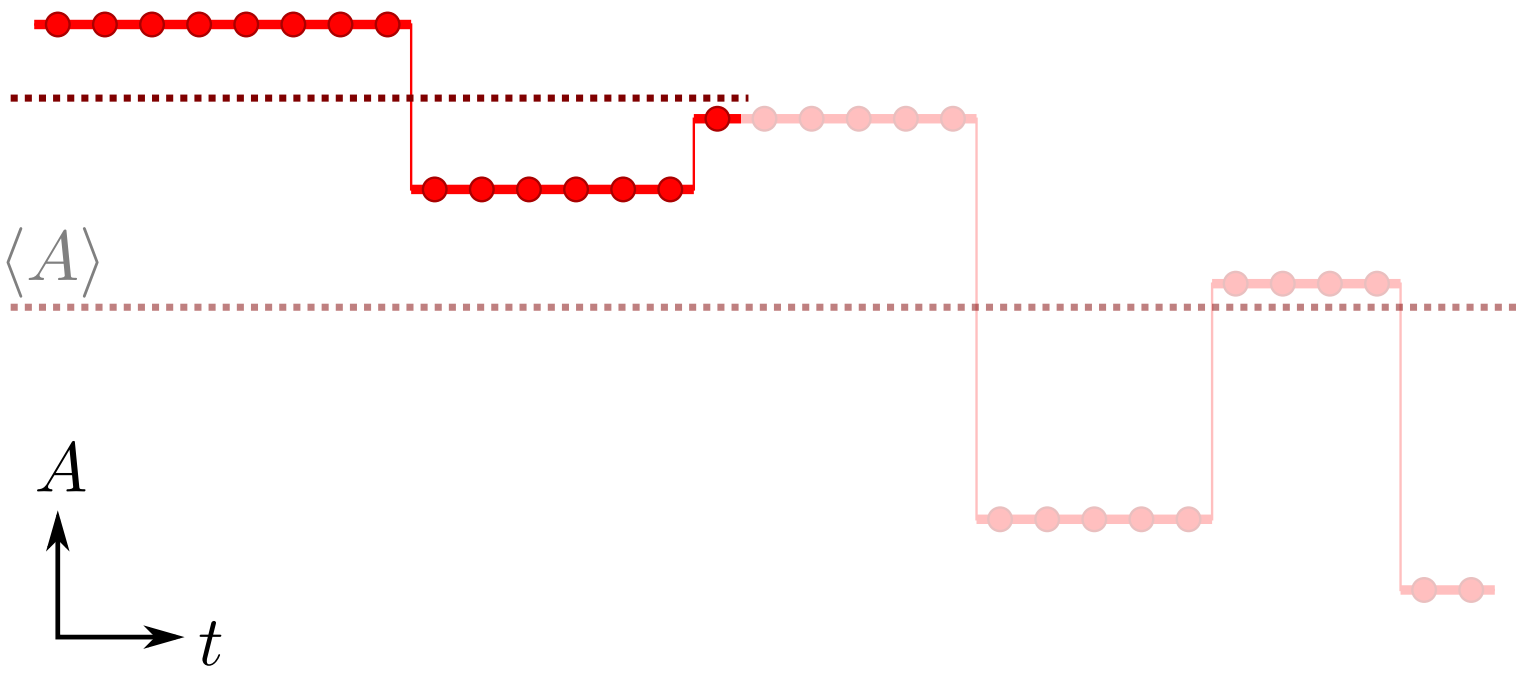
Practicalities of computing ACFs

- ACFs can be computed very efficiently with FFT
- There is a problem with the average entering the definition of c_{AA}
  - For $T \sim \tau_A$, $\langle A \rangle$ estimated from the trajectory is correlated with $A(t)$
  - The estimator for c_{AA} is biased (sort of $N \to N-1$ evaluating σ_A^2)
  - Use the exact (or most accurate) $\langle A \rangle$ whenever possible
  - In practice τ_A tends to always be underestimated



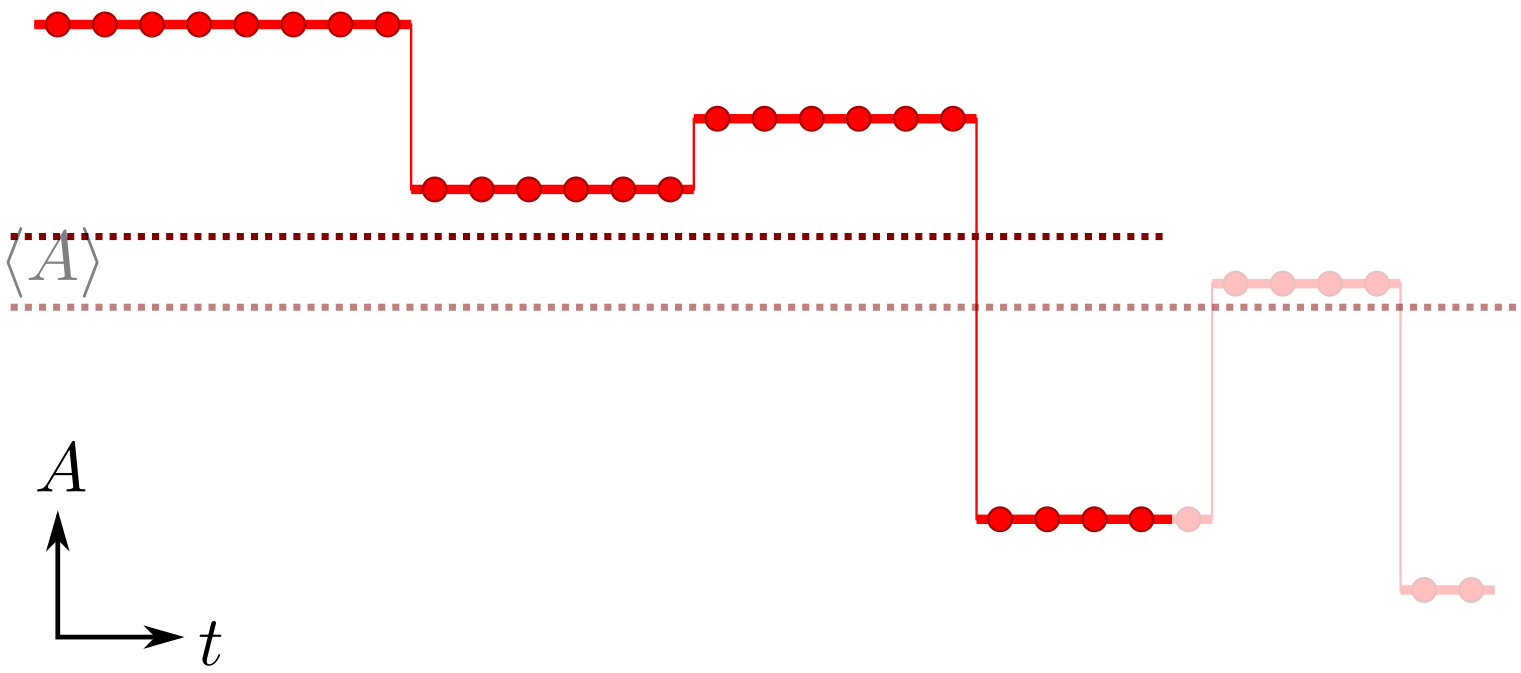
Practicalities of computing ACFs

- ACFs can be computed very efficiently with FFT
- There is a problem with the average entering the definition of c_{AA}
  - For $T \sim \tau_A$, $\langle A \rangle$ estimated from the trajectory is correlated with $A(t)$
  - The estimator for c_{AA} is biased (sort of $N \to N-1$ evaluating σ_A^2)
  - Use the exact (or most accurate) $\langle A \rangle$ whenever possible
  - In practice τ_A tends to always be underestimated



Practicalities of computing ACFs

- ACFs can be computed very efficiently with FFT
- There is a problem with the average entering the definition of c_{AA}
  - For $T \sim \tau_A$, $\langle A \rangle$ estimated from the trajectory is correlated with $A(t)$
  - The estimator for c_{AA} is biased (sort of $N \to N-1$ evaluating σ_A^2)
  - Use the exact (or most accurate) $\langle A \rangle$ whenever possible
  - In practice τ_A tends to always be underestimated



Wrap up and take home

- Monte Carlo and molecular dynamics are both effective ways to evaluate averages and generate configurations consistent with constant-temperature conditions
- Molecular dynamics entails an integration error, but is robust with growing system size. Problem with constant energy: next talk!
- Ergodicity is key to efficient evaluation of observables. Can be assessed by the autocorrelation time
- The autocorrelation function is hard to evaluate, in practice. Requires extremely well-converged statistics
- In most cases, autocorrelation time and therefore errors will be underestimated. Be careful!

Bibliography

Allen & Tildesley, "Computer Simulations of Liquids"
Frenkel & Smit, "Understanding Molecular Simulations"
Tuckerman, "Statistical Mechanics: Theory and Molecular Simulation"