Sampling Transition Paths: Some applications to biological systems

Henri Orland

IPhT, Université Paris-Saclay
CEA, CNRS
France

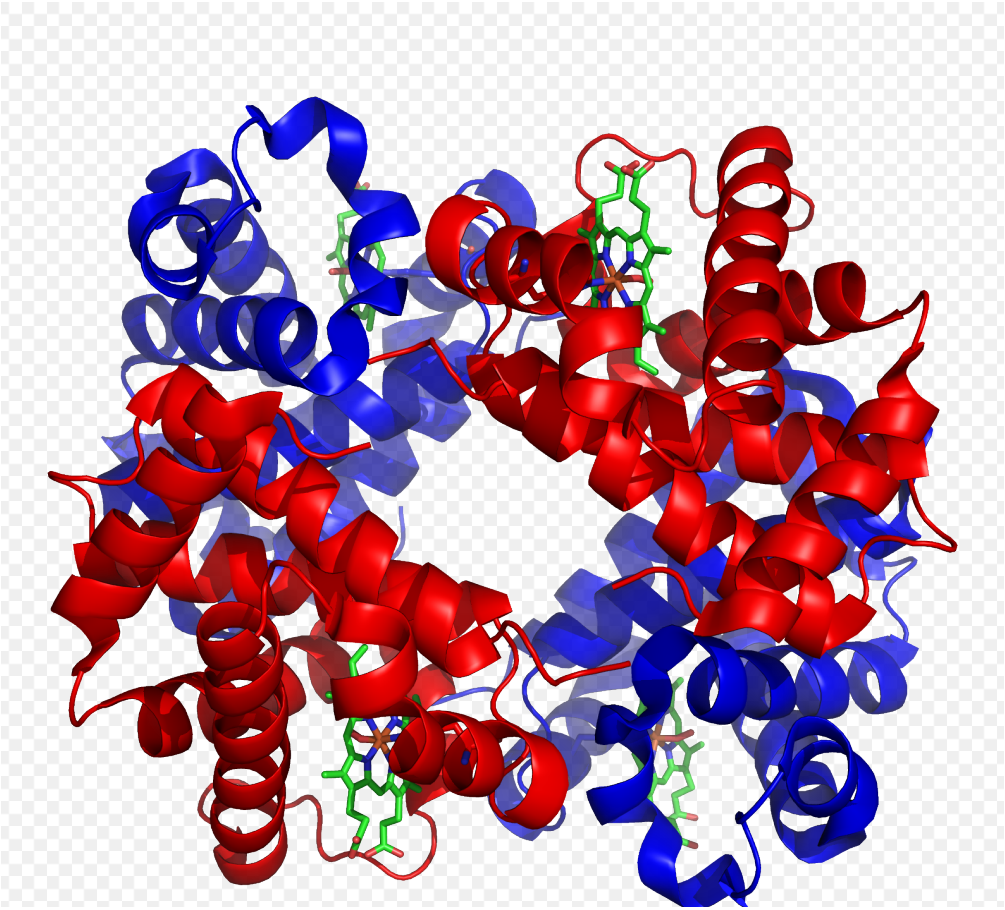
In collaboration with

- Satya Majumdar (Orsay, France)
- Cristian Micheletti (SISSA, Italy)
- Marc Delarue (Pasteur, Paris)
- Patrice Koehl (UC Davis, USA)

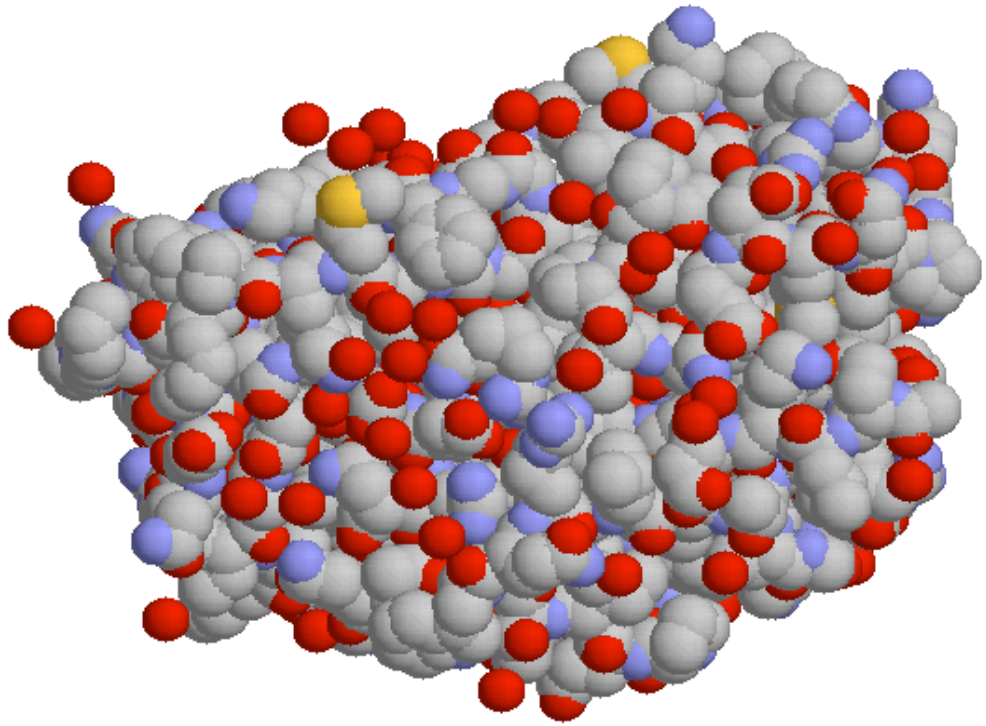
- The Folding Path problem
- Langevin dynamics and Path integral representation
- Brownian Bridges
  - some analytic examples
  - Knotting-unknotting of DNA with topoisomerase
- Transition Path Time distribution
- A useful form of the Bridge equation
- Allosteric transition of Adenylate Kinase

Proteins exist under 2 forms

- Proteins are polymers made of 20 amino-acids. They exist under 2 forms
- Folded or Native: globular unique conformation, biologically active
- Unfolded: random coil, biologically inactive
- Proteins are small objects: at equilibrium, they fluctuate (thermally) between the 2 forms.



Hemoglobin



HIV protease (199 residues)

The Protein Folding problem

- A sequence of amino-acids is given by the biologists.
- What is the 3d shape of the corresponding protein?
- To study this problem, use Molecular Dynamics: Karplus, Levitt and Warschel, Nobel prize in Chemistry 2013 ➡ ANTON
- More recently, use Machine Learning: AlphaFold2: Hassabis, Jumper, Nobel prize in Chemistry 2024

Molecular Dynamics

- Proteins are made of amino-acids, which are themselves made of atoms
- Each configuration of atoms $\{r_i\}$ has a certain energy
- Parametrize the interaction between constituent atoms (valence bond, Lennard-Jones, Coulomb, etc.)

Use Langevin or Newton equations

$$m_i \ddot{r}_i + \gamma_i \dot{r}_i = -\frac{\partial E}{\partial r_i} + \eta_i(t)$$

Friction coefficient

Interactions

Gaussian white noise

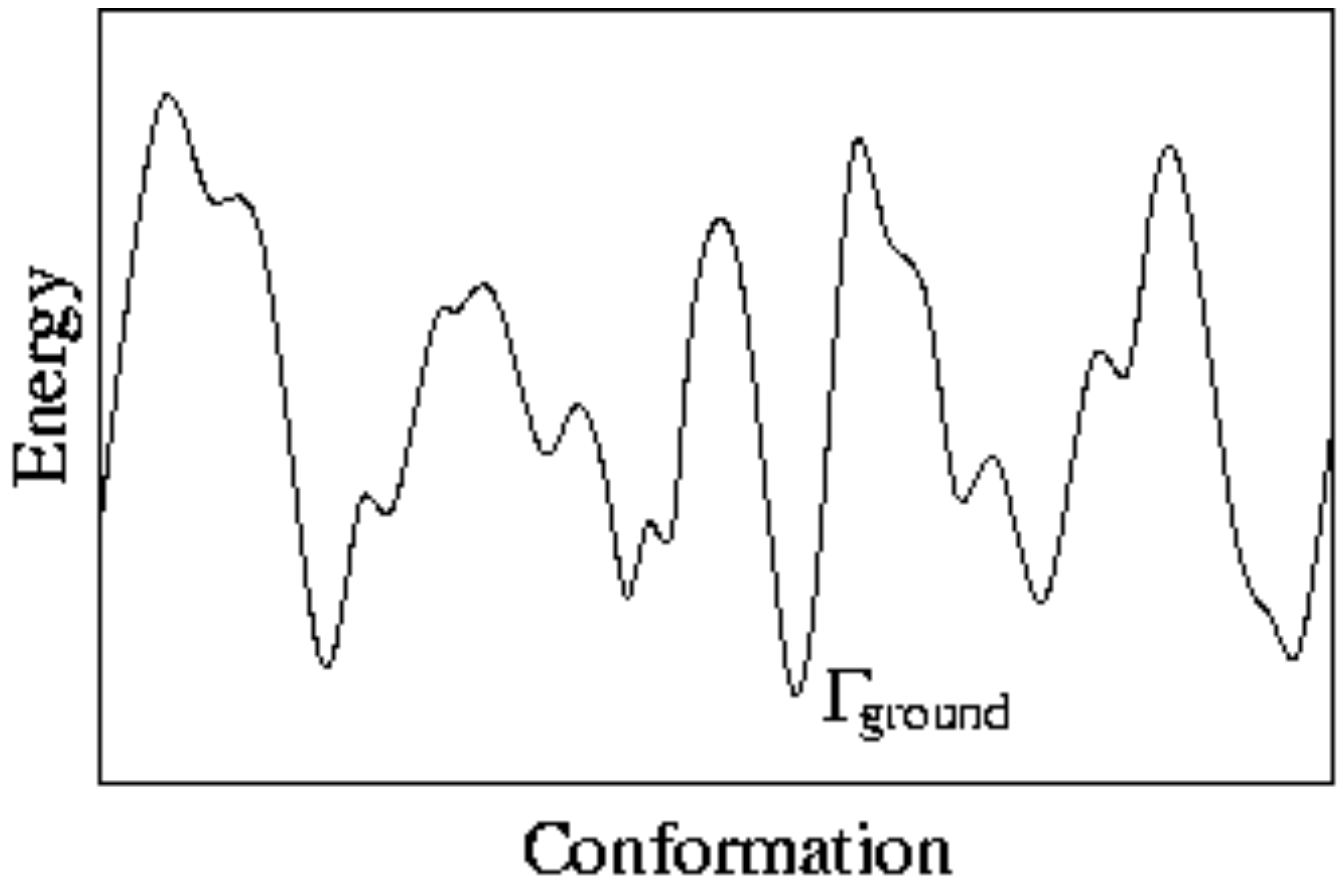
with Fluctuation-Dissipation relation

$$\langle \eta_i(t) \eta_j(t') \rangle = 2\gamma_i k_B T \delta_{ij} \delta(t - t')$$

Theorem: the probability distribution $P(\{r_i(t)\})$ converges to the Boltzmann distribution at large time $\frac{e^{-\beta E(\{r_i\})}}{Z}$

Why is it difficult?

- To discretize the equations, one must use time steps of the order of $10^{-15}s$
- Large number of degrees of freedom (a few thousand) plus few thousand water molecules
- Empirical force fields not necessarily accurate enough
- Longest runs: around $1\mu s$ << folding time 1ms- 1s
- Recently, runs of 1ms on short proteins (ANTON)
- Reason: Many metastable states and high barriers


Energy
Γ_{ground}
Conformation

The problem of protein structure prediction is too complicated for MD

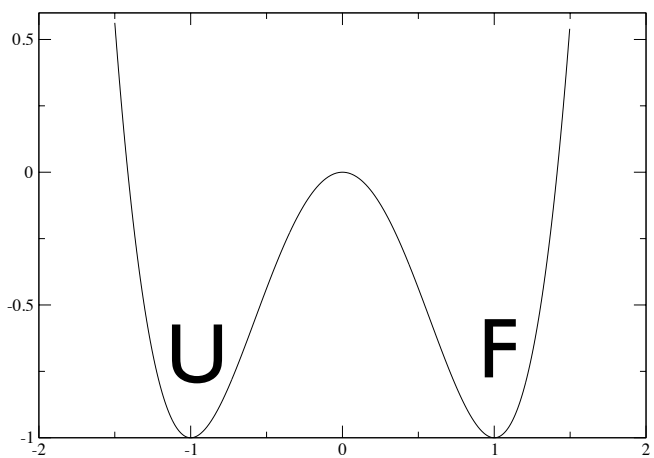
Solved by machine learning: AlphaFold 2

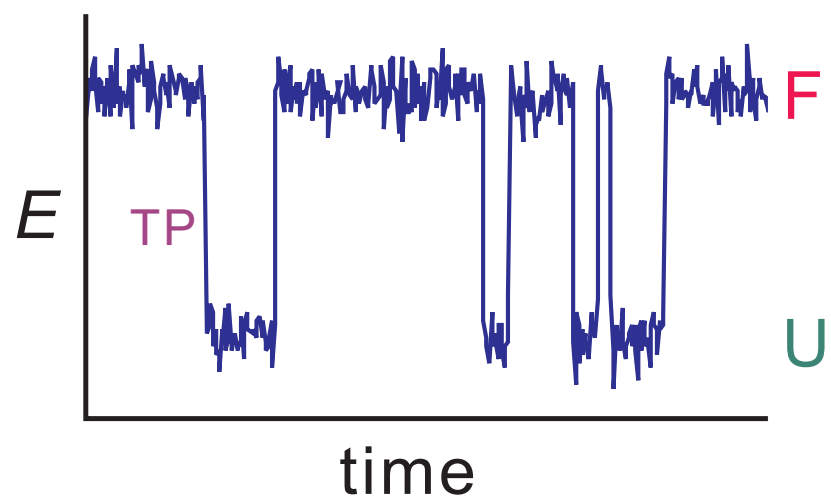
Other problem: How do proteins fold? How do they go from Unfolded to Native State?

(Assuming unfolded and native states are known)

Physical Picture

- In given denaturant conditions, a protein spends a fraction of its time in the native state and a fraction of its time in the denatured state.





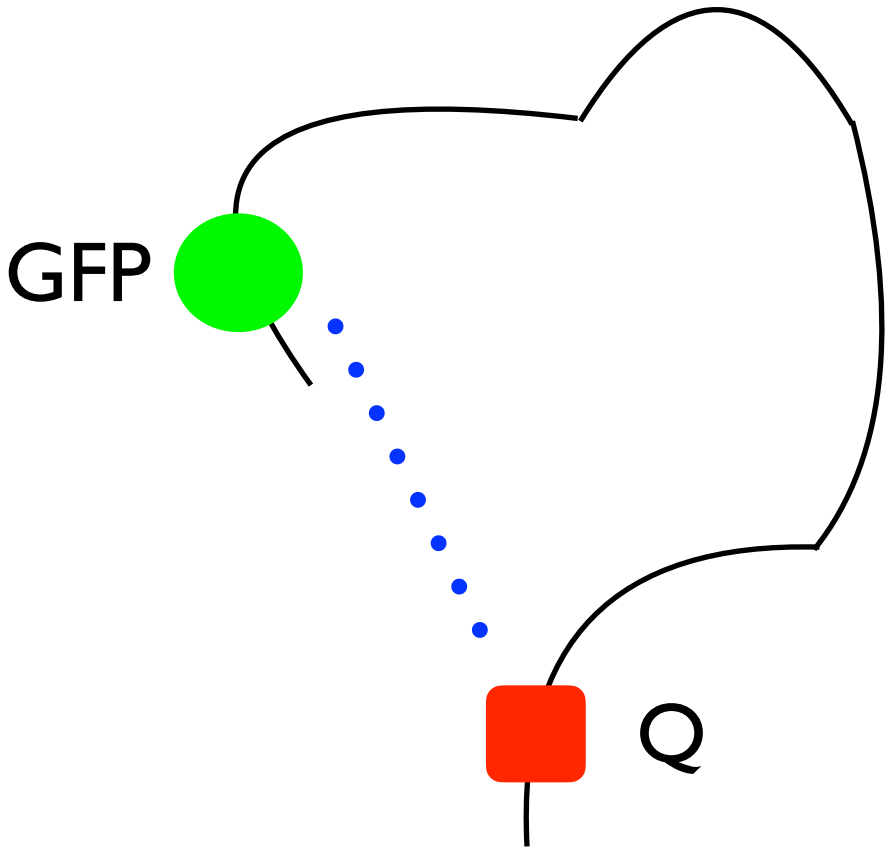
Motivation from single molecule experiments on proteins

W. Eaton (NIH), G. Haran (Weitzman), M. Woodside (U. Alberta),…

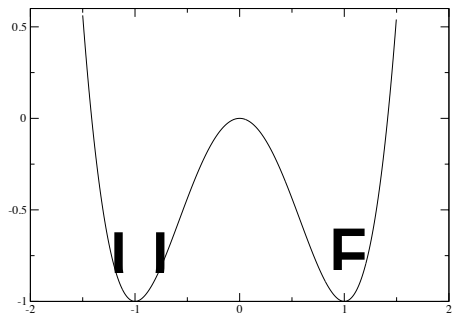
- Examples:
  - from denatured to native in native conditions
  - Allosteric transition between A and B

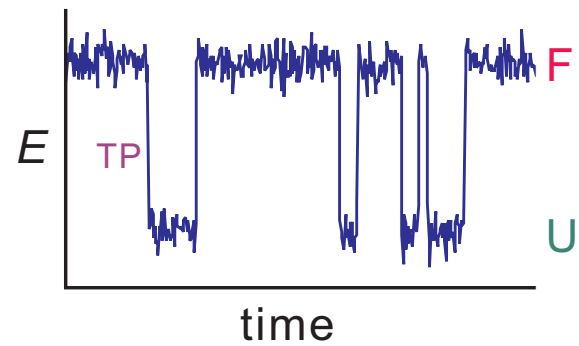
Difficulty: looking for exponentially rare events

Single Molecule Experiments



FRET experiments





The Transition Path Problem

- The problem: Assume a system can go (stochastic dynamics) from state A to state B:
  - liquid to solid; nucleation; phase changes,..
  - Chemical reactions,…
  - biopolymer folding: transition between denatured and native state, allostery,…
- Which pathways (or family of pathways) does the system use? What are the trajectories from A to B?

Related to the Schrödinger Bridge problem
Schrödinger (1931): On the reversal of Natural Laws

How do you go from an initial probability distribution P_0 to a final distribution P_f in time t_f if points follow Brownian Motion

Also related to Optimal Transport (Monge problem)

Langevin dynamics

- The case of one particle in a potential $U(x)$ at temperature T
- Use Langevin dynamics

$$m\frac{d^2x}{dt^2} + \gamma\frac{dx}{dt} + \frac{\partial U}{\partial x} = \zeta(t)$$

- where γ is the friction and $\zeta(t)$ is a random noise

$$< \zeta(t)\zeta(t') >= 2k_BT\gamma\delta(t-t')$$

Overdamped Langevin dynamics

- At large enough time scale, mass term negligible

$$m\omega^2 \approx \gamma\omega$$

$$\tau \approx 2\pi\frac{m}{\gamma}$$

$$\gamma = \frac{k_B T}{D}$$

$$\tau \approx 10^{-13}\mathrm{s}$$

$D = 10^{-5}\mathrm{cm}^2/\mathrm{s}$ $\qquad$ $m \approx 5.10^{-26}\mathrm{kg}$

- Take overdamped Langevin (Brownian) dynamics

$$\frac{dx}{dt} = -\frac{1}{\gamma}\frac{\partial U}{\partial x} + \eta(t)$$

- with Gaussian noise:

$$\langle \eta(t)\eta(t') \rangle = \frac{2k_B T}{\gamma}\delta(t-t')$$

- γ is the friction coefficient: $D = \frac{k_B T}{\gamma}$

Diffusion coefficient

$$\eta(t) = \sqrt{\frac{2k_B T}{\gamma}}\xi(t)$$

Normal variable

- Discretization of the Langevin equation

Euler-Maruyama

$$x_{k+1} = x_k - D\beta dt \frac{\partial U}{\partial x_k} + \eta_k dt$$

with

$$P(\eta_k) = \left(\frac{dt}{4\pi D}\right)^{d/2} e^{-\frac{dt}{4D}\eta_k^2}$$

If studying high (free) energy barrier crossing, crossing events are exponentially rare.

Folding time of proteins: 1ms-1s

Typical timestep: 10^{-15}s

Very long simulation to hopefully see one folding event due to high barrier and exponentially large number of metastable states

Special purpose computer: ANTON (D.E. Shaw)

- Discretization of the Langevin equation

Euler-Maruyama

$$x_{k+1} = x_k - D\beta dt \frac{\partial U}{\partial x_k} + \eta_k dt$$

with

$$P(\eta_k) = \left(\frac{dt}{4\pi D}\right)^{d/2} e^{-\frac{dt}{4D}\eta_k^2}$$

$$P(x_{k+1}, t+dt \,|\, x_k, t) = \left(\frac{dt}{4\pi D}\right)^{d/2} e^{-\frac{dt}{4D}\left(\frac{x_{k+1}-x_k}{dt} + D\beta\frac{\partial U}{\partial x_k}\right)^2}$$

Path integral representation

$$P(x_f,t_f|x_i,0)=\int\prod_{k=1}^{N}dx_k\exp\left(-\frac{dt}{4D}\sum_{k=1}^{N}\left(\frac{x_{k+1}-x_k}{dt}+\frac{D}{k_BT}\frac{\partial U}{\partial x_k}\right)^2\right)$$

- fixed end points x_i and x_f
- Continuous limit

Onsager-Machlup (Ito)

$$P(\mathbf{r}_f,t_f|\mathbf{r},t)=\int_{\mathbf{r}(t)=\mathbf{r}}^{\mathbf{r}(t_f)=\mathbf{r}_f}\mathcal{D}\mathbf{r}(\tau)e^{-\frac{1}{4D}\int_t^{t_f}d\tau\left(\dot{\mathbf{r}}+\frac{1}{\gamma}\nabla U\right)^2}$$

$$=e^{-\beta(U(\mathbf{r}_f)-U(\mathbf{r}))/2}\int_{\mathbf{r}(t)=\mathbf{r}}^{\mathbf{r}(t_f)=\mathbf{r}_f}\mathcal{D}\mathbf{r}(\tau)e^{-\int_t^{t_f}d\tau\left(\frac{\dot{\mathbf{r}}^2}{4D}+\frac{1}{D\gamma^2}V(\mathbf{r}(\tau))\right)}$$

Stratonovich

where the effective potential V is given by

$$V(\mathbf{r})=\frac{1}{4}(\nabla U)^2-\frac{k_BT}{2}\nabla^2U$$

Path Integral = Feynman Path integral = Schroedinger equation

$$D = \frac{k_B T}{\gamma}$$

- Dominant Paths: Saddle-Point expansion: Minimise action: Newton equation = Instanton $\ddot{r} = \frac{2}{\gamma^2}\frac{\partial V}{\partial r}$ with B.C. x_i, x_f

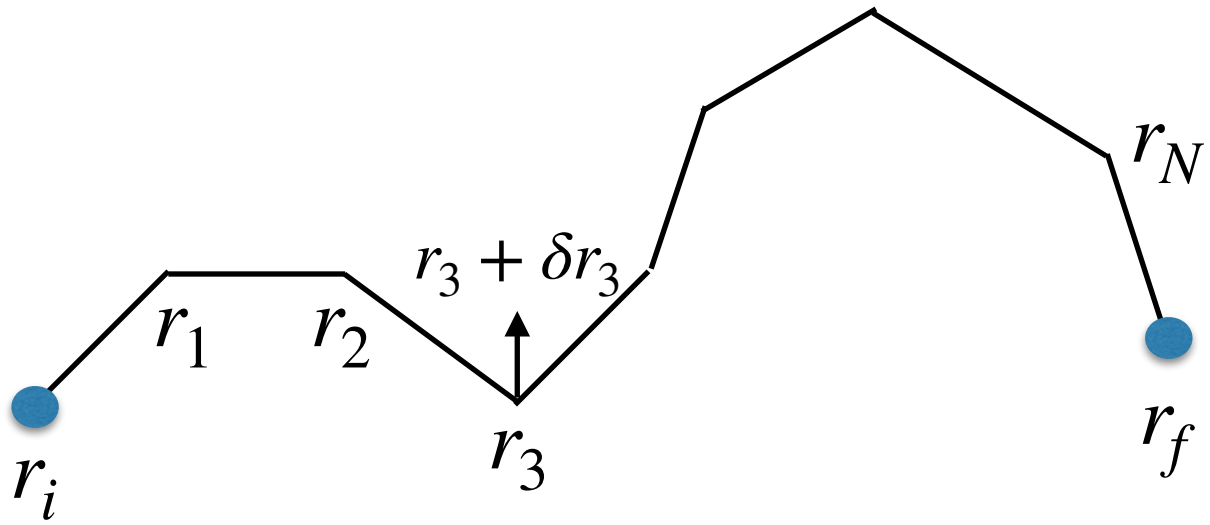
- Low dimension: solve Schrödinger equation

Transition path sampling

Path integral representation

$$P(x_f, t_f | x_i, 0) = \int \prod_{k=1}^{N} dx_k \exp\left(-\frac{dt}{4D} \sum_{k=1}^{N} \left(\frac{x_{k+1} - x_k}{dt} + \frac{D}{k_B T} \frac{\partial U}{\partial x_k} \right)^2 \right)$$

- Construct an initial trajectory with fixed end points x_i and x_f
- Deform the trajectory locally and accept or reject with a Monte Carlo algorithm



Initial trajectory from r_i to r_f

- Deform the trajectory locally and accept or reject with a Monte Carlo algorithm
  - Difficulties:
    - Huge sampling space
    - Depends very much on initial trajectory

Bridges (Doob)

- Consider paths starting at $(x_0, 0)$ and conditioned to end at (x_f, t_f)
- The conditional probability for such a path to be at (x, t) is given by

$$\mathcal{P}(x,t) = P(x,t|x_f,t_f \text{ and } x_0,0)$$

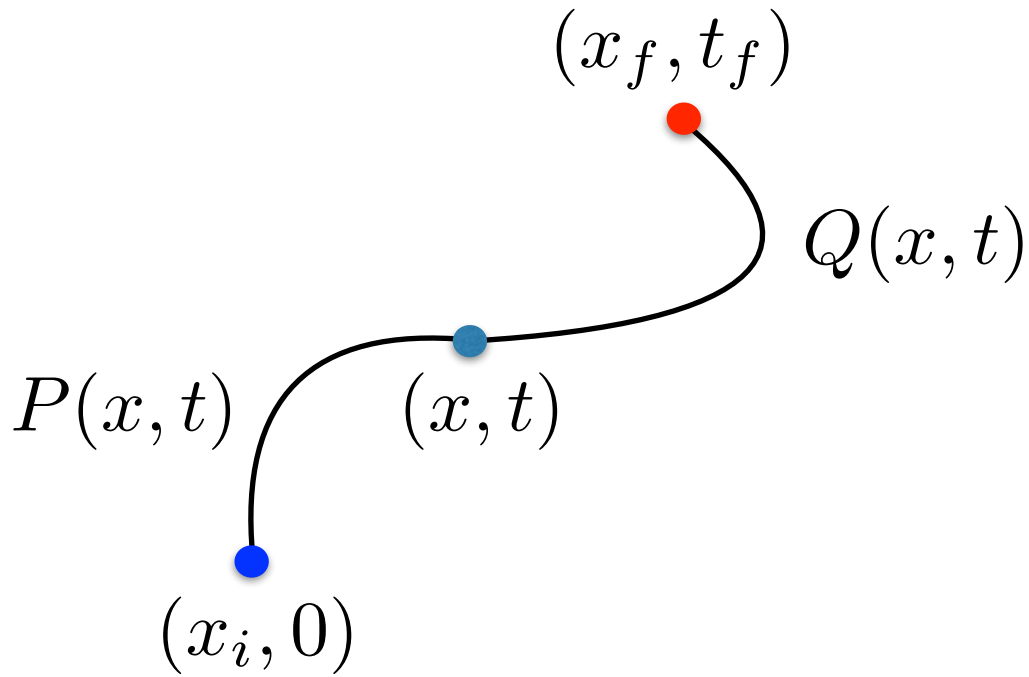
$$\mathcal{P}(x,t) = \frac{1}{P(x_f,t_f|x_0,0)} Q(x,t)P(x,t)$$

adjoint FP

FP

$$P(x,t) = P(x,t|x_0,0)$$

$$Q(x,t) = P(x_f,t_f|x,t)$$



$$\mathcal{P}(x,t) = \frac{1}{P(x_f,t_f|x_0,0)} Q(x,t) P(x,t)$$

FP equation

adjoint FP equation

$$P(x,t) = P(x,t|x_0,0)$$

$$Q(x,t) = P(x_f,t_f|x,t)$$

Fokker-Planck and adjoint

FP

$$\frac{\partial P}{\partial t} = D\frac{\partial}{\partial x}\left(\frac{\partial P}{\partial x} + \beta\frac{\partial U}{\partial x}P\right)$$

adjoint FP

$$\frac{\partial Q}{\partial t} = -D\frac{\partial^2 Q}{\partial x^2} + D\beta\frac{\partial U}{\partial x}\frac{\partial Q}{\partial x}$$

$$\dot{\mathcal{P}} = \frac{\dot{Q}P + Q\dot{P}}{P(x_f, t_f|x_0, 0)}$$

conditional probability

$$\frac{\partial \mathcal{P}}{\partial t} = D\frac{\partial}{\partial x}\left(\frac{\partial \mathcal{P}}{\partial x} + \frac{\partial}{\partial x}\left(\beta U - 2\ln Q\right)\mathcal{P}\right)$$

- Modified Langevin equation for conditioned paths

$$\frac{dx}{dt} = -\frac{D}{k_B T}\frac{\partial U}{\partial x} + 2D\frac{\partial \ln Q}{\partial x} + \eta(t)$$

where

$$Q(x,t) = P(x_f, t_f | x, t)$$

- Equation is Markovian (but depends through Q on the whole future of the trajectories!)
- No bias in the statistics of the trajectories
- Could be obtained from Girsanov theorem

- Example: Brownian bridges

$$U(x) = 0$$

$$Q(x,t) = P(x_f, t_f | x, t) = \sqrt{\frac{1}{4\pi D(t_f - t)}} e^{-\frac{(x_f - x)^2}{4D(t_f - t)}}$$

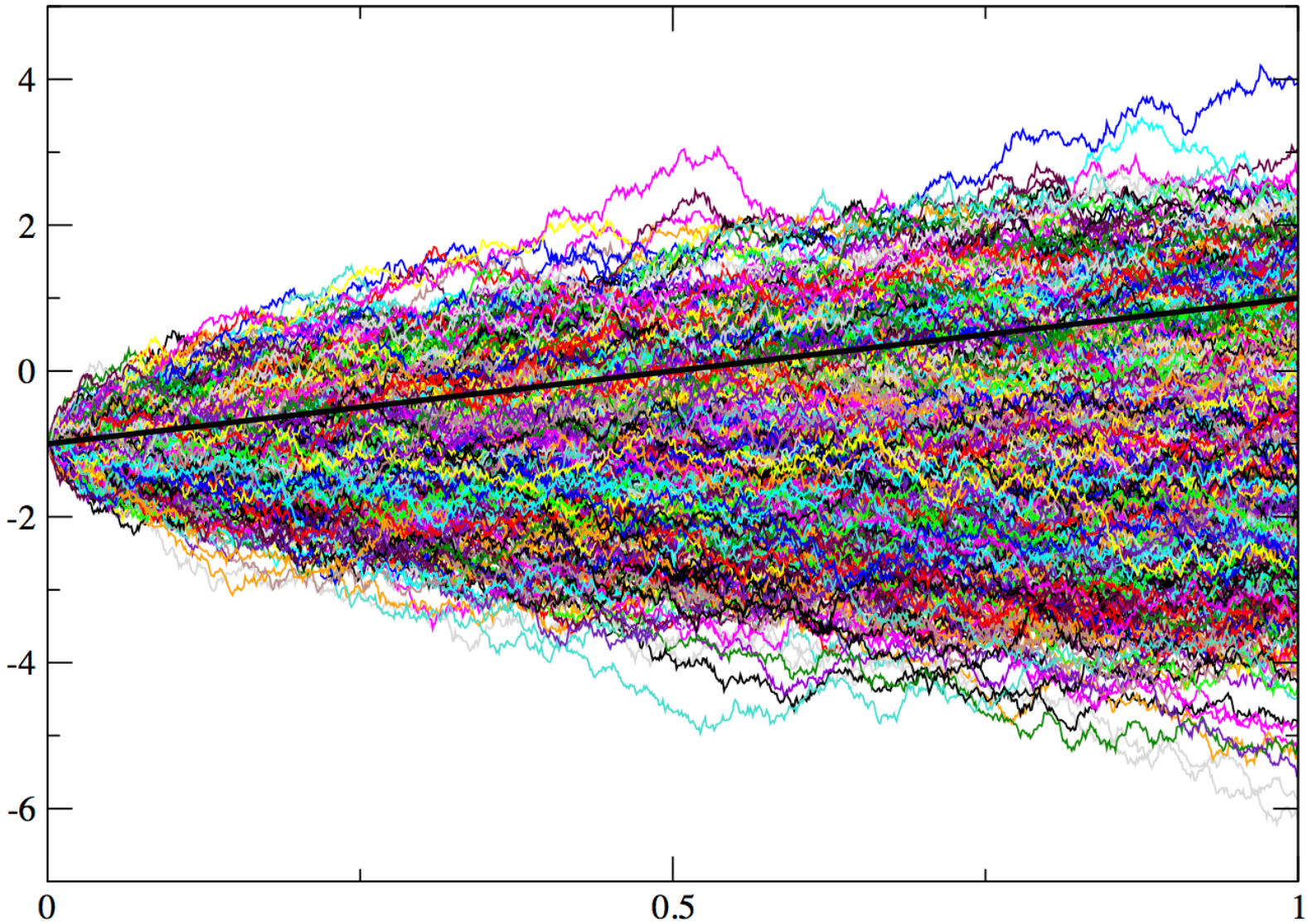
- Conditioned Langevin equation becomes

$$\frac{dx}{dt} = \frac{x_f - x}{t_f - t} + \eta(t)$$

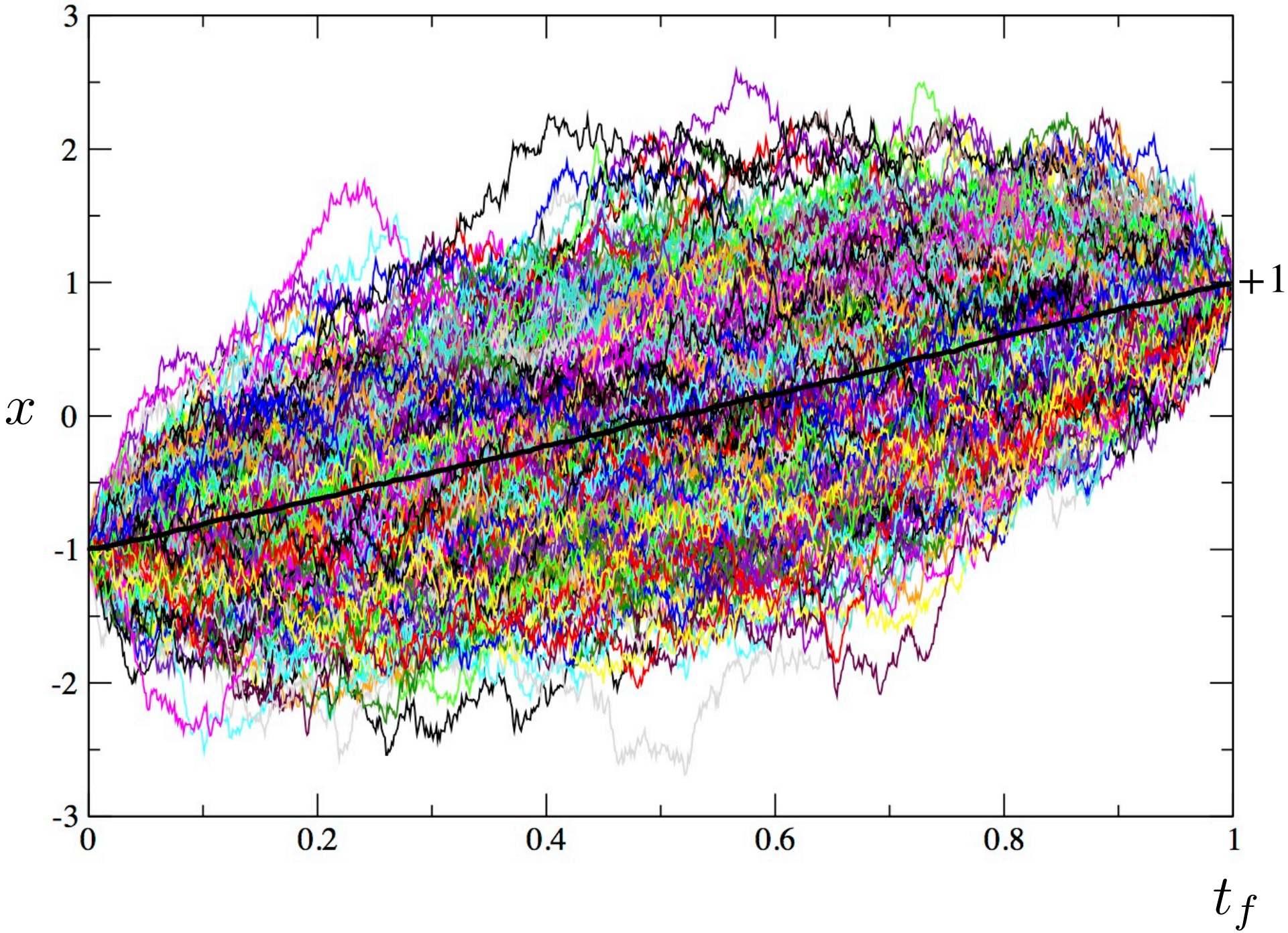
$$\frac{dX}{dt} = \frac{x_f - X}{t_f - t} \Longrightarrow$$

average is linear in time

Free Langevin



Constrained Langevin



- Example: Brownian meanders = Brownian walks constrained to stay with x>0 during fixed time. Use mirror image

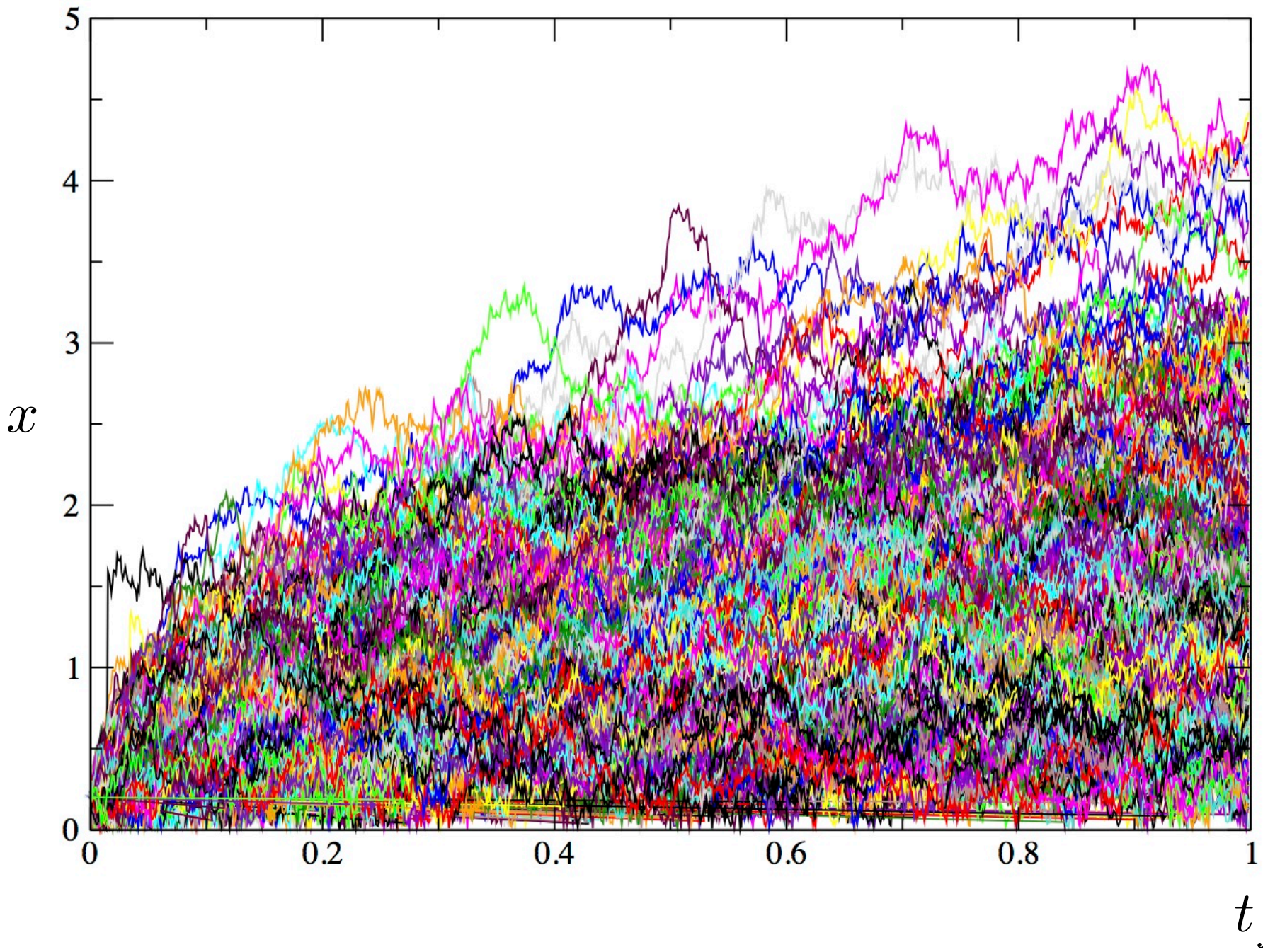
$$\dot{x} = 2D\frac{\partial \log Q(x,t)}{\partial x} + \eta(t)$$

$$= \frac{\left(\frac{x_f - x}{t_f - t}\right) e^{-\frac{1}{4D}\frac{(x_f-x)^2}{t_f-t}} + \left(\frac{x_f + x}{t_f - t}\right) e^{-\frac{1}{4D}\frac{(x_f+x)^2}{t_f-t}}}{e^{-\frac{1}{4D}\frac{(x_f-x)^2}{t_f-t}} - e^{-\frac{1}{4D}\frac{(x_f+x)^2}{t_f-t}}} + \eta(t)$$

- Integrate for

$$x_f > 0$$

$$\frac{dx}{dt} = \frac{2}{\sqrt{4\pi D(t_f - t)}} \frac{\exp\left(-\frac{x^2}{4D(t_f-t)}\right)}{\mathrm{Erf}\left(\frac{x}{\sqrt{4D(t_f-t)}}\right)} + \eta(t)$$


5
4
3
2
1
0
x
0
0.2
0.4
0.6
0.8
1
t_f

- Example: Brownian excursions: Brownian walks x>0, starting around x=0 and ending at x=0.

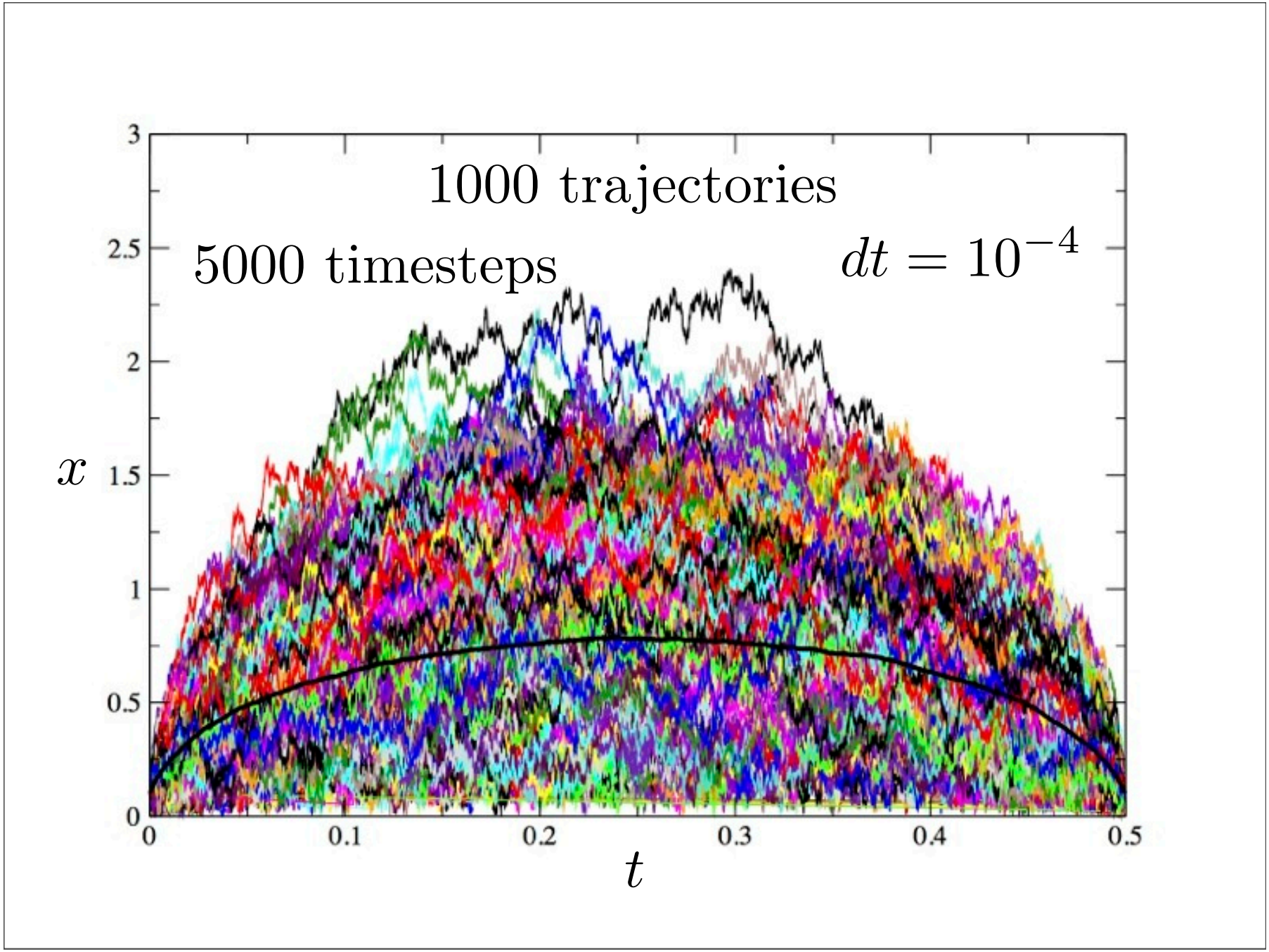
$$\dot{x} = 2D\frac{\partial \log Q(x,t)}{\partial x} + \eta(t)$$

$$= \frac{\left(\frac{x_f - x}{t_f - t}\right) e^{-\frac{1}{4D}\frac{(x_f-x)^2}{t_f-t}} + \left(\frac{x_f + x}{t_f - t}\right) e^{-\frac{1}{4D}\frac{(x_f+x)^2}{t_f-t}}}{e^{-\frac{1}{4D}\frac{(x_f-x)^2}{t_f-t}} - e^{-\frac{1}{4D}\frac{(x_f+x)^2}{t_f-t}}} + \eta(t)$$

- Take the limit $x_f \to 0$

$$\frac{dx}{dt} = \frac{2D}{x}\left(1 - \frac{x^2}{2D(t_f - t)}\right) + \eta(t)$$

Example: Brownian excursions above a line

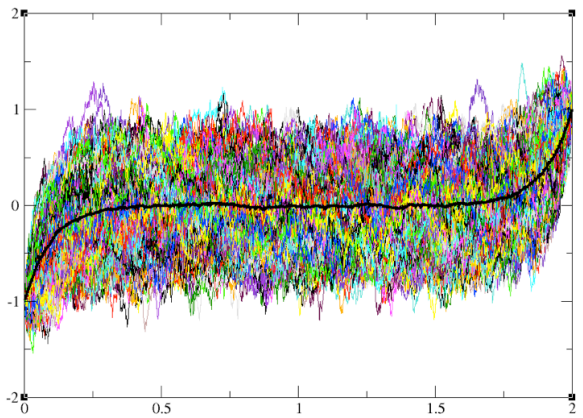


- Example: the Harmonic oscillator (Ornstein-Uhlenbeck process)
- $$U(x) = \frac{1}{2} K x^2$$
- $$\dot{x} = -\frac{1}{\gamma} K x + \eta(t)$$
- Bridge equation

$$\frac{dx}{dt} = \frac{K}{\gamma} \frac{x_f - x \cosh \frac{K}{\gamma}(t_f - t)}{\sinh \frac{K}{\gamma}(t_f - t)} + \eta(t)$$

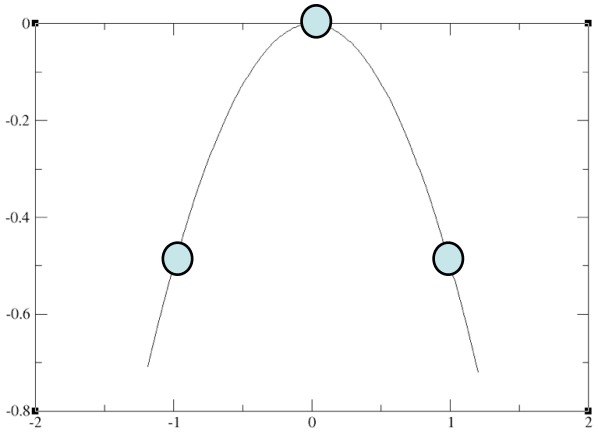
- Note that this equation does not depend on the sign of *K*: same conditioned trajectories

$t_f = 2$

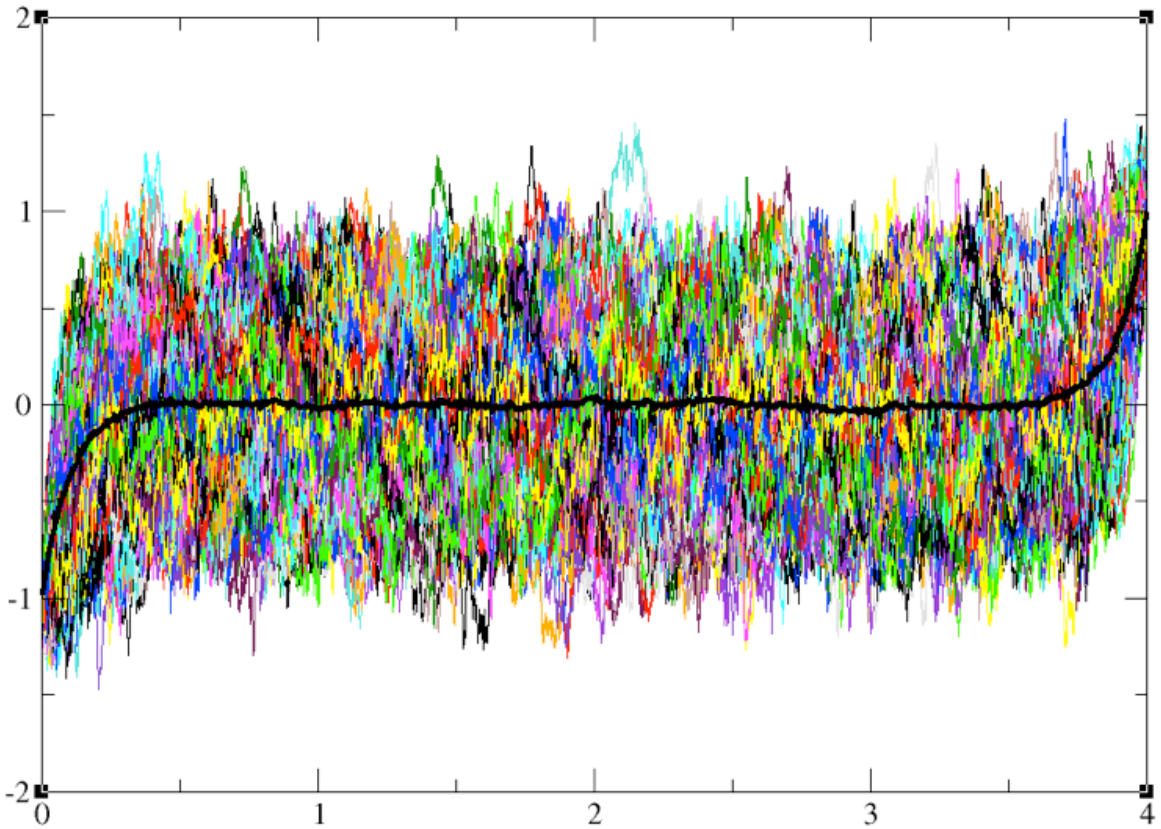


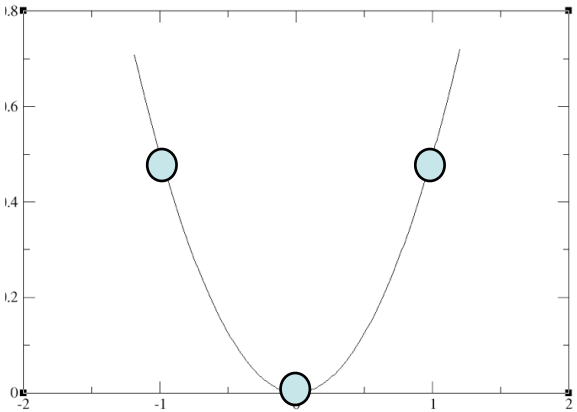
T=0.1
K=±1

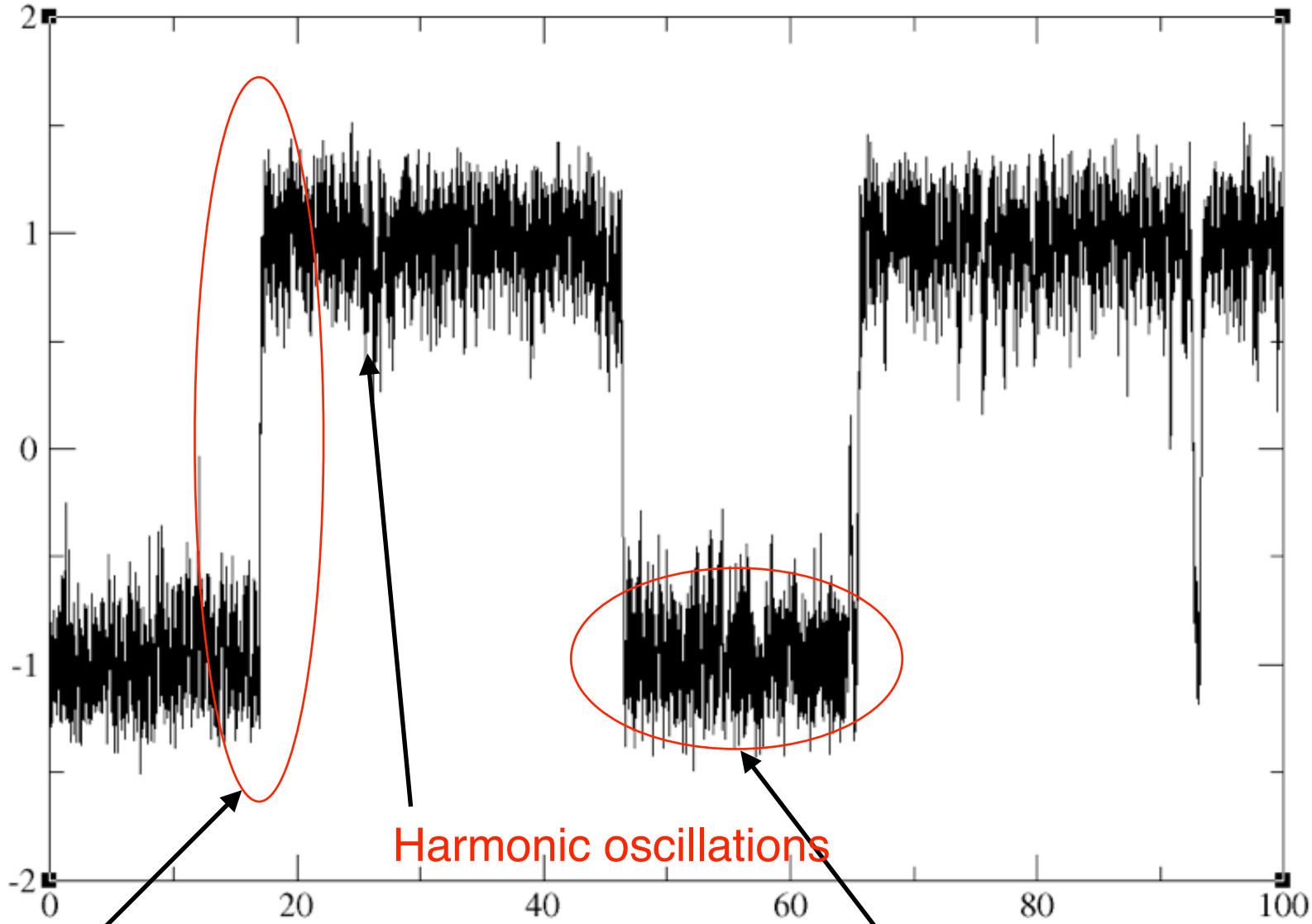
D=1



$t_f = 5$





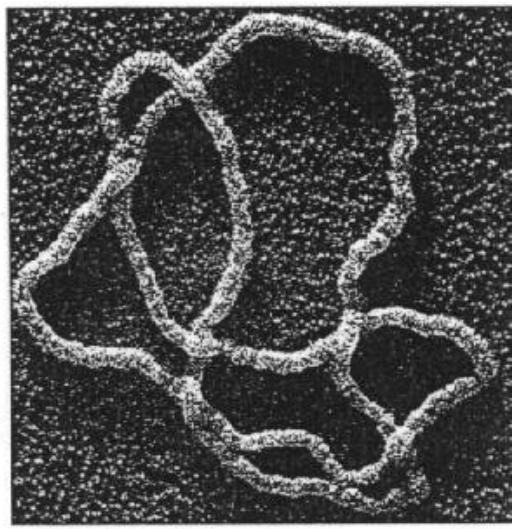


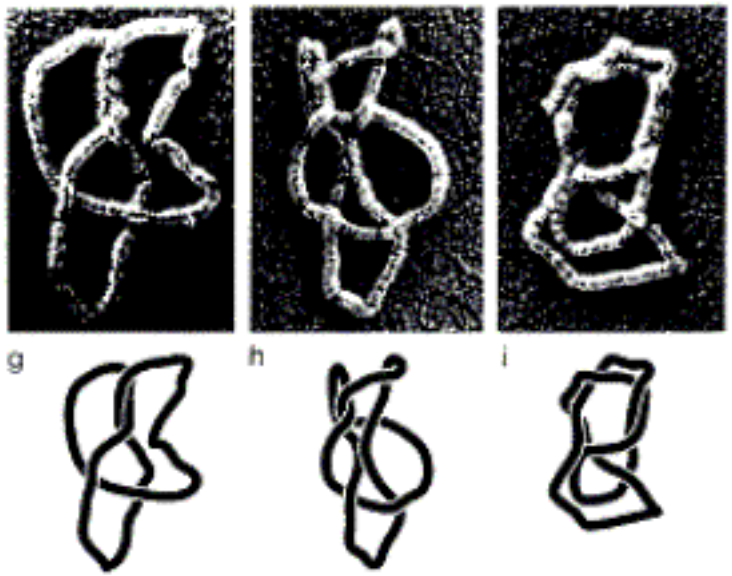
Transition Path Time

Kramers Time

Knotting-Unknotting of DNA, Vortices, Defect Lines, etc… (with C. Micheletti, SISSA, Italy)

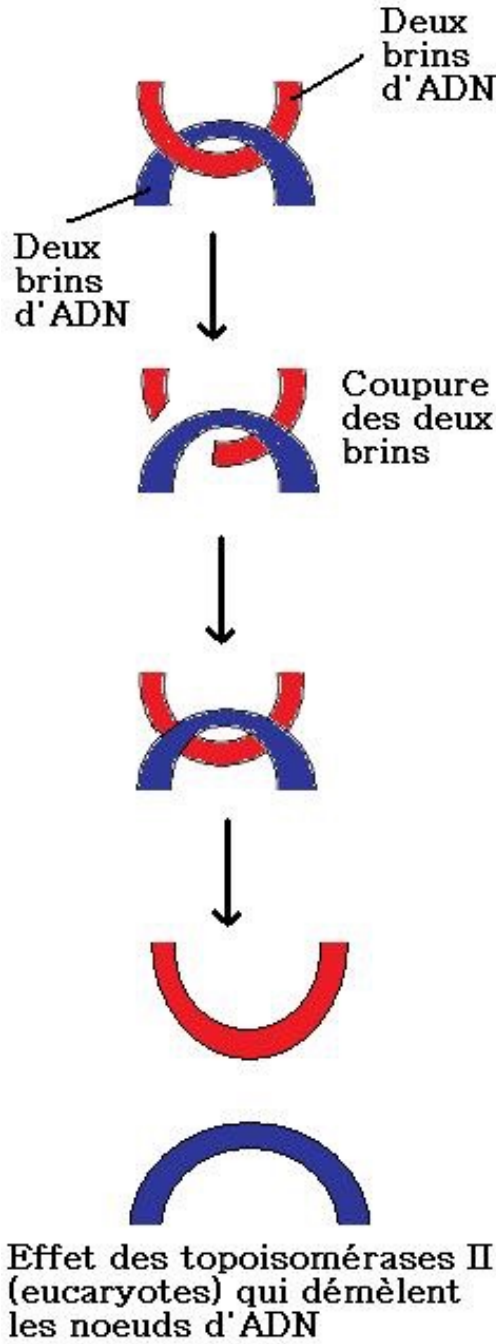
- Many knots in DNA





- Topoisomerase I and II can unknot DNA by passing two DNA strands through each other (cut and reconnect)

- How do knots transform in DNA?
- Mathematicians use a topological approach: what are the minimum sets of moves to go from one knotted structure to another
- This approach ignores the dynamics
- We propose an approach based on bridges



Consider a 1/2 flexible Gaussian chain

$$H = \frac{3}{2a^2}\int_0^N ds\left(\frac{dr}{ds}\right)^2 + \frac{K}{2}\int_0^N ds\left(\frac{d^2r}{ds^2}\right)^2 - \int_0^N F(s)r(s)$$

Elasticity

Bending energy

Langevin-Rouse dynamics

$$\frac{dr(s,t)}{dt} = D\left(\frac{3}{a^2}\frac{\partial^2 r}{\partial s^2} - K\frac{\partial^4 r}{\partial s^4} + \beta F(s)\right) + \eta(s,t)$$

Assume you know the chain configuration $r_0(s)$ at time 0 and the final chain configuration $r_f(s)$ at time t_f

The Langevin Bridge equation is

$$\frac{dr(s,t)}{dt} = D\left(\frac{3}{a^2}\frac{\partial^2 r}{\partial s^2} - K\frac{\partial^4 r}{\partial s^4}\right) + 2D\frac{\partial \log Q}{\partial r(s,t)} - \beta F(s) + \eta(s,t)$$

with

$$Q = P(r_f(s), t_f | r_0(s), 0)$$

Everything can be solved exactly in Fourier space

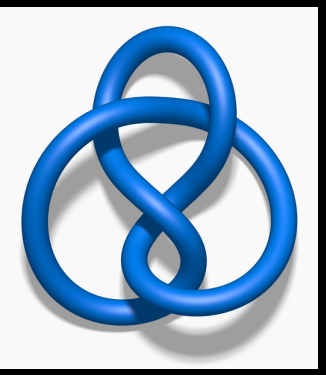
$$\frac{d\tilde{r}_n}{dt} = -D\Omega_n \tilde{r}_n + \frac{D\Omega_n}{\sinh D\Omega_n (t_f - t)} \left(\tilde{r}_n(t_f) - \tilde{r}_n(t) e^{-D\Omega_n (t_f - t)} \right) + \tilde{\eta}_n(t)$$

where

$$\Omega_n = \frac{3}{a^2}\omega_n^2 + K\omega_n^4$$

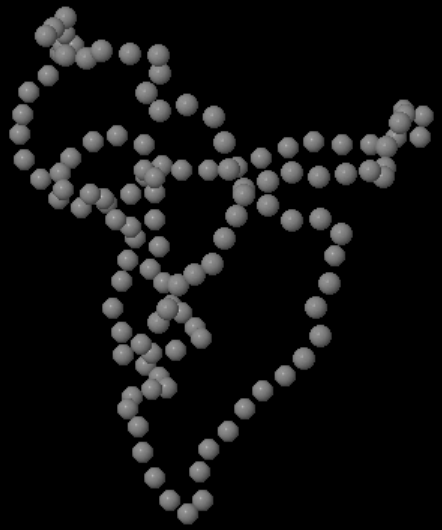
$$\omega_n = \frac{2\pi}{N} n$$

Solve numerically in Fourier space, then go back to real space

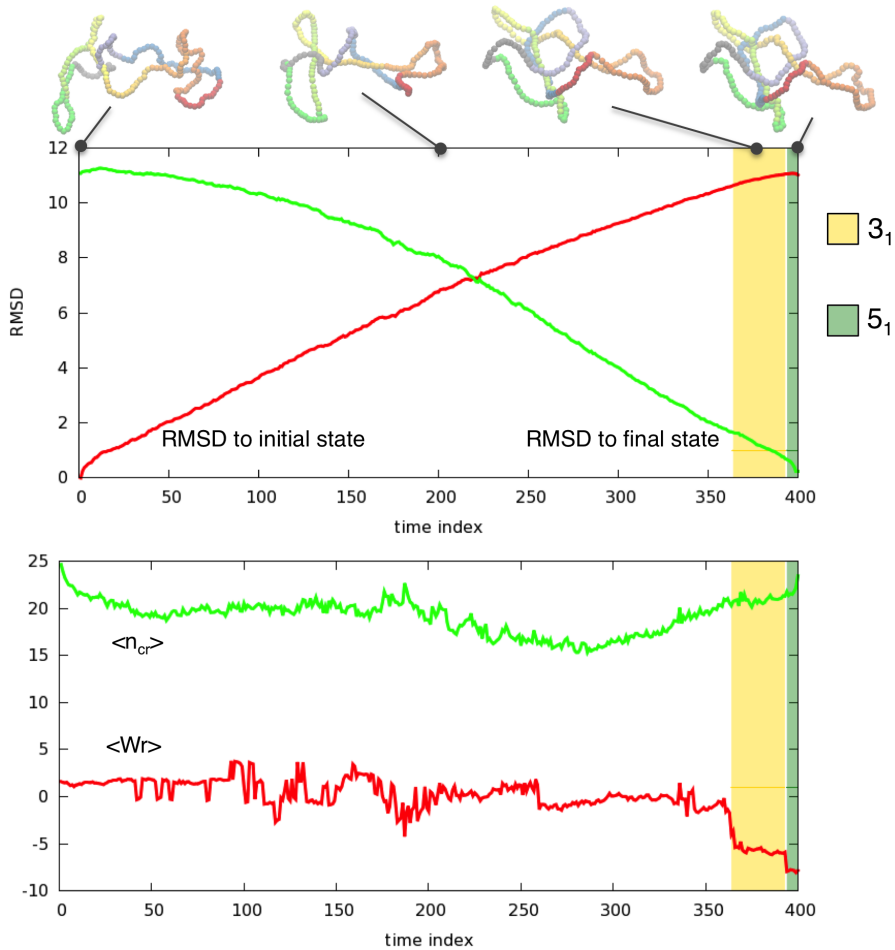


Jmol

Jmol

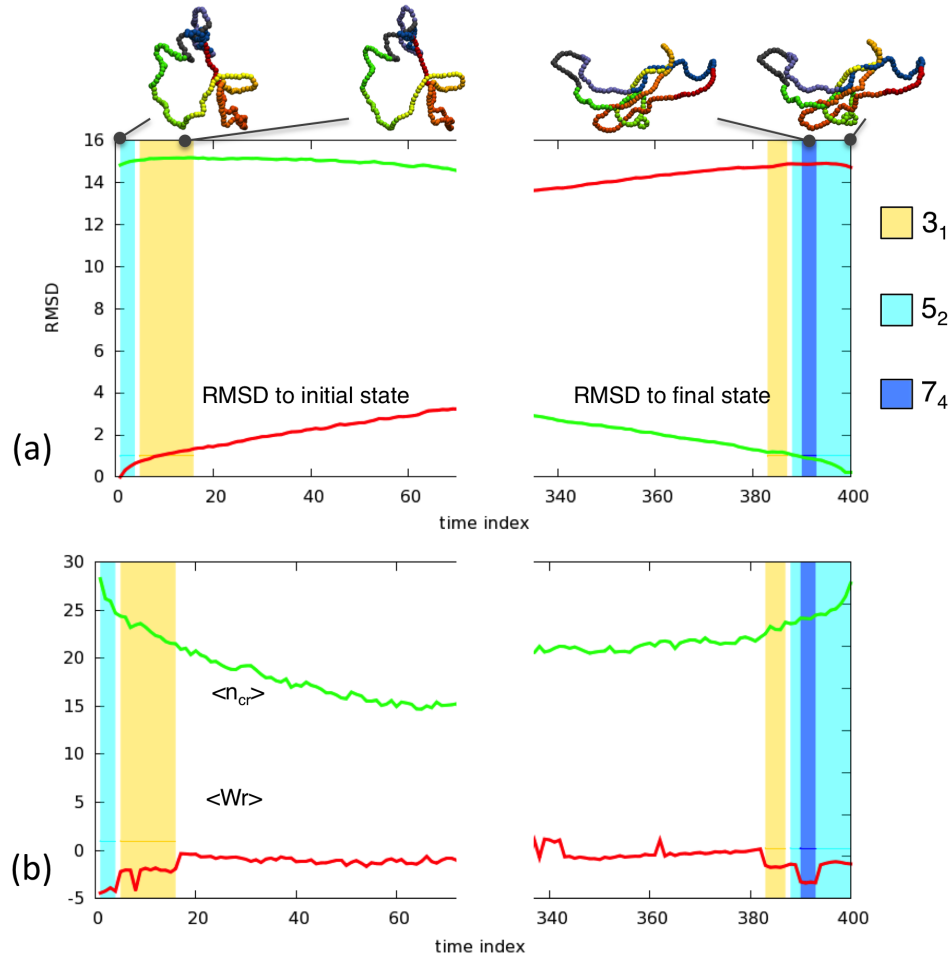


Jmol



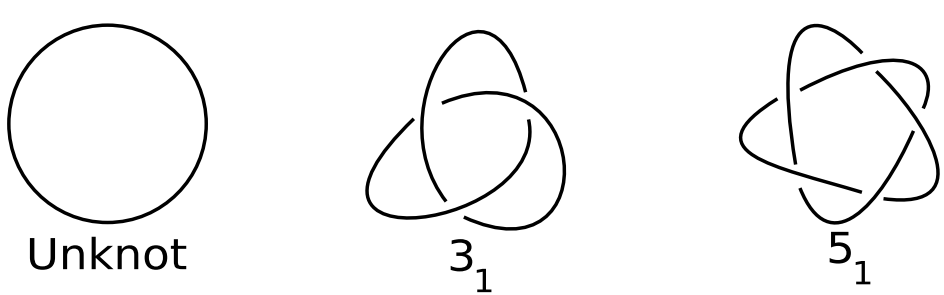
$$0 \to 5_1$$

$$0 \to 3_1 \to 5_1$$



$$5_2 \to 5_2$$

$$5_2 \to 3_1 \to 0 \to 3_1 \to 5_2 \to 7_4 \to 5_2$$



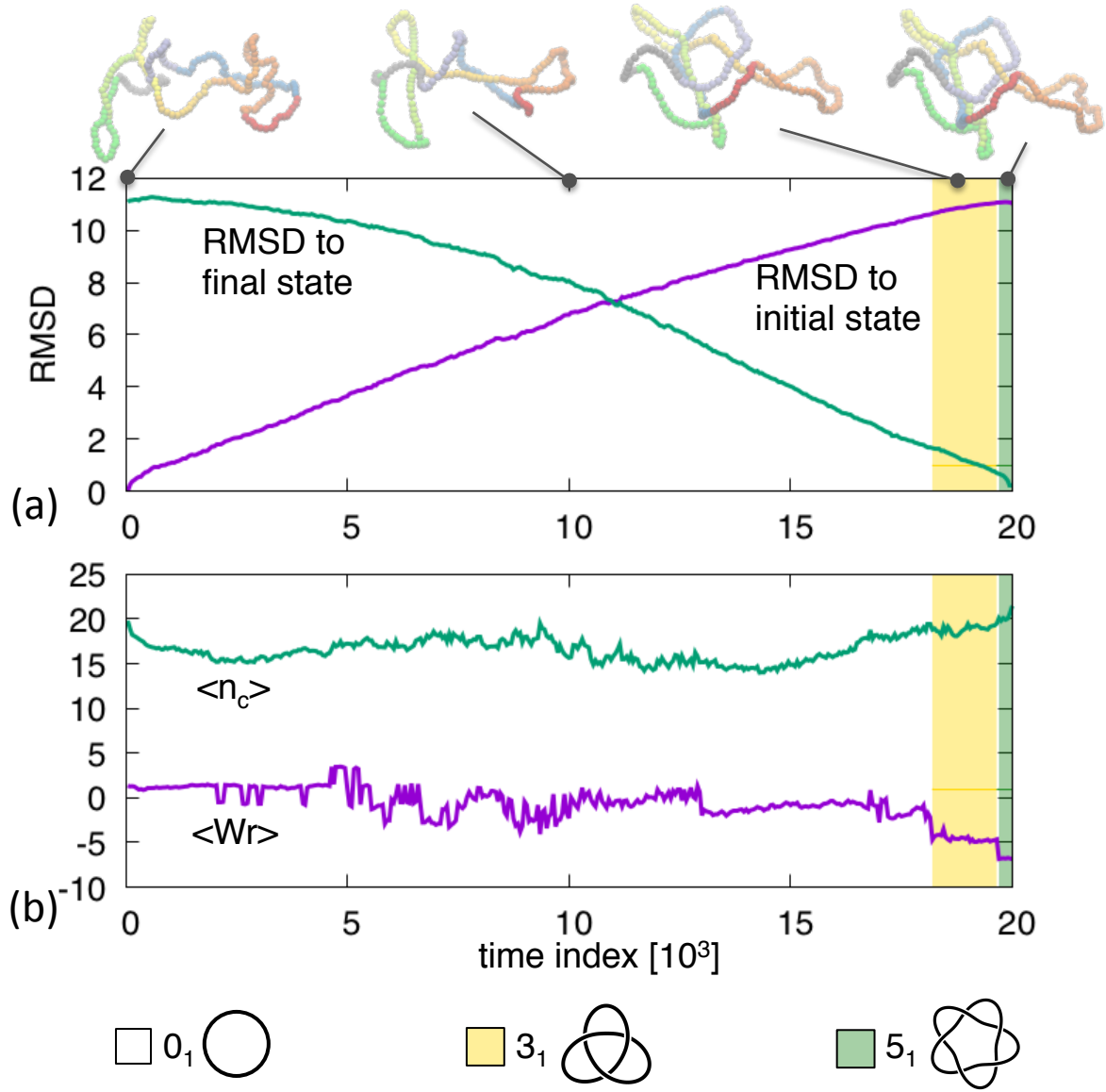


Figure 1. Transition pathway between an unknotted ring and a left-handed 5_1 knotted ring. The root-mean-square distance (RMSD) to the initial and final structures at various stages of the trajectory are shown in panel (**a**). Instantaneous configurations at selected times are highlighted. The average crossing number and writhe are shown in panel (**b**). The overlayed colored background indicates the non-trivial topological states, see legend.

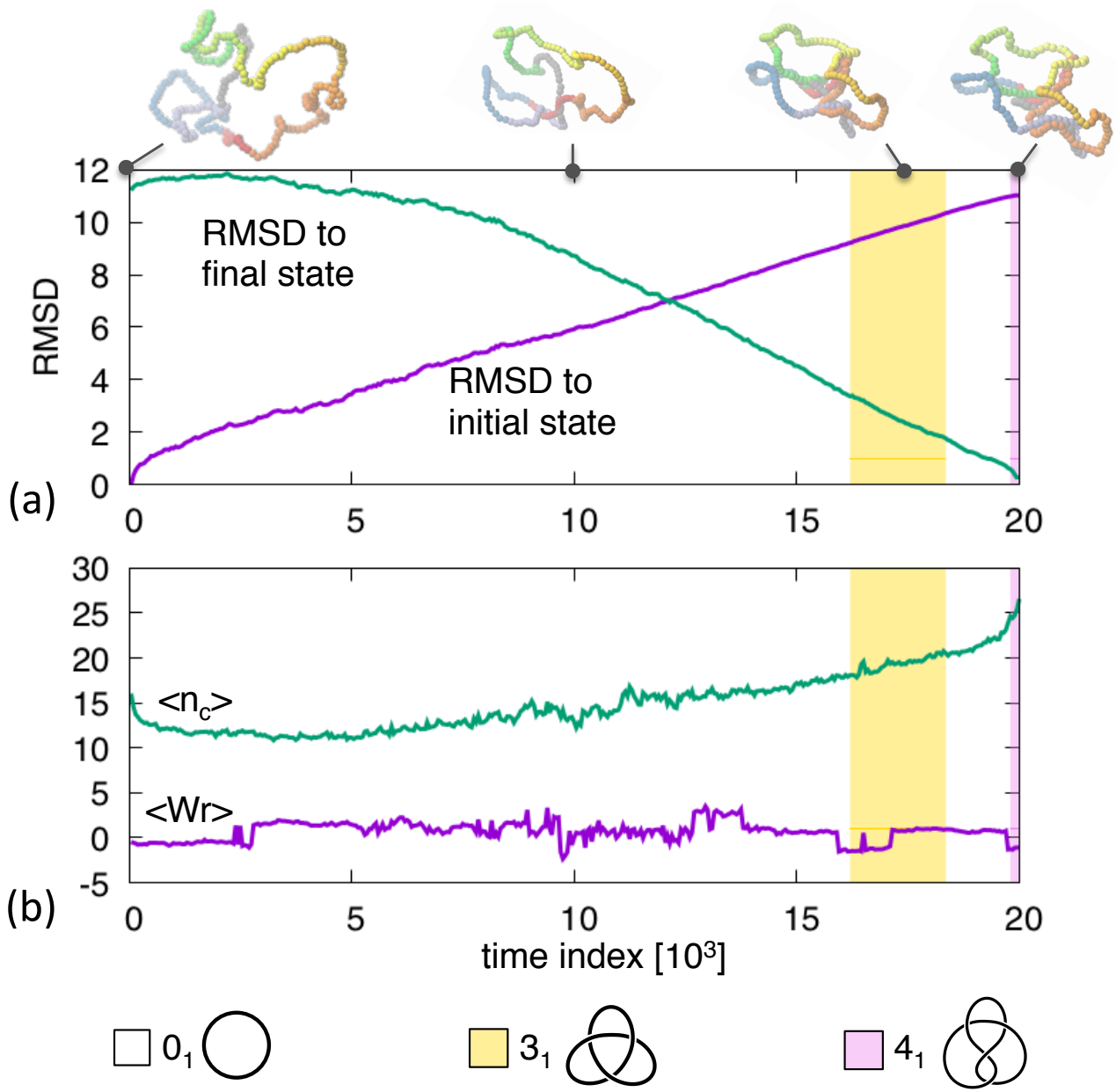


Figure 2. Transition pathway between an unknotted ring and a 4_1 knotted ring. The shown observables are the same as in Figure 1.

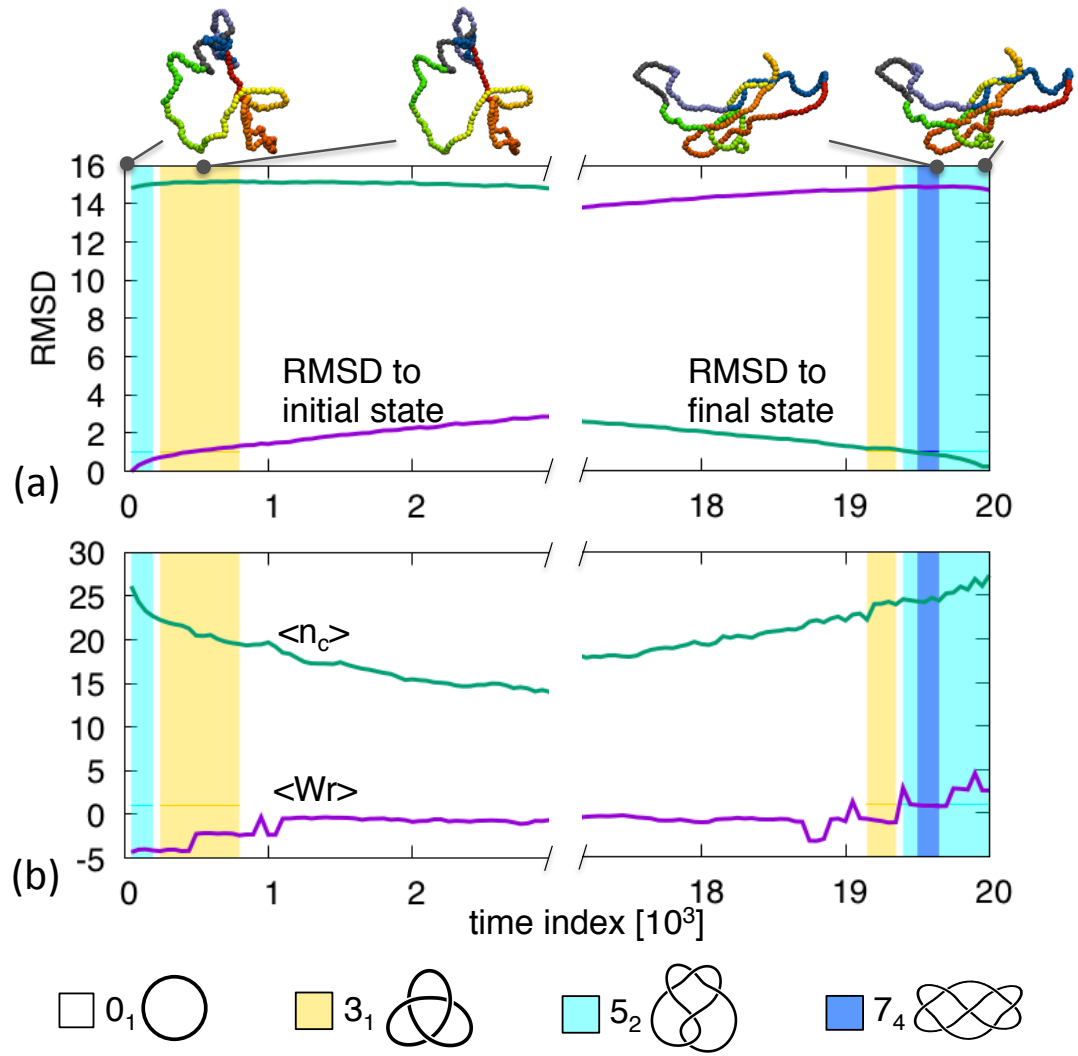


Figure 3. Transition pathway between two 5_2 knotted ring. The shown observables are the same as in Figure 1.

For low-dimensional system, it is possible to compute the function Q by computing the eigenvectors and eigenvalues of the Fokker-Planck operator and using its spectral decomposition. The exact bridge equation becomes

$$
\begin{aligned}
Q(x,t) &= P(x_f, t_f | x, t) \\
&= e^{-\beta(U(x_f)-U(x))/2} \langle x_f | e^{-H(t_f - t)} | x \rangle \\
&= e^{-\beta(U(x_f)-U(x))/2} \sum_\alpha e^{-E_\alpha (t_f - t)} \Psi_\alpha(x_f) \Psi_\alpha(x)
\end{aligned}
$$

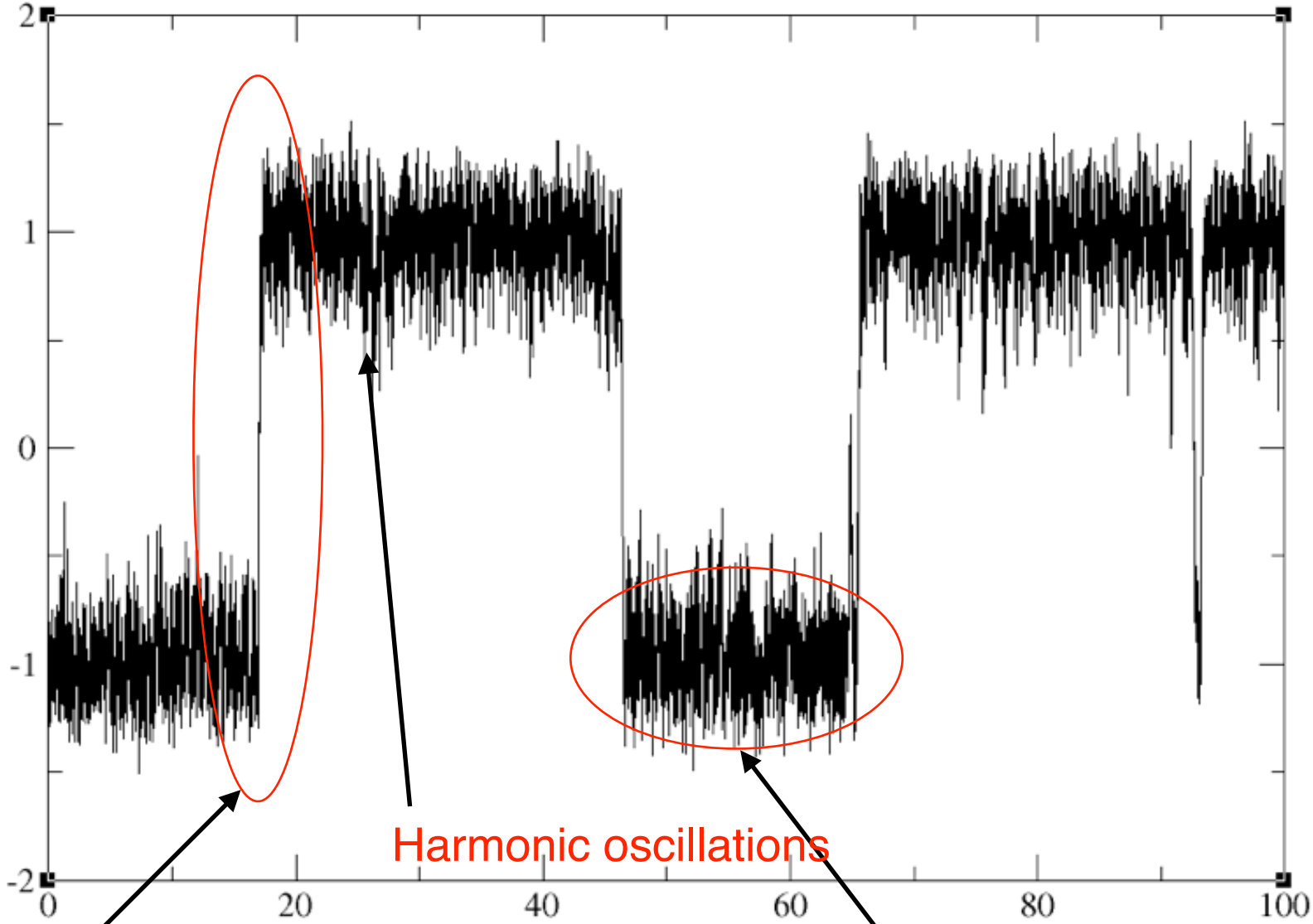
$$
\frac{dx}{dt} = 2D \frac{\sum_\alpha e^{-E_\alpha (t_f - t)} \Psi_\alpha(x_f) \frac{\partial \Psi_\alpha(x)}{\partial x}}{\sum_\alpha e^{-E_\alpha (t_f - t)} \Psi_\alpha(x_f) \Psi_\alpha(x)} + \eta(t)
$$

- For large number of degrees of freedom, we don't know how to calculate the function *Q(x,t)*. We need to make approximations.
- Some important requirements:
  - *Q(x,t)>0*
  - Detailed balance: $Q(x,t) = P(x_f, t_f | x, t)$ and we should have

$$\frac{P(x_f, t_f | x, t)}{P(x, t_f | x_f, t)} = e^{-\frac{U(x_f) - U(x)}{k_B T}}$$

Transition Path Time Distribution

Double Well Potential



Transition Path Time

Kramers Time

Short transition path time approximation

- In the Kramers picture, there are 2 time scales:
  - Kramers time, or waiting time, or folding time

$$\tau_K \approx e^{\frac{\Delta E}{k_B T}}$$ ← 1ms-1s

  - Transition path time (A. Szabo), Instantons

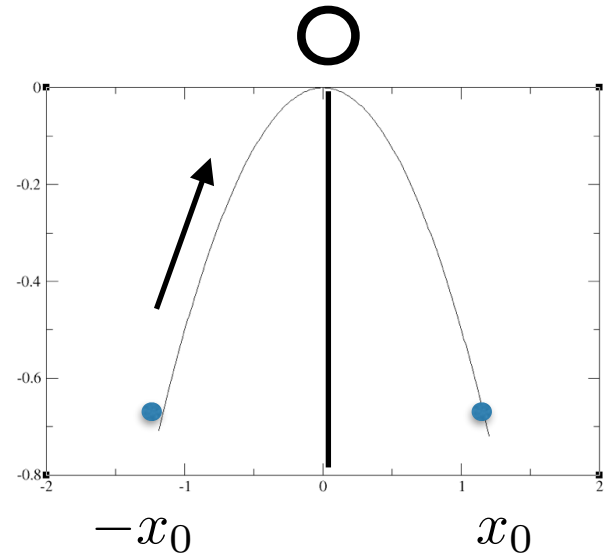
$$\tau_{TP} \approx \log \frac{\Delta E}{k_B T}$$ ← < 1μs

  - We have

$$\tau_{TP} << \tau_K$$

Simple calculation of the TPT distribution

Inverted parabolic potential



$$\frac{dx}{dt} = \frac{1}{\gamma}F(x) + \eta(t)$$

$$F(x) = -\frac{\partial U}{\partial x} = -Kx$$

Equation is linear: can be solved

$$x(t) = x_0 e^{-\Omega t} + \int_0^t d\tau e^{-\Omega(t-\tau)}\eta(\tau)$$

$$\Omega = K\beta D$$

The solution is a linear combination of Gaussians, so it is a Gaussian variable

$$Q_{\mathrm{A}}(t) \equiv \int_{x_0}^{\infty} \mathrm{d}x\ \mathrm{P}(x,t|-x_0,0)$$

Absorption Probability

$$= \langle \theta(t - t_{\mathrm{TP}}) \rangle$$

$$p_{\mathrm{TP}}(t) = \frac{dQ_{\mathrm{A}}}{dt}$$

$$p_{\mathrm{TP}}(t) = \frac{dQ_{\mathrm{A}}}{dt}$$

This equation neglects paths which return to the left side.
Very good approximation for high barrier

$$p_{\mathrm{TP}}(t) = -\frac{2}{\sqrt{\pi}} \frac{G'(t)e^{-G^2(t)}}{1 - \mathrm{Erf}(\sqrt{\beta E})}$$

Paths which never cross

$$G^2(t) = \beta E \frac{1 + \exp(-\Omega t)}{1 - \exp(-\Omega t)}$$

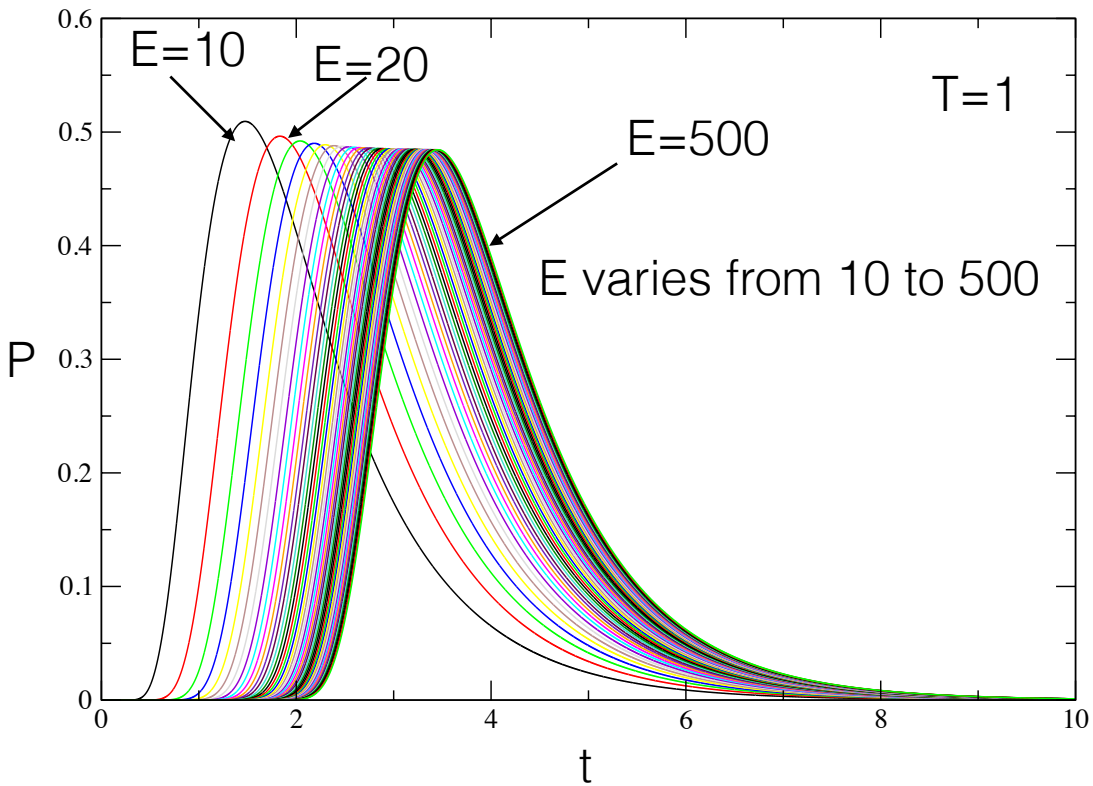
$$\Omega \equiv K/\gamma$$

$$p_{\mathrm{TPT}}(t) = \sqrt{\frac{\beta E}{\pi}} \frac{2\Omega e^{-2\Omega t}}{\left(1-e^{-2\Omega t}\right)^{3/2}} \frac{\exp\left[-\beta E\left(1-e^{-2\Omega t}\right)^{-1}\right]}{1-\mathrm{Erf}\left(\sqrt{\beta E}\right)}$$

where $\Omega = \beta D |U_0"|$

$$e^{-2z} = 1 - e^{-2\Omega t} \qquad \mathcal{P}_{TPT}(z) = 2\sqrt{\frac{\beta E_0}{\pi}} \frac{1}{1-\mathrm{Erf}\left(\sqrt{\beta E_0}\right)} e^{z-\beta E_0 e^{2z}}$$

Gumbel law



$$t_{TPT} = \frac{\gamma}{2|U_0''|} \log\left(e^C \beta E_0\right)$$

$$C = 0.57721$$

Euler constant

It is possible to do the calculation with inertia term
Expressions are more complicated when inertia present

$$p_{\mathrm{TP}}(t) \underset{t\to 0}{\sim} e^{-G^2(t)} = \exp\left(-\frac{2\beta m x_0^2}{t^2}\right)$$

$$p_{\mathrm{TP}}(t) \sim_{t\to\infty} \exp\left(-\lambda t\right)$$

$$\lambda = \frac{-\gamma + \sqrt{\gamma^2 + 4Km}}{2m}$$

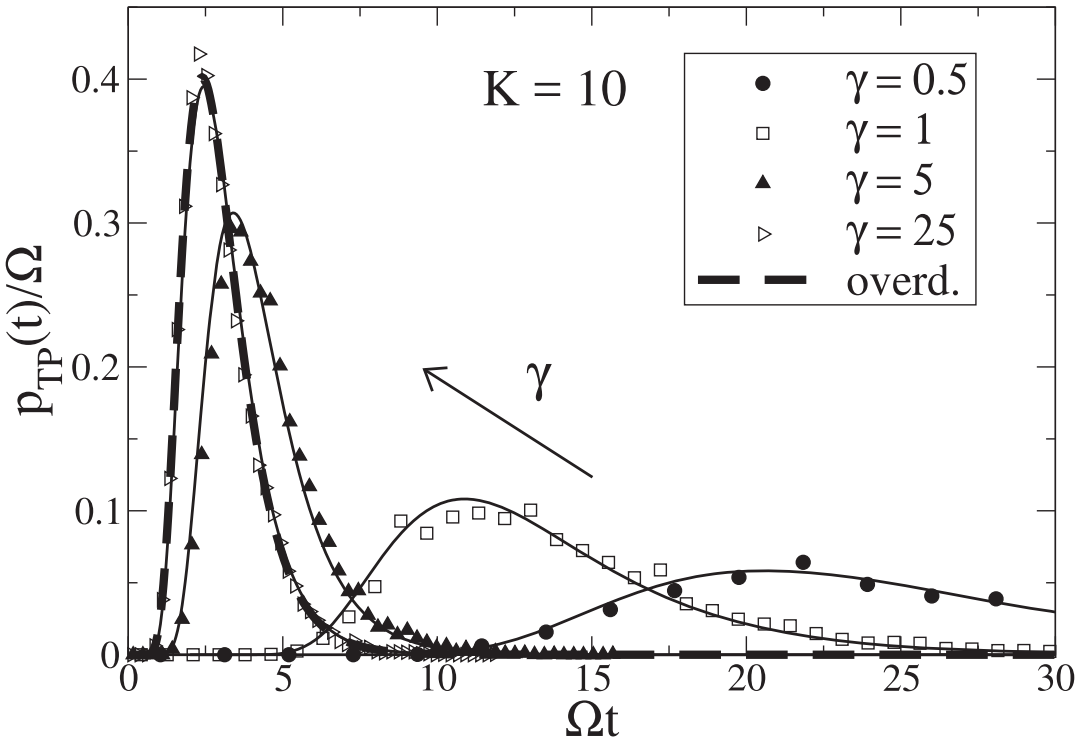
Underdamped

$$p_{\mathrm{TP}}(t) \sim_{t\to 0} \exp\left(-\frac{\beta\gamma x_0^2}{t}\right)$$

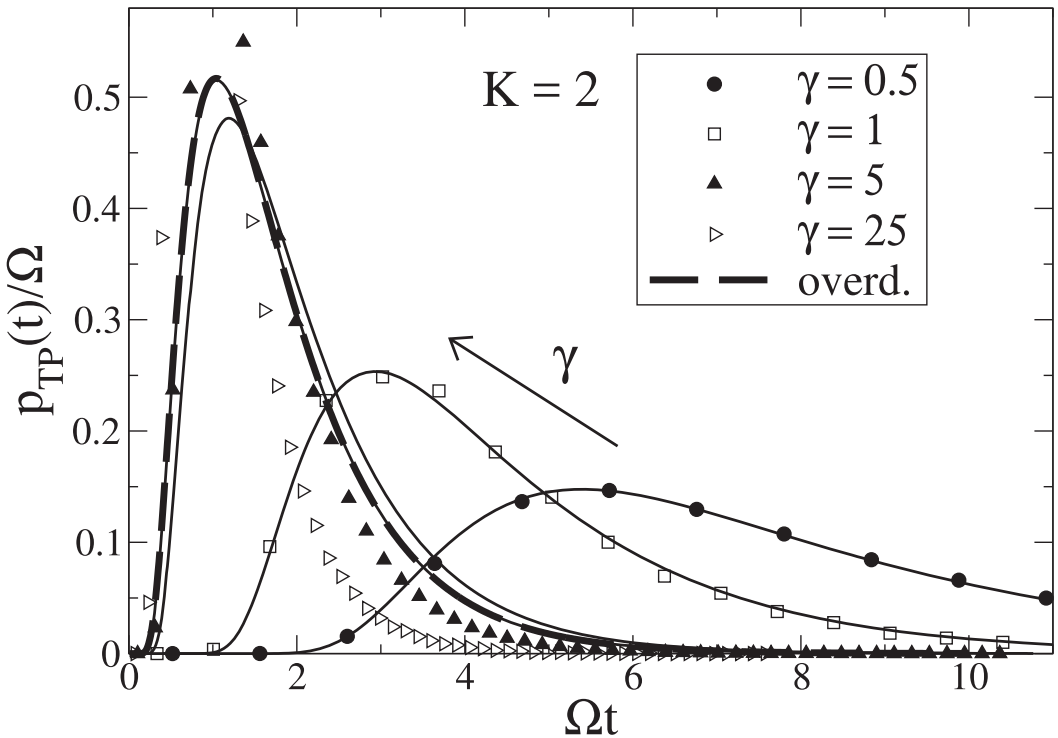
$$p_{\mathrm{TP}}(t) \sim_{t\to\infty} \exp\left(-\Omega t\right)$$

$$\Omega \equiv K/\gamma$$

Overdamped



High Barrier



Low Barrier

A useful form of the Bridge equation

Original bridge equation

$$\dot{\mathbf{r}} = -\frac{1}{\gamma}\nabla U + 2D\nabla \ln Q_{\cdot}(\mathbf{r},t) + \boldsymbol{\eta}(t)$$

Onsager-Machlup

Using the path-integral representation

$$\begin{aligned} Q_{\cdot}(\mathbf{r},t) &= P(\mathbf{r}_f, t_f|\mathbf{r},t) \\ &= \int_{\mathbf{r}(t)=\mathbf{r}}^{\mathbf{r}(t_f)=\mathbf{r}_f} \mathcal{D}\mathbf{r}(\tau) e^{-\frac{1}{4D}\int_t^{tf} d\tau \left(\dot{\mathbf{r}}+\frac{1}{\gamma}\nabla U\right)^2} \\ &= e^{-\beta(U(\mathbf{r}_f)-U(\mathbf{r}))/2} \int_{\mathbf{r}(t)=\mathbf{r}}^{\mathbf{r}(t_f)=\mathbf{r}_f} \mathcal{D}\mathbf{r}(\tau) e^{-\int_t^{tf} d\tau \left(\frac{\dot{\mathbf{r}}^2}{4D}+\frac{1}{D\gamma^2}V(\mathbf{r}(\tau))\right)} \end{aligned}$$

Stratonovich

where the effective potential is given by

$$V(\mathbf{r}) = \frac{1}{4}(\nabla U)^2 - \frac{k_B T}{2}\nabla^2 U$$

After some simple transformations, the bridge equation becomes:

$$\dot{\mathbf{r}} = \frac{\mathbf{r}_f - \mathbf{r}(t)}{t_f - t} - \frac{2}{\gamma^2}\int_t^{t_f} d\tau \left(\frac{t_f - \tau}{t_f - t}\right) \langle \nabla V(\mathbf{r}(\tau))\rangle_Q + \sqrt{\frac{2k_B T}{\gamma}}\xi(t)$$

$$V(\mathbf{r}) = \frac{1}{4}(\nabla U)^2 - \frac{k_B T}{2}\nabla^2 U$$

where the bracket $\langle \cdots \rangle_Q$ denotes the average over all paths joining $(\mathbf{r}, t)$ to $(\mathbf{r}_f, t_f)$ weighted by Q.

- If no potential V=0, one recovers the standard free Brownian bridge equation
- The only singular term is the free Brownian bridge term. It is the term which conditions the paths
- Highly non-Markovian and depends on the future of the trajectory

Zero Temperature

At T=0, only one trajectory and the equation becomes

$$\dot{\mathbf{r}}_0 = \frac{\mathbf{r}_f - \mathbf{r}_0(t)}{t_f - t} - \frac{2}{\gamma^2}\int_t^{t_f} d\tau \left(\frac{t_f - \tau}{t_f - t}\right) \nabla V_0(\mathbf{r}_0(\tau))$$

where $V_0(\mathbf{r}) = \frac{1}{4}\left(\nabla U\right)^2$

taking a time derivative of the above equation we get

$$\boxed{\ddot{\mathbf{r}}_0 = \frac{2}{\gamma^2}\nabla V_0(\mathbf{r}_0) = \frac{1}{\gamma^2}\nabla U(\mathbf{r}_0).\nabla^2 U(\mathbf{r}_0)}$$

boundary conditions $\mathbf{r}_0(0) = \mathbf{r}_i$ and $\mathbf{r}_0(t_f) = \mathbf{r}_f$

This is the exact zero temperature equation: Instanton Theory and String Method.

Weak Fluctuations

If fluctuations of trajectories are small (low T or Transition paths), use the approximation

$$\langle \nabla V(\mathbf{r}(\tau)) \rangle_Q \approx \nabla V(\langle \mathbf{r}(\tau) \rangle_Q)$$

The Bridge equation becomes

$$\dot{\mathbf{r}} = \frac{\mathbf{r}_f - \mathbf{r}(t)}{t_f - t} - \frac{2}{\gamma^2} \int_t^{t_f} d\tau \left(\frac{t_f - \tau}{t_f - t} \right) \nabla V(\mathbf{r}(\tau)) + \boldsymbol{\eta}(t)$$

Not Markovian.
Valid at order 1 in T

Iterative Solution

This is a Non-Linear, Integro-Differential, non Markovian Stochastic Equation

Solve it by iteration. Discretize à la Euler-Maruyama

$$\mathbf{r}^{(n+1)}(k+1) = \mathbf{r}^{(n)}(k)dt + \frac{\mathbf{r}_f - \mathbf{r}^{(n)}(k)}{t_f - kdt}dt - \frac{2}{\gamma^2}\sum_{k'=k}^{I-1}\left(\frac{t_f - k'dt}{t_f - kdt}\right)dt\nabla V(\mathbf{r}^{(n)}(k')) + \sqrt{(2Ddt)}\boldsymbol{\xi}(k)$$

$$\langle \xi_a(k) \rangle = 0$$
$$\langle \xi_a^2(k) \rangle = 1$$

Choose an initial trajectory $\{\mathbf{r}^{(0)}(k)\}$

Iterate above equation

Choice of Initial Trajectory

- Use a free Brownian bridge trajectory

$$\dot{r} = \frac{r_f - r}{t_f - t} + \eta(t)$$

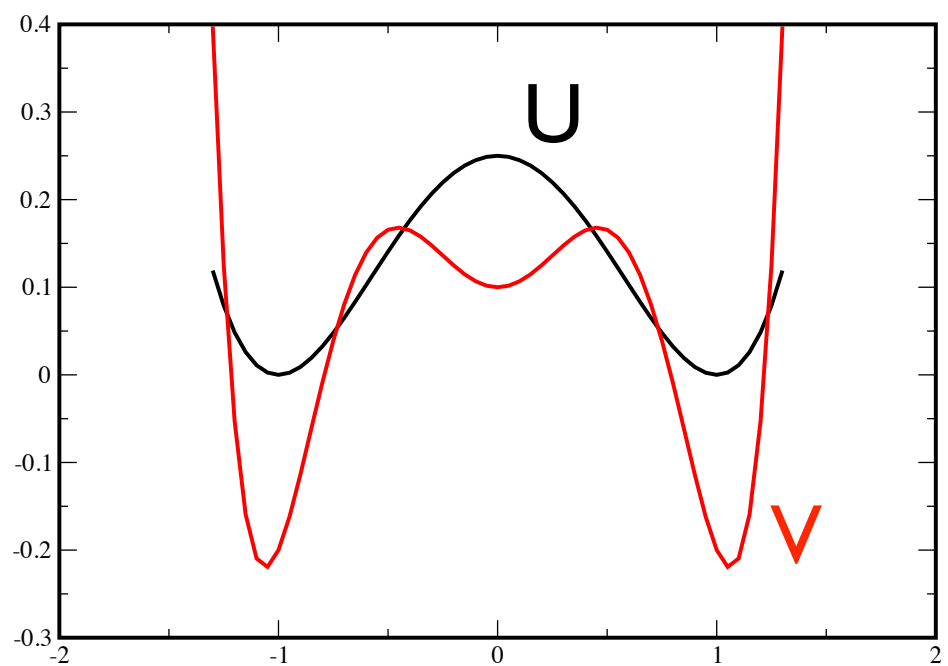
- Iterate from there

Example: Quartic double well

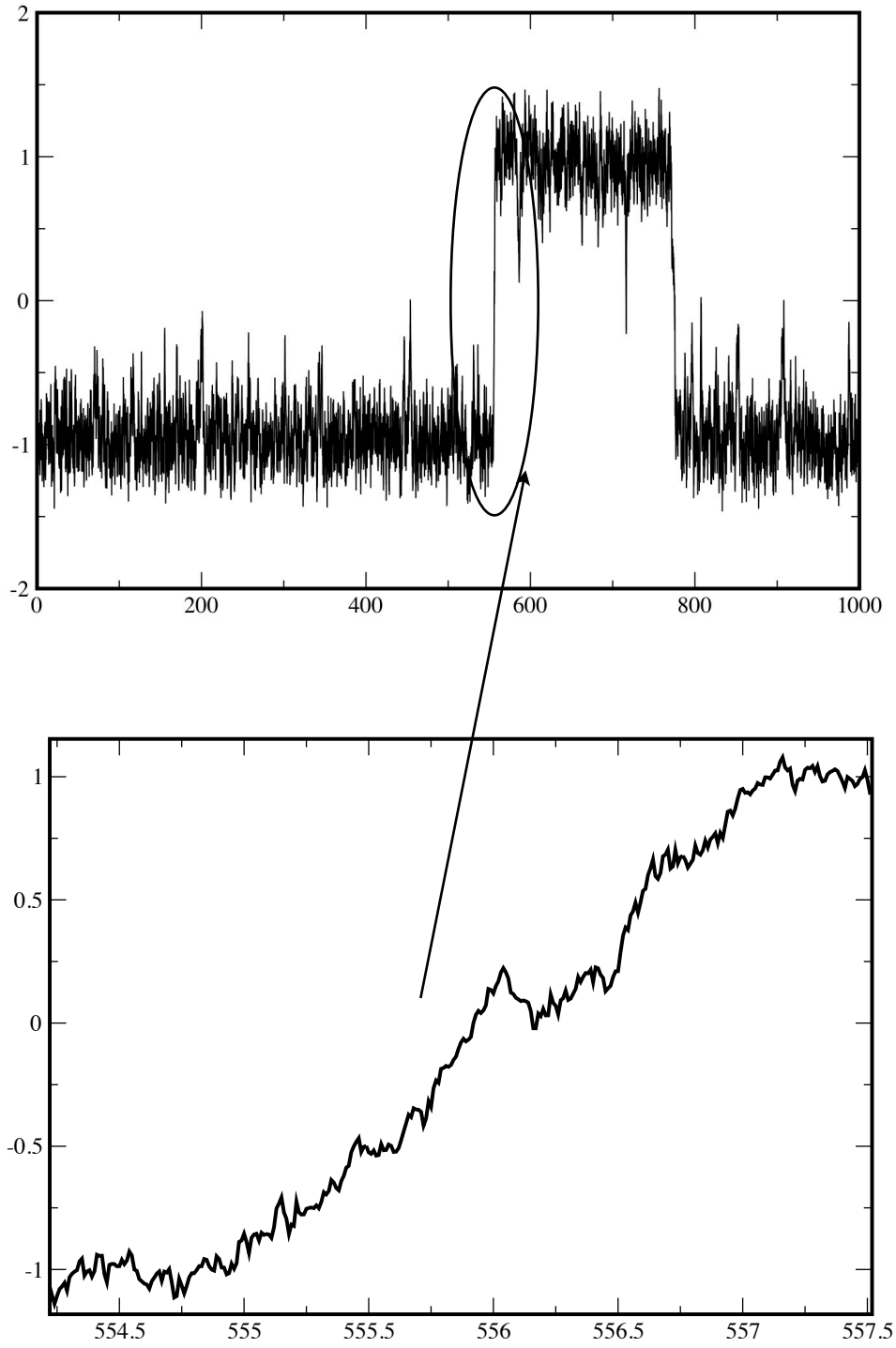
- We take

$$U(x) = \frac{1}{4}(x^2 - 1)^2$$

$$V(x) = \frac{1}{4k_BT}(U'^2(x) - 2k_BTU''(x))$$



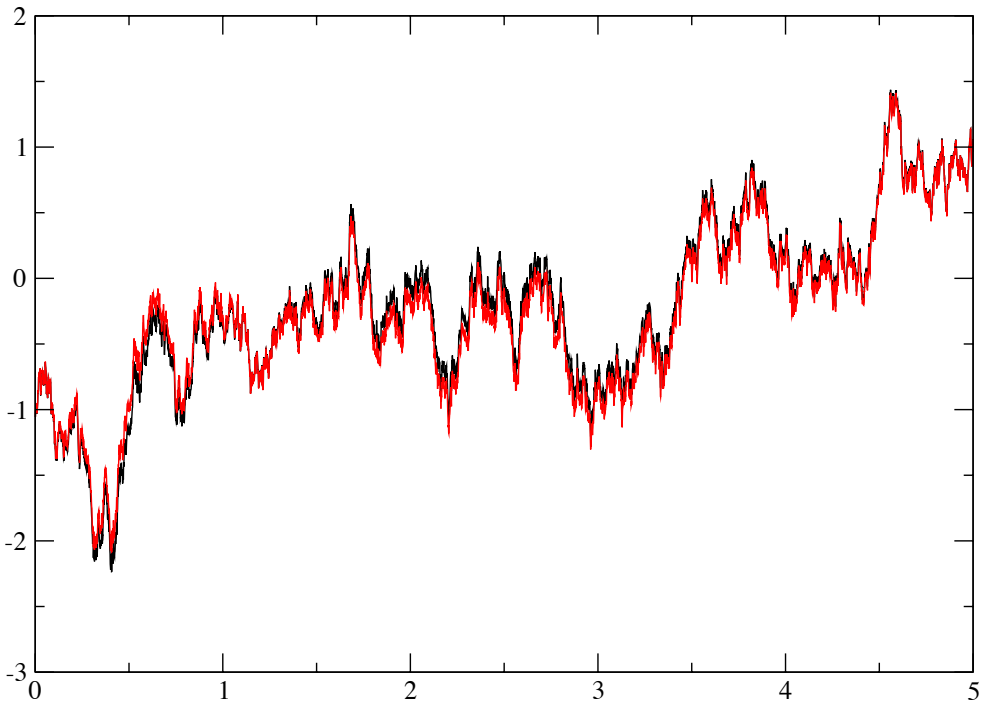
$T = 0.05$

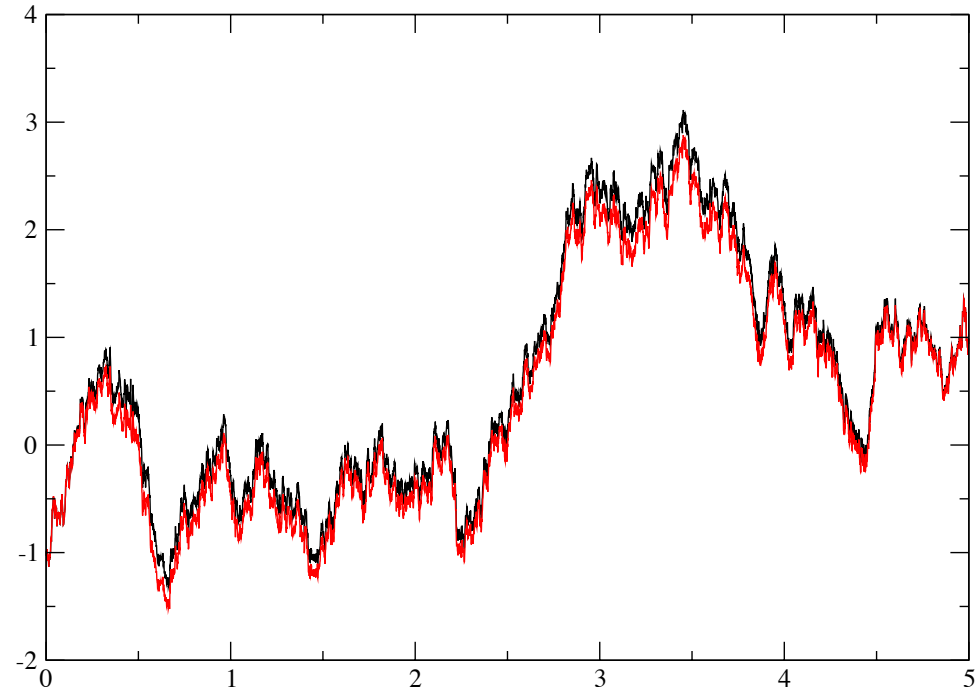


Langevin

transition region

Trajectories





$$\tau_{TP} \approx 3.5$$

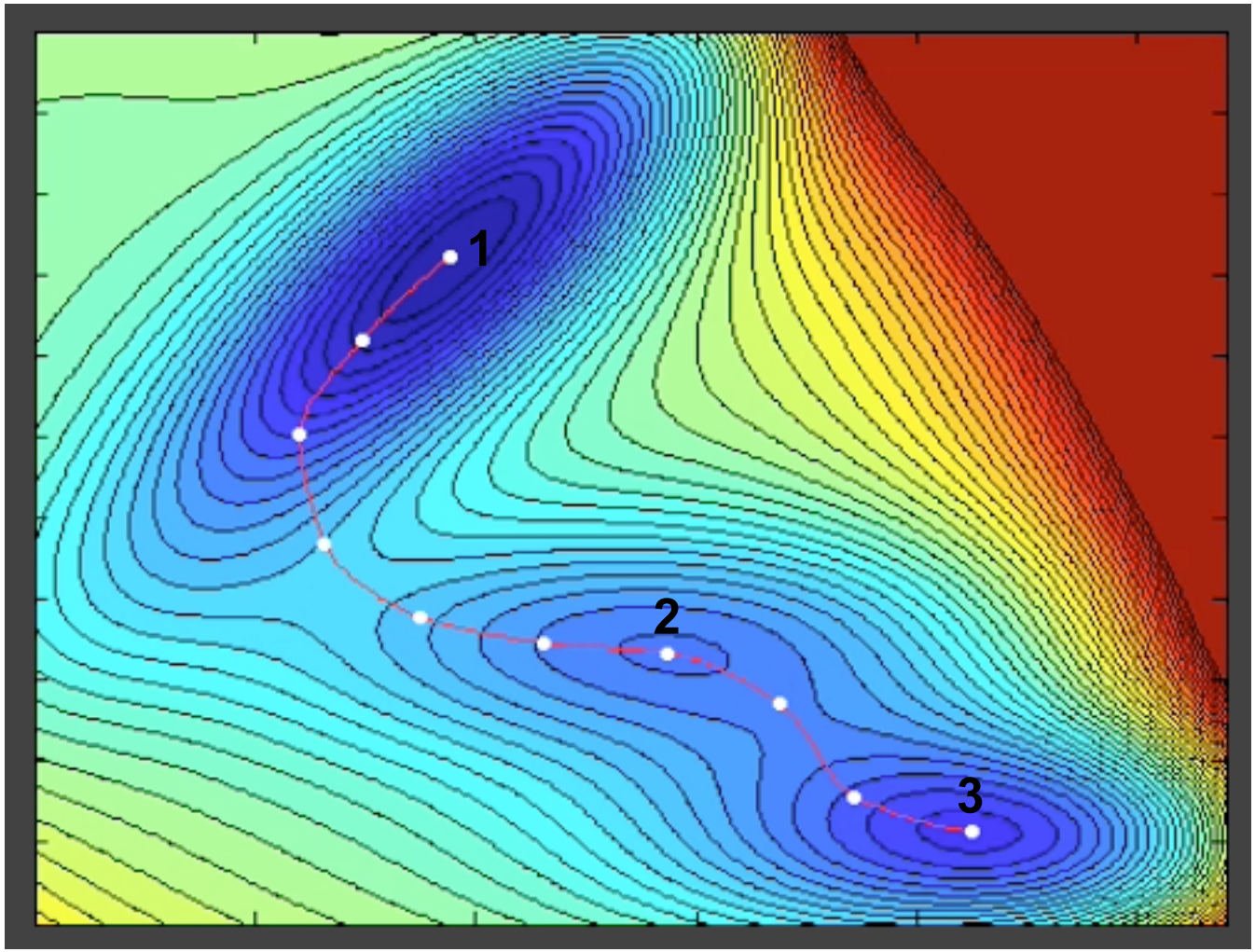
Exact trajectories (in black) obtained from spectral decomposition.

Approximate trajectories in red (with same noise)

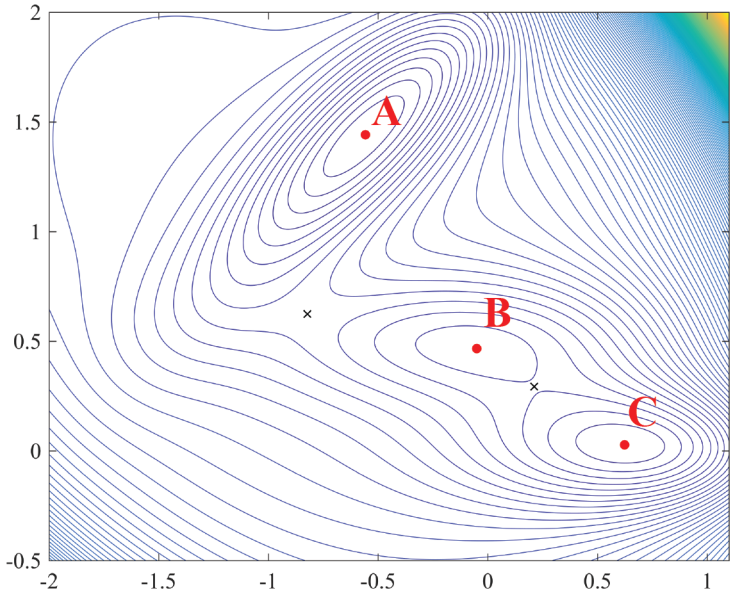
The Mueller potential

$$V(x,y) = \sum_{i=1}^{4} A_i \exp\left[a_i(x-x_i^0)^2 + b_i(x-x_i^0)(y-y_i^0) + c_i(y-y_i^0)^2\right]$$

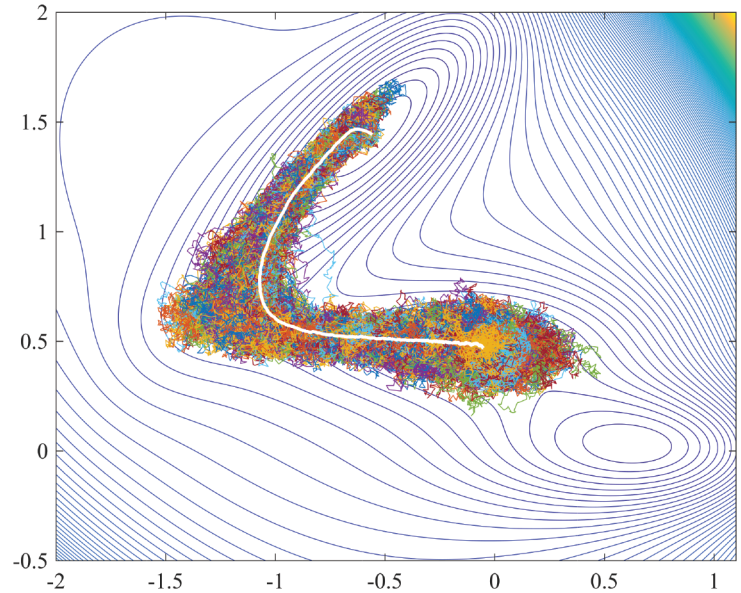
where $A = (-200, -100, -170, 15)$, $a = (-1, -1, -6.5, 0.7)$, $b = (0, 0, 11, 0.6)$,



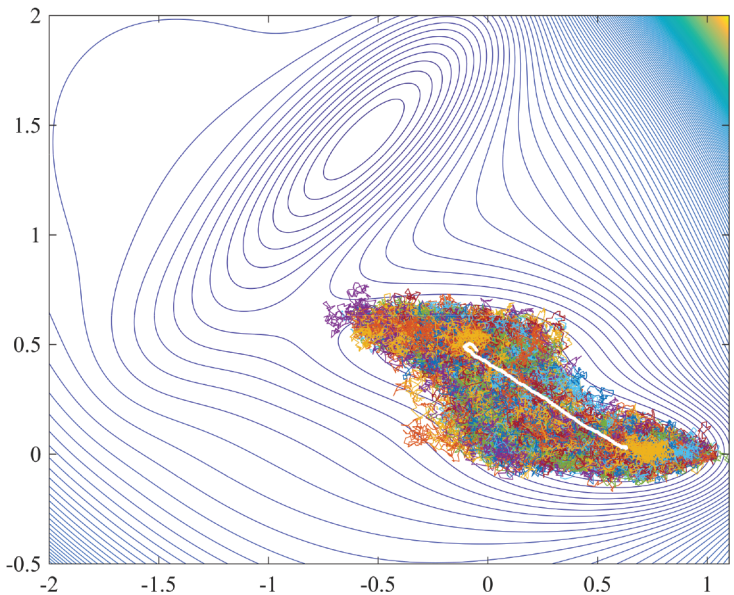
A) The Mueller potential



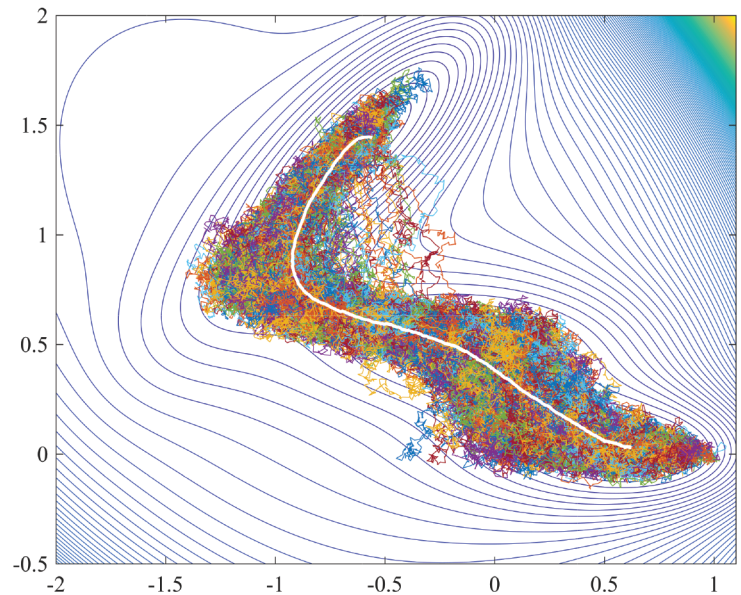
B) Traiectories A⟶B



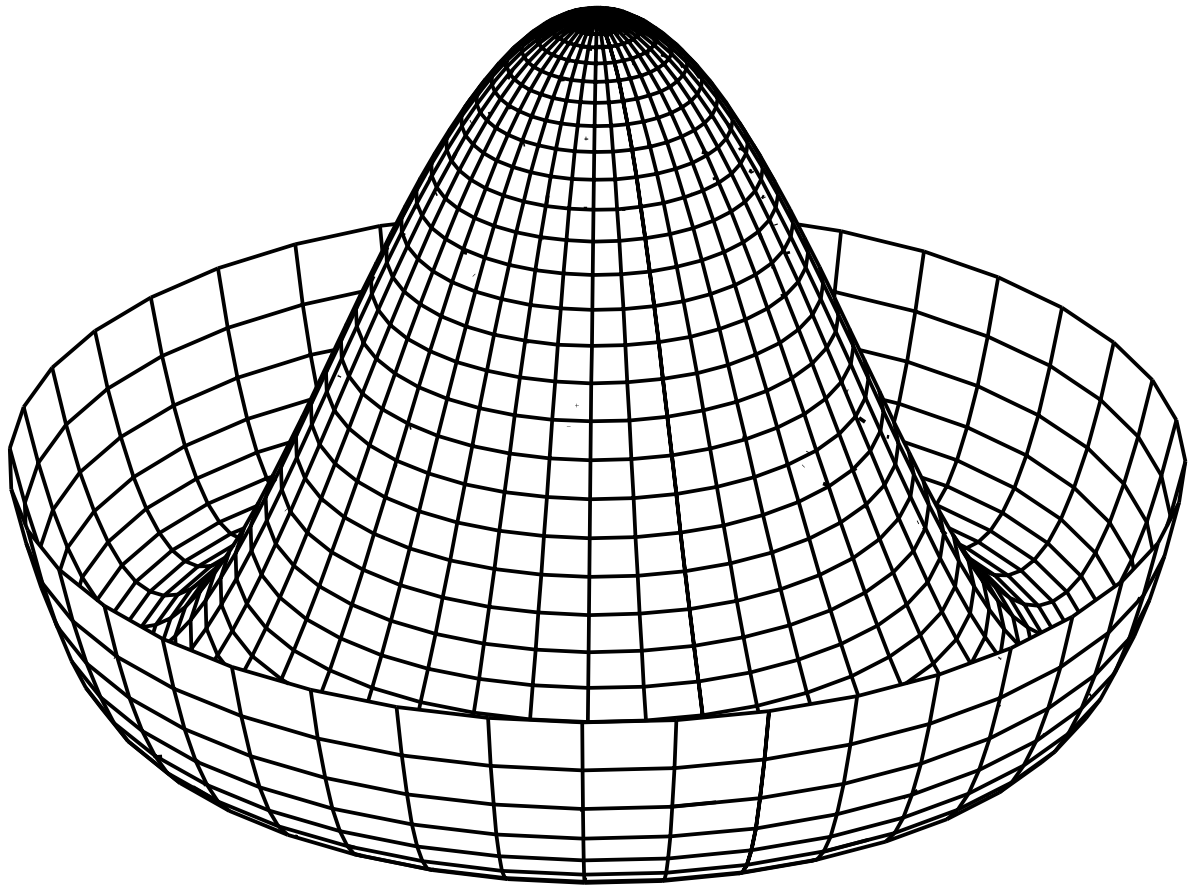
C) Trajectories B⟶C



D) Trajectories A⟶C

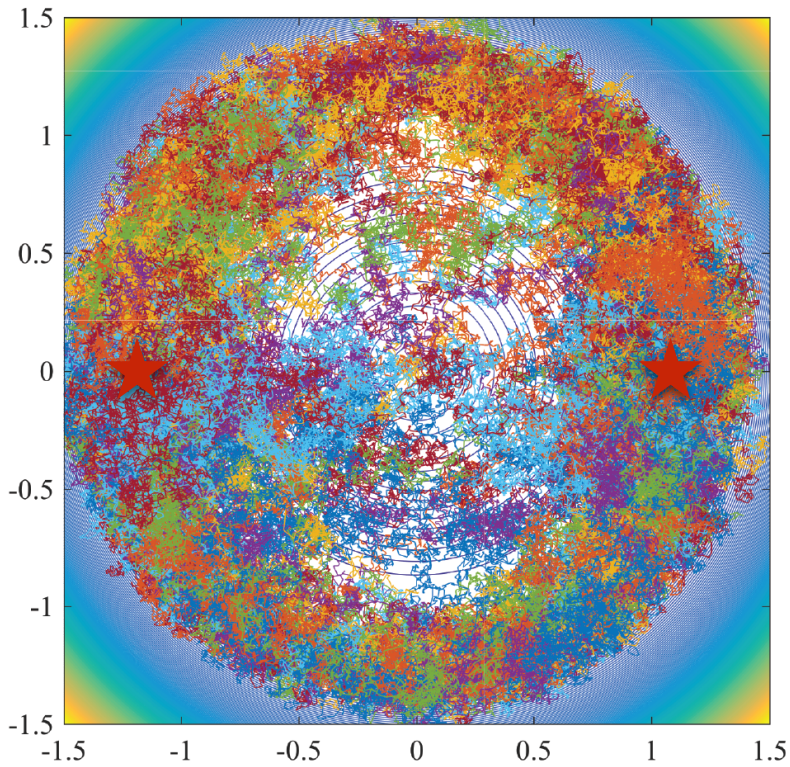


Mexican Hat Potential

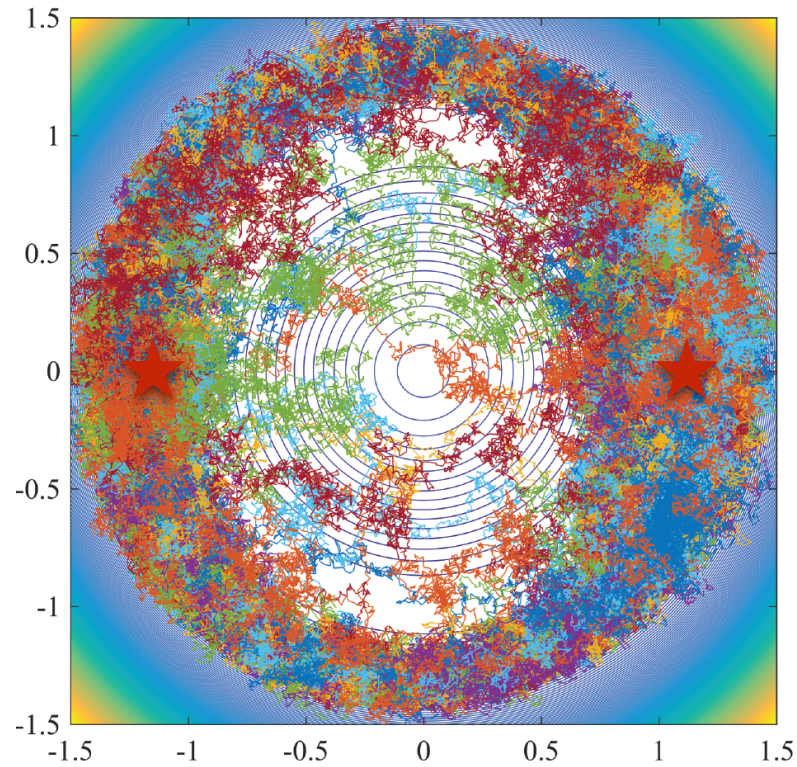


$$U(x,y) = \frac{1}{4}(x^2 + y^2 - 1)^2$$

A) 100 paths with T = 0.1 and $t_f = 7$



B) 100 paths with T = 0.1 and $t_f = 10$



Langevin bridge trajectories on the Mexican hat potential.

$$U(x, y) = \frac{1}{4}(x^2 + y^2 - 1)^2$$

Test of the method: Allosteric transition of Adenylate Kinase

- Work with P. Koehl (UC Davis) and M. Delarue (Institut Pasteur, Paris)
- Transition studied by several groups: good benchmark
- System has two (meta)stable configurations PDB 1AKE.pdb and 4AKE.pdb
- Make a Gaussian Elastic Network model for each structure

- Define mixed elastic network model by

$$U_{tot} = U_{Mix-ENM} + U_{collision}$$

$$U_{Mix-ENM} = -\frac{1}{\beta_m} log(e^{-\beta_m U_A} + e^{-\beta_m U_B})$$

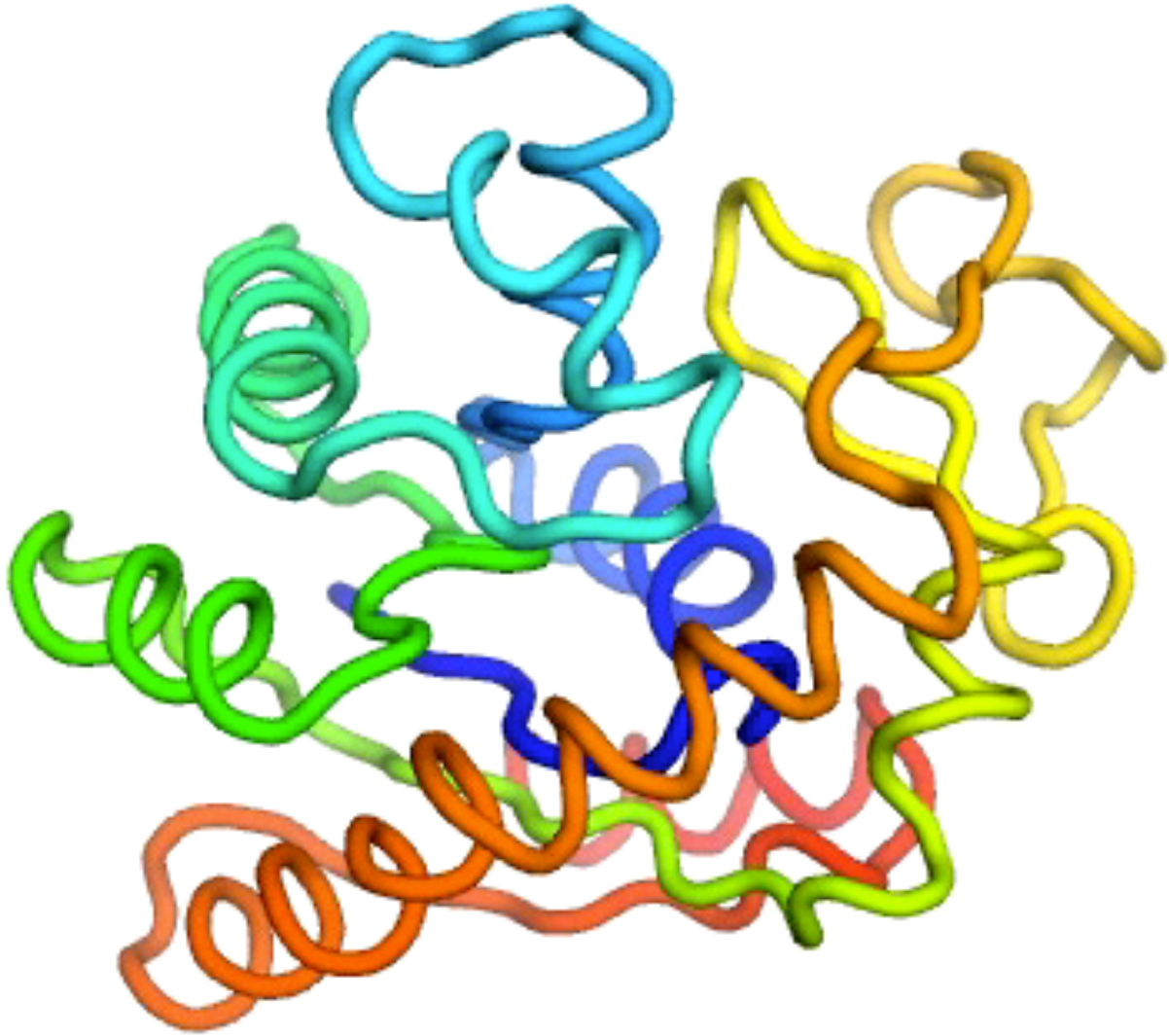
$$U_A = \sum_{ij} k_{ij} C_{ij} (d_{ij} - d_{ij}^A)^2$$

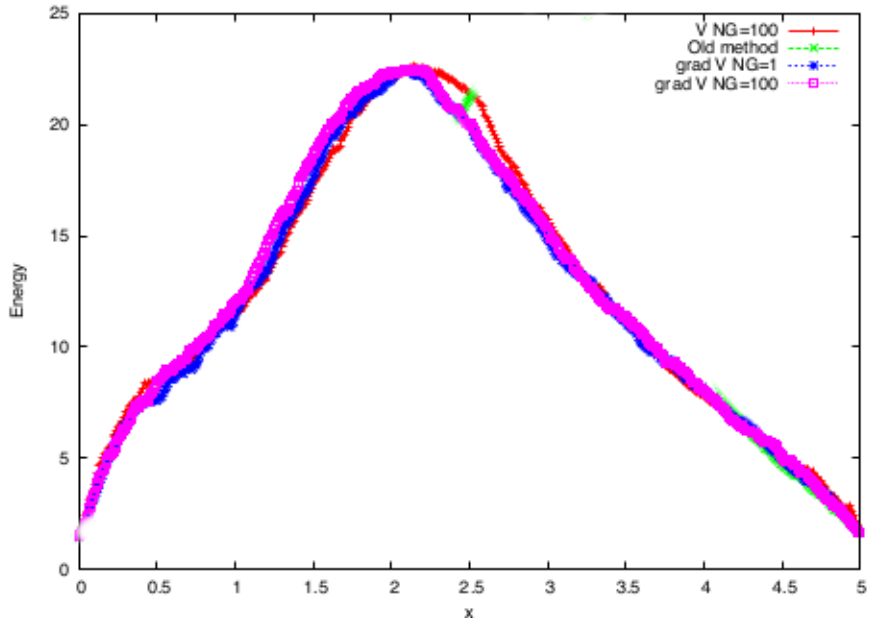
$$U_B = \sum_{ij} k_{ij} C_{ij} (d_{ij} - d_{ij}^B)^2 + \Delta U$$

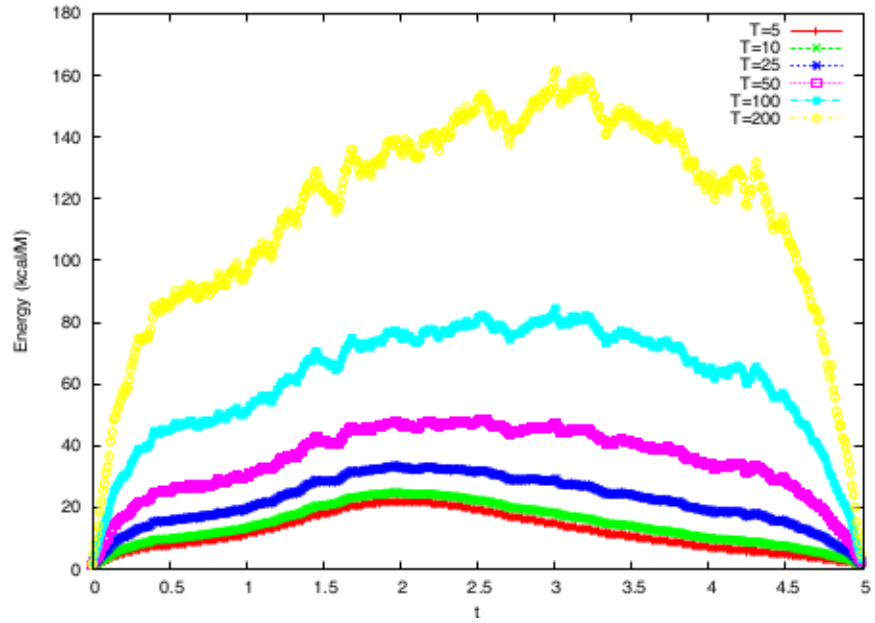
- Collision term to avoid steric clashes during the transition

$$U_{collision} = \epsilon \sum_{i,j} (\frac{\sigma}{d_{ij}})^{12} = \sum_{i,j} U_{ij}$$

- Solve the low temperature bridge equations and generate trajectories.







Energy as a function of time

Conclusion

- Method is efficient, fast and parallelizable
- All trajectories are statistically independent
- Possibility to include the solvent
- Can be generalized to discrete systems.
- Working on applying it to all-atom description.

APPENDIX: COMPUTING THE GRADIENT OF ln(Q)

In this appendix, we prove the central equation of this article, namely, Eq. (17). For that matter, we need to compute the gradient of the logarithm of Q.

We have

$$Q(\mathbf{r},t) = \int_{\mathbf{r}(t)=\mathbf{r}}^{\mathbf{r}(t_f)=\mathbf{r}_f} \mathcal{D}\mathbf{r}(\tau) e^{-\int_t^{t_f} d\tau \left(\frac{\dot{\mathbf{r}}^2}{4D} + \frac{1}{D\gamma^2} V(\mathbf{r}(\tau)) \right)}, \tag{A1}$$

which we discretize by splitting the time interval $t_f - t$ in $N-k$ intervals of size dt,

$$t_k = t < t_{k+1} = t + dt < t_{k+2} < \cdots < t_{N-1} < t_N = t_f$$

with

$$t_i = t + (i-k)dt$$

for all $i \in [k, N]$ and

$$t_f - t = (N-k)dt.$$

We write

$$Q(\mathbf{r},t) = \int_{\mathbf{r}_k=\mathbf{r}}^{\mathbf{r}_N=\mathbf{r}_f} d\mathbf{r}_{k+1} \cdots d\mathbf{r}_{N-1} e^{\left[-\sum_{l=k}^{N-1} \left(\frac{(\mathbf{r}_{l+1}-\mathbf{r}_l)^2}{4Ddt} + \frac{dt}{D\gamma^2} V(\mathbf{r}_l) \right) \right]}.$$

We have

$$\begin{aligned}
\nabla Q(\mathbf{r},t) &= \int_{\mathbf{r}_k=\mathbf{r}}^{\mathbf{r}_N=\mathbf{r}_f} d\mathbf{r}_{k+1} \cdots d\mathbf{r}_{N-1} e^{\left[-\sum_{l=k+1}^{N-1} \left(\frac{(\mathbf{r}_{l+1}-\mathbf{r}_l)^2}{4Ddt} + \frac{dt}{D\gamma^2} V(\mathbf{r}_l) \right) \right]} \nabla e^{\left[-\left(\frac{(\mathbf{r}_{k+1}-\mathbf{r})^2}{4Ddt} + \frac{dt}{D\gamma^2} V(\mathbf{r}) \right) \right]} \\
&= \int_{\mathbf{r}_k=\mathbf{r}}^{\mathbf{r}_N=\mathbf{r}_f} d\mathbf{r}_{k+1} \cdots d\mathbf{r}_{N-1} e^{\left[-\sum_{l=k+1}^{N-1} \left(\frac{(\mathbf{r}_{l+1}-\mathbf{r}_l)^2}{4Ddt} + \frac{dt}{D\gamma^2} V(\mathbf{r}_l) \right) \right]} \left[\frac{\mathbf{r}_{k+1}-\mathbf{r}}{2Ddt} - \frac{dt}{D\gamma^2} \nabla V(\mathbf{r}) \right] \times e^{\left[-\left(\frac{(\mathbf{r}_{k+1}-\mathbf{r})^2}{4Ddt} + \frac{dt}{D\gamma^2} V(\mathbf{r}) \right) \right]} \\
&= \int_{\mathbf{r}_k=\mathbf{r}}^{\mathbf{r}_N=\mathbf{r}_f} d\mathbf{r}_{k+1} \cdots d\mathbf{r}_{N-1} e^{\left[-\sum_{l=k+1}^{N-1} \left(\frac{(\mathbf{r}_{l+1}-\mathbf{r}_l)^2}{4Ddt} + \frac{dt}{D\gamma^2} V(\mathbf{r}_l) \right) \right]} \left[-\nabla_{\mathbf{r}_{k+1}} - \frac{dt}{D\gamma^2} \nabla V(\mathbf{r}) \right] \times e^{\left[-\left(\frac{(\mathbf{r}_{k+1}-\mathbf{r})^2}{4Ddt} + \frac{dt}{D\gamma^2} V(\mathbf{r}) \right) \right]}.
\end{aligned}$$

We may then integrate by part the term $\nabla_{\mathbf{r}_{k+1}}$ and obtain

$$\nabla Q(\mathbf{r},t) = \int_{\mathbf{r}_k=\mathbf{r}}^{\mathbf{r}_N=\mathbf{r}_f} d\mathbf{r}_{k+1}\cdots d\mathbf{r}_{N-1} e^{\left[-\sum_{l=k+2}^{N-1}\left(\frac{(\mathbf{r}_{l+1}-\mathbf{r}_l)^2}{4Ddt}+\frac{dt}{D\gamma^2}V(\mathbf{r}_l)\right)\right]} \times \left[\frac{\mathbf{r}_{k+2}-\mathbf{r}_{k+1}}{2Ddt}-\frac{dt}{D\gamma^2}(\nabla V(\mathbf{r}_{k+1})+\nabla V(\mathbf{r}))\right] \times e^{\left[-\left(\frac{(\mathbf{r}_{k+2}-\mathbf{r}_{k+1})^2}{4Ddt}+\frac{(\mathbf{r}_{k+1}-\mathbf{r})^2}{4Ddt}+\frac{dt}{D\gamma^2}(V(\mathbf{r}_{k+1})+V(\mathbf{r}))\right)\right]}. \quad \text{(A3)}$$

By repeating this procedure, we obtain

$$\nabla Q(\mathbf{r},t) = \int_{\mathbf{r}_k=\mathbf{r}}^{\mathbf{r}_N=\mathbf{r}_f} d\mathbf{r}_{k+1}\cdots d\mathbf{r}_{N-1}\left[\frac{\mathbf{r}_f-\mathbf{r}_{N-1}}{2Ddt}-\frac{dt}{D\gamma^2}\sum_{l=k}^{N-1}\nabla V(\mathbf{r}_l)\right] \times e^{\left[-\sum_{l=k}^{N-1}\left(\frac{(\mathbf{r}_{l+1}-\mathbf{r}_l)^2}{4Ddt}+\frac{dt}{D\gamma^2}V(\mathbf{r}_l)\right)\right]}. \quad \text{(A4)}$$

By summing these $(N-k)$ equations [Eqs. (A2), (A3), and (A4)] and dividing by $(N-k)$, we obtain

$$\nabla Q(\mathbf{r},t) = \frac{\mathbf{r}_f-\mathbf{r}}{2D(t_f-t)}Q(\mathbf{r},t) - \frac{1}{D\gamma^2}\int_{\mathbf{r}_k=\mathbf{r}}^{\mathbf{r}_N=\mathbf{r}_f} d\mathbf{r}_{k+1}\cdots d\mathbf{r}_{N-1} \times e^{\left[-\sum_{l=k+1}^{N-1}\left(\frac{(\mathbf{r}_{l+1}-\mathbf{r}_l)^2}{4Ddt}+\frac{dt}{D\gamma^2}V(\mathbf{r}_l)\right)\right]}\frac{1}{N-k}((N-k)dt\nabla V(\mathbf{r}) + (N-k-1)dt\nabla V(\mathbf{r}_{k+1})+\cdots+dt\nabla V(\mathbf{r}_{N-1})). \quad \text{(A5)}$$

Taking the continuous limit of Eq. (A5) yields

$$2D\nabla \ln Q(\mathbf{r},t) = \frac{\mathbf{r}_f-\mathbf{r}}{t_f-t}-\frac{2}{\gamma^2}\int_t^{t_f} d\tau\left(\frac{t_f-\tau}{t_f-t}\right)\langle\nabla V(\mathbf{r}(\tau))\rangle,$$

where the average $\langle\cdots\rangle$ is done over all the Langevin paths starting at $(\mathbf{r},t)$ and ending at $(\mathbf{r}_f,t_f)$,

$$\langle\nabla V(\mathbf{r}(\tau))\rangle = \frac{1}{Q(\mathbf{r},t)}\int_{\mathbf{r}(t)=\mathbf{r}}^{\mathbf{r}(t_f)=\mathbf{r}_f}\mathcal{D}\mathbf{r}(\tau) \times e^{-\int_t^{t_f}d\tau\left(\frac{\dot{\mathbf{r}}^2}{4D}+\frac{1}{D\gamma^2}V(\mathbf{r}(\tau))\right)}\nabla V(\mathbf{r}(\tau)). \quad \text{(A6)}$$

The Langevin bridge equation, thus, becomes

$$\frac{d\mathbf{r}}{dt} = \frac{\mathbf{r}_f-\mathbf{r}}{t_f-t}-\frac{2}{\gamma^2}\int_t^{t_f} d\tau\left(\frac{t_f-\tau}{t_f-t}\right)\langle\nabla V(\mathbf{r}(\tau))\rangle+\boldsymbol{\eta}(t), \quad \text{(A7)}$$

which is Eq. (17) of this article.