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 aminoacids. 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.

Probe
release behind the Documentation License.

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

with Fluctuation-Dissipation relation

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

Z

Theorem: the probability distribution $P({r_i(t)})$ converges $e^{-\beta E({r_i})}$

to the Boltzmann distribution at large time

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 $1\,\mathrm{ms}$ 1s
- Recently, runs of Ims on short proteins (ANTON)
- Reason: Many metastable states and high barriers

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)

times in protein folding from single molecule photon-by-photon 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. **Hoi Sung Chung*, John M. Louis, and William A. Eaton***

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 **Experimental determination of upper bound for transition path** GFP **times in protein folding from single molecule photon-by-photon** FRET experiments **Hoi Sung Chung*, John M. Louis, and William A. Eaton*** \blacksquare \overline{O} \overline{a} 0.5 Mayayaka Mariy Ma **MAM** F U F $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ 0 *E* TP -0.5 $U \setminus F$ \cup **x** $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ **x** -1 -2 -1 0 1 2

time

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 $P_{\rm 0}$ to a final distribution P_f in time t_f if points follow Brownian Motion

Also related to Optimal Transport (Monge problem)

Langevin dynamics

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

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

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

$$
\langle \zeta(t)\zeta(t') \rangle = 2k_B T \gamma \delta(t - t')
$$

Overdamped Langevin dynamics

• At large enough time scale, mass term negligible

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

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

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

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

 $D = 10^{-5}$ cm²/s *m* ≈ 5.10⁻²⁶kg

• 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: $D =$ k_BT γ Diffusion coefficient \angle

$$
\eta(t) = \sqrt{\frac{2k_B T}{\gamma}} \xi(t) \quad \longleftarrow \quad \textbf{Normal variable}
$$

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

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

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

Using eq.(4) and (6), one can easily see that *P*(r*, t*) satisfies the modified FP equation

 $\overline{\hat{\beta}_{t}^{t}}(x_{f}, t_{f} | x_{i}, 0) = \left(\prod_{i=1}^{N} x_{i}^{t} \right)$ *k*=1 $\frac{d}{dx}_k \exp\left(-\frac{2}{4L}\right)$ 4*D* \sum *N k*=1 $\int \frac{x_{k+1} - x_k}{x_k}$ $\frac{d}{dt}$ + *D* $k_B T$ ∂U ∂x_k $\langle \rangle^2$ Path integral representation deformation. The additional force term 2D v ming 1 (with the original Deutscher Hollows and the contract of \bullet Continy where the effective potential *V* is given by $\mathrm{Rtr}(t)$ et the conditioned system satisfies a brook fied Langevin equation given languary $\mathrm{Rtr}(t)$ \rightarrow $\boldsymbol{f}^{\mathbf{r}(t_f)=\mathbf{r}_f}$ $\begin{array}{c}\n\cdot & 4D \end{array}$ *^D*r(⌧)*e* ¹ \int_{-1}^{1} $\int_{0}^{t} f \, d\tau (\dot{\bf{r}} + 1 \nabla U)^2$ $\sum_{r=0}^{\infty} \frac{f(\mathbf{x}_f(t))}{\sum_{i=0}^{\infty} \mathbf{F}(t_i)} - U(\mathbf{r})^2 \int_{0}^{\infty} \int_{0}^{t} \int_{0}^{t} d\mathbf{x}$ $\int_{\gamma(t)}$ ^t *D*r(⌧)*e* $\int_0^t f \, d\tau$ $\int \mathbf{r}(t)dt = \int \mathbf{r}(t)dt$ $\gamma^{2}(v)$ ⁻¹ $V(\mathbf{r}) = \frac{1}{\sigma \mathbf{v}}$ $\frac{1}{4}$ $\left(\nabla U\right)$ $V(\mathbf{r}) = \frac{1}{k_B T^4} (\nabla U)^2 - \frac{k_B T}{2}$ $\frac{B^2}{2} \nabla^2 U$ Onsager-Machlup (Ito) Stratonovich Path integral representation $\begin{equation*} \begin{array}{l} \nabla P \ + \ \nabla \theta \end{array} \begin{array}{l} \nabla \theta \end{array} \begin{array}{l}$ @*P* @*t* $= D \nabla (\nabla P + \nabla \left(\beta U(\mathbf{r}) - 2 \ln Q_1 \right) P) N \over L$ $\eta_{\rm t}$ r (t) of the conditioned system satisfies a modified Langevin equation given β γ δU s⁺ $\beta \overline{W}$ $\ln Q_1 + \eta(t)$ (8) (8) Onsager-Machlup $P(\mathbf{r}_f$ ation (cite). Chiquones system satisfies a spot in Q1 (w.r.t. the original Latin and the diversity of the additional force term $2D\nabla$ in Q_1 (w.r.t. the original La This equation is called a bridge equation (cite). The additional force term 2*D*r ln *Q*¹ (w.r.t. the original Langevin \bullet Continuous limit \mathbb{R} in $Q_1 + \eta(t)$ changer-Machlup \int ⁷ $P(r_f, t_f | r_f)$ $\mathbf{r}(t) = \mathbf{r}$ $\mathbf{r}(t) = \mathbf{r}$ $(Tt)_{\mathbf{r}}$ $\frac{\partial \mathbf{r}}{\partial \mathbf{r}}$ *d*_(*x*) *d*_(*x*)) *d*_(*x*)) *d*_(*x*)) *d*_(*x*)) *d*_(*x*)) *d*_(*x*)) *d*_(*x*)) *d*_(*x*)) *d*₍ $\oint_{\mathbf{\hat{x}}(t)=\mathbf{r}}$ $\int_{\mathbf{\hat{x}}(t)=\mathbf{r}}$ effe where ω $\frac{1}{2}$ is given by $P(\mathbf{r}_f)$ ˆ ^r(*t^f*)=r*^f* ${\bf r}(t)$ = ${\bf r}$ $\hat{\mathcal{D}} \hat{\mathbf{r}}$ $\langle \overline{\tau} \rangle$ $e \hspace{-0.1cm} \int_{\mathbf{r}(t)}^{\frac{1}{4L}}$ $\frac{1}{4D}\int_{t}^{t}f^{'}d\mathcal{D}(\mathbf{r}+\tau) \nabla U_{t}^{12}$ $\int_{\mathcal{C}^{\rho}}-\beta(U(\mathbf{r}_f)-U(\mathbf{r}))\overline{\mathbf{r}}\int_{0}^{\mathbf{r}}\mathbf{f}(\theta)\overline{\mathbf{r}}\mathbf{r}_f^{\mathbf{r}}\theta$ $\mathbf{\hat{r}}(t)$ =r $\mathcal{D} \mathbf{r}(\tau)$ *e* $-\int_t^t f \, d\tau$ $\left(\mathbb{Z}_p^2 f(\tau) \right) = \sqrt{\frac{1}{2} \left(\frac{1}{2} \pi \right)^2}$ *v*en by */*2 \overrightarrow{T} $\int_{1}^{\mathbf{r}(t_{f})=\mathbf{r}_{f}}$ ${\bf r}(t)$ =r $\mathcal{D}\hspace{-1.5pt}/\mathrm{r}(\tau_{\gamma}^{\perp})\mathcal{C}U_{\gamma}^{\perp}$ $\frac{\lambda^{12}}{4D} \int_{t}^{t_{f}} d\tau \Big(\dot{\mathbf{r}}+\frac{1}{\gamma}\nabla U\Big)^{2}$ $=\oint \frac{\mathbf{r}(t_f - t_f)}{\mathbf{r}(t_f)} = \int \frac{\mathbf{r}(t_f - t_f)}{\mathbf{r}(t_f)} dt$ $\mathrm{\acute{r}}(t)$ =r $\frac{d\mathbf{p}_{\mathbf{F}}}{d\mathbf{B}}(\mathbf{r}_{\mathbf{p}\mathbf{S}})$ $-\int_{t_{\ell}}^{t_{f}} d\tau$ $\sqrt{2} \int_{-}^{\mathbf{r}(t_f)} \int_{0}^{t_f} d\tau \left(\mathbf{p}_{\mathbf{F}}^2(r) \right) e^{-\int_{0}^{t_f} d\tau} \sqrt{\frac{\dot{\mathbf{r}}^2}{4D} + \frac{1}{D\gamma^2} V(\mathbf{r}(\tau))}$ *where the effective* potential *v* is a *v* in the effective potential \mathbf{v} 2^V ^U

$$
V(\mathbf{r})=\frac{1}{4}\left(\nabla U\right)^2-\frac{k_BT_4}{2}\nabla^2 U
$$

Path Integral = Feynman Path integral = Schroedinger equation

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

- Dominant Paths: Saddle-Point expansion: Minimise action: Newton equation = Instanton $\ddot{r} = \frac{2}{2} \frac{\partial v}{\partial r}$ with B.C. 2 *γ*2 ∂*V* ∂*r* 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 *xⁱ* and *x^f*
- Deform the trajectory locally and accept or reject with a Monte Carlo algorithm

- 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 *dx^f e*−β(*U*(*x^f*)−*U*(*x*0))*/*² *Drobability for such a* 4*kBT* $\overline{}$ Z

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

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

 $P(x, t) = P(x, t|x_0, 0)$ $Q(x,t) = P(x_f, t_f | x, t)$ FP adjoint FP

27

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

 $P(x,t) = P(x,t|x_0,0)$ $Q(x,t) = P(x_f, t_f | x, t)$ FP equation adjoint FP equation

The equation satisfied by *P* is given in eq. 4, whereas the equation for *Q* is given by

Fokker-Planck and adjoint
\n
$$
\frac{\partial P}{\partial t} = D \frac{\partial}{\partial x} \left(\frac{\partial P}{\partial x} + \beta \frac{\partial U}{\partial x} P \right)
$$
\n
$$
\frac{\partial Q}{\partial t} = -D \frac{\partial^2 Q}{\partial x^2} + D\beta \frac{\partial U}{\partial x} \frac{\partial Q}{\partial x}
$$
\n
$$
\vec{P} = \frac{\dot{Q}P + Q\dot{P}}{P(x_f, t_f | x_0, 0)}
$$
\n
$$
\frac{\partial P}{\partial t} = D \frac{\partial}{\partial x} \left(\frac{\partial P}{\partial x} + \frac{\partial}{\partial x} (\beta U - 2 \ln Q) P \right)_{29}
$$

• Modified Langevin equation for conditioned paths one sees that it can be obtained from a model it can be obtained potential potential potential potential poten
Potential potential p

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

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

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

$$
\frac{dX}{dt} = \frac{x_f - X}{t_f - t} \xrightarrow{\text{average is linear}} \frac{dX}{dt}
$$

Free Langevin

Constrained Langevin

tf

- Example: Brownian meanders = Brownian *^Q*(*x, t*) = ¹ *e* Example. *Drowman mearigers* = Drowman
walks constrained to stay with x>0 during fixed time. Use mirror image \dot{x} = $2D \frac{\partial \log Q(x,t)}{\partial x}$ ∂x $\dot{x} = 2D \frac{\partial \cos \varphi(x, t)}{\partial x} + \eta(t)$ = $\int \frac{x_f-x}{f}$ t_f ^{t} $\overline{ }$ *e* $-\frac{1}{4D}$ $(x_f - x)^2$ $\frac{f(x)}{f}$ + $\frac{x}{f}$ + $\frac{x}{f}$ t_f ^{t} $\overline{}$ *e* $-\frac{1}{4D}$ $(x_f+x)^2$ $t_f - t$ *e* $-\frac{1}{4D}$ $(x_f - x)^2$ t_f ^{-t} - e $-\frac{1}{4D}$ $(x_f+x)^2$ $t_f - t$ $+ \eta(t)$ $\frac{1}{2}$ eanders = B_I ¹ 4*D tf t FIXEG time. Use mirror image* α $\left(\frac{f-x}{f-t}\right) e^{-\frac{1}{4D}\frac{(x-\mu)^2}{t_f-t}} +$ $\left(\frac{f}{t_f-t}\right)e^{-tL}$ $\frac{f}{t_f - t}$ - $e^{-\frac{1}{4D} \frac{t_f}{t_f}}$
	- Integrate for

$$
x_f > 0
$$

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

The case of a Brownian excursion, where the extremity *x^f* is fixed, is generated by the

• Example: Brownian excursions: Brownian walks $x > 0$, starting around $x = 0$ and ending at $x=0$. *dt* $\overline{ }$ ple: Brownian excursior $rownian$ at $x=0$.

$$
\dot{x} = 2D \frac{\partial \log Q(x, t)}{\partial x} + \eta(t)
$$
\n
$$
= \frac{\left(\frac{x_f - x}{t_f - t}\right) e^{-\frac{1}{4D} \frac{(x_f - x)^2}{t_f - t}} + \left(\frac{x_f + x}{t_f - t}\right) e^{-\frac{1}{4D} \frac{(x_f + x)^2}{t_f - t}}}{e^{-\frac{1}{4D} \frac{(x_f - x)^2}{t_f - t}} - e^{-\frac{1}{4D} \frac{(x_f + x)^2}{t_f - t}}} + \eta(t)
$$

• Take the limit $x_f \to 0$ e limit $x_f \rightarrow$

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

Example: Brownian excursions above a line

• Example: the Harmonic oscillator (Ornstein-Uhlenbeck process)

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

$$
\dot{x} = -\frac{1}{\gamma}Kx + \eta(t)
$$

• 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

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 $\,f_f$

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

Figure 1. Transition pathway between an unknotted ring and a left-handed 5₁ 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 41 knotted ring. The shown observables are the same as in Figure 1.

Figure 3. Transition pathway between two 5₂ 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 puting?
בחם kc are eigenvectors and eigenvalues
 a for and using its spectral

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

= $e^{-\beta (U(x_f) - U(x))/2} \langle x_f | e^{-H(t_f - t)} | x \rangle$
= $e^{-\beta (U(x_f) - U(x))/2} \sum_{\alpha} e^{-E_{\alpha}(t_f - t)} \Psi_{\alpha}(x_f) \Psi_{\alpha}(x)$

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

- 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

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

Short transition path time approximation

• In the Kramers picture, there are 2 time scales:

–Kramers time, or waiting time, or folding time ΔE $\frac{\Delta E}{k_B T}$ \longleftarrow $\textsf{ms-ls}$

–Transition path time (A. Szabo), Instantons $\tau_K \approx e$

$$
\tau_{TP} \approx \log \frac{\Delta E}{k_B T} \sim 1 \text{ }\mu\text{s}
$$

–We have

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

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

$$
Q_{A}(t) \equiv \int_{x_{0}}^{\infty} dx P(x, t | -x_{0}, 0)
$$
 Absorption Probability
= $\langle \theta(t - t_{TP}) \rangle$

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

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

This equation neglects paths which return to the left side. Very good approximation for high barrier a characteristic rate of the process. The process \mathbf{r} and \mathbf{r} are process. The limiting behaviors of the li **G20** This equation neglects paths **w** by the behavior of *^t*(*G*) close to the lower bound *^G* & ^p*E*, ft side. This equation neglects paths which return ¹

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

Paths which never cross

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

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

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

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

It is possible to do the calculation with inertia term Expressions are more complicated when inertia present The solid line in Fig. 2 shows a plot of the TPT distribution in Fig. 2 shows a plot of the TPT distribution in Fig. 2 shows a plot of the TPT distribution in Fig. 2 shows a plot of the TPT distribution in TPT distributio it is possible to go the calculation **v** in the limit *t* ! 0. From Eqs. (A10) and (A12), one finds at

VII. DISCUSSION VII. DISCUSSION Conformational transition molecular systems between Conformational transition molecular systems between High Barrier

aug Dannian $\overline{}$ band in colored we example. barrier is parabolic, the associated Langevin equation is linear Low Barrier

The main properties of the TPT distribution have been summa-

A useful form of the Bridge equation @*P* = *D*r (r*P* + r (*U*(r) 2 ln *Q*1) *P*) (7)

Original bridge equation from which we see that the position r(*t*) of the conditioned system satisfies a modified Langevin equation given by Ω riginal bridge equation **e** participations the pathologies that the pathologies that the pathologies \mathbf{r} is a path integral representations of \mathbf{r} is a path integral representation in the path integral representation in the path integra

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

Using the path-integral representation

$$
Q_{\uparrow}(\mathbf{r},t) = P(\mathbf{r}_{f},t_{f}|\mathbf{r},t)
$$

\n
$$
= \int_{\mathbf{r}(t)=\mathbf{r}}^{\mathbf{r}(t_{f})=\mathbf{r}_{f}} \mathcal{D}\mathbf{r}(\tau)e^{-\frac{1}{4D}\int_{t}^{t_{f}} d\tau (\dot{\mathbf{r}}+\frac{1}{\gamma}\nabla U)^{2}} \mathbf{r}(\dot{\mathbf{r}}+\frac{1}{\gamma}\nabla U)^{2}
$$

\n
$$
= e^{-\beta(U(\mathbf{r}_{f})-U(\mathbf{r}))/2} \int_{\mathbf{r}(t)=\mathbf{r}}^{\mathbf{r}(t_{f})=\mathbf{r}_{f}} \mathcal{D}\mathbf{r}(\tau)e^{-\int_{t}^{t_{f}} d\tau (\frac{\dot{\mathbf{r}}^{2}}{4D}+\frac{1}{D\gamma^{2}}V(\mathbf{r}(\tau)))}
$$

where the effective potential is given by where the effective potential \mathbf{r} is given by \mathbf{r}

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

After some simple transformations, the bridge equation becomes: r˙ = 2*D*r ln *Q* + ⌘(*t*) (14) ˆ ^r(*t^f*)=r*^f* r(*t*)=r ransfor 4*D* ´ *tf ^t ^d*⌧(r˙+ ¹ ^r*^U*) 2 $\frac{u}{u}$ **e** $\frac{dt}{u}$ /*l ^D*² *^V* (r(⌧))⌘ In the Supplemental Material, using the path integral representation (13) and performing several integrations by After some simple transformations, the bridge equation

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

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

where the bracket $\langle \cdots \rangle_Q$ denotes the average over all paths joining (\mathbf{r}, t) to (\mathbf{r}_f, t_f) . (13)
(13)
(13) weighted by Q . Machlup action (cite) ¹ ⇣ ⌘2 ˆ ^r(*t^f*)=r*^f* σ above equations are obtained by \mathscr{L} . weighted by Q 1 ´ *tf* **D**
*D*₂ *V* (*r*)

- hr*V* (r(⌧))i*^Q* = recovers the *Q*(r*, t*) Brownian bridge equation ˆ *^t^f* the constraint r(*t^f*) = r*^f* is satisfied. It is the only term which is singular at time *t^f* . In fact, in the case of a free **a** If no potential V=0, one recovers the standard free (10), when expanding the square, and using the square, and using the identity of stochastic calculus α *d*⌧r˙r*U*(r(⌧)) = *U*(r*^f*) *U*(r) *D*
	- The only singular term is the free Brownian bridge term. Brownian particle, the effective potential *V* vanishes, and we recover the standard equation for free Brownian bridges It is the term which conditions the paths ref *i* i i i s
ich condi *th* conditions the paths *the paths*
- \mathbf{r} the future of the \bm{t} is the fundamental equation of this article and will be used to generate constrained paths. This article and will be used to generate constrained paths. This article and will be used to generate constraints. This • Highly non-Markovian and depends on the future of the trajectory r(*t*)=r *D*r(⌧)*e* • Highly non-Markovian and depends on the future of the This equation is a non-linear stochastic equation. It is matched that the sense that the r.h.s. of (15) depends on α on r(*t*). However, the presence of the average over all future paths makes it difficult to use.

At $\mathsf{Zero}\ \mathsf{Temperature}$ At zero temperature, the noise term vanishes and the average in (15) reduces to a single trajectory r0(*t*). We thus At zero temperature, the noise term vanishes and the average in (15) reduces to a single trajectory r0(*t*). We thus At zero temperature, the noise term vanishes and the average in (15) reduces to a single trajectory r0(*t*). We thus At zero temperature, the noise term vanishes and the average in (15) reduces to a single trajectory r0(*t*). We thus

At T=0, only one trajectory and the equation becomes r*^f* r0(*t*) *t t tf tf**tf**tf**tf**tf***** *tf**tf**tf**tf**tf***** *tf*

$$
\dot{\mathbf{r}}_0 = \frac{\mathbf{r}_f - \mathbf{r}_0(t)}{t_f - t} - \frac{2}{\gamma^2} \int_t^{t_f} d\tau \left(\frac{t_f - \tau}{t_f - t}\right) \nabla V_0(\mathbf{r}_0(\tau))
$$
\nwhere $V_0(\mathbf{r}) = \frac{1}{4} (\nabla U)^2$

 t_{α} is α at time derivative of the above equation we get taking a time derivative of the above equation we get

$$
\ddot{\mathbf{r}}_0 = \frac{2}{\gamma^2} \nabla V_0(\mathbf{r}_0) = \frac{1}{\gamma^2} \nabla U(\mathbf{r}_0). \nabla^2 U(\mathbf{r}_0)
$$
\nditions $\mathbf{r}_0(0) = \mathbf{r}_1$ and $\mathbf{r}_0(t_0) = \mathbf{r}_0$

boundary conditions $\mathbf{r}_0(0) = \mathbf{r}_i$ and $\mathbf{r}_0(t_f) = \mathbf{r}_f$ 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 This is the exact zero temperature equation. Instanton ine ory and string method. The two equations (20) and (21) can of course have several solutions of course have s which die by the boundary components requirements rich the standard in the standard instance of t (cited) and string iteration. \overline{c} mentioned above, the non-linear equation (20) is solved iteratively, starting from an approximate solved iteratively, starting from an approximate solved iteratively, starting from an approximate solution \overline{c} This is the exact zero temperature equation: Instanton Theory and String Method. $\overline{}$ at low temperature, we can expand the trajectories in $\overline{}$ $\mathcal{L}_{\mathcal{L}}$ as mentioned above, the non-linear equation (20) is solved iteratively, starting from an approximate solved iteratively, starting from an approximate solved iteratively, starting from an approximate solutio At low temperature, we can expand the trajectories in (??) as Γ heer and Γ tring Mothed iteratively, starting from an approximate solved iteratively, starting from an approximate solved iteratively, starting from an approximate solved iteratively, starting from an approximate so which was will discuss later. The two equations (20) and (21) can of course have seen of course have s

Weak Fluctuations

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

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

With the approximation, denoting the average here are the average here is a bit of α in α , and β The Bridge equation becomes hr*V* (r(⌧))i*^Q* ⇡ r*V* (hr(⌧)i*Q*) (18) Within this approximation, denoting the average hr(⌧)i*^Q* by r(⌧), eq.(15) becomes

t

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

As mentioned above, this equation is valid when the trajectories are not too much dispersed around the average Valid at order 1 in T

valid at order 1 in T Not Markovian.

Iterative Solution \blacksquare the equations (23) into the form of a recursion of a recursion of a recursion of a recursion of a reconvention of a recursion of a reconvergence dependence depend \int *f* \int *f* \int *f* \int *d* \int *f* \int *d* \int *f* \int (*k*) T simplest method to solve the non-linear integro-differential stochastic equations (2

This is a Non-Linear, Integro-Differential, non Markovian **Computation** Stochastic Equation of the trajectory at time $\frac{1}{2}$ <u>of this is a thoff-lifear, integro-Differential, non-trafricovian</u>
Chaebeetic Caustian There are several ways to implement the interactions, depending on the way one splits on the way one splits on the way one splits on the way on this is a Non-Linear, Integro-Differential, non Markovian C_t on the chine Γ cuestion the equation is to use a Euler-Maruyama discretization scheme for the equation, dividing the time *t^f* in *I* intervals

Solve it by iteration. Discretize à la Euler-Maruyama $r_f = r^{(n)}(k)$ $2 \frac{I-1}{I}$ \mathbf{B} X \mathbf{a} Euler-Plater
 $\mathbf{b} = k'dt \times d$ Solve it by iteration. Discretize à la Euler-Maruyama computation. Denoting by r(n) (k) the *n*-th iteration of the trajectory at time *t* = *kdt*, we write the iteration as

cumulant expansion developed in [].

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

$$
\langle \xi_a(k) \rangle = 0
$$

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

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

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

$$
V(x) = \frac{1}{4k_B T}(U'^2(x) - 2k_B T U''(x))
$$

Trajectories

Exact trajectories (in black) obtained from spectral decomposition. Approximate trajectories in red (with same noise)

The **Mueller potential**

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

where *A* = (−200*,* −100*,* −170*,* 15), *a* = (−1*,* −1*,* −6*.*5*,* 0*.*7), *b* = (0*,* 0*,* 11*,* 0*.*6),

 $0.5\,$ $\overline{0}$ $\mathbf{1}$

Mexican Hat Potential

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

 $\overline{0}$ of the Mexican hat is given by $\overline{0}$ Langevin bridge trajectories on the Mexican hat potential.

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

and has therefore a circle of minima for *x*²+*y*² = 1 with *U* = 0 and a maximum at (0*,* 0) with

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 center position of the continued of the inverse of the inverse of the inverse original property $\frac{1}{2}$ of the mixing Temperature *T^m* [9]. \bullet Define mixed elastic network model by **U**
Cu Clastic HCLWOI K IIIOUCI DY

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

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

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

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

• Collision term to avoid steric clashes during **•** the transition the transition **● Collision term to avoid steric clashes during** of the mixing \overline{q} *ij ij*)

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

 $\overline{}$

Uij (33)

• 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) namely, a 1D potential and a 2D potential. We recognize that we **EXPRENDIX: COMPUTING THE GRADIENT OF IN(Q) the solvers for the interval of the interval of the splitting of the interval of**

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 $\forall x \in \text{have}$

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

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 tial equations with two boundary conditions, the *shooting* methods and the *relaxation* methods.⁴⁶ with $\frac{1}{\sqrt{2}}$

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

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

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

We write

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

$$
\nabla Q(\mathbf{r},t) = \int_{\mathbf{r}_{k}=\mathbf{r}}^{\mathbf{r}_{N}=\mathbf{r}_{f}} d\mathbf{r}_{k+1} \cdots d\mathbf{r}_{N-1} e^{-\sum_{l=k+1}^{N-1} \left(\frac{(\mathbf{r}_{l+1}-\mathbf{r}_{l})^{2}}{4Ddt} + \frac{dt}{Dy^{2}} V(\mathbf{r}_{l}) \right)} \nabla e^{-\left(\frac{(\mathbf{r}_{k+1}-\mathbf{r}_{l})^{2}}{4Ddt} + \frac{dt}{Dy^{2}} V(\mathbf{r}_{l}) \right)} \times e^{-\left(\frac{(\mathbf{r}_{k+1}-\mathbf{r}_{l})^{2}}{4Ddt} + \frac{dt}{Dy^{2}} V(\mathbf{r}_{l}) \right)} \nabla e^{-\left(\frac{(\mathbf{r}_{k+1}-\mathbf{r}_{l})^{2}}{4Ddt} + \frac{dt}{Dy^{2}} V(\mathbf{r}_{l}) \right)} \nabla e^{-\left(\frac{(\mathbf{r}_{k+1}-\mathbf{r}_{l})^{2}}{4Ddt} + \frac{dt}{Dy^{2}} V(\mathbf{r}_{l}) \right)} \times e^{-\left(\frac{(\mathbf{r}_{k+1}-\mathbf{r}_{l})^{2}}{4Ddt} + \frac{dt}{Dy^{2}} V(\mathbf{r}_{l}) \right)}.
$$

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

$$
\nabla Q(\mathbf{r},t) = \int_{\mathbf{r}_{k}=\mathbf{r}}^{\mathbf{r}_{N}=\mathbf{r}_{f}} d\mathbf{r}_{k+1} \cdots d\mathbf{r}_{N-1} e^{-\sum_{l=k+2}^{N-1} \left(\frac{(\mathbf{r}_{l+1}-\mathbf{r}_{l})^{2}}{4Ddt} + \frac{dt}{D\gamma^{2}} V(\mathbf{r}_{l}) \right)}
$$

$$
\times \left[\frac{\mathbf{r}_{k+2} - \mathbf{r}_{k+1}}{2Ddt} - \frac{dt}{D\gamma^{2}} \left(\nabla V(\mathbf{r}_{k+1}) + \nabla V(\mathbf{r}) \right) \right]
$$

$$
\times e^{-\left[-\left(\frac{(\mathbf{r}_{k+2} - \mathbf{r}_{k+1})^{2}}{4Ddt} + \frac{(\mathbf{r}_{k+1} - \mathbf{r})^{2}}{4Ddt} + \frac{dt}{D\gamma^{2}} \left(V(\mathbf{r}_{k+1}) + V(\mathbf{r}) \right) \right] \right]}.
$$
 (A3)

By repeating this procedure, we obtain

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

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

$$
\nabla Q(\mathbf{r},t) = \frac{\mathbf{r}_f - \mathbf{r}}{2D(t_f - t)} Q(\mathbf{r},t) - \frac{1}{D\gamma^2} \int_{\mathbf{r}_k = \mathbf{r}}^{\mathbf{r}_N = \mathbf{r}_f} d\mathbf{r}_{k+1} \cdots d\mathbf{r}_{N-1}
$$

$$
\times e^{-\left[-\sum_{l=k+1}^{N-1} \left(\frac{(r_{l+1} - r_l)^2}{4Ddt} + \frac{dt}{D\gamma^2} V(\mathbf{r}_l)\right)\right]} \frac{1}{N-k} ((N-k)dt \nabla V(\mathbf{r}) + (N-k-1)dt \nabla V(\mathbf{r}_{k+1}) + \cdots + dt \nabla V(\mathbf{r}_{N-1})). \tag{A5}
$$

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) ,

$$
\langle \nabla V(\mathbf{r}(\tau)) \rangle = \frac{1}{Q(\mathbf{r},t)} \int_{\mathbf{r}(t)=\mathbf{r}}^{\mathbf{r}(t_f)=\mathbf{r}_f} \mathcal{D}\mathbf{r}(\tau)
$$

$$
\times e^{-\int_{t}^{t_f} d\tau \left(\frac{\mathbf{r}^2}{4D} + \frac{1}{D\gamma^2} V(\mathbf{r}(\tau))\right)} \nabla V(\mathbf{r}(\tau)). \tag{A6}
$$

The Langevin bridge equation, thus, becomes

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

which is Eq. (17) of this article.

mum likelihood trajectory for large-scale structural transitions in a coarse grained

dimensional problems," J. Chem. Phys. **85**, 5045–5048 (1986).