Title
Statistical Learning Theory of Protein Dynamics

Author
Haas, Kevin

Publication Date
2013

Peer reviewed|Thesis/dissertation

Statistical Learning Theory of Protein Dynamics

by

Kevin Richard Haas

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Chemical and Biomolecular Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jhih-Wei Chu, Chair
Professor Berend Smit
Professor Douglas Clark
Professor Phillip Geissler

Spring 2013

Statistical Learning Theory of Protein Dynamics

Abstract

Statistical Learning Theory of Protein Dynamics

by

Kevin Richard Haas

Doctor of Philosophy in Chemical and Biomolecular Engineering

University of California, Berkeley

Professor Jhih-Wei Chu, Chair

This thesis establishes a comprehensive statistical learning framework to extract from single-molecule Förster resonance energy transfer (smFRET) experiments the potential of mean force and diffusion coefficient that characterize the measured protein dynamics. To enable a fundamental understanding of how deterministic mean force and stochastic diffusion combine to affect conformational transitions, we first developed a general trajectory entropy functional for over-damped Langevin dynamics. This functional allows for evaluation of the information content in the dynamic trajectory ensemble. Next, we present a path integral statistical learning approach to infer the hidden trajectory from the data of smFRET measurements of protein dynamics. This methodology also yields a likelihood for the parameters of the equation of motion that can then be optimized to deduce the most probable profiles of mean force and diffusion coefficient for describing the observed dynamical data.

To provide a solid foundation for regularizing the parameters derived from experimental trajectories through statistical learning, the Fisher information metric of Langevin dynamics trajectories is derived via an eigenbasis representation of the time propagator. Using this Fisher information, the maximum entropy distributions for various kinetic constraints can derived for the first time. Finally the knowledge of trajectory entropy and likelihood of smFRET measurements is combined to present a new calculus of representing the Information Thermodynamics in statistical learning. Bayesian analysis using this methodology shows that in the balance between entropy, likelihood, and fluctuations given at the critical point in the phase diagram of information, the ideal force profile and diffusion can be determined from smFRET experiments in a systematic and robust manner.

To my family,

For inspiring my love of science and engineering,
and supporting me through this journey.

Contents

List of Figures

List of Tables

List of Symbols

Notation	Description
$\langle\alpha(t)\|$	State bra of system at time t incorporating data $[0, t]$
B_d	Intensity of background on donor channel
B_a	Intensity of background on acceptor channel
$\|\beta(t)\rangle$	State ket of system at time t incorporating data $[t, t_{ons}]$
$\boldsymbol{C}$	Constraint opperator
$D(x)$	Diffusion coefficient
$\boldsymbol{S}$	Entropy operator
$S[]$	Entropy functional
Φ	Expectation of normalized eigenvectors
Φ'	Expectation of derivative of normalized eigenvectors
η_D	Diffusion "pressure"
η_F	Force "temperature"
$\mathbb{E}_{\mathbb{Y}}$	Expectation over FRET data
$F(x)$	Mean Force
$\Gamma_{i,j}$	Expected eigenvector transitions
$\boldsymbol{H}$	Hamiltonian operator
$\boldsymbol{H^0}$	Hamiltonian operator of system time propagation
$\boldsymbol{H^I}$	Hamiltonian operator for smFRET dark state
$\mathcal{I}$	Fisher Information of Trajectory
I_a^0	Intensity of the acceptor FRET chromophore

Notation	Description
I_d^0	Intensity of the donor FRET chromophore
k	Reaction or relaxation rate
k_B	Boltzmann constant
$\mathcal{L}$	Likelihood or Lagrangian free energy functional
λ_i	Eigenvalues of hamiltonian or hessian
ℓ	Log-likelihood
∇	Gradient
N_P	Number of photons in FRET experiment
$\Omega_{jl}^{ik}(\Delta t)$	Overlap Integral
θ	Parameters for system dynamics
$\Theta(x)$	Collection of Dynamic parameters $\{F(x), D(x)\}$
$\mathcal{P}[]$	Probability of path functional
p_{eq}	Equilibrium probability
ϕ_i	Relative eigenvector of hamiltonian $\psi = \rho_{eq}\varphi$
x	Relative Position $x = \|r\|/R_0$
$V(x)$	Potential of mean force
ψ_i	Eigenvector of hamiltonian
R_0	Föster radius
ρ	Density field $\rho_{eq}(x) = \sqrt{p_{eq}}(x)$
r	FRET dye distance
$\tau_{A\to B}$	Mean First Passage Time for transition from state A to state B
T	Temperature
t	Time
t_{exp}	Duration of FRET experiment
t_{obs}	Duration of trajectory
ω_i	Lagrange multipliers, weight of bundle i
W_t	Weiner Process
x_L	Left Boundary
$X(t)$	System trajectory in time
x_R	Right Boundary
x_t	Random variable for system position at time t
y	smFRET data observation
$\boldsymbol{y}^a$	smFRET acceptor arrival operator

Notation	Description
$Y(t)$	Experimental observation trajectory in time
y_t	Random variable for smFRET observation at time t
$\zeta(x)$	Porportion of smFRET transfer

Acknowledgements

This thesis is a testament to the power of deep collaboration between theoretical and experimental research groups and the complimentary approaches necessary to solving problems and understanding the physical world around us. I am deeply grateful to the assistance of Professor Haw Yang from Princeton University and his students Jeff Hanson, Yan-Wen Tan, and Tom Morrell. Professor Yang's group had been working for several years on collecting and analyzing data form single-molecule FRET experiments before approaching Professor Chu and myself with the challenge to extract the same information of dynamics we had been pulling out of our molecular dynamics simulations. Although the challenge proved exceptionally difficult, the work required to ultimately formulate the comprehensive solution presented in this thesis was motivated by the desire to provide practical solutions to understand real experiments.

I would like to highlight the assistance from Jeff Hanson who spent countless ours explaining the raw details of the FRET experiments and also thank Tom Morrell who worked to collaborate on implementing the statical learning algorithms as well as serving as unofficial tour guide for my visits to Princeton. From the Chu group at UC-Berkeley, I would like to recognize the assistance, collaboration, and friendship of my co-workers Jason Brokaw, Barry Shang, and Adam Gross. From the greater UC-Berkeley community, I would like to thank David Sivak, John Chodera, Gavin Crooks, Prof. Phil Geissler, Prof. Doug Clark, and Prof. Berend Smit for education, enlightening discussions, and guidance in my research.

Finally, I literally owe all of my successes in my PhD career to the tireless dedication and steadfast resolve of my thesis advisor Prof. Jhih-Wei Chu. He has been the soldier down in the trenches with me during the many difficult days struggling to work through seemingly impossible theory only to reach

many dead-ends. Through it all, he has been the biggest cheerleader and motivator for my success. This research project was huge risk in both the scope of the problems and the depth of knowledge in many disparate fields from biophysics to quantum mechanics and computer science. I am ever grateful that he took that chance with me and I'm very proud to offer this thesis in tribute to our great collaboration, partnership, and friendship over the many years of my PhD career.

CHAPTER 1

Introduction

Engineers are masters of manipulating the dynamics of systems. From the design of airplanes to the coordination of data through a computer network, we must understand the manner in which system evolve through time in order to control and direct these systems to achieve our desired ends. For chemical engineers, our domain is the regulation of reactions which transform the structure of molecules, and many of the applications of modern bimolecular engineering require a deep understating of protein dynamics. Some of the grand challenges in the field can be surmised as making an enzymatic reaction faster, for example with the goal of converting cellulosic biomass to alternative fuels[1], or conversely slowing down the rate of conformational change in such pathologies as the amyloid-beta peptide fragments which lead to Alzheimer's disease[2].

We develop our understanding of these complex systems by converting data into models for dynamics and this operation demands a robust and comprehensive methodology. Statistical Learning Theory is a branch of computer science and statistics that formulates mathematical algorithms which recreate human's natural ability to synthesize statistical data into models of speech, vision, and classification[3, 4]. This thesis presents a statistical learning theory for protein dynamics.

Fully comprehending the nature of condensed phases systems like proteins at first seems futile[5–8]. Thousands of protein atoms are acted upon by a myriad of different forces and millions of solvent molecules on timescales that span the femtoseconds of hydrogen bond vibrations to the milliseconds of large scale protein conformational change[9]. Even if it was computationally feasible to simulate such physics on biologically relevant timescales, the raw deluge of trajectory data would overwhelm our ability to contemplate the nature of the

dynamics. Yet out of this chaos comes clarity. These many degrees of freedom aggregate to produce an average affect on the time evolution of the reaction. For a generalized coordinate of interest x_t that evolves through time and tracks the progress of the reaction, one can conduct the necessary model reduction and construct a phenomenological Langevin dynamics description[10]

$$dx_t = \beta D \langle F \rangle_{x_t} dt + \sqrt{2D} dW_t \tag{1.1}$$

where the average force $\langle F \rangle_{x_t}$ at a given position is constructed from the additive effect of all other degrees of freedom in the system.[11] As the force pushes the system towards low energy states, the random impulses of the fluctuations dW_t energize the system to maintain thermal equilibrium. The diffusion D sets the timescale for this dynamics and may depend on positions[12]. Discovering this governing coordinate—or coordinates—and calculating the average force profile and diffusion along a reaction path is therefore of upmost importance.

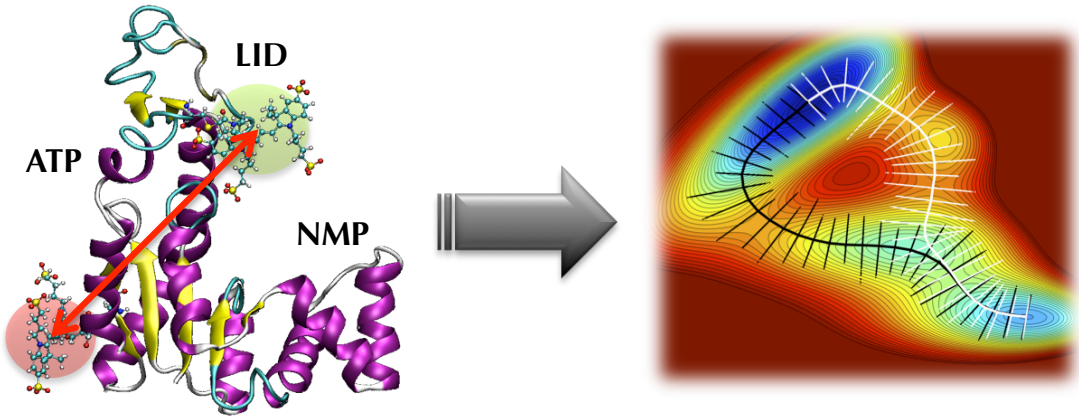

Figure 1.1: Cartoon of the transformation from a molecular dynamics model of the adynalate kinase (AK) protein to a free energy surface of two generalized coordinates. (left) The AK protein undergoes conformational transitions by which the LID and NMP binding domains collapse around the ATP active site. (Right) The hypothetical free energy contours for the reduced coordinates of LID-ATP and NMP-ATP distances. The dual reaction pathways are mapped out along with perpendicular planes which indicate the sampling of the degrees of freedom which contribute to the average force along this path. Also pictured are the red "acceptor" and green "donor" dyes from the subsequent smFRET study of this system[13].

Theoretical approaches to understanding these systems, I argue, have traditionally taken a generative model approach by which we include all atoms, forces, and physics that could possibly be relevant to the explanation of the desired system outcomes and observations. From these molecular/quantum dynamics simulations, one then perform some degree of importance sampling to extract out the most relevant statistics or interesting features of dynamics. This can be achieved by (for example):

- Sampling the transition path ensemble between meta-stable reactant and product states [14–16]
- Improving the speed of MD simulation to enable greater search of conformation space[17–19]
- Creating Markov state models between states to bridge timescales[20–22]
- Defining a reaction coordinate to compress the dimensionality of the system dynamics[23–25]
- Free energy and work decompositions including contributions from the author himself and co-workers [26–28]

Figure 1.1 is a cartoon representing the aim of this transformation. Although much progress has been made on these fronts, at the very least the demands on computational resources is immense and at the end of the day, there are still questions of the validity of the underlying physical parameters used to formulate the model.

However the statistical learning approach is different; we will flip the arrow to Figure 1.2 and work from data to a discriminative model. Groundbreaking experiments[29–31] of single-molecule force spectroscopy and single-molecule Förster resonance energy transfer[32, 33] (smFRET) have motivated a new class of statistical theory. In smFRET, the focus of this thesis, florescent dyes attached to different domains of the protein are excited with blue laser light. The donor dye absorbs this blue light and then can either release a lower energy green photon or transfer the energy to the acceptor dye which emits a red photon. The efficiency of this transfer is dependent on distance $E(x) = (r/R_0)^6$, therefore measuring the photon emissions gives a report on the current distance of the tagged coordinate. The time series measurement of photon emission is a statistical output of the underlying dynamic process we are interested in studying. However the observation time required for precision measurement from this experiment is on the order of the general relaxation time of the system meaning that the dynamics of the system is convoluted with the dynamics of

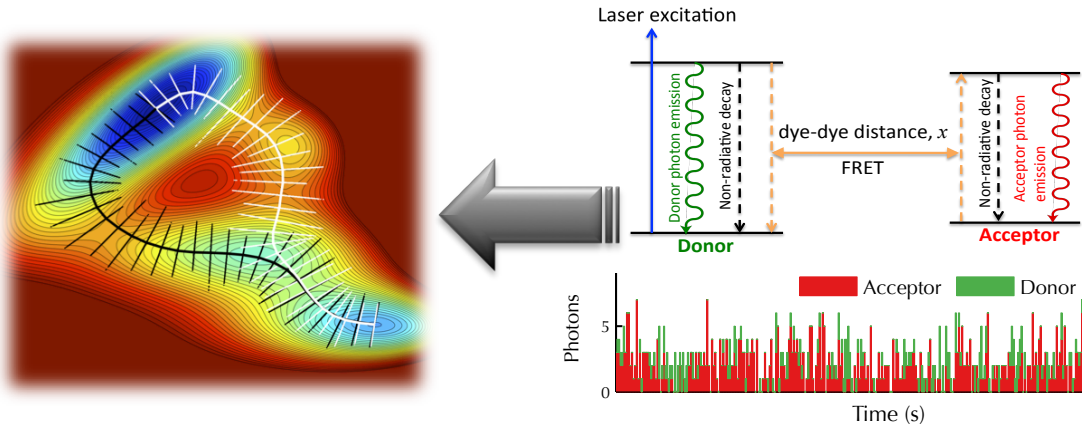

Figure 1.2: Cartoon of the transformation from experimentally observed smFRET time series data of dynamics to a phenomenological model which coresponds to the same free energy surface gleaned from molecular theory in Figure 1.1.

the experiment. As a result attempts to predict the distribution of equilibrium states is fraught with pitfalls and inaccuracy (see Appendix D and [34, 35]).

Our challenge is to develop a comprehensive, robust, and practical methodology to extract the potential of mean force and diffusion constant from single molecule experiment. This work will leverage many techniques from the fields of machine learning, statistical mechanics, and quantum mechanics.

This thesis is organized into five chapters which establish the comprehensive statistical framework necessary to solve this problem.

Entropy of trajectory ensemble (Chapter 2)

Before we set out to find the potential of mean force from smFRET, we would like to develop an understanding of what information is contained within the parameters we are seek to find. How do those functions of force and diffusion actual define the nature of dynamics for a system and what distinguishing features arise from such a parameterization? Through our efforts, we have discovered a fundamental and general property of all systems which obey Langevin dynamics: the trajectory entropy.

Path Integral Likelihood for FRET (Chapter 3)

How do the statistics of the time series data from experiment couple with the stochastic time evolution of the system? We account for the statistics of the Poisson process for photon emission and the Langevin dynamics

of the system with an exact and general path integral description which is solved using the eigenbasis of the time propagator. This gives an inferred system trajectory and the parameter likelihood which can then be optimized for force and diffusion profiles.

Fisher metric and max entropy models (Chapter 4)
Our solutions and dynamic models live in a parameter space characterized by the Fisher information metric. What is this metric for general Langevin dynamics and how can it be used to develop least informative models for given kinetic constraints?

Information Thermodynamics and Bayesian optimization (Chapter 5)
Combining our knowledge of entropy, likelihood, and Fisher information, we propose a new calculus of Information Thermodynamics. This allows us to perform Bayesian optimization to give the most robust and reliable estimates of parameters from experiment. The development also introduces the extension of the bundle method[36] to functional analysis to provide exponential convergence of our model parameters.

Conclusions and Future Perspectives (Chapter 6)
Armed with this statistical learning toolkit, we introduce and motivate future applications and protein engineering strategies. Finally we discuss the broader applicability of this approach to developing dynamic models from experimental data.

CHAPTER 2

Elements of the Trajectory Entropy in Continuous Stochastic Processes at Equilibrium

2.1 Abstract

We propose to define the trajectory entropy of a dynamic system as the Kullback-Leibler divergence of its path distribution against that of free diffusion. The space-time trajectory is now the dynamic variable and its path probability is given by the Onsager-Machlup action. For the time propagation of the over-damped Langevin equation, we solved the action path integral in the continuum limit and arrived at an exact analytical solution that emerged as a simple functional of the deterministic mean force and the stochastic diffusion.

2.2 Introduction

A dynamic system subjected to random influences explores its possible outcomes and evolves to exhibit a dispersion over state space and time that contains contributions from both deterministic and stochastic forces (Figure 2.1). One finds examples of this nature in areas including physics, chemistry, biology, as well economics. Resolving the physical origin of the space-time dispersion in dynamics and the quantification thereof will thus be illuminating. Here, we investigate this general problem using the over-damped Langevin equation as a model.

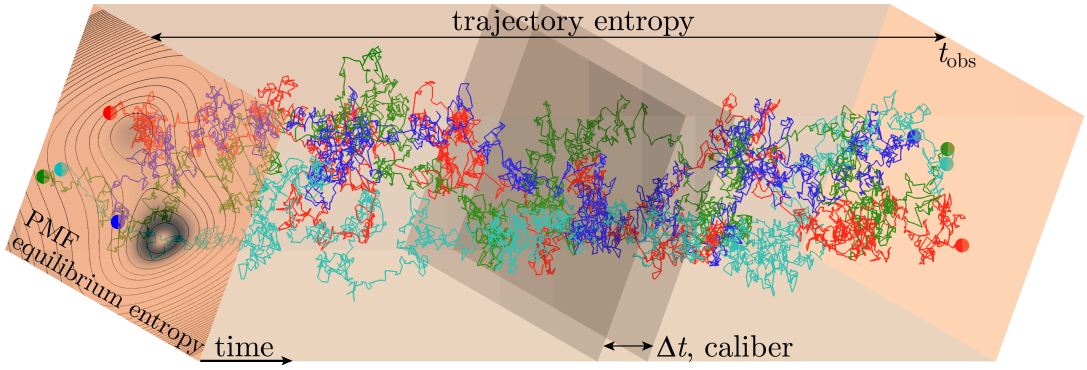

Figure 2.1: The space-time dispersion of a dynamic system. The equilibrium entropy is completely determined by the potential of mean force (PMF, the left-most plane with contour lines), which contains only the static dispersion (the shaded gradients on the PMF plane). Jaynes' caliber evaluates the extent over which the system can explore over a slice of finite width in time, Δt (the slice defined by the two gray panes in the middle). The trajectory entropy from this work resolves the system's static and dynamical dispersions over the full observation volume from $t = 0$ to t_{obs}, and quantifies them analytically.

The Langevin equation has been widely used to describe the dynamical behavior for a variety of applications. Let x represent a fluctuating degree of freedom (or a set of stochastic variables of interest) for which the stochastic differential equation is

$$dx_t = DF(x_t)dt + \sqrt{2D}dW_t \tag{2.1}$$

, where the energy is non-dimensionalized by the thermal energy $k_{\text{B}}T$, D is the diffusion coefficient, $F(x) = -dV(x)/dx$ the deterministic mean force, and $\sqrt{2D}dW_t$ the stochastic force of amplitude $\sqrt{2D}$ exerted by a Wiener process satisfying $\langle dW_t dW_{t'} \rangle = \delta(t - t')dt$. Evolving the Langevin equation generates a trajectory $X(t)$, which is a continuous but non-differentiable function of time that gives a value of x_t at time t with t starting from 0 and ending at t_{obs}. In the ergodic limit, the system reaches an equilibrium distribution, $p_{\text{eq}}(x) = \exp\left(-V(x)\right)/Z_{\text{eq}}$, where $V(x)$ is the potential of mean force (PMF), and Z_{eq} the equilibrium partition function [37]. The static dispersion of the system at equilibrium can be quantified using an entropy measure

$$S[p_{\text{eq}}(x)] = -\int dx p_{\text{eq}}(x) \ln p_{\text{eq}}(x). \tag{2.2}$$

The equilibrium entropy S_{eq} is thus completely determined by the PMF profile and contains no information regarding the dynamics [38–42].

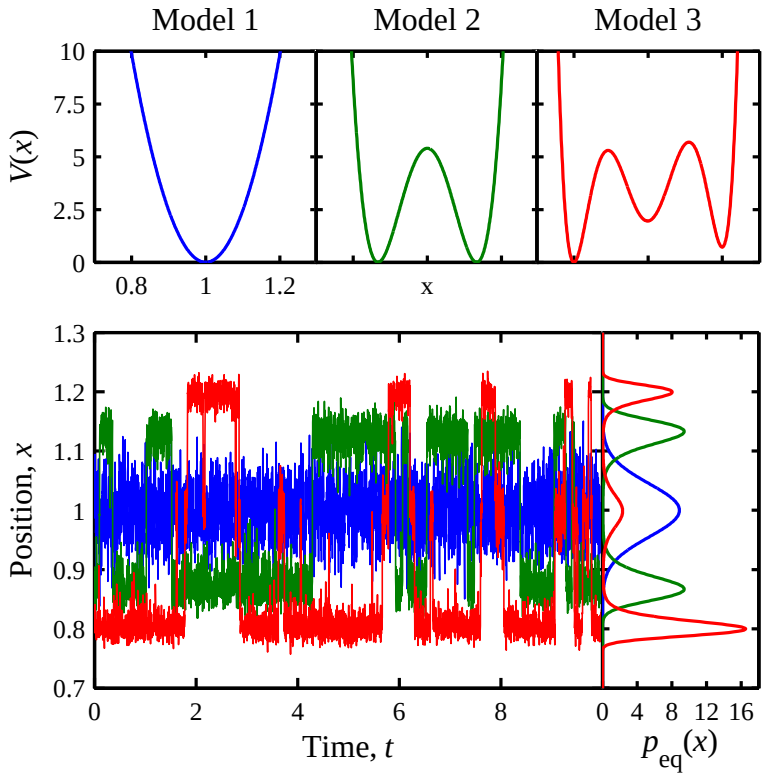

Figure 2.2: Model systems of the same equilibrium entropy but different dynamics information. (top) The PMF $V(x)$ of three systems with 1, 2, or 3 minima. (bottom-left) A simulated Langevin trajectory for the three models with $D = 1$ in dimensionless units. (bottom-right) The Boltzmann equilibrium probability density of the three models.

To consider the dispersion in dynamics, one may use Jaynes's caliber to evaluate the statistics of the conditional propagator, $p(x_{\Delta t}|x_0)$ [43–45]. Caliber was originally defined for finite-state Markov models as the conditional entropy of time propagation probability $S(\Delta t) \equiv \sum_{i,j}^{N} \pi_i r_{i,j} \log r_{i,j}$ [46]. The transition probability from state i to state j per unit time Δt is $r_{i,j}$ and the equilibrium probability of state i is π_i. Caliber may be generalized to the continuous space as the conditional entropy for the propagator $p(x_{\Delta t}|x_0)$ at a time resolution Δt,

$$S(\Delta t) = \int dx_0 dx_{\Delta t}\ p(x_{\Delta t}, x_0) \ln p(x_{\Delta t}|x_0). \tag{2.3}$$

The caliber thus quantifies both the static and the dynamical dispersion the system exhibits within a time slice of Δt.

The evaluation of Eq. (2.3) entails solving the corresponding Fokker-Planck equation or numerically averaging over stochastic trajectories [37]. While the caliber does take into account dynamics information, the relative contributions from the deterministic mean force and the stochastic random force is by no means apparent from the caliber integral. The problem is further exacerbated by the fact that $S(\Delta t)$ depends on the particular choice of Δt. For coarser time resolutions, the conditional probability density converges to $p_{\mathrm{eq}}(x)$ as $\Delta t \to \infty$ and the information about dynamics is lost. In the continuum limit of $\Delta t \to 0^+$, the conditional entropy diverges at a rate of $\sim \ln(D\Delta t)$ due to the non-differentiability of the Weiner process [47, 48]. In this chapter, we show how these issues can be overcome by using the Kullback-Leibler (KL) divergence to quantify the dispersion over the full volume of the trajectory space (rather than analyzing only a slice of it, cf. Figure 2.1). The resulting analytical expression reveals the manner by which the deterministic mean force and the stochastic diffusion contribute to the dynamical dispersion in equilibrium trajectories.

2.3 Kullback-Leibler divergence for trajectory entropy

The KL divergence represents the extra information required to encode a probability density relative to the reference distribution. It is strictly positive and becomes zero only when the queried distribution is identical to that of the reference and is often used to characterize the relaxation of non-equilibrium states back to equilibrium and the entropy production involved [49, 50]. We propose to define the trajectory entropy $\mathcal{S}$ as

$$\mathcal{S} \equiv -\int \mathcal{D}X(t)\ \mathcal{P}[X(t)] \ln \frac{\mathcal{P}[X(t)]}{\mathcal{Q}[X(t)]} \tag{2.4}$$

$\mathcal{P}(X(t))$ is the probability density of obtaining trajectory $X(t)$, $\mathcal{Q}(X(t))$ is the probability density for obtaining the same trajectory from the reference dynamics, and the integration $\mathcal{D}X(t)$ is a path-integral over all continuous functions. Immediately, one sees that the KL divergence removes the $\sim \ln \Delta t$ divergence of the path integral.

The choice of reference dynamics plays the role of further accentuating the information content in $\mathcal{P}(X(t))$ through the path integral of Eq. (2.4). Our

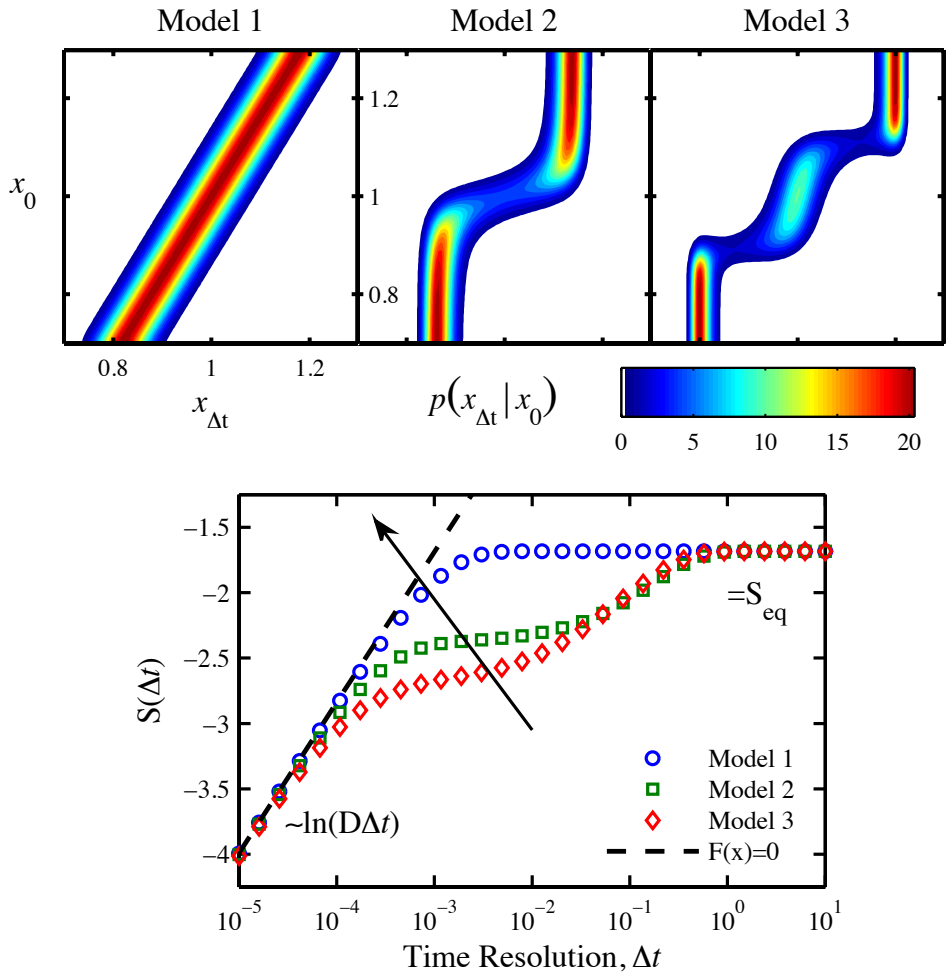

Figure 2.3: Conditional entropy of the time propagator. (top) The contours of the transition probability density $p(x_{\Delta t}|x_0)$ for the BD time propagation of Model 1, 2, and 3 with $\Delta t = 10^{-3}$. (bottom) The conditional entropy as a function of Δt for the 3 model systems. Annotated are the limiting values of conditional entropy when $\Delta t \to \infty$ and $\Delta t \to 0^+$. Details of the numerical calculation of the propagator probability density are given in the Supplementary Information. In all cases, the diffusion constant D is one in the dimensionless unit.

strategy is to use the most structureless Brownian dynamics with zero force $F_{\text{ref}}(x) = 0$. The diffusion coefficient D_{ref} for reference dynamics is then set to help in separating out each the elements in the Langevin equation that contribute to $\mathcal{S}$. The dissection of trajectory entropy outlined below can be understood as a two step thermodynamic integration along (1) the mean force coordinate as the KL divergence between the trajectory probabilities for the Langevin dynamics and those for the Brownian dynamics at the same diffusion coefficient $D_{\text{ref}} = D$ and (2) the diffusion coefficient coordinate as the KL divergence of the Brownian dynamics with D to that of D_{ref}. That is,

$$\mathcal{S}[F(x), D; D_{\text{ref}}] = \mathcal{S}_{F_{\text{ref}} \to F(x)}(D) + \mathcal{S}_{D_{\text{ref}} \to D}(F_{\text{ref}} = 0). \tag{2.5}$$

2.4 Path integral of the Onsager-Machlup (OM) action

The probability density of a Langevin trajectory of duration t_{obs} is proportional to the exponential of the OM action $E^{\text{OM}}[X(t)]$ [51–53]

$$\mathcal{P}[X(t)] = e^{-V(x_0)} / Z_{eq} \left(e^{-E^{\text{OM}}[X(t)]} / \mathcal{Z} \right) \tag{2.6}$$

$$E^{\text{OM}}[X(t)] = \frac{1}{4} \int_0^{t_{\text{obs}}} dt \; \frac{\dot{x}_t^2}{D} + DF^2(x_t) + 2DF'(x_t) \tag{2.7}$$

$$+ \frac{1}{2} \left(V(x_{t_{\text{obs}}}) - V(x_0) \right) \tag{2.8}$$

where the trajectory partition function of $\mathcal{P}[X(t)]$ is $\mathcal{Z} = \int \mathcal{D}X(t) e^{-E^{\text{OM}}[X(t)]}$.

Applying the definition of OM action into Eq. (2.5) with the Brownian dynamics reference results in the following expression for the trajectory entropy[1]

$$\mathcal{S} = \frac{1}{2} \left\langle V(x_0) + V(x_{t_{\text{obs}}}) \right\rangle_{X(t)} + \ln Z_{\text{eq}} \tag{2.9}$$

$$+ \frac{D}{4} \left\langle \int_0^{t_{\text{obs}}} dt \; F^2(x_t) + 2F'(x_t) \right\rangle_{X(t)} \tag{2.10}$$

$$+ \frac{1}{4} \left\langle \int_0^{t_{\text{obs}}} dt \; \frac{\dot{x}_t^2}{D} - \frac{\dot{x}_t^2}{D_{\text{ref}}} \right\rangle_{X(t)} + \ln \frac{\mathcal{Z}}{\mathcal{Z}_{\text{ref}}}. \tag{2.11}$$

[1] For the $F_{\text{ref}} = 0$ reference dynamics we have discarded V_{ref} and set $(Z_{\text{eq}})_{\text{ref}} = 1$ to ignore the unnecessary scalar offsets.

The expectations in Eqs. (2.9)–(2.11) are taken over the distribution of trajectories of the system dynamics. For an arbitrary functional of $X(t)$, $\langle g[X(t)]\rangle_{X(t)} = \int \mathcal{D}X(t)\mathcal{P}[X(t)]g[X(t)]$.

An important consequence of equilibrium dynamics in evaluating the terms of $\mathcal{S}$ in Eq. (2.5) is that for the expectation of any single-time function over the equilibrium trajectories, the result can be obtained by switching the order of integrating over time $\int dt$ and path $\int \mathcal{D}X(t)$ such that $\langle \int dt g(x_t)\rangle_{X(t)}$ becomes $\int dt \langle g(x_t)\rangle_{p_{\rm eq}(x_t)}$. The single-time terms of the trajectory entropy thus have an explicit extensivity of trajectory length $t_{\rm obs}$ timing an integrand that is time invariant after taking the path expectation. This result can also be reached via the Feynman-Kac theorem [54].

2.5 The trajectory entropy functional

The analytical form of the trajectory partition function can be obtained by taking its functional derivative with respect to the mean force

$$\frac{\delta \mathcal{Z}}{\delta F(y)} = \frac{D}{2}\left\langle \int_0^{t_{\rm obs}} dt\ F(x_t)\delta(x_t - y) + \delta'(x_t - y)\right\rangle_{X(t)} \tag{2.12}$$

where $\delta'(x)$ is the derivative of the delta function. Transforming the path expectations to integral over the equilibrium distribution of states and applying integration by parts to the δ' term cancels the term with $F(x_t)$ as $dp_{\rm eq}(x)/dx = F(x)p_{\rm eq}(x)$ and the functional derivative in Eq. (2.12) is thus zero. Since $\mathcal{Z}$ does not depend on $F(x)$, the partition functions of the Langevin trajectories are equivalent to those of the force-free Brownian dynamics [47, 48],

$$\mathcal{Z}(D) = (4\pi D\Delta t)^{(t_{\rm obs}/2\Delta t)}. \tag{2.13}$$

That is, Eq. (2.13) is the normalization for the Weiner process with variance $2D\Delta t$. The exponent, $t_{\rm obs}/\Delta t$, is the time steps used to discretize the trajectory and is the dimensionality of the path integral (see Supplementary Information for more details of this result).

Next, as the trajectory partition function is the cumulant generator of the OM action, the velocity-squared terms in Eq. (2.11) can be obtained by applying $d\ln(\mathcal{Z})/d(1/D)$ to the right hand side of Eq. (2.13) and imposing the definition

of $\mathcal{Z}$ to arrive at

$$D\frac{t_{\text{obs}}}{2\Delta t} = \frac{1}{\mathcal{Z}}\int \mathcal{D}X(t)\frac{d(E^{OM}[X(t)])}{d(1/D)}e^{-E^{OM}[X(t)]} \tag{2.14}$$

$$= \frac{1}{4}\left\langle \int_0^{t_{\text{obs}}} dt\ \dot{x}_t{}^2 - D^2F^2(x_t) - 2D^2F'(x)\right\rangle_{X(t)}. \tag{2.15}$$

Since integration by parts leads to $\langle F'(x)\rangle_{\text{eq}} = -\langle F^2(x)\rangle_{\text{eq}}$, solving Eq. (2.15) gives the result of the path integral of $\dot{x}_t^2$

$$\left\langle \overline{\dot{x}_t^2}\right\rangle_{X(t)} = \frac{2D}{\Delta t} - D^2\left\langle F^2(x)\right\rangle_{\text{eq}}. \tag{2.16}$$

The non-differentiability of Brownian trajectories does cause the velocity-squared term to diverge as expected.

For the components in Eq. (2.10), taking the expectation with the help of integration by parts yields another force squared factor $-t_{\text{obs}}D\left\langle F^2(x)\right\rangle_{\text{eq}}/4$. Finally, the boundary terms of the trajectory entropy in Eq. (2.9) are just the equilibrium entropy since $\langle V(x)\rangle_{\text{eq}} + \ln Z_{\text{eq}} = S_{\text{eq}}$.

Combining the path-integral results of Eqs. (2.9)–(2.11), the dependence of trajectory entropy on $F(x)$ can be recognized as

$$\mathcal{S}_{F_{\text{ref}}\to F(x)} = S_{\text{eq}} + t_{\text{obs}}D\left[\frac{D}{4D_{\text{ref}}} - \frac{1}{2}\right]\left\langle F^2(x)\right\rangle_{\text{eq}}. \tag{2.17}$$

Along the diffusion coefficient coordinate, taking the ratio of trajectory partition functions and adding the $\sim 2D/\Delta t$ terms from the path integral of square velocity leads to the asymptote

$$\mathcal{S}_{D_{\text{ref}}\to D} = \lim_{\Delta t\to 0^+}\frac{t_{\text{obs}}}{2\Delta t}\left[\ln\left(\frac{D}{D_{\text{ref}}}\right) + 1 - \frac{D}{D_{\text{ref}}}\right]. \tag{2.18}$$

It ought be noted that extending the result of Eqs. (2.17) and (2.18) to multiple dimensions only requires a generalized path action and careful integration by parts to give the force expectation term $\langle \vec{F}(x)\cdot\vec{F}(x)\rangle_{\text{eq}}$.

The arbitrariness of reference dynamics in the trajectory entropy functional derived above can in fact be eliminated by employing the most disordered dynamics of $F_{\text{ref}}(x) \to 0$ and $D_{\text{ref}} \to \infty$ as the reference model. Discarding the scalar constants irrelevant to $F(x)$ and D in Eqs. (2.18) and (2.17) gives the principal result of this chapter: the trajectory entropy functional,

$$\mathcal{S}[F(x), D] = S_{\text{eq}} - \frac{t_{\text{obs}}}{2}\left\langle DF^2(x)\right\rangle_{\text{eq}} + \lim_{\Delta t\to 0^+}\frac{t_{\text{obs}}}{2\Delta t}\ln D. \tag{2.19}$$

What emerges from the trajectory entropy in Eq. (2.19) is the firm connection between the spread in Langevin paths and their contributing elements, namely, the statically dispersing potential of mean force (S_{eq}), as well the dynamically dispersing mechanisms resulting from the deterministic mean force ($F(x)$) and the stochastic diffusion (D). The analytical expression clarifies how these two forces of different physical origins are coupled. The stochastic Wiener process with diffusion D alone causes the trajectory entropy to increase in a power-law order equal to the dimensionality of the temporal domain for dynamics. Information being the alter ego of entropy, Eq. (2.19) also sheds new light on the information content in the trajectories of a dynamic system: One sees that each element conducive to the apparent disorder carries a unique and quantifiable piece of information—and those pieces of information are additive.

2.6 Numerical studies of 3 model systems

As an independent validation of the analytical result via numerical calculations, we consider the three examples shown in Figure 2.2. The models were purposely designed to have identical equilibrium entropy, $S_{\text{eq}} = 1.683$, but markedly different dynamical behaviors even with the same diffusion coefficient. In addition to numerically verify the analytical functionals in Eq. (2.19), we show how the visually distinct dynamical features can be quantified.

We start by discussing the conventional caliber measure. Figure 2.3 shows the contours of the time propagation probability density $p(x_{\Delta t}|x_0)$ for the three examples at a fixed Δt and the corresponding caliber profiles, $S(\Delta t)$. The $\Delta t \to \infty$ and $\Delta t \to 0^+$ limits of $S(\Delta t)$ discussed earlier are clearly seen in the numerical results. Evidently, it is necessary to scan the Δt in the caliber measure to assess the dynamical contents. Importantly, the relative contributions due to elements of the dynamical system—the static PMF, the deterministic force, and the stochastic force—are scrambled in the integral. The issues of obscured physical origin and degeneracy remain even if one recasts the caliber measure in the form of KL divergence relative to the Brownian dynamics:

$$S_{\text{KL}}(\Delta t) = -\int dx_0 dx_{\Delta t} p(x_{\Delta t}, x_0) \ln \frac{p(x_{\Delta t}|x_0)}{q(x_{\Delta t}|x_0)} \tag{2.20}$$

(see insert in Figure 2.4).

The analytical expression (Eq. (2.19)) for the proposed trajectory entropy, on the other hand, is seen to be in quantitative agreement with the numerical simulations, summarized in Table 2.1 and Figure 2.4. The dynamical dispersion in Model 3 is approximately ten times more "complex" than that in Model 1.

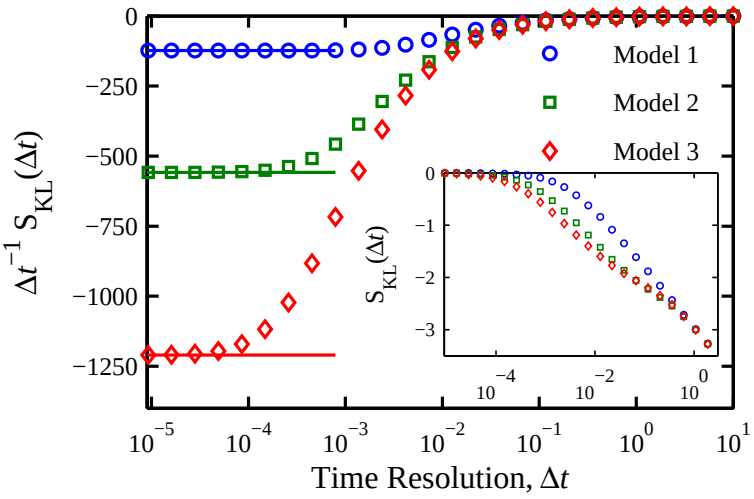

Figure 2.4: Comparison of numerical and analytical approaches of calculating $S_{\mathrm{KL}}(\Delta t)$. The values of $(1/\Delta t)S_{\mathrm{KL}}(\Delta t)$ for the three model systems at different levels of time resolution Δt. The horizontal lines indicate the analytic prediction of $-D/4\langle F^2(x)\rangle_{\mathrm{eq}}$ in the continuum limit.

Importantly, the trajectory entropy enables a direct and immediate identification for the origin of the differing complexity—the deterministic forces. Indeed, since the diffusion coefficients of the models are identical, it is due to the additional forces required for making three minima instead of one within the region. More illustrative examples can be found in Supplementary Information.

Table 2.1: Numerical calculations of $S_{\mathrm{KL}}(\Delta t)$ for the three model systems and analytic entropy functional. The reference state $q(x_{\Delta t}|x_0)$ is from the force-free BD with the same diffusion constant $D = 1$ in reduced units as the queried time propagator. When D_{ref}=D, the relation between the trajectory entropy and $S_{\mathrm{KL}}(\Delta t)$ leads to $\lim_{\Delta t\to 0} S_{\mathrm{KL}}(\Delta t)/\Delta t = -D/4\left\langle F^2(x)\right\rangle_{\mathrm{eq}}$ as derived in Supplementary Information.

Functionals	Model 1	Model 2	Model 3
$\int dx p_{\mathrm{eq}}(x) \ln p_{\mathrm{eq}}(x)$	1.683	1.683	1.683
[2] $-\Delta t^{-1} S_{\mathrm{KL}}(\Delta t)$	123.6	558.6	1210
[3] $D/4\langle F^2(x)\rangle_{\mathrm{eq}}$	123.5	558.7	1211

2.7 Concluding perspective

This work illustrates how the path integration of the entire continuous stochastic trajectory ensemble can be compressed into an analytic functional. The results provide a foundation for understanding the physical origins of the dynamical dispersion due to fluctuations in systems that can be modeled by an over-damped Langevin equation. The simplicity of the analytical results makes it appealing for immediate practical applications. In optimization and statistical learning, for instance, they may be used for Bayesian estimation of time-propagation parameters from data generated by experiments or simulations [55, 56], where the trajectory entropy may be used to deduce the most reduced dynamics representation in cases where the observables were obtained with insufficient information. Another potential application is to use them as design equations to engineer the information processing in dynamic systems. For systems that propagate quantum information, for example, the commonly employed modeling equations are isomorphic to the Langevin equation discussed here [57]. Finally, it shall be enlightening to generalize the ideas to non-equilibrium cases.

CHAPTER 3

Path Integral Statistical Learning Theory: Extracting Force and Diffusion from single-molecule FRET

3.1 Abstract

We present a comprehensive theoretical framework to extract the potential of mean force (PMF) and diffusion coefficient from single-molecule FRET (sm-FRET) experiments. The likelihood for such an experiment is developed by exactly solving a continuous path integral of the Poisson statistics of photon emission and the Langevin dynamics of the system under study. The solution is aided by an eigen-decomposition of the Fokker-Planck operator which governs the propagation of system probabilities. Optimization of the parameterizing PMF and diffusion constant for arbitrary functions required the adaptation of the expectation-maximization algorithm to the calculus of variations. The solution space is regularized by highlighting the parameters of dynamics which produce maximum entropy equilibrium trajectories. Results are presented for a hypothetical test cases which reproduces many experimentally relevant time and energy scales under feasible conditions for the smFRET apparatus.

3.2 Introduction

Direct observation of individual proteins is the most straightforward way of elucidating the driving forces of conformational dynamics. Without the information loss due to measuring the summed behaviors of an ensemble of entities, single-molecule fluctuations provide first-hand data of the free-energy landscape and stochastic diffusion underlying the transitions in biomolecule conformation. However, the currently available single-molecule methods rely on attaching probes to the system to direct the focus on one molecule at a time and only offer a convoluted view of the dynamics of the tagged molecule.[58, 59] As a result, the recorded data need to be decoded to resolve the time propagation of the protein of interest. This inference depends on the statistics of the dynamic coupling between the tracked degree of freedom and the reporting signal as well as the temporal resolution and duration of the trajectory recorded in experiments. The heavily fluctuating dynamics at the single-molecule level makes the data analysis of extracting maximum information from indirect measurements a very challenging task.[34, 60–64]

Taking smFRET as an example, a typical setup is using a pair of fluorescent dyes, a donor and an acceptor, to attach to the ends of a surface-immobilized protein, Figure 3.1. Following laser excitation at the relevant frequency, an excited donor dye can relax to its ground state by emitting a "green" photon or transferring the energy to the nearby acceptor dye that may then emit a "red" photon to go back to the ground state. The energy-transfer efficiency between dyes depends on the donor-acceptor distance r as $\zeta(x) = 1/(1+x^6)$ with $x = r/R_0$ and R_0 being the Förster radius for the acceptor-donor pair. In this case, the donor-acceptor distance r, or equivalently, x, is the protein dynamics degree of freedom that one wishes to learn about. The photons emitted from the tagged molecule can be captured by confocal microscopy and recorded by avalanche photodiode [33]. The statistics of photon arrival times follows that of a Poisson process with the density depending x parametrically:

$$I_a(x) = I_a^0 \zeta(x) + B_a \quad \text{(acceptor)} \tag{3.1}$$

$$I_d(x) = I_d^0 (1 - \zeta(x)) + B_d \quad \text{(donor)}. \tag{3.2}$$

Here, $I_{a,d}^0$ are the maximum intensities and $B_{a,d}$ are the background signals of the two types of photons.

Since the arrival time of each emitted photon is recorded,[65] the waiting-time distributions of the acceptor (a) and donor (d) photons, $\Delta t_{a,d}$, following the exponential probability density function with intensities $I_{a,d}$ describe the

statistical coupling between the latent variable and the observed signal:

$$p(\Delta t_{a,d}|I_{a,d}) = I_{a,d}e^{-I_{a,d}\Delta t_{a,d}}. \tag{3.3}$$

Within an infinitesimal time slice dt, one of the three observations would occur with the probability densities depending on the latent variable of the system state at the moment, x_t. The following list of the three outcomes and their probability characteristics involve the parameter of total intensity defined as $I(x_t) = I_d(x_t) + I_a(x_t)$.

1. An acceptor photon arrives, and the probability density of this event is:

$$p(\Delta t_a = dt|x_t)p(\Delta t_d > dt|x_t) = I_a(x_t)e^{-I(x_t)dt}. \tag{3.4}$$

2. A donor photon arrives, and the probability density of this event is:

$$p(\Delta t_d = dt|x_t)p(\Delta t_a > dt|x_t) = I_d(x_t)e^{-I(x_t)dt}. \tag{3.5}$$

3. No photon arrives. This dt instance is considered "dark", and the probability of this event is given by:

$$P(\Delta t_{a,d} > dt|x_t) = e^{-I(x_t)dt}. \tag{3.6}$$

The probability of observing both acceptor and donor photons in dt is extremely rare and this event is hence ignored.

Therefore, the information of protein dynamics along x is encoded in the sequence of the colors and arrival times of photons that depend on the system state probabilistically according to Eqs. 3.4-3.6. To what extent of the dynamics of x can be learned from the photon sequences recorded in a particular sm-FRET experiment? The guiding dynamics model we employ in this work for addressing this question is the over-damped Langevin equation:

$$dx_t = DF(x_t)dt + \sqrt{2D}dW_t. \tag{3.7}$$

In this model of time propagation, the potential of mean force (PMF) is related to the equilibrium probability density of x, $p_{\text{eq}}(x)$, as $V(x) = -\ln(p_{\text{eq}}(x))$. The mean force $F(x) = -\nabla V(x)$ is the deterministic component in the equation of motion. The stochastic force component is parameterized by the diffusion coefficient D and the Weiner process dW_t has the average $\langle dW_t \rangle = 0$ and the

variance $\langle dW_t \cdot dW_{t'} \rangle = \delta(t - t')dt$.[1] Although the theoretical development and numerical illustration presented in this work focus on the Langevin dynamics with a constant diffusion coefficient, generalization to x dependent diffusion is expected to be straightforward. The Langevin dynamics of Eq. 3.7 captures the continuous nature of protein conformational fluctuations and the low Reynolds number of biomolecular systems in the condensed phase.

Ideally, one would like to learn about the PMF and diffusion coefficient in Eq. 3.7 from the sequence of photon colors and arrival times recorded in a smFRET experiment. Under the assumption that x is stagnant until a sufficient number of photons is collected to estimate the value of the latent variable with a satisfactory certainty, a histogram of the values of x inferred at the specified level of uncertainty can be generated from a photon trajectory. Although this approach is general, free from limiting x to a set of discrete states, and readily applicable to process experimental data, information loss is inevitable by the single-value coarse-graining of the fluctuations during the counting time for reaching the set criterion of certainty. The ever-present movements in protein conformation can in theory be faithfully retained in the inference problem by setting up a statical learning model with x as the latent variable. However, the development of this approach has been limited to a coarse-grained description of system dynamics as jumping between discrete states[66–69] instead of the continuous stochastic dynamics of the Langevin equation. This treatment mixes the contributions from the deterministic and stochastic forces that govern protein dynamics into a rate constant matrix connecting different states. In addition to the information loss due to model simplification, the number of stable states along the x coordinate is in general unknown *a priori*. As a result, the practical applicability of discrete statistical learning in retrieving mechanistic understanding from smFRET data is severely limited.

Retrieving the dynamics parameters of protein conformational changes from single-molecule experiments thus relies on using a guiding model that can capture the essence of biomolecule fluctuations. The joining of deterministic and stochastic forces in the Langevin equation has a sound physical origin of the projection operator formalism and retains the spatial and temporal continuity of molecular mechanics and dynamics. However, the difficulties of infinite dimensionality, non-differentiability in time, and path integral need

[1] Throughout the text, the physical variables presented are nondimensionalized by the thermal energy $k_B T$ at a fixed temperature T as the characteristic energy, the Föroster radius R_0 as the characteristic length, and the timescale $\bar{t} = 1$s as the characteristic time. That is, $\bar{V}(r/R_0)/k_B T \to V(x)$, $\bar{F}(r/R_0)R_0/k_B T \to F(x)$, $\bar{D}\bar{t}/R_0^2 \to D$, and $\overline{I_{a,d}}\bar{t}(r/R_0) \to I_{a,d}(x)$. Variables with an overbar are the actual quantities before nondimensionalization.

to be overcome in order to employ Eq. 3.7 as the latent dynamics model for statistical learning. This work presents our analytical, numerical, and statistical developments that make possible this goal. Although the methodology was devised for the specific case of using smFRET to study protein conformational changes, the established foundation for statistical leaning of continuous stochastic dynamics may as well be applied to other single-molecule methods such as pulling using atomic force microscope or optical tweezer through molecular tags to transmit forces. Since the free-energy landscape and diffusion coefficient of conformational dynamics can also be constructed from the bottom up via performing molecular simulations with path-based methods of sampling and optimization, [27, 28, 70] the availability of experimental data of the same type can greatly facilitate cross-validation and tracing the atomistic origin and controlling amino acids of protein dynamics.

The rest of this paper is organized as the following. We first present the Bayesian inference framework that we employed for the statistical learning of Langevin dynamics from smFRET data. Theoretical developments for calculating the likelihood function of the PMF and diffusion coefficient through a trajectory path integral are presented next. This procedure can also be used to infer the trajectory probability densities of the latent variable, $X(t)$. We then derive the functional derivatives of the likelihood function with respect to the Langevin dynamics parameters given the observed photon trajectory. With these elements established, an expectation-maximization optimization of the Langevin model can be performed to derive the optimal PMF and diffusion coefficient that best describe the observed photon sequence. Application of this methodology to a highly non-trivial test case is presented at the end followed by the conclusion.

3.3 The Bayesian Inference Framework of smFRET for Continuous Stochastic Dynamics

Behind the scene of photon recording, the trajectory of the tagged protein degree of freedom, $X(t)$, was not observed directly. The statistics regarding the PMF profile and diffusion coefficient are thus not explicit in the photon trajectory. The structure of this convolution is best represented via a Bayesian Graphical Model (BGM) as shown in the bottom panel of Fig. 3.1. Vertical arrows in the BGM link the experimental observable at time t, y_t=donor, acceptor, or darkness, and the latent protein conformation variable at the same time, x_t, as the conditional probability density of $p(y_t|x_t)$. Following Eqs. 3.4 to 3.6, there

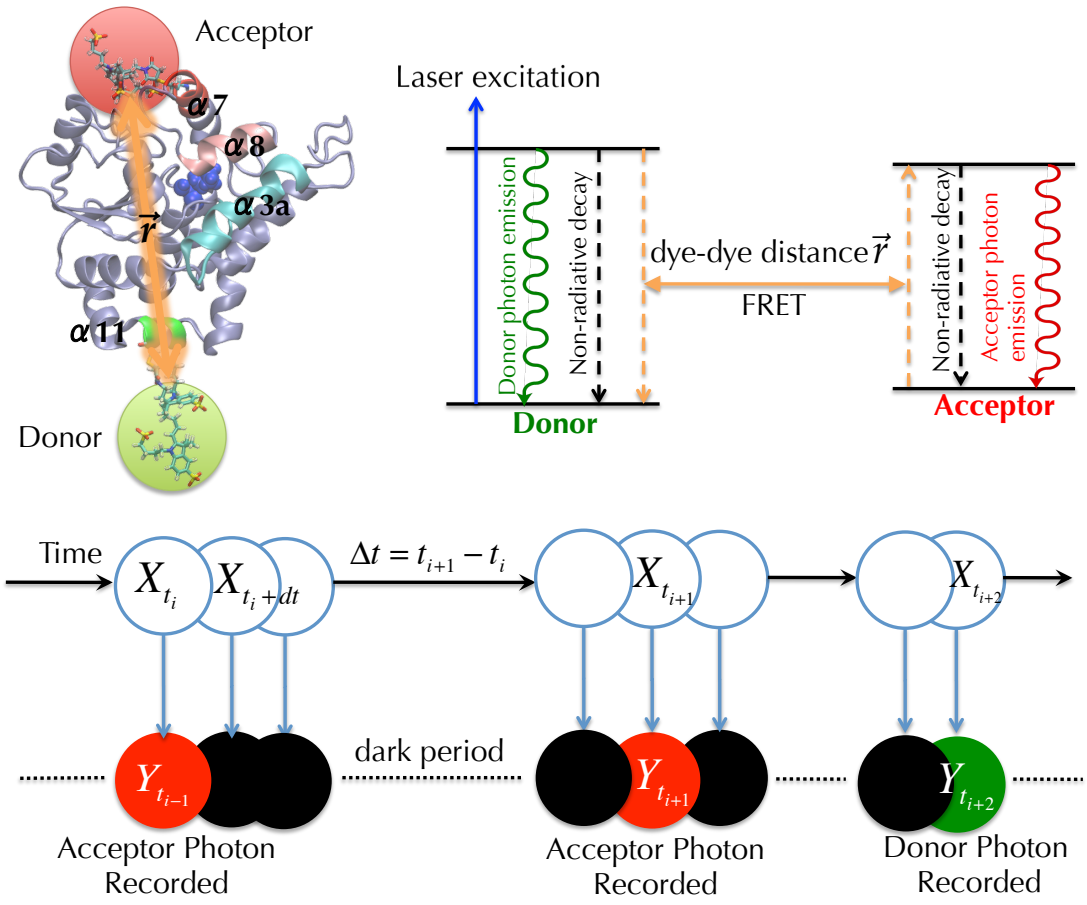

Figure 3.1: A schematic representation of smFRET experiments. (Top-left) The dye-attached structure of Mycobacterium tuberculosis protein tyrosine phosphatase, PtpB[71]. The structural segments of PtpB that cover the active site (blue balls) of the enzyme, $\alpha 7$, $\alpha 8$, and $\alpha 3a$, are highlighted. The acceptor and donor dyes are attached at the ends of $\alpha 7$ and $\alpha 11$, respectively. (Top-right) The Jablonski diagram of the energy states in the FRET flourophores and the energy transfer event. The efficiency of energy transfer depends on the inter-dye distance r and the Förster radius R_0. The dimensionless distance x is r/R_0. (Bottom) The graphical model of the continuous Bayesian inference for Langevin dynamics from smFRET measurements. Clear circles represent the latent observables of the system trajectory, $X(t)$. $X(t)$ is a continuous function of time that gives the value of x at a specific time t, i.e., x_t. The filled circles represent $Y(t)$, the experimentally recorded photon trajectory. At a specific time, the readout of the photon trajectory, y_t, is either a donor photon, an acceptor photon, or darkness. Horizontal arrows represent the conditional probability densities of the time evolution of the non-demensionalized inter-dye distance, $p(x_{t+dt}|x_t)$, and vertical arrows represent the probabilities of photon emission, $p(y_t|x_t)$.

are two different classes of observations. The instantaneous event of observing a photon is represented by taking the limit of $dt \to 0$, and the position dependent probability density functions of $p(y_t|x_t)$ are:

$$p(y_t|x_t) = \begin{cases} I_a(x_t) & y_t = \text{acceptor photon} \\ I_d(x_t) & y_t = \text{donor photon.} \end{cases} \tag{3.8}$$

Alternatively, if the state of darkness was observed over the infinitesimally small, but nonzero interval dt, performing time integration in the BGM framework spans a dark duration of the specified size along the trajectory. This observation also depends on x with the probability of:

$$P(y_t|x_t) = \begin{cases} e^{-I(x_t)dt} & y_t = \text{darkness.} \end{cases} \tag{3.9}$$

On the other hand, the horizontal arrows in the BGM indicate the conditional probability densities the time propagation of the latent variable, $p(x_{t+dt}|x_t)$, and embody the dynamics of interest. The Langevin parameters that determine $p(x_{t+dt}|x_t)$ can only be learned from the trajectory of y_t.

The inversion of smFRET measurements into the PMF and diffusion coefficient of the Langevin equation via the BGM framework comes down to solving the following two problems:

Inference What is the probability density of the dynamic trajectory of the protein degree of freedom of interest, i.e., $X(t)$, given a sequence of photon arrival times and colors recorded via smFRET, $Y(t)$? In other words, with a trial mean force profile $F(x)$ and diffusion coefficient D of the Langevin equation, one aims to calculate:

$$\mathcal{P}(X(t)|Y(t); F(x), D). \tag{3.10}$$

Optimization What is the optimal profiles of force $F(x)$ and diffusion coefficient D for describing the observed photon trajectory? The answer is finding the supremum of (maximizing) the likelihood functional:

$$\sup_{F(x),D} \mathcal{P}(Y(t); F(x), D). \tag{3.11}$$

Solving the inference and optimization problems stated above requires a path integral over the coordinate space of the probability density of a system trajectory $X(t)$ given the smFRET observation of $Y(t)$:

$$\mathcal{P}(Y(t); F(x), D) = \int \mathcal{D}X(t) \mathcal{P}(X(t), Y(t); F(x), D). \tag{3.12}$$

The differential volume of the trajectory space is $\mathcal{D}X(t)$. The theoretical development presented later illustrates how to perform such calculation based on the specific time stamps and colors of the arriving photons that also incorporates the intermediate times of "dark" periods into consideration. The capability of performing smFRET inference with the continuous profile of $F(x)$ as the basis eliminates the requirement of prior knowledge of the number of metastable conformational states. This information would simply emerge as a result of the optimization. Based on the BGM of Fig. 3.1 and the conditional independencies of probabilities prescribed therein, the joint probability density of the latent trajectory $X(t)$ and the recorded photon trajectory $Y(t)$ can be factorized as:

$$\mathcal{P}(X(t), Y(t); F(x), D) = \mathcal{P}(Y(t)|X(t))\mathcal{P}(X(t); F(x), D). \tag{3.13}$$

Although the theoretical developments are general, we perform analysis and illustration of the statistical learning algorithm with the model potential shown in Fig. 3.2. The PMF contains two barriers of around $5k_BT$ that are biologically relevant for proteins conformational changes. The two barriers connect two metastable states corresponding to a short and long inter-dye distance with an intermediate region locating at the value of the Föster radius, $x = r/R_0 \approx 1$. The diffusion coefficient is $D = 500$ in the dimensionless unit that is approximately $1 \times 10^{-14}\text{cm}^2/\text{s}$. Photon trajectories of smFRET experiments are simulated by propagating the Langevin equation with the aforementioned PMF and D coupled with a Kinetic Monte-Carlo (KMC) scheme for simulating the processes of photon emission; the Supplementary Information contains more details of this numerical procedure.

3.4 Calculation of the Trajectory Probability Density of Langevin Dynamics with Photon Data and the Likelihood Function

Eq. 3.13 indicates that the joint probability density of $X(t)$ and $Y(t)$ can be calculated based on the knowledge of the equation of motion which determines the trajectory probability distribution $\mathcal{P}(X(t); F(x), D)$ and the waiting time distributions of photon events using Eq. 3.8 and Eg. 3.9. Since both the Langevin equation and the dark snapshot probability (Eg. 3.9) do not have explicit time dependence, they can be propagated forward in time together in the calculation of $\mathcal{P}(X(t), Y(t))$ as shown later. At the specific instances of having bright milestones, $\{t_\tau \mid \forall\tau \in [0, N_P]\}$ where N_P is the total number of recorded photons, the probability density discussed in Eq. 3.8 is used to

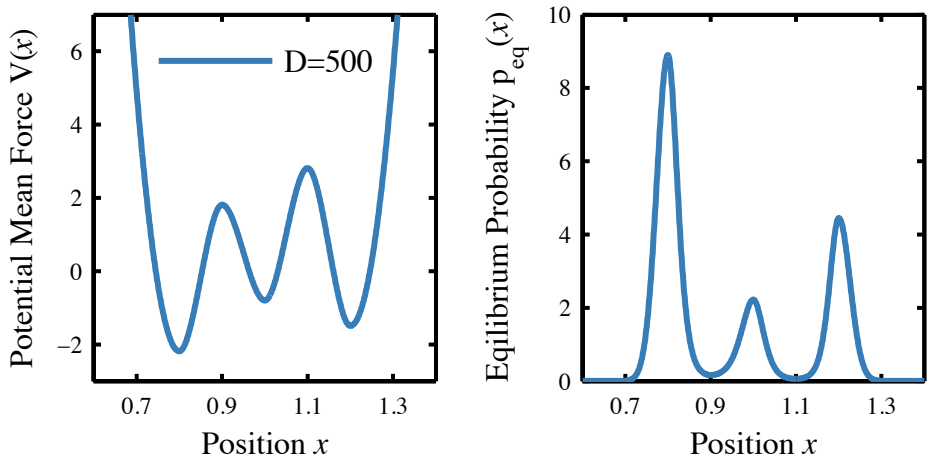

Figure 3.2: The potential of Mean force and the corresponding equilibrium probability density distribution of x used for simulating smFRET trajectories and applying the statistical learning algorithms developed in this work.

Table 3.1: The simulation parameters of smFRET employed in this work. These values were motivated by the typically encountered numbers in experiments. N_{P} is the number of photons observed before the first photo bleaching event occurred and $\langle t_{\mathrm{exp}} \rangle$ is the average duration of a trajectory with these intensities and the number of photons.

Intensity	(s^{-1})
I_d^0	15 000
I_a^0	8000
B_d	10
B_a	20
N_{P}	40 000
$\langle t_{\mathrm{exp}} \rangle$	3.3s

mark $\mathcal{P}(X(t), Y(t))$. The complete likelihood function, or $\mathcal{P}(X(t), Y(t))$, for a trajectory of duration t_{exp} can be expressed as:

$$\mathcal{P}(X(t), Y(t)) = p(x_{t_0}) \prod_{\tau=1}^{N_{\text{P}}} p(y_{t_\tau}^{a,d} | x_{t_\tau}) p(x_{t_\tau}, y_{[t_\tau, t_{\tau-1}]}^{\text{dark}} | x_{t_{\tau-1}}). \tag{3.14}$$

The notation of $y_{[t_\tau, t_{\tau-1}]}^{\text{dark}}$ in this equation indicates that during the time window between the arrivals of photon τ and photon $\tau - 1$, $\Delta t_\tau = t_\tau - t_{\tau-1}$, the recorded observation in the smFRET experiment is dark. On the other hand, $y_{t_\tau}^{a,d}$ denotes the photon color (acceptor or donor) observed at the time t_τ. A similar construction was offered in[72] for a desecrate state Markov representation.

An important message from the above analysis is that in the inference from smFRET measurements, observing the τ^{th} photon also involves the statistical information on the latent variable for the dark period of Δt_τ. Effective and accurate calculation of $p(x_{t_\tau}, y_{[t_\tau, t_{\tau-1}]}^{\text{dark}} | x_{t_{\tau-1}})$ is thus essential to the success of the statical learning algorithm. To evaluate this term, the dark period is divided into $\Delta t_\tau / dt$ slices to perform the path integral over x_{t_i}, $i = (1, ..., \Delta t_\tau / dt - 1)$, and $t_i = t_{\tau-1} + idt$:

$$p(x_{t_\tau}, y_{[t_\tau, t_{\tau-1}]}^{\text{dark}} | x_{t_\tau}) = \int \cdots \int \prod_{i=1}^{\Delta t_\tau / dt - 1} dx_{t_i} \\ p(y^{\text{dark}} | x_{t_i}) p(x_{t_i} | x_{t_{i-1}}) \delta(X(\Delta t) - x_{t_\tau}). \tag{3.15}$$

Here, $p(x_{t_i} | x_{t_{i-1}})$ is the dynamic propagator over a time step dt. Taking the limit of $dt \to 0$ and considering $p(y^{\text{dark}} | x_{t_i}) = \exp(-I(x_{t_i}) dt)$ is the exponential of a Riemann integral over time, the following path expectation is arrived:

$$p(x_{t_\tau}, y_{[t_\tau, t_{\tau-1}]}^{\text{dark}} | x_{t_{\tau-1}}) = \\ \mathbb{E}_{X(t)} \left[e^{-\int_0^{\Delta t_\tau} dt' I(X(t'))} \delta(X(\Delta t) - x_{t_\tau}) \mid X(0) = x_{t_{\tau-1}} \right]. \tag{3.16}$$

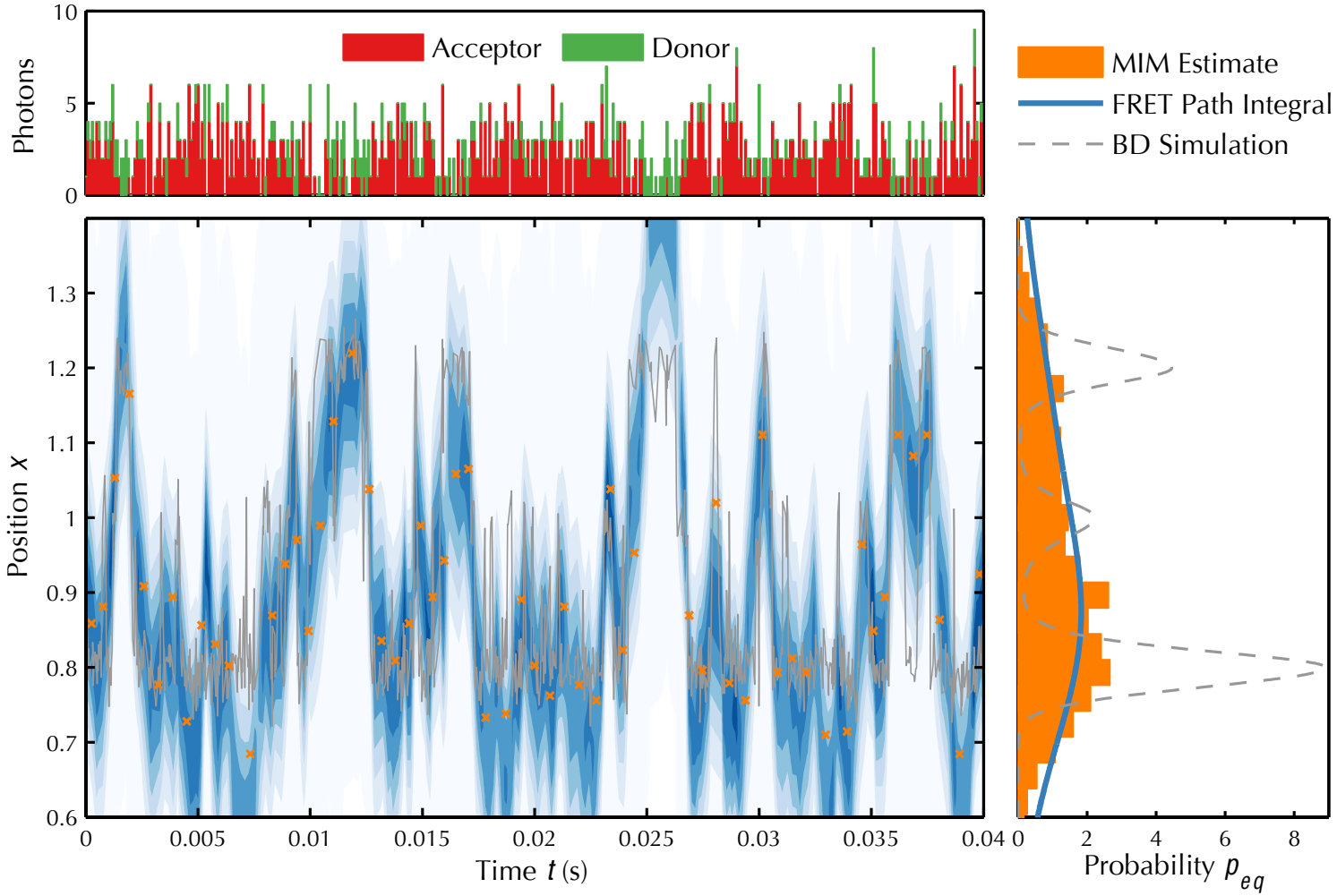

Figure 3.3: Comparison of Brownian Dynamics simulation and the resulting inferred trajectory from smFRET path integral with x2 resolution starting with $p_0 = \cos^2(x/L)$ initial distribution. (top-left) Trace of photon arrivals per millisecond for the donor and acceptor channel. (bottom-right) Time-averged position $p_{\text{eq}} = 1/t_{\text{rxn}} \int_0^{t_{\text{rxn}}} \delta(x - x')$. (bottom-left) Contours of the $\langle \alpha(t)|x\rangle$ and $\langle x|\beta(t)\rangle$ vectors in log space of color intensity. Lines are the raw Brownian dynamics simulation of the system on the free energy surface. Points (X) are the estimates from the MIM with standard error of $\sigma = 0.1$.

Following the Feynman-Kac theorem[34, 73], the probability density defined in Eq. 3.16 can also be obtained by solving the following partial differential equation (PDE) with the no-flux boundary conditions:

$$\frac{\partial p(x,t)}{\partial t} = \Big(D\nabla^2 - \nabla DF(x) - I(x)\Big)\, p(x,t). \tag{3.17}$$

Here, $p(x,t)$ is a shorthand notation of $p(x_{t_\tau}, y^{\text{dark}}_{[t_\tau,t_{\tau-1}]}|x_{t_{\tau-1}})$. The variable x at time t corresponds to x_{t_τ}, i.e., $t_\tau \equiv t$, and is the object of the gradient operators in Eq. 3.17. The initial distribution of probability density $p(x,0)$ represents the condition at $t_{\tau-1}$ in Eq. 3.16, and $t_{\tau-1} \equiv 0$ in this convention. Without the $I(x)$ term, Eq. 3.17 is the Fokker-Plank equation of the Langevin equation of motion defined in Eq. 3.7. The dark operator $I(x)$ implies $y^{\text{dark}}_{[t_\tau,t_{\tau-1}]}$ for the solution of Eq. 3.17.

A key advancement made this work is the recognition that a symmetric version of the PDE in Eq. 3.17 can drastically simplify the calculation of the likelihood function Eq. 3.12 via the path integral over $X(t)$. In particular, a new variable is defined as $\rho(x,t) = p(x,t)/\sqrt{p_{\text{eq}}(x)}$ to transform the PDE such that a Hermitian operator of time propagation emerges:

$$\frac{\partial}{\partial t}\rho(x,t) = -\boldsymbol{H}\rho(x,t) \tag{3.18}$$

$$\boldsymbol{H} = -D\nabla^2 + D\frac{\nabla F(x)}{2} + \frac{DF(x)^2}{4} + I(x). \tag{3.19}$$

The solution of this Hermitian PDE can be simply written as:

$$\rho(x,t) = e^{-\boldsymbol{H}t}\rho(x,0). \tag{3.20}$$

Along the same token, the photon arrival probability densities of Eq. 3.8 can be written in an operator form in the evaluation of the likelihood function of $(F(x), D) \equiv \theta(x)$. The bright operator, $\boldsymbol{y}_\tau$, would appear N_{P} times at the time stamps of $\{t_\tau, \tau \in [1, N_{\text{P}}]\}$:

$$\boldsymbol{y}_\tau = \begin{cases} \boldsymbol{y}^a \equiv I_a(x) & y_{t_\tau} = \text{acceptor photon} \\ \boldsymbol{y}^d \equiv I_d(x) & y_{t_\tau} = \text{donor photon.} \end{cases} \tag{3.21}$$

Performing the path integral of Eq. 3.12 via the factorization of Eq. 3.14 can now be represented via the Dirac notation[74] as a series of time propagations in the dark followed by the event of recording a photon:

$$\mathcal{P}(Y(t)|F(x),D) = \mathcal{P}(Y(t)|\theta(x)) = \mathcal{L}[\theta(x)] = \langle\alpha_0|e^{-\boldsymbol{H}\Delta t_1}\boldsymbol{y}_1 e^{-\boldsymbol{H}\Delta t_2}\boldsymbol{y}_2 \ldots e^{-\boldsymbol{H}\Delta t_{N_{\text{P}}}}\boldsymbol{y}_{N_{\text{P}}}|\beta_{t_{\text{exp}}}\rangle. \tag{3.22}$$

In this representation, the "bra" state $\langle \alpha_\tau |$ carries the probability amplitude of the system state at t_τ given all the photon data in between $[0, \tau]$ and the "ket" state $|\beta_\tau\rangle$ contains the probability density of the latent variable at the same time given all of the future points of the smFRET recordings. Path integral across the entire duration from smFRET initiation to the collection of the last photon is just the inner product of these "bra-ket" pairs. Therefore, inferring the trajectory of the latent variable x via all of the recorded photon data, i.e., solving the inference problem defined in Eq. 3.10, one can follow the Copenhagen interpretation in Quantum Mechanics[75] to obtain:

$$p(x_t|Y(t)) = \frac{\alpha(x_t, y_{[0,t]})\beta(y_{(t,t_{\text{exp}}]}|x_t)}{\mathcal{P}(Y(t))} = \frac{1}{\mathcal{L}}\langle \alpha_t|x\rangle\langle x|\beta_t\rangle. \tag{3.23}$$

The parametric dependence of the terms in Eq. 3.23 on $\theta(x)$ has been ignored to avoid over-complication of the notation.

Under the external-force free operation of smFRET, the initial and final state, $\langle \alpha_0|$ and $|\beta_{t_{\text{exp}}}\rangle$, respectively, are assumed to follow the equilibrium distribution $\rho_{\text{eq}}(x) = \sqrt{p_{\text{eq}}(x)}$:

$$\langle \alpha_0|x\rangle = \rho_{\text{eq}}(x) \quad and \quad \langle x|\beta_{t_{\text{exp}}}\rangle = \rho_{\text{eq}}(x). \tag{3.24}$$

The initial and final states can also be constructed by using the Hamiltonian propagator of the Langevin equation without the dark operator, $\boldsymbol{H}^0$, and extending the temporal domain to infinite times since

$$\langle 1|e^{-\boldsymbol{H}^0 t}|x\rangle|_{t\to\infty} \to \sqrt{p_{\text{eq}}(x)}. \tag{3.25}$$

As such, the likelihood function can be written as:

$$\mathcal{L} = \text{tr}\left[e^{-\boldsymbol{H}^0\infty}e^{-\boldsymbol{H}\Delta t_0}\left(\prod_\tau^{N_{\text{P}}}\boldsymbol{y}_\tau e^{-\boldsymbol{H}\Delta t_\tau}\right)e^{-\boldsymbol{H}^0\infty}\right]. \tag{3.26}$$

Much of this formulation resembles the structure of quantum dynamics in the form of the density matrix.[76]

Progress in evaluating the path integral of Eq. 3.22 or the trace operation of Eq. 3.26 can be made by seeking eigen-decomposition of the Hermitian operator of Eq. 3.18 to obtain the eigenbasis $\psi_i(x)$ that resolves the identity operator, $\mathbf{1} = \sum_i |\psi_i(x)\rangle\langle\psi_i(x)|$. Inserting this identity in between each operator in Eq. 3.22 transforms the path integral or trace operation into matrix multiplications for collecting statistics. This exploitation of the Hermitian nature of the time

propagator plays a critical role in making possible the statistical learning of continuous stochastic dynamics. Although the operation in Eq. 3.22 or Eq. 3.26 can be performed forward or backward in time, we generally start from time zero with the vector given by Eq. 3.24 and perform the matrix operation to the right. Next, we present the procedure we devised for eigen-decomposition of the Hermitian time propagator, evaluation of the likelihood function, and inference of the latent trajectory given the recorded photon sequence.

3.5 Diagonalization of $\boldsymbol{H}$ for eigenbasis

The procedure presented below for eigen-decomposition is not unique but allows for computational feasibility of the path integral. Diagonalization of the Hermitian operator was performed by using a spectral element method(sFEM).[77] First, we solve the symmetric Fokker-Planck equation of the Langevin dynamics listed below and use the resulting eigen-basis to solve the PDE of Eq. 3.18 with the dark operator.

$$\frac{\partial \rho(x,t)}{\partial t} = -\boldsymbol{H}^0 \rho(x,t) \tag{3.27}$$

$$\boldsymbol{H}^0 = -D\nabla^2 + V_{eff}(x) \tag{3.28}$$

$$V_{eff} = \frac{DF'(x)}{2} + \frac{DF^2(x)}{4} \tag{3.29}$$

The Hermitian $\boldsymbol{H}^0$ gives a set of orthonormal basis $\langle \psi_i^0 |$ that resolves the identity operator $\mathbf{1} = \sum_i |\psi_i^0\rangle\langle\psi_i^0|$. The time dependence of $\rho(x,t)$ in Eq. 3.27 can be accounted for as:

$$\rho(x,t) = \sum_i c_i \langle x|\psi_i^0\rangle e^{-\lambda_i^0 t}. \tag{3.30}$$

The coefficients c_i's are time invariant and can be determined, for example, based on the initial distribution of $\rho(x,0)$. The eigen-values satisfy:

$$\boldsymbol{H}^0|\psi_i^0\rangle = \lambda_i^0|\psi_i^0\rangle. \tag{3.31}$$

The finite-time propagation of $\boldsymbol{H}^0$ can be represented by constructing the matrix Ψ^0 that contains the eigenvectors as the columns and the diagonal matrix Λ^0 composed of the eigenvalues:

$$e^{-\boldsymbol{H}^0 \Delta t} = \Psi^0 e^{-\Lambda^0 \Delta t} \Psi^{0^\dagger}. \tag{3.32}$$

Given a PMF profile and diffusion coefficient for the Langevin equation, the eigenvalues and eigenvectors via a highly accurate spectral element method with localized polynomials as the interpolation function in the elements. Details of this numerical solution are provided in the Supplementary Information. We found robust convergence with $N_E = 256$ elements with $N_L = 7$ order polynomials for all of the systems we have analyzed. In particular, the spectral elements, $u_n(x)$'s, are used to expand the scaled eigen-vectors by the square root of the equilibrium probability density, $\rho_{eq}(x)$:

$$\psi_i^0(x) = \boldsymbol{H}^0(x)\phi_i^0(x) \tag{3.33}$$

$$\phi_i^0(x) = \sum_n c_n^0 u_n(x). \tag{3.34}$$

In this case, a generalized eigenvalue problem is solved with the zero-flux boundary conditions:

$$\sum_n \langle u_m \rho_{\text{eq}} | \boldsymbol{H}^0 | \rho_{eq} u_n \rangle c_n^0 = \lambda^0 \sum_n \langle u_m | p_{\text{eq}} | u_n \rangle c_n^0. \tag{3.35}$$

The Hamiltonian matrix $K_{nm}^0 = \langle u_m | \boldsymbol{H}^0 | u_n \rangle$ and the overlap matrix $S_{nm}^0 = \langle u_m | p_{\text{eq}} | u_n \rangle$ are then calculated with analytical differentiation and numerical integration to lead to the algebraic equation $K^0 c^0 = \lambda^0 S^0 c^0$ for solution.

The eigenbasis of $\boldsymbol{H}^0$ is then used to solve the eigenvalue problem involving the dark operator:

$$\left(\boldsymbol{H}^0 + \boldsymbol{H}^I\right) |\psi_i\rangle = \lambda_i |\psi_i\rangle. \tag{3.36}$$

Here, $\boldsymbol{H}^I = I(x)$,[2] and the new eigen-vector is then constructed as a linear combination of $|\psi_i^0\rangle$, $|\psi_i\rangle = \sum_j c_{ij} |\psi_j^0(x)\rangle$. The algebraic equation of this problem, $Kc = \lambda c$, is then solved. The matrix elements of K are:

$$K_{ij} = \langle \psi_i^0 | \boldsymbol{H}^0 + \boldsymbol{H}^I | \psi_j^0(x) \rangle = \lambda_i \delta_{ij} + \langle \psi_i^0 | \boldsymbol{H}^I | \psi_j^0 \rangle. \tag{3.37}$$

After obtaining the eigenbasis of $\boldsymbol{H}$, the photon arrival operators would adopt the matrix elements:

$$\langle \psi_i | \boldsymbol{y}_{a,d} | \psi_j \rangle = \int dx\ \psi_i(x) I_{a,d}(x) \psi_j(x). \tag{3.38}$$

[2]The summplemenary information details how the Jeffery's prior, or square root of fisher information for the smFRET experiment, can be added to $I(x)$ to account for the disparity in the information content from photons at different positions in the domain when the acceptor and donor intensities are not equivalent.

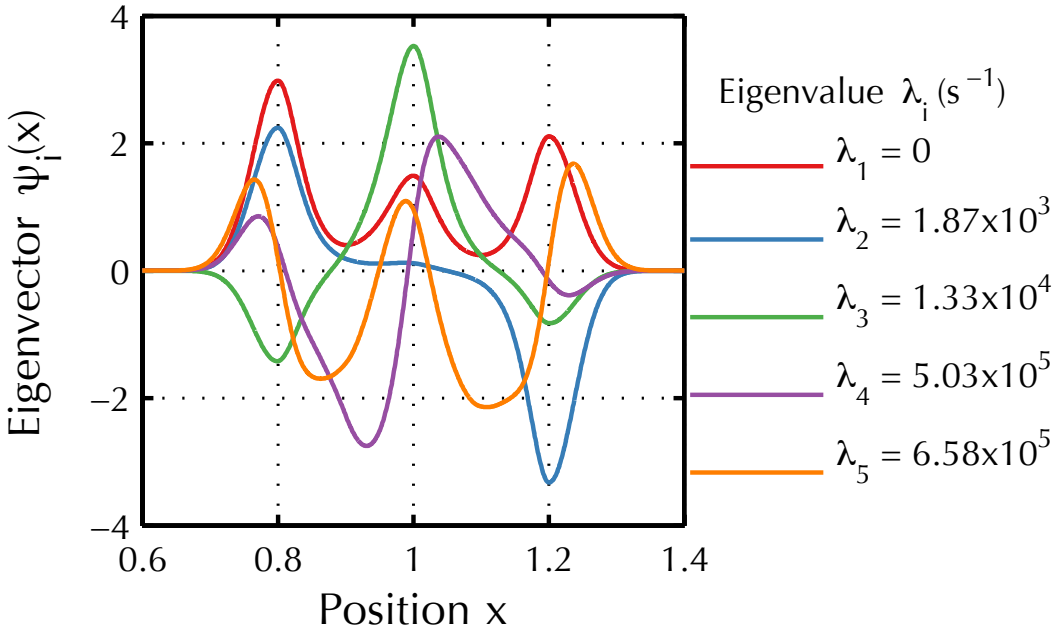

Figure 3.4: Eigenvectors and Eigenvalues for the test case potential $V(x)$ and $D = 500$. The first eigenvector is simply the equilibrium density $\rho_{\text{eq}}(x)$ with eigenvalue $\lambda_1 = 0$ for the stationary state. The second eigenvector shows the slowest process of system reaction between the state at $x = 0.8$ and $x = 1.2$; the second eigenvalue gives the overall system relaxation time $1/\lambda_2 = \tau \approx 0.5ms$. The third eigenvector is entrance and escape from the intermediate state of $x = 1.0$. Higher order modes show diffusion within each well.

With the elements developed thus far, the likelihood function of Eq. 3.22 can now be evaluated via a series of matrix algebra starting at either the α or β end. After each matrix-vector multiplication, the state vector α_t or β_t is normalized to prevent numerical underflow and these normalizations are collected according to Eq. 3.39 to record the log-likelihood as well as the inferred trajectory.[78]

$$\ell = \ln \mathcal{L} = \sum_{\tau=1} \ln \frac{\|\alpha_\tau\|}{\|\alpha_{\tau-1}\|} + \ln \frac{\langle \alpha_{N_P} | \beta_{N_P} \rangle}{\|\alpha_{N_P-1}\|}. \tag{3.39}$$

3.6 Inference of the Latent Trajectory from smFRET Data

Given the sequence of photon colors and arrival times in a specific round of smFRET measurement, the probability density of the latent variable at different times can be evaluated via Eq. 3.23. A simulated sequence of photon emission of a smFRET process using the PMF and diffusion coefficient outlined in Fig. 3.2 was used as the data set to perform the inference calculation. Other parameters of the photon trajectory data are listed in Table 3.1. The inferred result regarding the latent trajectory depends on the model parameters $F(x)$ and D of the Langevin dynamics used to calculate the likelihood function. Using a default trial profile for the equilibrium probability density, $p_0 = \cos^2(x/L)$, where L is the size of spatial domain of x, Fig. 3.3 plots the inferred trajectory as contour lines. In this inference, the diffusion coefficient of the actual dynamics model was used, i.e., perfect knowledge on D. The trial profile of $p_0 = \cos^2(x/L)$ gives the least informative dynamics without any prior knowledge of the time propagation of x.[79]

Comparing the contours inferred using $p_0(x)$ to the actual Langevin dynamics trajectory in the simulation of the smFRET output as shown in Fig. 3.2, it can be seen that the temporal instances of the transitions between metastable states can be captured rather accurately by the inference algorithm. This agreement would worsen if the knowledge on D deteriorates. However, the time-averaged distribution of x from the inferred trajectory probability density, $\overline{p}(x) \sim p_{\rm eq}(x)$, deviates significantly from the three-well PMF of the actual latent dynamics. It is because the trial PMF is far from that of the actual latent variable. As shown later, the agreement between the inferred and the actual trajectory is systematically improved by performing the maximization step of statistical learning, Eq 3.11. The Maximum Information Method (MIM) of constant uncertainty binning[60, 63] gives a similar profile of x histogram as that of the inferred result using p_0. Other details of the smFRET simulation and MIM analysis are reported in Supplementary Information.

The remaining sections of this paper detail the optimization step of the statistical learning algorithm after solving the inference problem. The optimal PMF upon the convergence of the expectation-maximization iteration is shown in Fig. 3.5 to illustrate the feasibility.

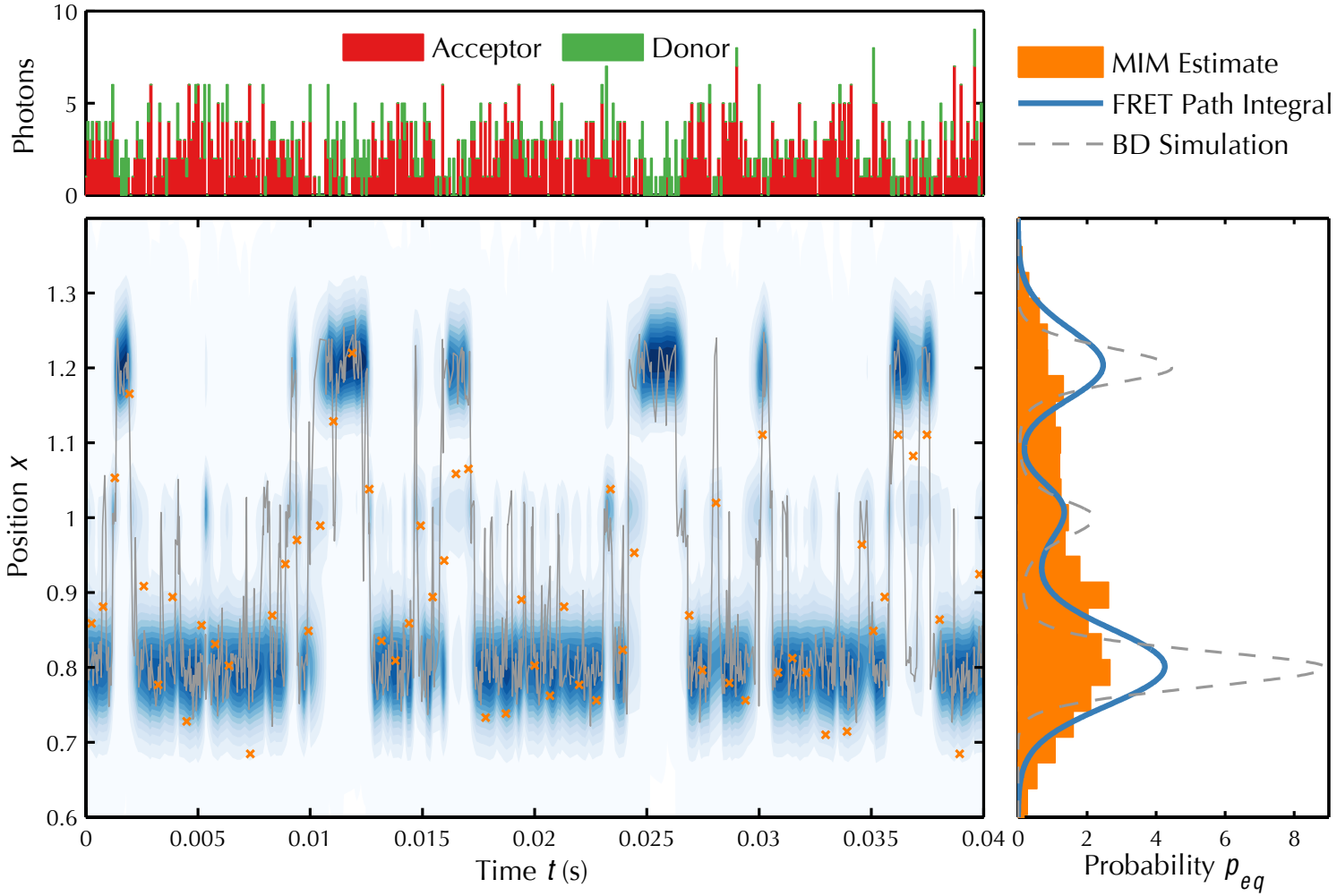

Figure 3.5: Comparison of Brownian Dynamics simulation and the resulting inferred trajectory from smFRET path integral with x2 resolution and converged parameters $F^*(x)$. (top-left) Trace of photon arrivals per millisecond for the donor and acceptor channel. (bottom-right) Time-averged position $p_{\text{eq}} = 1/t_{\text{rxn}} \int_0^{t_{\text{rxn}}} \delta(x - x')$. (bottom-left) Contours of the $\langle \alpha(t)|x \rangle$ and $\langle x|\beta(t) \rangle$ vectors in log space of color intensity. Lines are the raw Brownian dynamics simulation of the system on the free energy surface. Circles (o) are the estimates from time-binned maximum information method.

3.7 Expectation-Maximization Optimization of Langevin Dynamics from smFRET

Maximizing the log-likelihood for the optimal parameter set $\theta^* = \arg\max \ell[\theta]$ requires taking the derivative of the likelihood function in Eq. 3.22 and solves the Euler-Lagrange equation:

$$\frac{\delta \ell[\theta]}{\delta \theta(x)} = \frac{\delta}{\delta \theta} \ln \int \mathcal{D}X(t) \mathcal{P}(X(t), Y(t); \theta(x)) = 0. \tag{3.40}$$

This task appears nearly impossible because the functional dependence on the Langevin parameters is buried within the path integral. This operation can be recast by alternatively calculating the functional derivative of the expected logarithm of the joint probability for both the system position trajectory $X(t)$ and the photon trajectory $Y(t)$ and expected over the path probability

$$\frac{\delta}{\delta \theta(x)} \mathbb{E}_{X|Y} \left[\ln \mathcal{P}(X(t), Y(t); \theta(x)) \right] \tag{3.41}$$

This Expectation-Maximization algorithm[80] (EM) necessarily has two steps that are repeated until convergence:

1. Expect the trajectory of hidden state $X(t)$
2. Optimize the parameters θ to give the best likelihood for the expected trajectory

This derivative can be found more naturally from the immediate definition of the likelihood. What remans is to show that these operations are equivalent. First, the likelihood we seek to optimize can be split up into

$$\ln \mathcal{P}(Y|\theta) = \ln \mathcal{P}(Y, X|\theta) - \ln \frac{\mathcal{P}(Y, X|\theta)}{\mathcal{P}(Y|\theta)} \tag{3.42}$$

$$= \ln \mathcal{P}(Y, X|\theta) - \ln \mathcal{P}(X|Y, \theta) \tag{3.43}$$

where $Y = Y(t)$, $X = X(t)$ and $\theta = F(x), D$ to collapse notation. Then we can take the expectation over the hidden states to both sides $\mathbb{E}_{X|Y}[\cdot] = \int \mathcal{D}X(t) \mathcal{P}(X(t)|Y(t), \theta^k)$ for a particular parameter set at optimization step number k. The left hand size has no X dependence so acts like a constant through the expectation. The first term on the right hand side is the expected

log joint probability we proposed as a surrogate for optimization, and the second term on the right hand side is the "entropy" for the hidden states

$$\ln \mathcal{P}(Y|\theta) = Q(\theta;\theta^k) + S(\theta;\theta^k) \tag{3.44}$$

$$Q(\theta;\theta^k) = \int \mathcal{D}X\ \mathcal{P}(X|Y,\theta^k) \ln \mathcal{P}(Y,X|\theta) \tag{3.45}$$

$$S(\theta;\theta^k) = -\int \mathcal{D}X\ \mathcal{P}(X|Y,\theta^k) \ln \mathcal{P}(X|Y,\theta) \tag{3.46}$$

Because the relationship in Equation 3.44 holds for any value of θ including $\theta = \theta^k$ we subtract $\ln \mathcal{P}(Y|\theta^k) = Q(\theta^k;\theta^k) + S(\theta^k;\theta^k)$ from 3.44 to give the changes

$$\Delta\ell^k(\theta) = \Delta Q^k(\theta) + \Delta S^k(\theta) \tag{3.47}$$

where $\Delta Q^k(\theta) = Q(\theta;\theta^k) - Q(\theta^k;\theta^k)$. It can be shown from the Gibbs inequality that $\Delta S^k(\theta) \geq 0\ \forall\ \theta$ such that the change in log-likelihood must be greater than the change in the expected log-likelihood for the joint probability

$$\Delta\ell^k(\theta) \geq \Delta Q^k(\theta) \tag{3.48}$$

so we can achieve the updated parameters

$$\theta^{k+1} = \arg\max_{\theta} Q^k(\theta) \tag{3.49}$$

Translating these theoretical developments into a practical algorithm to calculate likelihoods, infer the trajectory, and optimize the parameters, recall that the likelihood is the abstract path integral from Equation 3.22 which after inserting a resolution of the identity between each operator becomes

$$\mathcal{L} = \sum_{\substack{\{i(\tau)\}\\\{j(\tau)\}}} \prod_{\tau}^{N_\mathrm{P}} \langle \phi_{i(\tau-1)}|\boldsymbol{y}_\tau|\phi_{j(\tau)}\rangle \langle \phi_{j(\tau)}|e^{-\boldsymbol{H}\Delta t_\tau}|\phi_{i(\tau+1)}\rangle \tag{3.50}$$

where the sum is taken over all indices of the eigenbasis at every photon arrival time τ. This means that the equivalent construction to the joint probability is just the product of all the bra-ket groupings $\mathcal{P}(Y,X|\theta) \leftarrow \prod_\tau \langle|\rangle$ as if the specific eigenstate at each time was known.[3] By translating Equation 3.42 to the eigenbasis, the log-likelihood $\mathcal{P}(Y|\theta) \leftarrow \ell = \ln \mathcal{L}$ can be re-write as

$$\ell = \ln \prod_\tau \langle|\rangle - \ln \frac{\prod_\tau \langle|\rangle}{\mathcal{L}} \tag{3.51}$$

[3] $\prod_\tau \langle|\rangle$ is shorthand for the summand in Equation 3.50

The expectation of states for a given parameter set θ^k now involves taking the following of the right hand side

$$\mathbb{E}_{X|Y}[\cdot] \leftarrow \sum_{\substack{\{i(\tau)\} \\ \{j(\tau)\}}} \frac{\prod_\tau \langle|\rangle^k}{\mathcal{L}^k} \tag{3.52}$$

where now the likelihood and bra-operator-ket terms are indexed by iteration step k. This leaves the equivalent operation to optimize the likelihood from the inference of eigenstates by inspecting the equivalent $Q^k(\theta)$ term

$$\theta^{k+1} = \arg\max_\theta \sum_{\substack{\{i(\tau)\} \\ \{j(\tau)\}}} \frac{\prod_\tau \langle|\rangle^k}{\mathcal{L}^k} \sum_{\tau'} \ln\langle|\rangle(\theta) \tag{3.53}$$

More concretely the functional dependence on the hamiltonian from the force and diffusion is parameters is realized in the exponential of the time propagation operator and the expectation gives back

$$Q^k(\theta) = \frac{1}{\mathcal{L}^k} \sum_\tau \langle \alpha_\tau^k | e^{-\boldsymbol{H}(\theta)\Delta t_\tau} | \beta_{\tau+1}^k \rangle \tag{3.54}$$

where the expected states

$$\langle \alpha_\tau^k| = \sum_j a_{j(\tau)}^k \langle \psi_j(x)| \tag{3.55}$$

$$|\beta_\tau^k\rangle = \sum_i b_{i(\tau)}^k |\psi_i(x)\rangle \tag{3.56}$$

are expanded in their coefficients of the eigenbasis inferred from the path integral.

$$a_{j(\tau')}^k = \sum_{\substack{\{i(\tau<\tau')\} \\ \{j(\tau<\tau')\}}} \prod_{\tau<\tau'} \langle|\rangle^k \langle \phi_{i(\tau-1)} | \boldsymbol{y}_\tau | \phi_{j(\tau)} \rangle \tag{3.57}$$

$$b_{i(\tau')}^k = \sum_{\substack{\{i(\tau>\tau')\} \\ \{j(\tau>\tau')\}}} \langle \phi_{i(\tau-1)} | \boldsymbol{y}_\tau | \phi_{j(\tau)} \rangle \prod_{\tau>\tau'} \langle|\rangle^k \tag{3.58}$$

The coefficients $\{a\}$ are derived purely from photon information before the current time step and thus convey forward information, the coefficients $\{b\}$ contain future information. This construction is known for finite Hidden Markov Models and the forward-backward Baum-Welch algorithm.[61]

Finally, it is shown in the supplementary information, how the necessary property analogous to the Gibbs inequality that $\Delta S^k(\theta) \geq 0 \; \forall \; \theta$ is preserved for the eigenbasis treatment which is fundamentally different from the typical analysis with probability distributions because the coefficients for each eigenvector at a particular time can be negative.

3.8 Pertubation theory for Hamiltonian and Time-Averaged Derivatives

The maximization step of the EM algorithm requires the maximization of terms of the form

$$Q^k(\theta) = \frac{1}{\mathcal{L}^k} \sum_{\tau} \langle \alpha_\tau | e^{-\boldsymbol{H}(\theta)\Delta t_\tau} | \beta_{\tau+1} \rangle \tag{3.59}$$

where the expected states α and β sandwich the exponential of the time propagator Hamiltonian $\boldsymbol{H}(\theta)$ which depends parametrically on the parameter set θ. This optimization of Q^k then requires a functional derivative with respect to the parameter θ which is subsequently equated to 0 to solve for the next optimal parameter set θ^*

$$0 = \frac{1}{\mathcal{L}^k} \sum_{\tau} \frac{\delta}{\delta\theta(x)} \langle \alpha_\tau | e^{-\boldsymbol{H}(\theta)\Delta t_\tau} | \beta_{\tau+1} \rangle \Big|_{\theta^*} \tag{3.60}$$

Unfortunately the functional dependence on θ is buried within the exponential of the operator inhibiting a direct extraction of the functional derivative. However, it is possible to develop an equality to a line integral of the derivative kernel[81]. (Details in Supplementary Information)

$$\begin{aligned} \frac{\delta}{\delta\theta(x)} \langle \alpha_\tau | e^{-\boldsymbol{H}(\theta)\Delta t_\tau} | \beta_{\tau+1} \rangle = \\ - \int_0^{\Delta t_\tau} dt' \langle \alpha_\tau | e^{-\boldsymbol{H}t'} \frac{\delta \boldsymbol{H}}{\delta\theta(x)} e^{-\boldsymbol{H}(\Delta t_\tau - t')} | \beta_{\tau+1} \rangle \end{aligned} \tag{3.61}$$

Essentially, the derivative of the operator exists within a moving window that integrates from one side of τ to the other $\tau + 1$ by adding onto one operators time and removing from the latter.

This integral can then be computed by inserting the eigenbasis $\mathbf{1} = \sum_i |\psi_i\rangle\langle\psi_i|$ around all of the operators and recalling that the diagonalization gives $\langle \psi_i | e^{-\boldsymbol{H}t} | \psi_j \rangle$

$= \delta_{i,j} e^{-\lambda_i t}$to give the derivative element

$$-\sum_{i,j} \int_0^{\Delta t_\tau} dt' \langle \alpha_\tau | \psi_i \rangle e^{-\lambda_i t'} \frac{\delta \langle \psi_i | \boldsymbol{H} | \psi_j \rangle}{\delta \theta(x)} e^{-\lambda_j (\Delta t_\tau - t')} \langle \psi_j | \beta_{\tau+1} \rangle \tag{3.62}$$

All of the time dependence is left to the two exponentials, so much of the complex functional derivative can be brought outside the integral

$$-\sum_{i,j} a_{i(\tau)} b_{j(\tau+1)} \frac{\delta}{\delta \theta(x)} \langle \psi_i | \boldsymbol{H} | \psi_j \rangle \underbrace{\int_0^{\Delta t_\tau} dt' e^{-\lambda_i t'} e^{-\lambda_j (\Delta t_\tau - t')}}_{\Gamma_{i,j}(\Delta t)} \tag{3.63}$$

The time integral can be computed directly to give

$$\Gamma_{i,j}^\tau = \begin{cases} \Delta t_\tau e^{-\lambda_i \Delta t_\tau} & i = j \\ \dfrac{e^{-\lambda_i \Delta t_\tau} - e^{-\lambda_j \Delta t_\tau}}{\lambda_j - \lambda_i} & i \neq j \end{cases} \tag{3.64}$$

Because the functional derivative is invariant to time in the path integral, we have derive the equivalent of a "transfer function" which can give the likelihood derivative as an expectation over the path integral

$$\frac{\delta Q^k}{\delta \theta(x)} = \frac{-1}{\mathcal{L}^k} \sum_{i,j} \frac{\delta \langle \psi_i | \boldsymbol{H} | \psi_j \rangle}{\delta \theta(x)} \mathbb{E}_{X|Y}^k [a_i b_j] \tag{3.65}$$

where the combined time-average quantity inferred at iteration k is

$$\mathbb{E}_{X|Y}^k [a_i b_j] = \sum_\tau \Gamma_{i,j}^\tau a_{i(\tau)}^k b_{j(\tau+1)}^k \tag{3.66}$$

3.9 Functional derivaitve of Hamiltonian for $F(x)$ and D

This functional derivative considers a perturbation to the function parameters $\epsilon \Theta'(x)$ which results in a perturbation to the Hamiltonian of dynamics $\epsilon \mathbf{H}'$. The functional derivative is defined a distribution such that for all test functions $f(x)$,

$$\left\langle \frac{\delta \ell[\Theta(x)]}{\delta \Theta(x)}, f(x) \right\rangle = \int \frac{\delta \ell[\theta(x)]}{\delta \theta(x')} f(x') dx' \tag{3.67}$$

$$= \lim_{\varepsilon \to 0} \frac{\ell[\theta(x) + \varepsilon f(x)] - \ell[\theta(x)]}{\varepsilon} \tag{3.68}$$

$$= \frac{d}{d\epsilon} \ell[\theta + \epsilon f] \Big|_{\epsilon=0}. \tag{3.69}$$

The steps to finding this distribution of interest requires two operations.

1. Carry out the operation of equation (3.69): take the derivative of the log posterior evaluated with the functional perturbation
2. Seperate out the pertubative function from the inner product in equation (3.67) to identify the remaining integrand as the functional derivative.

The Euler-Lagrange equation is best used to get derivative of path integral element with respect to the governing parameterizing functions

$$\frac{\delta \langle \psi_i | \boldsymbol{H} | \psi_j \rangle}{\delta \theta(x)} = \frac{\partial (\psi_i(x) \boldsymbol{H} \psi_j(x))}{\partial \theta(x)} - \frac{d}{dx} \frac{\partial (\psi_i(x) \boldsymbol{H} \psi_j(x))}{\partial \theta'(x)} \tag{3.70}$$

which plugging in the form of the hamiltonian gives the functional derivative with respect to force profile

$$\frac{\delta \langle \psi_i | \boldsymbol{H} | \psi_j \rangle}{\delta F(x)} = F(x) \frac{D}{2} \psi_i(x) \psi_j(x) - \frac{D}{2} \left(\psi_i(x) \psi_j(x) \right)' \tag{3.71}$$

Alternatively, the derivative with respect to diffusion constant, is a simple scalar derivative and because the Hamiltonian and corresponding eigenvalues scale linearly with the diffusion constant, the result is simply

$$\frac{d \langle \psi_i | \boldsymbol{H} | \psi_j \rangle}{dD} = \frac{\lambda_i}{D} \delta_{i,j} \tag{3.72}$$

3.10 Optimization Algorithm

After calculating the derivative with respect to force, the optimum is found by setting the expected derivative from equation 3.65 to 0

$$\mathbb{E}_{X|Y} \left[F^*(x) \frac{D}{2} \psi_i(x) \psi_j(x) - \frac{D}{2} \left(\psi_i(x) \psi_j(x) \right)' \right] = 0 \tag{3.73}$$

and solving for the optimal $F^*(x)$.

$$F^*(x) = \frac{d/dx(\mathbb{E}_{X|Y}[\delta(x-X)])}{\mathbb{E}_{X|Y}[\delta(x-X)]} \tag{3.74}$$

Substituting for the definition of equilibrium probability, can arrive at the update step where

$$p^*(x) = \mathbb{E}_{X|Y}\left[\delta(x-X)\right] = p^k_{\text{eq}}(x)\sum_{i,j}\mathbb{E}^k_{X|Y}[a_i b_j]\phi_i(x)\phi_j(x) \tag{3.75}$$

The optimization for diffusion coefficient is found by a simple line search[82] to optimize the likelihood after the new optimal $F^*(x)$ has been applied. Since the diffusion is just a single scalar number, this procedure is trivially fast.

Watkins and Yang showed how to use Bayesian information criteria to determine number of states from change point analysis [83]. Because the optimization is naturally underdetermined,there are thousands of photons per experiment but the parameters we are solving for are functions that have an uncountable number of points which dine fine them. Therefore, we add in a Bayesian prior $\mathcal{P}(F(x), D)$ which breaks the degeneracy in the parameter set and highlights the desired outputs from the algorithm [84].

$$\mathcal{L}(F(x), D) = \mathcal{P}(Y(t)|F(x), D)\frac{\mathcal{P}(F(x), D)}{\mathcal{P}(Y(t))} \tag{3.76}$$

$$\mathcal{P}(\theta(x)) = \exp\left(-\eta_F D\langle F^2(x)\rangle_{\text{eq}}\right) \tag{3.77}$$

where the ensemble average is approximated by the path expectation

$$-\eta_F D\langle F^2(x)\rangle_{\text{eq}} \simeq -\eta_F \mathbb{E}_{X|Y}\left[DF^2(x)\right] \tag{3.78}$$

such that it may naturally fit into the EM framework and give the modified update equation

$$p^*(x) = \left(\mathbb{E}_{X|Y}\left[p(x)\right]\right)^{1/(1+\eta_F)} \tag{3.79}$$

The calculation of averages $\mathbb{E}_{X|Y}[\delta(x-X)]$ and derivatives $d/dx(\mathbb{E}_{X|Y}[\delta(x-X)])$, can be quickly computed by matrix multiplication if the matrix Ψ is constructed with the eigenvectors oriented as columns in the matrix and the inferred state matrix $\mathbb{E}^k_{X|Y}[a_i b_j]$ as

$$p^*(x) = \text{tr}(\Psi\mathbb{E}^k_{X|Y}[a_i b_j]\Psi^\dagger) \tag{3.80}$$

Algorithm 1 smFRET Expectation-Maximization Learning algorithm for Force and Diffusion

procedure EM(Tad)
Initialize $\rho_{eq} \propto \cos(x)$
Diagonalize $\Psi, \Lambda \leftarrow$sFEM(F(x),D)
Photon matrix $Y_{a,d} = \Psi \boldsymbol{y}_{a,d} \Psi^{\dagger}$
Multiply matrices $Ye^{-\Lambda t}Y \ldots e^{-\Lambda t}Y$
Normalize coefficients $\ell = \sum_{\tau} |\alpha_{\tau}|$
Infer states $\langle \alpha(t)|, |\beta(t)\rangle$
Collect statistic $\mathbb{E}^{k}_{X|Y}[a_i b_j] = \sum_{\tau} \Gamma^{\tau}_{i,j}$
Run EM $\rho^{k+1}_{\text{eq}} \leftarrow p^{k}_{\text{eq}}(x) \sum_{i,j} \mathbb{E}^{k}_{X|Y}[a_i b_j] \phi_i(x) \phi_j(x)$
Line search $D^{k+1} = \arg\max \ell[\rho^{k+1}_{eq}, D]$
end procedure

3.11 Results of EM Estimation Test System

For 12 runs of synthetically produced experimental smFRET trajectories, we applied the EM algorithm for 10,000 steps of optimization beginning with an arbitrary initial guess of $F(x) = 0$ to arrive at the following results pictured in the Figures 5-10. The potential of mean force and the equilibrium probability are plotted for each trajectory along with the ensemble averaged result over experimental trials $\langle p(x) \rangle = \sum_n p^*_n(x)$ and $\langle V(x) \rangle = -\ln \langle p(x) \rangle$. These results are compared to the profile of the reference system used to generate the smFRET trajectories.

Although the low resolution of the experiment does allow for a few outlier solutions, the locations of meta-stable states, the number of these states, the relative population fractions of each well, and the barriers between wells are all accurately reproduced.

The converged results for the diffusion constant seen in Figure 3.8 shows systematic bias towards lower values caused by the Bayesian maximum entropy prior. Although the diffusion constants will change with the magnitude of the optimization temperature η_F, for a wide range of values from $10^{-4} \rightarrow 10^{-7}$, the results were not appreciably different. A rough bootstrap error estimate can be offered by comparing the average of the trajectories to the variance of the distribution in converged profiles.

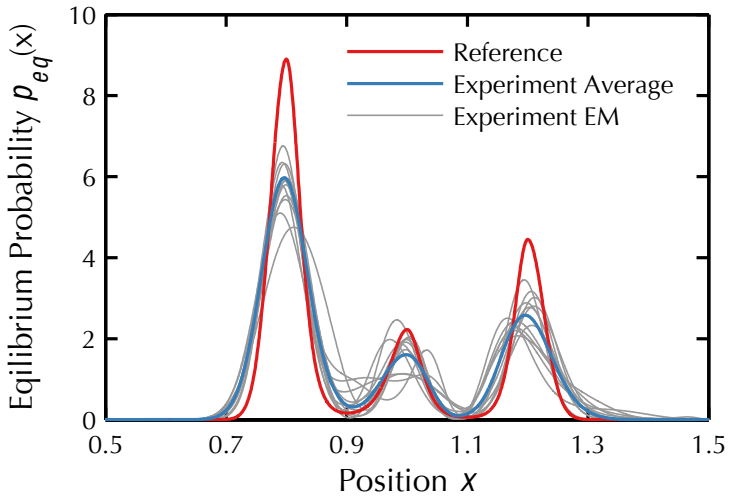

Figure 3.6: Converged $p^*_{\text{eq}}(x)$ from the EM algorithm on 12 test case data sets of 40,000 photons with $\eta_F = 1 \times 10^{-6}$ with 2x Resolution.

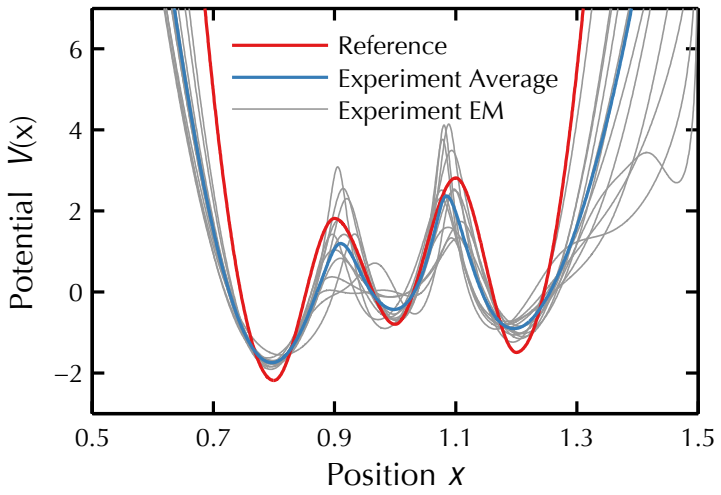

Figure 3.7: Converged $V^*(x)$ from the EM algorithm on 12 test case data sets of 40,000 photons with $\eta_F = 1 \times 10^{-6}$ with 2x Resolution.

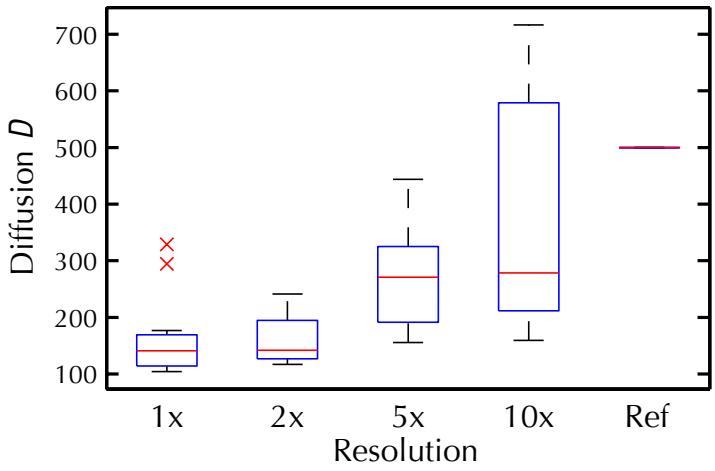

Figure 3.8: Converged D^* from the EM algorithm run on 12 test case data sets of 40,000 photons with $\eta_F = 1 \times 10^{-6}$ at all resolutions 1-10x.

3.12 Mean First Passage Time and Reaction Rates

The kinetic rates between states are determined by running a Mean First Passage Time calculation as a post-processing step on the optimized free energy surfaces at the optimum diffusion constant. From [37], mean first passage time from state (position) x_A to x_B is the nested cumulative integral

$$\tau_{\text{rxn}}(x_A \to x_B) = k^{-1}_{A\to B} = \frac{1}{D}\int_{x_A}^{x_B} dx\, e^{V(x)} \int_{x_L}^{x} dx'\, e^{-V(x')} \tag{3.81}$$

for the reverse direction the formula reads

$$\tau_{\text{rxn}}(x_B \to x_A) = k^{-1}_{B\to A} = \frac{1}{D}\int_{x_B}^{x_A} dx\, e^{V(x)} \int_{x_R}^{x} dx'\, e^{-V(x')} \tag{3.82}$$

which can be implemented using the previously computed integrals by flipping the bounds of integration and recasting the cumulative integral $\int_x^R = \int_L^R - \int_L^x$ to give

$$k^{-1}_{B\to A} = \frac{1}{D}\int_{x_A}^{x_B} dx\, e^{V(x)} \left(\int_{x_L}^{x_R} dx'\, e^{-V(x')} - \int_{x_L}^{x} dx'\, e^{-V(x')} \right) \tag{3.83}$$

The first passage time calculation on the test case gives relaxation rates between state "A" centered around $x = 0.8$ and state "B" centered around $x = 1.2$ shown in Table 3.2. These are then compared to the kinetic rates calculated off

of the converged potential of mean force for 12 different trajectories at various levels of resolution above that of the baseline FRET experimental emission rates. Box-plot summaries of the results are shown in Figure 3.9.

Table 3.2: Relaxation rates and reaction rates for model potential with $D = 500$.

$\tau_{A\to B}$	1.9592×10^{-3} s
$\tau_{B\to A}$	0.9394×10^{-3} s
$k_{A\to B}$	0.5104×10^{3} s^{-1}
$k_{B\to A}$	1.0645×10^{3} s^{-1}
$k_{A\to B}/k_{A\to B}$	0.4795

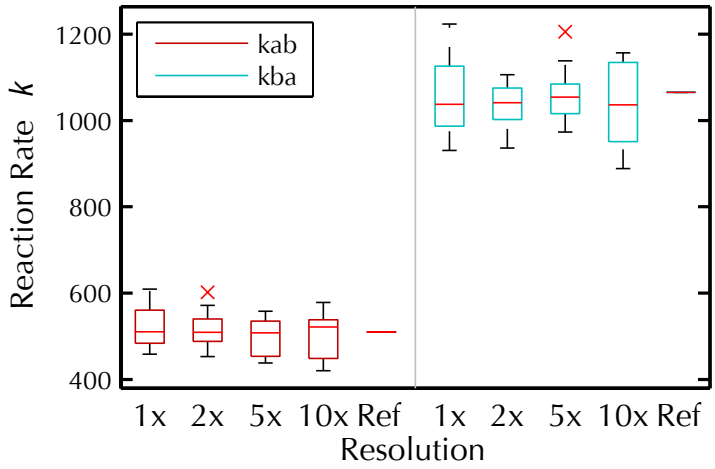

Figure 3.9: Reation rates $k_{a\to b}$ and $k_{b\to a}$ from Mean First Passage Time of the results on $V(x), D$ from the EM algorithm on test case data of 40,000 photons with $\eta_F = 1 \times 10^{-6}$ at all resolutions 1-10x.

3.13 Results Disscussion

The expectation-maximization has many features of self-consistent field theory optimization, in that convergence is attained when the derivative in the expected field of observed states is zero. The convergence is unfortunately sub-linear[85] meaning that over 10,000 iterations are required to converge

the optimized force profiles and diffusion constants when starting from an arbitrary guess potential of zero force. However, the fundamental property that likelihood is an every increasing quantity in the EM algorithm with iteration, means that these optimization are stable, robust and reliable.

When inspecting the test case results, the diffusion constant is systematically low because of the suppression caused by the entropy regularization term $\langle F^2(x) \rangle$. This biases the results towards profiles with lower barrier heights, and to preserve the correct reaction rates between states the diffusion constant must be lower. This Bayesian approach reduces the variance between possible solutions (which are infinite) but does induce bias into the final solution.[86] However, the reaction rate are within margin of error and without bias because D itself is determined by an unbiased maximum likelihood search for a given free energy surface. In effect, this D optimization seeks to find the correct reaction rates from the smFRET because $D \propto k$.

Although there can be a fair amount of variance at low resolutions for the diffusion constant, the reaction rates are accurately reproduced at even the lowest levels of resolution seemingly without bias. The precision in the reaction rate estimates actually doesn't increase with resolution because at at set number of photons N_P, the total trajectory time is reduced and there are less reactive events to make an estimate of this Poisson process ($\sigma_k^2 \sim 1/t_\mathrm{exp}$).

Most importantly, the resulting converged potentials $V(x)$ and equilibrium probabilities $p_\mathrm{eq}(x)$, show some variance between different experimental trajectories, but the sum of all runs shows very good agreement with the actual reference potential. The algorithm manifests a result with the correct number of states in the potential and the location of these states is quite accurate. The highest magnitude error is in the actual barrier height of between these states which couples to the diffusion constant to give alternate versions of the same rate constant (high diffusion and high barrier or low diffusion and low barrier).

3.14 Conclusion

Practically speaking, we have adopted the framework of the Hidden Markov Model, which traditionally exists for system with desecrate numbers of states[72, 87, 88] with a transition rate matrix and generalized it to the realm of continuum dynamics in space and time which is governed by the Fokker-Planck PDE. This translation is accomplished by making the eigenvectors the new desecrate states of our system with the eigenvalues serving as the rate constants. Alternatively, one can think of this development as extending the Kalman filter which is used for statistical inference on system with state space models,[3, 89]

to the realm of arbitrary and non-linear driving potentials while still allowing for exact calculation of both the inference and optimization problem in the limit of infinite basis set.

This general framework can be extended into multiple dimensions as experimental apparatus and the smFRET phlorophore dyes developed to allow simultaneous measurement of two or more distances [90, 91]. All of the difficulty is buried in the diagonalization of the system Hamiltonian to construct the eigenbasis. After that procedure, the entire problem maps to a familiar set of matrix-matrix multiplications.

Moreover the presentation in this article can be readily extended to other classes of experiment or simulation study. The only smFRET specific feature of the derivation was the exact nature of the information operator $\boldsymbol{y}$. For single-molecule pulling, our information is instead the approximate force and position of the bead at different time points.[92] For replica-exchange molecular dynamics, we would have many short bursts of trajectory information to feed into this dynamic model.[93] Due to the underlying similarities in the path integral development with many Von Neumann experimental systems, there is possibly great applicability to quantum systems and the discovery of the underlying potential energy of a system beyond the limitations of the traditional RKR inversion.[94]

Measuring the governing dynamics and driving forces for complex systems is now possible with the smFRET experiment by applying this statistical learning approach. The development is completely general to any system which undergoes over-damped Langevin dynamics. Although the experiment is still imprecise in establishing the exact profiles, signature features of the dynamics such as reaction rates and the number and location of any intermediate states have been reliably reproduced. Ultimately this serves to aid ones understanding of complex systems and the mechanisms and deriving forces which govern its dynamics. From engineering an increase in enzymatic activity or increasing the robustness of a protein machine[95], the interplay between forces and dynamics can now be elucidated through the smFRET experimental approach and this statistical learning algorithm.

CHAPTER 4

Fisher Information Metric for the Langevin Equation and Least Informative Models of Continuous Stochastic Dynamics

4.1 Abstract

The evaluation of the Fisher Information matrix for the probability density of trajectories generated by the over-damped Langevin dynamics at equilibrium is presented. The framework we developed is general and applicable to using an arbitrary potential of mean force as the parameter set. Leveraging an innovative Hermitian form of the corresponding Fokker-Plank equation allows for an eigenbasis decomposition of the time propagation probability density. This formulation motivates the use of the square root of the equilibrium probability density as the basis for evaluating the Fisher information of trajectories with the essential advantage that the Fisher information matrix in the specified parameter space is constant. This outcome greatly eases the calculation of information content in the parameter space via a line integral. In the continuum limit, a simple analytical form can be derived to explicitly reveal the physical origin of the information content in equilibrium trajectories. This methodology also allows deduction of least informative dynamics models from known or available observables that are either dynamical or static in nature. The minimum information optimization of dynamics is performed for a set of different constraints to illustrate the generality of the proposed methodology.

4.2 Introduction

Complex molecular systems are often studied by tracking the temporal evolution of important coordinates to reveal the hidden metastable states and to characterize the transitions between them. A central objective of many experimental and theoretical endeavors is thus to resolve the system dynamics from the measured data.[96–98] In this regard, the information content for the parameters of interest in a particular measurement is of prime interest and is the central focus of the present work. For biomolecular conformational changes, the over-damped Langevin equation is often employed as the model for a mechanistic understanding.[12, 99, 100] On top of the deterministic potential of mean force (PMF), the Langevin model incorporates the solvent-induced stochastic fluctuations via the diffusion coefficient[10] to describe system dynamics along a set of chosen coordinates or order parameters.[101–103] Unless specified otherwise, this work addresses the quantification of information content for Langevin dynamics.

A major difficulty of understanding biomolecular dynamics is that the commonly used characterization methods are often limited to processes with distinct temporal and spatial scales. For example, the computer simulation of molecular dynamics can be used to record the coordinates and velocities of the atoms of biomolecules as well as the surrounding solvent molecules but has limited ability to access structural transitions on longer time scales ($> \mu$s).[104–106] Different techniques of nuclear magnetic resonance (NMR) can be used to acquire the transition rates of the different aspects of biomolecule conformational changes on different time scales (usually with the necessary assumption of a two-state model)[107–109] but the ensemble averaging nature washes away the rich mechanistic details in molecular individualities. Single-molecule methods such as those via the Förster resonance energy transfer (FRET)[13, 32, 110] do away with the issues of ensemble averaging but face the challenge of the low signal-to-noise ratio convoluted with photon-counting statistics.[111, 112]

As a result, not only the data analysis in each category of measurements is complicated, but also the systematic combination of information across different techniques is a challenging issue. We reason that the foundation for the quantitatively integrating the knowledge from different data types could be based on an information measure for the dynamics parameters from the time trajectories. In this regard, the Fisher information provides a framework with a clear statistical picture and straightforward linkage to thermodynamics;[113–115] therefore, it is employed here to quantify the information underlying the Langevin dynamics model. With this information metric, the determination of the parameters of time propagation can serve as a common objective over

different data types for cross validation and knowledge integration between fields.[116]

Here, we develop numerical and analytical methods to determine the Fisher information of the PMF and diffusion coefficient in the trajectories of Langevin dynamics. This capability not only concerns single-molecule methods but also the other ways of characterizing biomolecule conformational changes mentioned above. Instead of devising various statistics to learn about the time-dependent properties of the system,[117, 118] Fisher information of trajectories (FIT) allows systematically deducing the optimal dynamics model under the constraints of observables.

Since FIT quantifies the disorder inscribed in the probability function of the data, or the likelihood function of model parameters, it can be applied to deduce the properties of the equation of motion such as the PMF and diffusion coefficient in Langevin dynamics under the constraints given by the selected statistics of observables.[42, 119, 120] If entropy was used as the metric, this approach results in the maximum entropy principle that is widely used to determine the static distribution of system states. For retrieving parameters relating to time propagation, the constrained optimization approach has been addressed for Markov states systems.[88, 121] For continuous stochastic dynamics, however, acquiring an explicit functional form of the information content in trajectories faces the challenges of infinite dimensionality, non-differentiability with respect to time of the Winer process, and path integral. This work provides numerical and analytical solutions for determining the FIT of Langevin dynamics at equilibrium.

The rest of the paper is organized as the following. The application of Fisher information to continuous but not differentiable trajectories at equilibrium is established in Section 4.3 with the selection of basis discussed in Section 4.4. Section 4.5 outlines a numerical procedure we developed to calculate FIT for continuous stochastic dynamics and Section 4.6 derives the analytical form of FIT in the continuum limit. Section 4.7 applies the analytic form of FIT for measuring information content in trajectories and the result of which is used in section 4.8 to derive the least informative dynamics under various constraints of based on the Langevin equation. The similarity of our particular approaches in evaluating FIT with the Quantum von Neumann entropy and matrix trace of power spectrum is then discussed in section 4.9 followed by our conclusion.

4.3 The Fisher Information Matrix for Langevin Dynamics

The Fisher information defines a size measure (Riemannian manifold) for the volume element of information content for a corresponding set of model parameters. The line integral of a parameter change with the Fisher information matrix is formally the dissipation function for moving in the space of parameters.[114, 122, 123] This metric translates between the parameter space of system dynamics and the information content of the resulting probability distribution of system trajectories. Therefore, Fisher information can be used to assess the manner by which changing the properties of time propagation may vary the information content in system dynamics.[124]

Using a general coordinate x to describe the dynamics of a system, the concern of our analysis is the information content for the mean force profile $F(x)$ and diffusion coefficient $D(x)$ contained in the Langevin trajectories $X(t)$ with $\dot{x} = \beta D(x)F(x) + \sqrt{2D(x)}dW_t$; t is time and β is one over the Boltzmann constant k_B multiplying the system temperature T. The Wiener process specifying the stochastic force in this equation of motion satisfies $\langle dW_t dW_{t'} \rangle = \delta(t - t')dt$. The profile of the deterministic mean force is related to the PMF, $V(x)$, as $F(x) = -dV(x)/dx$. The equilibrium distribution of system states in the continuous space of x is related to the PMF as $p_\mathrm{eq}(x) \propto \exp(-V(x)/k_\mathrm{B}T)$. A trajectory, $X(t);\ t \in [0, t_\mathrm{obs}]$, in this case is a continuous but non-differentiable function of time. In a measurement, this stochastic trajectory is generally realized at specific instances separated by a time resolution Δt to create a vector $\vec{X}_t = [X(0), X(\Delta t), X(2\Delta t), \ldots, X(t_\mathrm{obs})]$. This vector exists in a trajectory space of dimensionality $N = T/\Delta t$ with the coordinates denoted as the set $\{x_\tau | \tau = 0, 1, 2, \ldots, N\}$. With this setup, we aim to find the Fisher information of the deterministic and stochastic components in the Langevin equation in a multidimensional vector space. The the $\lim_{\Delta t \to 0}$ will be performed on the final results to recover the complete information content in trajectories in the continuum limit.

The collection of function parameters of the Langevin equation is now combined into $\vec{\theta} = \{\theta_i\}$ for the convenience of derivation. Here, i may go to infinity for describing the parameters associated with the dense set of points for a continuous functions. The Fisher information metric is defined as the expectation value for the product of the derivatives of the log probability density

of the trajectory with respect to $\vec{\theta}$:

$$\mathcal{I}_{i,j}(\vec{\theta}) = \mathbb{E}_{\vec{X}_t}\left[\frac{\partial \ln P(\vec{X}_t)}{\partial \theta_i}\frac{\partial \ln P(\vec{X}_t)}{\partial \theta_j}\middle|\vec{\theta}\right]. \tag{4.1}$$

The $\mathbb{E}_{\vec{X}_t}[\cdot]$ in the above equation denotes the expectation evaluated by path integration over $\vec{X}_t$, $\int D\vec{X}_t P(\vec{X}_t)[\cdot]$. The Fisher information is thus a matrix for the (i,j) pairs of parameters evaluated at the current values of $\vec{\theta}$.

In calculating this path integral, the Markovian nature of the Langevin equation can lead to tremendous simplification. In particular, the probability density of $\vec{X}_t$ can be factored via the probability densities of time propagation that connect two consecutive time slices,

$$P(\vec{X}_t) = p(x_0)\prod_{\tau=0}^{N-1} p(x_{\tau+1}|x_\tau). \tag{4.2}$$

In this equation, $p(x_0)$ is the static distribution of system states at time zero. For equilibrium trajectories, $p(x_0) \to p_{\text{eq}}(x_0)$ is employed for specifying the probability densities of the initial states. Therefore, each component in the Fisher information matrix becomes:

$$\mathcal{I}_{ij} = \sum_{\tau,\tau'=-1}^{t_{\text{obs}}/\Delta t}\int D\vec{X}_t\left[\frac{\partial \ln p(x_{\tau'+1}|x_{\tau'})}{\partial \theta_i}\frac{\partial \ln p(x_{\tau+1}|x_\tau)}{\partial \theta_j}\right]P(\vec{X}_t). \tag{4.3}$$

The contribution from $p_{\text{eq}}(x_0)$ is included by setting $p(x_0|x_{-1}) = p_{\text{eq}}(x_0)$. For the path integral in Eq. (4.3), the time indices that do not appear in the derivatives are marginalized out so that $P(\vec{X}_t) \to p(\tau, \tau+1, \tau', \tau'+1)$. Furthermore, unless $\tau = \tau'$, the other terms in the double sum of Eq. (4.3) contribute zero to $\mathcal{I}_{ij}$ due to the ability to isolate a normalization condition of the conditional probability densities $p(x_{\tau+1}|x_\tau), \tau = 0 \ldots N$:

$$\int dx \frac{\partial \ln p(x)}{\partial \theta}p(x) = \frac{\partial \int dx p(x)}{\partial \theta} = 0. \tag{4.4}$$

The only contributing terms to $\mathcal{I}_{ij}$ thus come from the Fisher information matrix of the equilibrium distribution, I_{eq}, and that of the conditional probability of time propagation, $I_{\Delta t}$:

$$I_{\text{eq}} = \int dx_0 \frac{\partial \ln p_{\text{eq}}(x_0)}{\partial \vec{\theta}}\frac{\partial \ln p_{\text{eq}}(x_0)}{\partial \vec{\theta}}p_{\text{eq}}(x_0) \tag{4.5}$$

$$I_{\Delta t} = \int dx_t dx_0 \frac{\partial \ln p(x_t|x_0)}{\partial \vec{\theta}}\frac{\partial \ln p(x_t|x_0)}{\partial \vec{\theta}}p(x_t, x_0). \tag{4.6}$$

Here, the notation of coordinates was simplified by implying that $t = \Delta t$ and noting that there are $t_{\text{obs}}/\Delta t$ equivalent terms of $I_{\Delta t}$. As a result, the Fisher information of $P(\vec{X}_t)$ is:

$$\mathcal{I} = I_{\text{eq}} + \frac{t_{\text{obs}}}{\Delta t} I_{\Delta t}. \tag{4.7}$$

The calculation of FIT via Eq. (4.7) thus comes down to evaluating the derivatives and integrals defined for the static distribution in Eq. (4.5) and the dynamic propagator in Eq. (4.6). An essential key toward achieving this goal is the selection of the parameter set. The strategy we followed is trying to eliminate the dependence of the matrix elements on other parameters because a constant Fisher information matrix is convenient for evaluating the information content via a line integral.[114, 122, 123] It turns out that the most natural basis for the deterministic components of the Langevin equation is the square root of the equilibrium probability density[125]:

$$\rho_{\text{eq}}(x) = \sqrt{p_{\text{eq}}(x)} \propto \sqrt{\exp\left(\int^x F(x)/k_{\text{B}}T\right)}. \tag{4.8}$$

The Fisher information is now defined with respect to a composite quantity of the functions for density and diffusion $\vec{\theta} \to \{\rho_{eq}(x), D(x)\}$. Because of this treatment, the derivatives are now functional derivatives and the Fisher information is a scalar field over the x, y arguments of the parameterizing functions rather than a matrix.

One starts from the following term for the equilibrium term:

$$I_{\text{eq}}(x, y) = \int dx_0 \frac{\delta \ln p_{\text{eq}}(x_0)}{\delta \rho_{\text{eq}}(x)} \frac{\delta \ln p_{\text{eq}}(x_0)}{\delta \rho_{\text{eq}}(y)} p_{\text{eq}}(x_0). \tag{4.9}$$

Using $p_{\text{eq}}(x) = \rho^2_{\text{eq}}(x)$, the functional derivatives in Eq. (4.9) can be easily calculated:

$$\frac{\delta \ln \rho^2_{\text{eq}}(x_0)}{\delta \rho_{\text{eq}}(x)} = 2\frac{\delta(x_0 - x)}{\rho_{\text{eq}}(x_0)}. \tag{4.10}$$

Therefore, the equilibrium Fisher information is just the integral of delta functions:

$$I_{\text{eq}} = 4 \int dx_0 \delta(x_0 - x)\delta(x_0 - y) = 4\delta(y - x). \tag{4.11}$$

Using $\rho_{\text{eq}}(x)$ to define the parameter space of the Fisher information, I_{eq} is thus constant in the sense that it is invariant to the values of the parameter set; that is, I_{eq} is not a functional of ρ.

As will be shown later, the property of $\rho_{\text{eq}}(x)$ in making I_{eq} constant also facilitates the evaluation of $I_{\Delta t}$ defined in Eq. (4.6). The remaining task of calculating FIT with respect to the deterministic components of the Langevin equation is then evaluating the Fisher information of the conditional probability density with respect to $\rho_{\text{eq}}(x)$:

$$I_{\Delta t}(x,y) = \int dx_t dx_0 \frac{\delta \ln p(x_t|x_0)}{\delta \rho_{\text{eq}}(x)} \frac{\delta \ln p(x_t|x_0)}{\delta \rho_{\text{eq}}(y)} p(x_t, x_0). \tag{4.12}$$

This Fisher information matrix is a rank 2 tensor field over the space coordinate and is also a functional of $\rho_{\text{eq}}(x)$.

For evaluating the functional derivatives of $p(x_t|x_0)$, we rely on the Fokker-Planck equation (FPE) that governs the temporal evolution of $p(x_t|x_0)$:

$$\frac{\partial p(x_t|x_0)}{\partial t} = -\nabla \cdot J(x_t), \tag{4.13}$$

and

$$J(x_t) = -\nabla \cdot \left(D(x_t) \nabla p(x_t|x_0) - \frac{D(x_t)(F(x_t)}{k_{\text{B}}T} p(x_t|x_0) \right). \tag{4.14}$$

In this formulation, x can in general be a multidimensional vector and the gradients in the FPE apply to the x_t coordinate. The initial condition of this partial differential equation is $p(x_t|x_0)|_{t=0} = \delta(x_t - x_0)$ and the no-flux boundary conditions $J(x_t = L) = J(x_t = -L) = 0$ are employed to conserve the total probability.

The FPE can be equivalently expressed in terms of $p_{\text{eq}}(x)$:

$$\frac{\partial p(x_t|x_0)}{\partial t} = \nabla \cdot \left(D(x_t) p_{\text{eq}}(x_t) \nabla \frac{p(x_t|x_0)}{p_{\text{eq}}(x_t)} \right). \tag{4.15}$$

Next, the unique features of expressing the FPE via the equilibrium density $\rho_{\text{eq}}(x)$ are discussed in preparation for the evaluation of $I_{\Delta t}(x,y)$.

4.4 The Hermitian Operator in the FPE Corresponding to Langevin Dynamics

With $\rho_{\text{eq}}(x)$ being the square root of $p_{\text{eq}}(x)$, the probability density of time propagation can symmetrized in time to become:

$$\rho(x_t, x_0) = p(x_t|x_0)\sqrt{\frac{p_{\text{eq}}(x_0)}{p_{\text{eq}}(x_t)}}, \tag{4.16}$$

such that the probability density of a trajectory also has the temporal symmetry:

$$P(\vec{X}_t) = \rho_{\text{eq}}(x_0)\prod_{\tau=0}^{N-1}\rho(x_{\tau+1}, x_\tau)\rho_{\text{eq}}(x_N). \tag{4.17}$$

Re-expressing Eq. (4.12) via the symmetric time propagator[122] gives additional factors for $I_{\Delta t}(x, y)$:

$$I_{\Delta t} = \overbrace{\int dx_t dx_0 \frac{\delta \ln \rho(x_t, x_0)}{\delta \rho_{\text{eq}}(x)} \frac{\delta \ln \rho(x_t, x_0)}{\delta \rho_{\text{eq}}(y)} p(x_t, x_0)}^{I^{\rho}_{\Delta t}} + 2\delta(x-y) - 2\rho(x, y). \tag{4.18}$$

Details of this derivation can be found in Appendix C.2.

In the continuum limit of $\Delta t \to 0^+$, the limit of $\rho(x_t, x_0) \to \delta(x_t - x_0)$ leads to cancellation of the last two terms in Eq. (4.18). Therefore, we will focus on the first term, an integration in a two-dimensional space, $I^{\rho}_{\Delta t}$. After expanding the functional logarithms therein, we will evaluate FIT according to:

$$I^{\rho}_{\Delta t}(x, y) = \int dx_t dx_0 \frac{\delta \rho(x_t, x_0)}{\delta \rho_{\text{eq}}(x)} \frac{\delta \rho(x_t, x_0)}{\delta \rho_{\text{eq}}(y)} \frac{\rho_{\text{eq}}(x_t)\rho_{\text{eq}}(x_0)}{\rho(x_t, x_0)}. \tag{4.19}$$

The FPE of the symmetric propagator $\rho(x_t, x_0)$ can be found by substituting the expression in Eq. (4.16) into Eq. (4.15):

$$\frac{\partial \rho(x_t, x_0)}{\partial t} = -\boldsymbol{H}\rho(x_t, x_0) \tag{4.20}$$

$$\boldsymbol{H} = -\frac{1}{\rho_{\text{eq}}(x)}\nabla \cdot \left(D(x)\rho^2_{\text{eq}}(x)\nabla \frac{1}{\rho_{\text{eq}}(x)} \right) \tag{4.21}$$

The boldface font is used to denote operators in this work.

With the proof shown in Appendix C.3, it can be seen that the $\boldsymbol{H}$ operator in the FPE of Eq. (4.20) is Hermitian. Therefore, $\rho(x_{\Delta t}, x_0)$ can be expressed via the matrix elements of $\boldsymbol{H}$ with the Dirac notation as:

$$\rho(x_{\Delta t}, x_0) = \langle x_{\Delta t} | e^{-\boldsymbol{H}\Delta t} | x_0 \rangle. \tag{4.22}$$

In the following section, a procedure for calculating $I^{\rho}_{\Delta t}$ based on the Hermitian version of the FPE of the Langevin equation is developed.

4.5 Numerical Calculations of FIT via an Eigenbasis Expansion

The Hermitian nature of $\boldsymbol{H}$ means that there exist real eigenvalues λ_i for which the eigenvectors ψ_i are orthogonal, $\langle \psi_i | \boldsymbol{H} | \psi_j \rangle = \lambda_i \delta_{ij}$. Based on this property, we performed an eigen-decomposition of the symmetric operator of the FPE (sFPE) via a spectral finite element method.[77] In particular, we assume a form for the eigenvectors $\psi_i(x) = \rho_{\mathrm{eq}}(x)\phi_i(x)$ with the modifying function $\phi_i(x)$ determined from the input of ρ_{eq} and diffusion coefficient by using spectral elements to solve the generalized eigenvalue problem $\langle \phi_i \rho_{\mathrm{eq}} | \boldsymbol{H} | \rho_{\mathrm{eq}} \phi_j \rangle = \lambda_i \delta_{ij} \langle \phi_i \rho_{\mathrm{eq}} | \rho_{\mathrm{eq}} \phi_j \rangle$.[1] By inserting the completeness property of $\sum_i |\psi_i\rangle\langle\psi_i| = \mathbf{1}$ that resolves the identity operator, and noting the orthogonality of the eigenbasis, $\rho(x_{\Delta t}, x_0)$ can now be decomposed as:

$$\rho(x_{\Delta t}, x_0) = \sum_{i,j} \langle x_{\Delta t} | \psi_i \rangle \langle \psi_i | e^{-\boldsymbol{H}\Delta t} | \psi_j \rangle \langle \psi_j | x_0 \rangle \tag{4.23}$$

$$= \sum_i \psi_i(x_{\Delta t}) e^{-\lambda_i \Delta t} \psi_i(x_0). \tag{4.24}$$

To illustrate the resulting values of $\rho(x_{\Delta t}, x_0)$ for a model system containing three wells separated by $5\,k_{\mathrm{B}}T$ barriers in the PMF (Fig. 4.1) the two-dimensional density field $\rho(x_{\Delta t}, x_0)$ calculated from the eigen-decomposition method discussed above is shown as log-contours in Fig. 4.2 at an intermediate time resolution of $\Delta t = 0.025$ s which is also on the same order of the relaxation time (Δt 0.036 s) out of the middle intermediate state.

For calculating the FIT according to Eq. (4.19), the functional derivative of the density field with respect to ρ_{eq} needs to be determined. By employing an

[1] The $\phi(x)$ modifying functions are in fact the eigenfunctions of the backward Kolmogrov equation

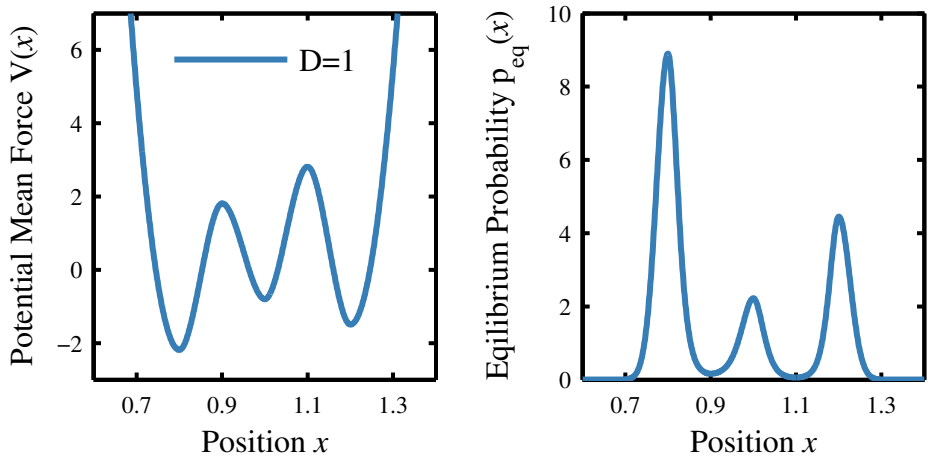

Figure 4.1: The reference potential of mean force and the corresponding equilibrium probability density for illustrating the calculations of the Fisher information of trajectories. The model has 3 local minima with approximately $5\,k_{\mathrm{B}}T$ barriers separating these metastable states

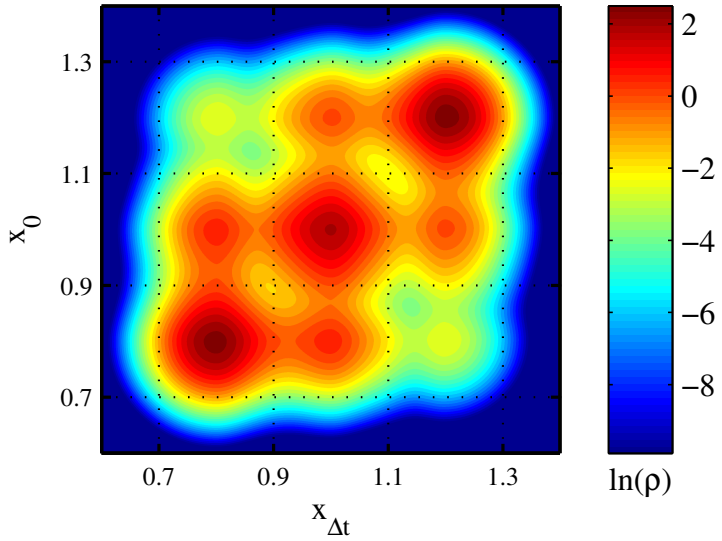

Figure 4.2: The contours of $\ln \rho(x_{\Delta t}, x_0)$ at $\Delta t = 0.025$ s for the reference PMF shown in Fig. 4.1 at $D = 1$ in the corresponding unit. The high-density regions along the diagonal axis around the three local minima are clear. However, off-diagonal densities representing transitions between states within the time window are also clear.

arbitrary test function $f(x)$, the functional derivative is defined as:

$$\int dx \frac{\delta \rho(x_{\Delta t}, x_0)}{\delta \rho_{\text{eq}}(x)} f(x) = \frac{d}{d\epsilon} \langle x_t | e^{-\Delta t \boldsymbol{H}[\rho_{\text{eq}} + \epsilon f]} | x_0 \rangle \Big|_{\epsilon=0}. \tag{4.25}$$

Since the reference Hamiltonian does not necessarily commute with the operator perturbed by $f(x)$, i.e., $[\boldsymbol{H}^0, \boldsymbol{H}'] \neq 0$, the exponential in Eq. (4.25) is not necessarily factorizable for taking the derivative. To overcome this difficulty, we employ the approximate Trotter splitting formalism:[126]

$$e^{-\Delta t(\boldsymbol{H}^0 + \epsilon \boldsymbol{H}')} = e^{-\Delta t \boldsymbol{H}^0/2} e^{-\epsilon \Delta t \boldsymbol{H}'} e^{-\Delta t \boldsymbol{H}^0/2} + \mathcal{O}(\Delta t^3), \tag{4.26}$$

wherein the errors due to approximation decrease as Δt goes towards the continuum limit. The effect of the perturbation is thus evaluated effectively at the time midpoint between x_t and x_0. Performing the derivative with respect to ϵ as prescribed by Eq. (4.25) to the approximate form in Eq. (4.26) leads to:

$$\begin{aligned} \int dx \frac{\delta \rho(x_{\Delta t}, x_0)}{\delta \theta(x)} f(x) &\approx -\Delta t \langle x_{\Delta t} | e^{-\Delta t \boldsymbol{H}^0/2} \boldsymbol{H}' e^{-\Delta t \boldsymbol{H}^0/2} | x_0 \rangle \\ &= -\Delta t \sum_{i,j} \psi_i(x_{\Delta t}) e^{-\lambda_i \Delta t/2} \langle \psi_i | \boldsymbol{H}' | \psi_j \rangle e^{-\lambda_j \Delta t/2} \psi_j(x_0). \end{aligned} \tag{4.27}$$

The completeness $\sum_i |\psi_i\rangle\langle\psi_i| = \mathbf{1}$ is inserted between all operators and orthogonality $\langle \psi_k | \boldsymbol{H}^0 | \psi_i \rangle = \lambda_i \delta_{ik}$ of the eigenbasis is applied to lead to the second line of Eq. (4.27).

The Hamiltonian of the first order perturbation with respect to ρ_{eq}, $\boldsymbol{H}'$, in Eq. (4.27) can be found by tracking ϵ in the total Hamiltonian:

$$\begin{aligned} \boldsymbol{H}[\rho_{\text{eq}} + \epsilon f] = \frac{-1}{(\rho_{\text{eq}} + \epsilon f)(x)} \nabla \cdot \\ \left(D(x)(\rho_{\text{eq}} + \epsilon f)^2(x) \nabla \frac{1}{(\rho_{\text{eq}} + \epsilon f)(x)} \right). \end{aligned} \tag{4.28}$$

The coefficient of the first order Taylor expansion of $\boldsymbol{H}$ can then be evaluated

to obtain $\boldsymbol{H}'$:

$$\boldsymbol{H}' = \frac{d}{d\epsilon}\boldsymbol{H}[\rho_{\text{eq}} + \epsilon f]\Big|_{\epsilon=0} \tag{4.29}$$

$$\boldsymbol{H}' = -\frac{2}{\rho_{\text{eq}}(x)}\nabla\cdot\left(D(x)f(x)\rho_{\text{eq}}(x)\nabla\frac{1}{\rho_{\text{eq}}(x)}\right) \tag{4.30}$$

$$+\frac{f(x)}{\rho_{\text{eq}}^2(x)}\nabla\cdot\left(D(x)\rho_{\text{eq}}^2(x)\nabla\frac{1}{\rho_{\text{eq}}(x)}\right) \tag{4.31}$$

$$+\frac{1}{\rho_{\text{eq}}(x)}\nabla\cdot\left(D(x)\rho_{\text{eq}}^2(x)\nabla\frac{f(x)}{\rho_{\text{eq}}^2(x)}\right). \tag{4.32}$$

With the knowledge of $\boldsymbol{H}'$, the error in Eq. (4.27) is purely due to the Trotter splitting approximation.

The only remaining factors for calculating FIT are the $\langle\psi_i|\boldsymbol{H}'|\psi_j\rangle$ terms in the second line of Eq. (4.27). For Eqs. (4.30) and (4.32), integration by parts can be performed to isolate the test function $f(x)$ in the integral. Utilizing the relation of $\phi_i(x) = \psi_i(x)/\rho_{\text{eq}}(x)$ and recognizing the existence of terms of the form $-\phi_i(x)\boldsymbol{H}\psi_j(x)$, one obtains:

$$\langle\psi_i|\boldsymbol{H}'|\psi_j\rangle = \int dx f(x)\big[2\nabla\phi_j(x)\cdot(D(x)\rho_{\text{eq}}(x)\nabla\phi_i(x)) - \phi_j(x)\lambda_i\psi_i(x) - \phi_i(x)\lambda_j\psi_j(x)\big]. \tag{4.33}$$

Therefore,

$$\frac{\delta\langle\psi_i|\boldsymbol{H}|\psi_j\rangle}{\delta\rho_{\text{eq}}(x)} = 2\nabla\phi_i(x)\cdot(D(x)\rho_{\text{eq}}(x)\nabla\phi_j(x)) - \rho_{\text{eq}}(x)(\lambda_i+\lambda_j)\phi_i(x)\phi_j(x). \tag{4.34}$$

Along the same token, one can find the functional derivative of $\boldsymbol{H}$ with respect to the diffusion coefficient:

$$\frac{\delta\langle\psi_i|\boldsymbol{H}|\psi_j\rangle}{\delta D(x)} = \nabla\phi_i(x)\cdot\rho_{\text{eq}}^2(x)\nabla\phi_j(x). \tag{4.35}$$

In the case that the diffusion coefficient is not a function of x, i.e., a constant, the functional derivative reduces to:

$$\frac{d\langle\psi_j|\boldsymbol{H}|\psi_i\rangle}{dD} = \frac{\lambda_i}{D}\delta_{ij}. \tag{4.36}$$

Combining the information in Eqs. (4.27) and (4.34), the functional derivative of $\rho(x_t, x_0)$ with respect to $\rho_{\text{eq}}(x)$ can be calculated as:

$$\frac{\delta\rho(x_t,x_0)}{\delta\rho_{\text{eq}}(x)} \approx -\Delta t \sum_{i,j} \psi_i(x_t) e^{-\lambda_i \Delta t/2} \frac{\delta\langle\psi_i|\boldsymbol{H}|\psi_j\rangle}{\delta\rho_{\text{eq}}(x)} e^{-\lambda_j \Delta t/2} \psi_j(x_0). \tag{4.37}$$

Armed with the knowledge of the functional derivatives from the eigenbasis construction, the key terms of Fisher information in Eq. (4.19) can be expressed:

$$\frac{I^{\rho}_{\Delta t}(x,y)}{(\Delta t)^2} = \sum_{i,j}\sum_{k,l} \frac{\delta\langle\psi_i|\boldsymbol{H}|\psi_j\rangle}{\delta\rho_{\text{eq}}(x)} \frac{\delta\langle\psi_k|\boldsymbol{H}|\psi_l\rangle}{\delta\rho_{\text{eq}}(y)} \Omega^{ik}_{jl}(\Delta t). \tag{4.38}$$

For all of the functional dependence on (x_t, x_0) that are integrated away in Eq. (4.38) is captured in the "overlap" integral over two consecutive time slices:

$$\begin{aligned}\Omega^{ik}_{jl}(\Delta t) \equiv \int dx_{\Delta t} dx_0 \; & e^{-(\lambda_i+\lambda_k)\Delta t/2} \psi_i(x_{\Delta t}) \psi_k(x_{\Delta t}) \\ & \left(\sum_n \phi_n(x_{\Delta t}) e^{-\lambda_n \Delta t} \phi_n(x_0)\right)^{-1} e^{-(\lambda_j+\lambda_l)\Delta t/2} \psi_j(x_0)\psi_l(x_0).\end{aligned} \tag{4.39}$$

According to Eqs. (4.38) and (4.39), Algorithm 2 below is used to calculate the Fisher information matrix of trajectories for the Langevin equation parametrized by the equilibrium density ρ_{eq} and constant diffusion coefficient D.

Algorithm 2 FIT calculation using the eigenbasis resolved via applying a spectral finite element method (sFEM) to solve the Hermitian FPE of Eq. 4.20.

procedure FISHER($\rho_{\text{eq}}(x), D; \Delta t$)
  Obtain λ_i and ϕ_i via sFEM
  Perform the 2D integral in $\Omega^{ik}_{jl}(\Delta t)$
  Calculate functional derivatives
  Go through the 4-loop summation for $I^{\rho}_{\Delta t}$
end procedure

For the reference model shown in Fig. 4.1, Algorithm 2 is used to calculate the Fisher information metric of Langevin trajectories at various time resolutions. Fig. 4.3 shows a contour plot of $I^{\rho}_{\Delta t}(x,y)/\Delta t$ as well as the cross derivatives with respect to the spatial coordinates of ρ_{eq} and the diffusion coefficient.

The matrix element of D alone is also shown as the actual numerical value in the figure.

At a low time resolution of $\Delta t \approx 0.15$ s that is on the order of the slowest timescale of system relaxation, the major component affecting the FIT is the low probability transition state in ρ_{eq} caused by the barrier located at $x = 1.1$ in the PMF. Therefore, reducing this barrier would lead to a faster system relaxation and hence a less prescriptive or informative dynamics model.

At an intermediate time resolution of $\Delta t \approx 0.02$ s, the dependence of FIT on the existence of the low density states at $x = 0.9$ and 1.1 is still significant, but off-diagonal negative couplings emerge. This pattern indicates that a flatter potential barrier would also result in a less informative time propagator.

At a high time resolution of $\Delta t \approx 0.002$ s, the importance and details of the underlying equilibrium probability of states seems to vanish, leading to a nearly tri-banded matrix with positives values along the diagonal terms and negative numbers for the off–diagonal elements. At this timescale, FIT begins to capture the behaviors of the diffusion processes within individual wells and the roles of ρ_{eq} are not as prominent in comparison. As we extend the analysis of FIT into the continuum limit of $\Delta t \to 0^+$ analytically, the origin of the tri-banded nature of the Fisher information matrix becomes explicit.

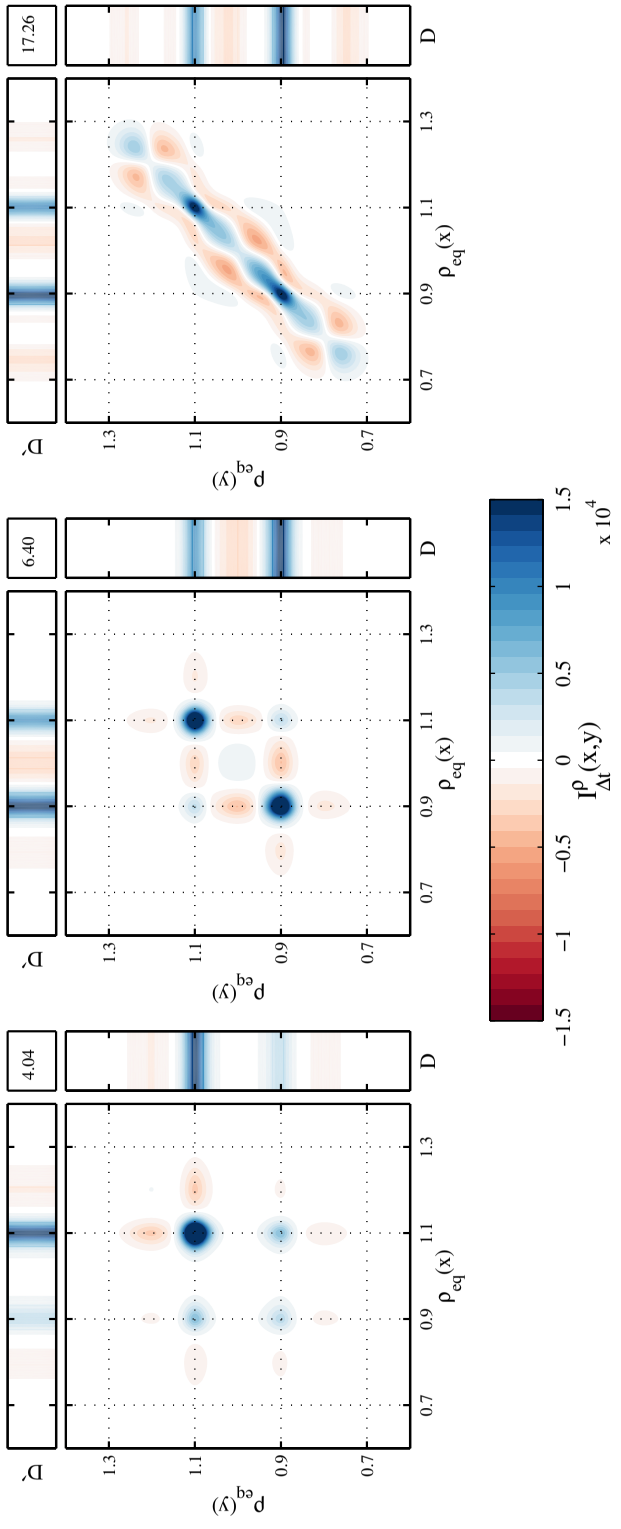

Figure 4.3: The contour of the Fisher information matrix of $I^{\rho}_{\Delta t}(x,y)/\Delta t$ for $\Delta t = 0.15$, 0.02, and 0.002 s. The upper band in contour is the Fisher information metric with respect to D and ρ, $\mathcal{I}(D,y)$. Corner value is the Fisher element for diffusion $\mathcal{I}(D,D')$.

4.6 Fisher Information of Trajectories in the Continuum Limit

The Fisher information of $p(x_{\Delta t}|x_0)$ can also be written as:[125]

$$I_{\Delta t} = \int dx_{\Delta t} dx_0 \frac{\delta^2 \ln p(x_{\Delta t}|x_0)}{\delta\rho_{\text{eq}}(x)\rho_{\text{eq}}(y)} p(x_{\Delta t}, x_0). \tag{4.40}$$

In the continuum limit of $\Delta t \to 0^+$, $p(x_{\Delta t}|x_0)$ looses its dependence on the parameters because it approaches to the delta function $p(x_{\Delta t}|x_0) = \delta(x_{\Delta t} - x_0)$. Therefore, $I_{\Delta t}$ goes to 0 in the continuum limit. Regarding the FIT for the entire trajectory, $\mathcal{I}$, though, the denominator of Eq. (4.7) also goes to zero as $\Delta t \to 0^+$. Applying the L'Hopital rule, utilizing the $\Delta t = t$ simplification again, and letting $t = 0$ leads to:

$$\mathcal{I} - I_{\text{eq}} = t_{\text{obs}} \int dx_t dx_0 \frac{\delta^2(\partial p(x_t|x_0)/\partial t)}{\delta\rho_{\text{eq}}(x)\rho_{\text{eq}}(y)} p_{\text{eq}}(x_0)\Big|_{t=0}. \tag{4.41}$$

The fact that $p(x_t|x_0)$ has no dependence on $\rho_{\text{eq}}(x)$ in the continuum is also employed in arriving at this equation.

Applying $p(x_t|x_0)|_{t=0} = \delta(x_t - x_0)$ into the FPE, one obtains:

$$\frac{\partial p(x_t|x_0)}{\partial t}\Big|_{t=0} = \nabla \cdot \left(D(x_t)\rho_{\text{eq}}^2(x_t)\nabla \frac{\delta(x_t - x_0)}{\rho_{\text{eq}}^2(x_t)} \right). \tag{4.42}$$

Using the result of Eq. (4.42) in Eq. (4.41), one obtains:

$$\mathcal{I} - I_{\text{eq}} = t_{\text{obs}} \int dx_t dx_0 \frac{\delta^2 \rho_{\text{eq}}^2(x_0)}{\delta\rho_{\text{eq}}(x)\rho_{\text{eq}}(y)} \nabla \cdot (D(x_t)\nabla\delta(x_t - x_0)). \tag{4.43}$$

With Eq. (4.43), the functional derivative therein becomes straightforward:

$$\mathcal{I} - I_{\text{eq}} = 2Dt_{\text{obs}} \int dx_t dx_0 \delta(x_t - y)\nabla \cdot (D(x_t)\nabla\delta(x_t - x_0))\delta(x_0 - x). \tag{4.44}$$

Finally, performing the integration gives the answer of FIT in the continuum limit:

$$\mathcal{I}(x, y)[\rho_{\text{eq}}] = 4\delta(y - x) + 2t_{\text{obs}}\nabla \cdot (D(x)\nabla\delta(x - y)). \tag{4.45}$$

This Laplacian kernel[127] is indeed seen in the numerical studies shown in Fig. 4.3 as the time resolution increases (Δt decreases).

4.7 Fisher information as a measure of change in information

If a $\xi \in [0,1]$ variable is employed to specify the state of model parameters and interpolate between two different sets of parameters, the Fisher information metric can be used to quantify the information change in varying the parameter set:[114, 128]

$$\mathcal{J} = \int_0^1 d\xi \sum_{i,j} \frac{\partial \theta_i}{\partial \xi} \mathcal{I}_{i,j}(\vec{\theta}) \frac{\partial \theta_j}{\partial \xi}. \tag{4.46}$$

The arc length along the ξ curve in the information space can also be calculated via the Fisher information:

$$\mathcal{S} = \int_0^1 d\xi \sqrt{\sum_{i,j} \frac{\partial \theta_i}{\partial \xi} \mathcal{I}_{i,j}(\vec{\theta}) \frac{\partial \theta_j}{\partial \xi}}. \tag{4.47}$$

The triangle inequality states that $\mathcal{J} \geq \mathcal{S}^2$; here, the equality is true if the integrand is constant along the curve. For FIT via the basis of ρ_{eq}, constant Fisher information matrix is satisfied in the continuum limit, Eq. (4.45).

To perform the line integral of FIT, we define a linear path interpolating a flat reference model with $\rho_{\text{eq}} = 0$ and the desired profile that is only denoted as $\rho^*_{\text{eq}}(x)$ for now. Similarly, a linear path can also be defined for the diffusion coefficient coordinate that goes from a reference value D_{ref} to the optimized D^*. Therefore,

$$\rho_{\text{eq}}(\xi) = \xi \rho^*_{\text{eq}}(x) \tag{4.48}$$

$$D(\xi) = D_{\text{ref}}(x) + (D^*(x) - D_{\text{ref}}(x))\xi, \tag{4.49}$$

where $\xi \in [0,1]$.

From the fisher information matrix in the continuum limit given by Eq. (4.45) where $\sum_{i,j} \to \int dxdy$, performing the line integral over $d\xi$ with integration by parts gives the analytical form of the information dissipation measure:

$$\mathcal{J} = \mathcal{S}^2 = \int dx \, \nabla \rho_{\text{eq}}(x) \cdot (D(x) \nabla \rho_{\text{eq}}(x)). \tag{4.50}$$

Expressing $\rho_{\text{eq}}(x)$ in terms of the mean force $F(x)$ in the case of a constant diffusion coefficient, Eq. (4.50) becomes:

$$\mathcal{J}[F(x)] = \frac{D}{4} \left\langle F^2(x) \right\rangle_{\text{eq}}. \tag{4.51}$$

The same information expression can also be obtained by using an entropy measure of the probability densities of trajectories and taking the continuum limit.[79]

4.8 The Equilibrium Distributions of Least Informative Dynamics

The analytical form of information can be used to deduce the least informative dynamics models (LID) with various constraints.[4, 125, 129] In fact, to the best of our knowledge, the LID models for continuous stochastic dynamics have never been resolved. In this section, we use several examples to illustrate how the analytical results derived in this work, e.g., Eq. (4.50) can be used for deducing such models.

Finding the least informative distribution under a set of constraints can be performed via minimizing a free energy functional or the Lagrangian:

$$\mathcal{L}(\rho(x), \vec{\omega}, \eta) = -\eta \underbrace{\langle\rho|\nabla^2|\rho\rangle}_{\text{Entropy}} + (\underbrace{\langle\rho|\boldsymbol{E}|\rho\rangle}_{\text{Energy}} - E) \tag{4.52}$$

$$+ \sum_i \omega_i \underbrace{(\langle\rho|\boldsymbol{C}_i|\rho\rangle - C_i)}_{\text{Constraint}}. \tag{4.53}$$

$\boldsymbol{E}$ is a log-likelihood energy operator for dynamic observables, η is the optimization temperature which includes the factor of $t_{\text{obs}}D$, ω_i are the Lagrange multipliers of constraints, $\boldsymbol{C}_i$ are the constraint operators, and E, C_i are the desired values for the energy and constraints, respectively. In the case of $\boldsymbol{C}_1 = \mathbf{1}$, i.e., the operator of summing all probabilities, $C_1 = 1$ is the targeted value to ensure proper normalization of the equilibrium distribution.

Minimization of the Lagrangian is accomplished by setting the functional derivative $\delta\mathcal{L}/\delta\rho$ to zero to reach a solution $\rho^*(x; \eta, \omega)$ that depends parametrically on the Lagrange multipliers:

$$\mathcal{L}^*(\vec{\omega}, \eta) = \inf_{\rho_{\text{eq}}(x)} \mathcal{L}(\rho(x), \vec{\omega}, \eta). \tag{4.54}$$

The solution requires the Lagrange multipliers to satisfy the Karush–Kuhn–Tucker (KKT) conditions of optimality.[82] Alternatively, the Lagrange multipliers can also be determined by optimizing the two-dimensional Lagrangian:

$$\omega^* = \arg\max_{\omega} \mathcal{L}^*(\vec{\omega}, \eta) \tag{4.55}$$

$$\eta^* = \arg\max_{\eta} \mathcal{L}^*(\vec{\omega}, \eta). \tag{4.56}$$

Typically, the desired value of "energy" is not known *a priori* and the "temperature" η is set instead. This is similar in flavor to the derivation of the maximum entropy Boltzmann distribution for the canonical ensemble.[130]

If all of the constraints can be expressed in the form of linear operators as in Eq. (4.52), the numerical techniques of solving partial differential equations can be utilized to determine the optimal profiles of LID. Applications of using the analytical form of FIT to derive the least informative parameters for Gaussian processes with a constant force and diffusion coefficient as well as the Ornstein-Uhlenbeck process are presented in Appendix C.4 and C.5, respectively. Several examples of using the procedure outlined above to deduce continuous PMF profiles is presented next to highlight the generality of the developed method. The diffusion coefficient is treated as a constant in these calculations. The least informative dynamics models also illustrate how the criterion of reducing trajectory information differs from the static objective of maximizing the entropy of the state distribution.

LID constraining on the domain bounds

If the only observation regarding the dynamics of a system is that x is bounded in between $[-L, L]$, the least informative dynamics model with constant diffusion is found by setting the functional derivative of the Lagrangian in Eq. (4.52) with respect to ρ to zero which gives:

$$\frac{\delta \mathcal{L}}{\delta \rho(x)} = -2\eta \nabla^2 \rho(x) + \omega_1 \rho = 0. \tag{4.57}$$

The normalization constraint is achieved by setting $\boldsymbol{C}_1 = \mathbf{1}$. With restriction of the boundary conditions $\rho(x)|_{-L,L} = 0$, the solution of this Strum-Louiville problem is:

$$\rho^*(x) = c_1 \cos(\omega' x) \tag{4.58}$$

$$\omega' = \frac{n\pi}{2L} \quad ; \quad n \in \mathbb{N}, \tag{4.59}$$

with n as natural numbers and c_1 set for normalization.

After substituting $\rho^*(x)$ into the definition for $\mathcal{L}$, the choice of ω' can be achieved by performing the maximization:

$$\omega^* = \arg\max_{\omega} \mathcal{L}^*(\vec{\omega}, \eta) = \arg\max_{\omega} -c_1 \left(\frac{n\pi}{2L}\right)^2. \tag{4.60}$$

This maximization simply selects the smallest possible n, and as a result,

$$p^*_{\text{eq}} = \frac{1}{L}\cos^2\left(\frac{x\pi}{2L}\right) \tag{4.61}$$

$$F^*(x) = \frac{\pi}{L}\tan\left(\frac{x\pi}{2L}\right). \tag{4.62}$$

Fig. 4.4 plots the least informative state distribution in comparison with the maximum entropy profile from the static entropy measure $S_{\text{eq}} = \int p_{\text{eq}}(x)\ln p_{\text{eq}}(x)$ which is a flat distribution in the domain and a force that diverges at the boundaries. Because the Fisher information of trajectories includes a measure of the deterministic force in dynamics, a diverging value at the boundaries is not desirable. Alternatively, a distribution $\cos^2(x)$ that smoothly decays to a zero gradient at the domain edges is selected instead.

LID constraining on the mean and variance of p_{eq}

The least informative dynamics model based the observed mean μ and variance σ^2 of x can be found by setting the two constraints $\boldsymbol{C}_1 = x\mathbf{1}$ and $\boldsymbol{C}_2 = x^2\mathbf{1}$. This optimization can be performed in the Fourier space as well and that approach is adopted here to illustrate the capability offered by having an analytical form of FIT. Because the Lagrangian is essentially a set of inner products, the Parseval's theorem[131] says that the Fourier transforms (FT) of functions $f(x)$ and $g(x)$, $\tilde{f}(k)$ and $\tilde{g}(k)$, respectively, satisfy:

$$\int dx\ f(x)g^\dagger(x) = \int dk\ \tilde{f}(k)\tilde{g}^\dagger(k). \tag{4.63}$$

The † sign indicates complex conjugation. Because the FT of $\nabla\rho(x)$ is $\mathrm{i}k\tilde{\rho}(k)$, where i is the imaginary number, the FT of the information measure in Eq. (4.50) reads:

$$\mathcal{J} = -D\int dk\ k^2\tilde{\rho}^2(k). \tag{4.64}$$

This functional form is isomorphic to the square curvature potential when using the Green's function analysis to determine the power spectrum of a density field, where the magnitude of the Fourier components decays with the wave number k as $\tilde{\rho}(k) \propto 1/(1+k^2)$.[132]

For the constraint on mean, $g(x) = x\rho(x)$, $g(k) = \mathrm{i}d\tilde{\rho}(k)/dk$ and variance $h(x) = x^2\rho(x)$, $\tilde{h}(k) = -d^2\tilde{\rho}(k)/dk^2$, the FT version of the Lagrangian is thus:

$$\mathcal{L}(\tilde{\rho}(k)) = \int dk\ \tilde{\rho}(k)\left(-\eta Dk^2\tilde{\rho}(k) - \omega_1\mathrm{i}\frac{d\tilde{\rho}}{dk}(k) - \omega_2\frac{d^2\tilde{\rho}}{dk^2}(k)\right). \tag{4.65}$$

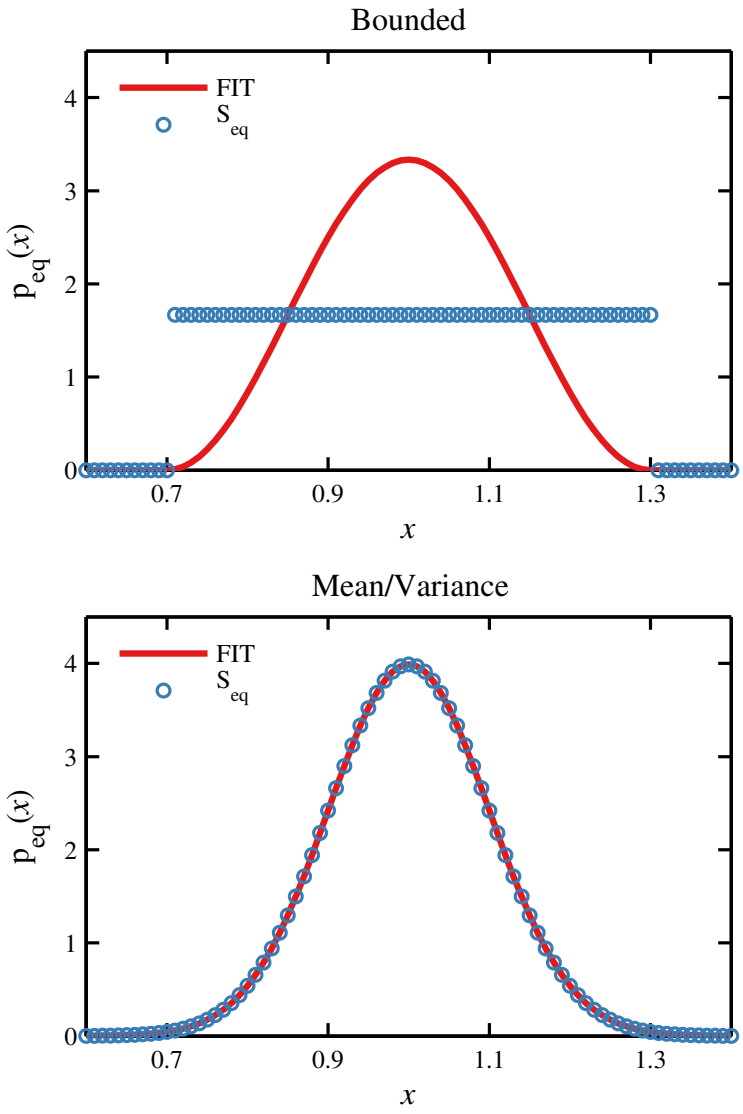

Figure 4.4: Maximum entropy distributions. (Top) Bounded domain constraint $x \in [0.7, 1.3]$ for the traditional equilibrium entropy $S_{\rm eq} = \langle \ln p_{\rm eq}(x) \rangle_{\rm eq}$ which gives a flat profile and the trajectory entropy $\mathcal{S}_{FIT} = \langle \rho | \nabla^2 | \rho \rangle$ which gives a distribution that scales as $p^*_{FIT} \sim \cos^2((x-\mu)\pi/2L)$. (Bottom) Maximum entropy distribution under the constraint on the average $\langle x \rangle = \mu$ and variance $\langle (x-\mu)^2 \rangle = \sigma^2$ which are equivalent for the static and trajectory information measure as $p^*_{FIT} = p^*_{S_{\rm eq}} \sim \exp(-(x-\mu)^2/2/\sigma^2)$.

The functional derivative $\delta\mathcal{L}/\delta\tilde{\rho}(k)$ in the Fourier space then gives the differential equation for optimization:

$$\omega_2 \frac{d^2\tilde{\rho}}{dk^2}(k) + \omega_1 \mathrm{i} \frac{d\tilde{\rho}}{dk}(k) + \eta D k^2 \tilde{\rho}(k) = 0. \tag{4.66}$$

By utilizing the translation operation in the Fourier space, we assume the functional form of the solution as $\hat{\rho}(k) = e^{-\omega_1 \mathrm{i}k/2\omega_2}\tilde{\rho}(k)$ to eliminate the first derivative in the equation:

$$\frac{d^2\hat{\rho}}{dk^2}(k) + \left(\frac{\eta D}{\omega_2}k^2 - \frac{\omega_1^2}{4\omega_2^2}\right)\hat{\rho}(k) = 0. \tag{4.67}$$

The solutions are the Hermite functions, same as those for the wave function of a quantum harmonic oscillator. The maximization step of Eq. (4.56) selects the temperature η for of the ground state vibrational mode and the mode with the least information of that has the form of a Gaussian:

$$\tilde{p}(k) \propto e^{-\omega_1' \mathrm{i}k} e^{-\omega_2' k^2}, \tag{4.68}$$

where the requirement of constraints is incorporated into ω' for simplicity. Inverse Fourier transform then takes Eq. 4.68 to the real-space solution:

$$\rho_{\mathrm{eq}}^*(x) = \mathcal{F}^{-1}(\tilde{p}(k)) \propto e^{-\omega_2''(x-\omega_1')^2}. \tag{4.69}$$

With the constants determined to satisfy the constraints of the observed mean and variance as well as probability normalization, the final result for the equilibrium probability is a Gaussian as shown in Fig. 4.4:

$$p_{\mathrm{eq}}^*(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}. \tag{4.70}$$

This result of LID model is the same as that obtained by maximizing the static entropy S_{eq} despite the use of trajectory information. However, it can be rationalized by inspecting the two different information measures and their corresponding inequality bounds which similarly are optimal when the equilibrium probability density becomes Gaussian:[133]

$$\text{Dynamic}: \left(\int dx\ x^2\rho^2(x)\right)\left(\int dk\ k^2\tilde{\rho}^2(k)\right) \geq \frac{1}{16\pi^2} \tag{4.71}$$

$$\text{Static}: -\int dx\ \rho^2(x)\ln\rho^2(x) - \int dk\ \tilde{\rho}^2(k)\ln\tilde{\rho}^2(k) \geq 1 - \ln(2). \tag{4.72}$$

The inequality of Eq. (4.71) can be viewed as an uncertainty principal of continuous stochastic dynamics.[133] It contains a static measure of the variance in position multiplied by the dynamic information $\mathcal{J}$. Here, $\mathcal{J}$ plays the analogous role of the variance of velocity in the Heisenberg uncertainty of quantum mechanics.[75] The inequality of Eq. (4.72) is the Hirschman uncertainty principle[134] which adds a static measure of equilibrium entropy $S_{\rm eq}$ with the Shannon information measure of the velocity distribution. In both cases a balance is reached between a static and dynamic measure of information, and in fact the Gaussian distribution is optimal and achieves equality in both instances for normalized distributions $p_{\rm eq}(x) \in L^2$.

LID constraining on the mean first passage time

If the transition between two meta-stable states locating at x_A, x_B with the reaction rate of $k_{A\to B}$ is known, the mean first passage time between the two states is $\tau_{\rm rxn}(x_A \to x_B) = k^{-1}_{A\to B}$. This knowledge can be used as the constraint to find the corresponding LID model. According to Gardiner[37] the mean first passage time from x_A to x_B can be expressed as

$$\tau_{\rm x_A\to x_B} = k^{-1}_{A\to B} = \frac{1}{D}\int_{x_A}^{x_B} dx\ \rho^{-2}(x) \int_{x_L}^{x} dx'\ \rho^2(x')$$
$$\tau_{\rm x_B\to x_A} = k^{-1}_{B\to A} = \frac{1}{D}\int_{x_B}^{x_A} dx\ \rho^{-2}(x) \int_{x_R}^{x} dx'\ \rho^2(x'). \quad (4.73)$$

using $\rho_{\rm eq}$ as the variable within the finite domain $x \in [x_L, x_R]$.

The constraints of Eq. (4.73) are non-linear and do not follow the advantageous linear mode of Eq. (4.52). Using $\boldsymbol{\tau}_{x_B\to x_A}$ and $\boldsymbol{\tau}_{x_A\to x_B}$ to represent the operation on ρ in Eq. (4.73), the Lagrangian of LID becomes:

$$\mathcal{L} = \eta \int dx(\nabla\rho(x))^2 + \omega_1\left(\int dx\rho^2(x) - 1\right)$$
$$+ \omega_2\left(\boldsymbol{\tau}_{x_B\to x_A}[\rho] - k^{-1}_{B\to A}\right) + \omega_3\left(\boldsymbol{\tau}_{x_A\to x_B}[\rho] - k^{-1}_{A\to B}\right). \quad (4.74)$$

The optimization requires setting the functional derivative with respect to ρ to zero:

$$2\eta\nabla^2\rho(x) \ + \ \omega_1 2\rho(x) \ + \ \omega_2\frac{\delta\boldsymbol{\tau}_{x_B\to x_A}}{\delta\rho} \ + \ \omega_3\frac{\delta\boldsymbol{\tau}_{x_A\to x_B}}{\delta\rho(x)} \quad = \quad 0. \quad (4.75)$$

Ultimately the functional derivatives of the mean first passage times depend on the actual place within the domain; only the integrals which are active in

the inner product with $f(x)$ at a given position will manifest in the functional derivative. The functional derivatives of the mean first passage time in this equation depend on the specific locations within the domain:

$$\frac{\delta \boldsymbol{\tau}_{x_A \to x_B}}{\delta \rho(x)} = \begin{cases} \frac{2}{D}\rho \int_{x_A}^{x_B} \rho^{-2}(x')dx' & x \leq x_A \\ \frac{2}{D}\rho(x) \int_{x}^{x_B} dx' \rho^{-2}(x') - \frac{2}{D}\rho^{-3}(x) \int_{x_L}^{x} dx' \rho^{2}(x') & x_A < x \leq x_B \\ 0 & x > x_B. \end{cases} \tag{4.76}$$

The derivation of this equation is shown in Appendix C.6. The non-linear optimization problem of Eq. (4.75) can be solved numerically by using a quasi-Newton method iteratively.[82] The resulting PMFs and equilibrium distributions for three sets of reaction rates are shown in Fig. 4.5.

The LID profiles constrained on mean first passage times generally follow Kramer's type of reactions[135] that have quadratic potentials for stable states connected by an inverted quadratic barrier, Fig. 4.5. The LID models also produce the Hammond-Leffler postulate that the transition state $x^{\ddagger}$ of an "endothermic" reaction is closer to the product.[136] This example shows that the analytical form of FIT can effectively utilize the information contained in dynamic observables to stitch together a least informative profile of the equilibrium densities. This approach strengthens our ability to model the phenomenology of reactive systems and estimate the underdetermined continuous parameters of a function from the subset of experimental data.

4.9 Similarity of FIT to Bures metric and trace of power spectrum matrix

Finally, we observe that the Hermitian formulation developed in this work for the Langevin equation adopts properties of the density matrix formulation of quantum mechanics. The symmetric operator of the FPE of a specific ρ_{eq} and diffusion coefficient, $\boldsymbol{H}$, defines the time propagation operator as $\boldsymbol{\rho} = e^{-\boldsymbol{H}\Delta t}$. The Hermitian FPE of the Langevin equation also delivers an eigenbasis that assembles the unitary matrix $\mathbf{1} = \boldsymbol{\Psi}^{\dagger}\boldsymbol{\Psi}$ such that

$$\boldsymbol{\rho} = \boldsymbol{\Psi} e^{-t\boldsymbol{\Lambda}} \boldsymbol{\Psi}^{\dagger}. \tag{4.77}$$

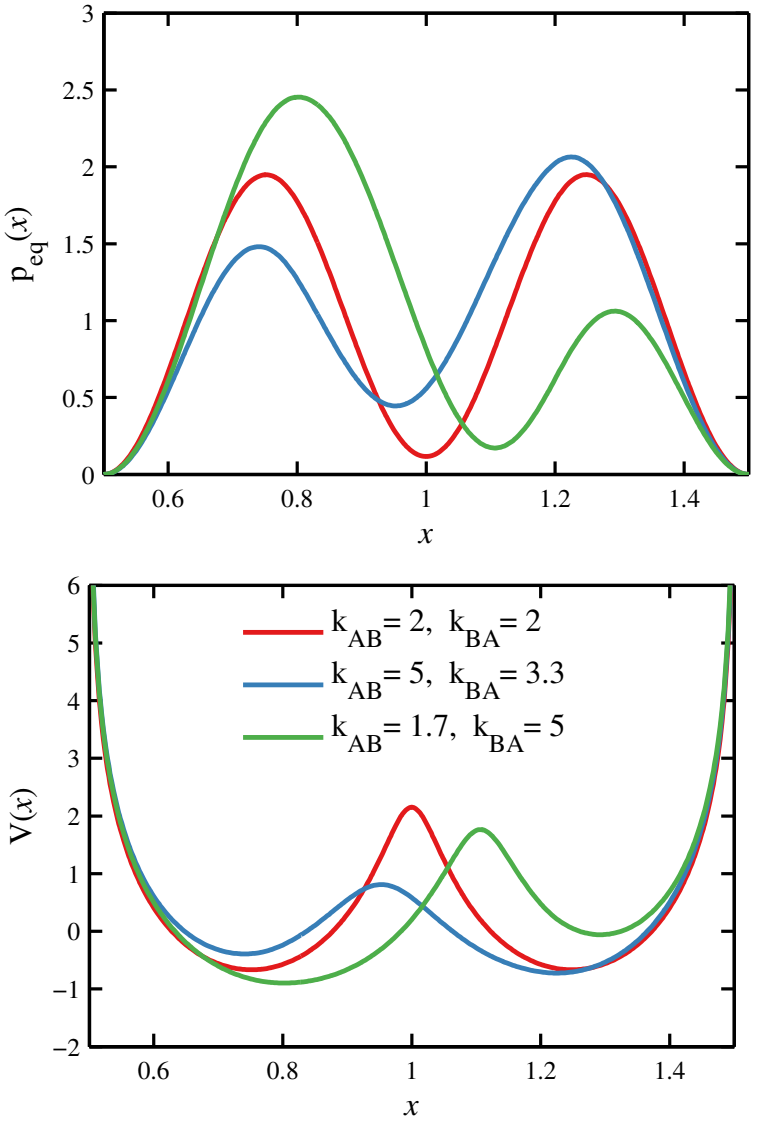

Figure 4.5: Least informative dynamic models for using the mean first passage times as constrains. (Top) The optimal equilibrium probability density $p^*_{\text{eq}}(x)$ for the constraint of rate constant / mean first passage time between two states located at $x_A = 0.8$ and $x_B = 1.2$ with diffusion constant $D = 1$. (Bottom) The corresponding potential of mean force $V^*(x) = -\ln p^*_{\text{eq}}(x)$. The constrained values of rate constants in the unit of s^{-1} are labeled as the legend of each profile.

The density matrix formulation of quantum mechanics (QM) has an isomorphic symmetrized form in the density operator time evolution:

$$\boldsymbol{\rho}(t) = e^{-\boldsymbol{H}t/2}\boldsymbol{\rho}(0)e^{-\boldsymbol{H}t/2} \tag{4.78}$$

when initially the operator is the diagonal delta function:

$$\boldsymbol{\rho}(0) = \hat{1} = \int dx\ |x\rangle\langle x| \quad ; \quad \langle x_t|\boldsymbol{\rho}(0)|x_0\rangle = \delta(x_t - x_0). \tag{4.79}$$

The power spectrum of the QM density operator, i.e., the collection of its eigenvalues, is an optimal basis from which to calculate the information content for mixed states with the Von Neumann entropy,[137, 138] or the Sinai entropy of transitions.[139, 140] Similarly from classical mechanics, Lyoponov stability analysis shows the importance of this spectral decomposition.[141–143] The QM version of the information metric between density matrices, the Bures distance, takes on the general features of the QM operator trace to give:[144–146]

$$\mathcal{S} = 2 - 2\,\mathrm{Tr}\left(\sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}}\right) \tag{4.80}$$

to measure the change in information content between two density matrices $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$.

The Bures metric distance is the QM analog of the Fisher information, which can be calculated from the eigenbasis of the density operator:[147]

$$\mathcal{I}^{\mathrm{QM}}(x,y) = \sum_{i,j} \frac{1}{2(\lambda_i + \lambda_j)} \frac{\delta\langle\psi_i|\boldsymbol{H}|\psi_j\rangle}{\delta\rho_{\mathrm{eq}}(x)} \frac{\delta\langle\psi_i|\boldsymbol{H}|\psi_j\rangle}{\delta\rho_{\mathrm{eq}}(y)}. \tag{4.81}$$

We note the correspondence with the FIT we calculated in the continuum limit with

$$\mathcal{I} = \rho_{\mathrm{eq}}(x)\rho_{\mathrm{eq}}(y)\mathcal{I}_{\mathrm{QM}}. \tag{4.82}$$

It can be seen in in Fig. 4.6 that the contours of I defined in Eq. (4.82) approach the shape and magnitude of the tri-banded matrix from the numerical calculations of the Laplacian kernel from the operator analysis presented in Eq. (4.45). Therefore, in the continuum limit of $\Delta t \to 0^+$, the information contents in QM and Langevin dynamics appear to be isomorphic.

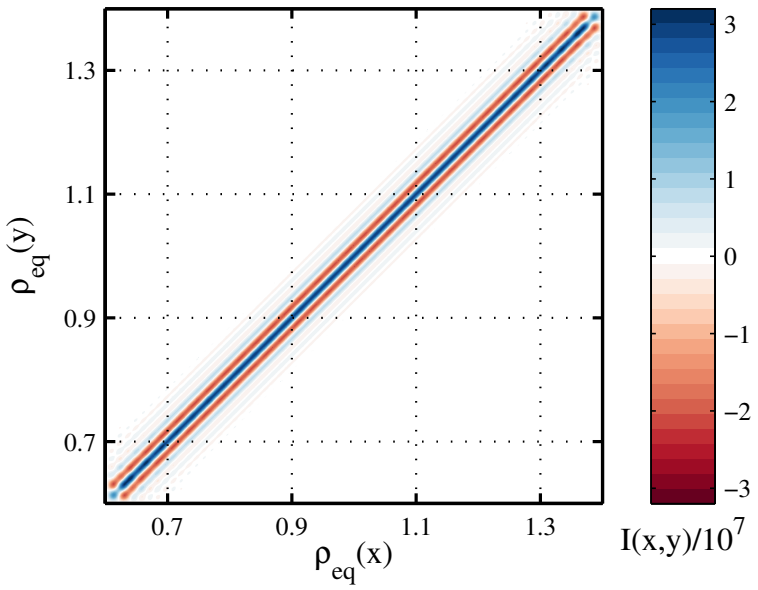

Figure 4.6: Contours of the QM Fisher Information $\mathcal{I} = \rho_{\text{eq}}(x)\rho_{\text{eq}}(y)\mathcal{I}_{\text{QM}}$, or Bures metric, calculated with 32 eigenvectors to illustrate the behavior of the tri-banded Laplacian kernel in approaching the continuum limit. Using a greater number of eigenvectors gives a narrower tri-banded matrix and closer resemblance to the Laplacian kernel. Eigenvectors were calculated for the reference three-well PMF shown in Fig. 4.1 but the results are invariant to choice of the potential since the Fisher and Bures metrics are constant.

4.10 Conclusion

For trajectories following the Langevin equation at equilibrium, a theoretical framework is developed in this work to evaluate the Fisher information. The corresponding Fokker-Planck equation of Langevin dynamics is transformed into a Hermitian form to allow an eigenbasis decomposition of the time propagator. Essential to the success of this approach is using the square root of the equilibrium probability density as the form on model parameters. This unique choice of basis not only symmetrizes the FPE, it also makes the Fisher information matrix constant, resulting in tremendous simplification in the calculation of the information metic of trajectories in the parameter space. In the continuum limit, we show that the analytical form of the Fisher information matrix for Langevin trajectories is a Laplacian kernel. With this constant-information matrix, a line integral in the parameter space can be devised to give an analytical

measure of dynamics information given the PMF and diffusion coefficient of the Langevin equation. An immediate impact of this derivation is enabling the imputation of least informative dynamics model constrained on the observables that are known or available. Although the examples of LID models presented in this work were limited to scalar constraints, generalization to using an arbitrary functional in describing the observable of interest, such as the likelihood of single-molecule FRET measurements,[60, 83, 148] other properties measured by single-molecule methods, or molecular simulation results, is expected to be equally applicable. The methodology developed in this work can be used to systematically utilize the measurable data to formulate a dynamics model based the principle of the least informative description.

CHAPTER 5

Information Thermodynamics for Bayesian Optimization: Applications to Single-Molecule FRET

5.1 Abstract

We present a systematic Bayesian framework for developing models of dynamics by extracting information from experimental or simulation data. These models are derived within the formalism of information thermodynamics which balances the inputs of empirical likelihood with the inherent entropy of the resulting model of dynamics. The specific choice of the balance between entropy and likelihood produces the equivalent of a phase diagram for optimization where the ideal results are generated at a critical point. Applications are presented specifically for deriving the potential of mean force (PMF), equilibrium probability density, and diffusion coefficient for systems modeled with over-damped Langevin dynamics. The single-molecule FRET experiment provides a statistical sampling of these dynamics and provides the likelihood information. The optimal PMFs and diffusion from this algorithm are shown for FRET data from the adenylate-kinase enzyme studied with and without inhibitor binding.

5.2 Introduction

In the statistical learning paradigm, we seek to use data to develop models of physical systems. Experiments and simulation offer insights into the pro-

gression of a systems dynamics and are statistical reporters to the underlying process that is the goal of the study. However there are inherent limitations to the precision and resolution of the experiment and the timescales accessible through simulation. Directly translating experimental results into a dynamic is therefore difficult because the full details and rich features of the true system are naturally blurred. Thus when performing the inverse mapping of data to model, the problem is inherently underdetermined. Ideally, a machine learning algorithm would produce the least descriptive model which still contains the full information from the data granting confidence that the true system's behavior is at least as complicated and feature rich as the optimal result. A prototypical example is the de-noising of photographs from a telescope for the reconstruction of details in the night sky.[149] Naturally the images are blurred and convoluted with specific optics of the instrument, but the de-convolution operation is an ill-posed problem because a myriad of complex features could give rise to the same blurred image. Therefore the operation requires a regularization to produce images which only show significant features that are confidently evidenced by the data.

Our focus is on studying the dynamics of condensed phase system and we seek to find parameters for the time evolution of general coordinates of interest. The over-damped Langevin equation which governs the coarse-grained dynamic behavior of the physical system is given by

$$\dot{x} = -D(x)\nabla\frac{V(x)}{k_B T} + \nabla D(x) + \sqrt{2D(x)}dW_t \tag{5.1}$$

where $V(x)$ is the potential of mean force for the reduce coordinate system x, and $D(x) = k_B T/\zeta$ is the diffusion coefficient with collision frequency ζ. The random motion is driven by the Weiner process dW_t with the expected average $\langle dW_t \rangle = 0$ and variance $\langle dW_t \cdot dW_{t'} \rangle = \delta(t - t')$. The diffusion coefficient can be position dependent for systems where barriers to reactivity aren't contained within the probed coordinate [12]. Because we don't necessarily study the actual reactive pathway, diffusion may be positionally varying and reflective of the non-markovian nature of coordinate [150].

5.3 Information Thermodynamics

Effectively the approach outlined in this work seeks a Bayesian optimization of the parameters for Langevin Dynamics [151]. The information from experiment comes in the form of an internal energy is also the negative log-likelihood $E[\theta] = -\ln P(Y(t)|\theta)$ where θ are the parameters and $Y(t)$ are the experimental

observations in time. For over-damped Langevin dynamics, the likelihood is a functional of the parameters of $V(x)$ which is the PMF of the underlying system giving rise to the microstates of the system which are the dynamic trajectories $X(t)$.

$$P(Y(t)|\theta) = \int DX(t)P(Y(t)|X(t))P(X(t)|\theta) \tag{5.2}$$

In this decomposition, the specific experimental outputs depend on the actual micro states of the system $P(Y(t)|X(t))$ while the time evolution of the system is governed by the parameters of interest. Maximum likelihood optimization would aim to optimize the parameter set θ to maximize this quantity. However, as discussed previously , this optimization is underdetermined and the degeneracy of solutions will create huge variance in the possible outcomes.

To make direct analogy with thermodynamics we can define the following function for internal energy $E(S, \tilde{D})$ where the arguments are the extensive variables of entropy S and log-diffusion $\ln(D) = \tilde{D}$ which sets the timescale for the dynamics. The differential of internal energy is

$$dU = \eta_F dS + \eta_D d\tilde{D} \tag{5.3}$$

where η_F is an intensive quantity of "temperature" and η_D is the diffusion temperature; both of which are intensive parameters of the system. As shown in [79], the trajectory entropy for dynamic systems can also be related to a functional of the order parameters by $S_F[\theta] = -D\langle F^2(x)\rangle$. This functional captures the actual degree of information generated by a given parameter set onto the ensemble of equilibrium trajectories.

The remaining thermodynamic functionals can then be found the following ensembles. First the Helmholtz free energy which is found by subtracting the factor of entropy for the Legendre transformation

$$A(\eta_F, D) = E - \eta_F S \tag{5.4}$$

$$dA = -Sd\eta_F + \eta_D d\tilde{D} \tag{5.5}$$

Now the thermodynamic function is parameterized by the temperature and diffusion coefficient and the parameterizing PMF is allowed to fluctuate to achieve its optimal equilibrium result. The Gibbs free energy is found by a second transformation recognizing that the diffusion part of trajectory entropy $S_D[\theta] = \ln D$ is equivalent to the extensive "volume" in an N-P-T system thus the term added to generate the free energy $G(\eta_F, \eta_D)$ is

$$G(\eta_F, \eta_D) = A - \eta_D \tilde{D} \tag{5.6}$$

$$dG = -Sd\eta_F - \tilde{D}d\eta_D \tag{5.7}$$

And for completeness the enthalpy is

$$H(S, \eta_D) = E - \eta_D \tilde{D} \tag{5.8}$$

$$dH = Sd\eta_F - -\tilde{D}d\eta_D \tag{5.9}$$

These thermodynamic functionals are essential the log of a hierarchical Bayesian model of the system parameters and the optimization of the free energy amounts to a maximum a posteriori estimation problem.

$$\arg\max_{\theta} p(\theta|y) = \frac{p(Y(t)|\theta)p(\theta)}{p(y)} \tag{5.10}$$

$$p(\theta) = \exp(-\eta S[\theta]) \tag{5.11}$$

Free energy A can also be composed of the functionals for likelihood and entropy for a given parameter set. At equilibrium, free energy is minimized meaning that we set a temperature η_F and allow force to fluctuate, however the force achieves an equilibrium value in the thermodynamic limit. This translates to solving an optimization problem over force

$$A(\eta_F, D) = \inf_{F(x)} \quad -\ln P(Y(t)|\theta) + \eta_F D\langle F^2(x)\rangle \tag{5.12}$$

5.4 Bundle Method Optimization

The likelihood can be approximated by the bundle method [36, 152] which constructs a convex hull approximation of the likelihood surface by a bundle of hyperplanes which are the derivatives evaluated at points along the course of the minimization

$$\ell(\theta) \leq l^{CP}(\theta) = \min_i \langle \theta(x), \delta_\theta^i \rangle + b_i \tag{5.13}$$

where θ are the parameters for the time propagation probability distributions, δ_θ^i are the gradients of log-likelihood evaluated at point θ^i and b_i are the intercepts $b_i = \ell(\theta^i) - \langle \theta(x), \delta_\theta^i \rangle$. The parameter $\epsilon = \max_i l(\Theta_i) - \sup_\Theta l^{CP}$ measures the precision of this convex hull and is strictly positive during the convergence of the convex optimization.

For the study of dissipative systems which undergo Langevin dynamics, the ideal parameter set for minimum free energy optimization is the square root of equilibrium probability $\rho_(x) = \sqrt{p_{\text{eq}}}$ and diffusion constant D[153] because

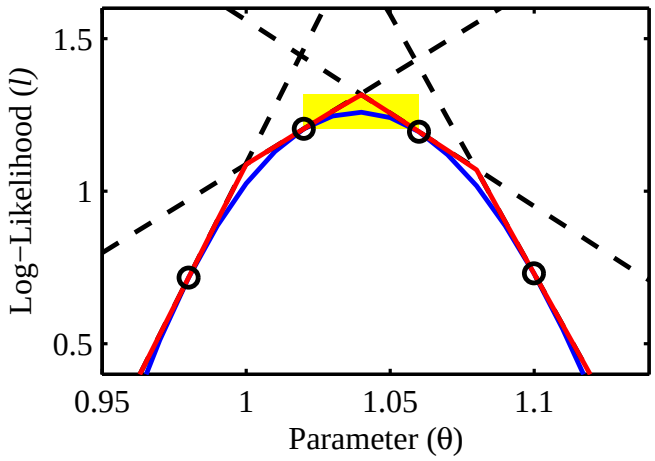

Figure 5.1: Bundle approximation to the likelihood constructed from δ_θ hyperplanes. The blue line is the actual likelihood of the data as a function of model parameter(s) $\ell(\theta)$ which in this case is a linear multiplier to the correct system force $F = \theta F^*$. The circles (o) indicate where a likelihood evaluation and gradient has been calculated. The dotted black lines (–) are the resulting bounding hyper-planes and the red line is the minimum of this bundle which gives the closest bound to the true likelihood. The yellow box shows the deviation in ϵ and the variance in the optimal θ^* in finding the maximum likelihood using the current set of bundles.

the entropy is a simple linear functional $S[\rho(x)] = \langle\rho|\nabla^2|\rho\rangle$. Therefore the likelihood bundle is constructed from

$$-\ln P(Y(t)|\theta) \approx \max_k \langle\delta^k_{\rho^2}|\rho^2\rangle + \langle\delta^k_{\tilde{D}}|\tilde{D}\rangle + b^k \tag{5.14}$$

which is an approximation of the likelihood surface by hyperplanes k defined by the slopes δ^k and intercept b^k.

After inserting into the free energy equation, the bundle method effectively solves the minimization by introducing the slack variable $\xi \leq \langle\delta^k_{\rho^2}|\rho^2\rangle + \langle\delta^k_{\tilde{D}}|\tilde{D}\rangle + b^k \; \forall k$ which is satisfied by introducing the Lagrange multipliers $\boldsymbol{\omega} > 0$ for each k. The resulting optimization problem is recast with the Lagrangian

$$\mathcal{A}[\rho(x), \xi, D, \boldsymbol{\omega}] = \xi + \eta_F\langle\rho|\nabla^2|\rho\rangle - \boldsymbol{\omega}\cdot\left(\xi\mathbf{1} - \langle\boldsymbol{\delta}_{\rho^2}|\rho^2\rangle - \langle\boldsymbol{\delta}_{\tilde{D}} - \boldsymbol{b}|\tilde{D}\rangle\right) \tag{5.15}$$

where the thermodynamic quantity $A(\eta_F, D) = \inf_\rho \mathcal{A}$ is achieved by performing the functional optimization which is the Fenchel dual of the fully specified functional $\mathcal{A}$.[3] where $\boldsymbol{\delta}_{\rho^2}$ is a matrix of where each row is the gradients $\boldsymbol{\delta}^k_{\rho^2}$ and $\boldsymbol{b}$ is the vector of intercepts. Using the principals of convex optimization, we first optimize with respect to ξ to give $1 - \boldsymbol{\omega}\mathbf{1} = 0$ which means $\|\omega\|_1 = 1$. Then we do the functional optimization over $\rho(x)$ by taking the functional derivative of the Lagrangian $\frac{\delta \mathcal{A}}{\delta \rho(x)}$ which is equivalent to finding by variational calculus

$$\inf_\rho \int \boldsymbol{\omega} \cdot \boldsymbol{\delta}_{\rho^2} \rho^2 - \eta_F D \int \left(\rho'\right)^2 + \lambda \left(\int \rho^2 - 1 \right) \tag{5.16}$$

which includes the condition that the distribution must sum to 1 with Lagrange multiplier λ. Preforming the functional derivative with respect to ρ and setting equal to 0 gives an eigenfunction differential equation

$$\eta_F D \rho'' - \boldsymbol{\omega} \cdot \boldsymbol{\delta}_{\rho^2} \rho = \lambda \rho \quad , \quad \rho|_{-L,L} = 0 \tag{5.17}$$

with Dirichlet boundary conditions. Using a spectral element method[154], the eigenfunctions and eigenvalues can be determined. Only the first eigenvector which contains no nodes satisfies the condition that $\rho > 0$ with the first eigenvalue $\lambda_0(\boldsymbol{\omega}; \eta_F, D)$ which depends on the bundle weights. Inserting this solution back into the Lagrangian gives the free energy desired up to the parametric dependence on bundle weights.

Finally these coefficients ω_k can be determined by solving the dual problem for finding the equilibrium free energy parameters

$$A(\eta_F, D) = \max_{\substack{\boldsymbol{\omega} \\ \|\boldsymbol{\omega}\|_1 = 1 \\ \boldsymbol{\omega} \succeq 0}} \quad \lambda_0(\boldsymbol{\omega}) + \langle \boldsymbol{\omega} \cdot \delta_{\tilde{D}} | \tilde{D} \rangle \tag{5.18}$$

The remaining optimization over the bundles ω is solved by a sequential quadratic program (SQP) in MATALB where free energy is approximated by its second order Taylor series at each iteration and solved by a quadratic program (QP) with linear constraints on ω to give an approximation of free energy at a given number of bundles k and the optimum bundle vector

$$\omega^k = \arg\max_{\substack{\boldsymbol{\omega} \\ \|\boldsymbol{\omega}\|_1 = 1 \\ \boldsymbol{\omega} \succeq 0}} \quad \omega \cdot \left(\nabla_\omega \lambda_0(\omega; \eta_F, D) + \langle \delta_{\tilde{D}} | \tilde{D} \rangle + \boldsymbol{b} \right) + \frac{1}{2} \omega^\top \left(\nabla_\omega \nabla_\omega \lambda_0(\omega; \eta_F, D) \, \omega \right. \tag{5.19}$$

The derivatives of the eigenvalues A_0 from the $\theta(x)$ optimization problem are given by perturbation theory using the higher order eigenvectors $\theta_n(x)$.

$$\nabla_\omega A_0(\omega;\eta_F) = \langle\theta_0|\delta_{\theta^2}|\theta_0\rangle \tag{5.20}$$

$$\nabla_\omega\nabla_\omega A_0(\omega;\eta_F) = \sum_{n=1}^{\infty}\frac{\langle\theta_n|\delta_{\theta^2}\otimes\delta_{\theta^2}|\theta_n\rangle}{A_n - A_0} \tag{5.21}$$

Convergence of the bundle is tested by tracking the difference between the actual and approximate values of free energy $\epsilon^k = A(\eta_F, D, \rho^*) - A^k(\eta_F, D)$. Since the problem is convex, this epsilon will always be greater than one and reach zero when the bundle accurately describes the likelihood surface. In practice, the convergence of ϵ is nearly exponential which is a marked improvement over the sub-linear convergence from the Expectation-Maximization (EM) algorithm used in the authors previous study of single-molecule FRET[155]. Typically 5 iterations are required for convergence of the bundle at a given temperature and the total operation of finding the optimal temperature stopping point requires 300 iterations. In contrast the EM algorithm required 50,000 iterations.

The optimal bundle coefficients give the next best guess for the optimal profile $\rho^{k+1}(x)$ which is then used to construct an improved representation of the bundle. This operation is iterated until convergence of the ϵ at which point the regularization temperature η_F is lowered to explore lower energy regions of the phase space.

The $\eta_D - \tilde{D} - \eta F$ equation of state (EOS) can be mapped out by calculating the derivative of A at each value of diffusion at a set force temperature

$$\frac{\partial A}{\partial \tilde{D}} = \eta_D \tag{5.22}$$

$$\eta_D = D\left(\boldsymbol{\omega}\cdot\boldsymbol{\delta_D} + \eta_F\frac{d\lambda_0}{dD}\right) \tag{5.23}$$

where the change in eigenvalue is given by perturbation theory

$$\frac{d\lambda_0}{dD} = \langle\rho^*|\nabla^2|\rho^*\rangle \tag{5.24}$$

5.5 Optimization of $G(\eta_F, \eta_D)$

To do optimization in the η_F, η_D ensemble, it is advantageous to introduce the composite variables $\theta(x) = \sqrt{D}\rho(x)$ such that the trajectory entropy $S =$

$D\langle\rho|\nabla^2|\rho\rangle$ is converted to $S = \langle\theta|\nabla^2|\theta\rangle$. The resulting Lagrangian is slightly modified to

$$\mathcal{G}(\eta_F, \eta_D; \omega, \theta, D) = \langle\omega\delta_{\theta^2}|\theta^2\rangle - \eta_F\langle\theta|\nabla^2|\theta\rangle + \lambda\left(\langle\theta|\theta\rangle - D\right) + \langle\omega\delta_D|D\rangle + \eta_D\ln(D) + \omega b \quad (5.25)$$

where essential the only difference is that the new θ should be normalized to D instead of 1. Doing the minimization over θ yield the similar eigenvalue problem

$$\left(\eta_F\frac{d^2}{dx^2} - \omega\cdot\delta_{\theta^2}\right)\theta = \lambda\theta \quad (5.26)$$

Naturally the eigenvector solution will have $\langle\theta^*|\theta^*\rangle = 1$ for eigenvalue λ_0 so the result is scaled by D to satisfy the constraint in the Lagrangian. Substituting this optimum θ^* yields

$$\mathcal{G}^*(\omega, D; \eta_F, \eta_D) = \inf_\theta \mathcal{G} = D\lambda_0(\omega;\eta_F) + \langle\omega\delta_D|D\rangle + \eta_D\ln(D) + \omega b \quad (5.27)$$

Performing the minimization over D involves inverting the equation

$$-\eta_D\frac{1}{D} = \lambda_0(\omega;\eta) + \omega\delta_D \quad (5.28)$$

and solving for the optimum diffusion

$$D^* = \frac{\eta_D}{\lambda_0(\omega;\eta_F) + \omega\cdot\delta_D} \quad (5.29)$$

and the resulting Lagrangian or "Free Energy" including the bundle intercepts E is

$$\mathcal{L}^*(\omega;\eta_F,\eta_D) = \eta_D - \eta_D\ln\eta_D + \eta_D\ln(\lambda_0(\omega;\eta_F) + \omega\cdot\delta_D) + \omega b \quad (5.30)$$

With these Lagrangians, can produce a quadratic approximation with respect to ω then apply the newton solver to maximize and get the ω^*. The gradients for the newton solver with respect to ω are

$$\nabla_\omega\mathcal{L}^*(\omega;\eta_F,\eta_D) = E + \eta_D\frac{\nabla_\omega\lambda_0(\omega;\eta_F) + \delta_D}{\lambda_0(\omega;\eta_F) + \omega\delta_D} \quad (5.31)$$

with Hessian

$$\nabla_\omega\nabla_\omega\mathcal{L}^*(\omega;\eta_F,\eta_D) = \eta_D\frac{\nabla_\omega\nabla_\omega\lambda_0(\omega;\eta_F)}{\lambda_0(\omega;\eta_F) + \omega\delta_D} - \eta_D\left(\frac{\nabla_\omega\lambda_0(\omega;\eta_F) + \delta_D}{\lambda_0(\omega;\eta_F) + \omega\delta_D}\right)\otimes\left(\frac{\nabla_\omega\lambda_0(\omega;\eta_F) + \delta_D}{\lambda_0(\omega;\eta_F) + \omega\delta_D}\right) \quad (5.32)$$

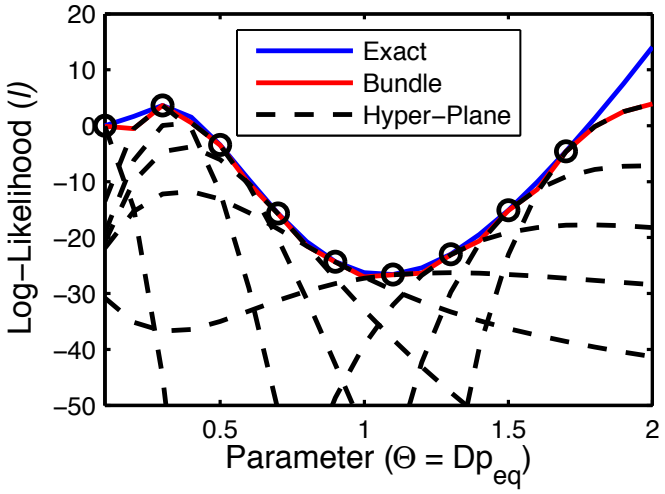

Figure 5.2: Application of bundle approximation to the FRET log-likelihood. Plotted is the log-likelihood loss for the 3-well model with $D = 500$ evaluated using exact path integral at parameters which are a linear fraction of the ideal 3-well potential and diffusion coefficient $\Theta = [0 \dots 2] D^* p^*_{eq}$. The gradients are evaluated at 10 of these points and then a linear extrapolation constructs a hyper-plane approximation of the likelihood that is always lower than the exact likelihood for convex problems. The maximum of these hyper-planes is the "Bundle".

5.6 The Hyper-plane derivatives from smFRET

The information thermodynamics formalism presented thus far is invariant to the specific type of simulation of experiment. The specifics of the approach is codified into the calculation of the actual likelihood derivatives. For the smFRET experiment, a robust means of calculating the likelihood was presented in[155] with the main results summarized here that the likelihood can be reprinted by a path integral

$$\ell = \ln \langle \alpha_0 | e^{-\boldsymbol{H}\Delta t_1} \boldsymbol{y}_1 e^{-\boldsymbol{H}\Delta t_2} \boldsymbol{y}_2 \dots e^{-\boldsymbol{H}\Delta t_{N_P}} \boldsymbol{y}_{N_P} | \beta_{t_{\exp}} \rangle. \tag{5.33}$$

with operators that describe the system dynamics $\boldsymbol{H}$ and depend on the parameterizations of ρ by

$$\boldsymbol{H} = \frac{1}{\rho(x)} \cdot \left(\nabla D(x) \rho^2(x) \nabla \frac{1}{\rho(x)} \right) \tag{5.34}$$

and the information input from the smFRET experiment $\boldsymbol{y}$.

The derivative of log-likelihood from this path integral with respect to an arbitrary parameter of the hamiltonian is

$$\frac{\delta \ell^k}{\delta \theta(x)} = \sum_{i,j} \frac{\delta \langle \psi_i | \boldsymbol{H} | \psi_j \rangle}{\delta \theta(x)} \mathbb{E}^k_{X|Y}[a_i b_j] \tag{5.35}$$

where $\mathbb{E}^k_{X|Y}[a_i b_j]$ is the magnitude of the coefficients for those eigenvectors during the inference of the dynamic trajectory.

Specifically for the factors explored in this work, the gradients with respect to $\Theta = \theta^2$ are

$$\delta_{\theta^2}(x) = \delta_\Theta(x) = \frac{1}{2}\frac{\Phi'(x)}{dx} + \frac{F(x)}{2}\Phi'(x) - \frac{\alpha_T(x) + \beta_0(x)}{2D\rho(x)} \tag{5.36}$$

The gradients with respect to $D(x)$ are

$$\delta_D(x) = \rho(x)\Phi''(x)\rho(x) + \frac{(\alpha_T(x) + \beta_0(x))\rho_{eq}(x)}{2D(x)} \tag{5.37}$$

where the derivative for constant D is found by taking the integral

$$\delta_D = \int dx \delta_D(x) \tag{5.38}$$

The symbols of Φ are shorthand for the path expectations of space derivatives of the eigenvectors for the dynamic operator.

$$\Phi'(x) = \sum_{i,j} (\phi_i'(x)\phi_j(x) + \phi_i(x)\phi_j'(x)) \mathbb{E}_Y[\alpha_i \beta_j] \tag{5.39}$$

and

$$\Phi''(x) = \sum_{i,j} (\phi_i'(x)\phi_j'(x)) \mathbb{E}_Y[\alpha_i \beta_j] \tag{5.40}$$

5.7 Temperature selection

Maximum entropy optimization is a balance between the entropy of the resulting dynamic model and energetic of the likelihood. The temperature sets the balance between these two terms and must be determined to satisfy our desire for the least informative model that is explained by the data. Stated differently,

we wish that the solution we find is universal in that if the experiment was repeated, we would arrive at a similarly detailed picture of the underlying dynamics of the process. To be avoided are situations where arbitrary additional states or dynamical features are presented that are beyond the possible resolution of the experiment itself.

Numerically, we seek the least perturbed solution of highest likelihood that still converges to one convex solution. This point represents a critical point in free energy space where for any lower temperature, the solution would bifurcate into multiple solutions who's extra detail is whole arbitrary. Effectively the ω's are like mole fractions so, by analogy with multicomponent distillation we are looking for the azeotrope.

Looking to approximate at each iteration empirical Bayesian model selection of temperature

$$p(\eta|y) = \frac{\int d\theta p(Y(t)|\theta,\eta)p(\theta|\eta)p(\eta)}{p(y)} \tag{5.41}$$

where $p(Y(t)|\theta,\eta)$ is the likelihood and $p(\theta|\eta) = \exp(\eta S[\theta])$ is the entropic regularizer. Then the task is to maximize the marginal with respect to temperature [1]

$$\eta^* = \arg\max_{\eta} p(\eta|y) \tag{5.42}$$

Because the logarithm is a convex function, it's equivalent to maximizing the logarithm to give

$$\eta^* = \arg\max_{\eta} \log \int d\theta \exp\left(l_y(\theta) + \eta S[\theta]\right) \tag{5.43}$$

Where $l_y(\theta)$ is the log-likelihood for the observed data.

Multivariate LaPlace approximation will allow this integration assuming gaussian statistics[156]

$$\int dx e^{-Nh(x)} \simeq e^{-Nh(\hat{x})}(2\pi)^{d/2}|\Sigma|^{1/2}N^{-d/2} \tag{5.44}$$

where $\Sigma = (\nabla\nabla h(\hat{x}))^{-1}$ is the inverse of the Hessian at the optimum $\hat{x} = \arg\min_x h(x)$.

[1] This approach is sometimes presented without the treatment of the temperature as a function parameter and instead the marginal $p(y)$ itself, called the experimental "evidence", is maximized. These methods are termed Maximum Evidence and have been used to select the number of states for a HMM treatment of FRET[67]

At critical point the Hessian $H \gtrsim 0$ is essentially singular and the determinant of covariance matrix can be express in terms of product of eigenvalues of matrix

$$|\Sigma| = \prod_{i=1}^{d} \sigma_{\lambda_i} = \prod (\sigma_H)^{-1} \tag{5.45}$$

When maximizing the log of marginal likelihood the determinant reduces to sum of log of eigenvalues to give

$$\eta^* = \arg\min_{\eta} A(\theta^*) + 1/2 \sum_i \log \sigma_A^i(\theta^*) \tag{5.46}$$

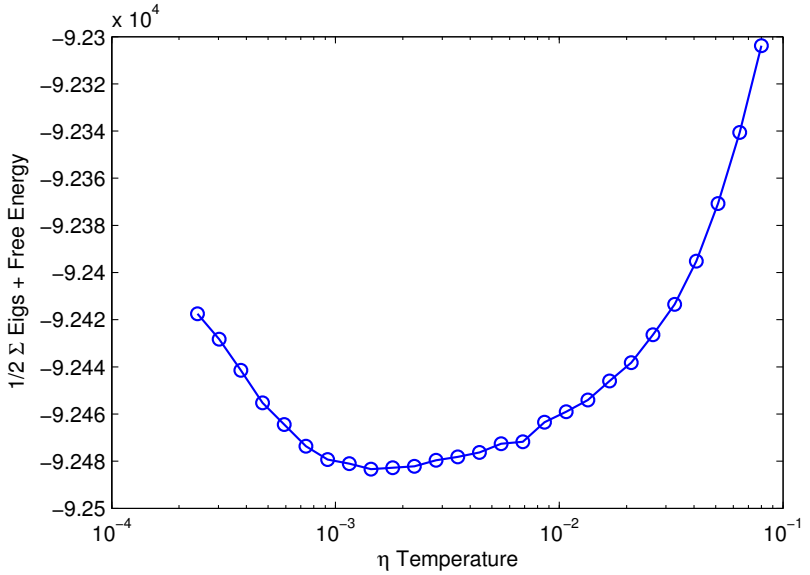

Figure 5.3: Trial optimization of temperature on 50 bundles using Equation 5.46. This would imply an ideal temperature of $\eta_F = 1x10^{-3}$. Because there is no divergence in this optimization plot, the system is still above the critical point.

Therefore at a critical point, there is the maximum in marginal likelihood and thus ideal temperature. Critical points are also the inception in a bifurcation of the solution into separate local minimum or "phases". By choosing a temperature above the critical point we are guaranteed a universal and global solution, but in balance, the lower the temperature the more detail in the proposed solution. Therefore the problem works to find the $\limsup_{\epsilon \to 0} \eta$.

To get the ideal temperature from empirical Bayesian theory we desire a balance between the maximum in free energy with the flatness of the posterior distribution so applying equation 5.46 we get

$$\eta^* = \arg\max_{\eta} \mathcal{L}^*(\eta, \omega^*) + \frac{1}{2} \ln |\nabla_\omega \nabla_\omega \mathcal{L}^*(\eta, \omega^*)| \tag{5.47}$$

For the diffusion temperature this gives

$$0 = -\ln \eta_D^* + \ln(\lambda_0(\omega^*) + \omega^* \cdot \delta_D) + \frac{N}{2\eta_D^*} \tag{5.48}$$

where N is the number of rows in the Hessian, we get to remove the constant multiple η_D. However, if the Hessian is singular, we would apply the pseudo-determinant in which case $N = rank(\nabla_\omega \nabla_\omega \mathcal{L})$ is the rank of the Hessian matrix.

5.8 Test for Binodal Decomposition

During the course of optimization, can test for the property that the minimization of free energy at constant D when $\delta A(\eta_F, D) \geq 0$ matches the optimization for free energy at constant η_D when $\delta G(\eta_F, \eta_D) \geq 0$. Inside of the binodal, when phase splitting exists, if the optimal $\boldsymbol{\omega}$ is fed between the two optimizations, should receive matching diffusion temperature and diffusion constant between them. Only situation when they should diverge is when in the binodal region of phase diagram in which the constant η_D optimization will change to one of the coexisting phases.

Points of binodal decomposition are situations in which multiple classes of solutions to the optimization problem exist. This is not desirable because one would like to have a unified picture of the possible system parameters. Going to a low temperature where phase splitting occurs means that features are beginning to appear in the solution space that are not necessarily implied by the data and may be spurious.

5.9 Critical Point Search

For the information thermodynamics formulation which works in the space of θ and the logarithm of diffusion LD, the critical point at which the empirical Bayes formula is a maximum due to the degeneracy of the correlations at all

length scales, can be found from normal thermodynamic manipulations. The critical point is when

$$\left(\frac{\partial \eta_D}{\partial LD}\right)_{\eta_F} = \left(\frac{\partial^2 \eta_D}{\partial LD^2}\right)_{\eta_F} = 0 \tag{5.49}$$

Thus by knowing the functional dependence of the free energy with respect to the two intensive temperatures, we can find the critical point by

1. Optimize at given η_F and D and converge bundle
2. Invert 5.30 to get η_D^*
3. Run again at $D \pm \Delta D$
4. Take the first and second derivatives of η_D w.r.t. D
5. If first derivative negative, within spinodal of mechanical instability. Raise η_F
6. If second derivative negative, reduce D. If second derivative positive, increase D to tend towards critical point when $\frac{d^2\eta_D}{d(LD)^2} = 0$ When setting just D the 1st and 2nd derivative are $\frac{d\eta_D}{d(LD)} = \frac{1}{D}\frac{d\eta_D}{dD}$ and $\frac{d^2\eta_D}{d(LD)^2} = \frac{1}{D}\left(\frac{-1}{D^2}\frac{d\eta_D}{dD} + \frac{1}{D}\frac{d^2\eta_D}{dD^2}\right)$

One can also leverage triple product rule expression to relate compressibility to derivatives with respect to η_F in the constant D and constant η_D ensembles

$$\left(\frac{\partial \eta_D}{\partial \tilde{D}}\right)_{\eta_F} = -\left(\frac{\partial \eta_D}{\partial \eta_F}\right)_{\tilde{D}} \Big/ \left(\frac{\partial \tilde{D}}{\partial \eta_F}\right)_{\eta_D} \tag{5.50}$$

5.10 Constrained Optimization Techniques

The actual optimization of the bundle coefficients is handled with a constrained Newton-Raphson solver in MATLAB where at each iteration the local approximation of the dual-space Lagrangian $\mathcal{L}^*(\omega)$ around ω^0 to give a proposed optimum point $\omega^* = \omega_0 + \Delta\omega$ then the Hessian and gradients are updated and a test for first-order optimality is ensured

$$\left\| \nabla \mathcal{L}(\omega) + \lambda_{\sum \omega < 1} \vec{1} - \vec{\lambda}_{\omega > 0} \right\|_\infty < 10^{-9} \tag{5.51}$$

and the method is cycled for self-consistency.

For numerical stability purposes, the constraint that $\sum \omega = 1$ is ensured by defining a modified variable ω' where $\omega_{[1,n-1]} = \omega'$ and $\omega_n = 1 - \sum \omega'$. This modified coordinate set has one less degree of freedom and the sum to 1 constraint is implemented as a restraint $\sum \omega' < 1$.

This also modifies the gradients and Hessian such that

$$\nabla_{\omega'} \mathcal{L}^* = \nabla_{\omega_{[1,n-1]}} \mathcal{L}^* - \nabla_{\omega_n} \mathcal{L}^* \tag{5.52}$$

and

$$\nabla_{\omega'_i} \nabla_{\omega'_j} \mathcal{L}^* = \nabla_{\omega_i} \nabla_{\omega_j} \mathcal{L}^* - \nabla_{\omega_i} \nabla_{\omega_n} \mathcal{L}^* - \nabla_{\omega_n} \nabla_{\omega_j} \mathcal{L}^* + \nabla_{\omega_n} \nabla_{\omega_n} \mathcal{L}^* \tag{5.53}$$

Algorithm 3 Minimum free energy optimization using the bundle method.

```
k ← 0
loop
    while ϵ > 1x10^−3 do
        Update k ← k + 1
        Calculate likelihood and derivative ℓ^k(ρ^k, D) , δ^k(ρ^k, D)
        while ∇_ω ≠ 0 do
            From sFEM get λ(ω)
            Get derivative ∇_ω L and Hessian ∇_ω ∇_ω L
            Run constrained Newton-Raphson to get ω*    ▷ Using MATLAB
        end while
        Update ρ^{k+1} ← arg min L and D^{k+1} ← arg min L
        Adjust η_D to remain near maximum likelihood D
    end while
    Adjust η_F towards critical point
end loop
```

5.11 Analysis of smFRET experiments of Adenylate Kinase

The Bundle method for optimization of the equilibrium probability and diffusion constant is applied to the experimental smFRET data captured when studying the conformational transitions of adynlate-kinase (AK) enzyme using prior methods in [13]. The AK enzyme binds ATP while undergoing a conformational transition of having a LID domain fold over the binding pocket.

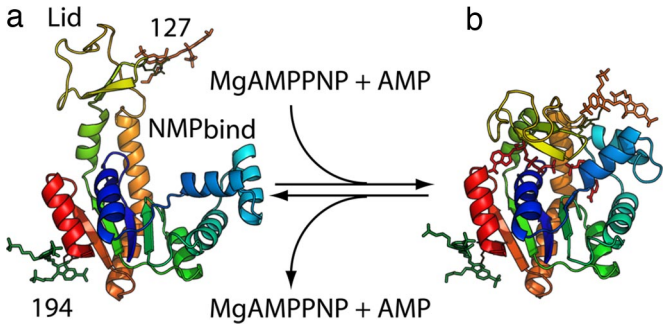

Figure 5.4: Adynalyate Kinase (AK) system which shows the conformational transition of (AK) and the location of the FRET dyes. Figure reproduced here from the original PNAS work by Jeff Hanson and co-workers.[13]

The distance of this LID domain to the core of the enzyme is tracked by the smFRET dyes. Figure 5.5 shows a comparison of these results and elucidates the wealth of additional knowledge that can be gleaned from this new information thermodynamics approach. We see a clear delineation between two states in PMF along with the diffusion coefficient which has converged to $D/R_0^2 = 3.77$ in reduced units or $D = 12x10^{-12}cm^2/s$ in physical units.

A mean first passage time calculation[37] is also run between the states $x_A = 0.8$ and $x_B = 1.0$ to give kinetic rate constants $k_{A\to B} = 18.5s^{-1}$ and $k_{B\to A} = 36.4s^{-1}$ for the conformational transition. The diffusion temperature is adjusted such that the diffusion constant remains in the region of maximum likelihood such that these kinetic estimates are most likely given the smFRET data.

5.12 Conclusion

In the quest to extract models of dynamics from the limited information of convoluted data, we must substitute our lack of data with a desire to produce models that do not reflect features beyond what can be substantiated by the data. Naturally the missing information should be substituted by the maximum entropy profiles. This this approach provides a practical and exponentially converging bundle method to achieve maximum free energy optimization. At the same time, there is now a fundamental understanding of the interplay of the likelihood information from the data and the Bayesian priors of maximum entropy to combine into a full information thermodynamics analysis. With this approach there is a fundamental justification for the derivation of ideal

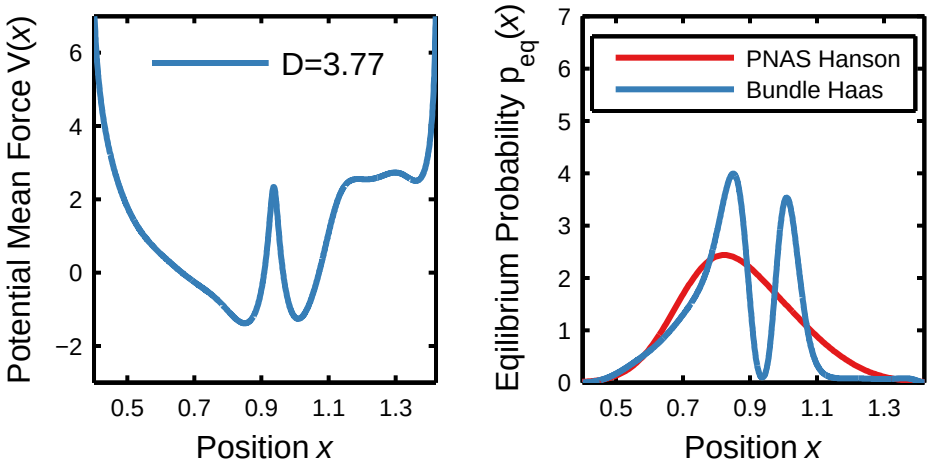

Figure 5.5: The preliminary results for minimum free energy optimization of smFRET data of AK applying the bundle method to 1 of the 415 smFRET trajectories used in the original publication and comparison with the profiles derived from that work.

temperatures and the resultant optimal PMF and diffusion constant from single-molecule FRET. The framework of information thermodynamics is general to all classes of problems where one seeks to generate dynamic models from data.

CHAPTER 6

Conclusion

Many of the developments in this work have been motivated by examples from protein dynamics, however the insights gleaned are universal and fundamental to uncovering properties of stochastic systems and extracting dynamic model data from time series measurements.

For starters, we have discovered a fundamental property of the entropy for the distribution of equilibrium trajectories from Langevin dynamics. This functional will enable a deep understanding of the nature of this very pervasive class of dynamics. Next, the path integral likelihood was developed for the smFRET experiment through a precise accounting of the statistics of the observed experimental process and the hidden stochastic trajectory of the system of interest. We also presented a robust toolkit of practical approaches to inferring theses trajectories and optimizing the paramaterizing force and diffusion through the eigenbasis expansion and the Expectation-Maximization algorithm. The Fisher information metric was derived into the limit of continuum trajectories to define the parameter space of dissipative dynamics and permit the wealth of analysis that is possible with this metric[157]. Finally these efforts synthesized into a comprehensive Bayesian analysis package that permuted exponential convergence of the optimal force profiles and diffusion constants for an ensemble of real experimental trajectories of the Adenylate Kinase enzyme.

Although this thesis was motivated as an alternative approach to molecular simulation, using statistical learning theory to extract model parameters from experiments should be used in harmony with molecular theory and simulation to produce the deepest understanding and insight about a problem. Our proposed approach to protein engineering is necessarily synergistic as depicted in Figure 6.1:

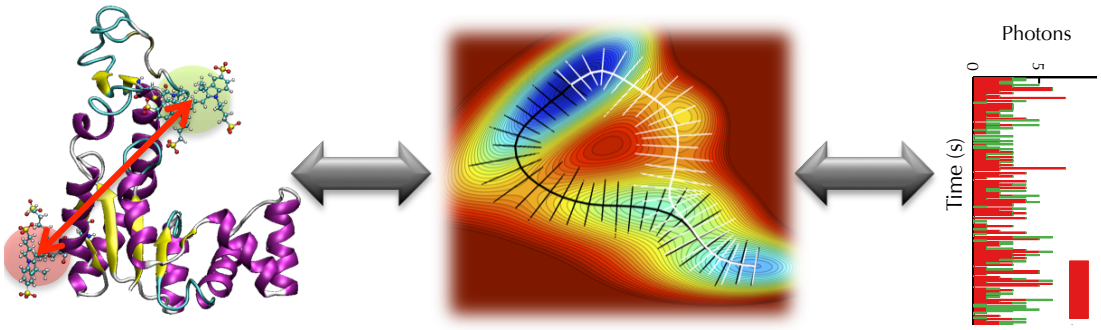

Figure 6.1: Cartoon of the integrated approach between molecular simulation and statistical learning theory of single molecule experiments to protein engineering.

- Target systems are studied with single molecule experiments to construct a dynamic model following the procedure from this work
- Free energy barriers are identified as targets for mutagenesis[30, 158]
- The coarse grained dynamic model is then reconstructed into it's all atom analog [132, 159] to identify the responsible amnio acid residues or secondary structures
- The relevance of the coordinate under study can be tested. Lack of free energy barriers and low diffusion constants are indicative that it is not the most relevant reactive coordinate [91, 160]
- Multi-coordinate smFRET[90] can also drive the search for the atomic level description of the reactive mechanism[161, 162]
- For confirmation, candidate alterations to the protein sequence can be simulated for their effects and these predictions verified by the single molecule data on the real system[163]

Ultimately this works makes possible a new level of detailed analysis of single molecule experiments starting from the very general and fundamental description of Langevin dynamics. We only hope in anticipation of the future advances and discoveries in biophysics that will be enabled by this work.

APPENDIX A

Elements of the Trajectory Entropy in Continuous Stochastic Processes at Equilibrium

A.1 Numerical Integration of Fokker-Planck Equation for $p(x_t|x_0)$

The conditional probabilities are calculated by integrating the Fokker-Planck equation (FPE) that specifies the time evolution of $p(x_t|x_0)$ as the divergence of a flux $J(x_t)$

$$\frac{\partial p(x_t|x_0)}{\partial t} = \nabla \cdot J(x) = \frac{\partial}{\partial x_t}\left(D\frac{\partial p(x_t|x_0)}{\partial x_t} - \frac{DF(x_t)}{k_B T}p(x_t|x_0)\right) \tag{A.1}$$

with initial condition $p(x_t|x_0)|_{t=0} = \delta(x_t - x_0)$ and the zero-flux boundary conditions $J(x_t = L) = J(x_t = -L) = 0$. To numerically integrate these equations we followed the procedure outlined in Gardiner [37] and performed an eigen-decomposition of the FP operator. The eigenvectors $\psi_i(x)$ for eigenvalue λ_i are found using a spectral finite element method [77]. The time dependent probability is then given by

$$p(x_{\Delta t}|x_0) = \sum_i \psi_i(x_{\Delta t})\frac{\psi_i(x_0)}{p_{\text{eq}}(x_0)}\exp\left(-\lambda_i \Delta t\right) \tag{A.2}$$

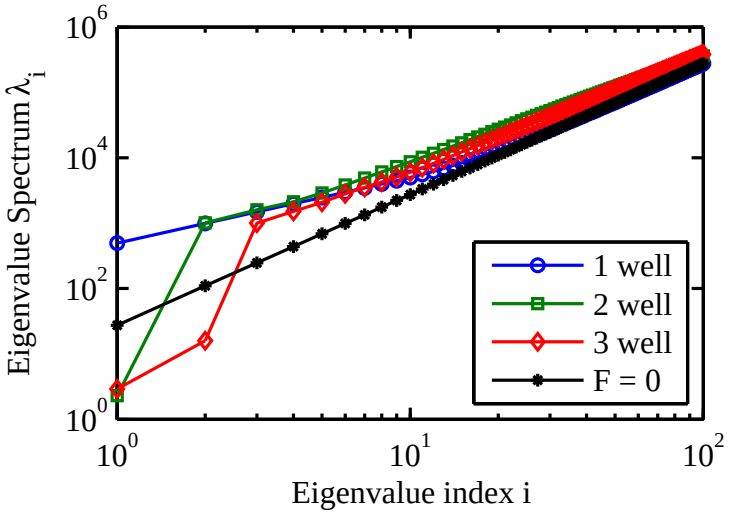

Figure A.1: Power spectrum of the eigenvalues of the statistical time evolution Hamiltonian for the 3 model systems with $D = 1$.

A.2 Calculation of Trajectory Partition Function $\mathcal{Z}$

The partition function for the trajectory has no force dependence because when evaluating the functional derivative

$$\frac{\delta \mathcal{Z}}{\delta F(y)} = \frac{\delta}{\delta F(y)} \int \mathcal{D}X(t) e^{-E_{OM}[X(t)]} \tag{A.3}$$

one can move the linear operator of derivative through the path integral,

$$\frac{\delta \mathcal{Z}}{\delta F(y)} = -\int \mathcal{D}X(t) \frac{\delta E_{OM}[X(t)]}{\delta F(y)} e^{-E_{OM}[X(t)]} \tag{A.4}$$

where the derivative of the OM action is given by the Euler-Lagrange equation

$$\frac{\delta E_{OM}[X(t)]}{\delta F(y)} = \frac{D}{2}\left(F(x_t)\delta(x_t - y) + \frac{d\delta(x_t - y)}{dx_t}\right) + \int^{x_0} dx_0' \, \delta(x_o' - y) - \int^{x_{t_{obs}}} dx_{t_{obs}}' \, \delta(x_{t_{obs}}' - y) \tag{A.5}$$

The path integral of the derivative times the exponent of the OM action is essentially the trajectory expectation up to the scalar

$$\frac{\delta \mathcal{Z}}{\delta F(y)} = -\mathcal{Z} \int \mathcal{D}X(t) \frac{\delta E_{OM}[X(t)]}{\delta F(y)} \mathcal{P}(X(t)) = -\mathcal{Z} \left\langle \frac{\delta E_{OM}[X(t)]}{\delta F(y)} \right\rangle_{X(t)} \tag{A.6}$$

Because derivative contains only terms which are local in time the $\int^x dx' \delta(x' - y)$ terms are equivalent and thus cancels each other. The resulting trajectory expectation thus reduces to just an expectation over the equilibrium distribution of position

$$\frac{\delta \mathcal{Z}}{\delta F(y)} = -\mathcal{Z} \int dx \frac{D}{2} \left(F(x)\delta(x-y) + \frac{d\delta(x-y)}{dx} \right) p_{\text{eq}}(x) \tag{A.7}$$

We apply integration by parts on the second term which contains the derivative of the delta function, noting that $dp_{\text{eq}}(x)/dx = F(x)p_{\text{eq}}(x)$ for the domain $x \in \mathbb{R}$ and normalized equilibrium probability $p_{\text{eq}}|_{-\infty,\infty} = 0$, the boundary term $F(x)p_{\text{eq}}(x)|_{-L,L} = 0$. Entropy formula can be extended to include $F'(x)p_{\text{eq}}(x)|_{-L,L} \neq 0$ if necessary. We finally arrive at the cancellation

$$\frac{\delta \mathcal{Z}}{\delta F(y)} = -\mathcal{Z} \frac{D}{2} \left(F(y) - F(y) \right) p_{\text{eq}}(x) = 0 \tag{A.8}$$

Since the partition function is not functionally dependent upon force $\mathcal{Z} \neq \mathcal{Z}[F(x)]$, the partition function at any force must be equal to the partition function at zero force which is the case of a purely diffusive process when the OM-action is simply a convolution of many Gaussian into a multivariate Gaussian Process. That partition function is know and is equal to

$$\mathcal{Z}(D) = (4\pi D \Delta t)^{(t_{\text{obs}}/2\Delta t)} \tag{A.9}$$

where $t_{\text{obs}}/\Delta t$ is essentially the number of steps in the Gaussian Process and $\sqrt{4\pi D \Delta t}$ is the partition function for each step with variance $2D\Delta t$.[47, 48]

A.3 Numerical Studies of Entropy on Model Systems

Validation plot of analytical trajectory entropy versus numerical 2-D integral at different diffusion constants for all 3 models. Agreement is reached in the continuum limit when $\Delta t \to 0$. For system in which $D = D_{\text{ref}}$ the theoretical curves approach from below to reach a finite limit. For system when $D < D_{\text{ref}}$ the curves approach from below but reach asymptotic convergence of the $1/\Delta t$ scaling. For systems when $D > D_{\text{ref}}$ the curves first overshoot and then converge from above. For system when $D \ll D_{\text{ref}}$ the reference diffusion term dominates.

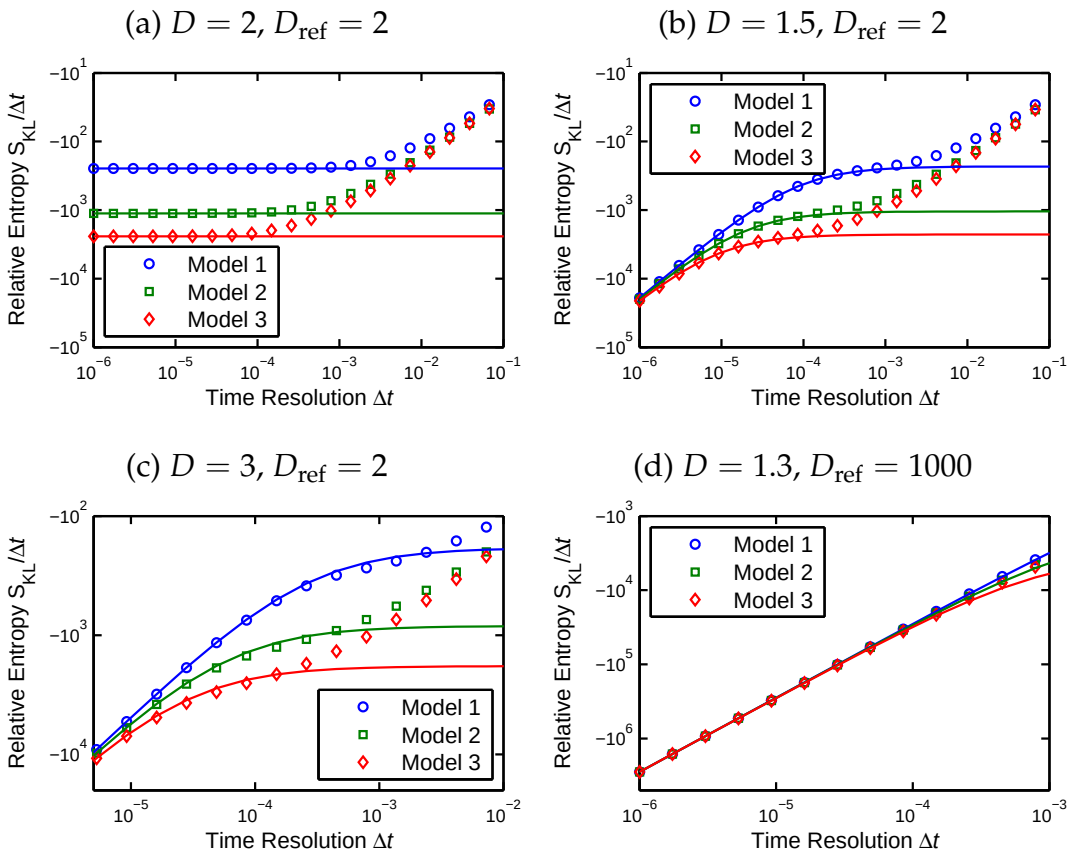

Figure A.2: Comparison between numerical calculation of $S_{\text{KL}}(\Delta t)$ represented by $(\circ, \square, \Diamond)$ and the analytic formula $S_{F_{\text{ref}} \to F(x)} + S_{D_{\text{ref}} \to D}$ show as curves for a few combinations of D and D_{ref}. Scale of axis is adjusted for each figure to highlight nature of convergence in continuum limit.

In the case where $D \ll D_{\text{ref}}$ we also plot the limiting trajectory entropy behavior where

$$\mathcal{S}[F(x), D] = S_{\text{eq}} - \frac{t_{\text{obs}}}{4}\left\langle DF^2(x)\right\rangle_{\text{eq}} + \lim_{\Delta t\to 0}\frac{t_{\text{obs}}}{2\Delta t}\ln D \tag{A.10}$$

in comparison to the numerical calculations with the diverging offset $(1 - \ln(D_{ref}))/(2\Delta t)$ removed.

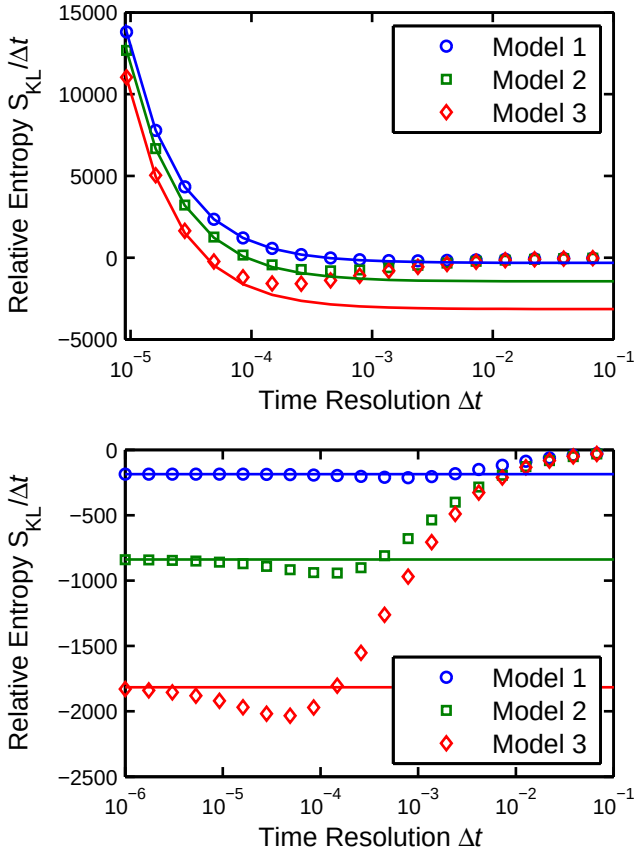

Figure A.3: Limiting behavior of trajectory entropy when $D \ll D_{\text{ref}}$. (top) Plotting against the numerical calculation $(S_{\text{KL}}(\Delta t) - 1/2 + \ln(D_{\text{ref}})/2)/\Delta t$ for $D = 1.3$, $D_{\text{ref}} = 1000$. (bottom) Comparison between numerical calculation of $S_{\text{KL}}(\Delta t) - S_{D_{\text{ref}}\to D}$ represented by $(\circ, \square, \Diamond)$ and the analytic formula $S_{F_{\text{ref}}\to F(x)}$ show as curves for $D = 3$ and $D_{\text{ref}} = 2$.

For a final numerical test, we can check just the evaluation of $\mathcal{S}_{F_{\text{ref}}\to F(x)}$ for instances in which $D \neq D_{\text{ref}}$ by removing the diffusion contribution $\mathcal{S}_{D_{\text{ref}}\to D}$ from both the theoretical and numerical calculations.

A.4 Functional Form of Model Potentials

The models were constructed from low order polynomials with coefficients manually adjusted in order to create three systems that had equivalent equilibrium entropy $S_{\text{eq}} = \int dx\ p_{\text{eq}}(x) \ln p_{\text{eq}}(x) = 1.683$. The barriers between local minimum are approximately $5k_{\text{B}}T$ and with a diffusion constant $D = 1$, the first order relaxation rate is approximately $\tau_{\text{rxn}} = 1\text{m}s$.

$$V(x) = \begin{cases} \underbrace{\frac{494.29836}{2}(x-1)^2}_{\text{Model 1}} \\ \underbrace{4.6139373\left(\frac{15000}{4}(x-1)^4 - \frac{265}{2}(x-1)^2\right)}_{\text{Model 2}} \\ \underbrace{\frac{1}{15.47}\left[\left(\frac{x-1}{13.9}\right)^6 - 15\left(\frac{x-1}{13.9}\right)^4 + 53\left(\frac{x-1}{13.9}\right)^2 + 2\left(\frac{x-1}{13.9}\right) - 15\right]}_{\text{Model 3}} \end{cases} \tag{A.11}$$

A.5 Derivation of $\lim_{\Delta t \to 0} \frac{S_{\text{KL}}(\Delta t)}{\Delta t} = -\frac{D}{4}\left\langle F^2(x)\right\rangle_{\text{eq}}$

The continuous stochastic trajectory $X(t);\ t \in [0, t_{\text{obs}}]$ of duration t_{obs} can be realized at specific instances in time separated by a time resolution Δt to create a segmented trajectory $\boldsymbol{X_t} = [X(0), X(\Delta t), X(2\Delta t), \ldots, X(t_{\text{obs}})]$. This composite vector exists in a space of dimensionality $N = T/\Delta t$ with coordinates $\{x_\tau | \tau = 0, 1, 2, \ldots, N\}$. The proposed formulation will utilize this time segmentation such that the problem is recast to finding the KL-Divergence of a multidimensional vector space with integrations taken at every time point. The the $\lim_{\Delta t \to 0}$ will be performed on the final results to recover the information metric in the continuum limit.

$$\mathcal{S} = \lim_{\Delta t \to 0} \int dx_0 \int dx_1 \cdots \int dx_T\ P(\mathbf{X_T}) \ln \frac{P(\mathbf{X_T})}{Q(\mathbf{X_T})} \tag{A.12}$$

The logarithm in the KL-divergence formula allows us to exploit this memoryless time factorization to reduce the effectively infinite product to a summation of logarithms of functions with two dependent coordinates $\ln P(\boldsymbol{X}_T) =$

$\ln p(x_0) + \sum_{\tau=0}^{N} \ln p(x_{\tau+1}|x_\tau)$. Installing the factorized distributions into Equation A.12 gives (Note the separate indexing τ' for the product and τ for the sum)

$$\mathcal{S} = \int \prod_{\tau'=0}^{N} dx_{\tau'}\ p(x_0)p(x_{\tau'+1}|x_{\tau'}) \left(\ln \frac{p(x_0)}{q(x_0)} + \sum_{\tau=0}^{N} \ln \frac{p(x_{\tau+1}|x_\tau)}{q(x_{\tau+1}|x_\tau)} \right) \tag{A.13}$$

By reorganization this equation to bring the summation outside the integral and explicitly grouping the integral operations with and without logarithm in the integrand gives the following

$$\begin{aligned} \mathcal{S} = \sum_{\tau=0}^{N} \Bigg(& \int \prod_{\substack{\tau' \neq \tau \\ \tau' \neq \tau+1}}^{N} dx_{\tau'}\ p(x_0)p(x_{\tau'+1}|x_{\tau'}) \Bigg) \\ & \int dx_{\tau+1}dx_\tau\ p(x_{\tau+1}|x_\tau) \ln \frac{p(x_{\tau+1}|x_\tau)}{q(x_{\tau+1}|x_\tau)} \\ & + \int dx_0 \left(\int \prod_{\tau'}^{N} dx_{\tau'}\ p(x_0)p(x_{\tau'+1}|x_{\tau'}) \right) \ln \frac{p(x_0)}{q(x_0)} \end{aligned} \tag{A.14}$$

For the products grouped by the parenthesis not containing the coordinates in the logarithms, the integration over the normalized probabilities leave only the marginalized probability $\int \prod_{\tau' \neq \tau\tau+1}^{N} dx_{\tau'}\ p(x_0)p(x_{\tau'+1}|x_{\tau'}) \to p(x_\tau)$. The remaining non-trivial integrals are formally invariant to the actual choice of index τ leaving N equivalent terms of 2-D integration to determine the trajectory entropy.

$$\mathcal{S} = N \int dx_0 dx_\tau p(x_\tau, x_0) \ln \frac{p(x_\tau|x_0)}{q(x_\tau|x_0)} + \int dx_0 p(x_0) \log \frac{p(x_0)}{q(x_0)} \tag{A.15}$$

For the Langevin dynamics considered in this work, dynamic parameters are time invariant in the equation of motion. The marginal probability of finding the system at a given position at a given time is equal to the equilibrium probability such that $p(x_0) = p_{\mathrm{eq}}(x)$ and the second term from Equation A.15 is equal to equilibrium entropy S_{eq}.

$$\mathcal{S} = \lim_{\Delta t \to 0} \frac{t_{\mathrm{obs}}}{\Delta t} \int dx_0 dx_\tau p(x_\tau, x_0) \ln \frac{p(x_\tau|x_0)}{q(x_\tau|x_0)} + S_{\mathrm{eq}} \tag{A.16}$$

The 1-D contribution from the equilibrium distributions itself can be separated from the contribution from purely 2-D dynamic integral. This 2-D integral is $S_{\mathrm{KL}}(\Delta t)$ defined in the main text.

In the short time limit at $\Delta t \to 0$ the two dynamic processes become equivalent in that at no time separation, one simply recovers the initial delta function condition for the Fokker-Planck equation $p(x_t|x_0) = q(x_t|x_0) = \delta(x_t - x_0)$ and consequently the remaining expectation in equation A.15 is taken over $\ln(\delta(x_t - x_0)/\delta(x_t - x_0)) = \ln 1 = 0$. What remains to be developed however, is how this limit reacts when calculating the full path KL-divergence when there is a competing factor of Δt in the denominator.

$$\lim_{\Delta t \to 0} \frac{S_{\mathrm{KL}}(\Delta t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{\int dx_0 dx_\tau p(x_\tau, x_0) \ln \frac{p(x_\tau|x_0)}{q(x_\tau|x_0)}}{\Delta t} \tag{A.17}$$

Because both the top and bottom of the fraction go to zero, we apply the L'Hopital's rule to realize the limit as the time derivative of the expectation evaluated at $t = 0^+$.

$$\lim_{\Delta t \to 0} \frac{S_{\mathrm{KL}}(\Delta t)}{\Delta t} = \frac{\partial}{\partial \Delta t} \int dx_0 dx_\tau p(x_\tau, x_0) \ln \frac{p(x_\tau|x_0)}{q(x_\tau|x_0)}\bigg|_{\Delta t = 0} \tag{A.18}$$

By inserting the definition of the KL-divergence, and taking the time derivative, this asymptotic gives

$$\lim_{\Delta t \to 0} \frac{S_{\mathrm{KL}}(\Delta t)}{\Delta t} = \mathbb{E}_{p(x_0)} \left[\frac{\partial}{\partial t} \int dx_t p(x_t|x_0) \ln \frac{p(x_t|x_0)}{q(x_t|x_0)}\bigg|_{t=0} \right] \tag{A.19}$$

$$= \mathbb{E}_{p(x_0)} \left[\int dx_t \frac{\partial p(x_t|x_0)}{\partial t} \ln \frac{p(x_t|x_0)}{q(x_t|x_0)} + \frac{\partial p(x_t|x_0)}{\partial t} - \frac{p(x_t|x_0)}{q(x_t|x_0)} \frac{\partial q(x_t|x_0)}{\partial t}\bigg|_{t=0} \right] \tag{A.20}$$

At the initial evaluation condition $t = 0^+$, the two conditional probabilities are equivalent $p(x_t|x_0) = q(x_t|x_0) = \delta(x_t - x_0)$. This makes the quotient of these two probabilities equal to unity and leaves the simplified equation

$$\lim_{\Delta t \to 0} \frac{S_{\mathrm{KL}}(\Delta t)}{\Delta t} = \mathbb{E}_{p(x_0)} \left[\int dx_t \, \frac{\partial p(x_t|x_0)}{\partial t} - \frac{\partial q(x_t|x_0)}{\partial t}\bigg|_{t=0} \right] \tag{A.21}$$

The time derivatives in Equation A.21 are given by the Fokker-Planck partial differential equation (Equation A.1). To aid in the calculation, we actually multiply each time derivative by $p_{\mathrm{eq}}(x_0)/p_{\mathrm{eq}}(x_t)$ (equaling to one in the continuum limit) to give the symmeterized Fokker-planck equation for $\rho(x_t, x_0) = p(x_t|x_0)p_{\mathrm{eq}}(x_0)/p_{\mathrm{eq}}(x_t)$

$$\frac{\partial \rho(x_t, x_0)}{\partial t} = D\nabla^2 \rho - \left(\frac{F^2(x)D}{4} + \frac{F'(x)D}{2} \right) \tag{A.22}$$

For the case in which the diffusion coefficient is equivalent from the reference process and the system of interest, $D = D_{\text{ref}}$, the second derivative terms cancel and the reference process $F_{ref} = 0$ leaving

$$\lim_{\Delta t \to 0} \frac{S_{\text{KL}}(\Delta t)}{\Delta t} = -\mathbb{E}_{p_{\text{eq}}(x_0)} \left[\int dx_t \left(\frac{F'(x_t)D}{2} + \frac{F^2(x_t)D}{4} \right) \delta(x_t - x_0) \right]$$

with the delta function initial condition inserted. Doing the x_t integral trivially eliminates the delta function to make $x_t = x_0$ and consequently

$$\lim_{\Delta t \to 0} \frac{S_{\text{KL}}(\Delta t)}{\Delta t} = -\mathbb{E}_{p_{\text{eq}}(x_0)} \left[\frac{F'(x_0)D}{2} + \frac{F^2(x_0)D}{4} \right]$$

Doing the expectation by integrating against equilibrium probability requires the integration

$$\lim_{\Delta t \to 0} \frac{S_{\text{KL}}(\Delta t)}{\Delta t} = -\int dx_0 p_{\text{eq}}(x_0) \left(\frac{F'(x_0)D}{2} + \frac{F^2(x_0)D}{4} \right)$$

where integrating the first term by parts and using $dp_{\text{eq}}(x)/dx = F(x)(p_{\text{eq}}(x))$ gives the same result as the path partition function analysis of the OM-action

$$\lim_{\Delta t \to 0} \frac{S_{\text{KL}}(\Delta t)}{\Delta t} = \int dx_0 p_{\text{eq}}(x_0) \left(\frac{F^2(x_0)D}{2} - \frac{F^2(x_0)D}{4} \right) \tag{A.23}$$

$$\lim_{\Delta t \to 0} \frac{S_{\text{KL}}(\Delta t)}{\Delta t} = -\frac{D}{4} \left\langle F^2(x) \right\rangle_{\text{eq}} \tag{A.24}$$

A.6 KL-Divergence for Finite-State Continuous Time Markov Process

A continuous-time Markov Process involves jumps between a finite number of N discrete states is parameterized by the transition rate matrix $\mathbf{P}$ with elements $p_{i,j}$ that give the probability of making a transition from state i to sate j per unit of time Δt such that $P\left(x_{t+\Delta t} = j | x_t = i\right) = p_{i,j}\Delta t + \mathcal{O}(\Delta t^2)$ under the condition that $p_{i,i} = -\sum_{j \neq i} p_{i,j}$. The time evolution of the system probability is given by the differential equation $\dot{p} = p\mathbf{P}$ which has the formal solution in terms of a matrix exponential $p_t = p_0 e^{\mathbf{P}t}$ with initial distribution over states p_0.

The KL-divergence, also know as relative entropy, for an continuous time Markov process $E_{p(x_0)} \left\{ D_{KL}(p(x_{\Delta t}|x_0)||q(x_{\Delta t}|x_0)) \right\}$ can be computed by inserting the definition of the time propagator probability to yield

$$D_{KL} = \lim_{\Delta t \to 0} \frac{T}{\Delta t} \sum_{i,j=1}^{N} \pi_i \left(e^{\mathbf{P}\Delta t} \right)_{i,j} \log \frac{\left(e^{\mathbf{P}\Delta t} \right)_{i,j}}{\left(e^{\mathbf{Q}\Delta t} \right)_{i,j}} \tag{A.25}$$

The ratios of exponentials above won't become a simple exponential of difference because the matrix operations don't normally commute, $[\mathbf{P}, \mathbf{Q}] \neq 0$. To perform the limit we take the short time expansion of the operator exponentials $e^{\mathbf{P}\Delta t} \simeq \mathbf{I} + \mathbf{P}\Delta t + \mathcal{O}(\Delta t^2)$ where $\mathbf{I} = \delta_{i,j}$ is the identity operator. Plugging the short time expansion into equation A.25 gives two groupings of terms for the summand for the diagonal and off-diagonal elements from the transition matrix

$$D_{KL} = \lim_{\Delta t \to 0} \frac{T}{\Delta t} \left\{ \sum_{i=j}^{N} \pi_i \left(1 + p_{i,j}\Delta t\right) \log \frac{\left(1 + p_{i,j}\Delta t\right)}{\left(1 + q_{i,j}\Delta t\right)} + \sum_{i \neq j}^{N} \pi_i p_{i,j} \Delta t \log \frac{p_{i,j}\Delta t}{q_{i,j}\Delta t} \right\} \tag{A.26}$$

where the higher order terms in Δt are omitted because we will eventually be taking the short time limit. Making use of the approximation $\log(1 + x) \simeq x$, we simplify to

$$D_{KL} = \lim_{\Delta t \to 0} \frac{T}{\Delta t} \left\{ \sum_{i=j}^{N} \pi_i \left(1 + p_{i,j}\Delta t\right) \left(p_{i,j}\Delta t - q_{i,j}\Delta t\right) + \sum_{i \neq j}^{N} \pi_i p_{i,j} \Delta t \log \frac{p_{i,j}}{q_{i,j}} \right\} \tag{A.27}$$

we are then allowed to take explicit cancelation of the Δt terms and preform the limit to give

$$D_{KL} = T \left\{ \sum_{i}^{N} \pi_i \left(p_{i,i} - q_{i,i}\right) + \sum_{i \neq j}^{N} \pi_i p_{i,j} \log \frac{p_{i,j}}{q_{i,j}} \right\} \tag{A.28}$$

Recalling that for conservation of probability there exists the restriction $p_{i,i} = -\sum_{j \neq i} p_{i,j}$, the first sum can be rewritten as $\sum_{i \neq j}^{N} \pi_i \left(q_{i,j} - p_{i,j}\right)$ given the modified formula

$$D_{KL} = T \sum_{i \neq j}^{N} \pi_i \left\{ \left(q_{i,j} - p_{i,j}\right) + p_{i,j} \log \frac{p_{i,j}}{q_{i,j}} \right\} \tag{A.29}$$

Again, as expected, for equivalent processes where $q_{i,j} = p_{i,j}$ we get $D_{KL} = 0$. Unlike the previous result for the general ODLD process, this divergence measure doesn't diverge at a rate $\lim_{\Delta t \to 0} \frac{T}{\Delta t}$. However, when attempting to relate the matrix elements $p_{i,j}$, we see that the matrix elements themselves are not constant with time-step and will similarly diverge.

APPENDIX B

Path Integral Statistical Learning Theory: Extracting Free Energy and Diffusion from single-molecule FRET

B.1 Analysis of Estimator Variance from Fisher Information Metric

We can calculate the Fisher Information Metric[114, 164] to get a sense of the order of accuracy of our EM method and approximate an error bound on the final results. The fisher information metric is a 2-D scalar field over the functional values of force and the diffusion constant evaluated at a given set of parameters θ^*

$$J_{\theta^*}[\theta(x), \theta(x')] = \mathbb{E}_Y \left[\left(\frac{\delta \ell}{\delta \theta(x)} \right) \left(\frac{\delta \ell}{\delta \theta(x')} \right) \Big| \theta^* \right] \tag{B.1}$$

where the expectation is taken over all possible smFRET experiments.

By inverting this metric, the Cramer-Rao bound[125] estimates the minimum variance of our maximum-likelihood optimized parameters

$$\langle \Delta\theta(x) \Delta\theta(x') \rangle \geq \left(J_{\theta^*}[\theta(x), \theta(x')] \right)^{-1} \tag{B.2}$$

It's probably most convenient to think of the Fisher information as a matrix where the rows/columns correspond to the Force values along the x axis and

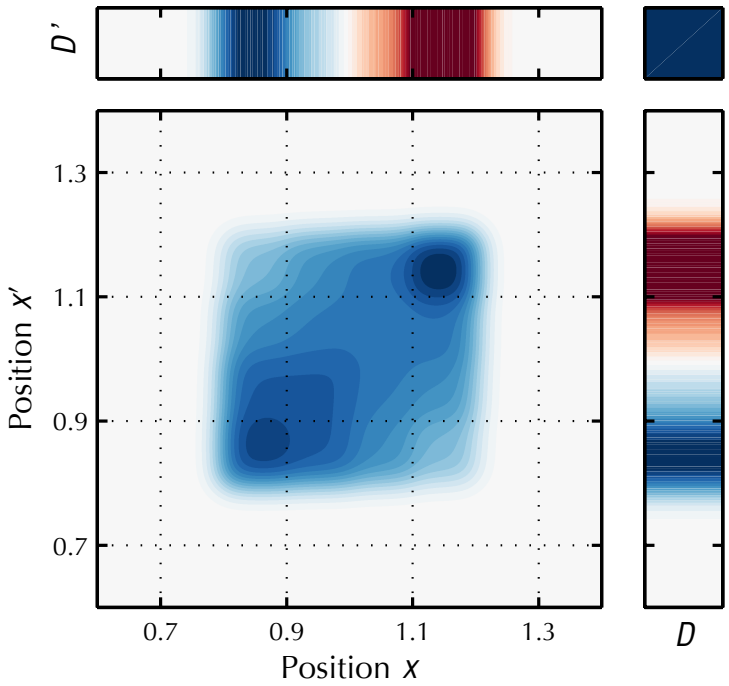

Figure B.1: The Fisher information for the smFRET likelihood constructed from evaluating the derivative at $p^* = p_{ref}$ from 48 simulated experiments to sample the distribution of experiments at x1 resolution. The coordinates for the Fisher information metric are the force function along the space coordinate $F(x)$ appended by the diffusion constant D. This four quadrant matrix is plotted (lower-left) $J_{\theta^{ref}}[F(x), F(x')]$ (lower-right) $J_{\theta^{ref}}[D, F(x')]$ (upper-left) $J_{\theta^{ref}}[F(x), D']$ (upper-right) $J_{\theta^{ref}}[D, D']$

the last entry is the diffusion constant. We calculate the fisher information matrix for our model system by evaluating the derivative for each of the 12 experimental runs at the correct $\theta^* = \theta_{ref}$ and average the results to approximate the expectation over experiments.

The strong correlation in force occurs between the points of $x = 0.9, 1.1$ which are the free energy barriers. These points also correlate strongly with the diffusion coefficient which governs relaxation rates. The largest amplitude eigenvector of the covariance matrix (which is not singular) reflects the physics that one can achieve the same relaxation rate between the metastable wells by having a small diffusion constant with small force (free energy barrier) or high diffusion and high free energy barrier at transition state.

When attempting to invert this matrix however, we note that it is very

singular due to the low probability in the boundaries of the domain. Therefore we could expect a pure maximum likelihood EM-algorithm to have nearly unbounded oscillation and variance of the force in the wings. Once we do add in the average of force squared, the likelihood is supplemented by the regularizer which has the derivative

$$\frac{\delta\langle F^2(x)\rangle}{\delta F(x)} = \int dx\ F^2(x)\frac{\exp(-\int^{x'} dx' F(x'))}{\int dx' \exp(\int^{x'} dx'' F(x''))} \tag{B.3}$$

$$= 2F(x)p_{\text{eq}}(x) + \int^{x} dx\ p_{\text{eq}}(x)\left(F^2(x) - \langle F^2(x)\rangle\right) \tag{B.4}$$

When adding in the regularization, the term in fisher information goes from likelihood $\ell \to \ell - \eta_F D\langle F^2(x)\rangle$, these added terms make the fisher information take on information in the wings remove the hard singularity to the solution. Although the addition of the maximum entropy term biases our EM algorithm to produce solutions with less force and less diffusion, there is less variance and we can be confident that the actual system contains at least as much complexity as is afforded by the optimization. This is the bias-variance tradeoff prevalent in statistical learning algorithms that fit a parameter space which could contain more information and precision than the actual raw statistically noisy data.

The choice of parameter η_F actually only requires making it large enough such that we can converge within a feasible amount of computation time. Fisher information says that the likelihood is actually degenerate for all of these singular fluctuations so any finite perturbation by the regularizer will produce a low-variance biases answer. However the rate of convergence suffers when η_F is too low.

B.2 Reference Potential of Mean Force

The potential of mean force $V(x)/k_{\text{B}}T$ is constructed using the piecewise cubic Hermite interpolation polynomial `pchip()` in MATLAB with the following points

Table B.1: Interpolation points for $V(x)$

Position	0.5	0.8	0.9	1	1.1	1.2	1.5
$V(x)$	60	0	4	$\ln(4)$	5	$\ln(2)$	60

B.3 Proof of $\Delta S \geq 0$ for Expectation-Maximization algorithm

For the EM algorithm to hold, the property that the hidden state "entropy" must be strictly greater than zero must be confirmed. From the EM development, we are left to show that $\Delta S(\theta) \geq 0 \; \forall \; \theta$ where

$$\Delta S(\theta) = \sum_{\substack{\{i(\tau)\} \\ \{j(\tau)\}}} \frac{\prod_\tau \langle | \rangle^k}{\mathcal{L}^k} \left(\ln \frac{\prod_\tau \langle | \rangle^k}{\mathcal{L}^k} - \ln \frac{\prod_\tau \langle | \rangle(\theta)}{\mathcal{L}(\theta)} \right) \tag{B.5}$$

$$= - \sum_{\substack{\{i(\tau)\} \\ \{j(\tau)\}}} \frac{\prod_\tau \langle | \rangle^k}{\mathcal{L}^k} \left(\ln \frac{\mathcal{L}^k}{\prod_\tau \langle | \rangle^k} \frac{\prod_\tau \langle | \rangle(\theta)}{\mathcal{L}(\theta)} \right) \tag{B.6}$$

and k is the index of the current parameters and the terms without a k superscript are functionally dependent on new parameters θ. To test the inequality, the relationship for logarithms that $\ln(x) \leq x - 1$ is used to remove the logarithm in the above equation and give

$$\Delta S(\theta) \geq \sum_{\substack{\{i(\tau)\} \\ \{j(\tau)\}}} \frac{\prod_\tau \langle | \rangle(\theta)}{\mathcal{L}(\theta)} - 1 = 0 \tag{B.7}$$

because the summation of the numerator is just the definition of the denominator.

B.4 Maximum Information Method

The Maximum Information Method [60] estimates the current position of the system is constructed by counting photons for a short time window t_w of approximately 1ms. The Poisson distribution for the number of acceptor photons N_a and donor photons N_d within the time window is maximized for the time-averaged position

$$\left(\overline{x^6} \right)_w^* = \left[\frac{B_d}{B_a} \frac{N_a - I_a^\beta N_d / B_d}{N_d - I_d^\beta N_a / B_a} \right] \tag{B.8}$$

The exact time window duration is set to give a constant standard error in the estimate $\Delta x = 0.1$.

The collection of these estimates $\{x(j)\}$ are weighted by the time window $t_w(j)$ when performing the histogram reconstruction so that time average matches the ensemble average.

$$h_i = \sum_j t_w(j) \left[x_i < x(j) < x_{i+1}\right] \tag{B.9}$$

B.5 Eigenvector Numerical Methods

The spectral element basis functions are constructed from a set of $N_E = 128$ non-overlapping spectral elements of polynomial order $N_L = 7$. The element are constructed of Lagrange interpolation polynomials $L_k(s)$ where $k \in [0 \dots N_L]$.

$$L_k(s) = \prod_{l \neq k} \frac{s - x_m}{s_l - s_m} \tag{B.10}$$

where s_k are the Legendre-Gauss-Lobatto points of numerical quadrature defined by the zeros of the derivative of the Legendre polynomial $P'_{N_L}(s)$ (see Figure B.2)

$$s_0 = -1 \quad ; \quad P'_{N_L}(s_{[1,N_L-1]}) = 0 \quad ; \quad s_{N_L} = 1 \tag{B.11}$$

which after substation gives the equivalent formula

$$L_k(s) = \frac{-1}{N_L(N_L+1)P_{N_L}(s_k)} \frac{(1-s^2)P'_{N_L}(s)}{s - s_k} \tag{B.12}$$

where $P_{N_L}(s)$ is the Legendre polynomial of order N_L and it derivative with respect to affine position $P'_{N_L}(s)$.

An affine mapping function $M^n : x \in [-x_L, x_R] \rightarrow s \in [-1, 1]^n$ takes real space positions and maps them to coordinates with the element s and indicates which element corresponds to this position n. Given the expansion coefficients $\{c_k^n\}$ for element n and polynomial k, and summing over elements and polynomials, functions are approximated by the interpolant

$$f(x) \simeq \sum_{n=1}^{N_E} \sum_{k=0}^{N_L} c_k^n L_k\left(M^n(x)\right) \tag{B.13}$$

For the set of node points over the entire domain $\{x_k^n\}$, the interpolant evaluated at those points are the Kronecker deltas $L_l\left(M^m(x_k^n)\right) = 1\delta_{n,m}\delta_{l,k}$ so that the expansion coefficients are simply given by

$$c_k^n = f(x_k^n) \tag{B.14}$$

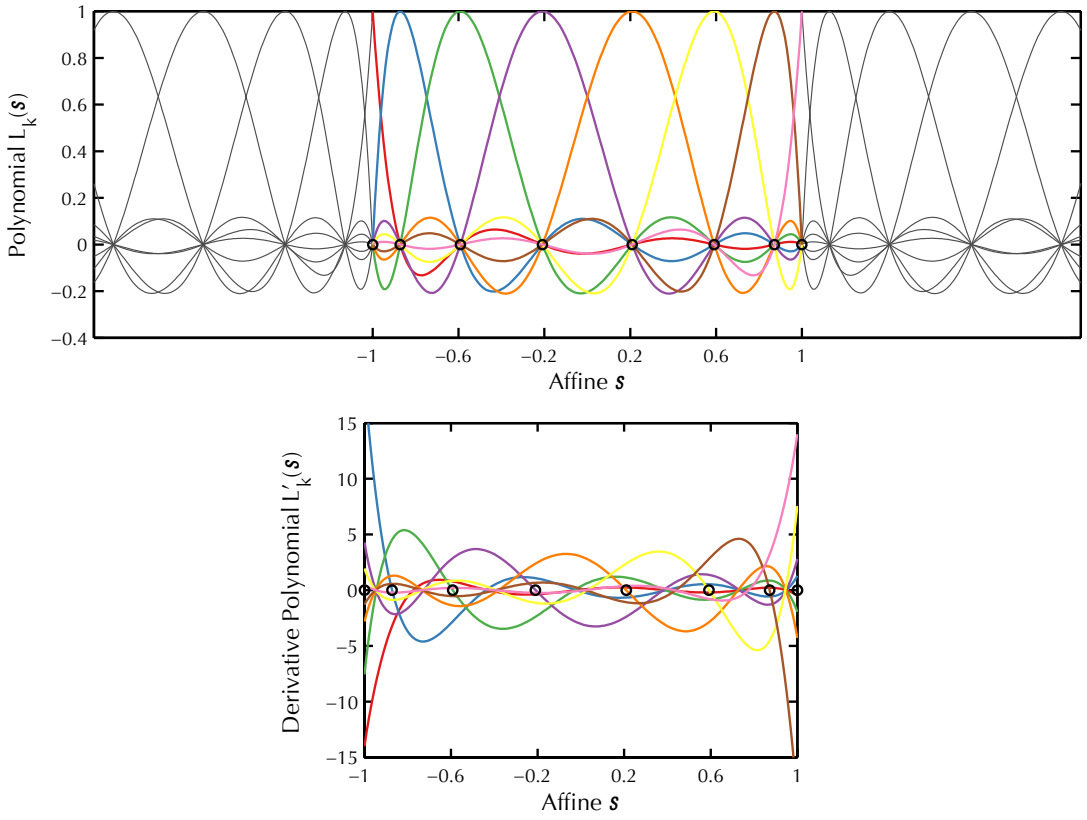

Figure B.2: (top) Spectral element constructed of Legendre-Gauss-Lobatto-Lagrange interpolation polynomials $L_i(s)$ for $N_L = 8$ and (O) the interpolation points s_i. (bottom) Derivative of the spectral element $L_i'(s)$.

The solution of the eigenvector problem are necessarily continuous functions of the $\mathcal{C}^0$ manifold so to ensure continuity, $c^n_{N_L} = c^{n+1}_0 \; \forall \; n \in [1 \ldots N_E]$.

Integration over the domain of an element is performed by Gauss-Lobatto integration (which is exact for polynomials up to order $2N_L - 1$)

$$\int_{-1}^{1} ds \; f(s) = \sum_k f(s_k) w_k \tag{B.15}$$

where the functional value at the interpolation points are weighted by

$$w_k = \frac{2}{N_L(N_L+1)} \frac{1}{(P_{N_L}(s_k))^2} \tag{B.16}$$

which are tabulated for $N_L = 7$ in Table B.2.

Table B.2: Legendre-Gauss-Lobatto-Lagrange element node points and the corresponding Gauss-Lobatto integration weights

Nodes s_k	Weights w_k
-1	0.035 714
$-0.871\,740$	0.210 704
$-0.591\,700$	0.341 123
$-0.209\,299$	0.412 459
0.209 299	0.412 459
0.591 700	0.341 123
0.871 740	0.210 704
1	0.035 714

For integration over the actual domain of the system, we sum over elements and nodes while multiplying by the inverse Jacobian of the affine mapping function

$$\int_{x_L}^{x_R} dx\ f(x) = M^{-1} \sum_l \sum_k f(x_k^l) w_k \tag{B.17}$$

B.6 Brownain Dynamics and smFRET simulation

The Brownian Dynamics trajectories are run by numerically integrating the differential equation

$$dx = DF(x)dt + \sqrt{2D}dW_t \tag{B.18}$$

with a time step $dt = 1 \times 10^{-9}$ by drawing a random number $R(t) \sim \mathcal{N}(0, \sigma^2)$ from a zero mean normal distribution with variance $\sigma^2 = (dt)^{-1}$ such that the time update equation is

$$x(t + dt) = x(t) + DF(x(t))dt + \sqrt{2D}R(t) \tag{B.19}$$

To simulate the Poisson process of photon emission, we employ the Next Reaction Method [165] version of Kinetic Monte Carlo which generates the events of donor/acceptor photon emission separated by a dwell time between donor/acceptor arrival. These $\{\Delta t_i \sim Exp(I_i) \mid i \in a, d\}$ are exponentially distributed according to the time varying rates which are dependent upon

position

$$I_a(x(t)) = I_a^0 \frac{1}{1+x^6(t)} + B_a \quad \text{(acceptor)} \tag{B.20}$$

$$I_d(x(t)) = I_d^0 \frac{x^6(t)}{1+x^6(t)} + B_d \quad \text{(donor)} \tag{B.21}$$

The method begins by selecting two uniformly distributed random numbers $u_a, u_d \in (0,1]$ then monitoring two separator integrations to find the $\Delta t_i \ \forall i$ such that

$$\int_0^{\Delta t_i} I_i(t')dt' = -\ln(u_i) \tag{B.22}$$

recording the event, then reseting that integration and u_i while the other continues unaffected. This procedure is summarized in Algorithm 4.

Algorithm 4 Brownian dynamics and smFRET simulation of length t_{exp}

procedure SMFRET(t_{exp})
  $u_0 \leftarrow \mathcal{U}(0,1)$
  $x \leftarrow$SOLVE($\int_{-\infty}^{x} p_{\text{eq}}(x')dx' = u_0$) ▷ Find initial starting point from equilibrium distribution
  $u_a, u_d \leftarrow \mathcal{U}(0,1)$
  $t, \Sigma_a, \Sigma_d \leftarrow 0$ ▷ Keep track of axproximate integral Σ for Equation B.22
  $dt = 1 \times 10^{-9}$
  repeat
    $R \leftarrow \mathcal{N}(0, dt^{-1})$
    $x \leftarrow x + DF(x)dt + \sqrt{2D}R$
    $\Sigma_a \leftarrow \Sigma_a + I_a(x)dt$
    $\Sigma_d \leftarrow \Sigma_d + I_a(x)dt$
    if $\Sigma_a > -\ln(u_a)$ **then** RECORD(t, x,"acceptor") ▷ Record photon detection times to output file
      $u_a \leftarrow \mathcal{U}(0,1)$
      $\Sigma_a \leftarrow 0$
    end if
    if $\Sigma_d > -\ln(u_d)$ **then** RECORD(t, x,"donor")
      $u_d \leftarrow \mathcal{U}(0,1)$
      $\Sigma_d \leftarrow 0$
    end if
    $t \leftarrow t + dt$
  until $t > t_{exp}$
end procedure

B.7 Jeffery's Prior for Exponential Distribution of Inter-Photon Times

The jeffery's prior is optimum for parameter estimation in reducing bias in experimental procedures where the probability information is a member of the exponential family of models. This prior is given by the Fisher information[86, 166]

$$\pi(x) \propto \sqrt{J(x)} \tag{B.23}$$

and estimations using this prior give Bayesian Maximum a posteriori estimation (MAP).

The Fisher information for estimating position based upon smFRET data is

$$J(x) = \mathbb{E}_{\Delta t}\left[\left(\frac{\partial \ln p(\Delta t|x)}{\partial x}\right)^2 |x\right] \tag{B.24}$$

For the exponential distribution $p(\Delta t|x) = I(x)e^{-I(x)\Delta t}$ with intensity $I(x)$, doing the derivative gives

$$J(x) = \mathbb{E}_{\Delta t}\left[\left(\frac{I'(x) - I(x)I'(x)\Delta t}{I(x)}\right)^2 |x\right] \tag{B.25}$$

Doing the expectation over the distribution of Δt gives the properties of the exponential distribution namely $\mathbb{E}[\Delta t] = (I(x))^{-1}$ and $\mathbb{E}[(\Delta t)^2] = 2(I(x))^{-2}$ gives the fisher information and Jeffery's Prior

$$J(x) = \left(\frac{I'(x)}{I(x)}\right)^2 \tag{B.26}$$

$$\pi(x) \propto \left|\frac{I'(x)}{I(x)}\right| \tag{B.27}$$

For acceptor intensity $I_a(x) = I_a^0/(1+x^6) + B_a$ and $I_a'(x) = -(6I_a^0 x^5)/(1+x^6)^2$ we derive acceptor contribution to the prior

$$\pi_a(x) = \frac{6I_a^0 x^5}{(1+x^6)}\frac{1}{I_a^0 + B_a(1+x^6)} \tag{B.28}$$

For donor intensity $I_d(x) = I_d^0 x^6/(1+x^6) + B_d$ and $I_d'(x) = -(6I_d^0 x^{11})/(1+x^6)^2 + 6I_d^0 x^5/(1+x^6)$ we derive donor contribution to the prior

$$\pi_d(x) = \frac{6I_d^0 x^5}{(1+x^6)}\frac{1}{I_d^0 x^6 + B_d(1+x^6)} \tag{B.29}$$

For the independent channels acceptor and donor, the Fisher information is just the addition $J(x) = J_a(x) + J_d(x)$ and therefore the log-prior distribution is an addition

$$\ln \pi(x) = \ln \sqrt{\pi_a^2(x) + \pi_d^2(x)} \tag{B.30}$$

which is subsequently added to the dark state operator such that the Jeffery's prior applied once per independent measurement.

By adding prior we reduce bias in our maximum likelihood procedure[86] because the inference step is now combines the Bayesian prior. The effective result is sharpening of the precision and suppression of oscillations in the wings of the domain.

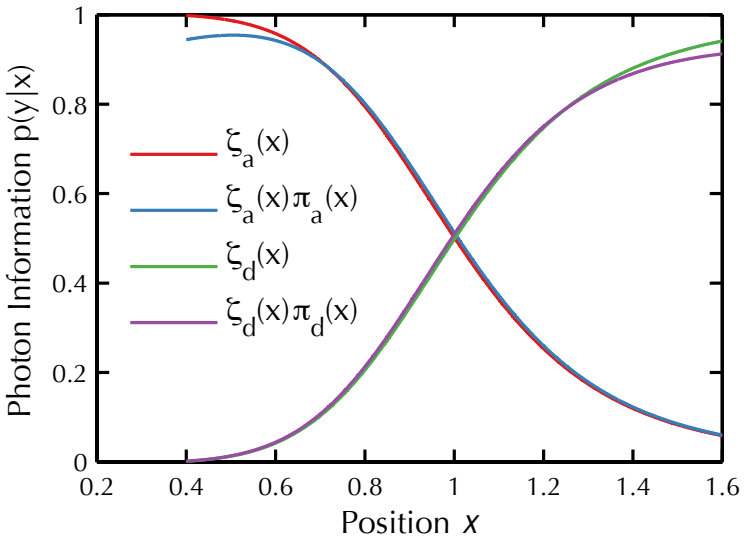

Figure B.3: Jefferys Prior for the acceptor and donor exponential process and the combined prior used for initializing $\pi(x)$.

B.8 Proof of the time integration perturbation theory

The proof of the time integration perturbation theory given by Wilcox [81] for scalar derivatives by realizing the equivalence of the expression $e^{-\Delta t(\mathbf{H}+\epsilon\mathcal{H}')} = \left(e^{-\frac{\Delta t}{n}(\mathbf{H}+\epsilon\mathcal{H}')}\right)^n \forall n$. To isolate the functional dependence of the perturbation

$\mathcal{H}'$, the derivative with respect to ϵ is applied. Applying the product rule to the derivative of the product results in the sum of many products where the perturbation operator $\mathbf{H}'$ is sandwiched. Because the operators generally do not commute, $[\mathbf{H}, \mathbf{H}'] \neq 0$, the ordering of terms from the product rule must be explicitly preserved.

$$\frac{d}{d\epsilon} e^{-\Delta t(\mathbf{H}+\epsilon\mathcal{H}')} = \sum_{m=1}^{n} \left\{ \left(1 - \frac{\Delta t}{n}(\mathbf{H} + \epsilon\mathcal{H}')\right)^{m-1} \left(-\frac{\Delta t}{n}\mathcal{H}'\right) \left(1 - \frac{\Delta t}{n}(\mathbf{H} + \epsilon\mathcal{H}')\right)^{n-m} \right\} \quad \text{(B.31)}$$

The derivative is then evaluated at $\epsilon = 0$,and the exponent for each product sandwiching $\mathbf{H}'$ is multiplied by $\frac{n}{n}$.

$$\left.\frac{d}{d\epsilon} e^{-\Delta t(\mathbf{H}+\epsilon\mathcal{H}')}\right|_{\epsilon=0} = -\lim_{n\to\infty} \sum_{m=1}^{n} \frac{\Delta t}{n} \left(e^{-\frac{\Delta t}{n}\mathbf{H}}\right)^{n\frac{m-1}{n}} \mathcal{H}' \left(e^{-\frac{\Delta t}{n}\mathbf{H}}\right)^{n-m} + \mathcal{O}((\Delta t/n)^4) \quad \text{(B.32)}$$

where the error is given by the Trotter splitting formula[126]. Taking the $\lim_{n\to\infty}$ given an effective Riemann integral in the form of a line integral from 0 to Δt.

$$\left.\frac{d}{d\epsilon} e^{-\Delta t(\mathbf{H}+\epsilon\mathcal{H}')}\right|_{\epsilon=0} = -\int_0^{\Delta t} dt e^{-t\mathbf{H}} \mathcal{H}' e^{-(\Delta t - t)\mathbf{H}} \quad \text{(B.33)}$$

allowing for the calculation of the functional derivatives from this perturbation theory of the exponential of operators.

APPENDIX C

Fisher Information Metric for the Langevin Equation and Least Informative Models of Continuous Stochastic Dynamics

C.1 Potential of Mean Force of the Reference Model

The potential of mean force $V(x)/k_BT$ is constructed using the piecewise cubic Hermite interpolation polynomial `pchip()` in MATLAB with the following points

Table C.1: Interpolation points for $V(x)$

Position	0.5	0.8	0.9	1	1.1	1.2	1.5
$V(x)$	60	0	4	$\ln(4)$	5	$\ln(2)$	60

C.2 Derivation of Eq. (4.18) of FIT for symmetric $\rho(x_t, x_0)$

The starting point is Eq. (4.12) and is restated here:

$$I_{\Delta t}(x,y) = \int dx_t dx_0 \frac{\delta \ln p(x_t|x_0)}{\delta \rho_{\rm eq}(x)} \frac{\delta \ln p(x_t|x_0)}{\delta \rho_{\rm eq}(y)} p(x_t, x_0). \tag{C.1}$$

Inputing $p(x_t|x_0) = \rho_{\rm eq}(x_t)\rho(x_t, x_0)/\rho_{\rm eq}(x_0)$ and $\ln p(x_t|x_0) = \ln \rho_{\rm eq}(x_t) + \ln \rho(x_t, x_0) - \ln \rho_{\rm eq}(x_0)$ leads to 9 terms:

$$I_{\Delta t}(x,y) = \int dx_t dx_0 \frac{\delta \ln \rho(x_t, x_0)}{\delta \rho_{\rm eq}(x)} \frac{\delta \ln \rho(x_t, x_0)}{\delta \rho_{\rm eq}(y)} p(x_t, x_0) \tag{C.2}$$

$$+ \int dx_t dx_0 \frac{\delta \ln \rho_{\rm eq}(x_t)}{\delta \rho_{\rm eq}(x)} \frac{\delta \ln \rho_{\rm eq}(x_t)}{\delta \rho_{\rm eq}(y)} p(x_t, x_0) \tag{C.3}$$

$$+ \int dx_t dx_0 \frac{\delta \ln \rho_{\rm eq}(x_0)}{\delta \rho_{\rm eq}(x)} \frac{\delta \ln \rho_{\rm eq}(x_0)}{\delta \rho_{\rm eq}(y)} p(x_t, x_0) \tag{C.4}$$

$$- \int dx_t dx_0 \frac{\delta \ln \rho_{\rm eq}(x_0)}{\delta \rho_{\rm eq}(x)} \frac{\delta \ln \rho_{\rm eq}(x_t)}{\delta \rho_{\rm eq}(y)} p(x_t, x_0) \tag{C.5}$$

$$- \int dx_t dx_0 \frac{\delta \ln \rho_{\rm eq}(x_t)}{\delta \rho_{\rm eq}(x)} \frac{\delta \ln \rho_{\rm eq}(x_0)}{\delta \rho_{\rm eq}(y)} p(x_t, x_0) \tag{C.6}$$

$$+ \int dx_t dx_0 \frac{\delta \ln \rho_{\rm eq}(x_t)}{\delta \rho_{\rm eq}(x)} \frac{\delta \ln \rho(x_t, x_0)}{\delta \rho_{\rm eq}(y)} p(x_t, x_0) \tag{C.7}$$

$$- \int dx_t dx_0 \frac{\delta \ln \rho_{\rm eq}(x_0)}{\delta \rho_{\rm eq}(x)} \frac{\delta \ln \rho(x_t, x_0)}{\delta \rho_{\rm eq}(y)} p(x_t, x_0) \tag{C.8}$$

$$+ \int dx_t dx_0 \frac{\delta \ln \rho(x_t, x_0)}{\delta \rho_{\rm eq}(x)} \frac{\delta \ln \rho_{\rm eq}(x_t)}{\delta \rho_{\rm eq}(y)} p(x_t, x_0) \tag{C.9}$$

$$- \int dx_t dx_0 \frac{\delta \ln \rho(x_t, x_0)}{\delta \rho_{\rm eq}(x)} \frac{\delta \ln \rho_{\rm eq}(x_0)}{\delta \rho_{\rm eq}(y)} p(x_t, x_0) \tag{C.10}$$

The collection of integrals of that make up $I_{\Delta t}(x,y)$ can be paired down because the density field is symmetric to exchange of $x_t \leftrightarrow x_0$, $\rho(x_t, x_0) = \rho(x_0, x_t)$. Performing this exchange on Eq. (C.10) gives and exact copy and cancels with Eq. (C.9). Likewise Eq. (C.8) cancels with Eq. (C.7). The remaining 4 integrals containing the $\rho_{\rm eq}$ term require the functional derivative:

$$\frac{\delta \ln \rho_{\rm eq}(x_t)}{\delta \rho_{\rm eq}(x)} = \frac{\delta(x_t - x)}{\rho_{\rm eq}(x)}. \tag{C.11}$$

Therefore, Eqs. (C.6) and (C.5) become the integral of two delta functions:

$$\int dx_t dx_0 \frac{\delta(x_t - x)}{\rho_{\text{eq}}(x)} \frac{\delta(x_0 - y)}{\rho_{\text{eq}}(y)} p(x_t, x_0) = \rho(x, y), \tag{C.12}$$

for which the property of $\rho(x, y) = p(x, y)/\rho_{\text{eq}}(x)\rho_{\text{eq}}(y)$ is utilized as well. Applying the same functional derivative to Eqs. (C.4) and (C.3) gives the alternative result:

$$\int dx_0 \frac{\delta(x_0 - x)}{\rho_{\text{eq}}(x)} \frac{\delta(x_0 - y)}{\rho_{\text{eq}}(y)} \int dx_t p(x_t, x_0) = \delta(x - y). \tag{C.13}$$

Here, the integration over x_t can be done from the onset due to the lack of its dependence in the delta functions and noting that $p(x_0)/\rho_{\text{eq}}(x_0)\rho_{\text{eq}}(x_0) = 1$. Combining these results gives Eq. (4.18):

$$I_{\Delta t}(x, y) = \int dx_t dx_0 \frac{\delta \ln \rho(x_t, x_0)}{\delta \rho_{\text{eq}}(x)} \frac{\delta \ln \rho(x_t, x_0)}{\delta \rho_{\text{eq}}(y)} p(x_t, x_0) + 2\delta(x - y) - 2\rho(x_t, x_0). \tag{C.14}$$

C.3 Proof of the Hermitian Property of the Symetrized Fokker-Planck Equation

The Hermitian property is established by showing that $\langle \psi_i | \boldsymbol{H} | \psi_j \rangle = \langle \psi_i | \boldsymbol{H} | \psi_j \rangle$. For the targeted operator defined in Eq. (4.21), the left-hand side is:

$$\int dx \frac{\psi_i(x)}{\rho_{\text{eq}}(x)} \nabla \cdot (D(x)\rho_{\text{eq}}^2(x) \nabla \frac{\psi_j(x)}{\rho_{\text{eq}}(x)}). \tag{C.15}$$

This equation can be transformed by an integration by parts to:

$$- \int dx (D(x)\rho_{\text{eq}}^2(x) \nabla \frac{\psi_j(x)}{\rho_{\text{eq}}(x)}) \cdot \nabla \frac{\psi_i(x)}{\rho_{\text{eq}}(x)}. \tag{C.16}$$

The zero boundary terms in the space of square integrable functions $\psi \subset L^2$ were applied. Another integration by parts on the $\nabla(\psi_j(x)/\rho_{\text{eq}}(x))$ term gives

$$\int dx \frac{\psi_j(x)}{\rho_{\text{eq}}(x)} \nabla \cdot (D(x)\rho_{\text{eq}}^2(x) \nabla \frac{\psi_i(x)}{\rho_{\text{eq}}(x)}), \tag{C.17}$$

which is recognized as $\langle \psi_i | \boldsymbol{H} | \psi_j \rangle$. The Hermitian property of the operator defined in Eq. (4.21) is thus verified.

C.4 FIT for the Gaussian Process of Constant Force

The time propagator for this process is:

$$p(x_{\Delta t}|x_0) = \frac{1}{\sqrt{4\pi D\Delta t}} \exp\left(-\frac{(x_{\Delta t} - x_0 - FD\Delta t)^2}{4D\Delta t}\right). \tag{C.18}$$

The calculation of FIT via Eq. 4.12 starts by taking derivatives of $p(x_{\Delta t}|x_0)$ with respect to the x-independent force F and diffusion coefficient D.

$$\frac{\partial \ln p(x_{\Delta t}|x_0)}{\partial F} = \frac{(\Delta x - FD\Delta t)}{2} \tag{C.19}$$

$$\frac{\partial \ln p(x_{\Delta t}|x_0)}{\partial D} = \frac{F(\Delta x - FD\Delta t)}{2D} + \frac{(\Delta x - FD\Delta t)^2}{4D^2\Delta t} - \frac{1}{2D}. \tag{C.20}$$

Here, $\Delta x = x_{\Delta t} - x_0$. Each element of the Fisher information matrix can then be found by multiplying these derivatives by one anther and taking the expectation over the joint probability $\int dx_{\Delta t} dx_0\ p(x_{\Delta t}, x_0)$ which by rotating the domain, can easily be done as a integration over Δx and $\Delta x' = x_{\Delta t} + x_0$.

$$I_{F,F} = \mathbb{E}\left[\frac{(\Delta x - FD\Delta t)^2}{4}\right] = \frac{D\Delta t}{2} \tag{C.21}$$

$$I_{F,D} = I_{D,F} = \frac{F\Delta t}{2} \tag{C.22}$$

$$I_{D,D} = \frac{1}{2D^2} + \frac{F^2\Delta t}{2D} \tag{C.23}$$

The following properties for the moments of the Gaussian distribution have also been used to calculate the matrix elements.

$$\mathbb{E}\left[(X-\mu)^N\right] = \begin{cases} 0 & N = 1 \\ \sigma^2 & N = 2 \\ 0 & N = 3 \\ 3\sigma^3 & N = 4 \end{cases} \tag{C.24}$$

The mean μ is $FD\Delta t$ and the variance σ^2 is $2D\Delta t$.

Taking the continuous limit for the Fisher information of trajectories, $\mathcal{I} = \lim_{\Delta t \to 0^+} t_{\text{obs}} I_{\Delta t}/\Delta t$, the terms with Δt cancels the first factor terms for $I_{D,D}$. To evaluate the information change between parameter sets $\mathcal{J}_{1\to 2}$, we integrate according to Eq. 4.46 for the path:

$$F(\xi) = F_1 + (F_2 - F_1)\xi \tag{C.25}$$

$$D(\xi) = D_1 + (D_2 - D_1)\xi, \tag{C.26}$$

where $\xi \in [0,1]$. After apply the sum over the two parameters and the Fisher information $\sum_{i,j} \frac{\partial \theta_i}{\partial \xi} \mathcal{I}_{i,j}(\vec{\theta}) \frac{\partial \theta_j}{\partial \xi}$ the following integral remains

$$\mathcal{J}_{1\to 2} = \frac{t_{\text{obs}}}{2} \int_0^1 d\xi \; (\Delta F)^2 \frac{D(\xi)}{2} \tag{C.27}$$

$$+(\Delta F \Delta D) F(\xi) + (\Delta D)^2 \left(\frac{F^2(\xi)}{2D(\xi)} + \frac{\Delta t^{-1}}{2D^2(\xi)} \right). \tag{C.28}$$

Here, $\Delta F = F_2 - F_1$ and $\Delta D = D_2 - D_1$. After performing the integration and collecting terms, the information change is:

$$\mathcal{J}_{1\to 2} = t_{\text{obs}} \frac{(F_2 D_2 - F_1 D_1)^2}{2D_2} + \lim_{\Delta t \to 0^+} \frac{t_{\text{obs}}}{\Delta t} \left[2 - \frac{D_2}{D_1} - \frac{D_1}{D_2} \right]. \tag{C.29}$$

This information change is 0 for the case when $D_1 = D_2$ and $F_1 = F_2$. However, the two groupings have different scaling behaviors as $\Delta t \to 0$. The second grouping which solely depends upon the ratio of diffusion constants gives a rate of divergence at $\propto 1/\Delta t$. This asymptotic behavior for Gaussian processes has also been worked out via the Kolmogorov-Siani entropy measure.[48] However, the term with the difference in the square force remains constant through different values of time resolution and grows linearly with the length of the trajectory t_{obs}.

C.5 FIT for the Ornstein-Uhlenbeck Process

The Ornstein-Uhlenbeck (OU) process is a system governed by a simple harmonic potential $V(x) = -1/2kx^2$ with the spring constant k reflecting the width of the harmonic PMF. The equilibrium probability density of the OU process is Gaussian:

$$p_{\text{eq}}(x) = \sqrt{\frac{k}{2\pi}} e^{-kx^2/2}. \tag{C.30}$$

The time propagation probability density for the OU process is:

$$p(x_{\Delta t}|x_0) = \frac{\sqrt{k}}{\sqrt{2\pi(1 - e^{-2Dk\Delta t})}} \exp\left(\frac{-k \left(x_{\Delta t} - x_0 e^{-kD\Delta t} \right)^2}{2 - 2e^{-2Dk\Delta t}} \right). \tag{C.31}$$

In the continuum limit, we take the Taylor expansion of the exponential up to linear order in time $\exp(-2Dk\Delta t) \simeq 1 - 2Dk\Delta t$ to give the approximate propagator:

$$p(x_{\Delta t}|x_0) = \frac{1}{\sqrt{4\pi D\Delta t}} \exp\left(\frac{-\left(x_{\Delta t} - x_0 + x_0 kD\Delta t\right)^2}{4D\Delta t}\right). \tag{C.32}$$

Taking the derivative of the log propagator with respect to k results in:

$$\frac{\partial \ln p(x_{\Delta t}|x_0)}{\partial k} = \frac{-x_0\left(x_{\Delta t} - x_0 + x_0 kD\Delta t\right)}{2}. \tag{C.33}$$

Then, the Fisher information is the expectation of this derivative squared:

$$\mathcal{I}_k = \frac{t_{\text{obs}}}{\Delta t}\mathbb{E}_{x_{\Delta t}, x_0}\left[\frac{x_0^2\left(x_{\Delta t} - x_0 + x_0 kD\Delta t\right)^2}{4}\right]. \tag{C.34}$$

First, taking the expectation over $x_{\Delta t}$ puts $2D\Delta t$ in place of what is essentially a term of $\mathbb{E}[(X-\mu)^2]$. Next, the remaining expectation over x_0^2 gives the variance of the equilibrium distribution which is $1/k$ for a final answer:

$$\mathcal{I}_k = \frac{t_{\text{obs}} D}{2k}. \tag{C.35}$$

To relate this result to the overall change in information from an initial state of $k_1 = 0$ to a final state of $k_2 = k$, we perform the line integral:

$$\mathcal{J} = \frac{-1}{2}\int_0^1 d\xi \Delta k \frac{t_{\text{obs}} D}{2k(\xi)}\Delta k = -\frac{t_{\text{obs}} D}{4}\Delta k \ln\frac{k_2}{k_1}. \tag{C.36}$$

Here, $\Delta k = k_2 - k_1 = k$. Comparing to the result via the general functional of Eq. (4.51), $\mathcal{J} = -D\int dx\ (\nabla\rho_{\text{eq}}(x))^2 = -(Dk^2/4)\int dx\ x^2 p_{\text{eq}}(x) = -Dk/4$, the result is different although the scaling is similar. It is because the Fisher information is not constant with the parameter k. Naturally, systems with a larger spring constants are more restrictive and also moves faster, i.e., the dynamics is more deterministic. Therefore, there is more information for dynamics models with a larger spring constant.

C.6 Functional derivative of mean first passage times

The functional derivatives for the mean first passage time give by:

$$\tau_{\mathrm{x_A}\to\mathrm{x_B}} = k^{-1}_{A\to B} = \frac{1}{D}\int_{x_A}^{x_B} dx\ \rho^{-2}(x)\int_{x_L}^{x} dx'\ \rho^2(x')$$
$$\tau_{\mathrm{x_B}\to\mathrm{x_A}} = k^{-1}_{B\to A} = \frac{1}{D}\int_{x_B}^{x_A} dx\ \rho^{-2}(x)\int_{x_R}^{x} dx'\ \rho^2(x') \quad \text{(C.37)}$$

is derived from finding the first order perturbation from an arbitrary function $f(x)$

$$\int dx \frac{\delta \boldsymbol{\tau}_{x_A\to x_B}}{\delta\rho(x)} f(x) = \frac{d}{d\epsilon}\tau_{\mathrm{x_A}\to\mathrm{x_B}}[\rho(x)+f(x)]\bigg|_{\epsilon=0}. \quad \text{(C.38)}$$

Performing the derivation of the above equation with respect to ϵ gives:

$$\int dx \frac{\delta \boldsymbol{\tau}_{x_A\to x_B}}{\delta\rho(x)} f(x) = \overbrace{\frac{2}{D}\int_{x_A}^{x_B} dx\ \rho^{-2}(x)\int_{x_L}^{x} dx'\ f(x')\rho(x')}^{\text{Parts}} \quad \text{(C.39)}$$
$$- \frac{2}{D}\int_{x_A}^{x_B} dx\ f(x)\rho^{-3}(x)\int_{x_L}^{x} dx'\ \rho^2(x'). \quad \text{(C.40)}$$

The perturbation $f(x)$ is isolated from the nested integral in Eq. (C.39) by performing an integration by parts with $u = \int_{x_L}^{x} dx' f(x')\rho(x')$ and $dv = \rho^{-2}(x)$ to give

$$\text{Parts} = \frac{2}{D}\left(\int_{x_L}^{x} f(x)\rho(x)\right)\left(\int_{x_A}^{x} dx'\rho^{-2}(x')\right)\bigg|_{x_A}^{x_B}$$
$$- \frac{2}{D}\int_{x_A}^{x_B} dx f(x)\rho(x)\int_{x_A}^{x} dx'\rho^{-2}(x'). \quad \text{(C.41)}$$

APPENDIX D

Convolution of system and measurement timescales in statistical estimation

For the smFRET experiment, the probability of observing N_a acceptor photons and N_d donor photons in a time window Δt is given by the independent Poisson probabilities parameterized by intensities I_a and I_d

$$p(N_a, N_d; I_a, I_d, \Delta t) = \frac{(-I_a \Delta t)^{N_a} e^{-I_a \Delta t}}{N_a!} \frac{(-I_d \Delta t)^{N_d} e^{-I_d \Delta t}}{N_d!} \tag{D.1}$$

These intensities are functions of the normalized distance between the phloro-phores x according to the equation

$$I_a(x) = \frac{I_a^0}{1 + x^6} \quad ; \quad I_d(x) = \frac{x^6 I_d^0}{1 + x^6} \tag{D.2}$$

By counting the number of photons for a time from the smFRET experiment, one can calculate a maximum likelihood estimate of the distance of the two dyes and normal error estimate[60] seen in Section B.4. The difficulty arrises in that the distance $x(t)$ is also a function of time according to the Langevin equation $dx_t = DF(x_t) + \sqrt{2D} dW_t$. Therefore the system is moving during the time at which one would like to make an estimate of its position. The Poisson probabilities for a system with time varying intensities is just given by those time average quantities $p(N_a, N_d, \langle I \rangle_a, \bar{I}_d, , \Delta t)$ where

$$\langle I \rangle_a (\Delta t) = \frac{I_a^0}{\Delta t} \int_0^{\Delta t} dt \, \frac{1}{1 + x^6(t)} \tag{D.3}$$

However, now the statistical estimator is finding the time smeared position of the system rather than the actual location. Performing a histogram reconstruction of these estimates to get an approximate equilibrium distribution of states can therefore be problematic.[63]

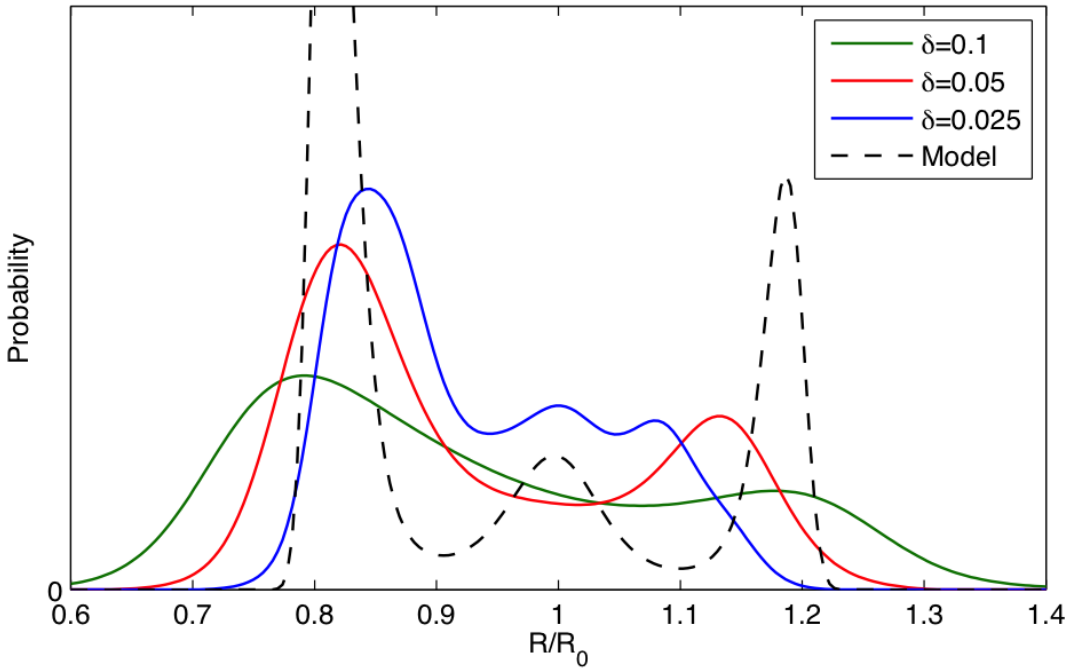

Figure D.1: The histogram of estimated states from the time windowed estimate at various precision levels $\delta = 0.1, 0.05, 0.025$ for a 3-well hypothetical model.

Finding the probability for the time averaged intensity, depends on the governing dynamics of the system, namely the force profile and diffusion constant. Gopich and Szabo have derived these distributions using a numerical inversion of the Fourier and Laplace transformed time propagator[34], we present an alternative approach using a 2-D Fokker planck equation and a finite element partial derivative solver. We can actually write the time evolution of this averaged intensity by taking the derivative with respect to Δt to give

$$\frac{d\langle I\rangle_a(t)}{dt} = \frac{1}{t}\left(\frac{I_a^0}{1+x^6(t)} - \langle I\rangle_a(t)\right) \tag{D.4}$$

which is purely deterministic dynamics with an effective information force $F_I(x,\langle I\rangle) = \left(\frac{I_a^0}{1+x^6(t)} - \langle I\rangle_a(t)\right)$ and time dependent diffusion $D_I(t) = \frac{1}{t}$ that

tends towards compiling the instantaneous values of intensity. Combining the stochastic differential equation for x_t with this ODE can give the time evolution of the joint probability

$$\frac{\partial p(x,\langle I\rangle_a,t)}{\partial t} = D\frac{\partial^2 p(x,\langle I\rangle_a,t)}{\partial x^2} - \frac{\partial DF(x)p(x,\langle I\rangle_a,t)}{\partial x} - \frac{\partial D_I(t)F_I(x)p(x,\langle I\rangle_a,t)}{\partial x} \tag{D.5}$$

This equation is integrated starting from the initial condition that $p(x,\langle I\rangle_a,t=0) = p_{eq}(x)\delta(x-I_a(x))$, using the finite element solver COMSOL (Shown in Figure D.2).

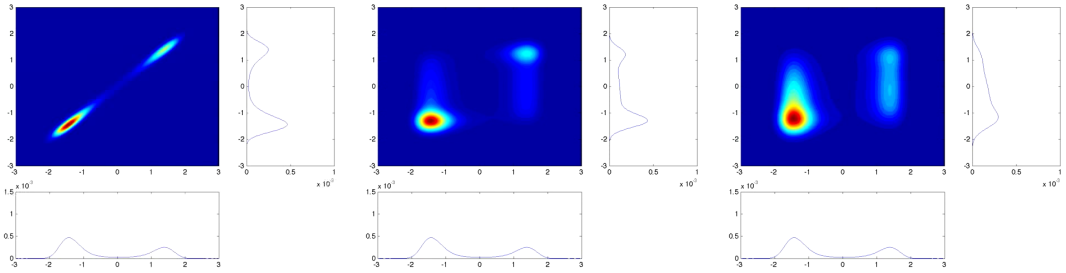

Figure D.2: Contours of the FEM Integration of $p(x,\langle I\rangle,t)$ from COMSOL with the marginalized probabilities plotted on the borders of the contour. From left to right $\Delta t = 0, 2, 5$.

The desired probabilities of the time averaged intensity can then be found though marginalizing the joint probability $p(\langle I\rangle_a) = \int dx p(x,\langle I\rangle_a; t=\Delta t)$ at the desired time window length. The complementary procedure is also followed for $p(\langle I\rangle_d)$. In Figure D.3 the Poisson distributions from D.1 are plotted for a 3-well system both with dynamics turned off $D=0$ and dynamics that decays on the timescale of the emission rates of photons. Only in the static system, can a clear separation between the two major states emission probabilities be found.

Finally, these intensity distributions are used to produce estimates of the results of applying the Maximum Information Method (MIM) point estimate and histogram analysis from [63]. In Figure D.1, the parameter δ adjusts the standard error of the point estimate from a particular time bin. As this error is reduced, one must wait and watch for photons longer and therefore the system is allowed to relax further and blur the expected state distribution. At short times, there is not enough experimental accuracy in the estimate to separate out states on the domain.

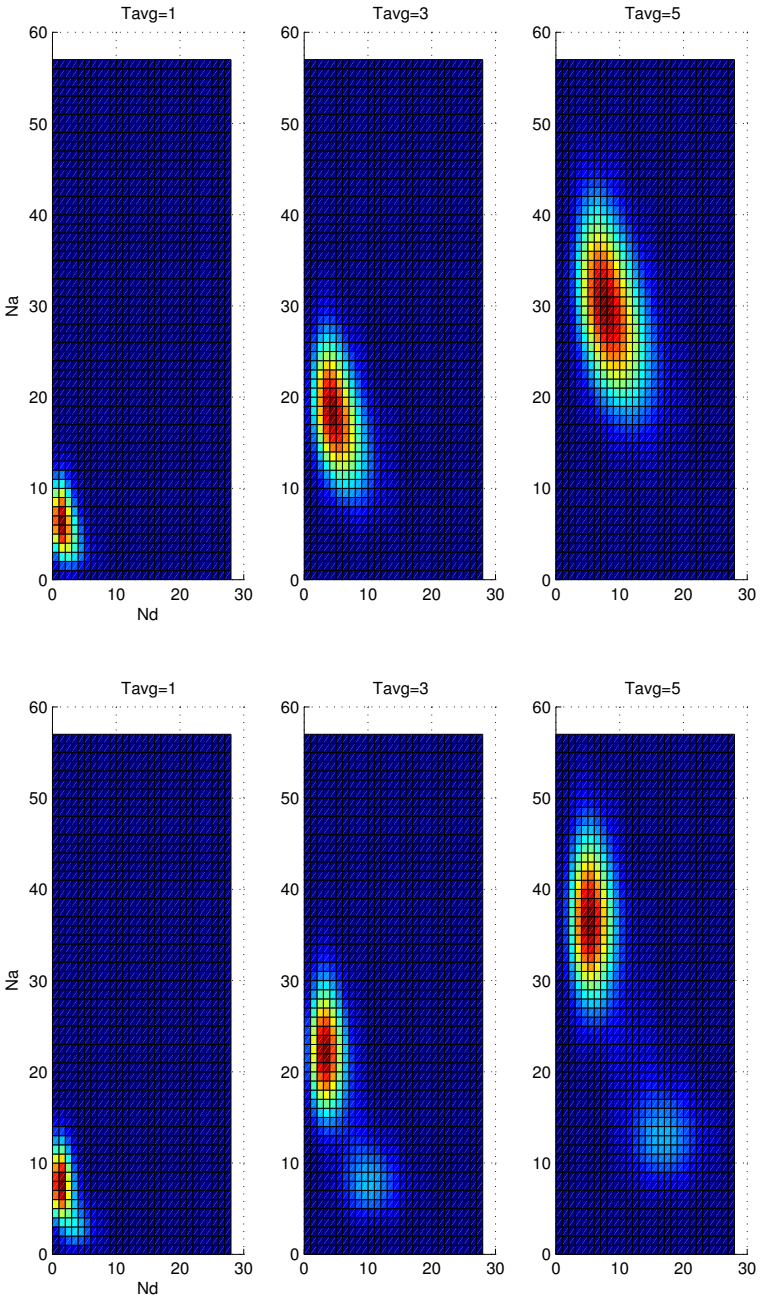

Figure D.3: Contours of the distributions of photon counts $p(N_a, N_d; \Delta t)$ at various time resolutions $Tavg = 1, 3, 5$ for a 3-well hypothetical model. (top) Dynamic system which relaxes on the timescale of $\tau = 1$ and thus blurs the system distribution into one likely intensity. Spread of contour is then due to Poisson statistics. (bottom) Static system which preserves 2 clustered contour peaks (3rd equilibrium states is too rare and is lost).

References

[1] Igarashi, K. *et al.* Traffic Jams Reduce Hydrolytic Efficiency of Cellulase on Cellulose Surface. *Science* **333**, 1279–1282 (2011).

[2] Sandberg, A. *et al.* Stabilization of neurotoxic Alzheimer amyloid-oligomers by protein engineering. *Proc. Natl. Acad. Sci. USA* **107**, 15595–15600 (2010).

[3] Wainwright, M. J. & Jordan, M. I. *Graphical Models, Exponential Families, and Variational Inference* (Now Publishers Inc, 2008), now publishers inc edn.

[4] Myung, J., Pitt, M. & Navarro, D. Model selection in cognitive science as an inverse problem. *Proc. SPIE* **5674**, 219–228 (2005).

[5] Hammes, G. G. How Do Enzymes Really Work? *Journal of Biological Chemistry* **283**, 22337–22346 (2008).

[6] McLaughlin, R. N., Jr, Poelwijk, F. J., Raman, A., Gosal, W. S. & Ranganathan, R. The spatial architecture of protein function and adaptation. *Nature* (2012).

[7] Halabi, N., Rivoire, O., Leibler, S. & Ranganathan, R. Protein Sectors: Evolutionary Units of Three-Dimensional Structure. *Cell* **138**, 774–786 (2009).

[8] Frauenfelder, H. *et al.* A unified model of protein dynamics. *Proc. Natl. Acad. Sci. USA* **106**, 5129 (2009).

[9] Henzler-Wildman, K. A. *et al.* A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450**, 913–916 (2007).

[10] Langevin, P. Sur la theorie du mouvement brownien mouvement brownien. *C. R. Acad. Sci.* **146**, 530–533 (1908).

[11] Vanden-Eijnden, E. & Venturoli, M. Revisiting the finite temperature string method for the calculation of reaction tubes and free energies. *J. Chem. Phys.* **130**, 194103 (2009).

[12] Best, R. B. & Hummer, G. Coordinate-dependent diffusion in protein folding. *Proc. Natl. Acad. Sci. USA* **107**, 1088–1093 (2010).

[13] Hanson, J. A. *et al.* Illuminating the mechanistic roles of enzyme conformational dynamics. *Proc. Natl. Acad. Sci. USA* **104**, 18055–18060 (2007).

[14] Bolhuis, P. G., Chandler, D., Dellago, C. & Geissler, P. L. T RANSITIONP ATHS AMPLING: Throwing Ropes Over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.* **53**, 291–318 (2002).

[15] Sabo, D., Meuwly, M., Freeman, D. L. & Doll, J. D. A constant entropy increase model for the selection of parallel tempering ensembles. *J. Chem. Phys.* **128**, 174109 (2008).

[16] Velez-Vega, C., Borrero, E. E. & Escobedo, F. A. Kinetics and reaction coordinate for the isomerization of alanine dipeptide by a forward flux sampling protocol. *J. Chem. Phys.* **130**, 225101 (2009).

[17] Zhang, C. & Ma, J. Comparison of sampling efficiency between simulated tempering and replica exchange. *J. Chem. Phys.* **129**, 134112 (2008).

[18] Shaw, D. E. *et al.* Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **330**, 341–346 (2010).

[19] Curuksu, J. & Zacharias, M. Enhanced conformational sampling of nucleic acids by a new Hamiltonian replica exchange molecular dynamics approach. *J. Chem. Phys.* **130**, 104110 (2009).

[20] Pan, A. C. & Roux, B. Building Markov state models along pathways to determine free energies and rates of transitions. *J. Chem. Phys.* **129**, 064107 (2008).

[21] Vanden-Eijnden, E. & Venturoli, M. Markovian milestoning with Voronoi tessellations. *J. Chem. Phys.* **130**, 194101 (2009).

[22] Prinz, J.-H. *et al.* Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **134**, 174105 (2011).

[23] Peters, B., Beckham, G. T. & Trout, B. L. Extensions to the likelihood maximization approach for finding reaction coordinates. *J. Chem. Phys.* **127**, 034109 (2007).

[24] Autieri, E., Faccioli, P., Sega, M., Pederiva, F. & Orland, H. Dominant reaction pathways in high-dimensional systems. *J. Chem. Phys.* **130**, 064106 (2009).

[25] Chelli, R., Marsili, S. & Procacci, P. Calculation of the potential of mean force from nonequilibrium measurements via maximum likelihood estimators. *Phys. Rev. E* **77**, 031104 (2008).

[26] Gullingsrud, J., Braun, R. & Schulten, K. Reconstructing potentials of mean force through time series analysis of steered molecular dynamics simulations. *Journal of Computational Physics* **151**, 190–211 (1999).

[27] Brokaw, J. B., Haas, K. R. & Chu, J.-W. Reaction path optimization with holonomic constraints and kinetic energy potentials. *J. Chem. Theory Comput.* **5**, 2050–2061 (2009).

[28] Haas, K. & Chu, J.-W. Decomposition of energy and free energy changes by following the flow of work along reaction path. *J. Chem. Phys.* **131**, 144105–144105–11 (2009).

[29] Liphardt, J. Reversible Unfolding of Single RNA Molecules by Mechanical Force. *Science* **292**, 733–737 (2001).

[30] Yu, H. *et al.* Energy landscape analysis of native folding of the prion protein yields the diffusion constant, transition path time, and rates. *Proc. Natl. Acad. Sci. USA* **109**, 14452–14457 (2012).

[31] Hummer, G. & Szabo, A. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci. USA* **98**, 3658 (2001).

[32] Förster, T. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys.* **437**, 55–75 (1948).

[33] Yang, H. *et al.* Protein conformational dynamics probed by single-molecule electron transfer. *Science* **302**, 262–266 (2003).

[34] Gopich, I. V. & Szabo, A. Single-Macromolecule Fluorescence Resonance Energy Transfer and Free-Energy Profiles. *J. Phys. Chem. B* **107**, 5058–5063 (2003).

[35] Hanson, J. A., Brokaw, J., Hayden, C. C., Chu, J.-W. & Yang, H. Structural distributions from single-molecule measurements as a tool for molecular mechanics. *Chemical Physics* **396**, 61–71 (2012).

[36] Teo, C., Vishwanthan, S., Smola, A. & Le, Q. Bundle methods for regularized risk minimization. *The Journal of Machine Learning Research* **11**, 311–365 (2010).

[37] Gardiner, C. W. *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences* (Springer Verlag, 2004).

[38] Tomé, T. & de Oliveira, M. J. Entropy production in irreversible systems described by a Fokker-Planck equation. *Phys. Rev. E* **82**, 021120 (2010).

[39] Latora, V. & Baranger, M. Kolmogorov-Sinai entropy rate versus physical entropy. *Phys. Rev. Lett.* **82**, 520–523 (1999).

[40] Pattanayak, A. Lyapunov exponents, entropy production, and decoherence. *Phys. Rev. Lett.* **83**, 4526–4529 (1999).

[41] Kleeman, R. Information theory and dynamical system predictability. *Entropy* **13**, 612–649 (2011).

[42] Plastino, A. R. & Plastino, A. Non-extensive statistical mechanics and generalized Fokker-Planck equation. *Phys. A* **222**, 347–354 (2002).

[43] Gaspard, P. Time-reversed dynamical entropy and irreversibility in Markovian random processes. *J Stat Phys* **117**, 599–615 (2004).

[44] Gómez-Gardeñes, J. & Latora, V. Entropy rate of diffusion processes on complex networks. *Phys. Rev. E* **78**, 065102 (2008).

[45] Ekroot, L. & Cover, T. M. The entropy of Markov trajectories. *Information Theory, IEEE Transactions on* **39**, 1418–1421 (1993).

[46] Jaynes, E. The minimum entropy production principle. *Annu. Rev. Phys. Chem.* **31**, 579–601 (1980).

[47] Paluš, M. On entropy rates of dynamical systems and Gaussian processes. *Phys. Lett. A* **227**, 301–308 (1997).

[48] Luschgy, H. & Pagès, G. Sharp asymptotics of the Kolmogorov entropy for Gaussian measures. *J. Func. Anal.* **212**, 89–120 (2004).

[49] Bach, A. & Dürr, D. Entropy density in function space and the Onsager-Machlup function. *Phys. Lett. A* **69**, 244–246 (1978).

[50] Roldán, É. & Parrondo, J. Estimating Dissipation from Single Stationary Trajectories. *Phys. Rev. Lett.* **105**, 150607 (2010).

[51] Onsager, L. & Machlup, S. Fluctuations and irreversible processes. *Physical Review* **91**, 1505–1512 (1953).

[52] Adib, A. B. Stochastic Actions for Diffusive Dynamics: Reweighting, Sampling, and Minimization . *J. Phys. Chem. B* **112**, 5910–5916 (2008).

[53] Hunt, K. & Ross, J. Path integral solutions of stochastic equations for nonlinear irreversible processes: The uniqueness of the thermodynamic Lagrangian. *J. Chem. Phys.* **75**, 976 (1981).

[54] Turgeman, L., Carmi, S. & Barkai, E. Fractional Feynman-Kac equation for non-brownian functionals. *Phys. Rev. Lett.* **103**, 190201 (2009).

[55] Golan, A., Judge, G. & Karp, L. A maximum entropy approach to estimation and inference in dynamic models or Counting fish in the sea using maximum entropy. *Journal of Economic Dynamics and Control* **20**, 559–582 (1996).

[56] Wu, D. *et al.* Trajectory Approach to Two-State Kinetics of Single Particles on Sculpted Energy Landscapes. *Phys. Rev. Lett.* **103**, 034119 (2009).

[57] Kastoryano, M. J., Wolf, M. M. & Eisert, J. Precisely Timing Dissipative Quantum Information Processing. *Phys. Rev. Lett.* **110**, 110501 (2013).

[58] Cornish, P. V. & Ha, T. A Survey of Single-Molecule Techniques in Chemical Biology. *ACS Chem. Biol.* **2**, 53–61 (2007).

[59] Merchant, K. A., Best, R. B., Louis, J. M., Gopich, I. V. & Eaton, W. A. Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc. Natl. Acad. Sci. USA* **104**, 1528–1533 (2007).

[60] Watkins, L. P. & Yang, H. Information Bounds and Optimal Analysis of Dynamic Single Molecule Measurements. *Biophys. J.* **86**, 4015–4029 (2004).

[61] Chodera, J. D. *et al.* Bayesian hidden Markov model analysis of single-molecule force spectroscopy: Characterizing kinetics under measurement uncertainty. *arXiv* (2011).

[62] Backović, M., Price, E. S., Johnson, C. K. & Ralston, J. P. A distribution-based method to resolve single-molecule Förster resonance energy transfer observations. *J. Chem. Phys.* **134**, 145101 (2011).

[63] Watkins, L. P., Chang, H. & Yang, H. Quantitative Single-Molecule Conformational Distributions: A Case Study with Poly-(l-proline). *J. Phys. Chem. A* **110**, 5191–5203 (2006).

[64] Gopich, I. V., Nettels, D., Schuler, B. & Szabo, A. Protein dynamics from single-molecule fluorescence intensity correlation functions. *J. Chem. Phys.* **131**, 095102 (2009).

[65] Witkoskie, J. B. & Cao, J. Analysis of the Entire Sequence of a Single Photon Experiment on a Flavin Protein . *J. Phys. Chem. B* **112**, 5988–5996 (2008).

[66] Zucchini, W. & MacDonald, I. L. *Hidden Markov Models for Time Series*. An Introduction Using R (Chapman and Hall/CRC, 2009).

[67] Bronson, J. E., Fei, J., Hofman, J. M., Gonzalez, R. L., Jr & Wiggins, C. H. Learning Rates and States from Biophysical Time Series: A Bayesian Approach to Model Selection and Single-Molecule FRET Data. *Biophys. J.* **97**, 3196–3205 (2009).

[68] Bronson, J. E., Hofman, J. M., Fei, J., Gonzalez, R. L. & Wiggins, C. H. Graphical models for inferring single molecule dynamics. *BMC Bioinformatics* **11**, S2 (2010).

[69] Liu, Y., Park, J., Dahmen, K. A., Chemla, Y. R. & Ha, T. A Comparative Study of Multivariate and Univariate Hidden Markov Modelings in Time-Binned Single-Molecule FRET Data Analysis. *J. Phys. Chem. B* **114**, 5386–5403 (2010).

[70] E, W., Ren, W. & Vanden-Eijnden, E. Finite Temperature String Method for the Study of Rare Events . *J. Phys. Chem. B* **109**, 6688–6693 (2005).

[71] Hanson, J., Alber, T. & Yang, H. Dynamic active-site protection by the M. tuberculosis protein tyrosine phosphatase PtpB lid domain. *Journal of the American ...* (2010).

[72] Cao, J. & al, e. Generic Schemes for Single-Molecule Kinetics. 1: Self-Consistent Pathway Solutions for Renewal Processes. *J. Phys. Chem. B* **112**, 12867–12880 (2008).

[73] Gopich, I. & Szabo, A. Theory of photon statistics in single-molecule Förster resonance energy transfer. *J. Chem. Phys.* **122**, 014707 (2005).

[74] Song, L., Boots, B., Siddiqi, S., Gordon, G. & al, e. Hilbert space embeddings of conditional distributions with applications to dynamical systems. *Citeseer* 961–968 (2009).

[75] Sakurai, J. J. *Modern Quantum Mechanics* (Addison Wesley Publishing Company, 1985).

[76] Hughes, R. I. G. Quantum Mechanics (1989).

[77] Pozrikidis, C. *Introduction To Finite And Spectral Element Methods Using Matlab* (CRC Press, 2005).

[78] Turner, R. ScienceDirect.com - Computational Statistics & Data Analysis - Direct maximization of the likelihood of a hidden Markov model. *Computational Statistics & Data Analysis* (2008).

[79] Haas, K. R., Yang, H. & Chu, J.-W. Elements of the Trajectory Entropy in Continuous Stochastic Processes at Equilibrium. *Phys. Rev. Lett.* Under Review (2013).

[80] Little, R. J. A. & Rubin, D. B. *Statistical analysis with missing data* (Wiley-Interscience, 2002).

[81] Wilcox, R. M. Exponential operators and parameter differentiation in quantum physics. *J. Math. Phys.* **8**, 962 (1967).

[82] Nocedal, J. & Wright, S. J. *Numerical Optimization* (Springer, 2006).

[83] Watkins, L. P. & Yang, H. Detection of Intensity Change Points in Time-Resolved Single-Molecule Measurements. *J. Phys. Chem. B* **109**, 617–628 (2005).

[84] Bayarri, M. & García Donato, G. Generalization of Jeffreys divergence-based priors for Bayesian hypothesis testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**, 981–1003 (2008).

[85] Xu, L. & Jordan, M. I. On Convergence Properties of the EM Algorithm for Gaussian Mixtures. *Neural Computation* **8**, 129–151 (1996).

[86] FIRTH, D. Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27–38 (1993).

[87] Sakurai, E. & Yamanishi, K. Comparison of dynamic model selection with infinite HMM for statistical model change detection 302–306 (2012).

[88] Xing, J., Wang, H. & Oster, G. From Continuum Fokker-Planck Models to Discrete Kinetic Models. *Biophys. J.* **89**, 1551–1563 (2005).

[89] Krishnamurthy, V. & Moore, J. B. On-line estimation of hidden Markov model parameters based on the Kullback-Leibler information measure. *Signal Processing, IEEE Transactions on* **41**, 2557–2573 (1993).

[90] Uphoff, S. *et al.* Monitoring multiple distances within a single molecule using switchable FRET. *Nat Meth* **7**, 831–836 (2010).

[91] Chodera, J. D. & Pande, V. S. Splitting probabilities as a test of reaction coordinate choice in single-molecule experiments. *Phys. Rev. Lett.* **107**, 098102 (2011).

[92] Mossa, A., de Lorenzo, S., Huguet, J. M. & Ritort, F. Measurement of work in single-molecule pulling experiments. *J. Chem. Phys.* **130**, 234116 (2009).

[93] Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **314**, 141–151 (1999).

[94] Wells, B. H., Smith, E. B. & Zare, R. N. The stability of the RKR inversion procedure to errors in the spectroscopic data: origin of the inner-wall ripple. *Chemical Physics Letters* **99**, 244–249 (1983).

[95] Wang, J., Xu, L. & Wang, E. Potential landscape and flux framework of nonequilibrium networks: Robustness, dissipation, and coherence of biochemical oscillations. *Proc. Natl. Acad. Sci. USA* **105**, 12271 (2008).

[96] Hummer, G. & Szabo, A. Free energy profiles from single-molecule pulling experiments. *Proc. Natl. Acad. Sci. USA* **107**, 21441–21446 (2010).

[97] Mittermaier, A. & Kay, L. E. Review - New tools provide new insights in NMR studies of protein dynamics. *Science* **312**, 224–228 (2006).

[98] Hanson, J. A. & Yang, H. Quantitative in vitro Biochemistry, one molecule at a time. In Wall, M. E. (ed.) *Quantitative Biology: From Molecular to Cellular Systems* (Taylor and Francis Group, New York, 2012).

[99] Bryngelson, J. D. & Wolynes, P. G. Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J. Phys. Chem. B* **93**, 6902–6915 (1989).

[100] Krivov, S. & Karplus, M. Diffusive reaction dynamics on invariant free energy profiles. *Proc. Natl. Acad. Sci. USA* **105**, 13841 (2008).

[101] Dellago, C., Bolhuis, P. G., Csajka, F. S. & Chandler, D. Transition path sampling and the calculation of rate constants. *J. Chem. Phys.* **108**, 1964 (1998).

[102] Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **5**, 789–796 (2009).

[103] Johnson, M. E. & Hummer, G. Characterization of a dynamic string method for the construction of transition pathways in molecular reactions. *J. Phys. Chem. B* **116**, 8573–8583 (2012).

[104] Brooks, B. *et al.* CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **30**, 1545–1614 (2009).

[105] Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).

[106] Schlick, T., Collepardo-Guevara, R., Halvorsen, L. A., Jung, S. & Xiao, X. Biomolecular modeling and simulation: a field coming of age. *Q. Rev. Biophys.* **44**, 191–228 (2011).

[107] Boehr, D. D., Dyson, H. J. & Wright, P. E. An NMR perspective on enzyme dynamics. *Chem. Rev.* **106**, 3055–3079 (2006).

[108] Palmer, A. G. NMR characterization of the dynamics of biomacromolecules. *Chem. Rev.* **104**, 3623–3640 (2004).

[109] Tzeng, S.-R. & Kalodimos, C. G. Protein dynamics and allostery: an NMR view. *Curr. Opin. Struct. Biol.* **21**, 62–67 (2011).

[110] Ha, T. *et al.* Single-molecule fluorescence spectroscopy of enzyme conformational dynamics and cleavage mechanism. *Proc. Natl. Acad. Sci. USA* **96**, 893–898 (1999).

[111] Yang, H. Model-Free Statistical Reduction of Single-Molecule Time Series. In Barkai, E., Brown, F. L., Orrit, M. & Yang, H. (eds.) *Theory and Evaluation of Single-Molecule Signals* (Scientific Publishing, 2008).

[112] Yang, H. Change-Point Localization and Wavelet Spectral Analysis of Single-Molecule Time Series. *Adv. Chem. Phys.* **146**, 129–143 (2011).

[113] Fisher, R. A. Cambridge Journals Online - Mathematical Proceedings of the Cambridge Philosophical Society - Abstract - Theory of Statistical Estimation. *Mathematical Proceedings of the Cambridge Philosophical Society* **22**, 700–725 (1925).

[114] Crooks, G. E. Measuring thermodynamic length. *Phys. Rev. Lett.* **99**, 100602 (2007).

[115] Frieden, B., Plastino, A., Plastino, A. & Soffer, B. Fisher-based thermodynamics: Its Legendre transform and concavity properties. *Phys. Rev. E* **60**, 48–53 (1999).

[116] Plastino, A. Fisher's information and the arrow of time. In *The twentieth International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, 38–48 (AIP).

[117] Dacunha-Castelle, D. & Florens-Zmirou, D. Estimation of the coefficients of a diffusion from discrete observations. *Stochastics* **19**, 263–284 (1986).

[118] Biane, P. & Speicher, R. Free diffusions, free entropy and free Fisher information. *Ann. Inst. Henri Poincare, B* **37**, 581–606 (2001).

[119] Haken, H. A new access to path integrals and Fokker Planck equations via the maximum calibre principle. *Z. Phys. B* **63**, 505–510 (1986).

[120] Tsallis, C. & Bukman, D. J. Anomalous diffusion in the presence of external forces: Exact time-dependent solutions and their thermostatistical basis. *Phys. Rev. E* **54**, 2197–2200 (1996).

[121] Stock, G., Ghosh, K. & Dill, K. A. Maximum Caliber: A variational approach applied to two-state dynamics. *J. Chem. Phys.* **128**, 194102 (2008).

[122] Plastino, A. R. & Plastino, A. Symmetries of the Fokker-Planck equation and the Fisher-Frieden arrow of time. *Phys. Rev. E* **54**, 4423–4426 (1996).

[123] Borland, L., Pennini, F., Plastino, A. R. & Plastino, A. The nonlinear Fokker-Planck equation with state-dependent diffusion-A nonextensive maximum entropy approach. *EPJ B* **12**, 285–297 (1999).

[124] Ganguli, S., Huh, D. & Sompolinsky, H. Memory traces in dynamical systems. *Proc. Natl. Acad. Sci. USA* **105**, 18970–18975 (2008).

[125] Luo, S. Maximum Shannon entropy, minimum Fisher information, and an elementary game. *Found. Phys.* **32**, 1757–1772 (2002).

[126] Drozdov, A. & Brey, J. Operator expansions in stochastic dynamics. *Phys. Rev. E* **57**, 1284–1289 (1998).

[127] Smola, A. J. & Kondor, R. Kernels and Regularization on Graphs. In *Learning Theory and Kernel Machines*, 144–158 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2003).

[128] Calmet, X. & Calmet, J. Dynamics of the Fisher information metric. *Phys. Rev. E* **71**, 056109 (2005).

[129] Bastian, M. R., Gunther, J. H. & Moon, T. K. A Simplified Natural Gradient Learning Algorithm. *Adv. Artificial Neural Syst.* **2011**, 1–9 (2011).

[130] Chandler, D. *Introduction To Modern Statistical Mechanics* (Oxford University Press, USA, 1987).

[131] Thiébaut, E. & Conan, J. M. Strict a priori constraints for maximum-likelihood blind deconvolution. *J. Opt. Soc. Am. A* **12**, 485 (1995).

[132] Reynolds, D. E. Construction of coarse-grained order parameters in nonequilibrium systems. *Phys. Rev. E* **79**, 061107 (2009).

[133] Pinsky, M. *Introduction to Fourier Analysis and Wavelets* (Amer Mathematical Society, 2002).

[134] Romera, E., Sánchez-Moreno, P. & Dehesa, J. S. Uncertainty relation for Fisher information of D-dimensional single-particle systems with central potentials. *J. Math. Phys.* **47**, 103504 (2006).

[135] Dekker, H. On the theory of activation: Kramers' reaction rate in bistable systems. *Phys. Lett. A* **112**, 197–200 (1985).

[136] Sánchez, I. E. & Kiefhaber, T. Non-linear rate-equilibrium free energy relationships and Hammond behavior in protein folding. *Biophys. Chem.* **100**, 397–407 (2003).

[137] Xiao-Yu, C. Perturbation theory of von Neumann entropy. *Chinese Phys. B* **19**, 040308– (2010).

[138] Garbaczewski, P. Differential entropy and time. *Entropy* **7**, 253–299 (2005).

[139] Sinai, I. On the Concept of Entropy for a Dynamic System. *Dokl Akad Nauk Sssr+* **124**, 768–771 (1959).

[140] Anantharaman, N. Entropy and the localization of eigenfunctions. *Ann. Math* **168**, 435–475 (2008).

[141] Pesin, Y. B. Characteristic Lyapunov exponents and smooth ergodic theory. *Russian Mathematical Surveys* **32**, 55–114 (1977).

[142] Bologna, M., Grigolini, P., Karagiorgis, M. & Rosa, A. Trajectory versus probability density entropy. *Phys. Rev. E* **64**, 016223 (2001).

[143] Pechukas, P. Kolmogorov entropy and" quantum chaos". *J. Phys. Chem.* **86**, 2239–2243 (1982).

[144] Rebolledo, R. & Orszag, M. *Quantum Probability and Related Topics*. Proceedings of the 30th Conference, Santiago, Chile, 23-28 November 2009 (World Scientific Publishing Company, 2011).

[145] Gianazza, U., Savaré, G. & Toscani, G. The Wasserstein Gradient Flow of the Fisher Information and the Quantum Drift-diffusion Equation. *Arch. Rational Mech. Anal.* **194**, 133–220 (2008).

[146] Ye, D. On the Bures volume of separable quantum states. *J. Math. Phys.* **50**, 083502–083502–14 (2009).

[147] Hübner, M. Explicit computation of the Bures distance for density matrices. *Phys. Lett. A* **163**, 239–242 (1992).

[148] Watkins, L. P., Chang, H. & Yang, H. Quantitative Single-Molecule Conformational Distributions: A Case Study with Poly-L-proline. *J. Phys. Chem. A* **110**, 5191–5203 (2006).

[149] Skilling, J. & Bryan, R. Maximum entropy image reconstruction-general algorithm. *Monthly Notices of the Royal Astronomical Society* **211**, 111 (1984).

[150] Hummer, G. Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations. *New J. Phys.* **7**, 34–34 (2005).

[151] Ensign, D. & Pande, V. Bayesian inference for Brownian dynamics. *Phys. Rev. E* **82** (2010).

[152] Smola, A., Vishwanathan, S. & Le, Q. Bundle methods for machine learning. *Advances in neural information processing systems* **20**, 1377–1384 (2007).

[153] Haas, K. R., Yang, H. & Chu, J.-W. Fisher Information Metric for the Langevin Equation and Least Informative Models of Continuous Stochastic Dynamics. *J. Chem. Phys.* Under Review (2013).

[154] Van de Vosse, F. & al, e. Spectral elements methods: Theory and applications. *alexandria.tue.nl* (1996).

[155] Haas, K. R., Yang, H. & Chu, J.-W. Path Integral Statistical Learning for Continuous Stochastic Dynamics: Expectation-Maximization Extraction of the Potential of Mean Force and Diffusion Coefficient of Protein Dynamics via Single-Molecule Förster resonance energy transfer (smFRET). *J. Phys. Chem. B* Under Review (2013).

[156] Jordan, M. I. Lecture 16 Laplace approximation review. *Stat260* 1–5 (2010).

[157] Sivak, D. A. & Crooks, G. E. Thermodynamic Metrics and Optimal Paths. *Phys. Rev. Lett.* **108**, 190602 (2012).

[158] Daily, M. D., Phillips, G. N., Jr & Cui, Q. Many Local Motions Cooperate to Produce the Adenylate Kinase Conformational Transition. *Journal of Molecular Biology* **400**, 618–631 (2010).

[159] Liu, P., Shi, Q., Lyman, E. & Voth, G. A. Reconstructing atomistic detail for coarse-grained models with resolution exchange. *J. Chem. Phys.* **129**, 114103 (2008).

[160] Peters, B., Bolhuis, P. G., Mullen, R. G. & Shea, J.-E. Reaction coordinates, one-dimensional Smoluchowski equations, and a test for dynamical self-consistency. *J. Chem. Phys.* **138**, 054106 (2013).

[161] Welsher, K. & Yang, H. Model-free analysis of time-dependent single-molecule spectroscopy: Dynamics of biological macromolecules. *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on* 921–923 (2012).

[162] Gin, B. C., Garrahan, J. P. & Geissler, P. L. The limited role of nonnative contacts in the folding pathways of a lattice protein. *Journal of Molecular Biology* **392**, 1303–1314 (2009).

[163] Maisuradze, G. G., Liwo, A. & Scheraga, H. A. How adequate are one-and two-dimensional free energy landscapes for protein folding dynamics? *Phys. Rev. Lett.* **102**, 238102 (2009).

[164] Comput Stat Data An 1996 RydenAn EM algorithm for estimation 1–17 (2003).

[165] Gibson, M. A. & Bruck, J. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A* **104**, 1876–1889 (2000).

[166] Svihlik, J. Model Parameters Estimation Using Jeffrey Divergence. *Radioelektronika* (2007).