Fisher information metric for the Langevin equation and least informative models of continuous stochastic dynamics

Kevin R. Haas, Haw Yang, and Jhih-Wei Chu

Articles you may be interested in
Lagrangian dynamics in stochastic inertia-gravity waves
Phys. Fluids **22**, 126601 (2010); 10.1063/1.3518137

Relation between Langevin type equation driven by the chaotic force and stochastic differential equation
AIP Conf. Proc. **519**, 359 (2000); 10.1063/1.1291585

From the Langevin equation to the fractional Fokker–Planck equation
AIP Conf. Proc. **502**, 375 (2000); 10.1063/1.1302409

Stochastic modelling of nonlinear dynamical systems
AIP Conf. Proc. **502**, 313 (2000); 10.1063/1.1302402

High-accuracy discrete path integral solutions for stochastic processes with noninvertible diffusion matrices. II. Numerical evaluation
J. Chem. Phys. **107**, 3505 (1997); 10.1063/1.474690

Fisher information metric for the Langevin equation and least informative models of continuous stochastic dynamics

Kevin R. Haas,[1] Haw Yang,[2,a)] and Jhih-Wei Chu[1,3,4,a)]

[1]*Department of Chemical and Biomolecular Engineering, University of California-Berkeley, Berkeley, California 94720, USA*
[2]*Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA*
[3]*Department of Biological Science and Technology, National Chiao Tung University, Hsinchu, Taiwan*
[4]*Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsinchu, Taiwan*

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 any arbitrary potential of mean force where the parameter set is now the full space dependent function. Leveraging an innovative Hermitian form of the corresponding Fokker-Planck 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. © *2013 AIP Publishing LLC.*
[http://dx.doi.org/10.1063/1.4820491]

I. 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.[1–3] 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.[4–6] On top of the deterministic potential of mean force (PMF), the Langevin model incorporates the solvent-induced stochastic fluctuations via the diffusion coefficient[7] to describe system dynamics along a set of chosen coordinates or order parameters.[8–10] 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).[11–13] 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)[14–16] 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)[17–19] do away with the issues of ensemble averaging but face the challenge of the low signal-to-noise ratio convoluted with photon-counting statistics.[20, 21] As a result, not only the data analysis in each category of these measurements is complicated, but also the systematic combination of information across different techniques is a challenging issue.

We reason that the foundation for quantitatively integrating the knowledge from different data types could be based on an information measure of dynamics parameters from the recorded time trajectories. In this regard, the Fisher information provides a framework with a clear statistical picture and straightforward linkage to thermodynamics;[22–24] 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.[25]

In order to extract maximum information of dynamics from time-dependent data, we aim to quantify the Fisher information of trajectories (FIT).[26, 27] This approach of evaluating information content in the path space is different from the typical methods of tracking the statistics of dynamic

---

a)Authors to whom correspondence should be addressed. Electronic addresses: hawyang@princeton.edu and jwchu@nctu.edu.tw

variables,[28,29] including the rate of changes of Fisher information matrices in the space of a single or a few time slices.[30,31] In this work, we devise numerical and analytical methods to determine the Fisher information of the PMF and diffusion coefficient in the trajectories of Langevin dynamics directly without the need of performing Monte Carlo simulations.[27] As a result, an explicit connection between parameters in the equation of motion and FIT is established, which demonstrates how dynamics information is encoded through time propagation.

Since FIT quantifies the disorder inscribed in the probability function of the data, or the likelihood function of the 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.[32–34] Therefore, minimizing FIT with such constraints would give the least informative dynamics (LID) model. For retrieving parameters relating to time propagation, this constrained optimization approach has been addressed for Markov states systems.[35,36] 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 Sec. II with our rationale of selecting the basis of representation discussed in Sec. III. Section IV outlines the numerical procedure we developed to calculate FIT for continuous stochastic dynamics and Sec. V derives the analytical form of FIT in the continuum limit. Section VI applies the analytic form of FIT for measuring information content in trajectories and the result of which is used in Sec. VII to derive the least informative dynamics under various constraints of based on the Langevin equation followed by our conclusion.

II. 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 model parameters.[23,37,38] 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.[26]

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 $p_{eq}(x) \propto \exp(-V(x)/k_B T)$. A trajectory, $X(t); t \in [0, t_{obs}]$, in this case is a continuous but non-differentiable function of time. In a measurement, this stochastic trajectory is generally recorded at specific instances separated by a time resolution Δt to create a vector $\vec{X}_t = [X(0), X(\Delta t), X(2\Delta t), \ldots, X(t_{obs})]$. This vector exists in a trajectory space of dimensionality $N = t_{obs}/\Delta t$ with t_{obs} the duration of observation and the coordinates denoted as the set $\{x_\tau | \tau = 0, 1, 2, \ldots, N\}$. Based on this setup, we aim to find the Fisher information of the deterministic and stochastic components in the Langevin equation in a multidimensional vector space. Then, $\lim_{\Delta t \to 0}$ 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 function. 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 the components in $\vec{\theta}$

$$\mathcal{I}_{i,j}(\vec{\theta}) = \mathbb{E}_{\vec{X}_t}\left[\frac{\partial \ln P(\vec{X}_t)}{\partial \theta_i}\frac{\partial \ln P(\vec{X}_t)}{\partial \theta_j}\bigg|\vec{\theta}\right]. \quad (1)$$

The $\mathbb{E}_{\vec{X}_t}[\cdot]$ in the above equation denotes the expectation evaluated by the 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,

$$P(\vec{X}_t) = p(x_0)\prod_{\tau=0}^{N-1} p(x_{\tau+1}|x_\tau). \quad (2)$$

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

$$\mathcal{I}_{ij} = \sum_{\tau,\tau'=-1}^{t_{obs}/\Delta t}\int D\vec{X}_t\left[\frac{\partial \ln p(x_{\tau'+1}|x_{\tau'})}{\partial \theta_i}\frac{\partial \ln p(x_{\tau+1}|x_\tau)}{\partial \theta_j}\right]P(\vec{X}_t). \quad (3)$$

The contribution from $p_{eq}(x_0)$ is included by setting $p(x_0|x_{-1}) = p_{eq}(x_0)$. For the path integral in Eq. (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. (3) contribute zero to $\mathcal{I}_{ij}$ due to the ability to isolate a normalization condition of

the conditional probability densities $p(x_{\tau+1}|x_\tau)$, $\tau = 0 \ldots N$:

$$\int dx \frac{\partial \ln p(x)}{\partial \theta} p(x) = \frac{\partial \int dx p(x)}{\partial \theta} = 0. \qquad (4)$$

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, $I_{\Delta t}$:

$$I_{eq} = \int dx_0 \frac{\partial \ln p_{eq}(x_0)}{\partial \vec{\theta}} \frac{\partial \ln p_{eq}(x_0)}{\partial \vec{\theta}} p_{eq}(x_0), \qquad (5)$$

$$I_{\Delta t} = \int dx_t dx_0 \frac{\partial \ln p(x_t|x_0)}{\partial \vec{\theta}} \frac{\partial \ln p(x_t|x_0)}{\partial \vec{\theta}} p(x_t, x_0). \qquad (6)$$

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

$$\mathcal{I} = I_{eq} + \frac{t_{obs}}{\Delta t} I_{\Delta t}. \qquad (7)$$

The connection of Eqs. (5)–(7) between an entire path integral and an integral over two closely spaced time points is a general result of Markovian dynamics. Therefore, if the FIT for a handful of parameters is of interest, the values can be calculated via a short-time expansion of the time propagator with Monte Carlo simulations for computing the expectation.[27] For cases that the parameterization of dynamics is in the space of continuous functions of infinite dimensionality, though, such as for quantifying the information of PMF in the Langevin equation, the approach of brute-force sampling becomes impractical. However, this difficulty could be resolved if an analytical expression of FIT was available. One of the main results of this work is Eq. (48) for the FIT of Langevin dynamics. This expression then allows a line integral to be performed in the parameter space for quantifying the dissipation measure of information (Eqs. (54) and (55)).

The calculation of FIT via Eq. (7) relies on evaluating the derivatives and integrals defined for the static distribution in Eq. (5) and the dynamic propagator in Eq. (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.[23,37,38] We found that the most natural basis for the deterministic components of the Langevin equation is the square root of the equilibrium probability density[39] (*vide infra*):

$$\rho_{eq}(x) = \sqrt{p_{eq}(x)} \propto \sqrt{\exp\left(\int^x F(x)/k_B T\right)}. \qquad (8)$$

Therefore, we now define Fisher information with respect to a composite quantity of the functions for density and diffusion, i.e., $\vec{\theta} \rightarrow \{\rho_{eq}(x), D(x)\}$. The Fisher information thus becomes functional derivatives and is a scalar field over the arguments (x, y) of the parameterizing functions rather than a matrix.

We start the evaluation of FIT from the equilibrium term

$$I_{eq}(x, y) = \int dx_0 \frac{\delta \ln p_{eq}(x_0)}{\delta \rho_{eq}(x)} \frac{\delta \ln p_{eq}(x_0)}{\delta \rho_{eq}(y)} p_{eq}(x_0). \qquad (9)$$

Since $p_{eq}(x) = \rho_{eq}^2(x)$, the functional derivatives in Eq. (9) are

$$\frac{\delta \ln \rho_{eq}^2(x_0)}{\delta \rho_{eq}(x)} = 2 \frac{\delta(x_0 - x)}{\rho_{eq}(x_0)}. \qquad (10)$$

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

$$I_{eq} = 4 \int dx_0 \delta(x_0 - x)\delta(x_0 - y) = 4\delta(y - x). \qquad (11)$$

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

As will be shown later, the property of $\rho_{eq}(x)$ in making I_{eq} constant also facilitates the evaluation of $I_{\Delta t}$ defined in Eq. (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 $\rho_{eq}(x)$:

$$I_{\Delta t}(x, y) = \int dx_t dx_0 \frac{\delta \ln p(x_t|x_0)}{\delta \rho_{eq}(x)} \frac{\delta \ln p(x_t|x_0)}{\delta \rho_{eq}(y)} p(x_t, x_0). \qquad (12)$$

This Fisher information matrix is a rank 2 tensor field over the space coordinate.

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 $p(x_t|x_0)$:

$$\frac{\partial p(x_t|x_0)}{\partial t} = -\nabla \cdot J(x_t) \qquad (13)$$

and

$$J(x_t) = -\left(D(x_t)\nabla p(x_t|x_0) - \frac{D(x_t)F(x_t)}{k_B T} p(x_t|x_0)\right). \qquad (14)$$

In this formula, 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 boundary conditions are zero flux ($J(\text{Boundary}) = 0$) for conserving the total probability.

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

$$\frac{\partial p(x_t|x_0)}{\partial t} = \nabla \cdot \left(D(x_t)p_{eq}(x_t)\nabla \frac{p(x_t|x_0)}{p_{eq}(x_t)}\right). \qquad (15)$$

Next, the unique features of expressing the FPE via the equilibrium density $\rho_{eq}(x)$ in Eq. (8) are discussed in preparing for the evaluation of $I_{\Delta t}(x, y)$.

III. THE HERMITIAN OPERATOR OF THE FPE OF LANGEVIN DYNAMICS

With $\rho_{eq}(x)$ being the square root of $p_{eq}(x)$, the probability density of time propagation can be symmetrized

in time:

$$\rho(x_t, x_0) = p(x_t|x_0)\sqrt{\frac{p_{eq}(x_0)}{p_{eq}(x_t)}}, \qquad (16)$$

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

$$P(\vec{X}_t) = \rho_{eq}(x_0) \prod_{\tau=0}^{N-1} \rho(x_{\tau+1}, x_\tau)\rho_{eq}(x_N). \qquad (17)$$

Expressing Eq. (12) via the symmetric time propagator[37] gives the factors of $I_{\Delta t}(x, y)$:

$$I_{\Delta t} = \overbrace{\int dx_t dx_0 \frac{\delta \ln \rho(x_t, x_0)}{\delta \rho_{eq}(x)} \frac{\delta \ln \rho(x_t, x_0)}{\delta \rho_{eq}(y)} p(x_t, x_0)}^{I_{\Delta t}^\rho}$$
$$+ 2\delta(x - y) - 2\rho(x, y). \qquad (18)$$

The details of deriving Eq. (18) are in Appendix A.

In the continuum limit of $\Delta t \to 0^+$, the fact that $\rho(x_t, x_0)$ approaches $\delta(x_t - x_0)$ leads to cancellation of the last two terms in Eq. (18). We will thus focus on the $I_{\Delta t}^\rho$ term, an integration in the two-dimensional space. After expanding the functional logarithms therein, we will evaluate FIT according to

$$I_{\Delta t}^\rho(x, y) = \int dx_t dx_0 \frac{\delta \rho(x_t, x_0)}{\delta \rho_{eq}(x)} \frac{\delta \rho(x_t, x_0)}{\delta \rho_{eq}(y)} \frac{\rho_{eq}(x_t)\rho_{eq}(x_0)}{\rho(x_t, x_0)}. \qquad (19)$$

The FPE of the symmetric propagator $\rho(x_t, x_0)$ can be found by substituting the expression in Eq. (16) into Eq. (15):

$$\frac{\partial \rho(x_t, x_0)}{\partial t} = -\boldsymbol{H}\rho(x_t, x_0),$$
$$\boldsymbol{H} = -\frac{1}{\rho_{eq}(x)} \nabla \cdot \left(D(x)\rho_{eq}^2(x)\nabla \frac{1}{\rho_{eq}(x)} \right). \qquad (20)$$

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

The proof detailed in Appendix B shows that $\boldsymbol{H}$ in the FPE of Eq. (20) is Hermitian. As a result, $\rho(x_{\Delta t}, x_0)$ can be expressed via the matrix elements of $\boldsymbol{H}$ with the Dirac notation

$$\rho(x_{\Delta t}, x_0) = \langle x_{\Delta t}|e^{-\boldsymbol{H}\Delta t}|x_0\rangle. \qquad (21)$$

In Sec. IV, a procedure for calculating $I_{\Delta t}^\rho$ based on the Hermitian version of the FPE of the Langevin equation is developed. This scheme is then employed to evaluate $I_{\Delta t}^\rho$ at different values of Δt. Since the presented formulation is general and does not rely on assuming extreme values of Δt, it can be used to capture the dynamic behaviors in different temporal resolutions as illustrated by the results presented in Sec. IV.

IV. CALCULATIONS OF FIT VIA AN EIGENBASIS EXPANSION

The Hermitian nature of $\boldsymbol{H}$ allows a complete set of orthogonal eigenvectors ψ_i's to be found with real eigenvalues λ_i's such that for each eigenbasis $\langle \psi_i|\boldsymbol{H}|\psi_j\rangle = \lambda_i \delta_{ij}$. Based on this property, we performed an eigen-decomposition of the symmetric operator of the FPE (sFPE) of Eq. (20). In
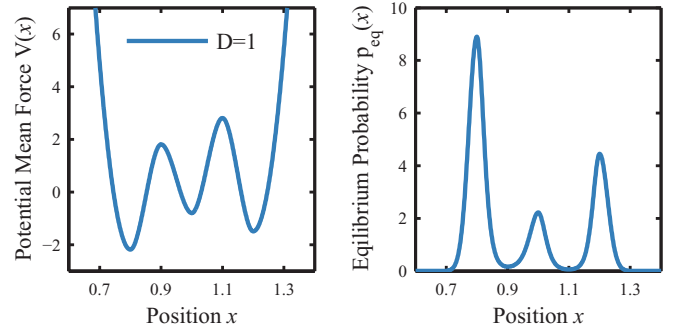

FIG. 1. The reference potential of mean force and the corresponding equilibrium probability density for illustrating the calculations of the Fisher information of trajectories. The model has 3 local minima with approximately 5 k_BT barriers separating these metastable states.

particular, we assume a form for the eigenvectors of $\psi_i(x) = \rho_{eq}(x)\phi_i(x)$ with the modifying function $\phi_i(x)$ determined from the input of ρ_{eq} and diffusion coefficient by using spectral elements[40] to solve the generalized eigenvalue problem $\langle \phi_i \rho_{eq}|\boldsymbol{H}|\rho_{eq}\phi_j\rangle = \lambda_i \delta_{ij} \langle \phi_i \rho_{eq}|\rho_{eq}\phi_j\rangle$. By inserting the completeness property of $\sum_i |\psi_i\rangle\langle\psi_i| = \boldsymbol{1}$ that resolves the identity operator, and noting the orthogonality of the eigenbasis, $\rho(x_{\Delta t}, x_0)$ can now be decomposed as

$$\rho(x_{\Delta t}, x_0) = \sum_{i,j} \langle x_{\Delta t}|\psi_i\rangle\langle\psi_i|e^{-\boldsymbol{H}\Delta t}|\psi_j\rangle\langle\psi_j|x_0\rangle$$
$$= \sum_i \psi_i(x_{\Delta t})e^{-\lambda_i \Delta t}\psi_i(x_0). \qquad (22)$$

The eigenvectors and eigenvalues themselves are found via a spectral finite element method[40] (sFEM) although the results of FIT calculations are invariant to the particular choice of the numerical method as long as the eigenvectors and eigenvalues are determined accurately.

To illustrate the resulting values of $\rho(x_{\Delta t}, x_0)$ for a non-trivial model system that contains three wells separated by 5 k_BT barriers in the PMF shown in Fig. 1, the two-dimensional density field $\rho(x_{\Delta t}, x_0)$ calculated from the eigen-decomposition method discussed above is plotted as log-contours in Fig. 2 at an intermediate time resolution of $\Delta t = 0.025$ s which is also on the same order of the relaxation time ($\Delta t = 0.036$ s) out of the middle intermediate state.

For calculating the FIT according to Eq. (19), the functional derivative of the density field with respect to ρ_{eq} needs to be determined. By employing an arbitrary test function $f(x)$, the functional derivative is defined as

$$\int dx \frac{\delta \rho(x_{\Delta t}, x_0)}{\delta \rho_{eq}(x)} f(x) = \frac{d}{d\varepsilon} \langle x_t|e^{-\Delta t \boldsymbol{H}[\rho_{eq}+\varepsilon f]}|x_0\rangle \Big|_{\varepsilon=0}. \qquad (23)$$

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. (23) is not necessarily factorizable for taking the derivative. To overcome this difficulty, we
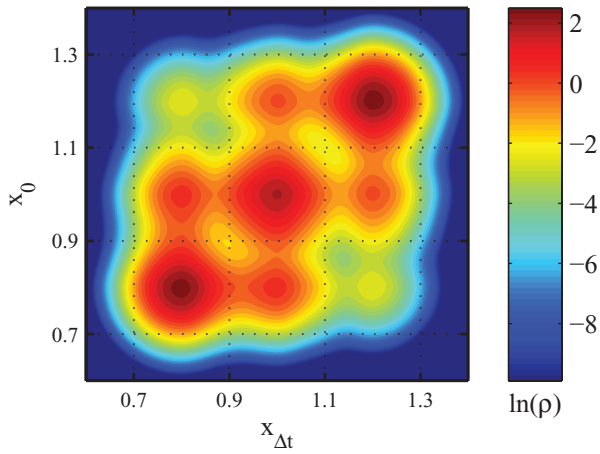
FIG. 2. The contours of $\ln \rho(x_{\Delta t}, x_0)$ at $\Delta t = 0.025$ s for the reference PMF shown in Fig. 1 at $D = 1$ in the corresponding unit. The high-density regions along the diagonal axis around the three local minima are clear. However, off-diagonal densities representing transitions between states within the time window are also clear.

employ the approximate Trotter splitting formalism[41]

$$e^{-\Delta t(\boldsymbol{H}^0 + \varepsilon \boldsymbol{H}')} = e^{-\Delta t \boldsymbol{H}^0/2} e^{-\varepsilon \Delta t \boldsymbol{H}'} e^{-\Delta t \boldsymbol{H}^0/2} + \mathcal{O}(\Delta t^3), \quad (24)$$

wherein the errors due to approximation decrease as Δt goes toward 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. (23) to the approximate form in Eq. (24) leads to

$$\int dx \frac{\delta \rho(x_{\Delta t}, x_0)}{\delta \theta(x)} f(x)$$
$$\approx -\Delta t \langle x_{\Delta t}| e^{-\Delta t \boldsymbol{H}^0/2} \boldsymbol{H}' e^{-\Delta t \boldsymbol{H}^0/2}|x_0\rangle$$
$$= -\Delta t \sum_{i,j} \psi_i(x_{\Delta t}) e^{-\lambda_i \Delta t/2} \langle \psi_i|\boldsymbol{H}'|\psi_j\rangle e^{-\lambda_j \Delta t/2} \psi_j(x_0).$$
$$(25)$$

After inserting the completeness property of the eigenbasis of $\sum_i |\psi_i\rangle\langle\psi_i| = \boldsymbol{1}$ in between all of the operators in the first line of Eq. (25) and applying orthogonality relationship between eigenvectors of $\langle \psi_k|\boldsymbol{H}^0|\psi_i\rangle = \lambda_i \delta_{ik}$, the second line of Eq. (25) can be obtained.

The Hamiltonian of the first order perturbation with respect to ρ_{eq}, $\boldsymbol{H}'$, in Eq. (25) can be found by tracking ε in the total Hamiltonian

$$\boldsymbol{H}[\rho_{\text{eq}} + \varepsilon f]$$
$$= \frac{-1}{(\rho_{\text{eq}} + \varepsilon f)(x)} \nabla \cdot \left(D(x)(\rho_{\text{eq}} + \varepsilon f)^2(x) \nabla \frac{1}{(\rho_{\text{eq}} + \varepsilon f)(x)} \right).$$
$$(26)$$

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

$$\boldsymbol{H}' = \frac{d}{d\varepsilon} \boldsymbol{H}[\rho_{\text{eq}} + \varepsilon f] \bigg|_{\varepsilon=0}, \quad (27)$$

$$\boldsymbol{H}' = -\frac{2}{\rho_{\text{eq}}(x)} \nabla \cdot \left(D(x) f(x) \rho_{\text{eq}}(x) \nabla \frac{1}{\rho_{\text{eq}}(x)} \right)$$
$$+ \frac{f(x)}{\rho_{\text{eq}}^2(x)} \nabla \cdot \left(D(x) \rho_{\text{eq}}^2(x) \nabla \frac{1}{\rho_{\text{eq}}(x)} \right)$$
$$+ \frac{1}{\rho_{\text{eq}}(x)} \nabla \cdot \left(D(x) \rho_{\text{eq}}^2(x) \nabla \frac{f(x)}{\rho_{\text{eq}}^2(x)} \right). \quad (28)$$

With the knowledge of $\boldsymbol{H}'$, the error in Eq. (25) 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. (25). For the first and third line of Eq. (28), 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

$$\langle \psi_i|\boldsymbol{H}'|\psi_j\rangle = \int dx f(x) [2\nabla\phi_j(x) \cdot (D(x)\rho_{\text{eq}}(x)\nabla\phi_i(x))$$
$$- \phi_j(x)\lambda_i\psi_i(x) - \phi_i(x)\lambda_j\psi_j(x)]. \quad (29)$$

Therefore,

$$\frac{\delta\langle\psi_i|\boldsymbol{H}|\psi_j\rangle}{\delta\rho_{\text{eq}}(x)} = 2\nabla\phi_i(x) \cdot (D(x)\rho_{\text{eq}}(x)\nabla\phi_j(x))$$
$$- \rho_{\text{eq}}(x)(\lambda_i + \lambda_j)\phi_i(x)\phi_j(x). \quad (30)$$

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

$$\frac{\delta\langle\psi_i|\boldsymbol{H}|\psi_j\rangle}{\delta D(x)} = \nabla\phi_i(x) \cdot \rho_{\text{eq}}^2(x)\nabla\phi_j(x). \quad (31)$$

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

$$\frac{d\langle\psi_j|\boldsymbol{H}|\psi_i\rangle}{dD} = \frac{\lambda_i}{D}\delta_{ij}. \quad (32)$$

Combining the information in Eqs. (25) and (30), the functional derivative of $\rho(x_t, x_0)$ with respect to $\rho_{\text{eq}}(x)$ can be calculated as

$$\frac{\delta\rho(x_t, x_0)}{\delta\rho_{\text{eq}}(x)}$$
$$\approx -\Delta t \sum_{i,j} \psi_i(x_t) e^{-\lambda_i \Delta t/2} \frac{\delta\langle\psi_i|\boldsymbol{H}|\psi_j\rangle}{\delta\rho_{\text{eq}}(x)} e^{-\lambda_j \Delta t/2} \psi_j(x_0).$$
$$(33)$$

Armed with the knowledge of the functional derivatives from the eigenbasis construction, the key terms of Fisher information in Eq. (19) can be expressed:

$$\frac{I_{\Delta t}^\rho(x, y)}{(\Delta t)^2} = \sum_{i,j} \sum_{k,l} \frac{\delta\langle\psi_i|\boldsymbol{H}|\psi_j\rangle}{\delta\rho_{\text{eq}}(x)} \frac{\delta\langle\psi_k|\boldsymbol{H}|\psi_l\rangle}{\delta\rho_{\text{eq}}(y)} \Omega_{jl}^{ik}(\Delta t).$$
$$(34)$$

For all of the functional dependence on (x_t, x_0) that are integrated away in Eq. (34) is captured in the "overlap" integral

ALGORITHM I. FIT calculation using the eigenbasis resolved via applying a spectral finite element method (sFEM) to solve the Hermitian FPE of Eq. (20).

---

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

---

over two consecutive time slices:

$$\Omega_{jl}^{ik}(\Delta t) \equiv \int dx_{\Delta t} dx_0 \, e^{-(\lambda_i + \lambda_k)\Delta t/2} \psi_i(x_{\Delta t})\psi_k(x_{\Delta t})$$

$$\times \left(\sum_n \phi_n(x_{\Delta t}) e^{-\lambda_n \Delta t} \phi_n(x_0) \right)^{-1}$$

$$\times e^{-(\lambda_j + \lambda_l)\Delta t/2} \psi_j(x_0)\psi_l(x_0). \tag{35}$$

According to Eqs. (34) and (35), Algorithm I 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.

For the reference model shown in Fig. 1, the general approach of eigenbasis expansion presented above is employed to calculate the Fisher information metric of Langevin trajectories at various time resolutions with Algorithm I. Fig. 3 shows a contour plot of $I_{\Delta t}^{\rho}(x, y)/\Delta t$ as well as the cross derivatives with respect to the spatial coordinates of $\rho_{eq}(x)$ and the diffusion coefficient. The matrix element of D alone is also shown as the actual numerical value in the figure. The behaviors of FIT at different time resolutions are discussed in the following to motivate the development of an analytical expression in the continuum limit.

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 of $\rho_{eq}(x)$ that is caused by the barrier located at $x = 1.1$ in the PMF. Therefore, reducing this free energy 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 as that in the previous case, but the off-diagonal negative couplings in the matrix start to emerge. Such pattern indicates that a flatter potential barrier would also result in a less informative time propagator. This result highlights the dependence of FIT on the time resolution used for investigation and the general applicability of our eigenbasis expansion approach as the framework of analysis.

At a high time resolution of $\Delta t \approx 0.002$ s, the importance and details of the underlying equilibrium probability of states vanishes, leading to a nearly tri-banded matrix with positives values along the diagonal terms and negative numbers for the off-diagonal elements. The reason is that FIT begins to capture the diffusion processes within individual wells at this timescale and the specific features of $\rho_{eq}(x)$ are not as prominent as in the previous cases. The origin of this tri-banded
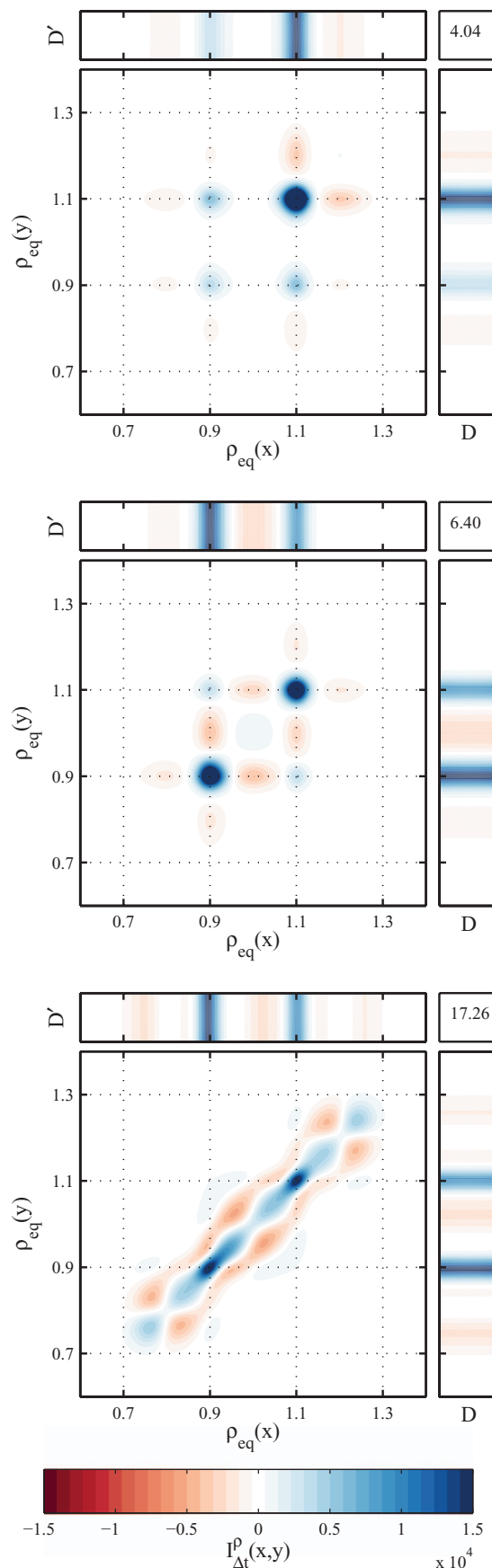

FIG. 3. The contour of the Fisher information matrix of $I_{\Delta t}^{\rho}(x, y)/\Delta t$ for $\Delta t = 0.15, 0.02$, and 0.002 s. The upper band in contour is the Fisher information metric with respect to D and ρ, $\mathcal{I}(D, y)$. Corner value is the Fisher element for diffusion $\mathcal{I}(D, D')$.

nature of FIT becomes explicit in the continuum limit of $\Delta t \to 0^+$.

V. FISHER INFORMATION OF TRAJECTORIES IN THE CONTINUUM LIMIT

In this derivation, it is convenient to employ the following equivalent expression of the Fisher information of $p(x_{\Delta t}|x_0)$:[39]

$$I_{\Delta t}(x, y) = \int dx_{\Delta t} dx_0 \frac{\delta^2 \ln p(x_{\Delta t}|x_0)}{\delta \rho_{\text{eq}}(x) \delta \rho_{\text{eq}}(y)} p(x_{\Delta t}, x_0). \quad (36)$$

In the continuum limit of $\Delta t \to 0^+$, $p(x_{\Delta t}|x_0)$ approaches the delta function $\delta(x_t - x_0)$ which has no dependence on $\rho_{\text{eq}}(x)$; therefore, the functional derivatives vanish and $I_{\Delta t}$ reaches zero in the continuum limit. On the other hand, the denominator of the FIT term of $I_{\Delta t}$ in Eq. (7) also goes to zero as $\Delta t \to 0^+$. After applying the L'Hopital rule, utilizing again $\Delta t = t$ to simplify the notation, and letting $t = 0$, Eq. (36) becomes

$$\lim_{\Delta t \to 0} \frac{I_{\Delta t}}{\Delta t} = \int dx_t dx_0 \frac{\delta^2 (\partial \ln p(x_t|x_0)/\partial t)|_{t=0}}{\delta \rho_{\text{eq}}(x) \delta \rho_{\text{eq}}(y)} p(x_t, x_0)$$
$$+ \int dx_t dx_0 \frac{\delta^2 \ln p(x_t|x_0)}{\delta \rho_{\text{eq}}(x) \delta \rho_{\text{eq}}(y)} \frac{\partial p(x_t, x_0)}{\partial t}\bigg|_{t=0}. \quad (37)$$

When $p(x_t|x_0)|_{t=0} \to \delta(x_t - x_0)$ in the continuum limit, the lack of $\rho_{\text{eq}}(x)$ dependence makes the second term zero. After applying the functional derivatives to the logarithm and canceling out the factor $\delta(x_t - x_0)$ from $p(x_t, x_0)|_{t=0} \to \delta(x_t - x_0) p_{\text{eq}}(x_0)$, Eq. (37) is simplified to

$$\lim_{\Delta t \to 0} \frac{I_{\Delta t}}{\Delta t} = \int dx_t dx_0 \frac{\delta^2 (\partial p(x_t|x_0)/\partial t)|_{t=0}}{\delta \rho_{\text{eq}}(x) \delta \rho_{\text{eq}}(y)} p_{\text{eq}}(x_0). \quad (38)$$

The functional dependence of the time derivative of the conditional probability is then given by the FPE:

$$\frac{\partial p(x_t|x_0)}{\partial t}\bigg|_{t=0} = \nabla_{x_t} \cdot \left(D(x_t) \rho_{\text{eq}}^2(x_t) \nabla_{x_t} \frac{\delta(x_t - x_0)}{\rho_{\text{eq}}^2(x_t)} \right), \quad (39)$$

where the specific coordinate upon which the gradient operator acts is indicated by the subscript on the gradient operator. Applying the product rule on the inner gradient, the FPE becomes

$$\frac{\partial p(x_t|x_0)}{\partial t}\bigg|_{t=0} = -2\nabla_{x_t} \cdot \left(D(x_t) \delta(x_t - x_0) \frac{\nabla_{x_t} \rho_{\text{eq}}(x_t)}{\rho_{\text{eq}}(x_t)} \right)$$
$$+ \nabla_{x_t} \cdot \left(D(x_t) \nabla_{x_t} \delta(x_t - x_0) \right). \quad (40)$$

The second term of Eq. (40) does not depend on $\rho_{\text{eq}}(x)$. The only contribution to the Fisher information thus comes from the first term in Eq. (40). With this understanding, Eq. (38) becomes

$$\lim_{\Delta t \to 0} \frac{I_{\Delta t}}{\Delta t} = -2 \int dx_t dx_0$$
$$\times \frac{\delta^2 [\nabla_{x_t} \cdot (D(x_t) \delta(x_t - x_0) \nabla_{x_t} \ln \rho_{\text{eq}}(x_t))]}{\delta \rho_{\text{eq}}(x) \delta \rho_{\text{eq}}(y)} \rho_{\text{eq}}^2(x_0). \quad (41)$$

With the exchange of linear operators of space and functional derivatives, the functional derivatives of the logarithm of the equilibrium density described below can be used to evaluate Eq. (41):

$$\frac{\delta^2 \ln \rho_{\text{eq}}(x_t)}{\delta \rho_{\text{eq}}(x) \delta \rho_{\text{eq}}(y)} = -\frac{\delta(x_t - x) \delta(x_t - y)}{\rho_{\text{eq}}^2(x_t)}. \quad (42)$$

After applying this result, Eq. (41) becomes

$$\lim_{\Delta t \to 0} \frac{I_{\Delta t}}{\Delta t} = 2 \int dx_t dx_0 \nabla_{x_t} \cdot \Bigg(D(x_t) \delta(x_t - x_0) \nabla_{x_t}$$
$$\times \frac{\delta(x_t - x) \rho_{\text{eq}}^2(x_0) \delta(x_t - y)}{\rho_{\text{eq}}^2(x_t)} \Bigg). \quad (43)$$

The factor $\rho_{\text{eq}}^2(x_0)$ is now moved inside the spatial derivatives of x_t. Integration of Eq. (43) over x_0 sets $x_0 = x_t$, and this equation can thus be reduced to

$$\lim_{\Delta t \to 0} \frac{I_{\Delta t}}{\Delta t} = 2 \int dx_t \nabla_{x_t} \cdot (D(x_t) \nabla_{x_t} \delta(x_t - x) \delta(x_t - y)). \quad (44)$$

In order to obtain an explicit expression for the Fisher information metric from the integral form of Eq. (44), we evaluate the effect on a twice-differentiable test function $G(x, y)$ that vanishes at domain boundaries. The integral of information evaluation with $G(x, y)$ is

$$\int dx dy G(x, y) \left[\lim_{\Delta t \to 0} \frac{I_{\Delta t}}{\Delta t} \right]$$
$$= 2 \int dx_t dx dy G(x, y) \nabla_{x_t} \cdot (D(x_t) \nabla_{x_t} \delta(x_t - x) \delta(x_t - y)). \quad (45)$$

After performing integrating by parts twice on x_t, the integration over x yields $x = x_t$ from the $\delta(x_t - x)$ term, and a change of variable of x_t to x transforms Eq. (45) into

$$\int dx dy G(x, y) \left[\lim_{\Delta t \to 0} \frac{I_{\Delta t}}{\Delta t} \right]$$
$$= 2 \int dx dy \delta(x - y) \nabla_x \cdot (D(x) \nabla_x G(x, y)). \quad (46)$$

Now reversing operations and performing integration by parts with respect to x twice to bring $G(x, y)$ outside of the gradient operators, we obtain

$$\int dx dy G(x, y) \left[\lim_{\Delta t \to 0} \frac{I_{\Delta t}}{\Delta t} \right]$$
$$= 2 \int dx dy G(x, y) [\nabla_x \cdot (D(x) \nabla_x \delta(x - y))]. \quad (47)$$

For an arbitrary $G(x, y)$ that satisfies the regularity conditions, the equality of Eq. (47) implies the equivalence between the bracketed items inside the integral. Therefore, one of the main results of this work, the analytical expression for the Fisher information metric of ρ_{eq}, is read off from Eq. (48) and applied to the total FIT of Eq. (7) as

$$\mathcal{I}(x, y)[\rho_{\text{eq}}] = 4\delta(x - y) + 2t_{\text{obs}} \nabla_x \cdot (D(x) \nabla_x \delta(x - y)). \quad (48)$$

In the case that the diffusion coefficient is a constant, Eq. (48) indicates that $\lim_{\Delta t \to 0} I_{\Delta t}/\Delta t$ is formally the Laplacian kernel[42] as indeed seen in the numerical example shown in Fig. 3 with the increase of time resolution (decrease of Δt). Thus, our Fisher information metric is truly a Fisher information operator whose net effect can only be realized once the operators are applied in Sec. VI. With the Fisher information metric presented in Eq. (48), the procedure of evaluating FIT is discussed in Sec. VI.

VI. 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[23,43]

$$\mathcal{J} = \int_0^1 d\xi \sum_{i,j} \frac{\partial \theta_i}{\partial \xi} \mathcal{I}_{i,j}(\vec{\theta}) \frac{\partial \theta_j}{\partial \xi}. \quad (49)$$

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

$$\mathcal{S} = \int_0^1 d\xi \sqrt{\sum_{i,j} \frac{\partial \theta_i}{\partial \xi} \mathcal{I}_{i,j}(\vec{\theta}) \frac{\partial \theta_j}{\partial \xi}}. \quad (50)$$

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. (48).

To perform the line integral of FIT, we define a linear path interpolating an essentially flat reference model with $\rho_{\text{eq}}^{\text{ref}}(x) = \lim_{L \to \infty} \sqrt{1/2L}$ and the desired profile that is only denoted as $\rho_{\text{eq}}^*(x)$ for now (with the boundary condition that $\rho_{\text{eq}}(x)|_{-L,L} = 0$). 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,

$$\rho_{\text{eq}}(\xi) = \xi \rho_{\text{eq}}^*(x) + (1 - \xi)\rho_{\text{eq}}^{\text{ref}}(x), \quad (51)$$

$$D(\xi) = \xi D^*(x) + (1 - \xi)D^{\text{ref}}(x), \quad (52)$$

where $\xi \in [0, 1]$. In this work, we concern specifically with the case that the diffusion constant is not varying due to the divergence in the information content in the continuum limit for cases in which $D^*(x) \neq D^{\text{ref}}(x)$.[44] (See Appendix D for a specific example of this divergence between two Gaussian processes.) Thus, we only require the Fisher information as a function of the equilibrium probability density given by Eq. (48).

To calculate the information change, the Fisher information matrix in the continuum limit given by Eq. (48) is applied to Eq. (49) where the sum is replaced by the corresponding integral in the space of equilibrium density functions $\sum_{i,j} \to \int dx dy$. Performing the line integral over $d\xi$ with integration by parts twice on the x coordinate gives the analytical form of

the information dissipation measure

$$\mathcal{J} = \int dx dy \, \rho_{\text{eq}}(y)\delta(x - y)\nabla_x \cdot (D(x)\nabla_x \rho_{\text{eq}}(x)). \quad (53)$$

Taking the integral over y eliminates the delta function to give the 1D integral of

$$\mathcal{J} = \int dx \, \rho_{\text{eq}}(x)\nabla \cdot (D(x)\nabla \rho_{\text{eq}}(x)). \quad (54)$$

Equation (54) is a primary result of this work. Based on the analytical expression we derived for the Fisher information metric in the chosen space for parametrizing the equation of motion, we illustrate how the information content of the dynamic parameters in equilibrium trajectories can be evaluated. This result is analytical and does not require numerical diagonalization of the time propagation operator. Furthermore, expressing $\rho_{\text{eq}}(x)$ in terms of the mean force $F(x)$ in the special case of a constant diffusion coefficient, Eq. (54) becomes

$$\mathcal{J}[F(x)] = \mathcal{S}^2[F(x)] = \frac{D\beta^2}{4}\langle F^2(x)\rangle_{\text{eq}}. \quad (55)$$

The same information expression can also be obtained by using an entropy measure of the probability densities of trajectories and taking the continuum limit.[45] Next, Eq. (54) will serve as the foundation for the analysis of dynamical information in Langevin trajectories.

VII. THE EQUILIBRIUM DISTRIBUTIONS OF LEAST INFORMATIVE DYNAMICS

With the analytical expression of FIT, the LID models under various constraints can be deduced via the constrained optimization approach.[39,46,47] To the best of our knowledge, deriving parameters of the equation of motion using dynamic information as the objective function has not yet been established for continuous stochastic processes. In this section, we use several examples to illustrate how Eq. (54) can be used to achieve this objective.

Finding the least informative distribution of $\rho_{\text{eq}}(x)$ under a set of constraints can be performed via minimizing the information objective function of $\mathcal{J}$ in Eq. (54) under the specified constraints. The resulting Lagrangian for this optimization is

$$\mathcal{L}(\rho_{\text{eq}}(x), \vec{\omega}) = \underbrace{t_{\text{obs}} D\langle \rho_{\text{eq}}|\nabla \cdot \nabla|\rho_{\text{eq}}\rangle}_{\text{FIT}}$$
$$- \sum_i \omega_i \underbrace{\left(\langle \rho_{\text{eq}}|\boldsymbol{C}_i|\rho_{\text{eq}}\rangle - C_i\right)}_{\text{Constraint}}. \quad (56)$$

In Eq. (56), ω_i's are the Lagrange multipliers for the constraints indexed here by i, $\boldsymbol{C}_i$'s are the constraint operators, and C_i's are the desired values for the constraints. For example, in the constraint of $\boldsymbol{C}_1 = \boldsymbol{1}$, 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 of Eq. (56) is accomplished by setting the functional derivative $\delta \mathcal{L}/\delta \rho_{\text{eq}}(x)$ to zero to reach a solution $\rho_{\text{eq}}^*(x)(x; \omega)$ that depends on the Lagrange

multipliers parametrically

$$\mathcal{L}^*(\vec{\omega}) = \inf_{\rho_{eq}(x)} \mathcal{L}(\rho_{eq}(x), \vec{\omega}). \tag{57}$$

The solution of Eq. (57) requires the Lagrange multipliers to satisfy the Karush–Kuhn–Tucker (KKT) conditions of optimality.[48] Alternatively, the Lagrange multipliers can also be determined by optimizing the Lagrangian

$$\omega^* = \arg\min_{\omega} \mathcal{L}^*(\vec{\omega}). \tag{58}$$

This is similar in favor to the derivation of the maximum entropy Boltzmann distribution for the canonical ensemble.[49] 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 (OU) process are presented in Appendixes D and E, respectively.

If all of the constraints are expressed in the form of linear operators as in Eq. (56), numerical techniques for solving partial differential equations can be utilized to determine the optimal profiles of LID. Several examples of using the procedure outlined above to deduce continuous PMF profiles is presented next to highlight the generality of the theoretical framework we developed. 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.

A. The LID model constraining on fixed domain bounds

If the only observation regarding the dynamics of a system is that x is bounded in the fixed domain between $[-L, L]$, the LID model with a constant diffusion coefficient is found by setting the functional derivative of the Lagrangian in Eq. (56) with respect to ρ to zero:

$$\frac{\delta\mathcal{L}}{\delta\rho(x)} = 2t_{obs}D\nabla^2\rho(x) - \omega_1\rho = 0. \tag{59}$$

The normalization constraint is achieved by setting $C_1 = 1$ and $C_1 = 1$. With the boundary conditions of $\rho(x)|_{-L, L} = 0$, the solution of this Strum-Liouville problem is

$$\rho^*(x) = C_1 \cos(\omega' x), \tag{60}$$

$$\omega' = \frac{n\pi}{2L} \quad ; \quad n \in \mathbb{N}. \tag{61}$$

Here, n's are natural numbers and C_1 is determined by the normalization condition.

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

$$\omega^* = \arg\min_{\omega} \mathcal{L}^*(\vec{\omega}) = \arg\min_{\omega} C_1 \left(\frac{n\pi}{2L}\right)^2. \tag{62}$$
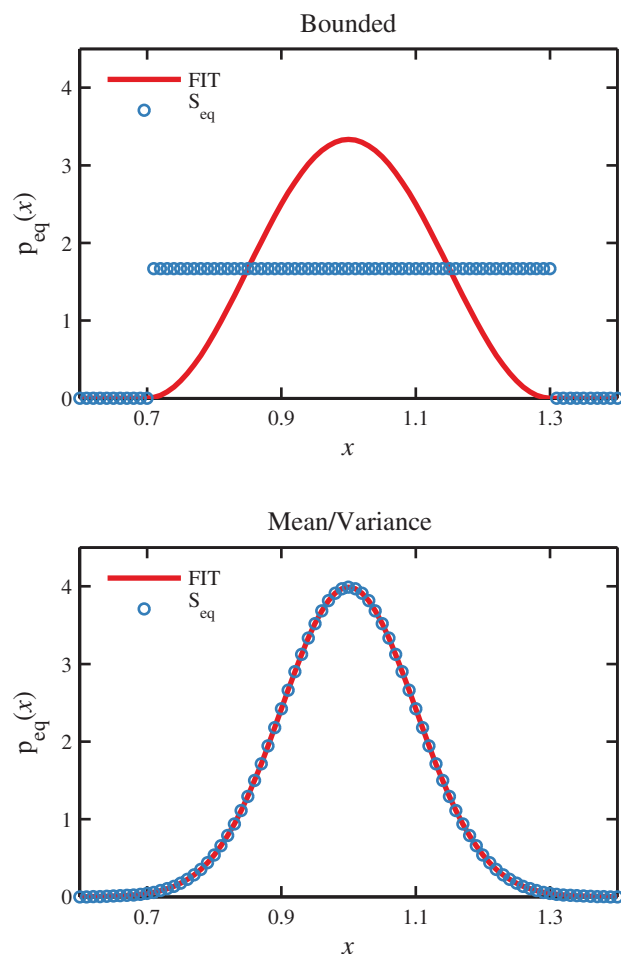

FIG. 4. LID (least informative dynamics) distributions. (Top) For the bounded domain constraint of $x \in [0.7, 1.3]$. The flat profile obtained via maximizing the equilibrium entropy $S_{eq} = \langle \ln p_{eq}(x) \rangle_{eq}$ is shown for comparison. The profile obtained by minimizing FIT (Fisher information of trajectories) has the functional form of $\cos^2((x - \mu)\pi/2L)$ instead. (Bottom) For the constraints of fixed average $\langle x \rangle = \mu$ and variance $\langle (x - \mu)^2 \rangle = \sigma^2$. In this case, maximizing S_{eq} and minimizing FIT give the same functional form of the profile as $\sim \exp(-(x - \mu)^2/2\sigma^2)$.

This optimization simply selects the smallest possible n. As a result,

$$p_{eq}^*(x) = \frac{1}{L}\cos^2\left(\frac{x\pi}{2L}\right), \tag{63}$$

$$F^*(x) = \frac{\pi}{L}\tan\left(\frac{x\pi}{2L}\right). \tag{64}$$

Fig. 4 plots the distribution of x based on the principle of LID in comparison with the maximum entropy profile of maximum entropy $S_{eq} = \int p_{eq}(x)\ln p_{eq}(x)$, which is a flat distribution in the domain with diverging forces at the boundaries. Because the Fisher information of trajectories includes a measure of the deterministic force in dynamics as indicated in Eq. (55), diverging values of $F(x)$ at the boundaries are not desirable. Alternatively, a distribution $\cos^2(x)$ that smoothly decays to a zero gradient at the domain edges is selected instead according to the criterion of LID.

B. The LID model constraining on fixed values of the mean and variance of p_{eq}

If the knowledge of mean μ and variance σ^2 of x are available, the LID distribution of x can be found by setting the two constraints $C_1 = x\mathbf{1}$ and $C_2 = x^2\mathbf{1}$. This optimization can be performed in the Fourier space and this approach is adopted here to illustrate the generality offered by having the analytical expression of FIT. Because the Lagrangian is essentially a set of inner products, the Parseval's theorem[50] says that the Fourier transforms (FT) of functions $f(x)$ and $g(x)$, $\tilde{f}(k)$ and $\tilde{g}(k)$, respectively, satisfy

$$\int dx \ f(x)g^{\dagger}(x) = \int dk \ \tilde{f}(k)\tilde{g}^{\dagger}(k). \quad (65)$$

The $\dagger$ sign indicates complex conjugation. Because the FT of $\nabla\rho(x)$ is $ik\tilde{\rho}(k)$, where i is the imaginary number, the FT of the information measure in Eq. (54) reads

$$\mathcal{J} = -D \int dk \ k^2 \tilde{\rho}^2(k). \quad (66)$$

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 $\tilde{\rho}(k) \propto 1/(1 + k^2)$.[51]

For the constraint on mean, $g(x) = x\rho(x)$, $g(k) = id\tilde{\rho}(k)/dk$, 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

$$\mathcal{L}(\tilde{\rho}(k)) = \int dk \ \tilde{\rho}(k)\Big(-t_{\text{obs}}DDk^2\tilde{\rho}(k) \\ -\omega_1 i\frac{d\tilde{\rho}}{dk}(k) - \omega_2\frac{d^2\tilde{\rho}}{dk^2}(k) \Big). \quad (67)$$

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

$$\omega_2\frac{d^2\tilde{\rho}}{dk^2}(k) + \omega_1 i\frac{d\tilde{\rho}}{dk}(k) + t_{\text{obs}}Dk^2\tilde{\rho}(k) = 0. \quad (68)$$

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

$$\frac{d^2\hat{\rho}}{dk^2}(k) + \Big(\frac{t_{\text{obs}}D}{\omega_2}k^2 - \frac{\omega_1^2}{4\omega_2^2} \Big)\hat{\rho}(k) = 0. \quad (69)$$

The solutions of this equation are the Hermite functions, the same as those for the wave function of a quantum harmonic oscillator. The ground state mode with the least information thus has the form of a Gaussian

$$\tilde{p}(k) \propto e^{-\omega_1' ik}e^{-\omega_2' k^2}, \quad (70)$$

where the requirement of constraints is incorporated into ω' for simplicity. Inverse Fourier transform then takes Eq. (70) to the real-space solution:

$$\rho_{\text{eq}}^*(x) = \mathcal{F}^{-1}(\tilde{p}(k)) \propto e^{-\omega_2''(x-\omega_1')^2}. \quad (71)$$

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:

$$p_{\text{eq}}^*(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\mu)^2/2\sigma^2}. \quad (72)$$

This result of LID model is the same as that obtained by maximizing the static entropy S_{eq} despite the use of FIT. However, this coincidence can be rationalized by inspecting the two different information measures and their corresponding inequality bounds to establish that the Gaussian equilibrium probability density is indeed optimal in both cases:[52]

$$\text{FIT}: \Big(\int dx \ x^2\rho^2(x) \Big)\Big(\int dk \ k^2\tilde{\rho}^2(k) \Big) \geq \frac{1}{16\pi^2}, \quad (73)$$

$$S_{\text{eq}}: \int dx \ \rho^2(x)\ln\rho^2(x) + \int dk \ \tilde{\rho}^2(k)\ln\tilde{\rho}^2(k) \leq \ln 2 - 1. \quad (74)$$

The inequality of Eq. (73) can be viewed as an uncertainty principle of continuous stochastic dynamics.[52] It contains a static measure of the variance in position multiplied by the dynamic information of $\mathcal{J}$. Here, $\mathcal{J}$ plays the analogous role of the variance of velocity in the Heisenberg uncertainty of quantum mechanics.[53] On the other hand, the inequality of Eq. (74) is the Hirschman uncertainty principle[54] which adds a static measure of equilibrium entropy S_{eq} with the Shannon information measure of the velocity distribution. In both cases, a balance can be reached between a static and dynamic measure of information and the Gaussian distribution is optimal and achieves equality in both scenarios among all of the normalized profiles of $p_{\text{eq}}(x) \in L^2$.

C. The LID model constraining on fixed values of the mean first passage times (MFPTs)

If the transition between two metastable states locating at x_A and x_B with the reaction rate of $k_{A\to B}$ is known, the mean first passage time between the two states is $\tau_{\text{rxn}}(x_A \to x_B) = k_{A\to B}^{-1}$. This knowledge can be used as the constraint for finding the corresponding LID model. According to Gardiner,[55] the mean first passage time from x_A to x_B can be expressed as

$$\tau_{x_A \to x_B} = k_{A\to B}^{-1} = \frac{1}{D}\int_{x_A}^{x_B} dx \ \rho^{-2}(x)\int_{x_L}^{x} dx' \ \rho^2(x'),$$

$$\tau_{x_B \to x_A} = k_{B\to A}^{-1} = \frac{1}{D}\int_{x_B}^{x_A} dx \ \rho^{-2}(x)\int_{x_R}^{x} dx' \ \rho^2(x'). \quad (75)$$

Here, we use $\rho = \sqrt{p}$ as the variable within the finite domain $x \in [x_L, x_R]$.

The constraints on MFPTs defined in Eq. (75) are nonlinear and do not follow the advantageous linear mode of Eq. (56). Using $\tau_{x_B \to x_A}$ and $\tau_{x_A \to x_B}$ to represent the operation on ρ in Eq. (75), the Lagrangian of LID becomes

$$\mathcal{L} = t_{\text{obs}}D\int dx(\nabla\rho(x))^2 + \omega_1\Big(\int dx\rho^2(x) - 1 \Big) \\ + \omega_2\big(\tau_{x_B \to x_A}[\rho] - k_{B\to A}^{-1} \big) + \omega_3\big(\tau_{x_A \to x_B}[\rho] - k_{A\to B}^{-1} \big). \quad (76)$$

The optimization of this Lagrangian is achieved by setting its functional derivative with respect to ρ to zero, and requires

the solution of the resulting equation

$$2t_{\text{obs}}D\nabla^2\rho(x) + \omega_1 2\rho(x) + \omega_2\frac{\delta\boldsymbol{\tau}_{x_B\to x_A}}{\delta\rho(x)} + \omega_3\frac{\delta\boldsymbol{\tau}_{x_A\to x_B}}{\delta\rho(x)} = 0.$$ (77)

However, there is no simple and closed form solution for this equation because the functional derivatives of the MF-PTs with respect to ρ depend on the specific locations within the domain in a nonlinear manner:

$$\frac{\delta\boldsymbol{\tau}_{x_A\to x_B}}{\delta\rho(x)} = \begin{cases} \frac{2}{D}\rho\int_{x_A}^{x_B}\rho^{-2}(x')dx' & x \le x_A \\ \frac{2}{D}\rho(x)\int_x^{x_B}dx'\rho^{-2}(x') - \frac{2}{D}\rho^{-3}(x)\int_{x_L}^x dx'\rho^2(x') & x_A < x \le x_B \\ 0 & x > x_B. \end{cases}$$ (78)

The derivation of this equation is shown in Appendix F. The nonlinear optimization problem of Eq. (77) can be solved numerically by using a quasi-Newton method iteratively.[48] The resulting PMFs and equilibrium distributions for three sets of reaction rates are shown in Fig. 5.
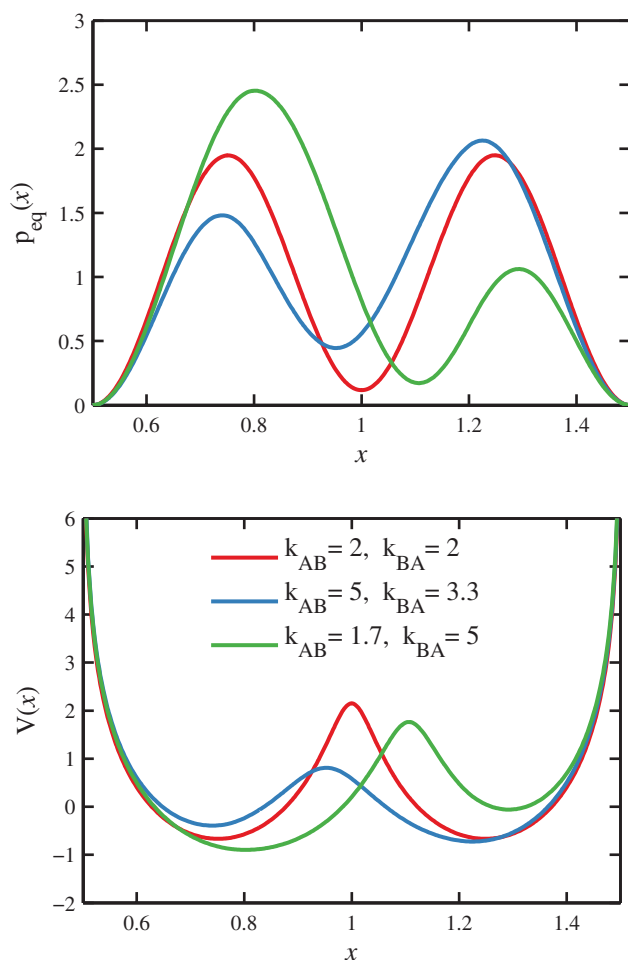

FIG. 5. Least informative dynamic models for using the mean first passage times as constrains. (Top) The optimal equilibrium probability density $p_{\text{eq}}^*(x)$ for the constraint of rate constant/mean first passage time between two states located at $x_A = 0.8$ and $x_B = 1.2$ with diffusion constant $D = 1$. (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.

It can be seen from Fig. 5 that the LID profiles of $p_{\text{eq}}(x)$ constrained on MFPTs generally follow those of Kramer's type of reactions[56] that have quadratic potentials for stable states connected by an inverted quadratic barrier. This result thus provides a framework for justifying Kramer's approach of modeling rare event processes. Observation of the LID models with MFPT constraints indicates that the results follow the Hammond-Leffler postulate based on a heuristic arguments, which states that the transition state $x^{\ddagger}$ of an "endothermic" reaction is closer to the product.[57] Therefore, the principle of LID is shown here to be consistent with the outstanding theories of chemical kinetics.

In this example of MFPT constraints, we illustrate that the analytical expression of FIT can be applied with the knowledge of kinetic observables to stitch together a least informative profile of the equilibrium probability density of state distribution. The theoretical framework developed in this work thus provides a systematic way of modeling dynamic systems that evolve with stochasticity. The diverse examples presented in this section establish that the principle of LID can also be used to estimate the underdetermined continuous parameters given the limited number of observables in the measured data.

VIII. 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 metric 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 co-efficient 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 derived from scalar constraints, generalization to using an arbitrary functional in describing the observable of interest, such as the likelihood of single-molecule FRET measurements,[58–60] 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 on the principle of the least informative description.

ACKNOWLEDGMENTS

APPENDIX A: DERIVATION OF EQ. (18) OF FIT FOR SYMMETRIC $\rho(x_t, x_0)$

The starting point is Eq. (12) and is restated here:

$$I_{\Delta t}(x, y) = \int dx_t dx_0 \frac{\delta \ln p(x_t|x_0)}{\delta \rho_{eq}(x)} \frac{\delta \ln p(x_t|x_0)}{\delta \rho_{eq}(y)} p(x_t, x_0). \tag{A1}$$

Inputting $p(x_t|x_0) = \rho_{eq}(x_t)\rho(x_t, x_0)/\rho_{eq}(x_0)$ and $\ln p(x_t|x_0) = \ln \rho_{eq}(x_t) + \ln \rho(x_t, x_0) - \ln \rho_{eq}(x_0)$ leads to 9 terms:

$$I_{\Delta t}(x, y) = \int dx_t dx_0 \frac{\delta \ln \rho(x_t, x_0)}{\delta \rho_{eq}(x)} \frac{\delta \ln \rho(x_t, x_0)}{\delta \rho_{eq}(y)} p(x_t, x_0) \tag{A2}$$

$$+ \int dx_t dx_0 \frac{\delta \ln \rho_{eq}(x_t)}{\delta \rho_{eq}(x)} \frac{\delta \ln \rho_{eq}(x_t)}{\delta \rho_{eq}(y)} p(x_t, x_0) \tag{A3}$$

$$+ \int dx_t dx_0 \frac{\delta \ln \rho_{eq}(x_0)}{\delta \rho_{eq}(x)} \frac{\delta \ln \rho_{eq}(x_0)}{\delta \rho_{eq}(y)} p(x_t, x_0) \tag{A4}$$

$$- \int dx_t dx_0 \frac{\delta \ln \rho_{eq}(x_0)}{\delta \rho_{eq}(x)} \frac{\delta \ln \rho_{eq}(x_t)}{\delta \rho_{eq}(y)} p(x_t, x_0) \tag{A5}$$

$$- \int dx_t dx_0 \frac{\delta \ln \rho_{eq}(x_t)}{\delta \rho_{eq}(x)} \frac{\delta \ln \rho_{eq}(x_0)}{\delta \rho_{eq}(y)} p(x_t, x_0) \tag{A6}$$

$$+ \int dx_t dx_0 \frac{\delta \ln \rho_{eq}(x_t)}{\delta \rho_{eq}(x)} \frac{\delta \ln \rho(x_t, x_0)}{\delta \rho_{eq}(y)} p(x_t, x_0) \tag{A7}$$

$$- \int dx_t dx_0 \frac{\delta \ln \rho_{eq}(x_0)}{\delta \rho_{eq}(x)} \frac{\delta \ln \rho(x_t, x_0)}{\delta \rho_{eq}(y)} p(x_t, x_0) \tag{A8}$$

$$+ \int dx_t dx_0 \frac{\delta \ln \rho(x_t, x_0)}{\delta \rho_{eq}(x)} \frac{\delta \ln \rho_{eq}(x_t)}{\delta \rho_{eq}(y)} p(x_t, x_0) \tag{A9}$$

$$- \int dx_t dx_0 \frac{\delta \ln \rho(x_t, x_0)}{\delta \rho_{eq}(x)} \frac{\delta \ln \rho_{eq}(x_0)}{\delta \rho_{eq}(y)} p(x_t, x_0). \tag{A10}$$

The collection of integrals of that makeup $I_{\Delta t}(x, y)$ 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. (A10) gives an exact copy and cancels with Eq. (A9). Likewise Eq. (A8) cancels with Eq. (A7). The remaining 4 integrals containing the ρ_{eq} term require the functional derivative

$$\frac{\delta \ln \rho_{eq}(x_t)}{\delta \rho_{eq}(x)} = \frac{\delta(x_t - x)}{\rho_{eq}(x)}. \tag{A11}$$

Therefore, Eqs. (A5) and (A6) become the integral of two delta functions

$$\int dx_t dx_0 \frac{\delta(x_t - x)}{\rho_{eq}(x)} \frac{\delta(x_0 - y)}{\rho_{eq}(y)} p(x_t, x_0) = \rho(x, y), \tag{A12}$$

for which the property of $\rho(x, y) = p(x, y)/\rho_{eq}(x)\rho_{eq}(y)$ is utilized as well. Applying the same functional derivative to Eqs. (A3) and (A4) gives the alternative result:

$$\int dx_0 \frac{\delta(x_0 - x)}{\rho_{eq}(x)} \frac{\delta(x_0 - y)}{\rho_{eq}(y)} \int dx_t p(x_t, x_0) = \delta(x - y). \tag{A13}$$

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 $p(x_0)/\rho_{eq}(x_0)\rho_{eq}(x_0) = 1$. Combining these results gives Eq. (18):

$$I_{\Delta t}(x, y) = \int dx_t dx_0 \frac{\delta \ln \rho(x_t, x_0)}{\delta \rho_{eq}(x)} \frac{\delta \ln \rho(x_t, x_0)}{\delta \rho_{eq}(y)}$$
$$\times p(x_t, x_0) + 2\delta(x - y) - 2\rho(x_t, x_0). \tag{A14}$$

APPENDIX B: PROOF OF THE HERMITIAN PROPERTY OF THE SYMMETRIZED 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. (20), the left-hand side is

$$\int dx \frac{\psi_i(x)}{\rho_{eq}(x)} \nabla \cdot (D(x)\rho_{eq}^2(x)\nabla \frac{\psi_j(x)}{\rho_{eq}(x)}). \tag{B1}$$

This equation can be transformed by an integration by parts to

$$- \int dx (D(x)\rho_{eq}^2(x)\nabla \frac{\psi_j(x)}{\rho_{eq}(x)}) \cdot \nabla \frac{\psi_i(x)}{\rho_{eq}(x)}. \tag{B2}$$

The zero boundary terms in the space of square integrable functions $\psi \subset L^2$ were applied. Another integration by parts on the $\nabla(\psi_j(x)/\rho_{eq}(x))$ term gives

$$\int dx \frac{\psi_j(x)}{\rho_{eq}(x)} \nabla \cdot (D(x)\rho_{eq}^2(x)\nabla \frac{\psi_i(x)}{\rho_{eq}(x)}), \tag{B3}$$

TABLE I. 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

which is recognized as $\langle \psi_i | \boldsymbol{H} | \psi_j \rangle$. The Hermitian property of the operator defined in Eq. (20) is thus verified.

APPENDIX C: POTENTIAL OF MEAN FORCE OF THE REFERENCE MODEL

The potential of mean force $V(x)/k_B T$ is constructed using the piecewise cubic Hermite interpolation polynomial `pchip()` in MATLAB with the following points (Table I):

APPENDIX D: FIT FOR THE GAUSSIAN PROCESS OF CONSTANT FORCE

The time propagator for this process is

$$p(x_{\Delta t}|x_0) = \frac{1}{\sqrt{4\pi D \Delta t}} \exp\left(-\frac{(x_{\Delta t} - x_0 - FD\Delta t)^2}{4D\Delta t}\right). \tag{D1}$$

The calculation of FIT via Eq. (12) starts by taking derivatives of $p(x_{\Delta t}|x_0)$ with respect to the x-independent force F and diffusion coefficient D

$$\frac{\partial \ln p(x_{\Delta t}|x_0)}{\partial F} = \frac{(\Delta x - FD\Delta t)}{2}, \tag{D2}$$

$$\frac{\partial \ln p(x_{\Delta t}|x_0)}{\partial D} = \frac{F(\Delta x - FD\Delta t)}{2D} + \frac{(\Delta x - FD\Delta t)^2}{4D^2\Delta t} - \frac{1}{2D}. \tag{D3}$$

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 another 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 an integration over Δx and $\Delta x' = x_{\Delta t} + x_0$:

$$I_{F,F} = \mathbb{E}\left[\frac{(\Delta x - FD\Delta t)^2}{4}\right] = \frac{D\Delta t}{2}, \tag{D4}$$

$$I_{F,D} = I_{D,F} = \frac{F\Delta t}{2}, \tag{D5}$$

$$I_{D,D} = \frac{1}{2D^2} + \frac{F^2\Delta t}{2D}. \tag{D6}$$

The following properties for the moments of the Gaussian distribution have also been used to calculate the matrix elements:

$$\mathbb{E}\left[(X - \mu)^N\right] = \begin{cases} 0 & N = 1 \\ \sigma^2 & N = 2 \\ 0 & N = 3 \\ 3\sigma^3 & N = 4 \end{cases}. \tag{D7}$$

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. (49) for the path

$$F(\xi) = F_1 + (F_2 - F_1)\xi, \tag{D8}$$

$$D(\xi) = D_1 + (D_2 - D_1)\xi, \tag{D9}$$

where $\xi \in [0, 1]$. After applying 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:

$$\mathcal{J}_{1\to 2} = \frac{t_{\text{obs}}}{2} \int_0^1 d\xi \, (\Delta F)^2 \frac{D(\xi)}{2} \tag{D10}$$

$$+ (\Delta F \Delta D) F(\xi) + (\Delta D)^2 \left(\frac{F^2(\xi)}{2D(\xi)} + \frac{\Delta t^{-1}}{2D^2(\xi)}\right). \tag{D11}$$

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

$$\mathcal{J}_{1\to 2} = t_{\text{obs}} \frac{(F_2 D_2 - F_1 D_1)^2}{2D_2} + \lim_{\Delta t \to 0^+} \frac{t_{\text{obs}}}{\Delta t} \left[2 - \frac{D_2}{D_1} - \frac{D_1}{D_2}\right]. \tag{D12}$$

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.[61] 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}.

APPENDIX E: FIT FOR THE ORNSTEIN-UHLENBECK PROCESS

The 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:

$$p_{\text{eq}}(x) = \sqrt{\frac{k}{2\pi}} e^{-kx^2/2}. \tag{E1}$$

The time propagation probability density for the OU process is

$$p(x_{\Delta t}|x_0) = \frac{\sqrt{k}}{\sqrt{2\pi(1 - e^{-2Dk\Delta t})}}$$

$$\times \exp\left(\frac{-k\left(x_{\Delta t} - x_0 e^{-kD\Delta t}\right)^2}{2 - 2e^{-2Dk\Delta t}}\right). \tag{E2}$$

In the continuum limit, we take the Taylor expansion of the exponential $\exp(-2Dk\Delta t) \simeq 1 - 2Dk\Delta t$ to give the

approximate propagator

$$p(x_{\Delta t}|x_0) = \frac{1}{\sqrt{4\pi D \Delta t}} \exp\left(\frac{-(x_{\Delta t} - x_0 + x_0 k D \Delta t)^2}{4D\Delta t}\right).$$
(E3)

Taking the derivative of the log propagator with respect to k results in

$$\frac{\partial \ln p(x_{\Delta t}|x_0)}{\partial k} = \frac{-x_0(x_{\Delta t} - x_0 + x_0 k D \Delta t)}{2}.$$
(E4)

Then, the Fisher information is the expectation of this derivative squared:

$$\mathcal{I}_k = \frac{t_{\text{obs}}}{\Delta t} \mathbb{E}_{x_{\Delta t},x_0}\left[\frac{x_0^2(x_{\Delta t} - x_0 + x_0 k D \Delta t)^2}{4}\right].$$
(E5)

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:

$$\mathcal{I}_k = \frac{t_{\text{obs}} D}{2k}.$$
(E6)

To relate this result to the overall change in information from an initial state of $k = k_1$ to a final state of $k = k_2$, we perform the line integral

$$\mathcal{J} = \frac{-1}{2}\int_0^1 d\xi \, \Delta k \frac{t_{\text{obs}} D}{2k(\xi)} \Delta k = -\frac{t_{\text{obs}} D}{4}\Delta k \ln\frac{k_2}{k_1}.$$
(E7)

Here, $\Delta k = k_2 - k_1$ and both k_1 and k_2 are positive and cannot be zero. Comparing to the result via the general functional of Eq. (55), $\mathcal{J} = -D\int dx \, (\nabla\rho_{\text{eq}}(x))^2 = -(Dk^2/4)\int dx \, x^2 p_{\text{eq}}(x) = -Dk/4$, the result is different in the limit that $k_1 \ll 1$ with pure free diffusion due to the nonergodic nature of this brownian diffusion which does not have a proper equilibrium distribution of states. This is also apparent because the Fisher information is not constant with the parameter k and thus the choice of integration path now alters the final form of the integral $\mathcal{J}$. Naturally, systems with a larger spring constant 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.

APPENDIX F: FUNCTIONAL DERIVATIVE OF MEAN FIRST PASSAGE TIMES

The functional derivatives for the mean first passage time are given by

$$\tau_{x_A \to x_B} = k_{A\to B}^{-1} = \frac{1}{D}\int_{x_A}^{x_B} dx \, \rho^{-2}(x)\int_{x_L}^{x} dx' \, \rho^2(x'),$$

$$\tau_{x_B \to x_A} = k_{B\to A}^{-1} = \frac{1}{D}\int_{x_B}^{x_A} dx \, \rho^{-2}(x)\int_{x_R}^{x} dx' \, \rho^2(x'),$$
(F1)

is derived from finding the first order perturbation from a test function $f(x)$ that vanishes at the boundaries of the domain and has bounded derivatives:

$$\int dx \frac{\delta\tau_{x_A \to x_B}}{\delta\rho(x)}f(x) = \frac{d}{d\varepsilon}\tau_{x_A \to x_B}[\rho(x) + \varepsilon f(x)]\bigg|_{\varepsilon=0}.$$
(F2)

Performing the differentiation of the above equation with respect to ε gives

$$\int dx \frac{\delta\tau_{x_A \to x_B}}{\delta\rho(x)}f(x) = \overbrace{\frac{2}{D}\int_{x_A}^{x_B} dx \, \rho^{-2}(x)\int_{x_L}^{x} dx' \, f(x')\rho(x')}^{\text{Parts}}$$
(F3)

$$- \frac{2}{D}\int_{x_A}^{x_B} dx \, f(x)\rho^{-3}(x)\int_{x_L}^{x} dx' \, \rho^2(x').$$
(F4)

The perturbation $f(x)$ is isolated from the nested integral in Eq. (F3) by performing an integration by parts with $u = \int_{x_L}^{x} dx' f(x')\rho(x')$ and $dv = \rho^{-2}(x)$ to give

$$\text{Parts} = \frac{2}{D}\left(\int_{x_L}^{x} f(x)\rho(x)\right)\left(\int_{x_A}^{x} dx' \rho^{-2}(x')\right)\bigg|_{x_A}^{x_B}$$

$$- \frac{2}{D}\int_{x_A}^{x_B} dx f(x)\rho(x)\int_{x_A}^{x} dx' \rho^{-2}(x').$$
(F5)

[1] G. Hummer and A. Szabo, Proc. Natl. Acad. Sci. U.S.A. **107**, 21441 (2010).
[2] A. Mittermaier and L. E. Kay, Science **312**, 224 (2006).
[3] J. A. Hanson, and H. Yang, in *Quantitative Biology: From Molecular to Cellular Systems*, edited by M. E. Wall (Taylor and Francis Group, New York, 2012).
[4] J. D. Bryngelson and P. G. Wolynes, J. Phys. Chem. B **93**, 6902 (1989).
[5] R. B. Best and G. Hummer, Proc. Natl. Acad. Sci. U.S.A. **107**, 1088 (2010).
[6] S. Krivov and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **105**, 13841 (2008).
[7] P. Langevin, C. R. Acad. Sci. **146**, 530 (1908).
[8] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, J. Chem. Phys. **108**, 1964 (1998).
[9] D. D. Boehr, R. Nussinov, and P. E. Wright, Nat. Chem. Biol. **5**, 789 (2009).
[10] M. E. Johnson and G. Hummer, J. Phys. Chem. B **116**, 8573 (2012).
[11] B. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus, J. Comput. Chem. **30**, 1545 (2009).
[12] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, J. Comput. Chem. **26**, 1781 (2005).
[13] T. Schlick, R. Collepardo-Guevara, L. A. Halvorsen, S. Jung, and X. Xiao, Q. Rev. Biophys. **44**, 191 (2011).
[14] D. D. Boehr, H. J. Dyson, and P. E. Wright, Chem. Rev. **106**, 3055 (2006).
[15] A. G. Palmer, Chem. Rev. **104**, 3623 (2004).
[16] S.-R. Tzeng and C. G. Kalodimos, Curr. Opin. Struct. Biol. **21**, 62 (2011).
[17] Th. Förster, Ann. Phys. **437**, 55 (1948).
[18] T. Ha, A. Y. Ting, J. Liang, W. B. Caldwell, A. A. Deniz, D. S. Chemla, P. G. Schultz, and S. Weiss, Proc. Natl. Acad. Sci. U.S.A. **96**, 893 (1999).
[19] J. A. Hanson, K. Duderstadt, L. P. Watkins, S. Bhattacharyya, J. Brokaw, J. W. Chu, and H. Yang, Proc. Natl. Acad. Sci. U.S.A. **104**, 18055 (2007).
[20] H. Yang, in *Theory and Evaluation of Single-Molecule Signals*, edited by E. Barkai, F. L. Brown, M. Orrit, and H. Yang (Scientific Publishing, 2008).
[21] H. Yang, Adv. Chem. Phys. **146**, 129 (2011).
[22] R. A. Fisher, Math. Proc. Cambridge Philos. Soc. **22**, 700 (1925).
[23] G. E. Crooks, Phys. Rev. Lett. **99**, 100602 (2007).
[24] B. Frieden, A. Plastino, A. Plastino, and B. Soffer, Phys. Rev. E **60**, 48 (1999).
[25] A. Plastino, in *Proceedings of the 20th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering* (AIP, 2000), pp. 38–48.
[26] S. Ganguli, D. Huh, and H. Sompolinsky, Proc. Natl. Acad. Sci. U.S.A. **105**, 18970 (2008).
[27] Y. Pantazis and M. A. Katsoulakis, J. Chem. Phys. **138**, 054115 (2013).

28D. Dacunha-Castelle and D. Florens-Zmirou, Stochastics **19**, 263 (1986).

29P. Biane and R. Speicher, Ann. Inst. Henri Poincare, Sect. B **37**, 581 (2001).

30A. Majda and X. Wang, Commun. Math. Sci. **8**, 145 (2010).

31M. Komorowski, M. J. Costa, D. A. Rand, and M. P. H. Stumpf, Proc. Natl. Acad. Sci. U.S.A. **108**, 8645 (2011).

32H. Haken, Z. Phys. B **63**, 505 (1986).

33A. R. Plastino and A. Plastino, Phys. A **222**, 347 (1995).

34C. Tsallis and D. J. Bukman, Phys. Rev. E **54**, R2197 (1996).

35G. Stock, K. Ghosh, and K. A. Dill, J. Chem. Phys. **128**, 194102 (2008).

36J. Xing, H. Wang, and G. Oster, Biophys. J. **89**, 1551 (2005).

37A. R. Plastino and A. Plastino, Phys. Rev. E **54**, 4423 (1996).

38L. Borland, F. Pennini, A. R. Plastino, and A. Plastino, Eur. Phys. J. B **12**, 285 (1999).

39S. Luo, Found. Phys. **32**, 1757 (2002).

40C. Pozrikidis, *Introduction to Finite and Spectral Element Methods Using Matlab* (CRC Press, 2005).

41A. Drozdov and J. Brey, Phys. Rev. E **57**, 1284 (1998).

42A. J. Smola and R. Kondor, in *Learning Theory and Kernel Machines* (Springer, Berlin, 2003), pp. 144–158.

43X. Calmet and J. Calmet, Phys. Rev. E **71**, 056109 (2005).

44R. Dahlhaus, Stochastic Proc. Appl. **62**, 139 (1996).

45K. R. Haas, J.-W. Chu, and H. Yang, "The Trajectory Entropy of Continuous Stochastic Processes at Equilibrium," Phys. Rev. Lett. (unpublished).

46M. R. Bastian, J. H. Gunther, and T. K. Moon, Adv. Artif. Neural Syst. **2011**, 1 (2011).

47J. Myung, M. Pitt, and D. Navarro, Proc. SPIE **5674**, 219 (2005).

48J. Nocedal and S. J. Wright, *Numerical Optimization* (Springer, 2006).

49D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, USA, 1987).

50E. Thiébaut and J. M. Conan, J. Opt. Soc. Am. A **12**, 485 (1995).

51D. E. Reynolds, Phys. Rev. E **79**, 061107 (2009).

52M. Pinsky, *Introduction to Fourier Analysis and Wavelets* (American Mathematical Society, 2002).

53J. J. Sakurai, *Modern Quantum Mechanics* (Addison Wesley Publishing Company, 1985).

54E. Romera, P. Sánchez-Moreno, and J. S. Dehesa, J. Math. Phys. **47**, 103504 (2006).

55C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences* (Springer Verlag, 2004).

56H. Dekker, Phys. Lett. A **112**, 197 (1985).

57I. E. Sánchez and T. Kiefhaber, Biophys. Chem. **100**, 397 (2003).

58L. P. Watkins and H. Yang, Biophys. J. **86**, 4015 (2004).

59L. P. Watkins and H. Yang, J. Phys. Chem. B **109**, 617 (2005).

60L. P. Watkins, H. Chang, and H. Yang, J. Phys. Chem. A **110**, 5191 (2006).

61H. Luschgy and G. Pagès, J. Funct. Anal. **212**, 89 (2004).