THE JOURNAL OF PHYSICAL CHEMISTRY A
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCA

Article

String Method with Swarms-of-Trajectories, Mean Drifts, Lag Time, and Committor

Published as part of The Journal of Physical Chemistry virtual special issue "125 Years of The Journal of Physical Chemistry".

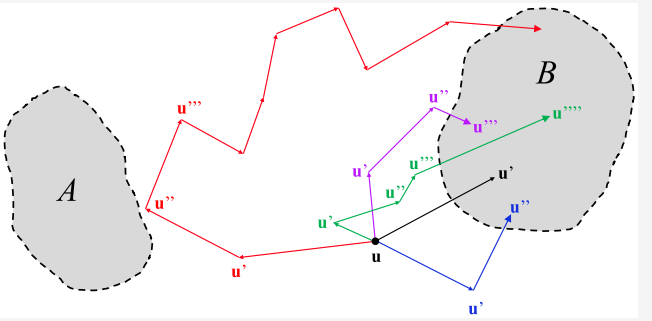
Benoît Roux*

Cite This: *J. Phys. Chem. A* 2021, 125, 7558–7571

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: The kinetics of a dynamical system comprising two metastable states is formulated in terms of a finite-time propagator in phase space (position and velocity) adapted to the underdamped Langevin equation. Dimensionality reduction to a subspace of collective variables yields familiar expressions for the propagator, committor, and steady-state flux. A quadratic expression for the steady-state flux between the two metastable states can serve as a robust variational principle to determine an optimal approximate committor expressed in terms of a set of collective variables. The theoretical formulation is exploited to clarify the foundation of the string method with swarms-of-trajectories, which relies on the mean drift of short trajectories to determine the optimal transition pathway. It is argued that the conditions for Markovity within a subspace of collective variables may not be satisfied with an arbitrary short time-step and that proper kinetic behaviors appear only when considering the effective propagator for longer lag times. The effective propagator with finite lag time is amenable to an eigenvalue-eigenvector spectral analysis, as elaborated previously in the context of position-based Markov models. The time-correlation functions calculated by swarms-of-trajectories along the string pathway constitutes a natural extension of these developments. The present formulation provides a powerful theoretical framework to characterize the optimal pathway between two metastable states of a system.

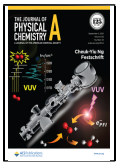


INTRODUCTION

A central problem in computational biophysics is the characterization of the long-time kinetic behavior of molecular systems. Many of the key concepts can be formulated by considering a prototypical system comprising two dominant metastable states A and B. In the context of a multistate Markov model, the steady-state flux from A to B can be expressed as the net sum of productive transitions across a dividing surface between the two end states. Further analysis shows that the steady-state probability of the states under such nonequilibrium conditions can be expressed as the product of the equilibrium probability of the states times the probability that a trajectory initiated at the same position will be reactive and first reach the state B before ever reaching the state A.[1] This "committor" probability, which can be determined on the basis of the backward dynamical propagation, then becomes a critical ingredient in efforts to formulate a theoretical framework seeking to treat such problems.[1,2]

This analysis leads to the observation that the principal lines of a reactive probability current between the states A and B are largely determined by the equilibrium probability times the local gradiant of the committor.[2] This observation provides a critical insight in the formulation of the string method,[3−5] which seeks to determine the dominant "reaction tube" that contains most of the probability current between A and B. Mathematically, the line of maximum probability flux is a curve in the phase space, and a calculation of the full committor is practically infeasible. To reduce the complexity of the problem, Maragliano et al.[6] assumed that the committor probability depends predominantly on a subset of collective variables (CVs), $\mathbf{z} \equiv \{z_1, z_2, ..., z_N\}$. These considerations provide the background that led to the development of the string method with CVs on the potential of mean force (PMF) surface $W(\mathbf{z})$.[6] The string method represents the curvilinear minimum free energy pathway (MFEP) linking the states A and B as a curve in the space of the collective variables. Inspired by pioneering work from Pratt[7] and Elber and

Received: May 8, 2021
Revised: July 26, 2021
Published: August 18, 2021



Karplus,[8] this curve (the string) is constructed iteratively as a chain of M copies of the system ("images"). In its original formulation, the information needed for each image is the mean force and the value of a metric tensor, both of which can be expressed in terms of conditional averages. This algorithm has been employed in a variety of applications.[9−17]

The problem of optimizing the path in a multidimensional space, nonetheless, remains arduous and computationally expensive. To circumvent these difficulties, a number of variations of the original mean force algorithm of Maragliano et al.[6] have been proposed, including the finite temperature string method,[18] the string method with swarms-of-trajectories,[19] and multiscale preconditioning.[20] The idea of the string method with swarms-of-trajectories[19] (also called the "drift" method[21]) is to rely on the outcome from a large number of short unbiased trajectories launched from the positions of the images along the curve to determine the optimal pathway. Because it relies on a large number of independent trajectories, the algorithm scales extremely well on large supercomputers,[22] and this becomes especially effective with applications to hybrid quantum mechanical-molecular mechanical (QM/MM) simulations due to the poor scaling displayed by ab initio codes.[23] The method has been used to characterize a conformational transition in very large macromolecular systems, including the activation of c-Src tyrosine kinase[24,25] cholesterol flip-flop in lipid membranes,[26] the movement of the voltage-sensor of K^+ channels,[27] the alternating-access mechanism in the sarcoplasmic reticulum calcium pump (SERCA),[28] and the chemo-mechanical coupling in V-type ATPases.[29]

It has been shown that the original mean force method of Maragliano et al.[6] and the swarms algorithm[19] are essentially equivalent in the limit of very short trajectories.[30] However, the significance of the swarms algorithm for longer trajectories remains unclear. Intuitive arguments relying on a physical picture of overdamped diffusional dynamics suggest that the mean drifts from the swarms-of-trajectories may have the ability to more correctly capture the effective behavior of the system supporting the construction of meaningful transition pathways.[19,21] But the lack of a rigorous theoretical treatment has prevented the analysis to go any further.

The goal of the present effort is to return to this unresolved issue and clarify the fundamental underpinning of the string method with swarms-of-trajectories. After a brief summary of the developments leading to the original mean force string method,[6] we begin by formulating the problem of the kinetics of a dynamical system comprising two metastable states A and B in terms of a finite-time propagator in phase space (position and velocity) adapted to the underdamped Langevin equation. This development is inspired by previous work based on propagators for position-based Markov models.[31−34] From this formulation, an effective propagator in the space of the CVs is defined, and its properties are clarified. It is shown that a formulation of the propagator with a finite lag time helps to clarify the conditions for Markovity within the subspace of the CVs. On the basis of this analysis, the significance of thea string method with swarms-of-trajectories with finite lag time is clarified.

■ ANALYSIS

String Method in Collective Variables. For the sake of clarity and completeness, we briefly recall the formal developments leading to the original mean force string method.[6] We first consider a system evolving according to the underdamped Langevin equation

$$m_i\dot{v}_i = -\frac{\partial U}{\partial x_i} - \gamma v_i + f_i(t)$$
$$\dot{x}_i = v_i \tag{1}$$

with $\langle f_i\rangle = 0$ and $\langle f_i(t)f_j(t')\rangle = 2\gamma k_B T\delta_{ij}\delta(t-t')$. The generic friction γ is assumed to be very small and is introduced primarily to act as a weak thermostat. The equilibrium distribution is $\rho_{eq}(\mathbf{x},\mathbf{v}) = e^{-H(\mathbf{x},\mathbf{v})/k_BT}/\int d\mathbf{x}\, d\mathbf{v}\, e^{-H(\mathbf{x},\mathbf{v})/k_BT}$, where $H(\mathbf{x},\mathbf{v}) = U(\mathbf{x}) + \sum_{i=1}^{N} 1/2 m_i v_i^2$ is the Hamiltonian.

It is assumed that the system comprises two metastable states A and B (defined by the user). The forward committor $q_+^B(\mathbf{x},\mathbf{v})$ is the probability that a path starting at $(\mathbf{x},\mathbf{v})$ ultimately reaches the state B before ever reaching the state A. The committor $q_+^B(\mathbf{x},\mathbf{v})$ is defined as $\mathcal{L}q_+^B(\mathbf{x},\mathbf{v}) = 0$, with the constraints $q_+^B(\mathbf{x},\mathbf{v}) = 0$ when $(\mathbf{x},\mathbf{v}) \in$ A, and $q_+^B(\mathbf{x},\mathbf{v}) = 1$ when $(\mathbf{x},\mathbf{v}) \in$ B, where $\mathcal{L}$ is the backward Kolmogorov propagator.

$$\mathcal{L} = \sum_{i=1}^{n} v_i\frac{\partial}{x_i} - \frac{\partial U(\mathbf{x})}{\partial x_i}\frac{\partial}{\partial v_i} + \frac{\gamma}{m_i}\left(-v_i\frac{\partial}{\partial v_i} + \frac{1}{k_BT}\frac{\partial^2}{\partial {v_i}^2}\right) \tag{2}$$

The forward committor $q_+^B(\mathbf{x},\mathbf{v})$ can be also obtained through a variational formulation by seeking to minimizes I[2,6]

$$I = \frac{1}{2}\int d\mathbf{x}\, d\mathbf{v}\, \rho_{eq}(\mathbf{x},\mathbf{v})|\mathcal{L}q|^2 \tag{3}$$

over all trial functions $q(\mathbf{x},\mathbf{v})$ satisfying the constraints for the state A and B.[6]

The string method of Maragliano et al.[6] represents the pathway linking A and B as a "chain of state", that is, a collection of M images located at the positions $\{\mathbf{z}^1, ..., \mathbf{z}^M\}$ in a subspace of collective variables (CVs). Here, $\mathbf{z}$ represents a vector-valued function, $\tilde{\mathbf{z}}(\mathbf{x}) = (\tilde{z}_1(\mathbf{x}), ..., \tilde{z}_N(\mathbf{x}))$, that maps every configuration $\mathbf{x}$ of the system on a set of values $\tilde{\mathbf{z}}(\mathbf{x})$. The central asantz to derive the string method in CVs is to assume a trial committor $\overline{q}$ that depends only on $\mathbf{z}$.[6] Inserting $\overline{q}(\mathbf{z})$ in eq 3, I is reduced as

$$I = \int d\mathbf{z}\, \overline{\rho}_{eq}(\mathbf{z})(\nabla\overline{q}\cdot\mathbf{M}\cdot\nabla\overline{q}) \tag{4}$$

with $\nabla\overline{q} = (\partial\overline{q}/\partial z_1, \partial\overline{q}/\partial z_2, ...)$. The reduced equilibrium probability $\overline{\rho}_{eq}(\mathbf{z})$ is defined in terms of $W(\mathbf{z})$, the PMF associated with these variables

$$\overline{\rho}_{eq}(\mathbf{z}) = \langle\delta(\tilde{\mathbf{z}}(\mathbf{x}) - \mathbf{z})\rangle$$
$$= \int d\mathbf{x}\, d\mathbf{v}\, \delta(\tilde{\mathbf{z}}(\mathbf{x}) - \mathbf{z})\rho_{eq}(\mathbf{x},\mathbf{v})$$
$$= \frac{e^{-W(\mathbf{z})/k_BT}}{\int d\mathbf{z}\, e^{-W(\mathbf{z})/k_BT}} \tag{5}$$

and the components of $\mathbf{M}(\mathbf{z})$ are given by the conditional average

$$M_{\alpha\gamma}(\mathbf{z}) = \left\langle \delta(\tilde{\mathbf{z}}(\mathbf{x}) - \mathbf{z})\sum_{i=1}^{3n}\frac{1}{m_i}\frac{\partial\tilde{z}_\alpha(\mathbf{x})}{\partial x_i}\frac{\partial\tilde{z}_\gamma(\mathbf{x})}{\partial x_i}\right\rangle \frac{1}{\langle\delta(\tilde{\mathbf{z}}(\mathbf{x}) - \mathbf{z})\rangle}$$
$$= \left\langle \sum_{i=1}^{3n}\frac{1}{m_i}\frac{\partial\tilde{z}_\alpha(\mathbf{x})}{\partial x_i}\frac{\partial\tilde{z}_\gamma(\mathbf{x})}{\partial x_i}\right\rangle_{\tilde{\mathbf{z}}(\mathbf{x})=\mathbf{z}} \tag{6}$$

https://doi.org/10.1021/acs.jpca.1c04110
J. Phys. Chem. A 2021, 125, 7558−7571

Having performed a dimensionality reduction leading to eq 4, one then seeks to minimize I over all functions $\overline{q}$ subject to the constraint $\overline{q} = 0$ when $\mathbf{z} \in$ A, and $\overline{q} = 1$ when $\mathbf{z} \in$ B. The Euler−Lagrange equation for the effective committor $\overline{q}(\mathbf{z})$ is

$$\nabla\cdot(e^{-W(\mathbf{z})/k_BT}\mathbf{M}(\mathbf{z})\cdot\nabla\overline{q}(\mathbf{z})) = 0 \tag{7}$$

subject to the constraint $\overline{q}(\mathbf{z}) = 0$ when $\mathbf{z} \in$ A and $\overline{q}(\mathbf{z}) = 1$ when $\mathbf{z} \in$ B. The MFEP corresponds to the curve (or line) linking the states A and B defined such that its tangent follows the direction of the vector $\mathbf{v}$.

$$\mathbf{v} = \mathbf{M}(\mathbf{z})\cdot\nabla W(\mathbf{z}) + k_BT\nabla\cdot\mathbf{M}(\mathbf{z}) \tag{8}$$

The curve can be generated, for example, by starting from a saddle point in the space $\mathbf{z}$ and following the direction $\mathbf{v}$ in small increment or, equivalently, by building up a curve such that the projection of $\mathbf{v}$ in the direction perpendicular to its tangent vanishes, that is, $[\mathbf{v}]^{\perp} = 0$. The term $\nabla\cdot\mathbf{M}(\mathbf{z})$ is often negligible, and in this case, the MFEP is a curve tangent to the vector $\mathbf{M}(\mathbf{z})\cdot\nabla W(\mathbf{z})$ in the space of the collective variables that connects the two states A and B. These results summarize the mean force string method in collective variables of Maragliano et al.[6] Conceptually, the developments leading to this theory represent one formal "route" to achieve a dimensionality reduction to the subspace of the CVs by inserting $\overline{q}(\mathbf{z})$ in eq 3.

For the sake of argument, let us assume that, by virtue of the particular system of interest, it is a physically valid approximation to treat $\mathbf{z}$ as a set of slow variables. In other words, the CVs genuinely undergo an overdamped dynamics on the free energy surface $W(\mathbf{z})$ with a diffusion coefficient $D_\alpha\ \gamma(\mathbf{z})$ according to the Smoluchowski equation. This statement of fact notwithstanding, both eq 7 for the effective committor $\overline{q}(\mathbf{z})$ and eq 8 defining the tangent to the curvilinear path underlying the mean force string method remain unchanged. On the one hand, the mean force string method based on eqs 5, (6), and (8) is effectively "blind" to the nature of the microscopic dynamics of the CVs. On the other hand, since we know with confidence that the CVs are indeed slow variables, the committor $\overline{q}(\mathbf{z})$ can be determined from the backward Kolmogorov equation (see eq 2.5 in ref 35)

$$\nabla\cdot(e^{-W(\mathbf{z})/k_BT}\mathbf{D}(\mathbf{z})\cdot\nabla\overline{q}(\mathbf{z})) = 0 \tag{9}$$

subject to the constraint $\overline{q}(\mathbf{z}) = 0$ when $\mathbf{z} \in$ A and $\overline{q}(\mathbf{z}) = 1$ when $\mathbf{z} \in$ B. Accordingly, the optimal path should be determined by

$$[-\mathbf{D}(\mathbf{z})\cdot\nabla W(\mathbf{z}) + k_BT\nabla\cdot\mathbf{D}(\mathbf{z})]^{\perp} = 0 \tag{10}$$

This treatment treating the CVs as slow variables represents a different formal "route" to achieve a dimensionality reduction to the subspace of the CVs. Clearly, a comparison of eqs 7 and (9) shows that the two routes do not yield the same result.

When the dimensionality is reduced by an insertion of the asantz $\overline{q}(\mathbf{z})$ in eq 3, the optimal path is determined from the mean force string method eq 8. In this case, the path is affected by the PMF and the matrix $\mathbf{M}$. In contrast, when the CVs are slow and the problem is first reduced to an overdamped diffusional dynamics, the optimal path is determined from eq 10. In this case, the path is affected by the potential W and the diffusion matrix $\mathbf{D}$. While the matrices $\mathbf{D}$ and $\mathbf{M}$ occupy a similar place in the equation defining the committor $\overline{q}(\mathbf{z})$, $\mathbf{D}$ is associated with dynamical dissipative effects, whereas $\mathbf{M}$ is a mass-weighted geometric factor reporting the change in the curvilinear collective variables elicited by a corresponding change in the Cartesian coordinates. For instance, the matrix $\mathbf{M}$ is simply constant and diagonal if the CVs are Cartesian variables. Specifically, the matrix $\mathbf{M}$ does not incorporate dynamical dissipative effects, which would be needed to characterize the CVs as slow variables.

Fundamentally, the discrepancy stems from the transition from underdamped inertial dynamics to overdamped diffusional dynamics. Somehow, the analysis that leads to eq 7 does not lead to eq 9 when the CVs are genuinely slow variables. Traditionally, the behavior of slow degrees of freedom is revealed through an analysis of an effective dynamics based on projection operators.[36−38] The expectation is that the slow dissipative dynamics of the CVs should be reflected in the time evolution of the system. However, to display the dynamical evolution of the CVs, the propagation of the system over a finite time is required, and this information is not available with the operator $\mathcal{L}$ in the variational principle expressed by eq 3. An alternative route is needed to perform this analysis. In the following, we will reconstruct the steps leading to the committor in a subspace of CVs in terms of a finite-time propagator for this purpose.

Microscopic Propagator. The underdamped Langevin dynamics eq 1 is prescribed by the Green's function propagator. The theoretical formulation seeking to build up the kinetic behavior from a finite-time propagator in phase space (positions and velocities) is inspired by previous developments of Markov model position-propagators.[31−34] Here, this general picture is expanded to underdamped Langevin dynamics. The probability density of the system at time t is expressed as $\rho(\mathbf{x},\mathbf{v};\ t)$, where $\mathbf{x}$ and $\mathbf{v}$ represent the set of coordinates x_i and velocities v_i, respectively. Using $\mathbf{u} \equiv (\mathbf{x},\mathbf{v})$ as a shorthand to represent the point in phase space, the forward propagation step $(\mathbf{u} \rightarrow \mathbf{u}')$ for the probability density from the time t to the time $t + \Delta t$ is

$$\rho(\mathbf{u}';\ t + \Delta t) = \int d\mathbf{u}\ \mathcal{P}(\mathbf{u}'|\mathbf{u};\ \Delta t)\rho(\mathbf{u};\ t) \tag{11}$$

The elementary propagator for a null time step, $\mathcal{P}(\mathbf{u}'|\mathbf{u};\ 0)$, is the identity $\delta(\mathbf{u}' - \mathbf{u})$, and the dynamical propagation, which we may formally represent as $\boldsymbol{\rho}(t + \Delta t) = \mathcal{P}(\Delta t)\cdot\boldsymbol{\rho}(t)$, obeys the Chapman-Kolmogorov equation for arbitrary times. The implication is that, while Δt is a microscopic time step (e.g., 1 fs), the propagator may be repeatedly applied an arbitrary number of times as $\boldsymbol{\rho}(t + \Delta t) = \mathcal{P}(n\Delta t)\cdot\boldsymbol{\rho}(t) \equiv \mathcal{P}(\Delta t)^n\cdot\boldsymbol{\rho}(t)$, with

$$\mathcal{P}(\Delta t)\cdot\mathcal{P}(\Delta t) \equiv \int d\mathbf{u}'\ \mathcal{P}(\mathbf{u}''|\mathbf{u}';\ \Delta t)\mathcal{P}(\mathbf{u}'|\mathbf{u};\ \Delta t) \tag{12}$$

The backward propagation step is $\mathcal{P}^{\dagger}(\mathbf{u}|\mathbf{u}';\ \Delta t) = \mathcal{P}(\mathbf{x}, -\mathbf{v}|\mathbf{x}', -\mathbf{v}';\ \Delta t)$. Probability is conserved by both the forward propagator $\int d\mathbf{u}'\ \mathcal{P}(\mathbf{u}'|\mathbf{u};\ \Delta t) = 1$ and backward propagator $\int d\mathbf{u}\ \mathcal{P}^{\dagger}(\mathbf{u}|\mathbf{u}';\ \Delta t) = 1$, implying that the unity function is a left eigenvector of these operators with an eigenvalue equal to 1. The total energy accumulated over the propagation correspond to the heat transferred from the bath to the system from the Langevin propagation, $Q = H(\mathbf{u}') - H(\mathbf{u})$. Because Q is the heat exchanged with the reservoir during the forward propagation step, it follows that $-Q$ is the heat exchanged with the reservoir for a propagation along the reverse step. Noting that $H(\mathbf{x},\mathbf{v}) = H(\mathbf{x},-\mathbf{v})$, and $\rho_{eq}(\mathbf{u}) = \rho_{eq}(\mathbf{x},-\mathbf{v})$, the forward−backward microscopic detailed balance relation

$$\mathcal{P}(\mathbf{u}'|\mathbf{u};\ \Delta t)\rho_{eq}(\mathbf{u}) = \mathcal{P}^{\dagger}(\mathbf{u}|\mathbf{u}';\ \Delta t)\rho_{eq}(\mathbf{u}') \tag{13}$$

https://doi.org/10.1021/acs.jpca.1c04110
J. Phys. Chem. A 2021, 125, 7558−7571

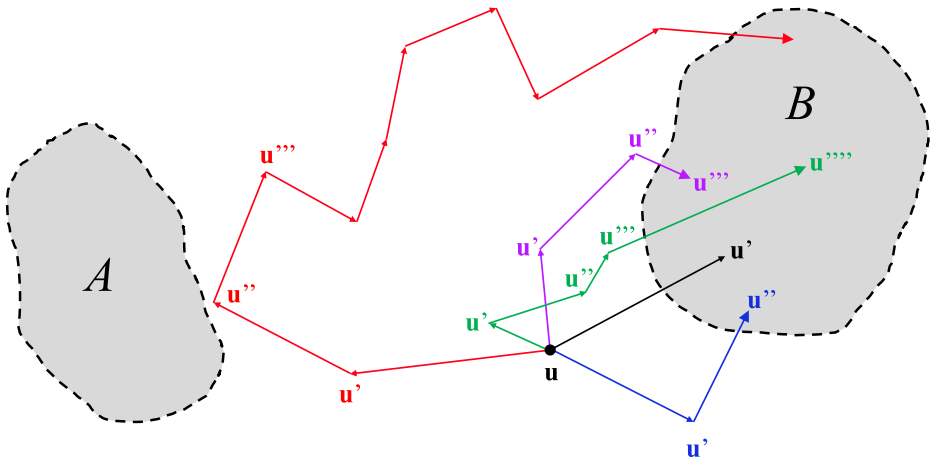


Figure 1. Schematic illustration of the forward committor $q_+^B(\mathbf{u})$ represented as the sum over all possible paths starting at $\mathbf{u}$ that ultimately reach the state B before ever reaching the state A. All paths start at $\mathbf{u}$ and end in B without entering the state A. The black path reaches the B region in one step ($\mathbf{u} \to \mathbf{u}'$), the blue path in two steps ($\mathbf{u} \to \mathbf{u}' \to \mathbf{u}''$), the purple path in three steps ($\mathbf{u} \to \mathbf{u}' \to \mathbf{u}'' \to \mathbf{u}'''$), and the green path in four steps ($\mathbf{u} \to \mathbf{u}' \to \mathbf{u}'' \to \mathbf{u}''' \to \mathbf{u}''''$). The restriction that the intermediate states resulting from all these steps are $\notin$ A,B applies to all paths. For example, the longer red path does not enter the boundary region of state A before reaching the state B in the final destination.

is satisfied. The momentum reversal requirement for microscopic reversibility is discussed in section 2.2.3 of ref 39. It follows that the canonical equilibrium distribution is invariant by propagation

$$\begin{aligned}\rho_{eq}(\mathbf{u}';\,\Delta t) &= \int d\mathbf{u}\,\mathcal{P}(\mathbf{u}'|\mathbf{u};\,\Delta t)\rho_{eq}(\mathbf{u}) \\ &= \int d\mathbf{u}\,\mathcal{P}^{\dagger}(\mathbf{u}|\mathbf{u}';\,\Delta t)\rho_{eq}(\mathbf{u}') \\ &= \left(\int d\mathbf{x}\int d\mathbf{v}\,\mathcal{P}^{\dagger}(\mathbf{x},-\mathbf{v}|\mathbf{x}',-\mathbf{v}';\,\Delta t)\right)\rho_{eq}(\mathbf{u}') \\ &= \rho_{eq}(\mathbf{u}')\end{aligned} \tag{14}$$

Thus, $\boldsymbol{\rho}_{eq} = \mathcal{P}(\Delta t)\cdot\boldsymbol{\rho}_{eq}$, showing that $\boldsymbol{\rho}_{eq}$ is a right-eigenvector with an eigenvalue equal to 1.

Committor Probabilities for Two Metastable States. Assuming two metastable states A and B, the forward committor $q_+^B(\mathbf{u})$ is the sum of the probability over all paths starting at $\mathbf{u}$ that ultimately reach the state B before ever reaching the state A. The probability of each of these paths is expressed as a product of discrete propagation steps $\mathcal{P}(\Delta t)\cdots\mathcal{P}(\Delta t)$, under the restriction that the intermediate states resulting from all these steps are $\notin$ A, B. The sum over paths is illustrated schematically in Figure 1. It follows that $q_+^B(\mathbf{u})$ is written explicitly as

$$\begin{aligned}q_+^B(\mathbf{u}) &= \int_{\in B} d\mathbf{u}'\,\mathcal{P}(\mathbf{u}'|\mathbf{u};\,\Delta t) \\ &+ \int_{\notin A,B} d\mathbf{u}'\int_{\in B} d\mathbf{u}''\,\mathcal{P}(\mathbf{u}''|\mathbf{u}';\,\Delta t)\mathcal{P}(\mathbf{u}'|\mathbf{u};\,\Delta t) \\ &+ \int_{\notin A,B} d\mathbf{u}'\int_{\notin A,B} d\mathbf{u}''\int_{\in B} d\mathbf{u}''' \\ &\mathcal{P}(\mathbf{u}'''|\mathbf{u}'';\,\Delta t)\mathcal{P}(\mathbf{u}''|\mathbf{u}';\,\Delta t)\mathcal{P}(\mathbf{u}'|\mathbf{u};\,\Delta t) \\ &+ \ldots\end{aligned} \tag{15}$$

$$\begin{aligned}&= \int_{\in B} d\mathbf{u}'\,\mathcal{P}(\mathbf{u}'|\mathbf{u};\,\Delta t) + \int_{\notin A,B} d\mathbf{u}' \\ &\left[\int_{\in B} d\mathbf{u}''\,\mathcal{P}(\mathbf{u}''|\mathbf{u}';\,\Delta t)\right. \\ &+ \int_{\notin A,B} d\mathbf{u}''\int_{\in B} d\mathbf{u}'''\,\mathcal{P}(\mathbf{u}'''|\mathbf{u}'';\,\Delta t)\mathcal{P}(\mathbf{u}''|\mathbf{u}';\,\Delta t) \\ &\left.+\ldots\right]\mathcal{P}(\mathbf{u}'|\mathbf{u};\,\Delta t) \\ &= \int_{\in B} d\mathbf{u}'\,\mathcal{P}(\mathbf{u}'|\mathbf{u};\,\Delta t) + \int_{\notin A,B} d\mathbf{u}'\,[q_+^B(\mathbf{u}')] \\ &\quad \mathcal{P}(\mathbf{u}'|\mathbf{u};\,\Delta t) \\ &= \int_{\in A} d\mathbf{u}'\,(0) + \int_{\in B} d\mathbf{u}'\,(1)\mathcal{P}(\mathbf{u}'|\mathbf{u};\,\Delta t) \\ &+ \int_{\notin A,B} d\mathbf{u}'\,q_+^B(\mathbf{u}')\mathcal{P}(\mathbf{u}'|\mathbf{u};\,\Delta t) \\ &= \int_{\in A} d\mathbf{u}'\,q_+^B(\mathbf{u}') + \int_{\in B} d\mathbf{u}'\,q_+^B(\mathbf{u}')\mathcal{P}(\mathbf{u}'|\mathbf{u};\,\Delta t) \\ &+ \int_{\notin A,B} d\mathbf{u}'\,q_+^B(\mathbf{u}')\mathcal{P}(\mathbf{u}'|\mathbf{u};\,\Delta t) \\ &= \int d\mathbf{u}'\,q_+^B(\mathbf{u}')\mathcal{P}(\mathbf{u}'|\mathbf{u};\,\Delta t)\end{aligned} \tag{16}$$

with the constraints of $q_+^B(\mathbf{u}) = 0$ if $\mathbf{u} \in$ A and $q_+^B(\mathbf{u}) = 1$ if $\mathbf{u} \in$ B. By construction, $0 \le q_+^B(\mathbf{u}) \le 1$. This expression may be related to eq 2 by recognizing that the operator $\mathcal{L}$ is defined in the limit of an infinitesimal time step.

$$\mathcal{L}q_+^B(\mathbf{u}) = \lim_{\Delta t\to 0}\int d\mathbf{u}'\,q_+^B(\mathbf{u}')\left(\frac{\mathcal{P}(\mathbf{u}'|\mathbf{u};\,\Delta t) - \delta(\mathbf{u}-\mathbf{u}')}{\Delta t}\right) \tag{17}$$

A forward committor $q_+^A(\mathbf{u})$ can also be defined, satisfying the constraints $q_+^A(\mathbf{u}) = 1$ if $\mathbf{u} \in$ A and $q_+^A(\mathbf{u}) = 0$ if $\mathbf{u} \in$ B and the identity $q_+^A(\mathbf{u}) + q_+^B(\mathbf{u}) = 1$. In a similar fashion, a backward committor probability $q_-^A(\mathbf{u})$ can be defined from the backward propagation as

https://doi.org/10.1021/acs.jpca.1c04110
J. Phys. Chem. A 2021, 125, 7558−7571

$$q_-^{\mathrm{B}}(\mathbf{u}) = \int \mathrm{d}\mathbf{u}'\, q_-^{\mathrm{B}}(\mathbf{u}')\mathcal{P}^{\dagger}(\mathbf{u}'|\mathbf{u};\, \Delta t) \tag{18}$$

with the constraints of $q_-^{\mathrm{B}} = 0$ if $\mathbf{u} \in \mathrm{A}$ and $q_-^{\mathrm{B}} = 1$ if $\mathbf{u} \in \mathrm{B}$. A backward committor $q_-^{\mathrm{A}}(\mathbf{u})$ can be defined, satisfying the identity $q_-^{\mathrm{A}}(\mathbf{u}) + q_-^{\mathrm{B}}(\mathbf{u}) = 1$. Furthermore, it can be verified that $q_-^{\mathrm{A}}(\mathbf{u}) = q_+^{\mathrm{A}}(\mathbf{x},-\mathbf{v})$ and $q_-^{\mathrm{B}}(\mathbf{u}) = q_+^{\mathrm{B}}(\mathbf{x},-\mathbf{v})$.

While the equations for the committor probabilities involve only the elementary propagator $\mathcal{P}(\mathbf{u}'|\mathbf{u};\, \Delta t)$ for the short time Δt, the fundamental validity of these equations is predicated upon the necessity to satisfy the Markovity of the dynamics as expressed by the Chapman-Kolmogorov equation $\mathcal{P}(n\Delta t) \equiv \mathcal{P}(\Delta t)^n$.

Steady-State Flux between Two States. This section largely follows the development previously laid out by Berezhkovskii, Hummer, and Szabo[40] for continuous-time discrete-state Markov models. A number of established relations are briefly recapped for the sake of completeness. To elicit a net nonequilibrium flux from a metastable state A to metastable state B, we construct steady-state conditions as[40]

$$\rho_{\mathrm{ss}}(\mathbf{u}') = \int \mathrm{d}\mathbf{u}\, \mathcal{P}(\mathbf{u}'|\mathbf{u};\, \Delta t)\rho_{\mathrm{ss}}(\mathbf{u}) \tag{19}$$

with $\rho_{\mathrm{ss}}(\mathbf{u}) = \rho_{\mathrm{eq}}(\mathbf{u})$ if $\mathbf{u} \in \mathrm{A}$ and $\rho_{\mathrm{ss}}(\mathbf{u}) = 0$ if $\mathbf{u} \in \mathrm{B}$. The states A and B are defined by the user via indicator functions in the space of the CVs. The steady-state density can be written as

$$\rho_{\mathrm{ss}}(\mathbf{u}) = \rho_{\mathrm{eq}}(\mathbf{u})q_-^{A}(\mathbf{u}) \tag{20}$$

where $q_-^{\mathrm{A}}(\mathbf{u})$ is the backward committor probability defined with the constraints $q_{\mathrm{A}} = 1$ if $\mathbf{u} \in \mathrm{A}$ and $q_{\mathrm{A}} = 0$ if $\mathbf{u} \in \mathrm{B}$. This is verified by a direct substitution in the steady-state equation

$$\begin{aligned} q_-^{\mathrm{A}}(\mathbf{u}') &= \frac{1}{\rho_{\mathrm{eq}}(\mathbf{u}')}\int \mathrm{d}\mathbf{u}\, \mathcal{P}(\mathbf{u}'|\mathbf{u};\, \Delta t)\rho_{\mathrm{eq}}(\mathbf{u})q_-^{\mathrm{A}}(\mathbf{u}) \\ &= \frac{1}{\rho_{\mathrm{eq}}(\mathbf{u}')}\int \mathrm{d}\mathbf{u}\, \mathcal{P}^{\dagger}(\mathbf{u}|\mathbf{u}';\, \Delta t)\rho_{\mathrm{eq}}(\mathbf{u}')q_-^{\mathrm{A}}(\mathbf{u}) \\ &= \int \mathrm{d}\mathbf{u}\, q_-^{\mathrm{A}}(\mathbf{u})\mathcal{P}^{\dagger}(\mathbf{u}|\mathbf{u}';\, \Delta t) \end{aligned} \tag{21}$$

The steady-state flux from A to B can be expressed as the net transitions across a dividing surface separating the two sides

$$\begin{aligned} J^{\mathrm{AB}} &= \frac{1}{\Delta t}\int_{\in \mathrm{A}'} \mathrm{d}\mathbf{u} \int_{\in \mathrm{B}'} \mathrm{d}\mathbf{u}'\, (\mathcal{P}(\mathbf{u}'|\mathbf{u};\, \Delta t)\rho_{\mathrm{ss}}(\mathbf{u}) \\ &\quad - \mathcal{P}(\mathbf{u}|\mathbf{u}';\, \Delta t)\rho_{\mathrm{ss}}(\mathbf{u}')) \\ &= \frac{1}{\Delta t}\int_{\in \mathrm{A}'} \mathrm{d}\mathbf{u} \int_{\in \mathrm{B}'} \mathrm{d}\mathbf{u}'\, (\mathcal{P}(\mathbf{u}'|\mathbf{u};\, \Delta t)\rho_{\mathrm{eq}}(\mathbf{u})q_-^{\mathrm{A}}(\mathbf{u}) \\ &\quad - \mathcal{P}(\mathbf{u}|\mathbf{u}';\, \Delta t)\rho_{\mathrm{eq}}(\mathbf{u}')q_-^{\mathrm{A}}(\mathbf{u}')) \\ J^{\mathrm{AB}} &= \frac{1}{\Delta t}\int_{\in \mathrm{A}'} \mathrm{d}\mathbf{u} \int_{\in \mathrm{B}'} \mathrm{d}\mathbf{u}'\, (q_+^{\mathrm{B}}(\mathbf{u}') - q_-^{\mathrm{B}}(\mathbf{u}))\mathcal{P}(\mathbf{u}' \\ &\quad |\mathbf{u};\, \Delta t)\rho_{\mathrm{eq}}(\mathbf{u}) \end{aligned} \tag{22}$$

where the integral over $\mathbf{u}$ and $\mathbf{u}'$ are restricted to be on two different sides A′ and B′ of a separating surface between the two end states A and B, though the steady-state flux does not depend on the position of this dividing surface. Equivalently, this expression can also be obtained by combining the probability that a transition occurring from $\mathbf{u}$ to $\mathbf{u}'$ along a reactive trajectory from A to B at equilibrium, $q_+^{\mathrm{B}}(\mathbf{u}')\mathcal{P}(\mathbf{u}'|\mathbf{u};\, \Delta t)\rho_{\mathrm{eq}}(\mathbf{u})q_-^{\mathrm{A}}(\mathbf{u})$, and the corresponding probability for the opposite transition from $\mathbf{u}'$ to $\mathbf{u}$, to express the net steady-state flux.

Collective Variables and Dimensionality Reduction. Following Maragliano et al.,[6] we introduce an approximation to the exact committor function in phase space $\mathbf{u}$ to achieve a dimensionality reduction to the subspace of the CVs $\mathbf{z}$. We use the asantz $\overline{q}(\mathbf{z}') \approx q_+^{\mathrm{B}}(\mathbf{u}')$ and $\overline{q}(\mathbf{z}) \approx q_-^{\mathrm{B}}(\mathbf{u})$ in eq 22 to express the steady-state flux J^{AB} and then integrate out the orthogonal degrees of freedom

$$\begin{aligned} J^{\mathrm{AB}} &= \frac{1}{\Delta t}\int \mathrm{d}\mathbf{z} \int \mathrm{d}\mathbf{z}' \int_{\in \mathrm{A}'} \mathrm{d}\mathbf{u} \int_{\in \mathrm{B}'} \mathrm{d}\mathbf{u}' \\ &\quad \delta(\tilde{\mathbf{z}}(\mathbf{x}') - \mathbf{z}')\delta(\tilde{\mathbf{z}}(\mathbf{x}) - \mathbf{z}) \\ &\quad (\overline{q}(\mathbf{z}') - \overline{q}(\mathbf{z}))\mathcal{P}(\mathbf{u}'|\mathbf{u};\, \Delta t)\rho_{\mathrm{eq}}(\mathbf{u}) \\ &= \frac{1}{\Delta t}\int_{\in \mathrm{A}'} \mathrm{d}\mathbf{z} \int_{\in \mathrm{B}'} \mathrm{d}\mathbf{z}'(\overline{q}(\mathbf{z}') - \overline{q}(\mathbf{z}))\overline{\mathcal{P}}(\mathbf{z}' \\ &\quad |\mathbf{z};\, \Delta t)\overline{\rho}_{\mathrm{eq}}(\mathbf{z}) \end{aligned} \tag{23}$$

where the effective reduced propagator is defined as

$$\begin{aligned} \overline{\mathcal{P}}(\mathbf{z}'|\mathbf{z};\, \Delta t)\overline{\rho}_{\mathrm{eq}}(\mathbf{z}) &= \int \mathrm{d}\mathbf{u} \int \mathrm{d}\mathbf{u}'\, \delta(\tilde{\mathbf{z}}(\mathbf{x}') - \mathbf{z}') \\ &\quad \mathcal{P}(\mathbf{u}'|\mathbf{u};\, \Delta t)\rho_{\mathrm{eq}}(\mathbf{u})\delta(\tilde{\mathbf{z}}(\mathbf{x}) - \mathbf{z}) \end{aligned} \tag{24}$$

The effective propagator obeys the detailed balance relation

$$\overline{\mathcal{P}}(\mathbf{z}'|\mathbf{z};\, \Delta t)\overline{\rho}_{\mathrm{eq}}(\mathbf{z}) = \overline{\mathcal{P}}(\mathbf{z}|\mathbf{z}';\, \Delta t)\overline{\rho}_{\mathrm{eq}}(\mathbf{z}') \tag{25}$$

as shown by integration of eq 13. Because the steady-state flux J^{AB} in eq 23 does not depend on the position of the dividing surface defining the A′ and B′ regions, it is convenient to choose a dividing surface corresponding to an isocommittor surface of $\overline{q}(\mathbf{z})$ at some arbitrary value q^*. The flux from A to B can then be written as transitions from the point $\mathbf{z}$ with committor $\overline{q}(\mathbf{z}) < q^*$, to the point $\mathbf{z}'$ with committor $\overline{q}(\mathbf{z}') > q^*$.

$$\begin{aligned} J^{\mathrm{AB}}(q^*) &= \frac{1}{\Delta t}\int \mathrm{d}\mathbf{z} \int \mathrm{d}\mathbf{z}'\, \theta(\overline{q}(\mathbf{z}') - q^*)\theta(q \\ &\quad * - \overline{q}(\mathbf{z})) \\ &\quad (\overline{q}(\mathbf{z}') - \overline{q}(\mathbf{z}))\overline{\mathcal{P}}(\mathbf{z}'|\mathbf{z};\, \Delta t)\overline{\rho}_{\mathrm{eq}}(\mathbf{z}) \end{aligned} \tag{26}$$

However, one can also write

$$J^{\mathrm{AB}} = \int_0^1 \mathrm{d}q^*\, J^{\mathrm{AB}}(q^*) \tag{27}$$

because the steady-state flux $J^{\mathrm{AB}}(q^*)$ does not actually depend on the specific value of q^*, which can be demonstrated by showing that there is no accumulation of probability in the region between the states A and B.[41] Performing the integration in the expression above affects only the term

$$\begin{aligned} &\int_0^1 \mathrm{d}q^*\, \theta(\overline{q}(\mathbf{z}') - q^*)\theta(q^* - \overline{q}(\mathbf{z})) \\ &\quad = (\overline{q}(\mathbf{z}') - \overline{q}(\mathbf{z}))\theta(\overline{q}(\mathbf{z}') - \overline{q}(\mathbf{z})) \end{aligned} \tag{28}$$

yield the quadratic expression

https://doi.org/10.1021/acs.jpca.1c04110
J. Phys. Chem. A 2021, 125, 7558−7571

$$J^{AB} = \frac{1}{\Delta t}\int d\mathbf{z}\int d\mathbf{z}'\,\theta(\bar{q}(\mathbf{z}') - \bar{q}(\mathbf{z}))$$
$$(\bar{q}(\mathbf{z}') - \bar{q}(\mathbf{z}))^2\bar{\mathcal{P}}(\mathbf{z}'|\mathbf{z};\,\Delta t)\rho_{eq}(\mathbf{u})$$
$$= \frac{1}{2\Delta t}\int d\mathbf{z}\int d\mathbf{z}'\,(\bar{q}(\mathbf{z}') - \bar{q}(\mathbf{z}))^2$$
$$\bar{\mathcal{P}}(\mathbf{z}'|\mathbf{z};\,\Delta t)\bar{\rho}_{eq}(\mathbf{z}) \tag{29}$$

(the factor of 2 is needed when the restriction $\bar{q}(\mathbf{z}') > \bar{q}(\mathbf{z})$ is removed). An integration of the step functions analogous to eq 28 was previously used by Krivov in a different situation.[42] A similar steady-state flux expression, quadratic in the committor difference, has also appeared in the context of discrete-state Markov models.[43] A transformation of the steady-state flux in the phase space eq 22 through steps equivalent to eqs 26−(29) could not be found by this author. eq 29 can be used as a variational principle. The minimization of J^{AB} with respect to the trial function $\bar{q}(\mathbf{z})$ subject to the constraint $\bar{q}(\mathbf{z}) = 0$ when $\mathbf{z} \in A$ and $\bar{q}(\mathbf{z}) = 1$ when $\mathbf{z} \in B$ yields

$$\bar{q}(\mathbf{z}) = \int d\mathbf{z}'\,\bar{q}(\mathbf{z}')\bar{\mathcal{P}}(\mathbf{z}'|\mathbf{z};\,\Delta t) \tag{30}$$

The relation to the mean force string method via the committor defined by eqs 4−(7) can be established by assuming that the time step Δt is extremely short and noting that $\bar{q}(\mathbf{z}') = \bar{q}(\mathbf{z}) + \nabla\bar{q}(\mathbf{z})\cdot(\mathbf{z}' - \mathbf{z}) + ...$, with $(\mathbf{z}' - \mathbf{z}) = \tilde{\mathbf{z}}(\mathbf{x} + \mathbf{v}\,\Delta t) - \tilde{\mathbf{z}}(\mathbf{x})$. Following through with eq 29 yields

$$J^{AB} = \frac{1}{2\Delta t}\int d\mathbf{z}\int d\mathbf{z}'\left(\sum_\alpha\sum_i \frac{\partial\bar{q}(\mathbf{z})}{\partial z_\alpha}\frac{\partial\tilde{z}_\alpha(\mathbf{x})}{\partial x_i}\Delta t v_i\right)$$
$$\left(\sum_\gamma\sum_j \frac{\partial\bar{q}}{\partial z_\gamma}\Delta t v_j\frac{\partial\tilde{z}_\gamma(\mathbf{x})}{\partial x_i}\right)\bar{\mathcal{P}}(\mathbf{z}'|\mathbf{z};\,\Delta t)\bar{\rho}_{eq}(\mathbf{z})$$
$$= \frac{\Delta t}{2}\int d\mathbf{z}\,\bar{\rho}_{eq}(\mathbf{z})(\nabla\bar{q}(\mathbf{z})\cdot\mathbf{M}(\mathbf{z})\cdot\nabla\bar{q}(\mathbf{z})) \tag{31}$$

The components of the matrix $\mathbf{M}(\mathbf{z})$, previously introduced in eq 6, are recovered from the conditional average

$$M_{\alpha\gamma}(\mathbf{z}) = \left\langle\sum_{ij} v_i v_j\frac{\partial\tilde{z}_\alpha(\mathbf{x})}{\partial x_i}\frac{\partial\tilde{z}_\gamma(\mathbf{x})}{\partial x_j}\right\rangle_{\tilde{\mathbf{z}}(\mathbf{x})=\mathbf{z}} \tag{32}$$

with $\langle v_i\,v_j\rangle_{\mathbf{z}} = \delta_{ij}k_BT/m_i$. The minimization of eq 31 with respect to all possible functions $\bar{q}(\mathbf{z})$ recovers eq 7, which leads to the mean force string method eq 8 of Maragliano et al.[6] The equivalence is based on the requirement that the time step Δt is extremely short.

The consequence of the dimensionality reduction to the subspace of the CVs is perhaps revealed most transparently with the following observations. The function $\bar{q}(\mathbf{z})$ is determined by the effective propagator $\bar{\mathcal{P}}(\mathbf{z}'|\mathbf{z};\,\Delta t)$ through eq 30. The effective propagator $\bar{\mathcal{P}}(\mathbf{z}'|\mathbf{z};\,\Delta t)$, according to eq 24, corresponds to a Boltzmann-weighted average of the microscopic propagator $\mathcal{P}(\mathbf{u}'|\mathbf{u};\,\Delta t)$ for a single step Δt with respect to the initial conditions $\mathbf{u}$ constrained by $\mathbf{z}$ and a simple sum over all states constrained $\mathbf{z}'$. Recalling eq 7 from the analysis of the conditions for optimal dimensionality reduction of multistate Markov state models by Hummer and Szabo,[44] eq 24 is plainly recognized as a short-time local equilibrium approximation to a dimensionality reduction of the transition matrix between microstates to an effective transition matrix between macrostates. While a number of factors may be taken into consideration in attempting an optimal dimensionality reduction of a multistate Markov model,[44,45] it is likely that this effective propagator may not be valid at a long time within the subspace of the CVs. It is legitimate to ask if the propagation within the subspace of the CVs is Markovian in the sense that $\bar{\boldsymbol{\rho}}(t + n\Delta t) \stackrel{?}{=} \bar{\mathcal{P}}(\Delta t)^n\cdot\bar{\boldsymbol{\rho}}(t)$. If the effective propagator is not Markovian, then eq 30 does not yield a genuine committor that represents the total probability as a sum over all possible paths starting at $\mathbf{z}$ that reach the state B before reaching the state A. For example, the forward committor defined by eq 16 implicitly assumes that a valid sum over paths can be performed as expressed by eq 15. The Markovity of the effective dynamical propagator within the reduced subspace is a necessary condition for eq 30 to yield a meaningful committor $\bar{q}(\mathbf{z})$.

Markovity of the Effective Propagator. Let us consider the effective propagator within the subspace of the CVs for a lag time τ.

$$\bar{\mathcal{P}}(\mathbf{z}'|\mathbf{z};\,\tau) = \frac{1}{\bar{\rho}_{eq}(\mathbf{z})}\int d\mathbf{u}\int d\mathbf{u}'\,\delta(\tilde{\mathbf{z}}(\mathbf{x}') - \mathbf{z}')$$
$$\mathcal{P}(\mathbf{u}'|\mathbf{u};\,\tau)\rho_{eq}(\mathbf{u})\delta(\tilde{\mathbf{z}}(\mathbf{x}) - \mathbf{z}) \tag{33}$$

The Markovity of the effective propagator within the subspace of the CVs implies that the Chapman-Kolmogorov equation is satisfied, $\bar{\boldsymbol{\rho}}(t + m\tau) = \bar{\mathcal{P}}(\tau)^m\cdot\bar{\boldsymbol{\rho}}(t)$. Only then, the dimensionality reduction to the subspace of CVs will yield a self-consistent representation of the dynamics of the system within this subspace (closure of the dynamical propagation). An important framework to examine this issue is to rely on a spectral decomposition of the effective dynamical propagator.[31,32] The right-eigenvectors $\psi^R_i(\mathbf{z})$ of the effective operator with lag time τ are defined as

$$\lambda_i\psi^R_i(\mathbf{z}') = \int d\mathbf{z}\,\bar{\mathcal{P}}(\mathbf{z}'|\mathbf{z};\,\tau)\psi^R_i(\mathbf{z}) \tag{34}$$

The eigenvector $\psi^R_1(i)$ with eigenvalue $\lambda_1 = 1$ corresponds to an invariant state and is equal to the equilibrium vector $\bar{\rho}_{eq}(\mathbf{z})$. There is also a set of associated orthogonal left-eigenvectors

$$\lambda_i\psi^L_i(\mathbf{z}) = \int d\mathbf{z}'\,\psi^L_i(\mathbf{z}')\bar{\mathcal{P}}(\mathbf{z}'|\mathbf{z};\,\tau) \tag{35}$$

with

$$\delta_{kl} = \int d\mathbf{z}\,\psi^L_i(\mathbf{z})\psi^R_l(\mathbf{z}) \tag{36}$$

and $\psi^L_i(\mathbf{z}) = \psi^R_i(\mathbf{z})\,\bar{\rho}_{eq}(\mathbf{z})^{-1}$. The first right-eigenvector is actually the equilibrium distribution $\psi^R_1(\mathbf{z}) = \bar{\rho}_{eq}(\mathbf{z})$.

In the context of the underlying microscopic dynamics, the validity of the effective propagator is revealed by examining the matrix element for an arbitrary time $n\tau$

$$\mathcal{P}_{ij}(n\tau) = \int d\mathbf{u}'\int d\mathbf{u}''\,\psi^L_i(\tilde{\mathbf{z}}(\mathbf{x}'))\mathcal{P}(\mathbf{u}'|\mathbf{u};\,n\tau)\psi^R_j(\tilde{\mathbf{z}}(\mathbf{x})) \tag{37}$$

in the context of the microscopic propagator. One can write formally the microscopic propagator as

$$\mathcal{P}(\mathbf{u}'|\mathbf{u};\,n\tau) = \sum_k \Psi^R_k(\mathbf{u}')e^{\omega_k n\tau}\Psi^L_k(\mathbf{u}) \tag{38}$$

https://doi.org/10.1021/acs.jpca.1c04110
J. Phys. Chem. A 2021, 125, 7558−7571

where $\Psi_k^R(\mathbf{u})$ and $\Psi_k^L(\mathbf{u})$ are the genuine right- and left-eigenvector of the microscopic propagator, with $\Psi_i^L(\mathbf{u}) = \Psi_i^R(\mathbf{u})\rho_{eq}(\mathbf{u})^{-1}$. As the underlying microscopic dynamics is generated from the underdamped Langevin eq 1, a spectral decomposition of the microscopic propagator $\mathcal{P}(\mathbf{u}'|\mathbf{u}; n\tau)$ is expected to be extremely complex, reaching far beyond the scope of the present analysis. To sketch the main features of the spectral decomposition of the microscopic propagator, it is helpful to draw from a treatment of Langevin modes in macromolecules.[46] Such an analysis indicates that the eigenvalues ω_k of the microscopic propagator can be real or complex. Complex eigenvalues come by pairs with their complex conjugate ω_k*. The real part of these eigenvalues is negative, implying that all modes are ultimately decaying. Furthermore, a formal spectral decomposition of the Kramers-Fokker−Planck propagator indicates that its largest eigenvalues, controlling the long-time behavior, are real.[47] One can imagine that complex eigenvalues are associated with high-frequency underdamped oscillatory motions, whereas real negative eigenvalues are associated with slow transitions and diffusive motions. There is also the existence of a time-independent stable equilibrium distribution, corresponding to a right-eigenvector with an eigenvalue of 0.

$$\mathcal{P}_{ij}(n\tau) = \int d\mathbf{u}' \int d\mathbf{u}'' \, \psi_i^L(\tilde{\mathbf{z}}(\mathbf{x}'))\mathcal{P}(\mathbf{u}'|\mathbf{u}; n\tau)\psi_j^R(\tilde{\mathbf{z}}(\mathbf{x})) = \sum_k (\psi_i^L \cdot \Psi_k^R) e^{\omega_k n\tau} (\Psi_k^L \cdot \psi_j^R) \tag{39}$$

where $(\psi_i^L \cdot \Psi_k^R) = \int d\mathbf{u}' \, \psi_i^L(\tilde{\mathbf{z}}(\mathbf{x}'))\, \Psi_k^R(\mathbf{u}')$, and $(\Psi_k^L \cdot \psi_j^R) = \int d\mathbf{u}\, \Psi_k^L(\mathbf{u})\, \psi_i^R(\tilde{\mathbf{z}}(\mathbf{x}))$. With an assumption of the existence of some cutoff ω*, it is hoped that the slowest modes of the effective propagator with $\| \omega_k\| \leq \omega^*$ will correspond accurately to the slowest modes of the microscopic propagator. It follows that $(\Psi_k^R \cdot \psi_i^L) \approx \delta_{ki}$ and $(\Psi_k^L \cdot \psi_j^R) \approx \delta_{jk}$, with the matrix elements $\mathcal{P}_{ij}(n\tau) = \delta_{ij}\lambda_i(n\tau)$, and the eigenvalues of the slow modes are consistent with the inherent time scales of the microscopic propagator $\lambda_i(n\tau) = e^{\omega_i n\tau}$. In this case, the effective propagator $\overline{\mathcal{P}}$ obeys the Chapman-Kolmogorov equation, and the propagation is Markovian within the subspace supported by this set of eigenvectors. The remaining eigenvectors correspond to faster motions with eigenvalues above the threshold, $\|\omega_k\| > \omega^*$. While those modes were determined as eigenvectors of the effective propagator with a lag time of τ, they may not accurately represent the true eigenvectors of the underlying microscopic propagator $\mathcal{P}$. These imperfect eigenvectors are expected to overlap with multiple fast modes of the microscopic propagator according to eq 38. As a result, a propagation within the subspace supported by those faster modes does not satisfy the Chapman-Kolmogorov equation and is not Markovian.

One solution to ensure Markovity is to project out the contribution from those modes and reconstruct the effective propagator only from the subspace of the slowest modes. This can be achieved by probing the microscopic propagator in eq 33 with a longer lag time τ, ensuring that the amplitude of the undesired contributions has sufficiently decayed away. Furthermore, as the density of fast modes is expected to be very high, a propagation for a relatively short time would rapidly cause a destructive dephasing of these contributions. In practice, one seeks to determine the smallest possible lag time that achieves Markovity for the effective propagator. The implication of this analysis is that we may have to consider a reduced propagator for a time τ longer than the microscopic time step Δt. This further clarifies the origin of the apparent discrepancy between the string in CVs defined by eq 8 and the string for slow CVs defined by eq 10. The mean force string method eq 8 is derived by starting from the variational principle of eq 3 expressed in terms of the $\mathcal{L}$ operator associated with an infinitesimal time step in the limit of $\Delta t \to 0$, as shown in eq 17. As displayed through eqs 31 and (32), this makes it impossible to subsequently consider the dynamical propagation of the system over a finite lag time, which is required to reveal the conditions required to achieve Markovity in the case of slow CVs.

Once the smallest possible τ that ensures Markovity of the effective propagator has been determined, we can correctly define the forward committor for the effective dynamics within the subspace $\mathbf{z}$

$$\overline{q}(\mathbf{z}) = \int d\mathbf{z}' \, \overline{q}(\mathbf{z}')\overline{\mathcal{P}}(\mathbf{z}'|\mathbf{z}; \tau) \tag{40}$$

with the constraints $\overline{q} = 0$ when $\mathbf{z} \in$ A and $\overline{q} = 1$ when $\mathbf{z} \in$ B. While the expression has the same form as eq 30, a critical difference is that the solution of eq 40 yields the genuine forward committor, because the effective propagator $\overline{\mathcal{P}}(\mathbf{z}'|\mathbf{z}; \tau)$ is Markovian. This also yields the steady-state flux from A to B.

$$J^{AB} = \frac{1}{2\tau} \int d\mathbf{z} \int d\mathbf{z}' \, (\overline{q}(\mathbf{z}') - \overline{q}(\mathbf{z}))^2 \overline{\mathcal{P}}(\mathbf{z}'|\mathbf{z}; \tau)\overline{\rho}_{eq}(\mathbf{z}) = \frac{1}{2\tau}\langle(\overline{q}(\tau) - \overline{q}(0))^2\rangle \tag{41}$$

A similar expression has been derived by Krivov and co-workers.[48,49]

It is interesting to note that the steady-state flux J^{AB} expressed as an equilibrium time-correlation function is related to the mean square deviation of the committor. Because the steady-state flux J^{AB} must be independent of the lag time τ, the form of eq 41 also indicates that the time-correlation function $\langle(\overline{q}(\tau) - \overline{q}(0))^2\rangle$ must increase linearly with τ. However, while this expression is reminiscent of the familiar expression defining the diffusion coefficient, it does not imply that the time-dependent dynamics along $\overline{q}$ is neither diffusive nor Markovian. Importantly, eq 41 can be used to construct a robust variational principle for the determination of the forward committor. Assuming a "trial" function $r(\mathbf{z})$, the correct committor $\overline{q}(\mathbf{z})$ as defined by eq 40 can be found by finding the minimum of J^{AB} under the constraints of $r = 0$ when $\mathbf{z} \in$ A and $r = 1$ when $\mathbf{z} \in$ B. This offers a powerful route to determine optimal approximations for the committor. Eqs 33, (40), and (41) are the central results of the present dynamical propagator formulation of the transition kinetics from state A to state B.

String Method with Swarms-of-Trajectories. In the implementation of the string method based on swarms-of-trajectories,[19] one launches an ensemble of short trajectories from a specific position $\mathbf{z}$ corresponding to each image m and then calculates the mean drift of those trajectories relative to their initial starting point, $\langle \Delta\mathbf{z}(\tau) \rangle_{\mathbf{z}}$. The present analysis clarifies the significance of this procedure. By definition, the mean drift starting from $\mathbf{z}$ for a trajectory of length τ is

https://doi.org/10.1021/acs.jpca.1c04110
J. Phys. Chem. A 2021, 125, 7558−7571

$$\langle \Delta \mathbf{z}(\tau) \rangle_{\mathbf{z}} = \int d\mathbf{u} \int d\mathbf{u}' \, (\tilde{\mathbf{z}}(\mathbf{x}') - \tilde{\mathbf{z}}(\mathbf{x})) \mathcal{P}(\mathbf{u}'|\mathbf{u}; \tau) \frac{\rho_{eq}(\mathbf{u}) \delta(\tilde{\mathbf{z}}(\mathbf{x}) - \mathbf{z})}{\int d\mathbf{x} \rho_{eq}(\mathbf{u}) \delta(\tilde{\mathbf{z}}(\mathbf{x}) - \mathbf{z})} = \int d\mathbf{z} \int d\mathbf{z}' \, (\mathbf{z}' - \mathbf{z}) \bar{\mathcal{P}}(\mathbf{z}'|\mathbf{z}; \tau) \tag{42}$$

It was shown previously that, in the limit of $\tau \to \Delta t$, the mean drifts $\langle \Delta \mathbf{z}(\Delta t) \rangle_{\mathbf{z}}$ is equal to $1/2 \, \Delta t^2 \, (-\mathbf{M} \cdot \nabla W + k_B T \nabla \cdot \mathbf{M})$.[30] It follows that the string optimized by using the mean drifts calculated with swarms-of-trajectories of length Δt is equivalent to the original mean force formulation of the string method with CVs.[6] However, as discussed above, there can be issues of non-Markovity with short trajectories in the limit of $\tau \to \Delta t$. Intuitively, there are reasons to believe that the mean drifts from trajectories of finite time τ may better capture dissipative factors that affect the optimal pathway.[19,21] The following analysis further clarifies this matter.

As an illustration of the consequence of a longer lag time, let us consider a case where the dynamics within the subspace of collective variables follows an overdamped Langevin equation. Presumably, the development should take the familiar form of the Smoluchowski equation. Assuming that $\mathbf{z}'$ is close to $\mathbf{z}$, we write $\bar{q}(\mathbf{z}') \approx \bar{q}(\mathbf{z}) + \nabla \bar{q}(\mathbf{z}) \cdot \Delta \mathbf{z}'(\tau)$ with $\Delta \mathbf{z}'(\tau) = \mathbf{z}'(\tau) - \mathbf{z}$. Following through with this expansion yields

$$J^{AB} = \frac{1}{2\tau} \int d\mathbf{z} \int d\mathbf{z}' \, (\nabla \bar{q}(\mathbf{z}) \cdot \Delta \mathbf{z}'(\tau))^2 \bar{\mathcal{P}}(\mathbf{z}'|\mathbf{z}; \tau) \bar{\rho}_{eq}(\mathbf{z})$$

$$J^{AB} = \frac{1}{2\tau} \int d\mathbf{z} \, \nabla \bar{q}(\mathbf{z}) \cdot \left(\int d\mathbf{z}' \Delta \mathbf{z}'(\tau) \Delta \mathbf{z}'(\tau) \bar{\mathcal{P}}(\mathbf{z}'|\mathbf{z}; \tau) \right) \cdot \nabla \bar{q}(\mathbf{z}) \bar{\rho}_{eq}(\mathbf{z}) = \int d\mathbf{z} \, (\nabla \bar{q}(\mathbf{z}) \cdot \mathbf{D} \cdot \nabla \bar{q}(\mathbf{z})) \bar{\rho}_{eq}(\mathbf{z}) \tag{43}$$

The components of the diffusion matrix $\mathbf{D}(\mathbf{z})$ are obtained by retaining only contributions to first order in τ in the conditional average

$$\langle \Delta \mathbf{z}(\tau) \Delta \mathbf{z}(\tau) \rangle_{\mathbf{z}} = \langle (\Delta \mathbf{z}(\tau) - \langle \Delta \mathbf{z}(\tau) \rangle_{\mathbf{z}})^2 \rangle_{\mathbf{z}} + \langle \Delta \mathbf{z}(\tau) \rangle_{\mathbf{z}} \langle \Delta \mathbf{z}(\tau) \rangle_{\mathbf{z}} \approx 2\tau \mathbf{D}(\mathbf{z}) \tag{44}$$

The last term is of order τ^2 because the mean drift is linear in τ

$$\langle \Delta \mathbf{z}(\tau) \rangle_{\mathbf{z}} = \left(-\frac{1}{k_B T} \mathbf{D}(\mathbf{z}) \cdot \nabla W(\mathbf{z}) + \nabla \cdot \mathbf{D}(\mathbf{z}) \right) \tau \tag{45}$$

The expansion in a small displacement also defines the forward committor for the effective dynamics within the subspace $\mathbf{z}$, recovering the backward Smoluchowski equation previously introduced in eq 9.[41,50] In the context of the swarms-of-trajectories, the optimized string can be determined by linking the states A and B such that the projection of the mean drifts perpendicular to the tangent of the curve vanishes.

$$[\langle \Delta \mathbf{z}(\tau) \rangle_{\mathbf{z}}]^{\perp} = 0 \tag{46}$$

The purpose of this analysis is to display the possible relationship of a formulation based on a dynamical propagator with lag time τ to the familiar form of the Smoluchowski diffusion equation. However, it must be emphasized that a reliance on the assumption that the CVs undergo slow diffusional dynamics is not necessary to utilize the dynamical propagator formulation embodied in eqs 33, (41), and (40).

For specific systems, it may be that the dynamics of the system within the subspace of the CVs is truly underdamped and well-characterized by an effective propagator based on a short-time approximation, while for other systems, it may be that the dynamics of the system within the subspace of the CVs is adequately represented as slow variables undergoing overdamped diffusional dynamics. The adequate dynamical regime is controlled by the underlying physics and must be ascertained for each specific system. More generally, the framework of a Markovian effective propagator with a finite lag time τ shall reflect faithfully the underlying dynamics of the system within the subspace of the CVs. In practice, the lag time τ must be chosen to represent the dynamics of the system within the subspace of the CVs as accurately as possible. The string method with swarms-of-trajectories offers a natural framework to capture the underlying dynamics with a finite lag time. A general approach to characterize the kinetics is from the perspective of a spectral analysis of the dynamical propagator $\mathcal{P}(\mathbf{u}'|\mathbf{u}; \tau)$.

Variational Principle and String. A variational principle can be formulated either in terms of the left- or right-eigenvectors of the dynamical propagator $\bar{\mathcal{P}}(\mathbf{z}'|\mathbf{z}; \tau)$ with the lag time τ.[31,32] Assuming that we order the eigenvalues as $\lambda_1 > \lambda_2 > \cdots > \lambda_k$, we consider the trial left-eigenvector $v(i)$ constructed to be orthogonal to the first $(n-1)$th left-eigenvector

$$\frac{\langle v(\tau) v(0) \rangle}{\langle v(0) v(0) \rangle} \leq \lambda_n$$

where $\langle v(\tau) v(0) \rangle$ is an equilibrium time-correlation function. This shows that we can systematically find the nth eigenvector and eigenvalue, by trying to maximize the normalized time-correlation function with respect to a trial function that is orthogonal to the first $(n-1)$th vectors. This defined the variational principle to solve the eigenvalue and eigenvector problem of the dynamical propagation operator. The formulation of the variational principle from left-eigenvector is particularly useful because the trial function is weighted by the equilibrium distribution before the propagation step, yielding directly an equilibrium time-correlation function of the trial left-eigenvector.

Let us consider the trial left-eigenvector $v(\mathbf{z})$, orthogonal to the equilibrium vector and expressed as a linear combination of basis functions

$$v(\mathbf{z}) = \sum_{\alpha} b_{\alpha} \delta z_{\alpha} \tag{47}$$

where $\delta z_{\alpha} = (z_{\alpha} - \langle z_{\alpha} \rangle)$. The orthogonality of this trial left-eigenvector to the first right-eigenvector, $\int d\mathbf{z} \, v(\mathbf{z}) \, \psi^{R}{}_{1}(\mathbf{z}) = 0$, is satisfied by construction. Such a linear combination of CVs is one element of the time-lagged independent component analysis (TICA),[51,52] which was proposed to highlight kinetically relevant information in high-dimensional data. Solving for the normalized left-eigenvector corresponds to the maximization of

$$\mathrm{Max} \frac{\langle v(\tau) v(0) \rangle}{\langle v(0) v(0) \rangle} = \mathrm{Max} \left\{ \frac{\sum_{\alpha\gamma} b_{\alpha} b_{\gamma} C_{\alpha\gamma}(\tau)}{\sum_{\alpha\gamma} b_{\alpha} b_{\gamma} C_{\alpha\gamma}(0)} \right\} \tag{48}$$

where the equilibrium time-correlation functions are

https://doi.org/10.1021/acs.jpca.1c04110
J. Phys. Chem. A 2021, 125, 7558−7571

$$C_{\alpha\gamma}(\tau) = \langle \delta z_\alpha(\tau)\delta z_\gamma(0)\rangle \tag{49}$$

Taking the first derivative with respect to the coefficient b_α yields

$$\mathbf{C}(\tau)\mathbf{b}_k = \lambda_k \mathbf{C}(0)\mathbf{b}_k \tag{50}$$

which is a generalized eigenvalue problem yielding the kth left-eigenvector with the associated eigenvalue λ_k.

In the context of the string method with swarms-of-trajectories,[19] the integral over $\mathbf{z}$ weighted by the equilibrium distribution $\overline{\rho}_{\mathrm{eq}}(\mathbf{z})$ can be converted into a discrete sum over the images weighted by the probability $\overline{p}_{\mathrm{eq}}(m)$. The equilibrium time correlation function may be expressed as

$$\begin{aligned} C_{\alpha\gamma}(\tau) &= \int \mathrm{d}\mathbf{z}\, \langle \delta z_\alpha(\tau)\delta z_\gamma(0)\rangle_{\mathbf{z}} \overline{\rho}_{\mathrm{eq}}(\mathbf{z}) \\ &\approx \sum_{m=1}^{M} \langle \delta z_\alpha(\tau)\delta z_\gamma(0)\rangle_{(m)} \overline{p}_{\mathrm{eq}}(m) \\ &= \sum_{m=1}^{M} \langle \Delta z_\alpha(\tau)\rangle_{(m)} (z_\alpha^m - \langle z_\alpha\rangle)\overline{p}_{\mathrm{eq}}(m) + \langle \delta z_\alpha \delta z_\gamma\rangle \end{aligned} \tag{51}$$

where $\langle \Delta z_\alpha(\tau)\rangle_{(m)}$ represents the mean drift calculated from the swarms-of-trajectories of length τ initiated at the mth image along the string. This suggests that the dominant dynamical behavior that determines the optimal choice of lag time is likely to arise from these terms.

Illustrative Calculations. The relationship between the mean drifts and the correlation function is illustrated with a simple one-dimensional system with two stable states separated by a free energy barrier as shown in Figure 2A. In one dimension, the trial left-eigenvector $v(z)$ defined by eq 47 is simply $\delta z = (z - \langle z\rangle)$. In this simple case

$$\frac{C(\tau)}{C(0)} = 1 + \sum_{m=1}^{M} \langle \Delta z(\tau)\rangle_{(m)} \frac{(z^m - \langle z\rangle)}{\langle \delta z \delta z\rangle} \overline{p}_{\mathrm{eq}}(m) \tag{52}$$

In this development, it is assumed that the equilibrium distribution of the images $\overline{p}_{\mathrm{eq}}(m)$ has been determined. Generally, this can be done efficiently using biased sampling methods.[53,54] Eq 52 clarifies the significance of the mean drifts $\langle \Delta z(\tau)\rangle_{(m)}$ in the string method with swarms-of-trajectories.

The behavior of the swarms is depicted in Figure 2B. With the quantity $[C(\tau)/C(0)] = \lambda_2$ to monitor the slowest eigenvalue, Markovity should mean that the implicit relaxation rate equal to $-\ln \lambda_2/\tau$ is independent of the lag time τ. However, the simple functional form δz is not a perfect trial eigenvector, and, as illustrated in Figure 2C, this condition is only met for a lag time larger than 10 reduced time units, which reflects mainly the relaxation of the two-state system. This is still much shorter than the slowest relaxation time of the double-well system, which is on the order of 1200 reduced time units. Nevertheless, this example shows that it is possible to calculate an equilibrium time-correlation function from controlled starting configurations.

The slow convergence suggests that it is important to go beyond a description that relies mainly on a linear combination of CVs to represent the dynamics along the string. In a more realistic multidimensional system, it is possible that the optimal string could be established for a shorter lag time using a richer basis set that relies on the images along the string pathway. For example, the images along the string can be used to define "one-hot" indicator functions in the subspace of the CVs, leading to

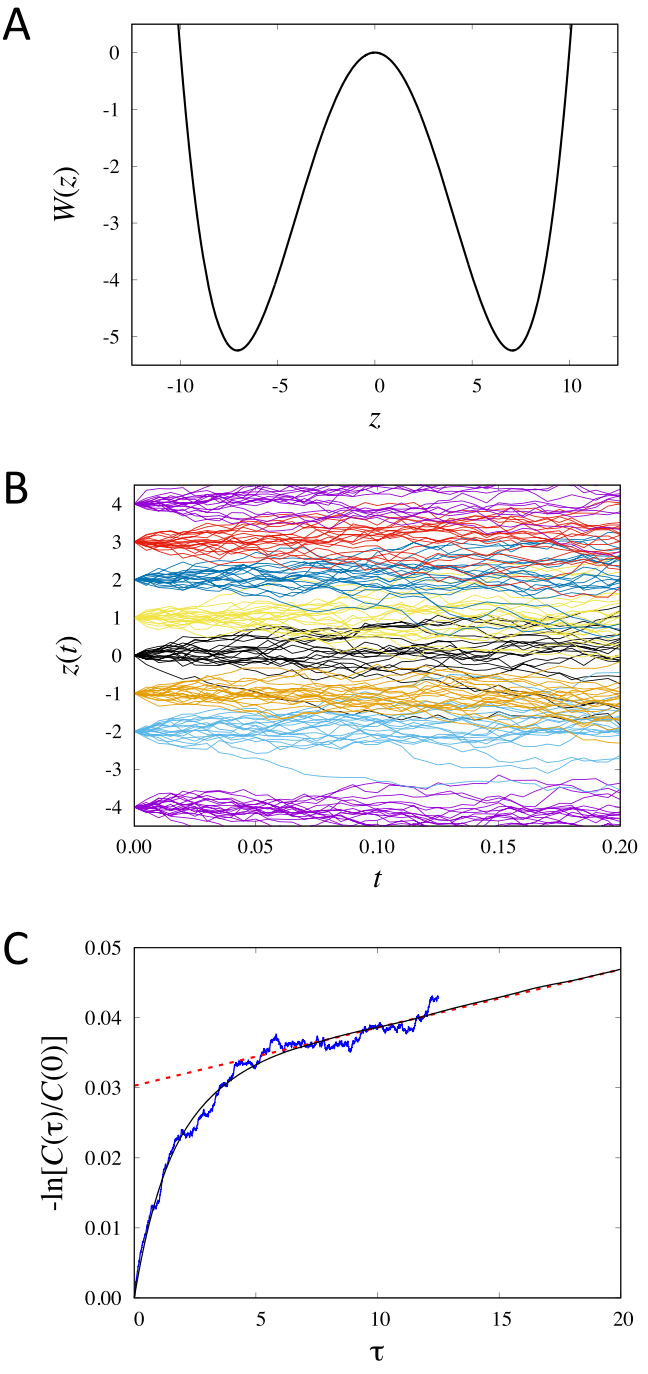


Figure 2. Simple one-dimensional system to illustrate the dynamics and the mean drift from swarms-of-trajectories. (A) Potential energy for the one-dimensional model. The energy surface is given by $W(z) = k_BT(0.002\,0968z^4 - 0.209\,68z^2)$ in units of k_BT. (B) Swarms of 20 short Brownian dynamics trajectories initiated at specific positions along the z axis (diffusion coefficient is equal to 1). (C) Correlation function calculated from eq 52 (blue line) and compared with a simple average from a very long unbiased trajectory of the same system (black line). *Method:* The potential $W(z)$ is expressed in units of k_BT; time is expressed in reduced time units; the diffusion coefficient D is expressed in reduced length units squared per reduced time units. The correlation function was calculated using 201 images (one every 0.1 from from −12 to +12 in reduced units) from a swarms of 200 trajectories of 2500 steps each with a time step of 0.005. The reference correlation function was calculated from a single long unbiased trajectory of 20 million steps with a time step of 0.005.

https://doi.org/10.1021/acs.jpca.1c04110
J. Phys. Chem. A 2021, 125, 7558−7571

the traditional Markov State (MSM) formulation.[55] Likewise, the images along the string can be used to support a Voronoi tessellation in the context of a milestoning algorithm.[56−58] A future application should exploit the position of the individual images as kernels to construct the basis set using nonlinear functions.

This exceedingly simple one-dimensional model serves to illustrate the relationship between the mean drifts and the correlation function reflecting the inherent time scales within the system. However, the directionality of the path is one critical feature cannot be captured in one dimension. This can be illustrated with the simple two-dimensional (2D) system with two stable states separated by a free energy barrier shown in Figure 3. The same model was previously used by Berezhkovskii

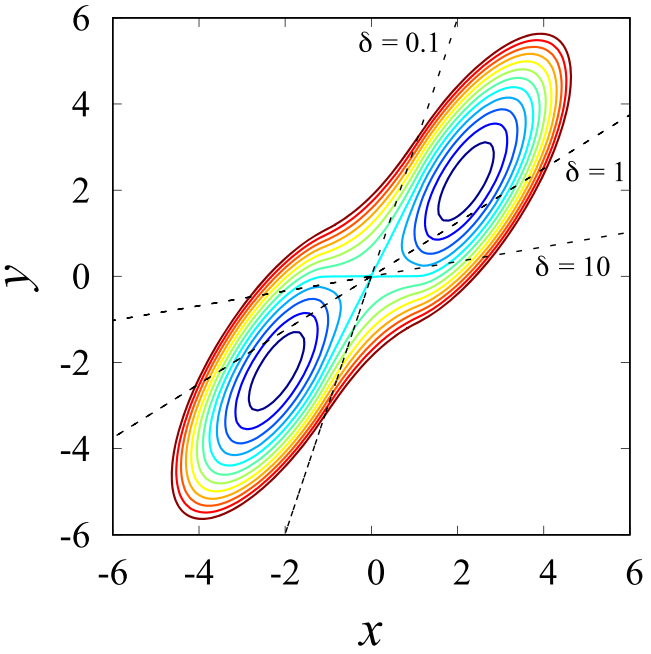


Figure 3. Potential $W(x,y)$ of the 2D system was taken from eqs (14a) and (14b) of Berezhkovskii and A. Szabo.[59] The energy surface is in units of k_BT, and the levels are at −4.0, −3.0, −2.0, −1.0, 0.005, 1.0, 2.0, 3.0, 4.0, 5.0, 6.0, and 7.0. The dash lines represent the direction of the dominant reactive mode for the different conditions: $\theta = 32°$ for $\delta = 1$, $\theta = 72°$ for $\delta = 0.1$, and $\theta = 10°$ for $\delta = 10$.

and Szabo to examine the effect of anisotropic diffusion[59] and also by Tiwary and Berne to illustrate the spectral gap optimization of order parameters (SGOOP) method for predicting reaction coordinates.[60] The magnitude of the anisotropy is characterized in terms of the parameter $\delta = D_y/D_x$, where D_x and D_y are the diffusion coefficients along the x and y axes, respectively. The dash lines in Figure 3 represent the direction of the dominant reactive mode for three different values of δ. The MFEP corresponds to the dominant reactive path ($\theta = 32°$) only when the diffusion is isotropic ($\delta = 1$).[59] To capture the directionality of the path through the saddle point, we model the committor in the space of the CVs by the simple function

$$\overline{q}(\mathbf{z}) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\mathbf{e}\cdot(\mathbf{z} - \mathbf{z}_m)}{\sigma}\right)\right] \tag{53}$$

with $\mathbf{z} \equiv (x,y)$ in two dimensions. The normal to the isocommittor surface is given by the unit vector $\mathbf{e}$; the width of the separatrix region ($q = 0.5$) located at $\mathbf{z}_m$ is given by the parameter σ. By construction, this form starts with $q = 0$ in the reactant region A and reaches $q = 1$ in the product region B. eq 53 can be derived from a quadratic barrier (see Appendix). Since we are only interested in the direction of the path here, we used the knowledge that the top of the barrier is located at $\mathbf{z}_m = (0,0)$ for the sake of simplicity. In two dimensions, the unit vector $\mathbf{e}$ can be characterized via the angle θ with the x-axis.[59,60] Similar functional forms were previously used by Peters and Trout to optimize reaction coordinates via likelihood maximization.[61,62]

On the basis of the model committor based on eq 53, we determine the direction of the optimal path by seeking to minimize the time-correlation function $\langle[q(\tau) - q(0)]^2\rangle$ as a function of the angle θ. Following previously work,[59,60] we compare three cases with $\delta = 1$ ($D_x = 1.0$ and $D_y = 1.0$), $\delta = 10$ ($D_x = 0.1$ and $D_y = 1.0$), and $\delta = 0.1$ ($D_x = 1.0$ and $D_y = 0.1$). Upon examination, the optimal value of the width parameter σ was found to vary only slightly between 0.8 and 1.0; it was kept equal to 1.0 in all cases for the sake of simplicity. The results are shown in Figure 4. In all cases, the minimum of $\langle[q(\tau) - q(0)]^2\rangle$ yields the correct direction of the reactive paths previously identified (Figure 4A).[59,60] It is noteworthy that this formal treatment based on the minimization of the time-correlation function $\langle[q(\tau) - q(0)]^2\rangle$ does not require any assumptions regarding the Markovity of the dynamics along the committor itself. This contrasts with other methods assuming that the dynamics along the putative optimal path is Markovian.[60−62] While the present formulation requires that the effective propagator $\overline{\mathcal{P}}$ within the subspace of the CVs be Markovian (see above), this is not generally true of the dynamics projected long the one-dimensional committor coordinate.[41]

Of particular interest is the dependence of the reactive path direction as a function of the lag time τ. As shown in Figure 4B, clearly all three reactive paths appears to converge to the isotropic case ($\delta = 1$ and $\theta = 32°$) when the lag time is extremely short. The small discrepancy between the three cases at $\tau \approx 0$ is due to the magnitude of the dissipation;it vanishes if D_x and D_y are increased while keeping δ constant. In the isotropic case, the reactive path captured by the short-time propagation corresponds to the MFEP. This path is equivalent to the mean-force string method as prescribed by eqs 5, (6), (8), (31), and (32).[6,30] The short-time propagator is effectively "blind" to the true dissipative nature of the microscopic dynamics of the CVs (Figure 4B). The optimal directionality of the path only emerges as the lag time is increased to 0.5 reduced time units. As prescribed by the minimization of the time-correlation function $\langle[q(\tau) - q(0)]^2\rangle$, the correct path follows the true slow reactive mode, which is dominated by the smallest diffusion coefficient; if $D_x < D_y$ the path follows the x-axis (θ is close to 0°), and if $D_x > D_y$ the path follows the y-axis (θ is close to 90°).

This analysis suggests a potentially practical "staging" strategy for characterizing the direction of the reactive paths in a multidimensional systems with a single dominant barrier. Operationally, the path is prescribed by minimizing the equilibrium time-correlation function $\langle[q(\tau) - q(0)]^2\rangle$. In effect, minimizing this quantity is associated with discovering the region in the subspace of the CVs where the fluctuations over q over a lag time τ are as infrequent as possible. Motivated by this perspective, we consider the harmonic restraining potential $u_m(\mathbf{z}) = 1/2k(\mathbf{z} - \mathbf{z}_m)^2$ and rewrite the time-correlation function for the reactive flux as

https://doi.org/10.1021/acs.jpca.1c04110
J. Phys. Chem. A 2021, 125, 7558−7571

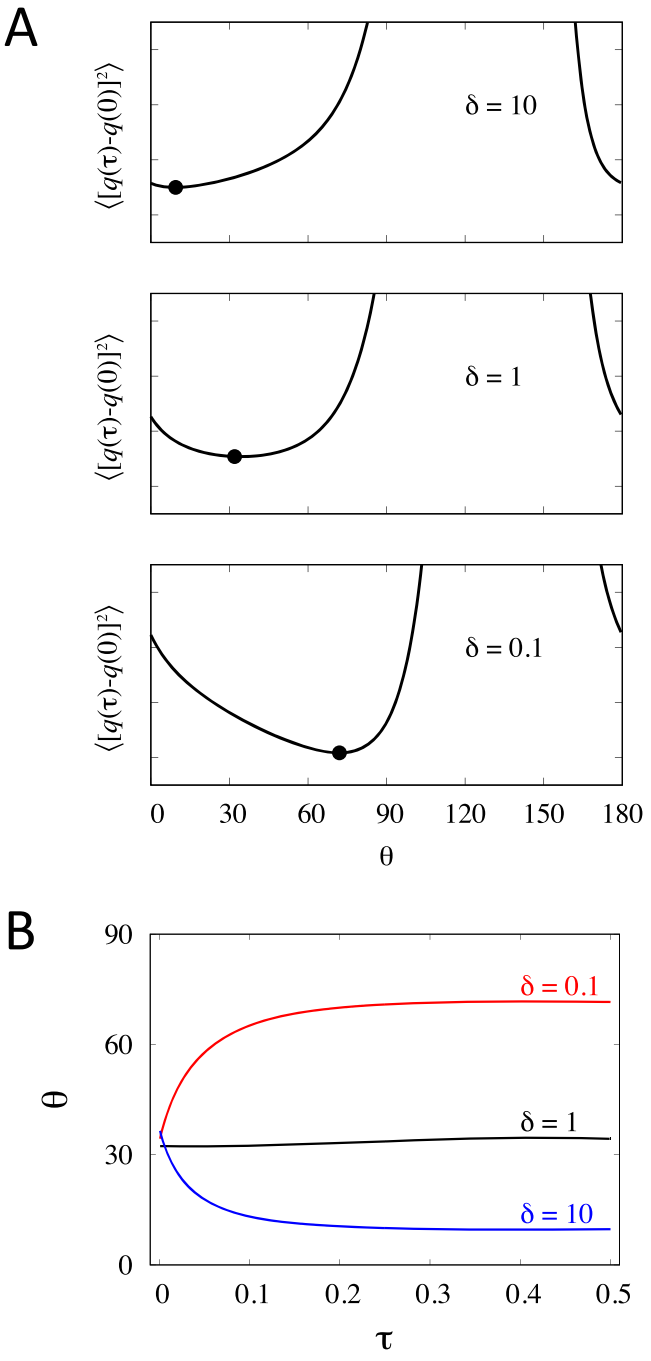


Figure 4. Results from using the reactive flux correlation function from eq 41 with the committor based on eq 53. (A) Dependence of the time-correlation reactive flux function $\langle [q(\tau) - q(0)]^2 \rangle$ on the angle θ for the direction of the path. The maximum of the curve occurring around $\theta = 130°$ (orthogonal to the axis linking the two wells) is not shown for clarity. The time-correlation function is shown for a fixed lag time $\tau = 0.5$. The black circle indicates the position of the minimum in the curves: $\theta = 32°$ for $\delta = 1$, $\theta = 72°$ for $\delta = 0.1$, and $\theta = 10°$ for $\delta = 10$. (B) Dependence of the time-correlation reactive flux function $\langle [q(\tau) - q(0)]^2 \rangle$ on the lag time τ (the optimal angle θ is used for each case). The width parameter σ is equal to 1.0 in all cases. *Method:* The time-correlation functions were calculated from a Langevin dynamics trajectory of 10 million steps generated via eq 1, with $\gamma_x = D_x/k_BT$ and $\gamma_y = D_y/k_BT$, for D_x and D_y taking values of 1.0 and 0.1 corresponding to an anisotropy $\delta = D_y/D_x$ of 0.1, 1.0, and 10.0. The mass was chosen to yield relaxation times γ/m of 0.08 and 0.008 reduced time unit, for diffusion coefficients of 1.0 and 0.1, respectively.

$$\begin{aligned}\langle [q(\tau) - q(0)]^2 \rangle &= \langle e^{-u_m(\mathbf{z})/k_BT} e^{u_m(\mathbf{z})/k_BT} [q(\tau) - q(0)]^2 \rangle \\ &= e^{-G_m/k_BT} \langle e^{u_m(\mathbf{z})/k_BT} [q(\tau) - q(0)]^2 \rangle_{(u_m)}\end{aligned} \tag{54}$$

where $\langle \cdots \rangle_{(u_m)}$ indicates a conditional average initially weighted by the restraining potential u_m, and $G_m = -k_BT \ln\langle \exp[-u_m(\mathbf{z})/k_BT] \rangle$ is the free energy for introducing the restraining potential u_m.[63] For example, similar restraining potentials are applied on the M images in the string method with swarms-of-trajectories.[19] In a staging strategy to minimize the reactive flux time-correlation function, one should first seek to locate the position $\mathbf{z}_m$ that yields the largest G_m and then determine the optimal direction of the reactive path from the conditional time-correlation function $\langle e^{u_m(\mathbf{z}(0))/k_BT} [q(\tau) - q(0)]^2 \rangle_{(u_m)}$ for a functional model of q such as that given by eq 53. Even though the biasing factor $\exp[u_m(\mathbf{z})/k_BT]$ in eq 54 increases the weight of the configuration at large distances from the top of the barrier, the conditional average is expected to converge locally because the quantity $[q(\tau) - q(0)]^2$, which goes as $\nabla q \cdot \mathbf{D} \cdot \nabla q$ in the diffusive limit according to eq 43, is nonzero only in the region where the committor varies abruptly. Correspondingly, the contribution from regions away from the barrier top where the committor is nearly constant are expected to be negligible. A framework akin to Chandler's reactive flux formalism is essentially recovered in the limit where the model committor varies abruptly as a Heaviside step-function.[64]

Simple tests with the 2D potential surface of Figure 3 indicate that the staging strategy can work. Using the restraining potential $u_m = 3k_BT(x^2 + y^2)$ to prepare initial conditions near the barrier top, followed by a swarms of 10 000 unbiased short trajectories of 0.5 time units (500 steps) show that the staging method can yield correct results for the direction of the reactive paths, with angles of $\theta = 32°$ for $\delta = 1$, $\theta = 68°$ for $\delta = 0.1$, and $\theta = 18°$ for $\delta = 10$. While these results on a simple model are encouraging, more work is needed to further develop the staging strategy into a fully practical and reliable method.

CONCLUSION

We have formulated the problem of the kinetics of a dynamical system comprising two metastable states A and B in terms of a finite-time propagator in phase space (position and velocity) adapted to the underdamped Langevin equation. The present development expands on a previous formulation based on propagators for position-based Markov models.[31−34] Starting from the full phase space representation with positions and velocities, the central ansatz assumes that the committor depends only on the variables of the selected collective variables. The dimensionality reduction to the subspace of CVs yields familiar expressions for the propagator, committor, and steady-state flux, as embodied in eqs 33, (41), and (40), respectively. These are the central results of the development. Importantly, eq 40 is a quadratic expression for the steady-state flux J^{AB} between the metastable states A and B that can serve as a robust variational principle to determine the optimal committor. These expressions can be exploited in the context of the string method, in which the dominant pathway between A and B is represented by a curve in the space of the CVs, to calculate dynamical correlation.

The analysis based on the effective dynamical propagator in the reduced space of the CVs, $\bar{\mathcal{P}}(\mathbf{z}'|\mathbf{z}; \tau)$, clarifies the theoretical foundation of the string method with swarms-of-trajectories to

https://doi.org/10.1021/acs.jpca.1c04110
J. Phys. Chem. A 2021, 125, 7558−7571

determine optimal transition pathways.[19] Most importantly, the analysis revealed the fundamental significance of the lag time to calculate the mean drift during short unbiased trajectories.[19,21] The lag time τ must be chosen so to satisfy the conditions of Markovity the effective propagator $\bar{\mathcal{P}}(\mathbf{z}'|\mathbf{z};\,\tau)$ within the subspace of the CVs. In practice, τ should be as short as possible but sufficiently large to ensure that $\bar{\mathcal{P}}(\mathbf{z}'|\mathbf{z};\,\tau)$ is Markovian. The string method with swarms-of-trajectories offers a natural framework to capture the underlying dynamics within the subspace of the CVs. If this dynamics experiences very little dissipative coupling and is near-inertial, then it is possible that Markovity is attained with a very short lag time (short swarms-of-trajectories). In this case, the mean drifts from the swarms-of-trajectories is essentially equivalent to the original mean force string method.[6,30] However, if the effective dynamics within the subspace of the CVs experiences complex dissipative effects, then a longer lag time is needed to guaranty the Markovity of the effective propagator—a necessary condition to determine the genuine forward committor $\bar{\mathcal{P}}(\mathbf{z})$. Without Markovity, the solution $\bar{q}(\mathbf{z})$ of eq 30 based on the effective short-time propagator $\bar{\mathcal{P}}(\mathbf{z}'|\mathbf{z};\,\Delta t)$ is not the forward committor in the subspace of the CVs. In both cases, the mean drift from the swarms-of-trajectories correctly captures the effective dynamical behavior of the system within the subspace of the CVs, supporting the construction of a meaningful transition pathway.[19,21] The formulation of the propagator with a finite lag time makes it possible to determine the necessary conditions for Markovity and the progression of the committor along the path. Furthermore, it circumvents the need to rely on the rank ordering of the different images along the string to define a one-dimensional effective reaction coordinate.[65,66] Interestingly, alternate methods built upon an optimized pathway, such as Milestoning[56,67] or MSM,[55] also offer the possibility to directly examine the natural unbiased dynamics within the subspace of the CVs. Furthermore, the effective propagator $\bar{\mathcal{P}}(\mathbf{z}'|\mathbf{z};\,\Delta t)$ with finite lag time τ is amenable to an eigenvalue-eigenvector spectral analysis as elaborated previously in the context of position-based Markov models.[31−34] The time-correlation functions calculated by swarms-of-trajectories along the string constitutes a natural extension of these developments. The present analysis is also related to a recent dynamical Galerkin approximation (DGA) formulated to predict the long-time scale behavior from short-trajectory.[68,69]

The present formulation strengthens the theoretical framework to determine the optimal pathway between the states A and B, that is, the proverbial "reaction coordinate", and to characterize the long-time kinetics of the system. Identifying the most relevant subspace of CVs and then determining the most representative reaction coordinates in this subspace remains the greatest challenge. In this endeavor, other related ideas may be brought to bear, including a dynamical self-consistency,[70] memory reduction,[71,72] multidimensional spectral gap optimization of order parameters (SGOOP),[73,74] and maximally predictive one-dimensional projection.[33] Of particular interest are the ideas of Krivov, who proposed a nonparametric variational optimization of reaction coordinates.[75] Future work will further explore these ideas.

APPENDIX: COMMITTOR AND QUADRATIC BARRIER

Following Berezhkovskii and Szabo,[41] we assume that the top of the barrier at $\mathbf{z} = 0$ is a multidimensional quadratic form, $\beta W = \mathbf{z}^t\mathbf{V}\mathbf{z}/2$ and that the committor can be written as $q(\mathbf{z}) \approx q(\mathbf{e}\cdot\mathbf{z})$ in the neighborhood of the saddle point (we drop the overline on q to simplify the notation). At this point, $\mathbf{e}$ is an unknown unit vector that has yet to be determined. The gradient Laplacians of q are $\nabla q(\mathbf{z}) = \mathbf{e}q'(\mathbf{e}\cdot\mathbf{z})$ and $\nabla^2 q(\mathbf{z}) = q''(\mathbf{e}\cdot\mathbf{z})$, respectively. Starting from the backward Kolmogorov eq 9,[41,50] in the case of a constant (independent of position) anisotropic diffusion matrix $\mathbf{D}$, we have

$$
\begin{aligned}
0 &= \nabla\cdot(e^{-\beta U}\mathbf{D}\mathbf{e}q') \\
0 &= -e^{-\beta U}(\nabla\beta U)\cdot(\mathbf{D}\mathbf{e}q') + e^{-\beta U}\nabla\cdot(\mathbf{D}\mathbf{e}q') \\
0 &= -(\nabla\beta U)\cdot(\mathbf{D}\mathbf{e})q' + \mathbf{e}\mathbf{D}\cdot(\nabla q') \\
0 &= -(\mathbf{z}^t\mathbf{V}\mathbf{D}\mathbf{e})q'(s) + (\mathbf{e}\mathbf{D}\mathbf{e})q''(s)
\end{aligned}
$$

where $s = \mathbf{z}\cdot\mathbf{e}$. To close this equation with respect to s, we require that the vector $\mathbf{e}$ be chosen to satisfy the condition

$$-\lambda\mathbf{e} = \mathbf{V}\mathbf{D}\mathbf{e}$$

($\lambda > 0$). Using $\mathbf{z}^t\mathbf{e} = \mathbf{e}\cdot\mathbf{z} = s$, we can write

$$0 = \lambda s q'(s) + (\mathbf{e}\mathbf{D}\mathbf{e})q''(s)$$

Letting $\mathbf{e}\cdot\mathbf{D}\cdot\mathbf{e} = D$, we write

$$
\begin{aligned}
Dq''(s) &= -\lambda s q'(s) \\
q'(s) &= e^{-\lambda s^2/2D}
\end{aligned}
$$

leading to

$$
\begin{aligned}
q(\mathbf{z}) &= \frac{\int_a^{\mathbf{z}\cdot} \mathrm{d}s\; e^{-\lambda s^2/2D}}{\int_a^b \mathrm{d}s\; e^{-\lambda s^2/2D}} \\
&= \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\mathbf{z}\cdot\mathbf{e}}{\sigma}\right)\right]
\end{aligned}
$$

with $\sigma^2 = 2D/\lambda$. This functional form is similar to eq 3.8 from Berezhkovskii and Szabo that was obtained following a different route.[41]

AUTHOR INFORMATION

Corresponding Author

Benoît Roux − *Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, Illinois 60637, United States; Department of Chemistry, The University of Chicago, Chicago, Illinois 60637, United States;* orcid.org/0000-0002-5254-2712; Email: roux@uchicago.edu

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.1c04110

Notes

The author declares no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (NSF), through Grant No. MCB-1517221. The author wishes to acknowledge useful discussions with A. Szabo, G. Ciccotti, S. Krivov, J. Weare, A. Dinner, T. Lelièvre, and C. Chipot.

REFERENCES

(1) Berezhkovskii, A. M.; Szabo, A. Committors, first-passage times, fluxes, Markov states, milestones, and all that. *J. Chem. Phys.* **2019**, *150*, 054106.
(2) E, W.; Vanden-Eijnden, E. Transition-path theory and path-finding algorithms for the study of rare events. *Annu. Rev. Phys. Chem.* **2010**, *61*, 391−420.
(3) E, W.; Ren, W.; Vanden-Eijnden, E. String method for the study of rare events. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2002**, *66*, 052301.
(4) Weinan, E.; Ren, W.; Vanden-Eijnden, E. Transition pathways in complex systems: Reaction coordinates, isocommittor surfaces, and transition tubes. *Chem. Phys. Lett.* **2005**, *413*, 242−247.
(5) E, W.; Ren, W.; Vanden-Eijnden, E. Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *J. Chem. Phys.* **2007**, *126*, 164103.
(6) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. String method in collective variables: minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.* **2006**, *125*, 24106.
(7) Pratt, L. A statistical-method for identifying transition-states in high dimensional problems. *J. Chem. Phys.* **1986**, *85*, 5045−5048.
(8) Elber, R.; Karplus, M. A Method for Determining Reaction Paths in Large Molecules: Application to Myoglobin. *Chem. Phys. Lett.* **1987**, *139*, 375−380.
(9) Miller, T.; Vanden-Eijnden, E.; Chandler, D. Solvent coarse-graining and the string method applied to the hydrophobic collapse of a hydrated chain. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 14559−14564.
(10) Zhu, F.; Hummer, G. Pore opening and closing of a pentameric ligand-gated ion channel. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 19814−19819.
(11) Ovchinnikov, V.; Trout, B. L.; Karplus, M. Mechanical Coupling in Myosin V: A Simulation Study. *J. Mol. Biol.* **2010**, *395*, 815−833.
(12) Ovchinnikov, V.; Karplus, M.; Vanden-Eijnden, E. Free energy of conformational transition paths in biomolecules: The string method and its application to myosin VI. *J. Chem. Phys.* **2011**, *134*, 085103.
(13) Kirmizialtin, S.; Nguyen, V.; Johnson, K. A.; Elber, R. *Structure* **2012**, *20*, 618−627.
(14) Stober, S. T.; Abrams, C. F. Energetics and Mechanism of the Normal-to-Amyloidogenic Isomerization of β2-Microglobulin: On-the-Fly String Method Calculations. *J. Phys. Chem. B* **2012**, *116*, 9371−9375.
(15) Vashisth, H.; Maragliano, L.; Abrams, C. F. "DFG-flip" in the insulin receptor kinase is facilitated by a helical intermediate state of the activation loop. *Biophys. J.* **2012**, *102*, 1979−1987.
(16) Matsunaga, Y.; Fujisaki, H.; Terada, T.; Furuta, T.; Moritsugu, K.; Kidera, A. Minimum Free Energy Path of Ligand-Induced Transition in Adenylate Kinase. *PLoS Comput. Biol.* **2012**, *8*, e1002555.
(17) Vashisth, H.; Abrams, C. F. All-atom structural models of insulin binding to the insulin receptor in the presence of a tandem hormone-binding element. *Proteins: Struct., Funct., Genet.* **2013**, *81*, 1017−1030.
(18) E, W.; Ren, W.; Vanden-Eijnden, E. Finite Temperature String Method for the Study of Rare Events. *J. Phys. Chem. B* **2005**, *109*, 6688−6693.
(19) Pan, A.; Sezer, D.; Roux, B. Finding transition pathways using the string method with swarms of trajectories. *J. Phys. Chem. B* **2008**, *112*, 3432−3440.
(20) Tempkin, J. O. B.; Qi, B.; Saunders, M. G.; Roux, B.; Dinner, A. R.; Weare, J. Using multiscale preconditioning to accelerate the convergence of iterative molecular calculations. *J. Chem. Phys.* **2014**, *140*, 184114.
(21) Johnson, M. E.; Hummer, G. Characterization of a dynamic string method for the construction of transition pathways in molecular reactions. *J. Phys. Chem. B* **2012**, *116*, 8573−8583.
(22) Jiang, W.; Phillips, J. C.; Huang, L.; Fajer, M.; Meng, Y.; Gumbart, J. C.; Luo, Y.; Schulten, K.; Roux, B. Generalized Scalable Multiple Copy Algorithms for Molecular Dynamics Simulations in NAMD. *Comput. Phys. Commun.* **2014**, *185*, 908−916.
(23) Melo, M. C. R.; Bernardi, R. C.; Rudack, T.; Scheurer, M.; Riplinger, C.; Phillips, J. C.; Maia, J. D. C.; Rocha, G. B.; Ribeiro, J. V.; Stone, J. E.; Neese, F.; Schulten, K.; Luthey-Schulten, Z. NAMD goes quantum: an integrative suite for hybrid simulations. *Nat. Methods* **2018**, *15*, 351−354.
(24) Gan, W.; Yang, S.; Roux, B. Atomistic view of the conformational activation of Src kinase using the string method with swarms-of-trajectories. *Biophys. J.* **2009**, *97*, L8−L10.
(25) Fajer, M.; Meng, Y.; Roux, B. The Activation of c-Src Tyrosine Kinase: Conformational Transition Pathway and Free Energy Landscape. *J. Phys. Chem. B* **2017**, *121*, 3352−3363.
(26) Jo, S.; Rui, H.; Lim, J. B.; Klauda, J. B.; Im, W. Cholesterol Flip-Flop: Insights from Free Energy Simulation Studies. *J. Phys. Chem. B* **2010**, *114*, 13342−13348.
(27) Lacroix, J. J.; Pless, S. A.; Maragliano, L.; Campos, F. V.; Galpin, J. D.; Ahern, C. A.; Roux, B.; Bezanilla, F. Intermediate state trapping of a voltage sensor. *J. Gen. Physiol.* **2012**, *140*, 635−652.
(28) Das, A.; Rui, H.; Nakamoto, R.; Roux, B. Conformational Transitions and Alternating-Access Mechanism in the Sarcoplasmic Reticulum Calcium Pump. *J. Mol. Biol.* **2017**, *429*, 647−666.
(29) Singharoy, A.; Chipot, C.; Moradi, M.; Schulten, K. Chemomechanical Coupling in Hexameric Protein-Protein Interfaces Harnesses Energy within V-Type ATPases. *J. Am. Chem. Soc.* **2017**, *139*, 293−310.
(30) Maragliano, L.; Roux, B.; Vanden-Eijnden, E. Comparison between Mean Forces and Swarms-of-Trajectories String Methods. *J. Chem. Theory Comput.* **2014**, *10*, 524−533.
(31) Nueske, F.; Keller, B. G.; Perez-Hernandez, G.; Mey, A. S. J. S.; Noe, F. Variational Approach to Molecular Kinetics. *J. Chem. Theory Comput.* **2014**, *10*, 1739−1752.
(32) Wu, H.; Nueske, F.; Paul, F.; Klus, S.; Koltai, P.; Noe, F. Variational Koopman models: Slow collective variables and molecular kinetics from short off-equilibrium simulations. *J. Chem. Phys.* **2017**, *146*, 154104.
(33) McGibbon, R. T.; Husic, B. E.; Pande, V. S. Identification of simple reaction coordinates from complex dynamics. *J. Chem. Phys.* **2017**, *146*, 044109.
(34) Krivov, S. V. Protein Folding Free Energy Landscape along the Committor - the Optimal Folding Coordinate. *J. Chem. Theory Comput.* **2018**, *14*, 3418−3427.
(35) Berezhkovskii, A. M.; Szabo, A. Effect of ligand diffusion on occupancy fluctuations of cell-surface receptors. *J. Chem. Phys.* **2013**, *139*, 121910.
(36) Zwanzig, R. Memory effects in irreversible thermodynamics. *Phys. Rev.* **1961**, *124*, 983.
(37) Mori, H. Transport, Collective Motion, and Brownian Motion. *Prog. Theor. Phys.* **1965**, *33*, 423−455.
(38) Kubo, R. The fluctuation-dissipation theorem. *Rep. Prog. Phys.* **1966**, *29*, 255−284.
(39) Lelievre, T.; Rousset, M.; Stoltz, G. *Free Energy Computations: A Mathematical Perspective;* Imperial College Press: London, UK, 2010.
(40) Berezhkovskii, A.; Hummer, G.; Szabo, A. Reactive flux and folding pathways in network models of coarse-grained protein dynamics. *J. Chem. Phys.* **2009**, *130*, 205102.
(41) Berezhkovskii, A. M.; Szabo, A. Diffusion along the splitting/commitment probability reaction coordinate. *J. Phys. Chem. B* **2013**, *117*, 13115−13119.
(42) Krivov, S. V. On Reaction Coordinate Optimality. *J. Chem. Theory Comput.* **2013**, *9*, 135−146.
(43) Vanden-Eijnden, E. Transition Path Theory. In *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology;* Ferrario, M., Ciccotti, G., Binder, K., Eds.; Springer, 2006; Vol. *2*; p 439.
(44) Hummer, G.; Szabo, A. Optimal Dimensionality Reduction of Multistate Kinetic and Markov-State Models. *J. Phys. Chem. B* **2015**, *119*, 9029−9037.
(45) Orioli, S.; Faccioli, P. Dimensional reduction of Markov state models from renormalization group theory. *J. Chem. Phys.* **2016**, *145*, 124120.
(46) Lamm, G.; Szabo, A. Langevin modes of macromolecules. *J. Chem. Phys.* **1986**, *85*, 7334−7348.

https://doi.org/10.1021/acs.jpca.1c04110
J. Phys. Chem. A 2021, 125, 7558−7571

(47) Herau, F.; Hitrik, M.; Sjoestrand, J. Tunnel effect and symmetries for Kramers-Fokker-Planck type operators. *J. Inst. Math. Jussieu* **2011**, *10*, 567−634.
(48) Banushkina, P. V.; Krivov, S. V. Nonparametric variational optimization of reaction coordinates. *J. Chem. Phys.* **2015**, *143*, 184108.
(49) Krivov, S. Blind Analysis of Molecular Dynamics. *J. Chem. Theory Comput.* **2021**, *17*, 2725−2736.
(50) Schulten, K.; Kosztin, I. *Lectures in Theoretical Biophysics*; https://www.ks.uiuc.edu/Services/Class/PHYS498/LectureNotes.html, 2000; pp 1−206.
(51) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **2013**, *139*, 015102.
(52) Schwantes, C. R.; Pande, V. S. Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9. *J. Chem. Theory Comput.* **2013**, *9*, 2000−2009.
(53) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(54) Souaille, M.; Roux, B. Extension to the Weighted Histogram Analysis Method: Combining Umbrella Sampling with Free Energy Calculations. *Comput. Phys. Commun.* **2001**, *135*, 40−57.
(55) Pan, A. C.; Roux, B. Building Markov state models along pathways to determine free energies and rates of transitions. *J. Chem. Phys.* **2008**, *129*, 064107.
(56) Faradjian, A.; Elber, R. Computing time scales from reaction coordinates by milestoning. *J. Chem. Phys.* **2004**, *120*, 10880−10889.
(57) Vanden-Eijnden, E.; Venturoli, M. Markovian milestoning with Voronoi tessellations. *J. Chem. Phys.* **2009**, *130*, 194101.
(58) Maragliano, L.; Vanden-Eijnden, E.; Roux, B. Free energy and kinetics of conformational transitions from Voronoi tessellated milestoning with restraining potentials. *J. Chem. Theory Comput.* **2009**, *5*, 2589−2594.
(59) Berezhkovskii, A.; Szabo, A. One-dimensional reaction coordinates for diffusive activated rate processes in many dimensions. *J. Chem. Phys.* **2005**, *122*, 14503.
(60) Tiwary, P.; Berne, B. J. J Chem PhysPredicting reaction coordinates in energy landscapes with diffusion anisotropy. *J. Chem. Phys.* **2017**, *147*, 152701.
(61) Peters, B.; Trout, B. L. Obtaining reaction coordinates by likelihood maximization. *J. Chem. Phys.* **2006**, *125*, 054108.
(62) Peters, B. Reaction Coordinates and Mechanistic Hypothesis Tests. *Annu. Rev. Phys. Chem.* **2016**, *67*, 669−690.
(63) Suh, D.; Jo, S.; Jiang, W.; Chipot, C.; Roux, B. String method for Protein-Protein Binding Free Energy Calculation. *J. Chem. Theory Comput* **2019**, *15*, 5829−5844, DOI: 10.1021/acs.jctc.9b00499.
(64) Chandler, D. Statistical mechanics of isomerization dynamics in liquids and the transition state approximation. *J. Chem. Phys.* **1978**, *68*, 2959−2970.
(65) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in free energy space. *J. Chem. Phys.* **2007**, *126*, 054103.
(66) Diaz Leines, G.; Ensing, B. Path Finding on High-Dimensional Free Energy Landscapes. *Phys. Rev. Lett.* **2012**, *109*, 020601.
(67) Vanden-Eijnden, E.; Venturoli, M.; Ciccotti, G.; Elber, R. On the assumptions underlying milestoning. *J. Chem. Phys.* **2008**, *129*, 174102.
(68) Thiede, E. H.; Giannakis, D.; Dinner, A. R.; Weare, J. Galerkin approximation of dynamical quantities using trajectory data. *J. Chem. Phys.* **2019**, *150*, 244111.
(69) Lorpaiboon, C.; Thiede, E. H.; Webber, R. J.; Weare, J.; Dinner, A. R. Integrated Variational Approach to Conformational Dynamics: A Robust Strategy for Identifying Eigenfunctions of Dynamical Operators. *J. Phys. Chem. B* **2020**, *124*, 9354−9364.
(70) 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.* **2013**, *138*, 054106.
(71) Guttenberg, N.; Dama, J. F.; Saunders, M. G.; Voth, G. A.; Weare, J.; Dinner, A. R. Minimizing memory as an objective for coarse-graining. *J. Chem. Phys.* **2013**, *138*, 094111.
(72) Guttenberg, N.; Dama, J. F.; Saunders, M. G.; Voth, G. A.; Weare, J.; Dinner, A. R. Identification of simple reaction coordinates from complex dynamics. *J. Chem. Phys.* **2013**, *138*, 094111.
(73) Tiwary, P.; Berne, B. J. Spectral gap optimization of order parameters for sampling complex molecular systems. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 2839−2844.
(74) Smith, Z.; Pramanik, D.; Tsai, S. T.; Tiwary, P. Multi-dimensional spectral gap optimization of order parameters (SGOOP) through conditional probability factorization. *J. Chem. Phys.* **2018**, *149*, 234105.
(75) Banushkina, P. V.; Krivov, S. V. Nonparametric variational optimization of reaction coordinates. *J. Chem. Phys.* **2015**, *143*, 184108.

https://doi.org/10.1021/acs.jpca.1c04110
J. Phys. Chem. A 2021, 125, 7558−7571