Markov state models based on milestoning

Christof Schütte,[1,a)] Frank Noé,[1,b)] Jianfeng Lu,[2,c)] Marco Sarich,[1,d)] and Eric Vanden-Eijnden[2,e)]
[1]*Institute of Mathematics, Freie Universitaet Berlin, D-14195 Berlin, Germany*
[2]*Courant Institute of Mathematical Sciences, New York University, New York, New York 10012, USA*

(Received 16 January 2011; accepted 21 April 2011; published online 24 May 2011)

Markov state models (MSMs) have become the tool of choice to analyze large amounts of molecular dynamics data by approximating them as a Markov jump process between suitably predefined states. Here we investigate "Core Set MSMs," a new type of MSMs that build on metastable core sets acting as milestones for tracing the rare event kinetics. We present a thorough analysis of Core Set MSMs based on the existing milestoning framework, Bayesian estimation methods and Transition Path Theory (TPT). We show that Core Set MSMs can be used to extract phenomenological rate constants between the metastable sets of the system and to approximate the evolution of certain key observables. The performance of Core Set MSMs in comparison to standard MSMs is analyzed and illustrated on a toy example and in the context of the torsion angle dynamics of alanine dipeptide.
 [doi:10.1063/1.3590108]

I. INTRODUCTION

Conformational transitions are essential to the function of proteins, nucleic acids, and other macromolecules. These transitions span large ranges of length scales, time scales, and complexity, and include processes as important as folding,[1,2] complex conformational rearrangements between native protein substates,[3,4] and ligand binding.[5] Molecular dynamics (MD) simulations are becoming increasingly accepted as a tool to investigate both the structural and the dynamical features of these transitions at a level of detail that is beyond that accessible in laboratory experiments.[6–8] Modern computing technologies, such as massively parallel simulation,[9] special-purpose high-performance computers,[10] and high-performance GPUs (Ref. 11) permit to generate MD data in amounts too large to be grasped by traditional "look and see" analyses. This calls for robust and automated methods to extract the essential structural and dynamical properties from these data in a manner that is little or not depending on human subjectivity.

To this end, a decade of work has led to the development of analysis techniques which rely on the partitioning of the conformation space into discrete substates and reduce the molecular kinetics to transitions between these states.[12–25] A particular successful class of methods of this type are Markov state models (MSMs), in which the transitions between the states in the partition are assumed to be memoryless jumps. Their kinetics is then described fully in terms of the transition probabilities that the system will have jumped from one state to another after a prescribed lag time τ.[21,23,24,26–32] These probabilities are estimated from the MD simulation data.

As yet, most MSMs have been based on discretizations that fully partition the molecular state space. Thorough analysis[13,33–35] has shown that these full-partition MSMs can approximate the original dynamics arbitrarily well, and their accuracy can be improved in two ways: (i) by increasing the lag time τ,[34] or (ii) by increasing the number of states in the partition.[13,33–35] Both procedures, however, have caveats: (i) reduces the time resolution of the model, whereas (ii) can be difficult to achieve in practice because the number of states is practically limited by available trajectory statistics.

For systems with strongly dominant metastable states,[31,33] a different approach to model the essential kinetic properties of the system may be adequate. In these systems the energy landscape is such that there are regions (the metastable states) in the vicinity of which a typical MD trajectory will remain for a long time before making a transition towards another such region. In these situations, Buchete and Hummer[19] have proposed to avoid a full partition and instead define a few *cores*, one for each dominant metastable state. Instead of finely subdividing the intervening transition regions, one only considers transitions of the MD trajectories between these cores and constructs a statistical model of the molecular kinetics from this information.

The aim of this paper is to give a solid theoretical foundation to such MSMs based on cores by treating these sets as *milestones* in the sense of Elber.[36–39] Specifically, we show how the framework of Markovian milestoning[40] can be combined with maximum estimation techniques and Bayesian sampling methods to construct a new type of core MSMs. This viewpoint helps the estimation of the statistical error of these new core MSMs due to finite sampling. It also permits to interpret the various quantities in these core MSMs using Transition Path Theory (TPT),[41–45] which indicates how the core MSMs can be used not only to estimate the rate of transitions between the cores but also to approximate the evolution of certain key observables in the system. One surprise of our analysis is that doing both requires the introduction of two

[a)]Electronic mail: Christof.Schuette@fu-berlin.de.
[b)]Electronic mail: frank.noe@fu-berlin.de.
[c)]Electronic mail: jianfeng@cims.nyu.edu.
[d)]Electronic mail: sarich@mi.fu-berlin.de.
[e)]Electronic mail: eve2@cims.nyu.edu.

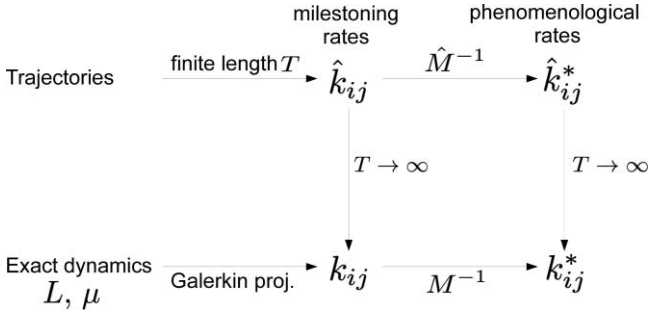

FIG. 1. Schematic illustration of the main quantities employed in this study. The exact dynamics described by the continuous generator L with stationary distribution μ gives rise to exact transition rates between milestones, k_{ij} which can be reweighted to phenomenological rates, k^*_{ij}. These quantities can also be approximated by proper counting based on trajectory data.

sets of rates: the *milestoning rates* $\hat{k}_{ij}$, which give the average number of transitions between core sets per unit time, and the *phenomenological rates* $\hat{k}^*_{ij}$, which are related to the measurable relaxation timescales of the observables. The relation between these rates is illustrated in Fig. 1 which also gives a roadmap to the main concepts discussed in this paper.

The remainder of this paper is organized as follows. In Sec. II, we first show how to estimate the milestoning rates $\hat{k}_{ij}$ from a given trajectory of finite length T. Furthermore, we outline how to convert these milestoning rates to the phenomenological rates $\hat{k}^*_{ij}$. In Sec. III, we use TPT to derive equations for the exact milestoning rates k_{ij} for $T \to \infty$, i.e., for the case of perfect sampling. On the one hand, these equations explain the formula for extraction of phenomenological rates from the milestoning rates. On the other hand, they explain when and why the milestoning process is approximately Markov. In Secs. IV and V, all the different quantities from Fig. 1 are illustrated on a simple test system and a small peptide example, and the performances of the new core MSMs are compared to those of standard full-partition MSMs. We give concluding remarks in Sec. VI. The most technical derivations are deferred to several appendices. For further mathematical details on how Markov models based on such cores may indeed approximate accurately the essential kinetic features of an MD system, and in some cases may even be superior to full partition Markov models, we refer the reader to Refs. 35 and 46.

II. TRANSITION RATES BETWEEN CORE SETS FROM MILESTONING

In this section, we assume that a few disjoint sets of state space have been identified as "cores," each one assigned to one of the dominant metastable states of the system. A core is the center of some metastable set in state space so that the system stays in the vicinity of this core for relatively long periods of time before making a transition towards another core. Below we derive an unbiased estimator for the transition rates between these core sets which is based on treating them as milestones.[19,36,40] Note that the magnitude of these transition rates may be sensitive to the precise definition of the core sets and are thus not phenomenological transition rates as, e.g., obtained by reactive flux theory[47] or comparable approaches. In order to calculate phenomenological rates the transition rates between the cores must be reweighted as described in Sec. II F. We emphasize that the following considerations do not depend on the dimensionality of the system, and that the accuracy only depends on the choice of the core sets and on the extend of sampling available.

A. Microscopic dynamics and core sets

Consider a state space Ω which contains all dynamical variables needed to describe the instantaneous state of the system. Ω may be discrete or continuous, and we treat the more general continuous case here. For molecular systems, Ω usually contains both positions and velocities of the species of interest and surrounding bath particles. $\boldsymbol{x}(t) \in \Omega$ will denote the microscopic dynamical process considered, which is continuous in space, and may be either time-continuous or time-discrete (e.g., when considering time-stepping schemes for computational purposes). We will adapt our notation to the time-continuous case and will add short remarks if the time-discrete case differs in some important aspect.

It is required that $\boldsymbol{x}(t)$ is uniformly ergodic, i.e., for $t \to \infty$ the trajectory will come arbitrarily close to each state $\boldsymbol{x}$ infinitely often. The fraction of time that the system spends in any of its states during an infinitely long trajectory can then be estimated from its unique, positive stationary density $\mu(\boldsymbol{x})$ that in molecular processes corresponds to the equilibrium probability density for some associated thermodynamic ensemble (e.g., NVT, NpT). For molecular dynamics at constant temperature T, the dynamics above yield a stationary density $\mu(\boldsymbol{x})$ that is a function of T, namely the Boltzmann distribution $\mu(\boldsymbol{x}) = Z(\beta)^{-1} \exp(-\beta H(\boldsymbol{x}))$ with Hamiltonian $H(\boldsymbol{x})$ and $\beta = 1/k_B T$ where k_B is the Boltzmann constant and $k_B T$ is the thermal energy. $Z(\beta) = \int \exp(-\beta H(\boldsymbol{x}))\, d\boldsymbol{x}$ is the partition function.

At this point we do neither assume that $\boldsymbol{x}(t)$ is Markovian nor that it fulfills detailed balance. However, when both these properties are fulfilled, then some particularly simple formal statements relating the microscopic and macroscopic dynamics can be made (see Sec. III).

In the following, $\boldsymbol{x}(t)$ shall also indicate a realization (trajectory) of the process under investigation. While it is assumed for simplicity that a single long trajectory is studied, the procedure is straightforwardly applicable to multiple trajectories provided that they are "sufficiently long" (see below for details).

B. Set-up: Core sets and milestoning

We are interested particularly in metastable dynamical systems and want to quantify the statistics of rare events in such systems. For this, we introduce disjoint *core sets* B_1, $B_2, \ldots, B_N$ that will act as *milestones*.[19,36,40] The core sets form a subset of state space, $\bigcup_i B_i \subset \Omega$, but in general they do *not* partition state space. The basic idea is that in order to describe the rare event statistics, it is sufficient to know which

milestones the system has visited at which times. Implicit in this idea are two assumptions:

1. The time to equilibrate within (the vicinity of) any one core set B_i is much faster than the mean time to transition between any two core sets B_i and B_j. That is, no core set must contain multiple subsets that are metastable on the timescale of interest.
2. When $\boldsymbol{x}(t)$ is outside any core ($\boldsymbol{x}(t) \notin \bigcup_i B_i$), it will "quickly" (compared to the slow timescales of interest) hit one of the cores. That is to say that all *dominant* metastable states of the system are characterized by core sets (there may be lots of additional metastable states whose life time is well below the timescales of interest).

Some comments on how to find good core sets in practice will be given below. For now, we assume that a trajectory $\boldsymbol{x}(t)$ and the core sets are given. Next, $\boldsymbol{x}(t)$ is mapped onto the core states by defining a coarse-grained trajectory $b(t)$ which contains, at any time t, the index of the last milestone $\boldsymbol{x}(t)$ hit:

$$b(t) = \text{index of the last milestone hit by } \boldsymbol{x}(t). \tag{1}$$

Thus, $b(t)$ is a piecewise constant function taking values in $\{1, 2, \ldots, N\}$ which jumps from one value to another each time the trajectory $\boldsymbol{x}(t)$ hits a new milestone (successive hits of the same milestone without hit of another one in between do not change $b(t)$).

The key assumption made in milestoning is that the evolution of $b(t)$ can be modeled by a continuous-time Markov jump process. The validity of this assumption will be discussed in Sec. III. For now we assume that $b(t)$ is a Markov process and explore the consequences of this assumption.

The evolution of some Markov process $b(t)$ in discrete, finite state space is completely specified by a set of rates $k_{i,j} \geq 0$ with $i \neq j$, each of which gives the average number of jumps from state i and into state j per unit time. For example, the probability that $b(t) = i$, which we denote as $\rho_i(t)$, evolves according to the Master equation[48]

$$\frac{d\rho_i(t)}{dt} = -\sum_{j\neq i} \rho_i(t) k_{i,j} + \sum_{j\neq i} \rho_j(t) k_{j,i}. \tag{2}$$

The first term at the right hand side of this equation accounts for changes in $\rho_i(t)$ due to the probability flux out of state i whereas the second one accounts for changes due to the flux into this state.

For later use, let us also characterize the probability that $b(t)$ follows a particular path

$$\text{path} = [(b_0, \Delta_0), (b_1, \Delta_1), \ldots, (b_M, \Delta_M)], \tag{3}$$

i.e., $b(t)$ was in state b_0 during the interval $[0, t_1)$ for a time $\Delta_0 = t_1$, then jumped to state b_1 where it stayed during $[t_1, t_2)$ for a time $\Delta_1 = t_2 - t_1$, etc., and then finally jumped to state b_M at time t_M where it stayed during $[t_M, T)$ for a time $\Delta_M = T - t_M$.

As we assume that $b(t)$ is a Markov process, the probability of this realization can be written as a product of probabilities of single jumps:

$$\mathbb{P}(\text{path}|\text{rates}) = \rho_{b_0,b_1}(\Delta_0)\rho_{b_1,b_2}(\Delta_1)\cdots\rho_{b_{M-1},b_M}(\Delta_{M-1}), \tag{4}$$

with the individual probabilities $\rho_{i,j}$ given by the rates $k_{i,j}$:

$$\rho_{i,j}(\Delta) = e^{-\sum_{l\neq i} k_{i,l}\Delta} k_{i,j} \quad (i \neq j), \tag{5}$$

as the basic theory of Poisson processes tells us.[48] Formula (4) is normalized such that if we sum the indices $b_1, b_2, \ldots, b_M$ from 1 to N and integrate all Δ_i from 0 to ∞, we obtain 1.

Formula (4) gives the probability $\mathbb{P}(\text{path}|\text{rates})$ to observe a specific path of $b(t)$ given the rates $k_{i,j}$. These rates, however, are unknown *a priori*. In order to estimate them from the available MD data what we need instead of Eq. (4) is the probability of the rates $k_{i,j}$ given the path. This inversion can be done via Bayes formula given in Sec. II C. The Bayes formula is the basis both for the maximum likelihood estimation (MLE) procedure presented in Sec. II D and the sampling strategy presented in Sec. II E.

C. Bayesian formalism

Suppose that we have observed the MD trajectory $\boldsymbol{x}(t)$ over times $[0, T]$ and thereby deduced a path $b(t)$ between core sets as given by Eq. (3). Neglecting the last step waiting time which is undetermined due to the termination of the trajectory at T, the probability of the path P is given by the product in Eq. (4), which can be rewritten as

$$\mathbb{P}(\text{path}|\text{rates}) = \prod_{\substack{i,j=1 \\ i\neq j}}^{N} k_{i,j}^{N_{i,j}^T} e^{-k_{i,j}R_i^T}. \tag{6}$$

Here $N_{i,j}^T$ is the number of transitions from state i to state j observed along $b(t)$ during the time interval $[0, T]$ and R_i^T is the total time during which $b(t) = i$ in $[0, T]$, i.e., $R_i^T = \int_0^T \delta_{i,b(t)} dt$. As we will see shortly, $N_{i,j}^T$ and R_i^T are the only quantities needed to estimate the rates.

The notation $\mathbb{P}(\text{path}|\text{rates})$ in Eq. (6) stresses that this quantity is the probability of the path $b(t)$, $t \in [0, T]$ given the rates $k_{i,j}$. As explained before, what we need to estimate these rates is the probability of $k_{i,j}$ given the specific, observed path, $\mathbb{P}(\text{rates}|\text{path})$. The two probabilities $\mathbb{P}(\text{path}|\text{rates})$ and $\mathbb{P}(\text{rates}|\text{path})$ can be related to each other via Bayes formula

$$\mathbb{P}(\text{path}|\text{rates})\mathbb{P}(\text{rates}) = \mathbb{P}(\text{rates}|\text{path})\mathbb{P}(\text{path}), \tag{7}$$

where $\mathbb{P}(\text{rates})$ and $\mathbb{P}(\text{path})$ are the probabilities of the rates $k_{i,j}$ and of the path $b(t)$, $t \in [0, T)$, respectively. $\mathbb{P}(\text{rates})$ is usually referred to as the prior probability and it incorporates any *a priori* information we have about the rates $k_{i,j}$. $\mathbb{P}(\text{path})$ does not depend on $k_{i,j}$, and so it can be incorporated in a normalization constant as far as sampling $\mathbb{P}(\text{rates}|\text{path})$ over the rates is concerned. Therefore combining Eqs. (6) and (7) we obtain that $\mathbb{P}(\text{rates}|\text{path})$ is simply given by

$$\mathbb{P}(\text{rates}|\text{path}) \propto \mathbb{P}(\text{rates}) \prod_{\substack{i,j=1 \\ i\neq j}}^{N} k_{i,j}^{N_{i,j}^T} e^{-k_{i,j}R_i^T}. \tag{8}$$

In the limit of poor statistics, the choice of the prior significantly influences the results. Hummer[19] has proposed to use

a prior that is uniform in the logarithm of the rates. Here, we assume for simplicity a uniform prior ($\mathbb{P}(\text{rates}) \propto 1$), which is acceptable in the limit of good statistics, and work out the results based on this choice. For a uniform prior, the posterior probability (8) is proportional to the likelihood, and is thus maximized by maximizing the likelihood.

D. Maximum likelihood estimate

The larger the amount of available data (i.e., the larger T and thus the longer the observed path), the larger $N_{i,j}^T$ and R_i^T become—both these quantities grow linearly in T when $T \to \infty$, see Sec. III B. Since $\mathbb{P}(\text{rates})$ is fixed (it does not depend on T) this implies that Eq. (8) becomes increasingly peaked around the values of $k_{i,j}$ that maximizes the product in Eq. (8). It is easy to see by differentiation of this product over $k_{i,j}$ and direct solution of the resulting equations that it is maximized by

$$\hat{k}_{i,j} = \frac{N_{i,j}^T}{R_i^T}, \quad i \neq j. \tag{9}$$

This is the so-called MLE for the rates $k_{i,j}$. The MLE $\hat{k}_{i,j}$ is unbiased, i.e., it converges to the exact rates of this process when $T \to \infty$. In practice, however, the available data are always finite, $T < \infty$, that is, we have to consider the finiteness of the *sampling* that underlies the MLE $\hat{k}_{i,j}$. As usual this finite sampling introduces some statistical errors. How to estimate these sampling errors based on the available data is discussed in Sec. II E.

E. Statistical uncertainty

When the amount of available data is finite, instead of maximizing the product in Eq. (8) we can sample this probability over the rates $k_{i,j}$. This permits to estimate the statistical errors in the rates and it is especially simple for a uniform prior. In this case, Eq. (8) shows that the rates $k_{i,j}$ are statistically independent, and each $k_{i,j}$ is distributed according to a gamma distribution with scale parameter $N_{i,j}^T + 1$ and shape parameter $1/R_i^T$. When it is desired to estimate statistical uncertainties of quantities derived from K, these distributions can be sampled straightforwardly using standard routines for generation of gamma distributed random numbers such as e.g. `gamrnd` in MATLAB. This is one aspect in which core MSMs based on milestoning are easier to work with than the standard MSM approaches discussed in Appendix D.

F. Phenomenological rates

While the rates k_{ij} quantify the number of transitions per time between core sets, one is often interested in the so-called *phenomenological rate matrix* K^* with entries $k_{i,j}^*$ that can be compared to the appropriately estimated reactive flux rate constants, see Refs. 47 and 49. The $k_{i,j}^*$ are the rate constants that are typically of interest to Chemical Physicists since the eigenvalues of K^* are the intrinsic relaxation rates $\lambda_1^r, \ldots, \lambda_m^r$ which can be probed experimentally. In the special case of a system with 2-state kinetics, there is a single relaxation rate $\lambda_2^r = k_{1,2}^* + k_{2,1}^*$.

In order to explain how these phenomenological rates can be computed from the k_{ij}, let us denote by $R_{i,j}^T$ the total time a trajectory generated on $[0, T]$ spent while being $i \to j$ reactive and assigned to B_i. Based on this, we define the so-called mass matrix $\hat{M}$ by

$$\hat{m}_{i,j} = \frac{R_{i,j}^T}{R_i^T}. \tag{10}$$

$m_{i,j}$ is the fraction of time the system is $i \to j$ reactive.

Given the mass matrix $\hat{M} = [\hat{m}_{i,j}]$, the matrix of phenomenological rate constants can be derived as

$$\hat{K}^* = \hat{M}^{-1}\hat{K}. \tag{11}$$

This equation is not intuitively obvious but will be derived in Sec. III.

III. RELATION BETWEEN MILESTONING DYNAMICS AND MICROSCOPIC DYNAMICS

In this section, we investigate the exact transition rates of milestoning dynamics between core sets, i.e., the rates obtained in the limit of an infinitely long trajectory (i.e., when $T \to \infty$) where the statistical uncertainty vanishes. This is done by expressing the milestoning rates in terms of the original microscopic dynamics in continuous state space *via* Transition Path Theory.[41–45] This relation provides the theoretical basis for further investigations on how the milestoning rates depend on the size and exact definition of the core sets. This will allow us to analyse the relation between the milestoning rates and phenomenological rates of the process under investigation which is treated in Sec. III D.

A. Microscopic dynamics and generator

We now define some properties of the microscopic dynamics $\boldsymbol{x}(t)$. For the present chapter we assume $\boldsymbol{x}(t)$ is a Markov process, i.e., the instantaneous change of the system $(d\boldsymbol{x}(t)/dt$ in time-continuous and $\boldsymbol{x}(t + \Delta t)$ in time-discrete dynamics with time step $\Delta t)$, is calculated based on $\boldsymbol{x}(t)$ alone and does not require the previous history. As a result of Markovianity in Ω, the transition probability density $p(\boldsymbol{x}, \boldsymbol{y}; \tau)$ is well defined: For every pair of state $\boldsymbol{x}, \boldsymbol{y} \in \Omega$ and a given lag time $\tau \in \mathbb{R}_{0+}$, it is given by

$$p(\boldsymbol{x}, \boldsymbol{y}; \tau)\, d\boldsymbol{y} = \mathbb{P}[\boldsymbol{x}(t + \tau) \in d\boldsymbol{y} \mid \boldsymbol{x}(t) = \boldsymbol{x}], \tag{12}$$

i.e., by the probability that a trajectory started at time t from the point $\boldsymbol{x} \in \Omega$ will be in an infinitesimal region $d\boldsymbol{y}$ around a point $\boldsymbol{y} \in \Omega$ at time $t + \tau$.

Furthermore, we assume that $\boldsymbol{x}(t)$ is reversible, i.e., $p(\boldsymbol{y}, \boldsymbol{x}; \tau)$ fulfills the condition of detailed balance:

$$p(\boldsymbol{x}, \boldsymbol{y}; \tau)\, \mu(\boldsymbol{x}) = p(\boldsymbol{y}, \boldsymbol{x}; \tau)\, \mu(\boldsymbol{y}), \tag{13}$$

i.e., in equilibrium, the number of systems transitioning from $\boldsymbol{x}$ to $\boldsymbol{y}$ per time is the same as the number of system transitioning from $\boldsymbol{y}$ to $\boldsymbol{x}$. Note that this "reversibility" is a more general concept than the time-reversibility of the dynamical equations as, e.g., encountered in Hamiltonian dynamics. For

example, Brownian dynamics on some potential are reversible as they fulfill Eq. (13), but are not time reversible in the same sense as Hamiltonian dynamics. Although detailed balance is not formally required to construct MSMs using milestoning, it is useful for the theoretical results of the present section; in Ref. 46 it has been worked out how to deal with the nonreversible case in a way very similar to what is outlined in the following. Note that detailed balance is expected to hold in equilibrium molecular dynamics due to basic physical arguments, although this is not true for all computer implementations of equilibrium molecular dynamics.[13]

With the definition of the transition probabilities in Eq. (12), we can easily write down how the probability density $\rho(\boldsymbol{x}, t = 0)$ of finding the system (or, more generally, an arbitrary function) in state $\boldsymbol{x}$ is propagated by the microscopic dynamics:

$$\rho(\boldsymbol{y}, t)\mu(\boldsymbol{y}) = \int_\Omega p(\boldsymbol{x}, \boldsymbol{y}; t)\rho(\boldsymbol{x}, t = 0)\mu(\boldsymbol{x})d\boldsymbol{x}, \tag{14}$$

where we used probability densities relative to the invariant measure which has technical advantages because of reversibility. The propagation equation (14) can be written in much simpler form since the Markov property of the underlying dynamics implies that is has a generator:

$$\frac{d\rho(\boldsymbol{x}, t)}{dt} = L\rho(\boldsymbol{x}, t), \tag{15}$$

such that $\rho(\boldsymbol{x}, t) = \exp(tL)\rho(\boldsymbol{x}, 0)$. This generator exists for all time-continuous Markov processes but can take significantly different from: For example, for diffusive dynamics the Eq. (15) is just the Fokker-Planck equation; this specific case and the associated formula for L (that gives L as a partial differential operator) is outlined in Appendix A. For Markov jump processes in discrete state space, L is a rate matrix instead. In the discrete-time setting, Eq. (15) has to be modified, see Ref. 46 for details.

B. Exact milestoning rates from transition path theory

Using TPT together with the generator description of the microscopic dynamics, we can derive exact expressions for some useful quantities that will help us to characterize the transition rates of the milestoning process:

$$\pi_i = \lim_{T\to\infty} \frac{R_i^T}{T}, \quad \nu_{i,j} = \lim_{T\to\infty} \frac{N_{i,j}^T}{T}, \quad (i \neq j). \tag{16}$$

Here π_i is the proportion of time during which the last milestone hit was B_i while $\nu_{i,j}$ is the average rate of transition between B_i and B_j $(i \neq j)$. Clearly, in terms of these limits, the exact rate $k_{i,j} = \lim_{T\to\infty} \hat{k}_{i,j}$ is given by

$$k_{i,j} = \frac{\nu_{i,j}}{\pi_i}, \quad (i \neq j). \tag{17}$$

π_i is the equilibrium distribution of the Markov jump process $b(t)$ (i.e., the stationary solution of Eq. (2)) provided that the microscopic dynamics $\boldsymbol{x}(t)$ is reversible. In that case, $\nu_{i,j} = \nu_{j,i}$, which from Eq. (17) implies a detailed balance condition of the milestoning dynamics in which π_i is the equilibrium distribution:

$$\pi_i k_{i,j} = \pi_j k_{j,i}. \tag{18}$$

To see how the limits in Eq. (16) can be computed using TPT, we need to recall a few key facts about this theory. In a nutshell, TPT analyzes the property of the reactive trajectories by which specific reactions occur. In the present context the reactions of interest are the N reactions where each of core set B_i is taken as a product state, and the union of all other core sets, $\cup_{j\neq i} B_j$, as reactant state. As shown in TPT, the statistical properties of the reactive trajectories associated with each of these reactions can by expressed solely in terms of the equilibrium probability density of the system $\mu(\boldsymbol{x})$, and the so-called committor functions (one per core set) which are the solutions q_i of

$$\begin{cases} Lq_i(\boldsymbol{x}) = 0 & \boldsymbol{x} \notin B_j \forall j, \\ q_i(\boldsymbol{x}) = 1 & \boldsymbol{x} \in B_i, \\ q_i(\boldsymbol{x}) = 0 & \boldsymbol{x} \notin B_j \forall j \neq i, \end{cases} \tag{19}$$

where the operator L again denotes the infinitesimal generator of the original process, the boundary conditions for q_i are specified by the second and third line.

The committor functions have a simple probabilistic interpretation: $q_i(\boldsymbol{x})$ gives the probability that the trajectory starting at $\boldsymbol{x}$ will reach B_i before it reaches any of the other sets, i.e., before $\cup_{j\neq i}^N B_j$ is entered. This property, together with the fact that the dynamics is Markovian can be used to get $\nu_{i,j}$ and π_i as follows. Suppose we ask what the equilibrium probability density $\mu_i(\boldsymbol{x})$ is to find the process in $\boldsymbol{x}$ given that the last milestone it came from was B_i? Using reversibility under time reversal, this is equivalent to ask what is the equilibrium probability density to find the process in $\boldsymbol{x}$ and that the next set it will hit is B_i and so $\mu_i(\boldsymbol{x})$ is given explicitly by

$$\mu_i(\boldsymbol{x}) = \mu(\boldsymbol{x})q_i(\boldsymbol{x}). \tag{20}$$

If we now integrate this quantity w.r.t. $\boldsymbol{x}$ over all the configuration space we obtain the equilibrium probability that the last set hit was B_i regardless of where the process actually is. From Eq. (16) this is precisely π_i and so

$$\pi_i = \int_\Omega \mu(\boldsymbol{x})q_i(\boldsymbol{x})d\boldsymbol{x}. \tag{21}$$

A similar argument can be used to obtain $\nu_{i,j}$ for all reversible processes[46] and show that it can be calculated by means of

$$\nu_{i,j} = \int_\Omega \mu(\boldsymbol{x})q_i(\boldsymbol{x})(Lq_j)(\boldsymbol{x})d\boldsymbol{x}. \tag{22}$$

We outline the details of the derivation of Eq. (22) for the case of diffusive dynamics in Appendix B.

Formulas (21) and (22) are the desired expressions for π_i and $\nu_{i,j}$. Therefore, whenever the (reversible) dynamics has a generator L and a unique equilibrium probability density, this density together with the committor functions uniquely determine the quantities π_i and $\nu_{i,j}$ that are needed to compute the exact rates $k_{i,j}$ via Eq. (17).

C. Galerkin projection interpretation

It is interesting to revisit the formula above from the view point of Galerkin projection. Therefore consider the space of square-integrable functions

$$H = \{f : \Omega \to \mathbb{R} : \int_\Omega |f(\boldsymbol{x})|^2 \mu(\boldsymbol{x}) d\boldsymbol{x} < \infty\}, \tag{23}$$

equipped with the scalar product

$$\langle f, g\rangle_\mu = \int_\Omega \mu(\boldsymbol{x}) f(\boldsymbol{x}) g(\boldsymbol{x}) d\boldsymbol{x}. \tag{24}$$

We consider the operator P that when acting on a function $f \in H$ returns its best approximation Pf in the finite-dimensional subspace

$$\mathbb{S} = \{f : \Omega \to \mathbb{R} : \ f = \sum_{i=1}^N \alpha_i q_i, \ \alpha_i \in \mathbb{R}\}, \tag{25}$$

that is spanned by the functions $q_1, \ldots, q_N$. As outlined in Appendix C the best approximation Pf can be computed from its property that the associated approximation error $f - Pf$ is orthogonal to $\mathbb{S}$, and thus has to satisfy

$$\langle f - Pf, q_j\rangle_\mu = 0, \quad \forall j = 1, \ldots, N. \tag{26}$$

Again from Appendix C we learn that therefore P can be expressed as

$$(Pf)(\boldsymbol{x}) = \sum_{i,j=1}^N q_i(\boldsymbol{x}) (S^{-1})_{i,j} \langle q_j, f\rangle_\mu, \tag{27}$$

where S^{-1} is the inverse of the matrix with entries

$$s_{i,j} = \langle q_i, q_j\rangle_\mu, \tag{28}$$

which exists since the q_i are linearly independent. Consider now the eigenvalue problem associated with the generator L of the process:

$$L\phi^e = \lambda^e \phi^e, \tag{29}$$

where the superscript "e" stands for exact. The Galerkin projection of this equation on the subspace $\mathbb{S}$ reads

$$PLP\phi = \lambda P\phi. \tag{30}$$

After a little algebra, it is easy to see that this equation can be written explicitly as

$$\sum_{j=1}^N \nu_{i,j} r_j = \lambda \sum_{j=1}^N s_{i,j} r_j, \tag{31}$$

where $\nu_{i,j}$ is given by

$$\nu_{i,j} = \langle q_i, Lq_j\rangle_\mu, \tag{32}$$

as already discussed and we defined

$$P\phi = \sum_{i=1}^N r_i q_i, \quad \text{s.t.} \quad r_i = \sum_{j=1}^N (S^{-1})_{i,j} \langle q_j, \phi\rangle_\mu. \tag{33}$$

Dividing both side of Eq. (31) by π_i, and defining the so-called mass matrix with entries

$$m_{i,j} = \frac{s_{i,j}}{\pi_i}. \tag{34}$$

Equation (31) can also be expressed as

$$\sum_{j=1}^N k_{i,j} r_j = \lambda \sum_{j=1}^N m_{i,j} r_j, \tag{35}$$

where $k_{i,j}$ is the rate matrix defined in Eq. (17). This means that (a) we can compute the exact transition rates $k_{i,j} = \lim_{T\to\infty} \hat{k}_{i,j}$ of the milestoning process by means of Galerkin projection of the generator *and* that (b) the associated eigenvalues λ_j of the generalized eigenvalue problem (35) will somehow be approximations of the original eigenvalues λ^e of the original eigenvalue problem $L\phi^e = \lambda^e \phi^e$.

What does the approximation by Galerkin projection entail in terms of evolution of observables and densities? Since for reversible processes L is a self-adjoint operator the propagation of observables *as well as* densities under the underlying microscopic dynamics is related to the action of the propagator $\exp(tL)$. That is, when starting with an observable or density f at time 0, its evolution in state space is given by $\exp(tL)f$. Suppose that we start from some $f(\boldsymbol{x})$ which belongs to the subspace spanned by $q_1, q_2, \ldots, q_N$, i.e., such that $f = Pf$. Then, the value at time t of this observable will be

$$\exp(tL)Pf, \tag{36}$$

and its projection on the subspace spanned by $q_1, q_2, \ldots, q_N$ is simply

$$P\exp(tL)Pf. \tag{37}$$

From the considerations above it is easy to see that the Galerkin projection amounts to approximating the evolution of the observable by

$$\exp(tPLP)f. \tag{38}$$

In Refs. 35 and 50 the associated approximation error is analyzed and bounds on the error are derived.

To sum up, what the above considerations show is that milestoning does more than give the Markov approximation of the evolution of the index process $b(t)$ discussed in Sec. II B. Milestoning also permits to approximate the evolution of observables in the system and, in particular, it gives estimates of the leading eigenvalues of the generator of the original process via solution of Eq. (35).

D. Exact phenomenological rate constants

We now combine the results of Secs. III A–III C in order to provide an estimator of phenomenological rate constants between the metastable states of the system. In Sec. II, it was described how the milestoning rate constants $k_{i,j}$ can be estimated. Those rates, however, are sensitive to the exact definition of cores within the metastable states. In order to arrive at phenomenological rate constants, the milestoning rate constants $k_{i,j}$ need to be corrected for the influence of the core definition. As was shown in Sec. III C this can be done by

introducing the masses $m_{i,j}$, see Eq. (35). We can rewrite this generalized Eigenvalue problem in matrix form as

$$Kr = \lambda Mr, \tag{39}$$

from which we see that the reweighted matrix $K^* = M^{-1}K$ with

$$K^*r = \lambda r, \tag{40}$$

has eigenvalues that approximate the original eigenvalues of the system and therefore its dominant intrinsic relaxation timescales.

Even though the derivation of Eq. (35) relies on the interpretation of $m_{i,j}$ in terms of the committor functions q_i, in milestoning we can obtain both $m_{i,j}$ without explicit knowledge of q_i. The factors $s_{i,j}$ defined in Eq. (34) give the proportion of time during which the trajectory is on its way from B_i to B_j, that is the last milestone hit was B_i and the next milestone hit will be B_j. It follows that $s_{ij} = \lim_{T\to\infty} R_{ij}^T$. With the approximate mass matrix $\hat{M}$ with entries $\hat{m}_{ij}$ as defined in Eq. (10), the entries of the exact mass matrix are thus given by

$$m_{i,j} = \lim_{T\to\infty} \hat{m}_{ij}, \tag{41}$$

which directly provides an estimator of $m_{i,j}$ from finite-time trajectories. That is, the mass matrix can be approximated from a long trajectory by monitoring transition times. This is remarkable because the calculation of q_i is a formidable task in high dimensional systems. Based on the estimate of $m_{i,j}$, K^* can be estimated as described in Sec. II F.

E. Consequences for the definition of the core sets

The approximation quality of the MSM based on milestoning depends on both the characteristics of the original dynamics and on the choice of the core sets $B_1, B_2, \ldots, B_N$. How to assess the error that the Markov assumption introduces is not obvious. Mathematical results in this direction are available in Refs. 35, 46, and 50, whose intuition can be understood from our interpretation via Galerkin projection discussed in Sec. III C. Indeed, this interpretation suggests that the MSM based on milestoning will work well if the space spanned by the eigenvectors corresponding to the low-lying eigenvalues of L is well approximated by the space spanned by the committor functions $q_1, \ldots, q_N$. In this case, the Galerkin projection PLP will approximate well the low-lying eigenvalues of the generator L, so that the long-time behavior will be captured, see the example in Sec. IV, and especially the discussion of Fig. 4.

Furthermore, it can be shown that if the Galerkin projection error is small, then also the Markov assumption for the process $b(t)$ is justified.[35] This will also be illustrated in one of the examples in Sec. IV.

How to use this assessment to constructively choose the core sets is a much harder question which goes beyond the scope of this paper. However, let us make a few comments. Technically, the above considerations require that the generator L of the original dynamics possesses a group of eigenvalues which are somewhat smaller in magnitude than all the other ones. The existence of small eigenvalues indicates that slow processes are taking place in the original state space. These slow processes are what the generalized eigenvalue problem in Eq. (35) is meant to capture, in the sense that the generalized eigenvalues should be close to the small eigenvalues of the original process. Clearly, this shows that we should choose N sets if there are N small eigenvalues, since a Markov jump process on N states has exactly N eigenvalues. In fact, if we assume that the original process has N small eigenvalues, then general results guarantee the existence of a good collection of sets $B_1, B_2, \ldots, B_N$. What this argument does not tell, however, is how to choose these sets, because in general we will not be able to compute the dominant eigenvectors and committor functions that would be needed to identify the sets based on the above insight. In other words, what these sets are is not given explicitly, except for the rather vague property that the trajectory $\boldsymbol{x}(t)$ should oscillate inside and around each for a long time before visiting another. How to use this criterion in a constructive way is the subject of current research, and we shall not dwell on this issue further here. However, the questions discussed above will be illustrated via examples in Secs. IV and V.

IV. ILLUSTRATIVE EXAMPLE: DOUBLE WELL POTENTIAL WITH DIFFUSIVE TRANSITION REGION

As a first example, Let us consider a one-dimensional overdamped diffusion processes. The associated equation of motion in the one-dimensional energy landscape $V(x)$ reads

$$\gamma\dot{x}(t) = -V'(x(t)) + \sqrt{2\beta^{-1}\gamma}\,\eta(t), \tag{42}$$

where γ is the friction coefficient, and $\beta = 1/k_BT$ where k_B is Botzmann's constant and T the system temperature, $\eta(t)$ a white-noise process, i.e., a Gaussian process with mean zero and covariance $\langle\eta(t)\eta(t')\rangle = \delta(t-t')$, and V' dnotes the first derivative of the energy landscape function. We consider $\gamma = 1$ and $\beta^{-1} = 0.25$ together with the potential $V(x)$ given by

$$V(x) = \begin{cases} (1-x^2)^2, & x \le 0; \\ \frac{4}{5} + \frac{1}{5}\cos(\pi x), & 0 \le x \le 8; \\ (1-(x-8)^2)^2, & x \ge 8. \end{cases} \tag{43}$$

The potential has two deep wells connected by an extended transition region with substructure, see Fig. 2.

The transition region between the two deep wells contains four smaller wells that each acts as dynamical trap for the transitions between these two deep wells. This can be seen from the equilibrium density also shown in Fig. 2. The minima in the two deep wells are located at $x_0 = -1$ and $x_1 = 9$, and the respective saddle points that separate the deep wells from the rest of the landscape are located at $x_0^s = 0$, and $x_1^s = 8$, respectively, with energy barrier equal to 1.

The associated generator L is given by Eq. (A3) in the appendix. Its eigenvalues can be obtained by solving Eq. (29) numerically. This gives the following estimates for the first 7

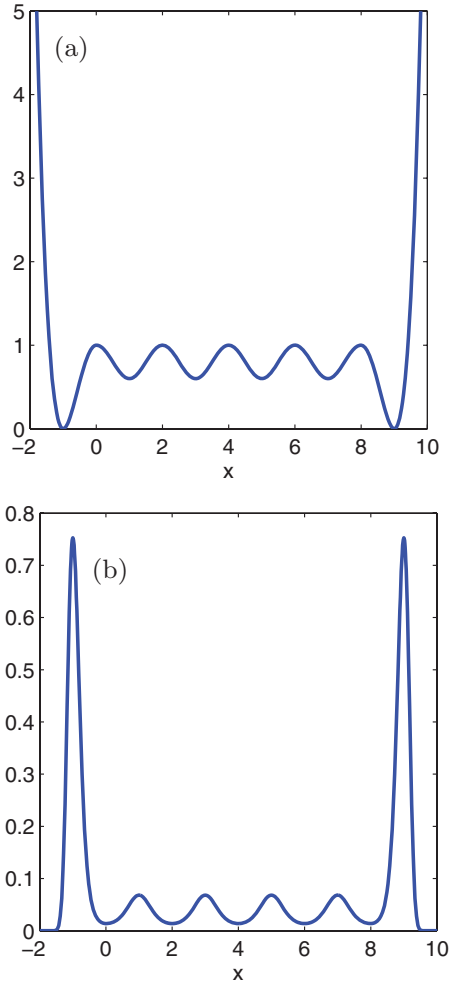

FIG. 2. (a) The potential $V(x)$ with extended transition region and (b) the associated equilibrium density $\mu(x) = \exp(-\beta V(x))$ for $\beta^{-1} = 0.25$.

eigenvalues with smallest amplitude: $\lambda_0^e = 0$ and

λ_1^e	λ_2^e	λ_3^e	λ_4^e	λ_5^e	λ_6^e
-0.0036	-0.0283	-0.0860	-0.1631	-0.2298	-1.3603

The eigenvalue λ_1^e measures the metastability between the two deep wells: the associated timescale is $|\lambda_1^e|^{-1} \approx 275.69$. The four next eigenvalues $\lambda_2^e, \ldots, \lambda_5^e$ measure metastable effects associated with switches between the four additional small wells. These effects make the example more challenging.

To build the MSM using either milestoning or the standard procedure based on a full partition of state space,[13] we generated a long trajectory from the overdamped equation. This trajectory was computed using the Euler-Maruyama discretization with the stepsize $\Delta t = 0.001$ in the time interval $[0, T]$ with $T = 100\,000$. This stepsize is so small that we can consider the discrete solution to be almost identical to a path of the solution such that we can consider it as "almost continuous."

A. Core MSM based on milestoning

We choose two core sets of the form

$$B_0^\delta = (-\infty, x_0 + \delta], \quad B_1^\delta = [x_1 - \delta, \infty). \tag{44}$$

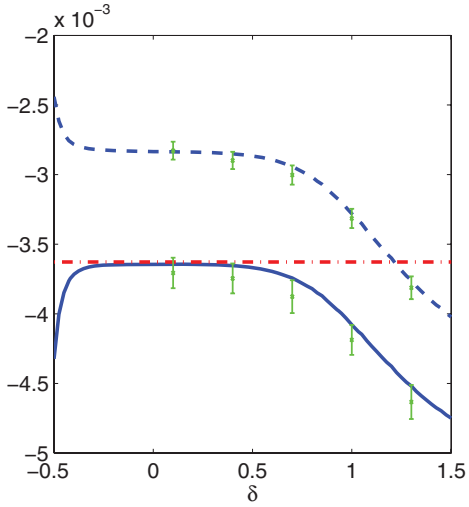

FIG. 3. Comparison of eigenvalues for different core MSM based on milestoning for the potential shown in Fig. 2. The flat dashed-dotted line indicates the first non-trivial eigenvalue of the original process, the dashed line the δ-dependent first non-trivial eigenvalue of K_δ, and the solid curved line indicates the first non-trivial eigenvalue of the corrected rate matrix K_δ^*. The vertical lines are the statistical error estimation based on 100 realizations of trajectories with length $T = 100\,000$.

Here δ is an adjustable parameter used to assess the robustness of the milestoning approximation with respect to the precise definition of the core sets. Note in particular that for $\delta < 0$, the core sets do not include the minima in the two deep well of the potential, whereas for $\delta \geq 1$ these sets include the first saddle point next to these minima. As a result, we expect that the accuracy of the core MSMs based on milestoning will deteriorate when either $\delta < 0$ or $\delta \geq 1$, but that these MSMs will be rather insensitive to the value of δ in the range $0 \leq \delta < 1$. The results below confirm this intuition.

First we compare the leading eigenvalue of the milestoning rate matrix K and the corrected eigenvalue due to Eq. (31) to the true leading eigenvalue. Figure 3 shows the result for different values of δ. It is observed that the eigenvalue of the uncorrected milestoning rate matrix K is biased. It significantly overestimates λ_1^e for small values of δ and then rapidly decays at large values of δ, with no significant range of δ values where λ_1^e is estimated correctly. In contrast, the corrected eigenvalue estimate using Eq. (31) is an excellent estimator for λ_1^e in the range $\delta \in [-0.25, 0.5]$, showing that the corrected milestoning rate matrix is a useful dynamical model. For very small and very large δ values, this estimate deteriorates as well.

To corroborate these observations, in Fig. 4 we show the eigenfunction of L corresponding to the first non-trivial eigenvalue, and its projection onto the subspace spanned by the committors q_0 and q_1 for different core sets depending on δ. We observe that for the core sets with $\delta = 0.1$, the eigenfunction is almost identical to its projection, while for $\delta = 1$ some discrepancy is visible which explains why the error in the eigenvalues in Fig. 3 for $\delta = 0.1$ is much smaller than for $\delta = 1$.

Finally, Fig. 5 exhibits the distribution of residence times r_1 in core set B_1^δ for $\delta = 1.0$; the residence times r_1 are the

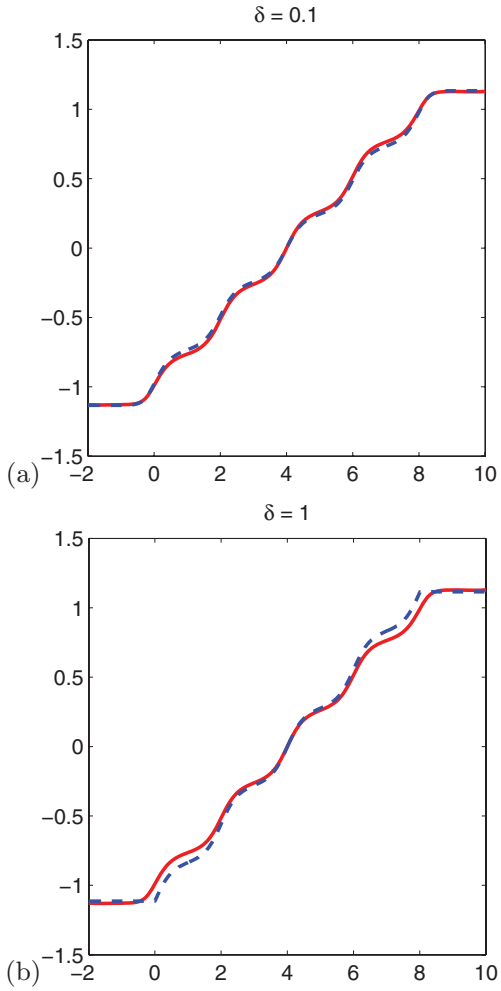

FIG. 4. Comparison of eigenfunction of L corresponding to the first non-trivial eigenvalue (solid lines) and its projection on the subspace spanned by the committors (dashed lines) for different core sets ((a) $\delta = 0.1$ and (b) $\delta = 1$) for the potential shown in Fig. 2.

length of the periods with $b(t) = 1$ along the time series. If $b(t)$ were a perfect Markov process the distribution of residence times in the core sets would be perfectly exponential with a decay rate given by the respective rate. We observe a distribution close to a single exponential and thus the deviations from the Markov property are small.

B. Full partition MSM

As recalled in Appendix D, in order to specify a standard MSM we have to specify a full partition of the state space into N sets and a lagtime τ. We will consider the following four partitions, A_1, A_2, A_3, and A_4, with N = 2, 3, 4, and 6:

$$A_1 = (-\infty, 4], (4, \infty),$$

$$A_2 = (-\infty, 0], (0, 8), [8, \infty),$$

$$A_3 = (-\infty, 0], (0, 4], (4, 8), [8, \infty),$$

$$A_4 = (-\infty, 0], (0, 2], (2, 4], (4, 6], (6, 8), [8, \infty),$$

where the refinements of the two-set partition are chosen such that the wells in the extended transition region are more and more resolved. For these partitions we computed the associated transition matrix P^* for different lagtimes τ based on the full observation. Figure 6 shows the estimated first non-trivial eigenvalue λ for the original process computed from the second-largest eigenvalue μ of P^* via $\lambda = \log(\mu)/\tau$ depending on the lagtime τ, compare Eq. (D4).

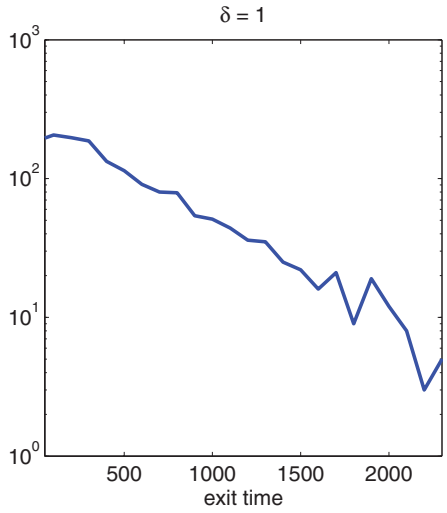

FIG. 5. Distribution of residence times in state $i = 1$ in a trajectory of length $T = 12\,000\,000$ for cores sets B_i^δ for $\delta = 1.0$: semi-logarithmic plot indicating almost singly exponential decay.

We observe that the quality of the approximation gets better as we increase the number of sets in the partition and/or lagtime. For large enough lagtime and fine enough partition the approximation quality is superior to that of the core MSM.

Finally, in Fig. 7 we compare the different MSM approaches and show the statistical (sampling) error of the eigenvalue estimate using 100 realizations of trajectories. The statistical error increases with the lagtime τ, since for larger τ the number of observed transitions decreases. In comparison to the core MSM, we observe that for small lagtime τ, the

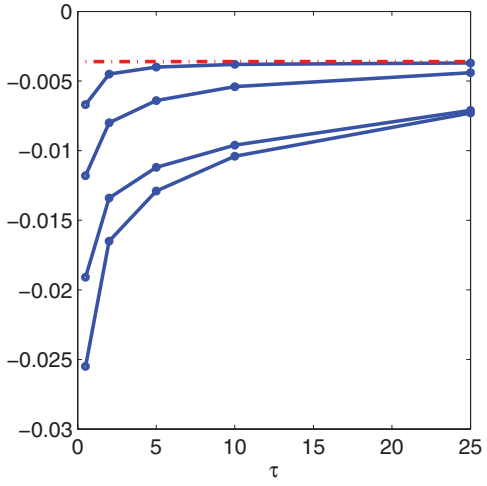

FIG. 6. Comparison of eigenvalue estimate for full-partition MSM with the different partitions A_1, A_2, ... for the potential shown in Fig. 2. The flat dashed-dotted line indicates the first non-trivial eigenvalue of the original process, and the four solid lines the estimate using P^* depending on the lag time τ. The higher the blue line, the more sets in the partition, from 2 for the bottom line (partition A_2) to 6 for the top one (partition A_4).

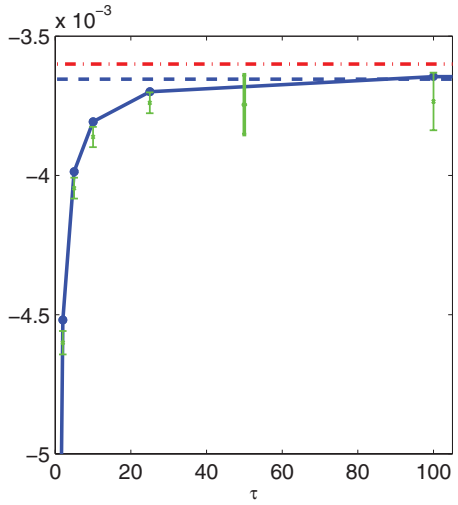

FIG. 7. Comparison of eigenvalue estimate using MSM based on milestoning and based on a full partition of state space into 6 sets. The flat dashed-dotted line indicates the first non-trivial eigenvalue of the original process, the flat dashed line the estimate using core set MSM with $\delta = 0.4$, and the solid (curved) line indicates the estimate using the standard MSM with 6 sets. The statistical errors are indicated with vertical lines: The one in the middle ($\tau = 50$) corresponds to the milestoning MSM, while the others are associated with the particular lagtime used.

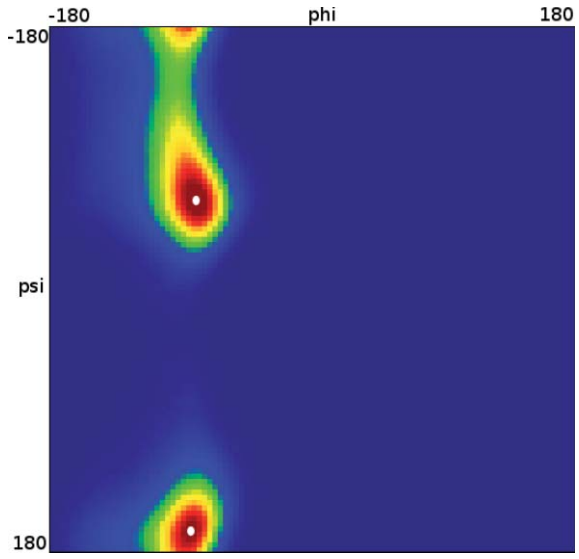

FIG. 8. Stationary distribution of alanine dipeptide in ϕ/ψ angle space. The two core centers are shown as white bullets.

statistical error of the full-partition MSM with fine partition is smaller, since for a trajectory with same length we have significantly more transitions between the boundaries of the sets in the partition than between the core sets. However, to achieve similar accuracy of the eigenvalue estimate, we need large lagtime ($\tau = 100$ in the figure), and the statistical error of the full-partition MSM then becomes similar to that of the core MSM. The reason is that the information in the trajectory that is useful to get the estimate of the first non-trivial eigenvalue is the transition between the core sets. While we have more transitions between the sets in the full partition when the number of these sets is large, they do not help in estimating the first non-trivial eigenvalue. Hence, for small τ, statistical errors are small but systematic bias is large, whereas for large τ, the systematic bias is reduced but the statistical errors become comparable to the ones of core set MSM.

V. PEPTIDE EXAMPLE

In this section the different ways to construct MSMs are compared by their ability to estimate the transition timescales of the $\alpha \rightleftharpoons \beta$ transition in explicitly solvated alanine dipeptide.

One molecule alanine dipeptide with termini ACE and CT3 using the CHARMM 27 force field was simulated in a box of 256 TIP3P water molecules using the program NAMD version 2.7b1. The box size was obtained by a short NPT equilibration and then held fixed followed by a 1 μs production run with Langevin dynamics at 300 K, using a friction constant of 5 ps^{-1} and options rigibonds all (all bond lengths fixed) and useSettle. Frames were saved every 10 fs. The dynamics of the system are monitored *via* its ϕ and ψ backbone dihedral angles, a density plot generated from a histogram of the simulation snapshots is shown in Fig. 8. This illustrates that the density is maximal at the α and β regimes, with a transition region between them at which trajectories cross between the two states.

We construct a full-partition MSM based on a partition of the ϕ/ψ coordinates. This partition was obtained by k-means clustering based on ϕ/ψ coordinates iterated to convergence. The use of ϕ/ψ is not essential here; we could also start from a clustering based on all or most dimensions of the state space. Here, we used k-means based on $k = 10$, 50, and 250 clusters which produced complete partitions of different resolution of the data set. Discrete trajectories were generated by mapping the continuous trajectories onto cluster numbers and counting transitions $c_{ij}(\tau)$ at several different lagtimes τ. The maximum likelihood transition matrices amongst these clusters were estimated *via* $\hat{p}_{ij}(\tau) = c_{ij}(\tau)/\sum_k c_{ik}(\tau)$, thus generating a series of standard Markov state models. Their slowest implied timescale is shown in Fig. 9 depending on the number of clusters used and the parameter τ. It is apparent that the ITS converges after $\tau \approx 20$ ps, i.e., the discretization error of the MSM requires a minimum lagtime of about 20 ps to decay. This convergence behavior depends on the spatial resolution of the partition used: convergence is faster when more clusters are used, thus being able to better approximate the transfer operator eigenfunctions. The standard MSM allows the timescale of the slowest process to be estimated, this being $\approx$19 ps based on the apparent convergence. We should, however, be aware that this estimate remains somewhat uncertain since it depends on where we agree to observe convergence. Statistical error estimates for the ITS were performed using the Bayesian error estimation algorithm described in Ref. 51, being on the order of the line thickness in Fig. 9. Thus, the statistical error is irrelevant in the current dataset, and significant deviations of the ITS estimates at small τ are due to discretization errors.

The full-partition MSM also provides an estimate of the transition rates between the two states if we consider the slowest transition process to be defined by the switching process

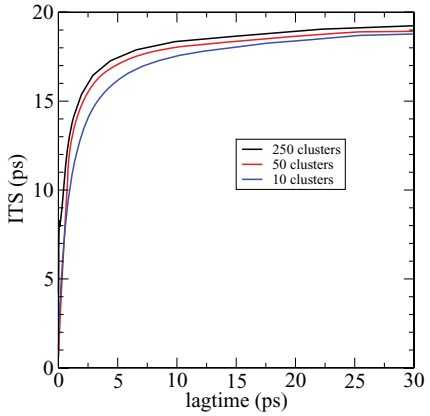

FIG. 9. Implied timescale of the process between the α and β conformations based on standard Markov state models as described in the text.

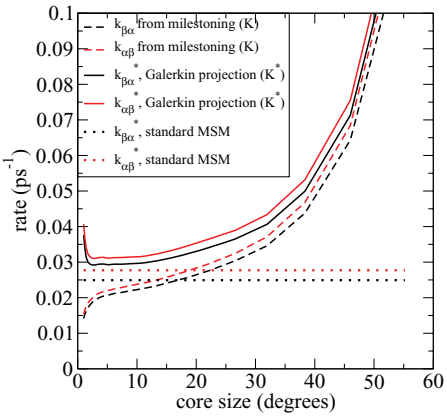

FIG. 11. Estimate of the implied timescales from the core-generator, the Galerkin projection, and the standard Markov state model.

between the α and β regions and using

$$k^* = k^*_{\alpha\beta} + k^*_{\beta\alpha},$$
$$\pi_\alpha k^*_{\alpha\beta} = \pi_\beta k^*_{\beta\alpha}, \tag{45}$$

$\pi_\alpha = 0.4735$ and $\pi_\beta = 0.5265$ were estimated from the stationary density of P where the β region was defined *via* the density minima to be $\psi \in [-130, 40]$ and the α region to be the rest. Using $k^* = 1/19$ ps, this provides the estimates:

$$k^*_{\alpha\beta} \approx 0.0277,$$
$$k^*_{\beta\alpha} \approx 0.0249, \tag{46}$$

that are shown in Fig. 11.

In order to obtain an estimation using a core MSM based on milestoning, two core centers were defined at angular coordinates $x_\alpha = (-80, -60)$ and $x_\beta = (-80, 170)$. Circular cores with a radius r were defined around these centers, and the milestoning MSMs were computed for different core sizes r. The results are shown in Fig. 10, where the 19-ps estimate

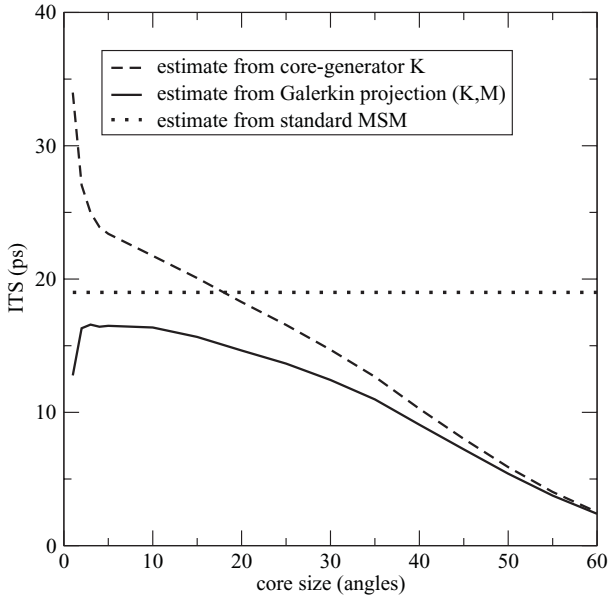

FIG. 10. Estimate of the implied timescales from the core-generator, the Galerkin projection, and the standard Markov state model.

from the full partitioning MSM is drawn as a dashed line, to indicate the reference solution. Both the straightforward core generator (K) and the Galerkin projection (K, M) generators strongly underestimate the ITS at large core sizes. This observation correspond to the observation that standard MSMs underestimate the ITS at small lagtimes: Large cores are unable to approximate the eigenfunctions corresponding to the ITS estimated, thus producing a discretization error that leads to underestimated ITS. The core generator K then has a roughly linearly increasing estimate of the ITS until the ITS is much too large for very small cores. This overestimation is due to the fact that at very small core sizes the core generator estimate misses some trajectories which have actually entered a given basin, but leave this basin before hitting the small core. For the core generator estimate there is no apparent indicator that would help to identify the core size that provides a correct estimate of the timescale. On the other hand, the Galerkin projection method does converge towards the MSM estimate of the ITS for small core sizes. Only at a core size of 1° also this estimate breaks down, presumably due to statistical reasons. The transition rates $k^*_{\alpha\beta}$ and $k^*_{\beta\alpha}$ are also directly obtained using the core generator and Galerkin projection approaches. Correspondingly to the ITS behavior, the rates are overestimated with large core sizes in both approaches, see Fig. 11. For small core sizes, the Galerkin projection converges towards a robust estimate close to the one of the full partition MSM, while the core MSM based on milrestoning does not.

These results indicate that the Galerkin projection method used with the core MSM based on milestoning is able to provide a robust estimate of the true ITS and rates of the system (A) without requiring a coarse time resolution as the full-partition MSM does and (B) based on just 2 sets instead of a full partition of state space.

This peptide example is still particularly well suited for the full-partition MSM approaches since our *a priori* knowledge about the importance of the two peptide angles allows an arbitrarily fine partition of the two-dimensional peptide angle sub-manifold. On the one hand this permits to approximate the rates rather precisely using a full-partition MSM and thus evaluate the accuracy of the core set MSM rates in comparison. On the other hand the reader should not forget that for a

more complicated molecular system such a low-dimensional essential sub-manifold may not be given or known *a priori* and thus any achievable partition will be in danger of being too coarse (especially in the transition regions). In such a case the core set MSM will have an advantage which cannot be underestimated: it "only" requires to find the centers of the relevant metastable sets for use as core sets and partition of the transition regions is not needed.

VI. CONCLUSION

The framrework of milestoning allows us to introduce a new type of Markov state model. This new class of MSM differs from standard MSM in three main aspects:

1. Instead of a finite partition of state space we only need some disjoint core sets which should be the cores of metastable sets of the process under consideration; the core sets can be pretty small neighborhoods around the energy minima in the most pronounced wells in the energy landscape.
2. We do not need to choose a lag time but can compute the generator characterizing the MSM directly from the MD timeseries; therefore core set MSMs do not introduce a lower bound on the resolvable timescales (which in case of standard MSM is given by the lag time).
3. The *a posteriori* estimators for the sampling error caused by the finiteness of the timeseries can be explicitly evaluated entrywise with core MSMs based on milestoning and it does not require sampling of multivariate, constrained distributions like for full-partition MSMs.

Here, we demonstrated how to compute maximum likelihood estimates of the generator of core MSM from the MD timeseries of finite length, constructed explicit expressions for this generator in the limit of infinitely long time series via TPT, and showed how to project the eigenvalue problem of the generator of the original process via a Galerkin ansatz. Numerical experiments on a model system with extended diffusive transition region and on a small peptide illustrated that the core MSMs based on milestoning allow to approximate the slowest timescales of the original process well, especially when based on the Galerkin projection ansatz. The results in Refs. 35, 46, and 50 form a solid basis for our approach to core MSMs in the sense that they show that the discretization error (core MSM compared to original dynamics on longest timescales) will be small for optimal core sets. The key questions in the context of core set MSMs thus is how to choose the core sets optimally: this question will be the topic of future investigations.

ACKNOWLEDGMENTS

We thank Luca Maragliano for providing us with the MD data used in Sec. V. Partial support by NSF grants DMS-0718172 and DMS-0708140, and ONR grant N00014-04-1-6046 is also acknowledged. CS, FN, and MS acknowledge support by the DFG research center MATHEON.

APPENDIX A: GENERATORS OF MARKOV PROCESSES

Here we give some details about generators for Markov processes.

Let us consider overdamped diffusion processes first. The general equation of motion for an overdamped diffusion in energy landscape $V = V(\boldsymbol{x})$ reads

$$\gamma \dot{\boldsymbol{x}}(t) = -\nabla V(\boldsymbol{x}(t)) + \sqrt{2\beta^{-1}\gamma}\,\boldsymbol{\eta}(t). \tag{A1}$$

Here γ is the friction coefficient, and $\beta = 1/k_B T$ where k_B is Botzmann's constant and T the system temperature. Finally $\boldsymbol{\eta}(t)$ is a white-noise process, i.e., a Gaussian process with mean zero and covariance $\langle \boldsymbol{\eta}(t)\boldsymbol{\eta}^T(t')\rangle = \delta(t-t')\,\mathrm{Id}$ where Id denotes the identity matrix. The associated invariant measure is

$$\mu(\boldsymbol{x}) = Z^{-1} e^{-\beta V(\boldsymbol{x})} \quad \text{where } Z = \int_\Omega e^{-\beta V(\boldsymbol{x})} d\boldsymbol{x}, \tag{A2}$$

and the corresponding infinitesimal generator

$$L = -\nabla V(\boldsymbol{x}) \cdot \nabla + \beta^{-1}\Delta. \tag{A3}$$

Here ∇ denotes the gradient operator with respect to $\boldsymbol{x}$ and Δ denotes the associated Laplacian.

The equations of motion of a system governed by Langevin dynamics are for $\boldsymbol{x} = (\boldsymbol{r}, \boldsymbol{v})$ and given by (using mass-weighted coordinates)

$$\begin{cases} \dot{\boldsymbol{r}}(t) = \boldsymbol{v}(t), \\ \dot{\boldsymbol{v}}(t) = -\nabla V(\boldsymbol{r}(t)) - \gamma \boldsymbol{v}(t) + \sqrt{2\beta^{-1}\gamma}\,\boldsymbol{\eta}(t), \end{cases} \tag{A4}$$

where $\boldsymbol{r}(t)$ and $\boldsymbol{v}(t)$ denotes positions and momenta, respectively, while V, γ, β, and η are as above in the case of overdamped diffusion. For simplicity we assumed the molecular mass matrix to be the identity, which can always be achieved by using mass-weighted coordinates. The equilibrium probability density associated with Eq. (A4) is

$$\mu(\boldsymbol{x}) = \frac{1}{Z} \exp\Big(-\beta\Big(V(\boldsymbol{r}) + \frac{1}{2}\boldsymbol{v}^T\boldsymbol{v}\Big)\Big),$$

where Z is a normalization constant. The generator associated with Eq. (A4) is

$$L = -\gamma \boldsymbol{v} \cdot \nabla_{\boldsymbol{v}} - \boldsymbol{v} \cdot \nabla_{\boldsymbol{r}} + \nabla_{\boldsymbol{r}} V(\boldsymbol{r}) \cdot \nabla_{\boldsymbol{v}} + \beta^{-1}\Delta_{\boldsymbol{v}}.$$

In case $\boldsymbol{x}(t)$ is a Markov jump process on discrete state space $\{1, \ldots, N\}$ we consider the time-t transition kernel between its discrete states,

$$p(t, i, j) = \mathbb{P}(\boldsymbol{x}(t) = j | \boldsymbol{x}(0) = i),$$

and get the generator L with entries

$$l_{i,j} = \lim_{t \to 0+} \frac{1}{t}(p(t, i, j) - \delta_{ij}),$$

where δ_{ij} denotes the Kronecker symbol. The transition kernel $\mathcal{P}_t$, i.e., the stochastic matrix with entries $p(t, i, j)$, relates to the generator as

$$\mathcal{P}_t = \exp(tL),$$

and the equilibrium probability distribution of the process solves $0 = \mu^T L$ or $\mu^T = \mu^T \mathcal{P}_t$.

APPENDIX B: COMPUTATION OF $v_{i,j}$

For simplicity let us focus on the overdamped case first and come back to the general situation at the end of this subsection.

Assume that the committor functions q_i from Eq. (19) are given, and μ_i and π_i have been computed from Eqs. (20) and (21). The starting point for obtaining the $v_{i,j}$ is the following expression for the probability current of reactive trajectories (these are all parts of an infinitely long trajectory that go from B_i to $\cup_{j\neq i} B_j$ directly, i.e., without going back to B_i before entering $\cup_{j\neq i} B_j$)

$$j_i(\boldsymbol{x}) = \mu(\boldsymbol{x})\nabla q_i(\boldsymbol{x}). \tag{B1}$$

The integral of this current through the boundary of set $B_j \neq B_i$ gives the net probability flux out of this set in the reaction from $\cup_{j\neq i} B_j$ to B_i. By invariance under time reversal this is also the net flux into $B_j \neq B_i$ in the reaction from B_i to $\cup_{j\neq i} B_j$. In other words

$$v_{i,j} = \int_{\partial B_j} \mu(\boldsymbol{x})\hat{\boldsymbol{n}}_j(\boldsymbol{x})\cdot\nabla q_i(\boldsymbol{x}) d\sigma_j(\boldsymbol{x}), \quad i \neq j, \tag{B2}$$

where ∂B_j denotes the boundary of B_j, $\hat{\boldsymbol{n}}_j(\boldsymbol{x})$ the unit normal pointing out of ∂B_j, and $d\sigma_j(\boldsymbol{x})$ is the surface element on ∂B_j. It is easy to see that Eq. (B2) can be expressed as

$$v_{i,j} = -\int_\Omega \mu(\boldsymbol{x})\nabla q_i(\boldsymbol{x})\cdot\nabla q_j(\boldsymbol{x}) d\boldsymbol{x}, \quad i \neq j \tag{B3}$$

Indeed Eq. (B2) is what remains if one integrates Eq. (B3) by parts and uses Eq. (19). Equation(B3) shows that $v_{i,j} \geq 0$ as needed. Another integration by parts indicates that Eq. (B3) can also be expressed as

$$v_{i,j} = \int_\Omega \mu(\boldsymbol{x}) q_i(\boldsymbol{x})(Lq_j)(\boldsymbol{x}) d\boldsymbol{x},$$

as we have outlined above.

This derivation can be transfered to other dynamics, too. For example, if the evolution is governed by the Langevin equation, then the positions $\boldsymbol{x}$ must simply be replaced by the set of positions and velocities, $(\boldsymbol{x}, \boldsymbol{v})$. For Markov jump processes in discrete state space see Ref. 46.

APPENDIX C: PROJECTION OPERATOR

The projection operator P maps a function f to its best approximation Pf in

$$\mathbb{S} = \left\{ f : \Omega \to \mathbb{R} : \ f = \sum_{i=1}^N \alpha_i q_i, \ \alpha_i \in \mathbb{R} \right\},$$

and is thus defined by $Pf = u \in \mathbb{S}$ with

$$\|u - f\| = \min_{v\in\mathbb{S}} \|v - f\|, \tag{C1}$$

where the norm is defined via the scalar product, $\|g\|^2 = \langle g, g\rangle_\mu$. For some real-valued scalar s and an arbitrary $v \in \mathbb{S}$ we get from Eq. (C1) that $Pf = u$ satisfies

$$\begin{aligned}\|u - f\|^2 &\leq \|(u + sv) - f\|^2 \\ &= \|u - f\|^2 + 2s\langle v, u - f\rangle_\mu + s^2\|v\|^2,\end{aligned}$$

which, for $s > 0$, reduces to $2\langle v, u - f\rangle_\mu + s\|v\|^2 \geq 0$ and in the limit $s \to 0$ yields $\langle v, u - f\rangle_\mu \geq 0$. Since we get $\langle v, u - f\rangle_\mu \leq 0$ for $s < 0$ and s is arbitrary it must be

$$\langle v, u - f\rangle_\mu = 0.$$

Since v also has been arbitrary, the last identity holds for all $v \in \mathbb{S}$, and thus

$$\langle q_j, f - Pf\rangle_\mu = 0, \quad \forall j = 1, \ldots, N.$$

Since $Pf \in \mathbb{S}$, we have $Pf = \sum_i \alpha_i q_i$ for some coefficients α_i. Using the orthogonality of the error, we get

$$\sum_i \alpha_i \langle q_j, q_i\rangle_\mu = \langle q_j, f\rangle_\mu,$$

which can be written as the system of linear equations $\sum_i \alpha_i s_{ij} = \langle q_j, f\rangle_\mu$ when introducing the matrix $S = (s_{ij})$ with $s_{ij} = \langle q_j, q_i\rangle_\mu$. Formal solution of this system of linear equations yields

$$(Pf)(\boldsymbol{x}) = \sum_{i,j=1}^N q_i(\boldsymbol{x})(S^{-1})_{i,j}\langle q_j, f\rangle_\mu.$$

APPENDIX D: COMPARISON WITH STANDARD MARKOV STATE MODELS

In this section, we recall the standard procedure used to build Markov state models and stress the differences with milestoning.

1. Set up

In contrast with milestoning, a standard MSM is typically based on a complete subdivision of state space Ω into disjoint sets $A_1, \ldots, A_N$ such that $\Omega = \cup_j A_j$. The index process $i(t)$ associated with these sets can be defined as before, which now results in the simpler relation

$$i(t) = j \quad \text{if } \boldsymbol{x}(t) \in A_j. \tag{D1}$$

A second difference with milestoning is that standard MSMs analyze the discrete-time process $i(k\tau)$, $\tau > 0$, $k = 0, 1, 2, \ldots$ instead of its continuous-time version. The key assumption made is that for appropriate choices of the lag time τ the process $i(k\tau)$ can be modeled by a discrete-time Markov process. The validity of this assumption depends on the choice of both the sets A_j and the lag time τ.

Assuming Markovianity, the evolution of the process $i(k\tau)$ is completely specified by a set of transition probabilities $p_{i,j} \geq 0$ which depends on τ and give the probabilities that $i((k+1)\tau) = j$ given that $i(k\tau) = i$ (when $i = j$, $p_{i,i}$ gives the probability of staying in state i after one step of length τ). This implies in particular that, given the initial state $i(0) = j_0$ and the transition probabilities $p_{i,j}$, the probability that the process $i(k\tau)$ evolves along the (discrete) trajectory $j_1, \ldots, j_n$ is (compare Eq. (4))

$$\mathbb{P}(\text{path}|\text{prob}) = p_{j_0,j_1}\cdots p_{j_{n-1},j_n}, \tag{D2}$$

2. Bayesian formalism, MLE, and error estimates

Formula (D2) gives the probability to observe a path given the transition probabilities $p_{i,j}$. As before, what is available from MD data is the path $i(k\tau)$, $k = 0, 1, 2, \ldots$, and not the transition probabilities $p_{i,j}$ and to estimate $p_{i,j}$ we need their probability given the path. Combining Bayes formula with Eq. (D2), we get

$$\mathbb{P}(\text{prob}|\text{path}) = C\,\mathbb{P}(\text{prob}) \prod_{i,j=1}^{N} p_{i,j}^{N_{i,j}^T}, \tag{D3}$$

where C is a normalization constant, $\mathbb{P}(\text{prob})$ is the prior in the set of all possible transition probabilities, and $N_{i,j}^T$ is the number of transitions from state i to state j observed along $i(k\tau)$ during the time interval $[0, T]$ ($N_{i,i}^T$ counts the number of events such that a $i(k\tau) = i$ and $i((k+1)\tau) = i$).

The maximum likelihood estimate for the transition probabilities $p_{i,j}$ can be computed analytically by maximizing Eq. (D3) over all $p_{i,j}$ subject to the constraints that $p_{i,j} \geq 0$ and $\sum_j p_{i,j} = 1$:

$$p_{i,j}^* = \frac{N_{i,j}^T}{N_i^T}, \tag{D4}$$

where N_i^T is the number of visits of the process $i(k\tau)$ in state i, i.e., the number of discrete times with $i(k\tau) = i$ in $[0, T]$. In order to access the statistical uncertainty of the MLE (D4) one has to construct an ensemble of transition matrices that is distributed according to $\mathbb{P}(\text{prob}|\text{path})$. Because of the constraints $p_{i,j} \geq 0$ and $\sum_j p_{i,j} = 1$, there is no simple formula that permits to compute this uncertainty entrywise (as is the case for the rates in the milestoning, see Eq. (8)). However, there are sampling algorithms that generate an ensemble of transition matrices distributed according to Eq. (D3).

3. Exact formulas

We can again ask what is the exact representation formula for the transition probabilities $p_{i,j}$. It is simply:

$$p_{i,j} = \frac{1}{\mu_{A_i}} \int_{A_i} d\boldsymbol{x} \int_{A_j} d\boldsymbol{y}\, \mu(\boldsymbol{x}) p(\boldsymbol{x}, \boldsymbol{y}; \tau), \tag{D5}$$

where $\mu(\boldsymbol{x})$ is the equilibrium probability density of the process $\boldsymbol{x}(t)$, $\mu_{A_i} = \int_{A_i} \mu(\boldsymbol{x}) d\boldsymbol{x}$, and $p(\boldsymbol{x}, \boldsymbol{y}; \tau)$ denotes the transition probability density already introduced in Eq. (13). Denoting by $P_\tau = \exp(\tau L)$ the transition kernel at lag-time τ, we can express Eq. (D5) as

$$p_{i,j} = \frac{\langle \chi_i, P_\tau \chi_j \rangle_\mu}{\langle \chi_i, \chi_i \rangle_\mu}, \tag{D6}$$

where χ_i is the indicator function of the set A_i: $\chi_i(\boldsymbol{x}) = 1$ if $\boldsymbol{x} \in A_i$, $\chi_i(\boldsymbol{x}) = 0$ otherwise.

4. Galerkin projection interpretation

A standard MSM can be also understood as a Galerkin approximation using the space spanned by $\chi_1, \ldots, \chi_N$. The eigenvalue problem associated with the transition matrix $\mathcal{P}_\tau$ is (compare Eq. (29)):

$$P_\tau \varphi^e = \mu^e \varphi^e, \tag{D7}$$

where $\mu^e = e^{\lambda^e \tau}$. The projected version of this equation is (compare Eq. (30))

$$P_\chi P_\tau P_\chi \varphi = \mu \varphi, \tag{D8}$$

where the projection operator P_χ is the equivalent of Eq. (27) with q_i replaced by χ_i

$$(P_\chi f)(\boldsymbol{x}) = \sum_{i=1}^{N} \xi_i(\boldsymbol{x}) \langle \xi_i, f \rangle_\mu. \tag{D9}$$

After a little algebra, it is easy to see that Eq. (D8) can be written as

$$\sum_{j=1}^{N} p_{i,j} r_j^\chi = \mu r_i^\chi, \tag{D10}$$

where r_i^χ is the equivalent of Eq. (33) with q_i replaced by χ_i

$$r_i^\chi = \frac{\langle \chi_i, \varphi \rangle_\mu}{\langle \chi_i, \chi_i \rangle_\mu}. \tag{D11}$$

Compared with a Markov state model based on milestoning, we see that the mass matrix $m_{i,j}$ is simply the identity matrix since the functions χ_i are orthogonal with each other and the weights $\langle \chi_i, \chi_i \rangle_\mu$ are included in the $p_{i,j}$.

Finally, let us comment on the approximation error in standard MSMs. The relevant error measure the deviation of the probability transport described by the MLE $p_{i,j}^*$, that has been computed based on a finite trajectory in $[0, T]$ via Eq. (D4), and the probability transport of the underlying process $\boldsymbol{x}(t)$. This error can be decomposed into the deviation between the probability transport of $p_{i,j}^*$ from the transport given by the exact representation $p_{i,j}$ from Eq. (D5), and the deviation between $p_{i,j}$ and the original probability transport of $\boldsymbol{x}(t)$. The former is the *statistical error* which depends on the length of the finite trajectory and for which we have an *a posteriori* estimator via the likelihood as described above. The latter is the *discretization error* that depends on the choice of the sets $A_1, \ldots, A_N$ (spatial discretization error) as well as the lag time τ (temporal discretization error). In Ref. 50 an estimate for the discretization error is given which shows that it can be made small if the lag time is large enough *and* the sets are chosen appropriately. In particular this result shows that for large enough lag times, only the spatial discretization error remains while the temporal error vanishes.

[1] M. Jäger, Y. Zhang, J. Bieschke, H. Nguyen, M. Dendle, M. E. Bowman, J. P. Noel, M. Gruebele, and J. W. Kelly, Proc. Natl. Acad. Sci. U.S.A. **103**, 10648 (2006).
[2] A. Y. Kobitski, A. Nierth, M. Helm, A. Jäschke, and G. U. Nienhaus, Nucleic Acids Res. **35**, 2047 (2007).
[3] S. Fischer, B. Windshuegel, D. Horak, K. C. Holmes, and J. C. Smith, Proc. Natl. Acad. Sci. U.S.A. **102**, 6873 (2005).
[4] F. Noé, D. Krachtus, J. C. Smith, and S. Fischer, J. Chem. Theo. Comp. **2**, 840 (2006).
[5] A. Ostermann, R. Waschipky, F. G. Parak, and U. G. Nienhaus, Nature (London) **404**, 205 (2000).
[6] D. D. Schaeffer, A. Fersht, and V. Daggett, Curr. Opin. Struct. Biol. **18**, 4 (2008).
[7] F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl, Proc. Natl. Acad. Sci. U.S.A. **106**, 19011 (2009).

[8]W. van Gunsteren, J. Dolenc, and A. Mark, Curr. Opin. Struct. Biol. **18**, 149 (2008).
[9]V. A. Voelz, G. R. Bowman, K. Beauchamp, and V. S. Pande, J. Am. Chem. Soc. **132**, 1526 (2010).
[10]D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers, Science **330**, 341 (2010).
[11]J. E. Stone, J. C. Phillips, P. L. Freddolino, D. J. Hardy, L. G. Trabuco, and K. Schulten, J. Comput. Chem. **28**, 2618 (2007).
[12]C. Schuette, *Conformational dynamics: modelling, theory, algorithm, and applications to biomolecules*, Habilitation thesis (Fachbereich Mathematik und Informatik, FU Berlin, 1998).
[13]J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Fischbach, M. Held, J. D. Chodera, Ch. Schütte, and F. Noé, J. Chem. Phys. **134**, 174105 (2011).
[14]D. J. Wales, *Energy Landscapes* (Cambridge University Press, Cambridge, 2003).
[15]F. Noé and S. Fischer, Curr. Opin. Struc. Biol. **18**, 154 (2008).
[16]M. E. Karpen, D. J. Tobias, and C. L. Brooks, Biochemistry **32**, 412 (1993).
[17]I. A. Hubner, E. J. Deeds, and E. I. Shakhnovich, Proc. Natl. Acad. Sci. U.S.A. **103**, 17747 (2006).
[18]M. Weber, ZIB Report 03-04, 2003.
[19]N. V. Buchete and G. Hummer, J. Phys. Chem. B **112**, 6057 (2008).
[20]F. Rao and A. Caflisch, J. Mol. Biol. **342**, 299 (2004).
[21]S. Muff and A. Caflisch, Proteins **70**, 1185 (2007).
[22]B. de Groot, X. Daura, A. Mark, and H. Grubmüller, J. Mol. Biol. **301**, 299 (2001).
[23]V. Schultheis, T. Hirschberger, H. Carstens, and P. Tavan, J. Chem. Theory Comput. **1**, 515 (2005).
[24]A. C. Pan and B. Roux, J. Chem. Phys. **129**, 064107 (2008).
[25]C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard, J. Comput. Phys. **151**, 146 (1999).
[26]S. V. Krivov and M. Karplus, Proc. Nat. Acad. Sci. U.S.A. **101**, 14766 (2004).
[27]F. Noé, I. Horenko, C. Schütte, and J. C. Smith, J. Chem. Phys. **126**, 155102 (2007).
[28]J. D. Chodera, K. A. Dill, N. Singhal, V. S. Pande, W. C. Swope, and J. W. Pitera, J. Chem. Phys. **126**, 155101 (2007).
[29]W. C. Swope, J. W. Pitera, and F. Suits, J. Phys. Chem. B **108**, 6571 (2004).
[30]P. Deuflhard, W. Huisinga, A. Fischer, and C. Schuette, Linear Algebra Appl. **315**, 39 (2000).
[31]C. Schuette and W. Huisinga, in *Handbook of Numerical Analysis*, Vol. 10 (Elsevier, Amsterdam, 2003), pp. 699–744.
[32]P. Deuflhard and M. Weber, Linear Algebra Appl. **398**, 161 (2005) [Special issue on matrices and mathematical biology].
[33]W. Huisinga, S. Meyn, and C. Schuette, Ann. Appl. Probab. **14**, 419 (2004).
[34]M. Sarich, F. Noé, and C. Schütte, SIAM Multiscale Model. Simul. **8**, 1154 (2010).
[35]N. Djurdjevac, M. Sarich, and Ch. Schütte, "Estimating the eigenvalue error of Markov State Models," Multiscale Model. Simul. (to be published).
[36]A. K. Faradjian and R. Elber, J. Chem. Phys. **120**, 10880 (2004).
[37]D. Shalloway and A. K. Faradijan, J. Chem. Phys. **124**, 054112 (2006).
[38]R. Elber, Biophys. J. **92**, L85 (2007).
[39]A. M. A. West, R. Elber, and D. Shalloway, J. Chem. Phys. **126**, 145104 (2007).
[40]E. Vanden-Eijnden and M. Venturoli, J. Chem. Phys. **130**, 194101 (2009).
[41]W. E and E. Vanden-Eijnden, J. Stat. Phys. **123**, 503 (2006).
[42]E. Vanden-Eijnden, in *Computer Simulations in Condensed Matter: From Materials to Chemical Biology*, edited by M. Ferrario, G. Ciccotti, and K. Binder (Springer, Berlin, 2006), Vol. 1, pp. 439–478.
[43]P. Metzner, C. Schütte, and E. Vanden-Eijnden, J. Chem. Phys. **125**, 084110 (2006).
[44]P. Metzner, C. Schütte, and E. Vanden-Eijnden, Multiscale Model. Simul. **7**, 1192 (2009).
[45]W. E and E. Vanden-Eijnden, Annu. Rev. Phys. Chem. **61**, 391 (2010).
[46]N. Djurdjevac, M. Sarich, and C. Schütte, "On Markov state models for metastable processes," in *Proceeding of the ICM* 2010.
[47]D. Chandler, J. Chem. Phys. **68**, 2959 (1978).
[48]N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, 4th ed. (Elsevier, Amsterdam, 2006).
[49]E. Vanden-Eijnden and F. Tal, J. Chem. Phys. **123**, 184103 (2005).
[50]M. Sarich, F. Noé, and C. Schuette, Multiscale Model. Simul. **8**, 1154 (2010).
[51]F. Noé, J. Chem. Phys. **128**, 244103 (2008).