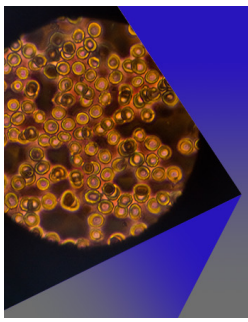
RESEARCH ARTICLE | SEPTEMBER 29 2023

Arbitrarily accurate, nonparametric coarse graining with Markov renewal processes and the Mori–Zwanzig formulation

David Aristoff ; Mats Johnson ; Danny Perez

AIP Advances 13, 095131 (2023)
https://doi.org/10.1063/5.0162440



Arbitrarily accurate, nonparametric coarse graining with Markov renewal processes and the Mori–Zwanzig formulation

Cite as: AIP Advances 13, 095131 (2023); doi: 10.1063/5.0162440
Submitted: 31 July 2023 • Accepted: 7 September 2023 •
Published Online: 29 September 2023

David Aristoff,[1,a)] Mats Johnson,[1] and Danny Perez[2]

AFFILIATIONS

[1] Colorado State University, Fort Collins, Colorado 80523, USA
[2] Theoretical Division T-1, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

[a)] Author to whom correspondence should be addressed: aristoff@rams.colostate.edu

ABSTRACT

Stochastic dynamics, such as molecular dynamics, are important in many scientific applications. However, summarizing and analyzing the results of such simulations is often challenging due to the high dimension in which simulations are carried out and, consequently, due to the very large amount of data that are typically generated. Coarse graining is a popular technique for addressing this problem by providing compact and expressive representations. Coarse graining, however, potentially comes at the cost of accuracy, as dynamical information is, in general, lost when projecting the problem in a lower-dimensional space. This article shows how to eliminate coarse-graining error using two key ideas. First, we represent coarse-grained dynamics as a Markov renewal process. Second, we outline a data-driven, non-parametric Mori–Zwanzig approach for computing jump times of the renewal process. Numerical tests on a small protein illustrate the method.

I. INTRODUCTION

Stochastic dynamics play a critical role in the study of complex systems across various scientific domains. Molecular dynamics (MD), for instance, simulate the motion of collections of atoms over time. MD simulations have found applications in materials science, chemistry, biology, and physics.[1–6] Analyzing the immense volume of data generated, and navigating the high-dimensional space in which these simulations operate, creates significant challenges. Indeed, a single snapshot of an MD trajectory resides in a continuous $3N_{\text{atom}}$-dimensional space, with N_{atom} ranging from hundreds to billions. This makes model reduction highly desirable. Effective model reduction not only enhances interpretability but also allows for upscaling results to inform higher-fidelity models.

Another challenge comes from metastability,[7] where stochastic trajectories are confined to small regions of space for long times, punctuated by rare but fast transitions between regions. Metastability is typical in MD, where such regions might represent folded and unfolded states of a protein. In addition to MD, metastable stochastic dynamics arise in climate models,[8–11] granular flows,[12] neural evolution,[13–18] hydrodynamics,[19] and power networks,[20] in addition to various ordinary differential equations models.[21–24]

Coarse graining is a common approach for handling high dimensionality or metastability. It is based on dividing the original high-dimensional space of *microstates* into a discrete set of *macrostates*. Usually, the dynamics on the macrostates is modeled as a Continuous Time Markov Chain (CTMC)[25] or a discrete time Markov chain (DTMC).[25,26] In MD, such CTMC models are called chemical reaction networks or kinetic Monte Carlo models,[27,28] and the DTMCs are called Markov state models.[29] These Markovian models offer advantages such as formal simplicity, compact representation, and ease of use with ready-made algorithms, such as BKL[30] or Gillespie,[31] for simulation.

The Markov assumption underlying CTMC and DTMC models is significantly flawed if macrostates are not carefully chosen,[32] if temperatures are not sufficiently low,[33] or if time scales are not long enough. Even with careful choices of all these parameters, some degree of departure from exact Markovian behavior generally remains.[34–37]

Meanwhile, recent findings show that the Markov assumption can be weakened, with an arbitrarily accurate representation achievable using Markov Renewal Processes[38] (MRPs) by simply adjusting a scalar parameter.[39] This scalar parameter, called τ below, is a *decorrelation time* chosen to allow the underlying dynamics to reach local equilibrium in the macrostates, inheriting the Markov property at each such time. MRPs differ from Markov processes in only having the Markov property at certain times (called *jump times below*). Despite having formal simplicity, a complete MRP parameterization for N macrostates would require N^2 scalars and N^2 functions of time. Accurately representing these functions from limited, short-time length data poses a challenge.[39] This article proposes a new technique, rooted in first principles, to efficiently model this MRP with a few $N \times N$ matrices.

A. Contributions

Below, we propose a compact, data-driven parameterization for the MRP model described in Ref. 39. Our methods, rooted in Mori–Zwanzig (MZ) theory, are simple and data-driven, and our contributions are practical and theoretical.

On the practical side, we propose a mathematically principled representation of the MRP derived from Mori–Zwanzig theory. In our formulation, the MRP is represented by, and can be generated from, a (small) number of memory kernels. These memory kernels are $N \times N$ matrices, where N is the number of macrostates. We propose a new method to obtain the kernels by solving a certain linear system comprised of correlation matrices. Efficient, scalable solvers designed for positive semidefinite systems can then be used to obtain the kernels (e.g., RPCholesky[40,41] uses randomized low-rank approximation). Numerical results on alanine dipeptide, a small protein, illustrate the promise of the method.

On the theoretical side, we show that these methods become exact as the number of memory kernels and the decorrelation time grow. This demonstration takes the following steps. To start, we give the first proof that coarse-grained dynamics described in Ref. 39 converges to a MRP (Theorem A.1). Then, we represent the transition probabilities of the MRP in terms of memory kernels using the discrete Mori–Zwanzig equation (3). Lastly, we prove that Eq. (3) is exact (Theorem D.2). This equation first appeared in Ref. 42 in a different setting (without the decorrelation). There, it was derived as an approximation of a continuous time Mori–Zwanzig equation. We give the first full derivation of (3) that shows it is exact for any choice of dynamical lag (we use lag τ in our setup). As τ can be significantly longer than the time step of the underlying dynamical integrator, exactness at the discrete time level is important.

In addition, we provide exact expressions for the memory kernels in terms of an orthogonal dynamics (Appendix D). While these expressions cannot directly be put to practical use, they help lend explainability to the kernels and could potentially be exploited to quantify their decay in time. Our novel data-driven method for actually computing the memory kernels, based on the linear solve (5), can also be explained in terms of inter-macrostate correlations.

Finally, we show that our Mori–Zwanzig equation is optimal, in the sense that the representation is compact when the MRP representation is almost fully Markovian. We actually prove an ideal case of this, showing that all but one of the memory kernels vanishes in the case where the MRP representation is, in fact, Markovian.

TABLE I. Definitions of symbols used in this work.

Symbol	Definition
$X(t)$	Underlying Markov chain on microstates
x, y, z	Microstates
I, J, L	Macrostates
N	Number of macrostates
τ	Macroscopic time step
$R(t)$	Macroscopic jump process
r, s, t	Times [multiples of τ, when associated with $R(t)$]
s_-, t_-	Preceding times: $s_- = s - \tau$, $t_- = t - \tau$
τ_I	Decorrelation time in macrostate I
η_I	QSD in macrostate I
$\mathcal{T}(s,t)$	Transition probability matrix
$\mathcal{T}(t)$	Transition matrix of renewal process
$\mathcal{P}(t)$	Jump probability matrix
$\mathcal{K}(t)$	Memory kernel matrix
$C(t)$	Consecutive time in current macrostate
P, Q	Projector and complementary projector
χ_I	Characteristic function of macrostate I
n, m	Non-negative integers

This article is organized as follows. We summarize our notation in Table I. In Sec. II, we review how we discretize the underlying dynamics, following Ref. 39. In Sec. III, we introduce the Mori–Zwanzig equation and explain how we use it to estimate memory kernels nonparametrically from short time simulations. We also show how the memory kernels can be used to infer longer time information. In Sec. IV, we give an outline of our proof that the discretized dynamics converges to a MRP (the proof is in Appendix A). In Sec. V, we illustrate our method on alanine dipeptide. We show that we can reduce errors arising from ordinary spatial discretization, recovering accurate dynamics with a relatively small number of memory kernels. All proofs, including the derivation of the Mori–Zwanzig equation and the proof of convergence to a MRP, are in Appendixes A–D.

II. MARKOV CHAINS AND MARKOV RENEWAL PROCESS

Throughout, $X(t)$ is an underlying Markov process evolving in a space of *microstates*. This process can be discrete or continuous in both time and space. We consider a division of microstates into finitely many *macrostates* I, J, etc.

Our work focuses on a discrete time jump process $R(t)$ on these macrostates, with time step τ, defined from the underlying process and a set of *decorrelation times*, written as τ_I, τ_J, etc. The jumps occur when $X(t)$ spends consecutive time τ_J in some macrostate J. Specifically, $R(t)$ jumps from I to J at time t if $X(t-c)$ is in macrostate J for $0 \le c \le \tau_J$. Jumps only occur among distinct states ($J \neq I$) and at multiples of the time step ($t = n\tau$ for integer n). See Fig. 1 for an illustration.

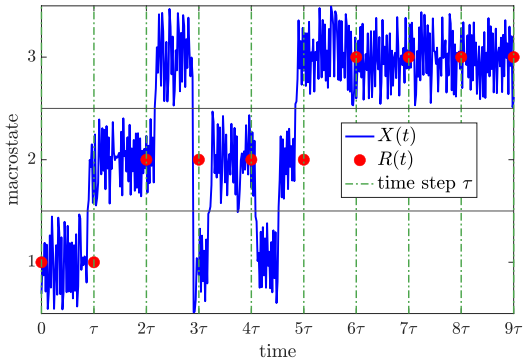


FIG. 1. Illustration of $X(t)$ and $R(t)$, with three macrostates labeled 1–3, when the decorrelation times are $\tau_1 = \tau_2 = \tau_3 = \tau$. Solid horizontal lines divide the macrostates. For illustrative purposes, we show an example where $X(t)$ makes several transitions that are not recorded by $R(t)$ due to a failure to decorrelate in macrostates.

To describe the evolution of $R(t)$, we define $\mathcal{T}_{IJ}(s,t)$ as the probability for $R(t)$ to be in J at time $s+t$, assuming that there was a jump into I at time s. That is,

$$\mathcal{T}_{IJ}(s,t) = \mathbb{P}(R(s+t) = J | R(s_-) \neq I,\, R(s) = I), \tag{1}$$

where we use the shorthand $s_- = s - \tau$.

The introduction of decorrelation times allows the underlying Markov process to reach a local equilibrium within each macrostate. Conceptually, when τ_J is large enough, $X(t)$ loses memory of how it entered J by the time that $R(t)$ jumps into macrostate J. This makes $R(t)$ into a MRP, which means it has the Markov property at jump times.[39] Note that $R(t)$ does not retain information about what occurs on timescales shorter than the decorrelation times and τ. This is a modeling assumption that may lead to the loss of relevant dynamical information if important transition events occur on such timescales. On the other hand, information loss will be minimal when the typical residence time in a macrostate is much longer than both τ and the decorrelation time.

Assuming that $R(t)$ is, in fact, a MRP, we can write $\mathcal{T}(s,t) = \mathcal{T}(t)$, where $\mathcal{T}(t)$ is a standard transition matrix for each t. These transition matrices together satisfy a *renewal equation* defined by a *jump probability matrix* $\mathcal{P}(t)$, where $\mathcal{P}_{IJ}(t)$ is the probability for $R(t)$ to jump from I to J in time t,

$$\mathcal{P}_{IJ}(t) = \mathbb{P}(R(s+t) = J | R(s_-) \neq I,\, R(s') = I,\, s \leq s' < s+t).$$

The renewal equation is[38]

$$\mathcal{T}(t) = \sum_{0<s\leq t} \mathcal{P}(s)\,\mathcal{T}(t-s) + \mathcal{F}(t), \tag{2}$$

where $\mathcal{F}_{IJ}(t) = \delta_{I=J} \sum_L \sum_{s>t} \mathcal{P}_{IL}(s)$. Here, $\delta_{I=J} = 1$ if $I = J$ and $\delta_{I=J} = 0$ otherwise. The time arguments here are multiples of τ, and we continue with this convention for other equations associated with $R(t)$ below.

The Markov renewal framework of (2) is exact in the limit of large decorrelation times (Theorem A.1). Below, we outline how to estimate $\mathcal{T}(t)$ in a principled, parameter-free way using Mori–Zwanzig theory. Once $\mathcal{T}(t)$ is estimated, Eq. (2) can be used to compute the jump time distribution $\mathcal{P}(t)$. This provides a principled way to describe—and simulate—the process $R(t)$, which exactly reflects the macroscopic behavior of $X(t)$.

Our setup above allows for situations where the decorrelation times are state-dependent: there is a (potentially different) decorrelation time τ_I for each macrostate I. For simplicity, in the numerical examples and ensuing discussion in Sec. V, we take all the decorrelation times to be the same and equal to τ, i.e., $\tau_I = \tau$ for each I.

III. NONPARAMETRIC ESTIMATION OF TRANSITION PROBABILITIES

Using Mori–Zwanzig theory,

$$\mathcal{T}(t) = \sum_{0<s\leq t} \mathcal{K}(s)\,\mathcal{T}(t-s), \tag{3}$$

where $\mathcal{K}(s)$ are *memory kernels* that can be estimated from data, as we describe below. Equation (3) is derived as an approximation of a continuous-time Mori–Zwanzig equation in Ref. 42, while different discrete time Mori–Zwanzig equations are described in Refs. 43 and 44. We will give a short proof of exactness of (3) in Appendix D (Theorem D.2) and provide more details on the memory kernel structure there.

Equations (2) and (3) appear superficially similar but are quite different. While $\mathcal{P}(s)$ defines jump probabilities of the MRP, $\mathcal{K}(s)$ involves quantities associated with a so-called *orthogonal dynamics*. Roughly speaking, this dynamics describes situations where $X(t)$ transitions between macrostates without decorrelating in them. We arrived at (3) by choosing a Mori–Zwanzig projector that leads to very compact representations (i.e., fast time decay of memory kernels) when $R(t)$ is nearly Markovian. Indeed, in Appendix D, we show that if $R(t)$ is actually Markovian, only one memory kernel is nonzero: $\mathcal{K}(s) = 0$ for $s > \tau$. Meanwhile, if $R(t)$ is Markovian, then $\mathcal{P}(s)$ is geometric in s with rates in inverse proportion to the mean jump times between macrostates [resulting in a slow decay of $\mathcal{P}(s)$ for large mean jump times].

While Eq. (3) could be used to solve for the memory kernels directly given enough sampling,[42] we find that the following setup is superior in practice. In order to nonparametrically estimate $\mathcal{K}(t)$, we introduce a loss function

$$\mathcal{L}(\mathcal{K}) = \sum_{t\leq t_{\max}} \left\| \mathcal{T}(t) - \sum_{0<s\leq \min\{t, t_{mem}\}} \mathcal{K}(s)\,\mathcal{T}(t-s) \right\|^2, \tag{4}$$

where t_{mem} is a cutoff time for the memory matrices, $t_{\max}$ is a cutoff time for the transition matrices, and $\|\cdot\|$ represents the Frobenius norm.

By setting the gradient of the loss function equal to zero, we get the following symmetric positive semidefinite linear system that can be solved for the memory matrices (see Appendix E):

$$\sum_{0<s\leq t_{mem}} \mathcal{K}(s)\,\mathcal{A}(s,t) = \mathcal{B}(t), \quad 0 < t \leq t_{mem}, \tag{5}$$

where $\mathcal{A}$ and $\mathcal{B}$ are the correlation matrices,

$$\mathcal{A}(s,t) = \sum_{r \le t_{\max}} \mathcal{T}(r-s)\mathcal{T}(r-t)^T,$$
$$\mathcal{B}(s) = \sum_{r \le t_{\max}} \mathcal{T}(r)\mathcal{T}(r-s)^T, \tag{6}$$

and where, by convention, $\mathcal{T}(s) = 0$ for $s < 0$. (Various regularizations, including ridge regression that penalizes the Frobenius norms of the memory kernels, can easily be applied if desired.)

The memory kernels $\mathcal{K}(t)$ can then be obtained as follows. First, we can estimate $\mathcal{T}(t)$ for $t \le t_{\max}$ from data of the underlying Markovian dynamics. Then, we can estimate the matrices $\mathcal{A}$ and $\mathcal{B}$ in (6). Finally, we solve the linear system (5) to obtain $\mathcal{K}(t)$ for $0 < t \le t_{mem}$.

With the memory kernels in hand, the transition probabilities can be estimated by repeatedly applying the equation

$$\mathcal{T}(t) \approx \sum_{0<s\le \min\{t,t_{mem}\}} \mathcal{K}(s)\mathcal{T}(t-s), \tag{7}$$

while incrementally increasing t. Note that this allows for estimation up to any time, including beyond $t_{\max}$. The memory kernels carry N^2k entries in total, with N being the number of macrostates and $k = t_{mem}/\tau$ being the number of memory kernels. We find good results even with a relatively small number of kernels; see Sec. V. Once $\mathcal{T}(t)$ is in hand, $\mathcal{P}(t)$ can be computed by unrolling the renewal equation (2).

In Appendix II, we show that if $R(t)$ is actually a Markov chain—that is, if it has the Markov property at *all* times, not just at jump times—then $\mathcal{K}(t) = 0$ for $t > \tau$. In this case, $\mathcal{T}(n\tau) = \mathcal{K}(\tau)^n = \mathcal{T}(\tau)^n$, and the estimation of the system only depends on the underlying Markov chain dynamics at lag τ. Equation (7) provides an extension of this to allow for non-Markovian behavior.

Other methods for estimating memory kernels have been recently described in Refs. 42, 45, 46, and 47. We find that our method significantly outperforms applying a direct solve[42] in Eq. (3), while inheriting the simplicity of least squares[45] and interpretability in terms of time correlation matrices.

IV. QUASISTATIONARY DISTRIBUTIONS AND CONVERGENCE TO A MARKOV RENEWAL PROCESS

For large enough decorrelation times, the underlying process reaches a local equilibrium each time that $R(t)$ makes a jump, leading to a Markov property for $R(t)$. We now make this precise using *quasistationary distributions* (QSDs).

The QSD of $X(t)$ in I is defined by the condition that if $X(t)$ is initially distributed as the QSD in I, then conditionally on staying in I, it remains distributed as the QSD. Writing η_I for the QSD in I,

$$\eta_I(\cdot) = \int \eta_I(dx)\mathbb{P}(X(t) \in \cdot | X(0) = x,\ X(s) \in I,\ s \le t), \tag{8}$$

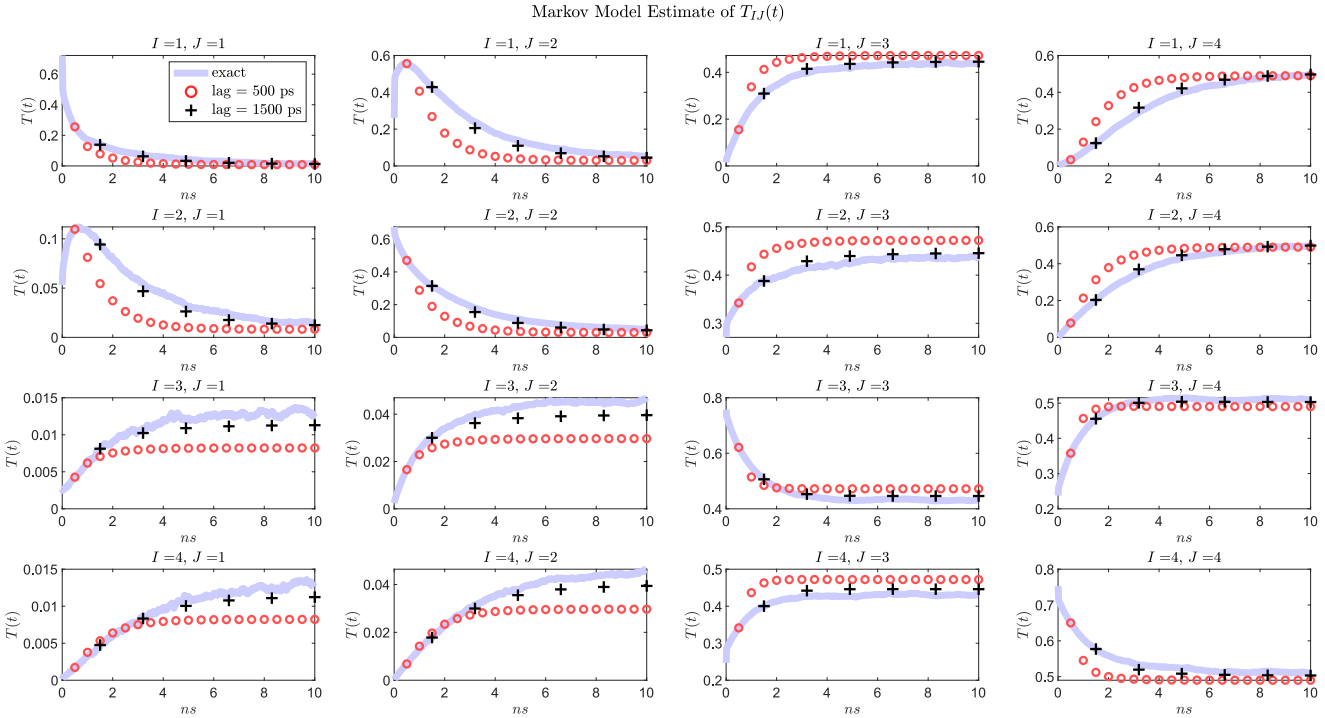
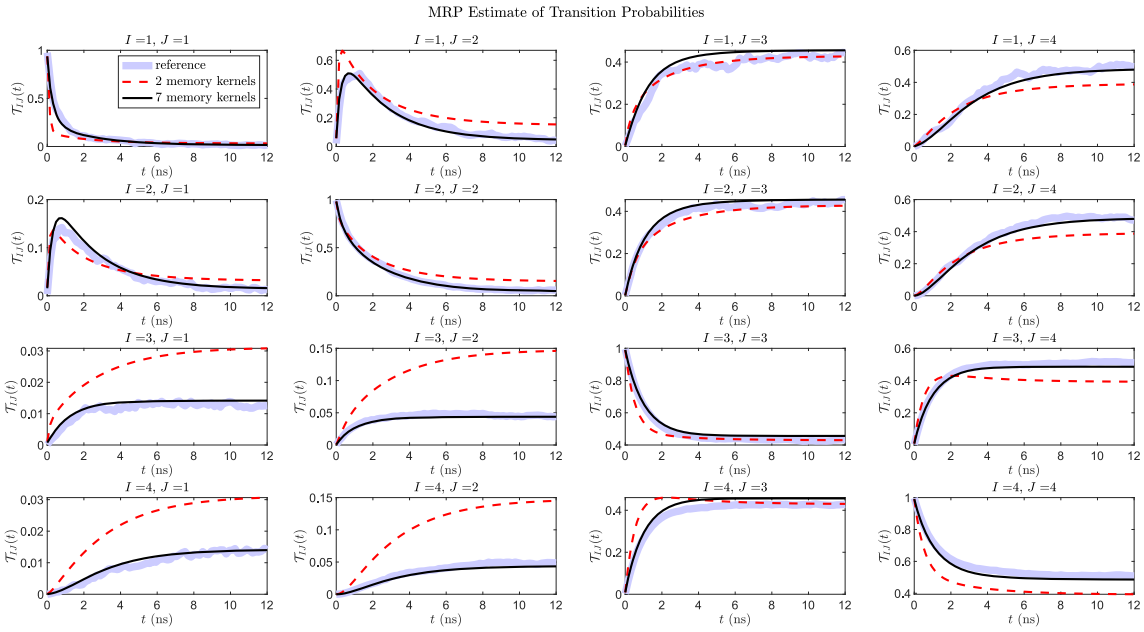
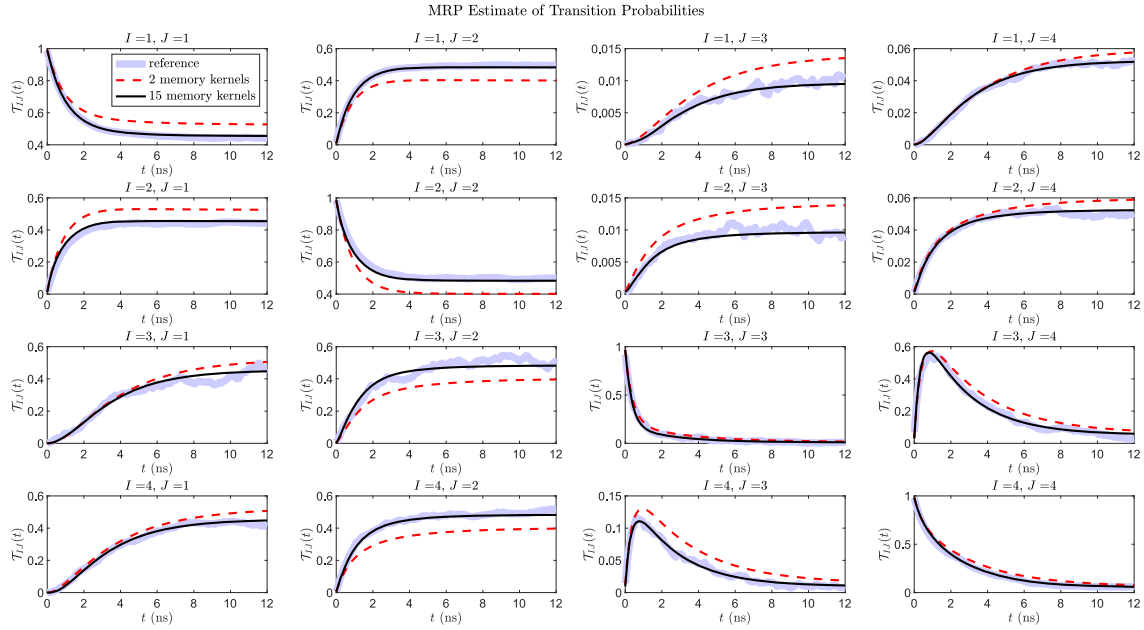


FIG. 2. Building a Markov model for alanine dipeptide, using states defined through PCCA. Except at very long lags, the Markov model is considerably less accurate than what we obtain with our methods (see Fig. 3). This is because a simple coarse graining of $X(t)$ into these states is not sufficiently Markovian. Each Markov model is based on a single transition matrix, computed from counts of transitions of $X(t)$ between macrostates at the specified lag time. Transitions at longer lags are computed using powers of this single matrix.



(a) PCCA states



(b) Rectangular states

FIG. 3. Results from our method for parameterizing the MRP using (a) states defined by PCCA and (b) equal rectangular states. A very good parameterization is achieved in each case with 7 and 15 memory kernels in (a) and (b), respectively. Plotted are transition probabilities inferred using (7) with a smaller number of memory kernels (dashed line) and a larger number of memory kernels (solid line). The kernels are computed using (5) and (6) with cutoff time twice the memory length ($t_{max} = 2t_{mem}$). We use the macroscopic time steps, $\tau = 8$ ps (a) and $\tau = 30$ ps (b), that define the "good" decorrelation times. [See Fig. 5 for the choice of τ in (a).] Results are clearly improved with the larger number of memory kernels.

where the variable x represents microstates of $X(t)$.

Under mild assumptions,[48,49]

$$\|\eta_I - \mathbb{P}(X(t) \in \cdot \,|\, X(s) \in I,\, s \le t)\| \le c_I \delta_I^t, \tag{9}$$

where c_I and $\delta_I < 1$ are constants, and the norm is the total variation of measures. Informally, given that $X(t)$ remains in macrostate I, it converges to η_I at a geometric rate.

In Theorem A.1 of Appendix A, we show that

$$\mathcal{T}(s,t) = O(t\delta^\sigma) + \sum_{0<r\le t} \mathcal{P}(r)\,\mathcal{T}(s,t-r) + \mathcal{F}(t), \tag{10}$$

where $\mathcal{P}(t)$ is the jump probability matrix of a Markov renewal process, $\mathcal{F}_{IJ}(t) = \delta_{I=J}\sum_L \sum_{s>t} \mathcal{P}_{IL}(s)$, and $\delta = \max_I \delta_I$, $\sigma = \min_I \tau_I$. It follows that the transition matrices $\mathcal{T}(s,t)$ converge to the transition matrices of a Markov renewal process defined by the jump time distribution $\mathcal{P}(t)$ at a geometric rate in terms of the decorrelation times.

V. NUMERICAL RESULTS

To demonstrate the potential of our method, we apply it to alanine dipeptide, using an MD trajectory[39] of length about 70 ms. Positions in ϕ–ψ space were saved at every 2 ps. The macrostates are either chosen by using PCCA (Perron-Cluster Cluster Analysis) or by dividing ϕ–ψ space into four equal rectangles. While the PCCA states are highly metastable, the rectangular states are not. A finite spatial discretization limits the accuracy of Markov models, as seen in Fig. 2, which shows that a Markov model does not accurately represent the discretized alanine dipeptide dynamics, except at long timescales.

We use a decorrelation time $\tau_I = \tau$ ps that is the same for all states $I = 1, 2, 3, 4$. These decorrelation times were chosen to be large enough to obtain good numerical accuracy of the renewal Eq. (2); see Fig. 5. Then, we construct a trajectory $R(t)$ as described in Sec. II (see also Fig. 1) and apply our method. The alanine MD trajectory was split in half into a training set and a test (or reference) set. We use the former to create our model of $R(t)$ and the latter to create reference results.

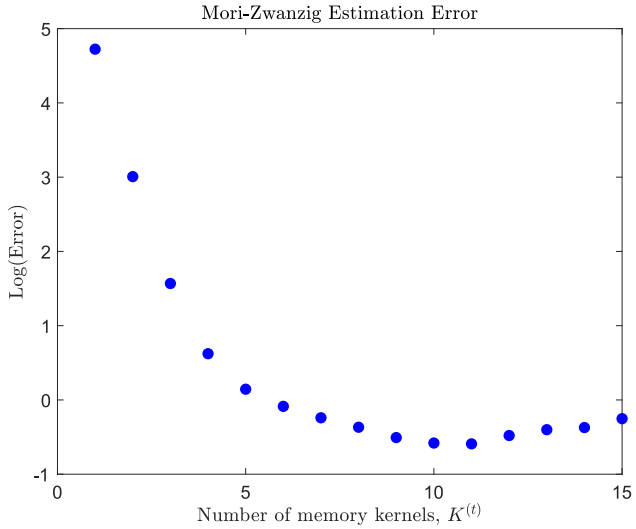
To assess our method, we compare it with a reference that uses the indicated value of τ. The reference results are based on simple counts of transitions. Figure 3 compares reference counts with our method's estimates of $\mathcal{T}(t)$. Figure 4 shows the error in $\mathcal{P}(t)$. To mitigate noise effects from finite sampling, we use the error measurement

$$\text{Error} = \sum_{I,J} \int_0^\infty \left(\int_0^t \frac{[\mathcal{P}_{IJ}(s) - \hat{\mathcal{P}}_{IJ}(s)]}{Z_{IJ}}\, ds \right)^2 \frac{\mathcal{P}_{IJ}(t)}{Z_{IJ}}\, dt, \tag{11}$$

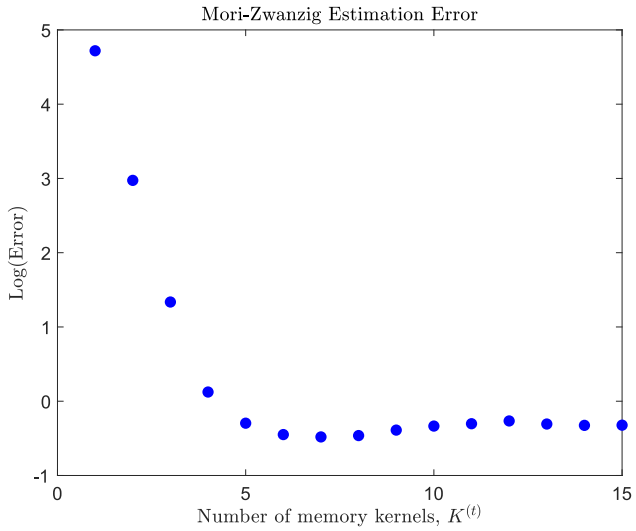
where $Z_{IJ} = \int_0^\infty \mathcal{P}_{IJ}(t)\, dt$ and where $\hat{\mathcal{P}}(t)$ is our estimate on training data, with $\mathcal{P}(t)$ being the reference. This is a slight variation on the Cramer–von Mises criterion.[50] Figures 3 and 4 show that the approach outlined in Sec. III gives good agreement with the reference, with just a few memory kernels.

A. Practical considerations

Our method requires a choice of macrostates and of scalar parameters τ, t_{mem}, and $t_{\max}$. Here, we discuss how these parameters might be chosen. Briefly, the microstates should be chosen as metastable states associated with timescales of interest; the parameter τ should be large enough for the Markov property to (nearly) hold, but no larger; and t_{mem} and $t_{\max}$ should be as large as needed to accurately parameterize the model, given constraints on how much data are available. We discuss all this in more detail below. In this discussion, as in the numerical simulations, we assume that all the decorrelation times equal τ, that is, $\tau_I = \tau$ for each macrostate I.

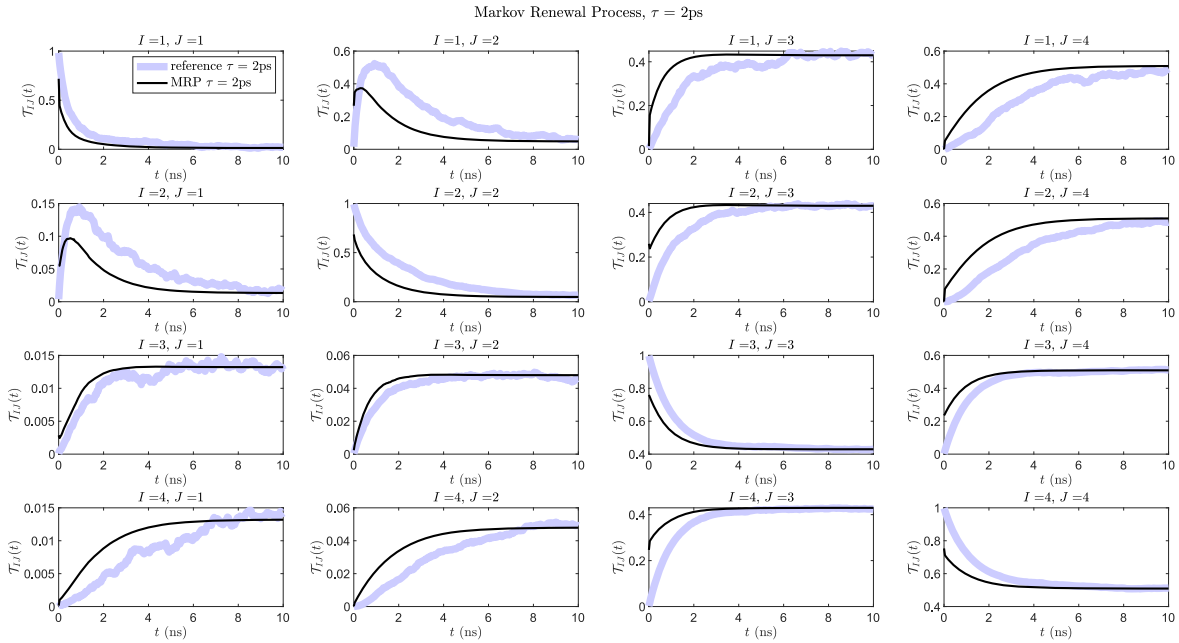


(a) PCCA States, $\tau = 8$ ps

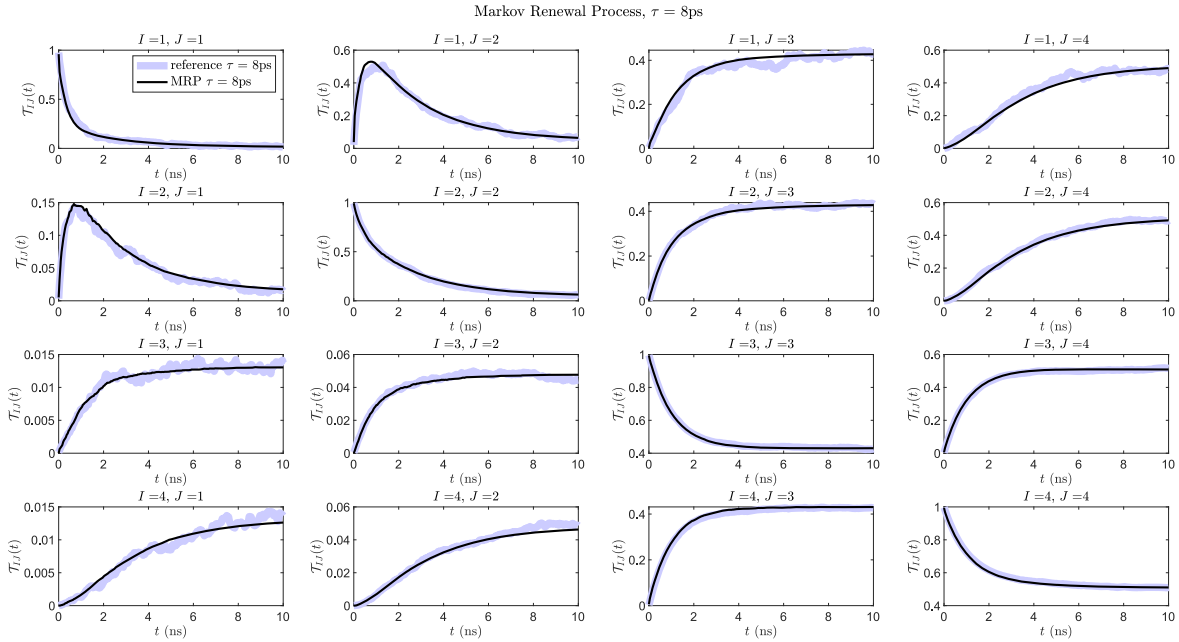


(b) Rectangular States, $\tau = 30$ ps

FIG. 4. The error in our method vs the number of memory kernels. Memory kernels are estimated at multiples of τ, and the error is defined by (11). The cutoff times, (a) $t_{\max} = 120$ ps and (b) $t_{\max} = 900$ ps, are chosen by applying the rule $t_{mem} = 0.5 \times t_{\max}$ to the largest t_{mem} pictured [for instance, in (b), $t_{mem} = 450$ ps, corresponding to 15 memory kernels]. (a) PCCA states, $\tau = 8$ ps. (b) Rectangular states, $\tau = 30$ ps.



(a) $\tau = 2$ps



(b) $\tau = 8$ps

FIG. 5. Verifying that $R(t)$ is approximately a MRP for large enough decorrelation times, for PCCA states. Plotted are reference transition probabilities computed from simple counts of $R(t)$, for τ = 2 ps in (a) and τ = 8 ps in (b), compared to probabilities computed from the renewal equation [Eq. (2)]. (In the renewal equation, the jump probability matrix, $\mathcal{P}$, is similarly computed from simple counts.) The (constant) decorrelation time must be chosen long enough to allow local equilibration within the macrostates. There is significant disagreement using τ = 2 ps in (a), while the larger value, τ = 8 ps, in (b) gives good agreement without being unnecessarily large. (a) τ = 2 ps. (b) τ = 8 ps.

We first consider τ, t_{mem}, and $t_{\max}$. With enough data, increasing τ, t_{mem}, and $t_{\max}$ will systematically improve results; in practice, though, there are tradeoffs. (Caveat: a too large τ causes modeling problems; see below.) Clearly, there need to be enough sampled transitions at each time lag. That is, we need enough samples of $\mathcal{T}_{IJ}(t)$ for each I, J and $t \leq t_{\max}$. Hence, for example, if data come in the form of many short trajectories of $X(t)$, then increasing $t_{\max}$ lowers transition counts and can improve model fidelity only to the extent that the number of sampled transitions does not get too low. The parameter t_{mem} defines the number of memory kernels, and we found good results when pairing it to $t_{\max}$ using the rule $t_{mem} \approx 0.5 \times t_{\max}$. In practice, $t_{\max}$ (and/or t_{mem}) could be chosen with standard techniques, such as cross-validation.

The macrostates and the parameter τ are more fundamental (though they are also subject to similar considerations concerning transition counts). Unlike t_{mem} and $t_{\max}$, which are parameters used to obtain the memory kernels that generate an *approximation* of $R(t)$, the macrostates and τ actually *define* $R(t)$. They must be chosen carefully to yield good results. For a given set of macrostates, a minimum value of τ is set by the requirement that $R(t)$ is approximately a MRP; the required value can be found empirically by using a plot as in Fig. 5 (we simply chose one "by eye" from such plots). Good macrostates are ones in which decorrelation occurs on a time scale much smaller than the typical escape time, i.e., good macrostates are metastable.[36] In practice, they could be chosen by standard techniques, such as PCCA.[29]

A bad choice of macrostates cannot be rescued by a good choice of τ. Indeed, $R(t)$ does not retain any events that occur on timescales smaller than τ. As a result, if τ is close to or larger than typical transition times between macrostates, then $R(t)$ can miss such transitions (as shown in Fig. 1), resulting in a potentially accurate but uninformative model. A good choice of both the macrostates and of τ is, therefore, important. For the purposes of this article, we think of the macrostates as already being given, and we choose τ by looking at plots such as Fig. 5, increasing τ until we find a good match.

Figure 2 shows an ordinary Markov model based on PCCA states. These PCCA states are the same as in Ref. 39. A lag of 1500 ps is needed for accuracy comparable to our methods. [Compare with Fig. 3(a).] This lag is on the order of the longest mean transition time, roughly 1000 ps. Particularly for macrostates 1 and 2, this lag sacrifices knowledge of shorter timescale (but still physically relevant) state-to-state transitions. Although these states are considered very good (Markovian) states, our methods still provide significant improvement over Markov models, as illustrated in Fig. 3.

Figure 3(a) shows results from our methods when using the PCCA states. There, we use a decorrelation time $\tau = 8$ ps. This serves as the fundamental time step of our coarse-grained model and is small enough that few transitions are missed. To build our model, we use many short trajectories of length 112 ps, smaller than the shortest mean transition time of 175 ps. (This trajectory length corresponds to using $\tau = 8$ ps, with $t_{mem} = 7$ memory kernels and $t_{\max} = 2 \times t_{mem}$.) In contrast, a similarly accurate Markov model in Fig. 2 requires trajectories of length 1500 ps. Recall that the longest mean transition time is around 1000 ps. In sum, the renewal model requires significantly shorter trajectories and is more accurate than the Markov model on all timescales.

Figure 3(b) shows analogous results for unphysical macrostates (defined as equal rectangles in ϕ-ψ coordinates, divided by the lines $\phi = 0, \pm\pi$ and $\psi = 0, \pm\pi$). Although these states are no longer metastable, results are similar to Fig. 3(a). (In this case, τ needs to be larger, however, resulting in our model missing some transitions, as discussed above.) We find good accuracy when $\tau = 30$ ps, $t_{mem} = 15$ memory kernels, and trajectories have length 900 ps. A Markov model would require a lag of 5000 ps for similar accuracy. Smaller Markov model lags of ~1000 ps result in widely inaccurate estimates for even a few time steps' prediction.

VI. DISCUSSION

The methodology introduced in this paper allows for the systematic exploitation of a rich set of trade-offs between compactness, expressiveness, and accuracy. It is particularly well suited to cases where the system contains a relatively small number of metastable states, but where metastability is insufficient for a Markovian assumption to be accurate. In contrast to conventional approaches, such as Markov state models, where accuracy can be improved by increasing the number of states at the cost of interpretability, the accuracy of the approach proposed is, instead, controlled by increasing the decorrelation time to ensure the convergence to a MRP. Doing so, however, comes with its own trade-off, as the expressiveness of $R(t)$ decreases when the decorrelation time exceeds the shortest transition time. However, the geometric convergence rate to an MRP makes this trade-off particularly advantageous as a small increase in decorrelation time yields a large increase in accuracy. Therefore, even for modestly metastable systems, it should be possible to produce very accurate MRPs using decorrelation times that are short compared to typical transition times, hence minimizing the loss of kinetic information. In this situation, the MZ approach described above will also yield a compact representation in terms of a limited number of kernel matrices. As shown above, this approach allows one to obtain compact and accurate models even with sub-optimal state definitions, which is very useful, given that optimizing state definitions in high dimension is generally difficult. That being said, the approach cannot fix state definitions where most of the states are not at least somewhat metastable, as accuracy would demand very long decorrelation times, which would then entail low expressiveness. It is arguable, however, that no representation in terms of jump processes would be appropriate in such a scenario.

ACKNOWLEDGMENTS

D. Aristoff and M. Johnson gratefully acknowledge the support from the National Science Foundation via Award No. DMS 2111277. D. Perez was supported by the Laboratory Directed Research and Development program of Los Alamos National Laboratory under Project No. 20220063DR. Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of U.S. Department of Energy (Contract No. 89233218CNA000001).

D. Aristoff and M. Johnson acknowledge illuminating discussions with D.M. Zuckerman, J. Copperman, J. Russo, G. Simpson, and R. J. Webber.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

David Aristoff: Conceptualization (equal); Data curation (equal); Formal analysis (lead); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (lead); Resources (lead); Software (supporting); Supervision (lead); Validation (supporting); Visualization (supporting); Writing – original draft (lead); Writing – review & editing (lead). **Mats Johnson**: Conceptualization (supporting); Data curation (supporting); Formal analysis (supporting); Investigation (equal); Methodology (equal); Software (lead); Validation (lead); Visualization (lead); Writing – original draft (supporting); Writing – review & editing (supporting). **Danny Perez**: Conceptualization (equal); Data curation (equal); Formal analysis (supporting); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (supporting); Resources (supporting); Software (supporting); Supervision (supporting); Validation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX A: CONVERGENCE TO A MARKOV RENEWAL PROCESS

We begin by introducing some notation. Let

$$E_{IJ}(s,t) = \{R(s+t) = J,\, R(s') = I,\, s \le s' < s+t\}$$

be the event of switching to from I to J after a time t, starting from time s. Let

$$E_J(t) = \{R(t) = J\}, \qquad E_J^c(t) = \{R(t) \neq J\}$$

be the events that $R(t) = J$ and $R(t) \neq J$, respectively.

We use $\sim$ to indicate equality in distribution, for example, $X(s) \sim \eta_I$ indicates that $X(s)$ is distributed as η_I.

The following result demonstrates convergence in distribution of $R(t)$ to a Markov renewal process as the decorrelation times grow.

Theorem A.1 (exactness of the renewal equation). *Assume that each macrostate* I *has a QSD* η_I, *and assume that (9) holds. Define*

$$\mathcal{T}_{IJ}(t) = \mathbb{P}(E_J(s+t)|E_I(s),\, X(s) \sim \eta_I). \tag{A1}$$

Then, $\mathcal{T}(s,t)$ *defined by (1) converges to* $\mathcal{T}(t)$ *defined by (A1) as* $\min_I \tau_I \to \infty$. *Moreover, the limit* $\mathcal{T}(t)$ *is the unique solution to the renewal equation (2) when* $\mathcal{P}$ *is defined by*

$$\mathcal{P}_{IJ}(t) = \delta_{I\neq J}\mathbb{P}(E_{IJ}(s,t)|E_I(s),\, X(s) \sim \eta_I). \tag{A2}$$

Proof. Using the law of total probability,

$$\begin{aligned}\mathcal{T}_{IJ}(s,t) &= \mathbb{P}(E_J(s+t)|E_I^c(s_-),\, E_I(s)) \\ &= \sum_{K\neq I}\sum_{0<r\le t} \mathbb{P}(E_J(s+t)|E_I^c(s_-), E_{IK}(s,r)) \\ &\quad \times \mathbb{P}(E_{IK}(s,r)|E_I^c(s_-),\, E_I(s)) \\ &\quad + \delta_{I=J}\mathbb{P}(E_{II}(s,t)|E_I^c(s_-),\, E_I(s)).\end{aligned} \tag{A3}$$

Let $\sigma = \min_I \tau_I$, and in the notation of (9), define

$$c = \max_I c_I, \quad \delta = \max_I \delta_I.$$

Using (9), (A1), and the Markov property of $X(t)$,

$$\begin{aligned}\mathcal{T}_{IJ}(s,t) &= \mathbb{P}(E_J(s+t)\,|\,E_I^c(s_-),\, E_I(s)) \\ &= \int \mathbb{P}(E_J(s+t)\,|\,E_I^c(s_-),\, E_I(s),\, X(s) = x) \\ &\quad \times \mathbb{P}(X(s) \in dx\,|\,E_I^c(s_-),\, E_I(s)) \\ &= \int \mathbb{P}(E_J(s+t)\,|\,E_I(s),\, X(s) = x)\eta_I(dx) + \varepsilon_I \\ &= \mathcal{T}_{IJ}(t) + \varepsilon,\end{aligned} \tag{A4}$$

where $|\varepsilon| \le c\delta^\sigma$. Similar calculations show that

$$\begin{aligned}\delta_{I\neq K}\mathbb{P}(E_J(s+t)|E_I^c(s_-), E_{IK}(s,r)) &= \mathcal{T}_{KJ}(t-r) + \varepsilon, \\ \delta_{I\neq K}\mathbb{P}(E_{IK}(s,r)|E_I^c(s_-),\, E_I(s)) &= \mathcal{P}_{IK}(r) + \varepsilon, \\ \mathbb{P}(E_{II}(s,t)|E_I^c(s_-),\, E_I(s)) &= \mathcal{F}_{II}(t) + \varepsilon,\end{aligned}$$

where each ε is different but $|\varepsilon| \le c\delta^\sigma$, and

$$\begin{aligned}\mathcal{F}_{IJ}(t) &= \delta_{I=J}\mathbb{P}(E_{II}(s,t)|E_I(s),\, X(s) \sim \eta_I) \\ &= \delta_{I=J}\textstyle\sum_L \sum_{s>t} \mathcal{P}_{IL}(s).\end{aligned}$$

Combining the previous three displays with (A3),

$$\mathcal{T}(s,t) = O(t\delta^\sigma) + \sum_{0<r\le t} \mathcal{P}(r)\,\mathcal{T}(s+r, t-r) + \mathcal{F}(t). \tag{A5}$$

Now, from (A4), we conclude that $\mathcal{T}(s,t)$ converges to $\mathcal{T}(t)$ as $\sigma \to \infty$. Meanwhile, using (A5), it is readily shown from a standard renewal equation representation[38] (Proposition 4.2) that $\mathcal{T}(t)$ is the unique solution to Eq. (2). □

The proof shows that the convergence rate is geometric in σ on finite time intervals, suggesting that large decorrelation times are not needed in order to model $R(t)$ as a Markov renewal process at least for reasonably defined states.

APPENDIX B: ACTIONS OF PROJECTOR AND MARKOV KERNELS

Below, we introduce another process $C(t)$ that counts the consecutive time that $X(t)$ has spent in its current macrostate, where the count stops at τ_J if $X(t) \in J$.

To develop the Mori–Zwanzig theory, we introduce the augmented Markov chain $[X(t), R(t), C(t)]$ on augmented states (x, I, s), where x and I represent the current values of $X(t)$ and $R(t)$ and s is the consecutive time that $X(t)$ has spent in the macrostate in

which it currently resides up to the decorrelation time. This Markov chain has time step τ.

Below, let $\mathbb{P}^{x,I,s}$ denote the probability for the augmented Markov chain that starts at $(X(0),R(0),C(0)) = (x,I,s)$. Let T be the Markov kernel of this augmented chain,

$$T(x,I,s;dy,J,t) = \mathbb{P}^{x,I,s}[(X(\tau),R(\tau),C(\tau)) = (dy,J,t)]. \quad \text{(B1)}$$

We will also make use of more broadly defined kernels $S(x,I,s;\ dy,J,t)$ by relaxing the nonnegativity and unit normalization properties of T. Specifically, such a kernel acts on functions $f = f(x,I,s)$ of augmented space according to the rule

$$Sf(x,I,s) = \int \sum_{J,t} S(x,I,s;dy,J,t) f(y,J,t).$$

We define a projector P on functions $f = f(x,I,s)$ of augmented states, that is, a mapping satisfying $P^2 = P$, by

$$Pf(x,I,s) = \int \eta_I(dz) f(z,I,\tau_I). \quad \text{(B2)}$$

APPENDIX C: PRINCIPAL AND ORTHOGONAL DYNAMICS AND MARKOVIAN CASE

The Mori–Zwanzig theory is characterized by a *principal* and *orthogonal* dynamics. The principal dynamics is driven by PT, defined by

$$PT(x,I,s;dy,J,t) = \int \eta_I(dz) T(z,I,\tau_I;dy,J,t).$$

The orthogonal dynamics is driven by QT, where $Q = \mathrm{Id} - P$ and Id is the identity operator, that is, $QT = T - PT$.

We consider a special *Markovian case*, in which the underlying dynamics instantaneously reaches the QSD in whatever macrostate it resides in, with associated decorrelation times $\tau_I = 0$ for all I. In this case, $T = PT$, so the orthogonal dynamics vanish, $QT = 0$, and all but one of the memory kernels is zero; see Appendix D.

APPENDIX D: DERIVATION OF THE MORI-ZWANZIG EQUATION

The following lemma applies to any transition kernel T and projector P, although we have in mind the Markov kernel T in (B1) and the projector P in (B2).

Lemma D.1. For any projector P and its complementary projector $Q = \mathrm{Id} - P$, where Id is the identity mapping, we have

$$PT^n = \sum_{m=1}^{n} K(m) PT^{n-m} + F(n), \quad \text{(D1)}$$

where $K(n) = PT(QT)^{n-1}$ and $F(n) = PT(QT)^{n-1}Q$.

Proof. Start with the self-evident equations,

$$PT^{n+1} = PTPT^n + PTQT^n, \quad \text{(D2)}$$

$$QT^{n+1} = QTPT^n + QTQT^n. \quad \text{(D3)}$$

Using induction in (D3),

$$QT^n = \sum_{m=1}^{n} (QT)^m PT^{n-m} + (QT)^n Q.$$

Plugging this back into (D2) yields the result. □

Below, we will make use of functions χ_J defined by

$$\chi_J(x,I,s) = \delta_{I=J}.$$

Theorem D.2 (exactness of the MZ equation). *Let $K(n)$ be as in Lemma D.1, where P is the projector from (B2) and T is defined in (B1). Define*

$$\mathcal{K}_{IJ}(n\tau) := K(n)\chi_J(x,I,s). \quad \text{(D4)}$$

Then, with $\mathcal{T}(t)$ as in (A1),

$$\mathcal{T}(n\tau) = \sum_{m=1}^{n} \mathcal{K}(m\tau)\mathcal{T}((n-m)\tau). \quad \text{(D5)}$$

Proof. Multiply (D1) on the right by $\chi_J(x,I,s)$. Note that $P\chi_J = \chi_J$ so that $Q\chi_J = 0$ and $F(n)\chi_J(x,I,s) = 0$. Thus,

$$PT^n\chi_J(x,I,s) = \sum_{m=1}^{n} K(m) PT^{n-m}\chi_J(x,I,s). \quad \text{(D6)}$$

Recalling $\mathcal{T}(t)$ defined in (A1), we compute

$$\begin{aligned} PT^n\chi_J(x,I,s) &= \int \eta_I(dx) T^n\chi_J(x,I,\tau_I) \\ &= \int \eta_I(dx) \mathbb{E}^{x,I,\tau_I}[\chi_J(X(n\tau),R(n\tau),C(n\tau))] \\ &= \int \eta_I(dx) \mathbb{P}^{x,I,\tau_I}[R(n\tau) = J] \\ &= \mathcal{T}_{IJ}(n\tau). \end{aligned}$$

Below, write $S_m = T(QT)^{m-1}$, and note that $PT^{n-m}\chi_J(y,L,t)$ does not depend on y or t. Thus,

$$\begin{aligned} &\sum_L \mathcal{K}_{IL}(m\tau)\mathcal{T}_{LJ}((n-m)\tau) \\ &\quad = \sum_L K(m)\chi_L(x,I,s) PT^{n-m}\chi_J(y,L,t) \\ &\quad = \sum_L \int \eta_I(dz) \left[\int \sum_t S_m(z,I,\tau_I;dy,L,t) \right] PT^{n-m}\chi_J(y,L,t) \\ &\quad = \int \eta_I(dz) \left[\int \sum_{L,t} S_m(z,I,\tau_I;dy,L,t) PT^{n-m}\chi_J(y,L,t) \right] \\ &\quad = K(m) PT^{n-m}\chi_J(x,I,s). \end{aligned}$$

Combining the last two displays with (D6) gives (D5). □

Next, we show that all but one of the memory kernels vanishes in the case where $R(t)$ is Markovian.

Theorem D.3. *Suppose that $\tau_I = 0$ for all I and that*

$$T(x,I,s;dy,J,t) = \int \eta_I(dz) T(z,I,\tau_I;dy,J,t).$$

Then, $\mathcal{K}(n\tau) = 0$ *for* $n > 1$.

Proof. The assumption on T implies that $PT = T$, so $QT = 0$ and the result follows from the formula

$$\mathcal{K}_{IJ}(n\tau) = PT(QT)^{n-1}\chi_J(x, I, s). \qquad \square$$

APPENDIX E: MINIMIZING THE LOSS FUNCTION

The gradient of the loss function (4) is

$$\frac{1}{2}\nabla_{\mathcal{K}(t)}\mathcal{L}(\mathcal{K}) = \sum_{r\le t_{\max}} \mathcal{T}(r)\mathcal{T}(r-t)^T - \sum_{0<s\le t_{mem}} \mathcal{K}(s)\sum_{r\le t_{\max}} \mathcal{T}(r-s)\mathcal{T}(r-t)^T,$$

where by definition $\mathcal{T}(s) = 0$ for $s < 0$.

This immediately leads to the linear system reported in (5).

REFERENCES

[1] M. Karplus and G. A. Petsko, Nature **347**, 631 (1990).
[2] M. Karplus and J. A. McCammon, Nat. Struct. Biol. **9**, 646 (2002).
[3] T. Hansson, C. Oostenbrink, and W. van Gunsteren, Curr. Opin. Struct. Biol. **12**, 190 (2002).
[4] J. D. Durrant and J. A. McCammon, BMC Biol. **9**, 71 (2011).
[5] A. Hospital, J. R. Goñi, M. Orozco, and J. L. Gelpí, Adv. Appl. Bioinform. Chem. **8**, 37 (2015).
[6] S. A. Hollingsworth and R. O. Dror, Neuron **99**, 1129 (2018).
[7] L. T. Chong, A. S. Saglam, and D. M. Zuckerman, Curr. Opin. Struct. Biol. **43**, 88 (2017).
[8] J. Weare, J. Comput. Phys. **228**, 4312 (2009).
[9] R. J. Webber, D. A. Plotkin, M. E. O'Neill, D. S. Abbot, and J. Weare, Chaos **29**, 053109 (2019).
[10] J. Finkel, R. J. Webber, E. P. Gerber, D. S. Abbot, and J. Weare, J. Atmos. Sci. **80**, 519 (2023).
[11] J. Finkel, E. P. Gerber, D. S. Abbot, and J. Weare, AGU Adv. **4**, e2023AV000881 (2023).
[12] G. Seiden and P. J. Thomas, Rev. Mod. Phys. **83**, 1323 (2011).
[13] A. A. Fingelkurts and A. A. Fingelkurts, Int. J. Neurosci. **114**, 843 (2004).
[14] C. Haldeman and J. M. Beggs, Phys. Rev. Lett. **94**, 058101 (2005).
[15] P. J. Hellyer, G. Scott, M. Shanahan, D. J. Sharp, and R. Leech, J. Neurosci. **35**, 9050 (2015).
[16] A. Córdova-Palomera, T. Kaufmann, K. Persson, D. Alnæs, N. T. Doan, T. Moberget, M. J. Lund, M. L. Barca, A. Engvig, A. Brækhus *et al.*, Sci. Rep. **7**, 40268 (2017).
[17] S. Naik, A. Banerjee, R. S. Bapi, G. Deco, and D. Roy, Trends Cognit. Sci. **21**, 509 (2017).
[18] F. Cavanna, M. G. Vilas, M. Palmucci, and E. Tagliazucchi, Neuroimage **180**, 383 (2018).
[19] Y. Pomeau, Physica D **23**, 3 (1986).
[20] C. Matthews, B. Stadie, J. Weare, M. Anitescu, and C. Demarco, arXiv:1806.02420 (2018).
[21] D. B. Duncan and R. M. Dunwell, Proc. Edinburgh Math. Soc. **45**, 701 (2002).
[22] X. Sun and M. J. Ward, Eur. J. Appl. Math. **10**, 27 (1999).
[23] D. Estep, Nonlinearity **7**, 1445 (1994).
[24] P. Groisman, S. Saglietti, and N. Saintier, Stochastic Process. Appl. **128**, 1558 (2018).
[25] J. R. Norris, *Markov Chains* (Cambridge University Press, 1998), Vol. 2.
[26] R. Durrett and R. Durrett, *Essentials of Stochastic Processes* (Springer, 1999), Vol. 1.
[27] D. Angeli, in *2009 European Control Conference (ECC)* (IEEE, 2009), pp. 649–657.
[28] A. F. Voter, *Radiation Effects in Solids* (Springer, 2007), pp. 1–23.
[29] J. D. Chodera and F. Noé, Curr. Opin. Struct. Biol. **25**, 135 (2014).
[30] A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, J. Comput. Phys. **17**, 10 (1975).
[31] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[32] B. E. Husic and V. S. Pande, J. Am. Chem. Soc. **140**, 2386 (2018).
[33] G. Di Gesù, T. Lelièvre, D. Le Peutrec, and B. Nectoux, Faraday Discuss. **195**, 469 (2016).
[34] S. M. Ross, *Stochastic Processes* (John Wiley and Sons, 1995).
[35] P. Brémaud, *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues* (Springer Science and Business Media, 2001), Vol. 31.
[36] T. Lelièvre, Eur. Phys. J.: Spec. Top. **224**, 2429 (2015).
[37] T. Lelièvre, *Handbook of Materials Modeling: Methods: Theory and Modeling* (Springer, 2020), p. 773.
[38] E. Cinlar, Manage. Sci. **21**, 727 (1975).
[39] A. Agarwal, S. Gnanakaran, N. Hengartner, A. F. Voter, and D. Perez, arXiv:2008.11623 (2020).
[40] Y. Chen, E. N. Epperly, J. A. Tropp, and R. J. Webber, arXiv:2207.06503 (2022).
[41] M. Díaz, E. N. Epperly, Z. Frangella, J. A. Tropp, and R. J. Webber, arXiv:2304.12465 (2023).
[42] S. Cao, A. Montoya-Castillo, W. Wang, T. E. Markland, and X. Huang, J. Chem. Phys. **153**, 014105 (2020).
[43] E. Darve, J. Solomon, and A. Kia, Proc. Natl. Acad. Sci. U. S. A. **106**, 10884 (2009).
[44] Y. T. Lin, Y. Tian, D. Livescu, and M. Anghel, SIAM J. Appl. Dyn. Syst. **20**, 2558 (2021).
[45] Y. T. Lin, Y. Tian, D. Perez, and D. Livescu, SIAM J. Appl. Dynam. Syst. (to be published).
[46] A. J. Dominic III, T. Sayer, S. Cao, T. E. Markland, X. Huang, and A. Montoya-Castillo, Proc. Natl. Acad. Sci. U. S. A. **120**, e2221048120 (2023).
[47] A. J. Dominic III, S. Cao, A. Montoya-Castillo, and X. Huang, J. Am. Chem. Soc. **145**, 9916 (2023).
[48] P. Collet, S. Martínez, and J. San Martín, *Quasi-Stationary Distributions: Markov Chains, Diffusions and Dynamical Systems* (Springer, 2013), Vol. 1.
[49] N. Champagnat and D. Villemonais, Electron. J. Probab. **28**, 1 (2023).
[50] T. W. Anderson, Ann. Math. Stat. **33**, 1148 (1962).