On approximating a weak Markovian process as Markovian: Are we justified when discarding longtime correlations

Cite as: J. Chem. Phys. 150, 085101 (2019); doi: 10.1063/1.5056242

Submitted: 13 September 2018 • Accepted: 31 January 2019 •

Published Online: 25 February 2019

Kai-Yang Leong and Feng Wang^{a)}

AFFILIATIONS

Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, Arkansas 72701, USA

^{a)}fengwang@uark.edu

ABSTRACT

The effect for removing weak longtime correlation is studied using a model system that contains a driven atom at liquid density under strong thermal fluctuations. The force that drives the tagged particle is about 1% of the average random force experienced by the particle. The tagged particle is allowed to assume a range of masses from 1/8 to 80 times that of a surrounding particle to study the effects of inertia. The driving force is indefinitely correlated but much weaker than “random” fluctuations from the environment. From this study, it is shown that the environmental influence is not fully random leading to the force autocorrelation function being a poor metric for detecting the correlated driving force. Although the velocity autocorrelation function shows stronger correlation for systems with higher inertia, the velocity autocorrelation function decays to a very small value of 2.5×10^{-3} even for the most massive driven particle. For systems with small inertia, our study reveals that discarding longtime correlation has negligible influence on the first passage time (FPT) estimate, whereas for particles with large inertia, the deviation can indeed be appreciable. It is interesting that the Markov State Model (MSM) still produces reasonable estimates on the FPT even when a very short lag time that clearly violates the Markovianity assumption is used. This is likely a result of favorable error cancellations when the MSM transition probability matrices were constructed using trajectories that are much longer than the lag time.

Published under license by AIP Publishing. <https://doi.org/10.1063/1.5056242>

I. INTRODUCTION

Many real life processes may be influenced by longtime correlations. For example, it has been argued that protein folding occurs along a single funnel in the free energy landscape;^{1,2} thus, there is a possibility that the entire folding process is weakly correlated over microsecond timescale or longer. It is probably reasonable to assume that correlations over extended periods of time will be weak. Under this assumption, it is a common practice to ignore longtime correlations. For example, when a long trajectory is simulated by a series of shorter ones, occasionally new initial velocities are generated at restarts. Certain stochastic thermostats will cause loss of longtime correlations in velocity.³ Many enhanced sampling methods, especially those based on piecing together shorter trajectories, ignore correlations over an

extended period of time.^{4–17} For example, the Markov State Model (MSM) assumes that all correlations are lost over a lag time.^{18–25} It is anticipated that, by choosing sufficiently long lag times, the evolution is approximately Markovian between updates. Several methods exist for verification of Markovianity after a lag time.^{26–30} In practice, however, the detection of weak correlations is hampered by statistical noises that are omnipresent for finite temperature simulations in the condensed phase. The most challenging problems are those where the correlation is significantly smaller than thermal fluctuations.

Protein folding is strongly influenced by random fluctuations. Longtime correlations, if present, are anticipated to be much smaller than such fluctuations. The objective of this work is to study the effect of very weak longtime correlations. We investigate to what degree ignoring such weak

correlations will affect the overall kinetics. Our model system will be indefinitely correlated by design. However, long-time correlation is much smaller than fluctuations. Particular attention will be paid to the popular MSM method, although the insights from this study are applicable to many other enhanced sampling algorithms,^{4–17} where longtime kinetics are derived from shorter trajectories while ignoring longtime correlations.

Molecular dynamics (MD) is fully deterministic and Markovian in the phase space. However, for a driven system such as those studied here, the Markovianity can be violated by discarding velocity information. The removal of velocities leads to a projection to a lower dimensional space that is not Markovian.^{31–33} In this work, longtime correlations are removed in one case by randomly resetting the system velocities and in another case by construction of MSM without the velocity information. We investigate whether such a treatment of a weakly non-Markovian process as Markovian will introduce an appreciable error. This paper is organized as follows: The model system is described in Sec. II. The results are presented in Sec. III, and summary and conclusion are given in Sec. IV.

II. MODEL SYSTEM

In this study, we will investigate the first passage time (FPT) of a driven atom in super-critical Ar at a relatively high density. The simulation box contains 328 Ar atoms at a density of 34.04 mol/L and a temperature of 200 K.

The Ar atoms are modeled with a Lennard-Jones potential V_{LJ}

$$V_{LJ}(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right], \quad (1)$$

with an atomic size parameter, σ , of 0.3405 nm and a well depth, ϵ , of 0.9980 kJ/mol.³⁴ The Ar model has a critical point of 162.92 K and 15.14 mol/L.³⁵ The 200 K simulation temperature is thus above the critical point and the box density corresponds to the density of a liquid below the critical temperature.³⁶ Unlike systems with slow hydrogen bond relaxations, a rare gas system intrinsically decorrelates much faster; thus, we are able to study the effect of applied correlated force with relatively short trajectories.

At these simulation conditions, the root mean square force is measured to be 7.847 kJ/(mol Å). For one of the 328 Ar atoms, we apply a constant biasing force of 0.0965 kJ/(mol Å), which is slightly more than 1% of the root mean square force resulting from random collisions at this temperature. The special particle with the constant driving force will be referred to as the tagged particle. The simulation is performed in an orthorhombic box with dimensions of 2 nm × 2 nm × 4 nm. The weak biasing force is applied along the longer dimension of the box, z . The FPT is defined as the time it takes for the tagged particle to travel the full length of 4 nm along z for the first time. Since the driving force is very small, it is easy to understand that the particle has almost equal probability of traveling along or opposite to the direction of the force over short distances. However, the integrated driving force for travelling over 4 nm will be 3.86 kJ/mol, which is

more than two times the thermal energy RT at 200 K. Thus, over a larger distance, the effect of the bias can be substantial. Since the system is always near equilibrium, the probability can be approximated with the Boltzmann distribution. With the 4 nm box size, the probability of travelling the opposite direction for 4 nm is about 1% (9.6×10^{-3}) of that for the particle to travel along the direction of the driving force for 4 nm. We note that the particle trajectory is continuously tracked until it passes the 4 nm mark in the positive direction. We do allow the particle to travel as far in the opposite direction as it needs.

Although a small work is being done to the system as a result of the external driving force, this work is negligibly small when compared to the total energy. In addition, the system temperature is maintained with the Nosé-Hoover

FIG. 1. (a) The force autocorrelation function, C_F , of the tagged particle. Panel (b) presents a zoomed-in view to show the fluctuations of the tail of the C_F . Panel (c) shows a 10 ps running average of the C_F to better reveal the asymptotic limit.

thermostat^{37,38} with a 2 ps relaxation time. The equation of motion is integrated with a 5 fs timestep.

The mass of the tagged particle should play an important role since it determines the inertia. A heavier particle is anticipated to amplify the effect of the driving force. A protein may have a relatively large mass although the driving force for any dynamical events, such as folding, might be small at any instant. In order to study the effect of inertia, the tagged particle is allowed to assume masses other than the atomic mass of Ar. In this study, we investigate tagged particles with masses of 5 g/mol, 40 g/mol, 320 g/mol, and 3200 g/mol.

Figure 1 shows the force autocorrelation function,

$$C_{ff}(t) = \frac{\langle \mathbf{f}(t) \cdot \mathbf{f}(0) \rangle}{\langle \mathbf{f}(0) \cdot \mathbf{f}(0) \rangle}, \quad (2)$$

for the tagged particle. In Eq. (2), the angle brackets indicate ensemble averages and bold letters indicate vectors. If the force, $\mathbf{f}$, is a summation of a correlated component $\mathbf{f}_c$ being the driving force and a random component $\mathbf{f}_r$ being the environmental force, which is assumed to be a white noise that is uncorrelated at longtime, one would anticipate that $C_{ff}(t)$ approaches $\langle \mathbf{f}_c / \mathbf{f} \rangle^2$ at large t (see proof in the [supplementary material](#)). In this case, the longtime asymptote would be

$$\langle \mathbf{f}_c / \mathbf{f} \rangle^2 = 0.012^2 = 1.44 \times 10^{-4}. \quad (3)$$

This long time limit is marked as a horizontal line in Fig. 1. It can be seen from Fig. 1 that the fluctuation of $C_{ff}(t)$ is orders of magnitude larger than the average at longtime, t , making the detection of any asymptotic behavior challenging. In order to remove the fluctuations and obtain the long time asymptote, 10 ps running averages of $C_{ff}(t)$ are calculated and plotted in Fig. 1(c). It is clear that the true asymptote is not strongly affected by particle mass and is orders of magnitude smaller than that predicted by Eq. (3). Figure S1 of the

[supplementary material](#) shows that when the driving force is as large as the root mean square random forces, a vanishing asymptote is also observed in such a case.

Such a small asymptotic average is actually not surprising. At the high simulation density, each particle has a free mean path of 0.335 Å, which leads to constant collisions. The collective effect of the collisions acts as a frictional resistance and almost completely cancels the driving force; thus, the force from the environment could not be considered as a white noise. The white noise assumption that leads to Eq. (3) is not valid. This violation leads us to observe a longtime average of $C_{ff}(t)$ much smaller than the naïve anticipation. The vanishing asymptote is analogous to an object falling through the atmosphere under the force of gravity. Due to frictional force, such an object is expected to reach a terminal velocity with a zero net force, which would result in a long time force correlation of zero. The vanishing asymptote along with the large fluctuations makes the force correlation an ineffective metric for detecting longtime correlation. It is also worth noting that the longtime average of $C_{ff}(t)$ approaches zero similarly for particles with different masses, which would seem to suggest that particles with different masses would behave similarly when correlations are ignored. However, it is clear from our subsequent calculations that removing longtime correlations from particles with larger inertia will have a much larger effect on transport kinetics, further suggesting that focusing only on the tail of $C_{ff}(t)$ is likely to lead to improper conclusions.

Although the force autocorrelation functions average to virtually zero at large t , the longtime correlation is more visible with the velocity autocorrelation function

$$C_{vv}(t) = \frac{\langle \mathbf{v}(t) \cdot \mathbf{v}(0) \rangle}{\langle \mathbf{v}(0) \cdot \mathbf{v}(0) \rangle}. \quad (4)$$

FIG. 2. (a) The velocity autocorrelation function, C_{vv} , of the tagged particle along with non-tagged Ar particles. Panels (b) and (c) are zoomed-in to better show fast decorrelation. Panel (d) shows the same functions with a semi-log scale.

Figure 2 reports the C_{vv} for both the tagged particle and a non-tagged particle. It can be seen that the effect of the biasing force is almost negligible for lighter tagged particles. The C_{vv} for the 40 g/mol tagged particle is indistinguishable from that of the non-biased particles for the first 5 ps. After 5 ps, a faint tail in C_{vv} can be observed. The lighter 5 g/mol particle decorrelates much faster showing a residual correlation comparable to simulation noises in our study. The 320 g/mol tagged particle retains a stronger correlation in C_{vv} as expected. However, even for this mass, the correlation becomes less than 0.1% after 15 ps. The C_{vv} for the 3200 g/mol tagged particle drops to about 0.3% after 60 ps, clearly demonstrating the effect of the increased inertia. It is worth emphasizing that even for this massive particle, the longtime asymptote of C_{vv} averages to about 2.5×10^{-3} . Although such a weak correlation of C_{vv} is detectable in our study, for a complex system, such as protein folding, the degree of freedom is likely to be a collective variable that is not known *a priori*. Detecting weak velocity correlation in such a collective variable will be much harder.

III. RESULTS AND DISCUSSION

Table I reports the mean FPT for the tagged particle to travel 4 nm. All the numbers were averaged over 10 000 repetitions from random initial conformations obtained from equilibrium NVT simulations without the driving force. The initial conformations are separated by 100 ps to eliminate any possible correlation between them. The error bars reported in Table I are the standard error of the mean. The reference FPT was obtained by continuous simulations fully considering longtime correlations. It can be seen that heavier particles have longer FPTs. While the 3200 g/mol particle has a 2320 ps FPT, the lightest 5 g/mol particle has a FPT of 1045 ps. The longer FPT for heavier particles is not surprising since such particles are moving at a lower average velocity at the same temperature. At the same time, it is interesting to note that a 20% increase in FPT is observed for each factor of 8 increase in mass from 5 g/mol to 320 g/mol. A large 60% increase in FPT is observed for the 3200 g/mol particle when compared to the 320 g/mol particle, the long acceleration period for the 3200 g/mol particle to reach the terminal speed is likely to be a contributing factor to the much longer time needed to travel 4 nm.

TABLE I. The mean first passage time in ps for tagged particles of various masses. The reference values are obtained from continuous simulations. t_r indicates restart intervals at which new random velocities are assigned to each particle in the system.

t_r (ps)	Mass of tagged particle (g/mol)			
	5	40	320	3200
Reference ($t_r = \infty$)	1045 $\pm$ 8	1210 $\pm$ 13	1478 $\pm$ 10	2320 $\pm$ 14
$t_r = 12.5$	1063 $\pm$ 6	1238 $\pm$ 6	1588 $\pm$ 13	4090 $\pm$ 42
$t_r = 25.0$	1037 $\pm$ 6	1220 $\pm$ 6	1534 $\pm$ 10	3147 $\pm$ 16
$t_r = 50.0$	1021 $\pm$ 7	1244 $\pm$ 12	1489 $\pm$ 17	2686 $\pm$ 10
$t_r = 100.0$	1027 $\pm$ 10	1212 $\pm$ 6	1478 $\pm$ 7	2520 $\pm$ 12

In order to understand the effect of ignoring correlations, we measured the FPT by running trajectories where the correlations were removed at constant intervals. This was accomplished by restarting the trajectories with the velocity of each particle randomized. A series of restarting intervals were studied ranging from 12.5 ps to 100 ps. For the 5 g/mol and 40 g/mol particles, the C_{vv} have reached their respective asymptotes even for the shortest 12.5 ps intervals studied. For the 320 g/mol tagged particle, the C_{vv} is already very small at 0.17% after 12.5 ps and reaches the longtime plateau at 25 ps.

With the lighter 5 g/mol and 40 g/mol tagged particles, our results indicate that discarding the longtime correlation has negligible effect on the FPT. Considering that the one sigma error bar reported corresponds to a 68% confidence interval, all the estimated mean FPTs agree with each other within statistical noise. For the 320 g/mol tagged particle, while an excellent agreement is obtained with a 100 ps restart frequency, the measured FPT increases as the correlations are discarded more frequently. With a restart interval of 25 ps, the FPT is overestimated by 56 ps, which is significantly larger than the statistical noise. Although 56 ps is a fairly small deviation, the residual of C_{vv} is only 0.057% after 25 ps. The test case with the heavy 3200 g/mol mass shows a quite significant 200 ps error, almost 10% over-estimation, even with the 100 ps restart time. This is fairly large considering the residual velocity correlation is only 0.25%. With a 12.5 ps restart time, the error grows to almost 100% despite an undetectable $C_{ff}(t)$ and C_{vv} being about 0.2. We note the Ar test system was simulated far above the critical point, where strongly oscillatory forces from the environment lead to a very fast decorrelation. For a more complex system with hydrogen bonds and their associated slower dynamics, a much slower decorrelation is anticipated. Thus, the timescale used for our model system should not be taken literally. This study suggests that neglecting longtime correlation could, in certain cases, lead to fairly appreciable errors and should be done with care.

MSM is widely used to derive longtime evolution from a collection of short trajectories. With this approach, short-trajectories are simulated independently with no correlation between them. A transition matrix is constructed by studying transition probabilities between microstates after a lag time. The Chapman-Kolmogorov property is assumed for the application of the transition matrix. In other words, the system is assumed to be Markovian after each lag time.

Although the MSM assumes the system is not correlated after a lag time, it is not the only method with such an approximation. Any algorithms based on construction of overall kinetics from short-trajectories^{12-17,39-42} potentially suffer the loss of longtime correlation. It is not within the scope of the present work to do a general survey of short-trajectory based methods. In this study, we will investigate the effect of forcing Markovianity on a system with weak longtime correlations with the MSM.

During the application of a MSM, the validation of the Markovian approximation can be tested with several

FIG. 3. Schematic representation of microstate assignments for the construction of MSM. Particles always start from state 41. Transitions from state 81 to any other state are not allowed.

methods, such as the Swope-Pitera eigenvalues,³⁰ information entropy,²⁹ or Chapman-Kolmogorov analysis,^{19,26,30,43,44} Such validation methods are much more practical for complex systems than force or velocity correlation functions. In this work, the lack of Markovianity will be checked with the Chapman-Kolmogorov analysis, which is one of the most straight-forward methods for validating Markovianity.

For each tagged particle, a series of MSM matrices were constructed with different lag times from 5 ps to 50 ps. Each MSM transition matrix was constructed from 10 000 simulations, where particles are allowed to travel a full box length. The trajectories used to construct the MSM matrices were restarted once every 100 ps with random velocities used at restarts. We made this decision to better reflect real world application of MSM, where the full evolution of interest is generally longer than what can be done in one continuous simulation. Thus, the construction of the MSM is typically based on shorter trajectories. We note that for all test systems, the correlation at 100 ps is already sufficiently small. A deviation for the observed FPT is only seen for the 3200 g/mol tagged particle. Using shorter trajectories to construct a MSM will only lead to larger deviations and are thus not tested.

For our model system, the MSM matrices were constructed using 81 microstates. For each of the 10 000 trajectory, the particle always start at state 41. All the states are

evenly spaced according to the displacement of the tagged particle along the z-direction from the initial position. If the particle travels along the direction of the driving force by 1 Å, the particle will be in microstate 42. If the particle travels by 40 Å or more opposing to the driving force, the particle will be in state 1. Only a small number of trajectories visit state 1. However, it is sufficient to obtain a reasonable transition vector from state 1. If the particle passes the 4 nm mark along the direction of the driving force, the particle will be in state 81. Since the FPT is to be measured, no transitions out of state 81 are allowed. The schematic picture of the state assignments is provided in Fig. 3. We refrain from using too fine a state density since increasing that state density will make the determination of converged matrix elements challenging. Too coarse a grid density is generally not recommended due to discretization errors.²⁶ We do not lump microstates into macrostates to avoid lumping errors^{30,45,46} that may lead to additional complications not directly related to the longtime correlation problem we are trying to investigate. We believe 81 states should be sufficiently high, since the width of each bin is significantly smaller than the diameter of the Ar atom. The stationary eigen vector of the MSM transition matrix is confirmed to be a one in state 81 and zero in all other states. This is expected since after “equilibrium,” all particles will pass the 4 nm mark.

As mentioned previously, the driven system in the full phase space including position and momentum degrees

FIG. 4. The Chapman-Kolmogorov test for tagged particles of various masses. The population of the initial microstate (41) is shown for the simulation for constructing the MSM matrices and those predicted with MSM with various lag times τ . Panels (a)–(d) correspond to particle masses of 5 g/mol, 40 g/mol, 320 g/mol, and 3200 g/mol, respectively.

of freedom is actually Markovian. The removal of the momentum of the tagged particle leads to a loss of memory, making the position space evolution Non-Markovian. Conceivably, a MSM can be constructed keeping velocity of the tagged particle. However, for the purpose of this study, we will intentionally discard the memory by applying the Markov state approximation and discard momentum information.

Figure 4 compares the MD probability of the tagged particle in the initial state to the probability predicted with MSM. For MSM, the initial state is state 41. The probability reported is the survival probability in state 41 predicted by MSM. For MD, this probability reflects the fraction of particles that travelled less than $0.5 \text{ \AA}$ to either direction since time starts. Such a test of Chapman-Kolmogorov property is frequently referred to as the Chapman-Kolmogorov analysis.^{19,26,30,43,44} The reference curves were produced with the 100 ps restart simulations used to construct the MSM. It is clear that all MSM predictions agree within the error of the MD simulations for any tagged particle with masses up to 320 g/mol. For the 3200 g/mol mass, the Chapman-Kolmogorov test indicates good agreement only for the 50 ps lag time, and the agreement with the 25 ps lag time can be considered fair. The agreement with the 12.5 ps lag time should probably be considered poor and would indicate a problem.

The FPT obtained with all MSM studies are summarized in Table II. It is not surprising that a MSM works well with the 5 g/mol and 40 g/mol particles without much inertia. All the FPT predictions agree with the reference calculations within statistical uncertainty. For the 320 g/mol tagged particle, the FPT predicted by the MSM with a 5 ps lag time is about 49 ps longer than the reference values. Although 49 ps is not significant, considering the Chapman-Kolmogorov test suggests excellent Markovianity for this test case. The fact that the error is larger than statistical uncertainty is still worth mentioning.

The MSM predicted FPTs for the 3200 g/mol tagged particle are most interesting. For this mass, the 100 ps trajectories on which the MSM were trained has a 2520 ps FPT. Thus, the 2520 ps FPT is the best the MSM could possibly achieve. The MSM predicts a FPT of 2495 ps and 2554 ps

with a 50.0 ps and 25.0 ps lag time, respectively. This very good agreement considering the Chapman-Kolmogorov test shows a minor deviation. With a lag time of 12.5 ps, the MSM only results in a 410 ps overestimation of the FPT. This is much better than the over-estimation of 1770 ps observed when velocity correlation is discarded every 12.5 ps, as shown in Table I. Even the 5 ps lag time MSM simulation leads to a much smaller 658 ps overestimation when compared to simply discarding velocity correlation at a longer 12.5 ps interval.

The much better-than-expected performance of the MSMs can only be explained by considering that the 100 ps trajectories based on which the MSM transition matrices were constructed have their motions fully correlated during the entire simulation length. The tagged particles are continuously accelerated by the driving force during the whole 100 ps. This led to accelerated transitions between microstates, which are built into the transition probability matrix elements. Consequently, the use of MSMs led to better estimates of the FPT even when shorter lag times are used.

IV. SUMMARY AND CONCLUSION

The effect of removal of weak longtime correlation was studied by simulating the motion of a tagged particle under a weak constant driving force. The particle is surrounded by super-critical Ar atoms at high density. The driving force is only about 1% of the random force experienced by each particle due to thermal fluctuations. The tagged particle is allowed to have a range of masses from $1/8$ to 80 times that of an Ar atom. The higher mass simulations will likely better mimic macromolecules with large inertia under a weak driving force.

Regardless of the particle masses studied, the force auto-correlation function shows little evidence of long-time correlation as a result of the frictional force exerted by the dense environment. Particles with different inertia demonstrate similar longtime decay of the force auto-correlation functions. These observations indicate that the force autocorrelation function is a poor metric for detecting longtime correlations. A small tail can be observed for the velocity autocorrelation functions, which depends on the mass of the particle. However, for the model system studied, the longtime asymptote is still very small, being no more than 2.5×10^{-3} , even for the most massive 3200 g/mol tagged particle.

The FPT for the tagged particle to travel a distance of 4 nm along the direction of the driving force is measured. When the particles have little inertia, discarding the longtime correlation has little influence on its FPT. When the particle has a fairly large inertia, discarding longtime correlation starts to affect the observed FPT, slowing down the motion of the particle along the direction of the driving force. Appreciable change of the FPT can be observed even when the velocity correlation function shows negligible residual correlation. For example, for the 3200 g/mol test case, while randomizing velocity once every 50 ps leads to a 16% overestimation of the FPT, the velocity correlation being discarded is only about 0.49% at the 50 ps mark.

TABLE II. The mean FPT in ps for tagged particles of various masses predicted by MSMs. The transition matrices for MSM evolution were constructed using a series of trajectories with a 100 ps restart time. τ indicates the lag time chosen for the MSM. Both the true reference FPT and the FPT values based on a 100 ps restart are reported since the latter is the best the MSM is expected to be able to reproduce.

τ (ps)	Mass of tagged particle (g/mol)			
	5	40	320	3200
Reference	1045 $\pm$ 8	1210 $\pm$ 13	1478 $\pm$ 10	2320 $\pm$ 14
$t_r = 100.0$	1027 $\pm$ 10	1212 $\pm$ 6	1478 $\pm$ 7	2520 $\pm$ 12
$\tau = 5.0$	1029	1210	1527	2978
$\tau = 10.0$	1054	1231	1513	2814
$\tau = 12.5$	1063	1240	1512	2730
$\tau = 25.0$	...	...	...	2554
$\tau = 50.0$	...	...	...	2495

It is interesting that simulation with MSMs shows very good agreement with the reference FPT even with very short lag times. For example, while forcing Markovianity by randomizing velocities once every 12.5 ps will lead to a 100% overestimation of FPT, the MSM with a 12.5 lag time only lead to a 18% overestimation. Such a better-than-anticipated performance is a result of the transition matrices being constructed with the much longer 100 ps trajectories. Although the way time-dependent probabilities are derived in the MSM assumes that the system completely loses memory after the lag time, in practice, a cancelation of errors could arise when the trajectory for transition matrix construction is correlated for a duration much longer than the lag time. This leads to a cautious optimism for using MSM with a high time resolution but constructed longer trajectories.⁴⁷

We note that the model system decorrelates quickly due to the lack of slow local relaxation, such as those present in hydrogen-bonded systems or with local frustrations. The specific restart intervals and lag times used for this work should not be taken literally for more complex systems in a more viscous environment. However, the insight into the effect of longtime correlation should be applicable to more complex systems with a much slower dynamical relaxation.

SUPPLEMENTARY MATERIAL

See [supplementary material](#) for a derivation of Eq. (3) and the force autocorrelation function for a system with a strong driving force.

ACKNOWLEDGMENTS

The research was supported by the National Institutes of Health (NIH) (Grant No. NIGMS 1R01GM120578) and the Arkansas Bioscience Institute. The computer resources for this study were provided by the Arkansas High Performance Computational Center through Grant No. MRI-R2 0959124 provided by the NSF.

REFERENCES

- ¹K. A. Dill and H. S. Chan, *Nat. Struct. Biol.* **4**, 10–19 (1997).
- ²D. Bryngelson Joseph, N. Onuchic José, D. Succi Nicholas, and G. Wolynes Peter, *Proteins: Struct., Funct., Bioinf.* **21**(3), 167–195 (1995).
- ³G. Bussi, D. Donadio, and M. Parrinello, *J. Chem. Phys.* **126**(1), 014101 (2007).
- ⁴D. Hamelberg, J. Mongan, and J. A. McCammon, *J. Chem. Phys.* **120**(24), 11919–11929 (2004).
- ⁵A. F. Voter, *Phys. Rev. Lett.* **78**(20), 3908–3911 (1997).
- ⁶A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.* **99**(20), 12562–12566 (2002).
- ⁷R. C. Bernardi, M. C. R. Melo, and K. Schulten, *Biochim. Biophys. Acta, Gen. Subj.* **1850**(5), 872–877 (2015).
- ⁸S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, *Science* **220**(4598), 671 (1983).
- ⁹C. Tsallis and D. A. Stariolo, *Physica A* **233**(1), 395–406 (1996).
- ¹⁰H. Szu and R. Hartley, *Phys. Lett. A* **122**(3), 157–162 (1987).
- ¹¹N. Metropolis and S. Ulam, *J. Am. Stat. Assoc.* **44**(247), 335–341 (1949).
- ¹²A. K. Faradjian and R. Elber, *J. Chem. Phys.* **120**(23), 10880–10889 (2004).
- ¹³D. Shalloway and A. K. Faradjian, *J. Chem. Phys.* **124**(5), 054112 (2006).
- ¹⁴R. Elber, *Biophys. J.* **92**(9), L85–L87 (2007).
- ¹⁵A. M. A. West, R. Elber, and D. Shalloway, *J. Chem. Phys.* **126**(14), 145104 (2007).
- ¹⁶E. Vanden-Eijnden and M. Venturoli, *J. Chem. Phys.* **130**(19), 194101 (2009).
- ¹⁷J. M. Bello-Rivas and R. Elber, *J. Chem. Phys.* **142**(9), 094102 (2015).
- ¹⁸V. S. Pande, K. Beauchamp, and G. R. Bowman, *Methods* **52**(1), 99–105 (2010).
- ¹⁹J. Chodera, W. Swope, J. W. Pitera, and K. Dill, *Multiscale Model. Simul.* **5**(4), 1214–1226 (2006).
- ²⁰W. Wang, S. Cao, L. Zhu, and X. Huang, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **8**, e1343 (2018).
- ²¹B. E. Husic and V. S. Pande, *J. Am. Chem. Soc.* **140**(7), 2386–2396 (2018).
- ²²N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland Physics Publishing, New York, 1992).
- ²³C. Schütte and M. Sarich, *Metastability and Markov State Models in Molecular Dynamics: Modeling, Analysis, Algorithmic Approaches* (American Mathematical Society, 2013).
- ²⁴G. R. Bowman, V. S. Pande, and F. Noé, *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation* (Springer, The Netherlands, 2014).
- ²⁵A. A. Markov, *The Theory of Algorithms* (USSR Academy of Sciences, Moscow–Leningrad, 1954).
- ²⁶J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, *J. Chem. Phys.* **134**(17), 174105 (2011).
- ²⁷F. Noé, *J. Chem. Phys.* **128**(24), 244103 (2008).
- ²⁸F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl, *Proc. Natl. Acad. Sci. U. S. A.* **106**(45), 19011 (2009).
- ²⁹S. Park and V. S. Pande, *J. Chem. Phys.* **124**(5), 054118 (2006).
- ³⁰W. C. Swope, J. W. Pitera, and F. Suits, *J. Phys. Chem. B* **108**(21), 6571–6581 (2004).
- ³¹H. Mori, *Prog. Theor. Phys.* **33**(3), 423–455 (1965).
- ³²D. J. Evans and G. P. Morriss, *Statistical Mechanics of Nonequilibrium Liquids* (Academic Press, London, 1990).
- ³³R. Zwanzig, in *Systems Far From Equilibrium*, edited by L. Garrido (Interscience, New York, 1980).
- ³⁴N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark, and F. v. G. Wilfred, *Eur. Biophys. J.* **40**(7), 843–856 (2011).
- ³⁵L. Verlet and D. Levesque, *Physica* **36**, 254 (1967).
- ³⁶J.-P. Hansen and L. Verlet, *Phys. Rev.* **184**(1), 151–161 (1969).
- ³⁷S. Nosé, *Mol. Phys.* **52**(2), 255–268 (1984).
- ³⁸W. G. Hoover, *Phys. Rev. A* **31**(3), 1695–1697 (1985).
- ³⁹T. S. van Erp, D. Moroni, and P. G. Bolhuis, *J. Chem. Phys.* **118**(17), 7762–7774 (2003).
- ⁴⁰D. Moroni, P. G. Bolhuis, and T. S. van Erp, *J. Chem. Phys.* **120**(9), 4055–4065 (2004).
- ⁴¹G. A. Huber and S. Kim, *Biophys. J.* **70**(1), 97–110 (1996).
- ⁴²B. W. Zhang, D. Jasnow, and D. M. Zuckerman, *J. Chem. Phys.* **132**(5), 054107 (2010).
- ⁴³R. D. Malmstrom, C. T. Lee, A. T. Van Wart, and R. E. Amaro, *J. Chem. Theory Comput.* **10**(7), 2648–2657 (2014).
- ⁴⁴J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill, and W. C. Swope, *J. Chem. Phys.* **126**(15), 155101 (2007).
- ⁴⁵K. H. Hoffmann and P. Salamon, *Appl. Math. Lett.* **22**(9), 1471–1475 (2009).
- ⁴⁶J. D. Chodera and F. Noé, *Curr. Opin. Struct. Biol.* **25**, 135–144 (2014).
- ⁴⁷T. J. Lane, G. R. Bowman, K. Beauchamp, V. A. Voelz, and V. S. Pande, *J. Am. Chem. Soc.* **133**(45), 18413–18419 (2011).