The Journal of Chemical Physics

RESEARCH ARTICLE | JUNE 16 2015

Preserving correlations between trajectories for efficient path sampling FREE

Todd R. Gingrich ; Phillip L. Geissler

J. Chem. Phys. 142, 234104 (2015)
https://doi.org/10.1063/1.4922343
CHORUS

AIP Publishing

THE JOURNAL OF CHEMICAL PHYSICS **142**, 234104 (2015)

Preserving correlations between trajectories for efficient path sampling

Todd R. Gingrich[1] and Phillip L. Geissler[1,2]
[1]*Department of Chemistry, University of California, Berkeley, California 94720, USA*
[2]*Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

(Received 21 April 2015; accepted 22 May 2015; published online 16 June 2015)

Importance sampling of trajectories has proved a uniquely successful strategy for exploring rare dynamical behaviors of complex systems in an unbiased way. Carrying out this sampling, however, requires an ability to propose changes to dynamical pathways that are substantial, yet sufficiently modest to obtain reasonable acceptance rates. Satisfying this requirement becomes very challenging in the case of long trajectories, due to the characteristic divergences of chaotic dynamics. Here, we examine schemes for addressing this problem, which engineer correlation between a trial trajectory and its reference path, for instance using artificial forces. Our analysis is facilitated by a modern perspective on Markov chain Monte Carlo sampling, inspired by non-equilibrium statistical mechanics, which clarifies the types of sampling strategies that can scale to long trajectories. Viewed in this light, the most promising such strategy guides a trial trajectory by manipulating the sequence of random numbers that advance its stochastic time evolution, as done in a handful of existing methods. In cases where this "noise guidance" synchronizes trajectories effectively, as the Glauber dynamics of a two-dimensional Ising model, we show that efficient path sampling can be achieved for even very long trajectories. [http://dx.doi.org/10.1063/1.4922343]

I. INTRODUCTION

Recent advances in non-equilibrium statistical mechanics have given fresh perspectives on computational procedures applied to fluctuating molecular systems. The Jarzynski work relation and the Crooks fluctuation theorem, for instance, provide routes to compute equilibrium quantities from non-equilibrium measurements.[1–6] Here, we demonstrate that traditional Metropolis-Hastings Markov Chain Monte Carlo (MCMC) can be similarly viewed as a procedure to extract equilibrium sampling from generically non-equilibrium processes. Monte Carlo (MC) trial moves drive a system away from the steady state distribution, and an entropy production can be assigned to these driven transformations. This interpretation provides an elegant way to understand challenges encountered in MCMC sampling, one that is especially revealing for MCMC sampling of trajectories. Path sampling methods suffer routinely from profound inefficiency when trajectories of interest become long. From a non-equilibrium perspective on MCMC, we provide simple and quantitative ways to understand the inefficiency.

Importance sampling of trajectories has enabled studies of a myriad of dynamical processes in physics and chemistry.[7–12] In particular, reaction rates and mechanisms can be found by transition path sampling (TPS), which examines the subensemble of trajectories that complete a reaction.[13] The practicality of TPS depends intimately on the design of the MC move set. Namely, the moves must generate correlated trajectories so that a trial trajectory is likely to exhibit similar dynamical behavior as the previously sampled trajectory. Chaotic divergence and microscopic reversibility of equilibrium dynamics inform the construction of two such moves, the so-called "shooting" and "shifting" moves.[13] These methods generate correlated trajectories by propagating alternative histories from highly correlated initial configurations. For sufficiently short trajectories, the imposed correlation at one time serves to strongly correlate the trajectories at all times. Long trajectories, however, are problematic: trial trajectories either lose all useful correlation with the reference path, or else they coincide so closely with the reference that changes are impractically small.[14] In both cases the efficiencies of shooting and shifting moves plummet as trajectories grow longer. Sampling trajectories that involve slow molecular rearrangements and diffusive processes stand to benefit significantly from alternative methods of generating correlated trajectories.

We consider three different ways to guide long trajectories: introducing auxiliary forces; selecting among a series of short trial segments, as in Steered Transition Path Sampling (STePS);[15] and advancing stochastic integrators with correlated random numbers (which we refer to as "noise guidance").[16,17] Of the three, only noise guidance yields a MC entropy production which is subextensive in the trajectory length. The other schemes, which accumulate extensive entropy production, cannot efficiently extend to sampling of long trajectories. Strong noise guidance is not, however, a panacea; correlated noises need not imply correlated trajectories. We illustrate this point by considering Glauber dynamics of a two-dimensional Ising model and Langevin dynamics of a two-dimensional Weeks-Chandler-Andersen (WCA) fluid. Only when microscopic degrees of freedom have a small number of discrete possibilities, as in the lattice dynamics, it is possible to generate correlated long time scale trajectories by tuning the noise.

The structure of the paper is as follows. First, we introduce and discuss the perspective of MC moves as non-equilibrium

processes which produce entropy, detailing the consequences of constraints analogous to fluctuation theorems and the second law of thermodynamics. Next, we review transition path sampling in stochastic dynamics and demonstrate the challenge posed by long trajectories in the context of trajectory sampling of a one-dimensional random walker. We then analyze alternative strategies to correlate long trajectories of a one-dimensional single particle system and of a two-dimensional Ising model. Finally, we explore how strongly noise guidance correlates trajectories in example systems, and then conclude.

II. MARKOV CHAIN MONTE CARLO ENTROPY PRODUCTION

We start by discussing a very general perspective on traditional Metropolis-Hastings MCMC sampling.[18,19] Consider the problem of sampling a configuration, $\mathbf{x}$, according to probability distribution $P(\mathbf{x})$. For example, $\mathbf{x}$ could be a vector of the coordinates and momenta of N hard spheres, the state of spins in an Ising model, or all coordinates of a classical trajectory. The Metropolis-Hastings algorithm generates a Markov chain, which can be thought of as a dynamics through configuration space with the steady-state distribution $P(\mathbf{x})$. This dynamics obeys detailed balance but is not necessarily physical.

One typically splits each Monte Carlo move into two steps. First, a change from $\mathbf{x}$ to a new state $\tilde{\mathbf{x}}$ is proposed according to a generation probability, $P_{\text{gen}}[\mathbf{x} \to \tilde{\mathbf{x}}]$. Throughout this paper we will refer to $\mathbf{x}$ as a reference and $\tilde{\mathbf{x}}$ as the trial. This trial is conditionally accepted with probability

$$P_{\text{accept}}[\mathbf{x} \to \tilde{\mathbf{x}}] = \min\left[1, e^{-\omega}\right], \tag{1}$$

where

$$\omega = \ln \frac{P(\mathbf{x})P_{\text{gen}}[\mathbf{x} \to \tilde{\mathbf{x}}]}{P(\tilde{\mathbf{x}})P_{\text{gen}}[\tilde{\mathbf{x}} \to \mathbf{x}]}. \tag{2}$$

Together, these two steps ensure detailed balance, guaranteeing that the equilibrium distribution $P(\mathbf{x})$ is stationary under the MC protocol. Lacking the conditional acceptance step, such a MC procedure would generally drive a system away from its equilibrium distribution. We find it instructive to view this notional process as a genuine non-equilibrium transformation, one that would generate nonzero entropy in most cases. In the formalism of stochastic thermodynamics, the resulting entropy production corresponds precisely to the quantity ω defined in Eq. (2).[20]

The MC acceptance step effectively filters realizations of this non-equilibrium process, with a bias towards low values of ω. By construction, the bias exactly negates the tendency of trial move generation to drive a system out of equilibrium. From this perspective, the countervailing tendencies of proposal and acceptance are akin to the operation of a Maxwellian demon, which by contrast filter realizations of *equilibrium* dynamics with a bias that creates a *non-equilibrium* state.

Viewing the procedure in the language of entropy production distributions reveals an important asymmetry of $P(\omega)$. Following the more general demonstration of an entropy production fluctuation theorem,[21] note that

$$\begin{aligned} P(\omega) &= \int d\mathbf{x}\, d\tilde{\mathbf{x}}\, P(\mathbf{x})P_{\text{gen}}(\mathbf{x} \to \tilde{\mathbf{x}})\delta(\omega - \omega(\mathbf{x}, \tilde{\mathbf{x}})) \\ &= \int d\mathbf{x}\, d\tilde{\mathbf{x}}\, e^{\omega(\mathbf{x},\tilde{\mathbf{x}})}P(\tilde{\mathbf{x}})P_{\text{gen}}(\tilde{\mathbf{x}} \to \mathbf{x})\delta(\omega + \omega(\tilde{\mathbf{x}}, \mathbf{x})) \\ &= e^{\omega}P(-\omega), \end{aligned} \tag{3}$$

with $\omega(\mathbf{x}, \tilde{\mathbf{x}})$ representing the entropy produced by a proposal move from $\mathbf{x}$ to $\tilde{\mathbf{x}}$ and δ denoting the Dirac delta function. We are more likely to propose moves with positive entropy production than we are to choose their negative counterparts. The straightforward corollary, $\langle\omega\rangle \geq 0$, is by analogy a statement of the second law, and the equality is satisfied if and only if $P(\omega) = \delta(\omega)$. A further consequence of Eq. (3) relates the MC acceptance rate to the probability of attempting a move with a negative value of ω, which we call $p_<$. Specifically,

$$\left\langle P_{\text{accept}}\right\rangle = \int d\omega\, P(\omega) \min\left[1, e^{-\omega}\right] = 2p_<, \tag{4}$$

which has been noted in the related context of replica exchange Monte Carlo.[22] As $\langle\omega\rangle$ increases, $p_<$, and therefore $\left\langle P_{\text{accept}}\right\rangle$, tends to decrease. We will see that $\langle\omega\rangle$ scales with the number of driven degrees of freedom, such that Monte Carlo sampling of chain molecules or of long trajectories becomes especially challenging.

We focus below on the sampling of dynamical pathways (rather than individual configurations). In this case, ω is an "entropy production" only by analogy, since the "non-equilibrium transformations" affected by MC trial moves occur in the more abstract space of trajectories. Lessons from Eq. (3) are nevertheless illuminating in the context of this abstract space.

III. TRANSITION PATH SAMPLING WITH STOCHASTIC DYNAMICS

A. Trajectory space and trajectory subensembles

Let us now specialize to the sampling of discrete-time stochastic trajectories with a fixed number of steps, t_{obs}. The probability of observing a trajectory, $\mathbf{x}(t) \equiv \left\{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{t_{\text{obs}}}\right\}$, can be written as

$$P_0[\mathbf{x}(t)] \propto \rho_{\text{init}}(\mathbf{x}_0) \prod_{t=0}^{t_{\text{obs}}-1} p(\mathbf{x}_t \to \mathbf{x}_{t+1}), \tag{5}$$

where ρ_{init} is a distribution for the initial time point, frequently an equilibrium or steady state distribution. The probability of each time propagation step is denoted as $p(\mathbf{x}_t \to \mathbf{x}_{t+1})$, the form of which depends on details of the stochastic dynamics. We refer to this propagation as the natural dynamics. Representative trajectories can be generated by sampling the initial state and propagating natural dynamics.

In many contexts, it is useful to study a biased trajectory ensemble constructed to highlight particular rare events. Common examples include the reactive subensemble,

$$P_{\text{reactive}}[\mathbf{x}(t)] \propto P_0[\mathbf{x}(t)]h_{\text{A}}(\mathbf{x}_0)h_{\text{B}}(\mathbf{x}_{t_{\text{obs}}}), \tag{6}$$

and the so-called tilted ensemble,

$$P_{\text{tilted}}[\mathbf{x}(t), s] \propto P_0[\mathbf{x}(t)]e^{-sK[\mathbf{x}(t)]}. \tag{7}$$

In the former case, h_A and h_B are indicator functions which constrain the endpoints of the trajectory to fall in regions of phase space corresponding to reactants and products of a chemical reaction or other barrier crossing process.[13] In the latter case, $K[\mathbf{x}(t)]$ is an order parameter reporting on dynamical properties of the trajectory (e.g., the current,[23] activity,[12,24] or entropy production[25–27]) and s sets the strength of bias.[28] These biased ensembles highlight classes of trajectories only rarely sampled by the natural dynamics. To effectively sample them, a Markov chain of correlated trajectories is constructed. The correlations between subsequent steps of the Markov chain ensure that newly generated trajectories are likely to share the rare features that made the prior trajectory a good representative of the biased ensemble.

B. Sampling with shooting moves

One of the most general and effective methods for generating a trial trajectory is the shooting move, which is particularly well-suited to sampling equilibrium dynamics.[13] The move proceeds as follows. A discrete time, t_{shoot}, between 0 and t_{obs} is uniformly selected and designated the shooting time. The state of the system at t_{shoot}, perhaps slightly modified, is then propagated forward and backward in time with natural dynamics to yield a trial trajectory, $\tilde{\mathbf{x}}(t)$. The probability of generating this trial takes the form

$$P_{\text{gen}}[\mathbf{x}(t) \to \tilde{\mathbf{x}}(t)] = p_{\text{gen}}\left(\mathbf{x}_{t_{\text{shoot}}} \to \tilde{\mathbf{x}}_{t_{\text{shoot}}}\right) \times \prod_{t=0}^{t_{\text{shoot}}-1} \bar{p}(\mathbf{x}_{t+1} \to \mathbf{x}_t) \prod_{t=t_{\text{shoot}}}^{t_{\text{obs}}-1} p(\mathbf{x}_t \to \mathbf{x}_{t+1}), \tag{8}$$

where $\bar{p}$ is the transition probability for time-reversed dynamics and p_{gen} is the probability of the perturbation at the shooting time. In the language of Sec. II, the entropy produced by this trial move is given by

$$\omega = \ln \frac{\rho_{\text{init}}(\mathbf{x}_0) h_A(\mathbf{x}_0) h_B(\mathbf{x}_{t_{\text{obs}}}) p_{\text{gen}}(\mathbf{x}_{t_{\text{shoot}}} \to \tilde{\mathbf{x}}_{t_{\text{shoot}}})}{\rho_{\text{init}}(\tilde{\mathbf{x}}_0) h_A(\tilde{\mathbf{x}}_0) h_B(\tilde{\mathbf{x}}_{t_{\text{obs}}}) p_{\text{gen}}(\tilde{\mathbf{x}}_{t_{\text{shoot}}} \to \mathbf{x}_{t_{\text{shoot}}})} + \sum_{t=0}^{t_{\text{shoot}}-1} \ln \frac{p(\mathbf{x}_t \to \mathbf{x}_{t+1}) \bar{p}(\tilde{\mathbf{x}}_{t+1} \to \tilde{\mathbf{x}}_t)}{\bar{p}(\mathbf{x}_{t+1} \to \mathbf{x}_t) p(\tilde{\mathbf{x}}_t \to \tilde{\mathbf{x}}_{t+1})}. \tag{9}$$

For trial trajectories to be accepted by the MCMC scheme, ω must be small. However, when p and $\bar{p}$ are not equal (as is the case in driven processes), the sum in Eq. (9) has order t_{obs} nonvanishing terms.[29] Consequently, $\langle\omega\rangle$ scales linearly with t_{obs}, and $P(\omega)$ adopts the long-time form,

$$P(\omega) \sim \exp\left[-t_{\text{obs}} I(\omega/t_{\text{obs}})\right], \tag{10}$$

with large deviation rate function $I(\omega/t_{\text{obs}})$. From this asymptotic expression for the entropy produced by a TPS move, one might generally expect that the corresponding acceptance rate decreases exponentially as t_{obs} grows long.

This extensive growth of $\langle\omega\rangle$ with time has an important and general exception, namely, the case of microscopically reversible dynamics. Under those conditions, the sum in Eq. (9) vanishes and the only entropy production is contributed from the endpoints of the trajectory (e.g., h_A and h_B). Since this entropy production is subextensive in time, long trajectories appear no more difficult to sample than short ones. The acceptability of trial trajectories, however, is also subject to biases like those expressed in Eqs. (6) and (7). Because long trajectories typically decorrelate strongly from one another, the rare, biased qualities of a reference trajectory (e.g., reactivity or inactivity) are recapitulated in the trial path with a probability that also decays with t_{obs}.

We conclude that the challenges for efficiently sampling long trajectories are twofold. The TPS move must produce entropy that is subextensive in observation time or the method will not scale to long trajectories. Additionally, one must preserve strong correlations between $\mathbf{x}(t)$ and $\tilde{\mathbf{x}}(t)$, so that rare properties of interest are retained in the trial trajectory. In Sec. IV, we show that these two goals are often conflicting. In particular, we examine three general schemes for engineering correlations between reference and trial trajectories in shooting-like moves. Two of the schemes fail to exhibit subextensive entropy production scaling while the remaining scheme can only maintain strong trajectory correlations in special cases.

IV. GUIDED DYNAMICS OF A 1D RANDOM WALKER

We explore the three methods for trial trajectory generation in the specific context of a one-dimensional discrete-time random walker with equation of motion

$$x_{t+1} = x_t + \xi_t, \tag{11}$$

where at the tth time step, the noise ξ_t is drawn from the normal distribution with zero mean and variance σ^2. As a simple illustration focusing on the effects of entropy production, suppose we want to sample the unbiased trajectory distribution,

$$P_0[x(t)] \propto \delta(x_0) \exp\left[-\sum_{t=0}^{t_{\text{obs}}-1} \frac{(x_{t+1} - x_t)^2}{2\sigma^2}\right], \tag{12}$$

where the initial position is set to zero without loss of generality.[30] To construct a reference trajectory $x(t)$, we draw a value for ξ_t at each time step and propagate the walker's position according to Eq. (11). A trial trajectory is then generated by evolving dynamics from the same initial configuration (with a different realization of the noise or perhaps even a different equation of motion).

We imagine that it is desirable for the trial trajectory to retain a significant correlation with the reference path. This goal is motivated by the challenges of sampling biased ensembles as discussed above, but for the sake of simplicity, we do not include such a bias here. To ensure this correlation, we employ shooting moves that differ from the conventional procedure described in Sec. III B. Specifically, we implement and scrutinize three distinct ways to engineer correlation over long times: (a) adding artificial forces that pull the trial trajectory closer to the reference, (b) preferentially selecting among sets of otherwise unbiased short path segments, or (c) using correlated histories of noises. We assess the influence of these three biasing methods on the MCMC efficiency by characterizing the distribution $P(\omega)$.

A. Guiding forces

We first consider effecting correlations with guiding forces, i.e., artificial contributions to the effective potential that tend to lead the trial trajectory toward the reference. This strategy is equivalent to using steered molecular dynamics[31] to generate new trajectories. The trial trajectory $\tilde{x}(t)$ is grown with the equation of motion

$$\tilde{x}_{t+1} = \tilde{x}_t + \tilde{\xi}_t + k(x_t - \tilde{x}_t). \tag{13}$$

We denote $\tilde{\xi}_t$ as the trial trajectory noise at time step t, also drawn from a Gaussian with mean zero and variance σ^2. The linear spring constant k adjusts the strength of correlation between reference and trial trajectories. The probability that this guided dynamics generates a particular trial from the reference is given by

$$P_{\text{gen}}[x(t) \to \tilde{x}(t)] \propto \exp\left[-\sum_{t=0}^{t_{\text{obs}}-1} \frac{(\tilde{x}_{t+1} - \tilde{x}_t - k(x_t - \tilde{x}_t))^2}{2\sigma^2}\right]. \tag{14}$$

The entropy production associated with the trial move depends also on the probability of generating the reverse TPS move, growing the reference trajectory with extra forces pulling it close to the trial. It is straightforward to compute ω from Eq. (2),

$$\omega = -\frac{k}{\sigma^2} \sum_{t=0}^{t_{\text{obs}}-1} (x_t - \tilde{x}_t)(x_{t+1} + \tilde{x}_{t+1} - x_t - \tilde{x}_t). \tag{15}$$

In this approach, ω can be physically interpreted as the difference between two work values: that expended by the artificial force to guide the trial trajectory, versus the work that would be required to conversely guide the reference. The resulting distribution of entropy production, obtained from numerical sampling, is shown in Fig. 1(a).

Since ω is given by a sum over all t_{obs} time steps, $P(\omega)$ adopts a large deviation form as in Eq. (10), and $\langle\omega\rangle \propto t_{\text{obs}}$. These scaling properties are demonstrated numerically in Fig. 2(a) and analytically in Appendix A. In the Appendix, we re-express ω in terms of the ξ and $\tilde{\xi}$ variables, which can be integrated over to yield

$$\langle\omega\rangle = \frac{2}{(k-2)^2}\left[(2-k)\,k t_{\text{obs}} - 1 + (k-1)^{2t_{\text{obs}}}\right]. \tag{16}$$

Indeed, for $0 < k < 2$, this expression gives the anticipated long time scaling with t_{obs},

$$\langle\omega\rangle \sim \frac{2kt_{\text{obs}}}{2-k}. \tag{17}$$

As seen in Fig. 1(a), the negative-ω tail of $P(\omega)$, which gives rise to MCMC acceptances, becomes correspondingly small for large t_{obs}.

B. Guiding choices

In both Secs. III B and IV A, we showed that time-extensive entropy production arises generically when we do not use natural (forward) dynamics to generate a trajectory. Dynamical biases can alternatively be achieved by preferentially selecting among different examples of natural dynamics. At a high level, conventional TPS[13] is just such an approach, constructing biased trajectory ensembles through selection rather than artificial forces. Can this strategy be used effectively to impose resemblance between reference and trial trajectories?

We consider a scheme very similar in spirit to the STePS algorithm.[15] Like configurational-bias MC sampling of a polymer,[32] the STePS procedure generates a long trajectory by piecing together short segments, as illustrated in Fig. 1(b). To generate a new segment, one starts at the end of the previous

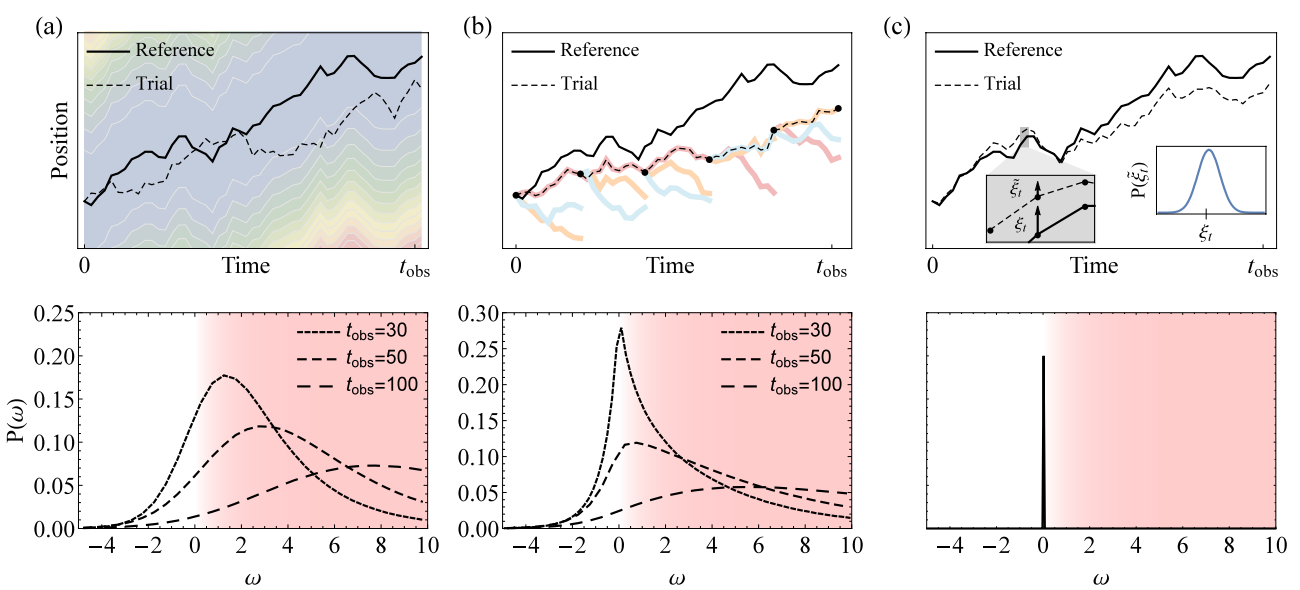

FIG. 1. Three guidance schemes for generating a trial trajectory that maintains proximity to a reference trajectory. For the specific case of a one-dimensional random walker, upper panels illustrate the consequences of (a) artificial corralling forces, (b) preferential selection of short trial branches, and (c) correlated noise histories. Bottom panels show the corresponding distributions of trajectory space entropy production ω. The intensity of red shading reflects the probability that trial moves are rejected. For cases (a) and (b), the average entropy production is nonzero and grows with trajectory length t_{obs}. With an appropriately designed noise guidance scheme (c), symmetric selection of noise variables results in identically zero entropy production for all trajectory lengths.

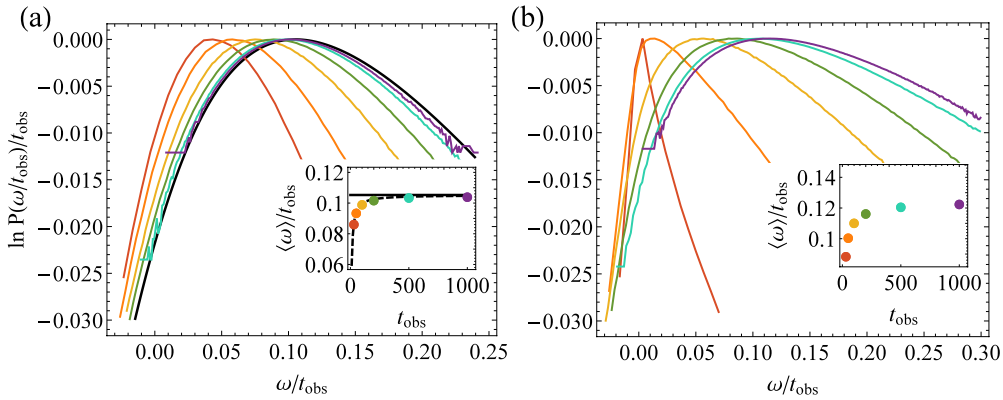

FIG. 2. Entropy production statistics for trial moves with guiding forces (a) and guiding choices (b), as discussed in Secs. IV A and IV B of the main text. Results in (a) are shown for $k = 0.1$; black curves indicate the long-time behavior determined analytically in Appendix A. Results in (b) are shown for $n = 3, \tau = 10$ and with $f(x) = e^{-|x|}$. In both panels, different colors indicate different trajectory lengths, ascending from left to right: $t_{\text{obs}} = 30$ (red), 50 (orange), 100 (yellow), 200 (green), 500 (cyan), and 1000 (purple).

segment and samples a collection of short, unbiased trajectories according to the natural dynamics, which we will refer to as branches. One of these branches is selected as the next segment of the trial trajectory, with a preference for branches that stay close to the reference trajectory. (Proximity could be judged in different ways, e.g., through Euclidean distance in the full phase space, or with respect to an order parameter.) Though each branch is grown with natural dynamics, the added segment is biased. To show that this bias affects acceptance rates in the same manner as the guiding forces bias, we compute the entropy produced by a TPS move.

Starting at the initial condition of the reference trajectory, we grow n branches of length τ according to

$$x_{t+1}^{(\alpha)} = x_t^{(\alpha)} + \xi_t^{(\alpha)}, \tag{18}$$

where α is an index over the n independent samples of the natural dynamics. Of these n possibilities for the ith segment of the trial trajectory, we select branch α with probability

$$P_{\text{select}}(\alpha) = \frac{f(|x_{i\tau}^{(0)} - x_{i\tau}^{(\alpha)}|)}{\sum_{\gamma=1}^{n} f(|x_{i\tau}^{(0)} - x_{i\tau}^{(\gamma)}|)}, \tag{19}$$

where f is a weighting function with a maximum when its argument is zero, for example, a Gaussian centered on zero. The reference trajectory is indicated by a superscript (0). Starting from the end of the chosen branch, the growth procedure is repeated with n new branches of length τ.

While each time propagation step uses segments of unbiased natural dynamics, the selection of preferred branches exerts a bias which ultimately leads to a nonvanishing entropy production,

$$\omega = -\sum_{i=1}^{t_{\text{obs}}/\tau} \ln \frac{\sum_{\gamma\neq 0} f(|x_{i\tau}^{(\gamma)} - x_{i\tau}^{(0)}|)}{\sum_{\gamma\neq\alpha_i} f(|x_{i\tau}^{(\gamma)} - x_{i\tau}^{(\alpha_i)}|)}, \tag{20}$$

where α_i is the index of the selected branch for the ith segment. The calculation of this entropy production requires generation of the backwards TPS move, in which the (0) branch is always selected.

In Sec. IV A, we discussed that the entropy produced by guiding forces could be thought of in terms of a work performed by the bias. From that perspective, this guiding choices scheme trades work for information. We bias the dynamics not by applying explicit forces but instead by selecting the preferred branches based on information about the likelihood that a branch stays close to the reference. In particular, ω is a difference between the Shannon information associated with selecting the set of trial branches which produced the trajectory $\tilde{x}(t)$ and the information associated with selecting the reference branches in a reverse TPS move. As with biasing forces, the trajectory space entropy production exhibits large deviation scaling with $\langle\omega\rangle \propto t_{\text{obs}}$. Numerical demonstrations of this scaling are provided in Fig. 2(b). Consequently, acceptance probabilities drop precipitously in the long time limit.

C. Guiding noises

As a third scheme for engineering path correlation, we consider generating a trial trajectory with natural dynamics but with biased noises. Rather than trying to corral trajectories to proceed along similar paths, one may impose much simpler correlations between their underlying noise histories.[16] Consider a TPS trial move which consists of re-propagating dynamics from the initial time step using new noises $\tilde{\xi}$ that differ only slightly from the old noises ξ,

$$\tilde{\xi}_t = \alpha\xi_t + \sqrt{1-\alpha^2}\eta_t, \tag{21}$$

where η_i is sampled from a Gaussian distribution with zero mean and variance σ^2. In Sec. IV B, the symbol α was used as an index, but here we redefined α to be the parameter controlling noise correlations. Guiding Gaussian noise variables in this manner has been referred to as a Brownian tube proposal move.[17] Unlike the prior two kinds of moves, the Brownian tube proposal produces strictly vanishing entropy production ω for all trials regardless of trajectory length. The cancellation results from some algebra after writing the path weights and generation probabilities in terms of the noise variables,

$$\frac{P(\boldsymbol{\xi})P_{\text{gen}}(\boldsymbol{\xi}\to\tilde{\boldsymbol{\xi}})}{P(\tilde{\boldsymbol{\xi}})P_{\text{gen}}(\tilde{\boldsymbol{\xi}}\to\boldsymbol{\xi})} = \frac{\exp\left[-\frac{\boldsymbol{\xi}^2}{2\sigma^2}\right]\exp\left[-\frac{(\tilde{\boldsymbol{\xi}}-\alpha\boldsymbol{\xi})^2}{2(1-\alpha^2)\sigma^2}\right]}{\exp\left[-\frac{\tilde{\boldsymbol{\xi}}^2}{2\sigma^2}\right]\exp\left[-\frac{(\boldsymbol{\xi}-\alpha\tilde{\boldsymbol{\xi}})^2}{2(1-\alpha^2)\sigma^2}\right]} = 1, \tag{22}$$

where $\boldsymbol{\xi}$ is a vector detailing noises at all times.

The vanishing entropy production is achieved by *independently sampling* the noise variables. In the previously discussed approaches, the bias applied to any one noise variable depended on how far astray the trial trajectory had drifted from its reference up to that point in time. Such history-dependent biasing coupled the sampling of one noise variable to all of the previous noises, ultimately giving rise to the nonvanishing ω. By sampling all noises independently, we can perturb the ξ variables in a symmetric manner. For noises drawn from a Gaussian distribution, this symmetric perturbation was given in Eq. (21), but the strategy of symmetrically sampling independent noises generalizes to other choices of stochastic dynamics. For example, Hartmann has applied these methods with uniform random variable noises to Monte Carlo dynamics in the form of Wolff dynamics of a two-dimensional Ising model.[33]

Using correlated noise histories to sample Monte Carlo trajectories avoids the time-extensive bias that arose from guiding paths in configuration space, but this merit comes at a cost. If the reactive or tilted ensembles are to be sampled, it is important that the guidance scheme produces highly correlated trajectories. That is to say, the x coordinates, not just the ξ coordinates, must be correlated. When will similar noise histories produce similar trajectories? In the remainder of the paper, we address this question in the context of two dynamical systems, one on-lattice and the other off-lattice.

V. EFFICACY OF NOISE GUIDANCE

In Sec. IV C, we noted that sampling trajectories with noise-guided shooting moves avoids a time-extensive MC entropy production. However, we seek correlated trajectories, not just correlated noises. When trajectories with correlated noises synchronize, efficient path sampling of long trajectories can be achieved. But under what conditions should such synchronization be expected? We investigate this question by studying lattice dynamics of a two-dimensional Ising model and off-lattice dynamics of a WCA fluid, also in two dimensions. We show that synchronization can be achieved with a suitable treatment of Ising dynamics. This success does not extend to our example of off-lattice dynamics.

A. Ising dynamics

Let us first consider a two-dimensional Ising model consisting of N spins. The ith spin, denoted σ_i, takes the value ± 1. The lattice evolves, at inverse temperature β, under single spin-flip Glauber dynamics with Hamiltonian

$$H = -h \sum_i \sigma_i - J \sum_{\langle ij \rangle} \sigma_i \sigma_j. \tag{23}$$

The spins interact in the usual Ising manner; they couple to nearest neighbors with coupling constant J and to an external field h. Each spin-flip trial move requires us to choose two random numbers uniformly from $[0,1)$. One random number, ξ_{site}, determines which site will be flipped. The other random number, ξ_{acc}, determines whether to accept or reject the flip. Given ξ_{site} and ξ_{acc}, the spin-flip move is deterministic.

1. Choose spin $i = \text{ceiling}(\xi_{\text{site}} N)$ to act on.
2. Construct a trial state by flipping spin i.
3. Compute the energy difference, ΔE, between the original configuration and the trial.
4. Accept the spin flip if $\xi_{\text{acc}} < (1 + \exp(\beta \Delta E))^{-1}$.

By carrying out t_{obs} sequential MC sweeps, each consisting of N spin flip moves, we construct an Ising trajectory, $\sigma(t)$. The effective unit of time is taken to be a MC sweep.

Now consider a noise-guided trial TPS move designed to generate a trajectory $\tilde{\sigma}(t)$ which is correlated with $\sigma(t)$. At every MC step we alter ξ_{site} and ξ_{acc} to some trial values, $\tilde{\xi}_{\text{site}}$ and $\tilde{\xi}_{\text{acc}}$. There is significant freedom in doing so, while producing zero entropy,[34] and we analyze one particular choice. We focus first on updating the noise that chooses which spin to flip. With probability $1 - \epsilon_{\text{site}}$, we reuse the old noise, i.e., $\tilde{\xi}_{\text{site}} = \xi_{\text{site}}$. Otherwise, we uniformly draw a new value of $\tilde{\xi}_{\text{site}}$ from the unit interval. The tunable parameter ϵ_{site} controls the correlation between noise histories of the reference and trial trajectories. We update the noises that control conditional acceptance, ξ_{acc}, in an analogous manner. Another parameter, ϵ_{acc}, is the probability of drawing new noise for $\tilde{\xi}_{\text{acc}}$.

Starting with the initial configuration of $\sigma(t)$, we construct $\tilde{\sigma}(t)$ by performing spin flips with the new trial noise history. The trial and reference trajectories start in identical configurations, but we expect the correlation to decay as MC time advances. To monitor the similarity between reference and trial, we study the site-wise product between σ and $\tilde{\sigma}$ as illustrated in Fig. 3. The average of this product over all spins,

$$\sigma \cdot \tilde{\sigma} = \frac{1}{N} \sum_{i=1}^{N} \sigma_i \tilde{\sigma}_i, \tag{24}$$

is a measure of correlation between σ and $\tilde{\sigma}$. Decorrelated configurations return a value of zero while identical configurations return one. Fig. 4(a) shows that the average correlation between $\sigma(t)$ and $\tilde{\sigma}(t)$ decays to zero at long times. The rate of this decay is tuned by ϵ_{acc} and ϵ_{site}, the parameters controlling the extent of noise correlation. At long times, $\langle \sigma \cdot \tilde{\sigma} \rangle$ eventually approaches zero, even with the strongest noise guidance. The corresponding "uncorrelated" configurations, however, bear a subtler resemblance. Some regions are significantly correlated, while others are significantly anticorrelated, averaging to give $\sigma \cdot \tilde{\sigma} \approx 0$.

Motivated by this subtler resemblance, we introduce a minor alteration in the implementation of Glauber spin-flip dynamics. Below we detail this modification and show that it does in fact enable the preservation of correlation between trajectories over very long times. As observed in the context of damage spreading, different choices of Ising model Monte Carlo dynamics can result in identical equilibrium states yet different dynamical properties.[35,36] The effectiveness of correlated noises in guiding trajectories is, in effect, one such dynamical property.

In particular, we replace step 2 of the spin-flip move to include a directionality. We introduce another random number, $\xi_{\text{dir}} \in [0,1)$, used to decide the trial state. If $\xi_{\text{dir}} < 0.5$, the trial state is a down spin; otherwise it is up. Whereas the

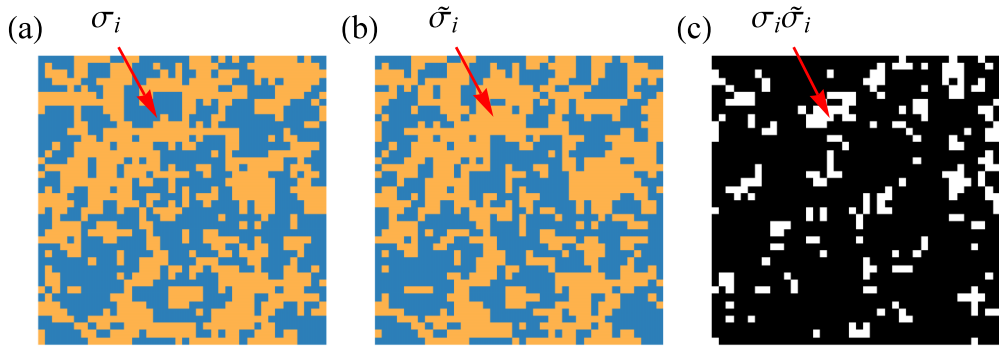

FIG. 3. Correlations between a reference trajectory and a trial trajectory generated by the noise guidance method described in Sec. V A using $\epsilon_{\rm dir} = \epsilon_{\rm site} = 10^{-3}$ and $\epsilon_{\rm acc} = 0.1$. The two trajectories begin with identical initial conditions and evolve through 100 sweeps of push up/push down Monte Carlo steps at $\beta J = 0.3, h = 0$. Final configurations of reference and trial trajectories are shown in (a) and (b), respectively. The site-wise overlap between these two configurations is depicted in (c), where black indicates spin alignment and white indicates anti-alignment.

conventional trial move affects an attempted spin flip, this trial move can be viewed as an attempt to push the spin either up or down, depending on the state of $\xi_{\rm dir}$. As with the other noise variables, correlations between $\xi_{\rm dir}$ and $\tilde{\xi}_{\rm dir}$ are tuned by the probability $\epsilon_{\rm dir}$ of resampling the noise.

The addition of directionality to the spin-flip dynamics results in moves which are trivially ineffectual. For example, half of the spin moves attempt to "push up" a spin which is already up. These moot moves are absent from the traditional implementation of single-spin-flip Glauber dynamics, which attempts a spin flip at every step of MC time. In every other respect, the two schemes generate Markov chains with identical statistics. They can therefore be made identical by excising moot moves, or, equivalently on average, by scaling time by a factor of two.

As Fig. 4(b) illustrates, the push up/push down implementation of single-spin-flip Glauber dynamics allows the trial trajectories, $\tilde{\sigma}(t)$ to remain close to $\sigma(t)$ for long times. By incorporating information about spin change directionality into the noise history, the noises signal not just how likely a spin is to change but in what direction it will change. Appropriately chosen ϵ parameters can maintain tunable correlations between trajectories for arbitrarily long time. When averaged over the whole lattice, steady state correlation is maintained, but the correlations are spatially homogeneous. As MC time progresses, the regions in which two trajectories are highly correlated move throughout the lattice, ensuring ergodic exploration of the trajectory space.

For push up/push down dynamics, noise guidance does not merely preserve correlations that existed at time zero. We find that correlated noise histories can in fact *induce* synchronization between trajectories. To illustrate this synchronization effect, we have characterized the correlation between trajectories that share similar noise histories only intermittently. As shown in Fig. 5, such paths acquire similarity during periods of strong noise guidance. This similarity degrades during periods without noise guidance, but can be recovered by re-introducing guidance, regardless of how significantly correlations have decayed. Indeed, even very different initial configurations, propagated with correlated noises, become more similar with time, their ensemble-averaged correlation $\langle \sigma \cdot \tilde{\sigma} \rangle$ converging to the same value as for trajectories that are identical at time zero.

A nonzero steady state value of $\langle \sigma \cdot \tilde{\sigma} \rangle$ is the quantitative signature of synchronization. The origins of this finite asymptotic correlation are transparent in the limit of weak coupling,

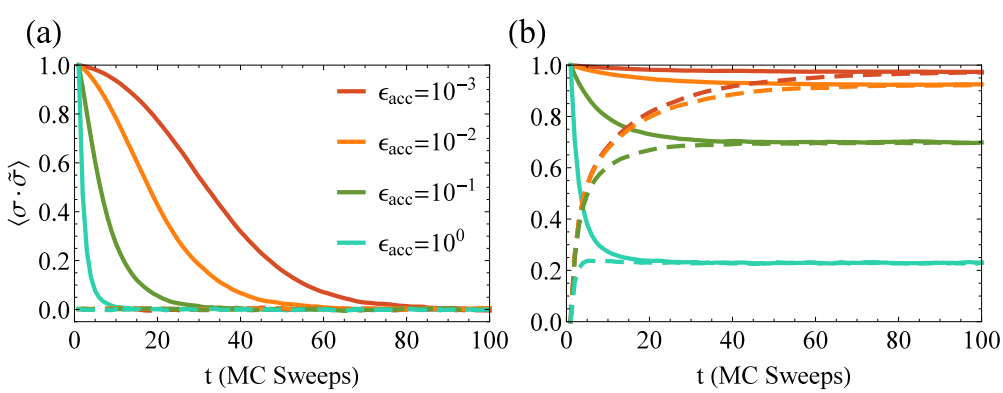

FIG. 4. Average overlap between reference and trial trajectories of a 40×40 two-dimensional Ising model with $\beta J = 0.3$. Results are shown in (a) for ordinary Glauber spin flip dynamics and in (b) for the modified directional dynamics described in Sec. V A. Two trajectories with identical initial conditions and different-but-correlated noise histories (solid lines) maintain a nonzero steady state overlap at long times only for the case of push up/push down dynamics. The same steady state values are obtained when the two trajectories evolve from very different initial conditions (dashed lines) generated by independently assigning each spin at random. Different colors indicate different values of the noise guidance parameter $\epsilon_{\rm acc}$. Ensemble-averaged results are shown for $\epsilon_{\rm site} = \epsilon_{\rm dir} = 0.001$, with averages performed over 500 independent pairs of trajectories.

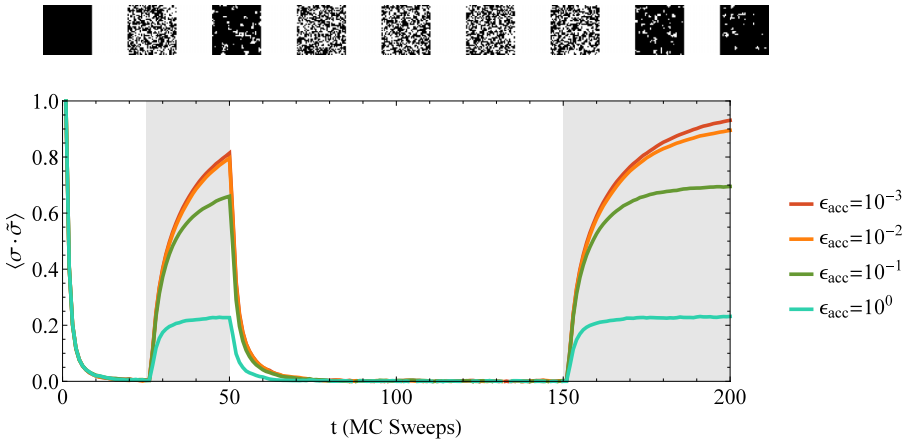

FIG. 5. Correlation between reference and trial trajectories of a 40×40 Ising model with $\beta J = 0.3$. Plotted lines are averages over 500 independent pairs of trajectories that evolve by push up/push down dynamics. Noise histories of each pair are generated in a correlated manner with $\epsilon_{site} = \epsilon_{dir} = 10^{-3}$ during the intervals $t = 25$–50 and $t = 150$–200 (the shaded regions); noise guidance is absent at all other times. The site-wise correlation between an example trajectory pair (with $\epsilon_{acc} = 10^{-3}$) is shown above the graph, with time advancing from left to right and adjacent configurations separated by 25 MC sweeps.

$\beta J = 0$. With the additional simplification $h = 0$, each attempted spin flip is accepted with probability 1/2 based on the value of ξ_{acc}, regardless of the states of neighboring spins. In this case, the steady state overlap can be calculated analytically. To do so, we derive an equation of motion for the probability $p(\tau)$ that a given spin has identical values in the reference and trial trajectories after τ MC steps. Note that τ differs from the unit of MC time, t, by a factor of N. The long-time, steady-state limit of this time evolution, $p_{ss} = \lim_{\tau\to\infty} p(\tau)$, yields $\langle\sigma\cdot\tilde{\sigma}\rangle_{ss} = 2p_{ss} - 1$. In Appendix B, we tabulate the various ways that a selected spin can become identical in reference and trial trajectories after a single time step. From this enumeration, and the corresponding probabilities, we find

$$p(\tau+1) = \frac{N-1}{N}p(\tau) - \frac{\epsilon_{dir}}{4N}\left(1-\epsilon_{site}+\frac{\epsilon_{site}}{N}\right)\left(1-\frac{\epsilon_{acc}}{2}\right) + \frac{1}{2N}\left(1-\frac{\epsilon_{acc}}{2}\right)\left(1-\epsilon_{site}+\frac{\epsilon_{site}}{N}\right)p(\tau) + \frac{1}{2N}. \tag{25}$$

The various terms of Eq. (25) describe the different ways that a single MC move can impact the state of an arbitrarily chosen spin in trial and reference trajectories. Since each move of our MC dynamics acts on a single site of the lattice, some moves do not involve the tagged spin at all, but instead some other lattice site; the first term in Eq. (25) reflects this possibility. The second term accounts for the decrease in overlap when reference and trial trajectories accept a spin-flip at the same site but in opposite directions. The third term results from the constructive action of correlated noises on the tagged spin, either maintaining existing correlation or inducing synchronization, as detailed in Appendix B. The final term accounts for random alignment of the tagged spin despite uncorrelated noise variables, a possibility particular to degrees of freedom with a limited number of discrete states. Equating $p(\tau)$ and $p(\tau+1)$ gives the steady state probability,

$$p_{ss} = \frac{1-\frac{\epsilon_{dir}}{2}\left(1-\frac{\epsilon_{acc}}{2}\right)\left(1-\epsilon_{site}+\frac{\epsilon_{site}}{N}\right)}{2-\left(1-\frac{\epsilon_{acc}}{2}\right)\left(1-\epsilon_{site}+\frac{\epsilon_{site}}{N}\right)}. \tag{26}$$

Analytically calculating steady-state overlap at finite temperature is not straightforward. Numerical results, shown for $\beta J = 0.4$ in Fig. 6, indicate that the dependence of overlap on strengths of noise perturbation is generically similar to the $\beta J = 0$ case analyzed above. Increasing βJ from zero does, however, slow the rate of convergence to the steady state while decreasing the degree of steady state correlation. For all coupling strengths we have examined, $\langle\sigma\cdot\tilde{\sigma}\rangle_{ss}$ can be made arbitrarily close to unity by decreasing the various ϵ parameters. This level of control ensures that one can generate trial trajectories which are correlated with a reference for all times, an essential capability for efficient path sampling of long trajectories.

B. WCA dynamics

The success we have achieved in synchronizing Ising dynamics with noise guidance should not be expected for complex dynamical systems in general. We demonstrate this limitation for the specific case of a two-dimensional WCA

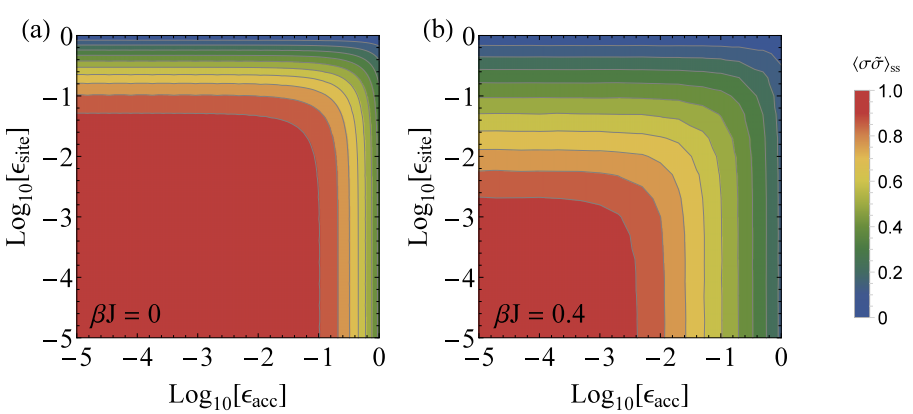

FIG. 6. Steady state correlations between reference and trial trajectories of a 40×40 Ising model, as a function of noise guidance strength. Results are shown for push up/push down dynamics with $\epsilon_{dir} = 0.001$. The high-temperature limit Eq. (26) is shown in (a). Finite temperature behavior (b) was obtained by sampling 4×10^4 pairs of trajectories for 340 MC sweeps each, with $\beta J = 0.4$.

fluid[38] evolving by underdamped Langevin dynamics. The purely repulsive particles are propagated using an integration scheme that requires generating a collection of Gaussian random variables.[39,40] These noises are guided by a Brownian tube proposal, Eq. (21). The similarity between trial and reference noise histories is controlled by a parameter α that ranges from zero (no noise guidance) to one (complete noise guidance).

Starting from identical initial configurations, we propagate dynamics with correlated noise histories and monitor the difference between trial and reference as

$$\langle |\mathbf{x} - \tilde{\mathbf{x}}| \rangle = \frac{1}{N} \sum_{i=1}^{N} \left| \mathbf{x}^{(i)} - \tilde{\mathbf{x}}^{(i)} \right|, \tag{27}$$

where $\mathbf{x}^{(i)}$ and $\tilde{\mathbf{x}}^{(i)}$ are the positions of particle i in the reference and trial, and $|\cdot|$ is the two-dimensional Euclidean distance. At short times, the difference between trial and reference trajectories is small, but this difference grows exponentially, a hallmark of chaotic dynamics. Even with exceptionally strong noise guidance, trajectories cannot be held arbitrarily close to each other for long times, as shown in Fig. 7.

Why are we unable to guide the evolution of WCA particles as effectively as we guided Ising dynamics? A principal difference between the two systems is the likelihood of spontaneous local recurrences. In either case, trajectories with similar initial conditions but different noise histories wander away from one another in a global sense, eventually exploring very different regions of configuration space. Correlating their noise histories generally acts to defer, but not defeat, this divergence. In the case of Ising dynamics, however, a given small block of adjacent spins will occasionally align spontaneously in two trajectories with reasonable probability. Following such a spontaneous, local recurrence, correlated noises again work to hold trajectories close. Global spontaneous recurrence of a large number of spins is, of course, highly improbable, but the noise guidance seems to "lock in" local correlations every time they spontaneously reoccur.

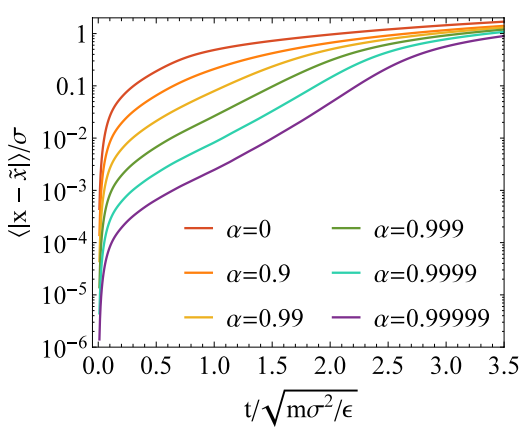

FIG. 7. Divergence between reference and trial trajectories of a two-dimensional WCA fluid with Brownian tube noise guidance of strength α. Average distance between the two trajectories, as defined in Eq. (27), is shown for a system of 400 particles with mass m and diameter σ, in a square box with side length 24 σ (i.e., density $\rho\sigma^2 = 0.694$). Underdamped Langevin dynamics was propagated with inverse temperature $\beta = 0.2$ and friction coefficient $\gamma = 0.1$ using a time step of 0.005 τ, where $\tau = \sqrt{m\sigma^2/\epsilon}$ and ϵ is the Lennard-Jones interaction energy scale.[37] Data are averaged over 500 independent trial trajectories.

VI. CONCLUSION

Transition path sampling has proven useful for a variety of equilibrium, as well as non-equilibrium, problems in chemical dynamics. The problem of sampling long trajectories, particularly those with multiple intermediates, has hindered a variety of extensions and applications of the methodology. We have outlined a modern physical perspective from which to assess and address these challenges. We have demonstrated and discussed successful trajectory guidance in the case of Monte Carlo dynamics of an Ising model. Substantial difficulties remain for systems with continuous degrees of freedom.

Our results suggest that effective noise guidance of long trajectories requires a nonnegligible probability of spontaneous local recurrence, i.e., a significant chance that reference and trial trajectories transiently align within small regions of space. Such synchronization could be particularly helpful for sampling reactive trajectories that traverse metastable intermediate states, for example, the coarsening or assembly of colloidal systems as they organize on progressively larger scales. In such cases, trial trajectories in the course of path sampling should maintain correlations with the reference while passing through the intermediates, not just at the endpoints. Even without identifying metastable configurations, correlated noises could be applied during some intervals but not others. A tendency to synchronize would enable trial trajectories to explore widely during unguided periods but to be reined in globally by intermittent guidance.

We anticipate that these noise-guidance methods will be effective for other lattice systems as well, but their usefulness could depend sensitively on the exact manner in which the noise influences dynamics. In particular, without incorporating directionality into proposed spin changes, we were not able to guide long Ising trajectories. Furthermore, Ising dynamics can exhibit spontaneous recurrence, i.e., transient local alignment between two trajectories regardless of their noise histories. Because small blocks of Ising spins can adopt only a modest number of configurations, such random local synchronization occurs with an appreciable probability. The probability of recurrence will likely be lower for models with a larger collection of possible local configurations, e.g., a Potts model or an Ising model with more neighbors. We thus expect that the application of noise-guided path sampling could face substantial challenges for long trajectories of these more intricate lattice models.

ACKNOWLEDGMENTS

We acknowledge many useful discussions with Michael Grünwald about strategies for sampling long trajectories. T.R.G. acknowledges support from the NSF Graduate Research Fellowship and the Fannie and John Hertz Foundation. P.L.G. was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, through the Chemical Sciences

Division (CSD) of the Lawrence Berkeley National Laboratory (LBNL), under Contract No. DE-AC02-05CH11231.

APPENDIX A: ENTROPY PRODUCTION STATISTICS FOR A ONE-DIMENSIONAL RANDOM WALKER WITH GUIDING FORCES

1. Mean entropy production

Here, we analytically characterize the entropy production distribution, $P(\omega)$, for shooting moves generated with guiding forces as discussed in Sec. IV A. It is useful to first rewrite Eq. (2) in terms of the noise variables, ξ and $\tilde{\xi}$. For a one-dimensional random walk, the position x_{t+1} and the difference between reference and trial trajectories, $x_{t+1} - \tilde{x}_{t+1}$, can be compactly expressed in terms of the noises,

$$x_{t+1} = \sum_{u=0}^{t} \xi_u, \tag{A1}$$

$$x_{t+1} - \tilde{x}_{t+1} = \sum_{u=0}^{t} (1-k)^{t-u}(\xi_u - \tilde{\xi}_u). \tag{A2}$$

After straightforward algebra, it is possible to express ω as

$$\omega = \frac{1}{\sigma^2} \sum_{t=0}^{t_{\text{obs}}-1} \left(S_t^2 - S_t \xi_t^+\right), \tag{A3}$$

where for convenience we have defined $\xi_t^+ \equiv \tilde{\xi}_t + \xi_t + 2\mu$, $\xi_t^- \equiv \tilde{\xi}_t - \xi_t$, and

$$S_t \equiv \sum_{u=0}^{t-1} k(1-k)^{t-1-u} \xi_u^-. \tag{A4}$$

The main text presents results for a random walk without drift, i.e., with ξ drawn from a distribution with mean zero. Here, we consider the more general case with nonzero mean μ. Noting that $\langle \xi_t^+ \xi_u^+ \rangle = \langle \xi_t^- \xi_u^- \rangle = 2\sigma^2 \delta_{tu}$ and $\langle \xi_t^+ \xi_u^- \rangle = 0$, the average entropy production is found to be

$$\langle \omega \rangle = 2 \sum_{t=0}^{t_{\text{obs}}-1} \sum_{u=0}^{t-1} k^2 (1-k)^{2(t-1-u)}. \tag{A5}$$

The two geometric series are summed to yield

$$\langle \omega \rangle = \frac{2}{(k-2)^2} \left[(2-k)\, k t_{\text{obs}} - 1 + (k-1)^{2t_{\text{obs}}} \right]. \tag{A6}$$

When $k > 2$, $\langle \omega \rangle$ grows exponentially in t_{obs}. This superlinear scaling results from coupling between trajectories so strong that the trial trajectory rapidly tends to infinity due to a numerical instability, much like the instability that arises in conventional molecular dynamics simulations performed with an excessively large integration time step. For the useful range of coupling strength, $k < 2$, $\langle \omega \rangle \propto t_{\text{obs}}$ in the long time limit. The marginal $k = 2$ case is well-behaved ($\langle \omega \rangle = 4 t_{\text{obs}} (t_{\text{obs}} - 1)$) but uninteresting for our purposes.

2. Cumulant generating function

The behavior of the higher-order cumulants can be extracted from the cumulant generating function $\ln \langle e^{-\lambda \omega} \rangle$. This average requires integration over all of the Gaussian ξ and $\tilde{\xi}$ variables at all times, which can be performed inductively. We define $\phi(\lambda, f, g, h, t)$ as

$$\phi(\lambda,f,g,h,t) = f\left(\frac{1}{2\sigma\sqrt{\pi}}\right)^{2t} \int d\xi_0^+ \ldots d\xi_{t-1}^+ d\xi_0^- \ldots d\xi_{t-1}^- \exp\left[\frac{1}{\sigma^2}\sum_{i=0}^{t-2}\left(\frac{-(\xi_i^+)^2 - (\xi_i^-)^2}{4} + \lambda S_i \xi_i^+ - \lambda S_i^2\right)\right]$$
$$\times \exp\left[\frac{1}{\sigma^2}\left(\frac{-(\xi_{t-1}^+)^2 - (\xi_{t-1}^-)^2}{4} + \lambda\left(S_{t-1}\xi_{t-1}^+ - h S_{t-1}^2 + 2gk(1-k) S_{t-1}\xi_{t-1}^- + gk^2(\xi_{t-1}^-)^2\right)\right)\right]. \tag{A7}$$

The integral ϕ is defined such that $\langle e^{-\lambda\omega} \rangle = \phi(\lambda, 1, 0, 1, t_{\text{obs}})$. By introducing f, g, and h, we can derive recursion relations as we sequentially integrate out Gaussian noises at the latest remaining time step. In particular, integration over ξ_{t-1}^- then ξ_{t-1}^+ returns an integral of the same form. That is to say $\phi(\lambda, f_i, g_i, h_i, t) = \phi(\lambda, f_{i+1}, g_{i+1}, h_{i+1}, t-1)$ with

$$f_{i+1} = \frac{f_i}{\sqrt{1 - 4\lambda g_i k^2}}, \tag{A8}$$

$$g_{i+1} = \lambda - h_i + \frac{4\lambda g_i^2 (1-k)^2 k^2}{1 - 4\lambda g_1 k^2}, \tag{A9}$$

$$h_{i+1} = 1 - (1-k)^2 \left((\lambda - h_i) + \frac{4\lambda g_i^2 (1-k)^2 k^2}{1 - 4\lambda g_i k^2} \right). \tag{A10}$$

Iterating the map t_{obs} times corresponds to integrating over all of the $2t_{\text{obs}}$ integrals in Eq. (A7). After some algebraic simplification,

$$\ln \langle e^{-\lambda\omega} \rangle = -\frac{1}{2} \sum_{i=0}^{t_{\text{obs}}-1} \ln(1 - 4\lambda g_i k^2), \tag{A11}$$

where $g_0 = 0$ and

$$g_{i+1} = \lambda - 1 + \frac{(1-k)^2 g_i}{1 - 4\lambda g_i k^2}. \tag{A12}$$

The scaled cumulant generating function is then given by

$$\lim_{t_{\text{obs}} \to \infty} \frac{1}{t_{\text{obs}}} \ln \langle e^{-\lambda\omega} \rangle = -\frac{1}{2} \ln(1 - 4\lambda g^* k^2), \tag{A13}$$

where g^* is a fixed point of the map given in Eq. (A12). Specifically, it is the lesser of the two roots of the quadratic equation obtained when $g_i = g_{i+1} \equiv g^*$ is inserted into Eq. (A12). The numerical Legendre transform of this scaled cumulant generating function gives the solid black curve in Fig. 2(a), toward which the results of numerical sampling should converge for long t_{obs}.

APPENDIX B: ISING MODEL STEADY STATE CORRELATIONS

To derive the steady state correlation between reference and trial trajectories, we examine the time evolution of the probability $p(\tau)$ that reference and trial overlap at site i and MC step τ. Without loss of generality, we focus on a particular site, $i = 1$. Push up/push down moves can be grouped into four classes: (i) the reference and trial each act on spin 1, (ii) the reference acts on spin 1 while the trial acts on a different spin, (iii) the trial acts on spin 1 while the reference acts on a different spin, or (iv) neither reference nor trial acts on spin 1. For each case, we derive a transition matrix which maps the vector $(p(\tau), 1 - p(\tau))$ to its state at MC step $\tau + 1$. The full transition matrix for a step of dynamics is the sum of these transition matrices, weighted by the probability of each case,

$$T = \left(1 - \frac{1}{N}\left(1 + \epsilon_{\text{site}} - \frac{\epsilon_{\text{site}}}{N}\right)\right) I + \frac{1}{N}\left(\epsilon_{\text{site}} - \frac{\epsilon_{\text{site}}}{N}\right) Q + \frac{1}{N}\left(\epsilon_{\text{site}} - \frac{\epsilon_{\text{site}}}{N}\right) R + \frac{1}{N}\left(1 - \epsilon_{\text{s}} + \frac{\epsilon_{\text{site}}}{N}\right) S. \tag{B1}$$

The transition matrix for case (iv) is the identity matrix, since this case cannot alter the overlap at site 1. The transition matrices for cases (i), (ii), and (iii) are Q, R, and S, respectively, the forms of which we now derive.

When reference and trial act on different spins, only one copy of spin 1 (the reference or the trial) can change its state. When the two copies differ at site i after τ steps, overlap is induced with probability 1/4 (i.e., the probability that any given move results in a change of spin state). For initially aligned copies, loss of overlap similarly occurs with probability 1/4. This logic applies equally well to cases (ii) and (iii), so

TABLE I. Enumeration of moves yielding $\sigma' = \tilde{\sigma}'$. Without loss of generality, we only list the moves which start with $\sigma = 1$. The moves starting with $\sigma = -1$ are analogous. d is the direction of a push with 1 meaning up. a indicates whether the move is accepted (1) or rejected (0). An asterisk indicates that both options yield the same result.

σ	d	a	σ'	$\tilde{\sigma}$	$\tilde{d}$	$\tilde{a}$	$\tilde{\sigma}'$	Probability
1	1	*	1	1	1	*	1	$\frac{p}{2}\left(1 - \frac{\epsilon_{\text{dir}}}{2}\right)$
1	1	*	1	1	−1	0	1	$\frac{\epsilon_{\text{dir}} p}{8}$
1	1	*	1	−1	1	1	1	$\frac{1-p}{4}\left(1 - \frac{\epsilon_{\text{dir}}}{2}\right)$
1	−1	0	1	1	1	*	1	$\frac{\epsilon_{\text{dir}} p}{8}$
1	−1	0	1	1	−1	0	1	$\frac{p}{4}\left(1 - \frac{\epsilon_{\text{dir}}}{2}\right)\left(1 - \frac{\epsilon_{\text{acc}}}{2}\right)$
1	−1	0	1	−1	1	1	1	$\frac{1-p}{4}\left(\frac{\epsilon_{\text{dir}}}{2}\right)\left(\frac{\epsilon_{\text{acc}}}{2}\right)$
1	−1	1	−1	1	−1	1	−1	$\frac{p}{4}\left(1 - \frac{\epsilon_{\text{dir}}}{2}\right)\left(1 - \frac{\epsilon_{\text{acc}}}{2}\right)$
1	−1	1	−1	−1	1	0	−1	$\frac{1-p}{4}\left(\frac{\epsilon_{\text{dir}}}{2}\right)\left(\frac{\epsilon_{\text{acc}}}{2}\right)$
1	−1	1	−1	−1	−1	*	−1	$\frac{1-p}{4}\left(1 - \frac{\epsilon_{\text{dir}}}{2}\right)$

$$Q = R = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}. \tag{B2}$$

When both reference and trial act on site 1, we must account for correlated influence on the two copies. As a result, S depends on ϵ_{acc} and ϵ_{dir}. To enumerate these correlated changes, we denote states of spin 1 at step τ in the reference and trial as σ_1 and $\tilde{\sigma}_1$, respectively. After the MC step, these spins are given by σ_1' and $\tilde{\sigma}_1'$. Table I lists the possible transformations which result in overlapping spins ($\sigma_1' = \tilde{\sigma}_i'$) after $\tau + 1$ steps. Collecting terms in the table and making use of the fact that S is a probability-conserving transition matrix, we find

$$S = \frac{1}{2}\begin{pmatrix} 1 + \left(1 - \frac{\epsilon_{\text{dir}}}{2}\right)\left(1 - \frac{\epsilon_{\text{acc}}}{2}\right) & 1 - \frac{\epsilon_{\text{dir}}}{2}\left(1 - \frac{\epsilon_{\text{acc}}}{2}\right) \\ 1 - \left(1 - \frac{\epsilon_{\text{dir}}}{2}\right)\left(1 - \frac{\epsilon_{\text{acc}}}{2}\right) & 1 + \frac{\epsilon_{\text{dir}}}{2}\left(1 - \frac{\epsilon_{\text{acc}}}{2}\right) \end{pmatrix}. \tag{B3}$$

Propagation according to the transition matrix T gives the overlap probability after a single MC step,

$$\begin{pmatrix} p(\tau + 1) \\ 1 - p(\tau + 1) \end{pmatrix} = T \begin{pmatrix} p(\tau) \\ 1 - p(\tau) \end{pmatrix}. \tag{B4}$$

The first row of this matrix equation reads, after some algebra,

$$p(\tau + 1) = p(\tau) - \frac{p(\tau)}{N} + \frac{\epsilon_{\text{site}} - \frac{\epsilon_{\text{site}}}{N}}{2N} + \frac{1 - \epsilon_{\text{site}} + \frac{\epsilon_{\text{site}}}{N}}{2N}\left[1 - \frac{\epsilon_{\text{dir}}}{2}\left(1 - \frac{\epsilon_{\text{acc}}}{2}\right) + \left(1 - \frac{\epsilon_{\text{acc}}}{2}\right) p(\tau)\right]. \tag{B5}$$

We are interested in the steady state solution, found by setting $p(\tau) = p(\tau + 1)$. Multiplying the equation through by N, followed by algebraic simplification, yields Eq. (26) of the main text.

[1]C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[2]J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante, Science **296**, 1832 (2002).
[3]D. Collin, F. Ritort, C. Jarzynski, S. Smith, I. Tinoco, and C. Bustamante, Nature **437**, 231 (2005).
[4]S. Park, F. Khalili-Araghi, E. Tajkhorshid, and K. Schulten, J. Chem. Phys. **119**, 3559 (2003).
[5]H. Oberhofer, C. Dellago, and P. L. Geissler, J. Phys. Chem. B **109**, 6902 (2005).
[6]J. P. Nilmeier, G. E. Crooks, D. D. Minh, and J. D. Chodera, Proc. Natl. Acad. Sci. U. S. A. **108**, E1009 (2011).
[7]P. G. Bolhuis, C. Dellago, and D. Chandler, Proc. Natl. Acad. Sci. U. S. A. **97**, 5877 (2000).
[8]P. L. Geissler, C. Dellago, D. Chandler, J. Hutter, and M. Parrinello, Science **291**, 2121 (2001).
[9]R. J. Allen, P. B. Warren, and P. R. Ten Wolde, Phys. Rev. Lett. **94**, 018104 (2005).

[10]R. J. Allen, C. Valeriani, S. Tănase-Nicola, P. R. Ten Wolde, and D. Frenkel, J. Chem. Phys. **129**, 134704 (2008).
[11]J. Hu, A. Ma, and A. R. Dinner, Proc. Natl. Acad. Sci. U. S. A. **105**, 4615 (2008).
[12]L. O. Hedges, R. L. Jack, J. P. Garrahan, and D. Chandler, Science **323**, 1309 (2009).
[13]P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, Annu. Rev. Phys. Chem. **53**, 291 (2002).
[14]M. Grünwald, C. Dellago, and P. L. Geissler, J. Chem. Phys. **129**, 194101 (2008).
[15]N. Guttenberg, A. R. Dinner, and J. Weare, J. Chem. Phys. **136**, 234103 (2012).
[16]G. E. Crooks and D. Chandler, Phys. Rev. E **64**, 026109 (2001).
[17]G. Stoltz, J. Comput. Phys. **225**, 491 (2007).
[18]N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[19]W. Hastings, Biometrika **57**, 97 (1970).
[20]U. Seifert, Phys. Rev. Lett. **95**, 040602 (2005).
[21]G. E. Crooks, "Excursions in statistical dynamics," Ph.D. thesis (University of California, 1999).
[22]S. Park, Phys. Rev. E **77**, 016709 (2008).
[23]B. Derrida and J. L. Lebowitz, Phys. Rev. Lett. **80**, 209 (1998).
[24]V. Lecomte, C. Appert-Rolland, and F. van Wijland, J. Stat. Phys. **127**, 51 (2007).
[25]J. L. Lebowitz and H. Spohn, J. Stat. Phys. **95**, 333 (1999).
[26]J. Mehl, T. Speck, and U. Seifert, Phys. Rev. E **78**, 011123 (2008).
[27]U. Seifert, Rep. Prog. Phys. **75**, 126001 (2012).
[28]J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland, J. Phys. A: Math. Theor. **42**, 075007 (2009).
[29]Since t_{shoot} is uniformly selected between 0 and t_{obs}, t_{shoot} is of order t_{obs}. Furthermore, while neighboring terms in the sum may be correlated, there will still be order t_{obs} independent terms for sufficiently long t_{obs}.
[30]Typically TPS procedures concurrently sample the initial configuration, but that complication is not necessary for our illustration.
[31]B. Isralewitz, M. Gao, and K. Schulten, Curr. Opin. Struct. Biol. **11**, 224 (2001).
[32]D. Frenkel and B. Smit, in *Understanding Molecular Simulation*, 2nd ed. (Academic, New York, 2002).
[33]A. K. Hartmann, Phys. Rev. E **89**, 052103 (2014).
[34]We want symmetric proposal probabilities, $p(\xi \to \xi') = p(\xi' \to \xi)$.
[35]H. Hinrichsen and E. Domany, Phys. Rev. E **56**, 94 (1997).
[36]A. Mariz, H. Herrmann, and L. de Arcangelis, J. Stat. Phys. **59**, 1043 (1990).
[37]D. A. Sivak, J. D. Chodera, and G. E. Crooks, J. Phys. Chem. B **118**, 6466 (2014).
[38]J. D. Weeks, D. Chandler, and H. C. Andersen, J. Chem. Phys. **54**, 5237 (1971).
[39]M. Athènes and G. Adjanor, J. Chem. Phys. **129**, 024116 (2008).
[40]D. A. Sivak, J. D. Chodera, and G. E. Crooks, Phys. Rev. X **3**, 011007 (2013).