Efficiency in nonequilibrium molecular dynamics Monte Carlo simulations

Brian K. Radak[1,a)] and Benoît Roux[2,3,b)]
[1]*Leadership Computing Facility, Argonne National Laboratory, Argonne, Illinois 60439-8643, USA*
[2]*Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois 60637-1454, USA*
[3]*Center for Nanoscale Materials, Argonne National Laboratory, Argonne, Illinois 60439-8643, USA*

(Received 15 August 2016; accepted 21 September 2016; published online 7 October 2016)

Hybrid algorithms combining nonequilibrium molecular dynamics and Monte Carlo (neMD/MC) offer a powerful avenue for improving the sampling efficiency of computer simulations of complex systems. These neMD/MC algorithms are also increasingly finding use in applications where conventional approaches are impractical, such as constant-pH simulations with explicit solvent. However, selecting an optimal nonequilibrium protocol for maximum efficiency often represents a non-trivial challenge. This work evaluates the efficiency of a broad class of neMD/MC algorithms and protocols within the theoretical framework of linear response theory. The approximations are validated against constant pH-MD simulations and shown to provide accurate predictions of neMD/MC performance. An assessment of a large set of protocols confirms (both theoretically and empirically) that a linear work protocol gives the best neMD/MC performance. Finally, a well-defined criterion for optimizing the time parameters of the protocol is proposed and demonstrated with an adaptive algorithm that improves the performance on-the-fly with minimal cost. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4964288]

I. INTRODUCTION

A general aim of molecular dynamics (MD) and Monte Carlo (MC) simulations is to sample configurations from a probability distribution dictated by equilibrium statistical mechanics. Nonequilibrium MD/MC (neMD/MC) simulations expand on these techniques by considering configurations generated by a nonequilibrium procedure.[1–7] That is, these neMD configurations are then treated as candidates to be accepted or rejected on the basis of an appropriately constructed criterion to recover the proper equilibrium distribution. An especially interesting family of neMD/MC algorithms are those that sample from expanded ensembles comprised of a set of potentials defined over a discrete parameter space.[1,3–5] While neMD/MC methods have distinct advantages over conventional MD and MC, achieving improved sampling efficiency is not always straightforward. In the worst case, a poorly designed neMD/MC algorithm can even lead to adverse performance.

The efficiency of neMD/MC depends on maximizing the production of uncorrelated configurations while minimizing the effort expended in generating them. The main drawback of neMD/MC simulations is that each new attempted MC move requires a nonequilibrium MD simulation; a low acceptance rate necessarily implies that a large fraction of computer time is discarded by the algorithm. A direct solution for this problem is not obvious. On the one hand, long neMD trajectories might yield a high acceptance probability, but are computationally prohibitive. Conversely, short neMD trajectories are computationally inexpensive but expected to yield vanishingly low acceptance probabilities. The most efficient algorithm is obviously a compromise balancing these opposing factors, but it is generally unclear as to how to systematically achieve this balance. While the optimization of nonequilibrium protocols has been considered previously by seeking to minimize the mean work[9,10] or mean error,[11] the issue of balancing the acceptance probability against the cost of the neMD trajectories, which is critical to maximize the efficiency of neMD/MC algorithms, has not been explored systematically.

The goal of this article is to assess and mitigate efficiency problems in a broad class of neMD/MC algorithms. The results of the present analysis are quite general, but a specific focus is given to neMD/MC schemes designed to generate constant-pH simulations with explicit solvent molecules.[1,5] The efficiency of neMD/MC simulations is shown to depend on specific dynamic properties of the system being studied and these are in turn related to the controllable parameters, including the nonequilibrium protocol. The theoretical results and approximations are tested against constant-pH simulations of propionic acid in water, a simple but realistic molecular system. Finally, a well-defined efficiency metric is developed and an adaptive procedure for maximizing the performance is proposed.

II. THEORY

Consider a classical system evolving according to the Hamiltonian $H(\boldsymbol{x};\lambda)$, with coordinates and conjugate momenta $\boldsymbol{x}$ and a state parameter λ varying over the interval $[0,1]$. The neMD/MC algorithm consists of carrying out a

a)Electronic mail: brian.radak@anl.gov
b)Electronic mail: roux@uchicago.edu. URL: http://thallium.bsd.uchicago.edu/RouxLab/.

short nonequilibrium MD trajectory of length τ to switch the state ($\lambda \to \lambda'$) in order to generate a candidate configuration ($\boldsymbol{x} \to \boldsymbol{x}'$) that must subsequently be accepted or rejected according to a Metropolis criterion

$$P_{\mathrm{acc}}(W) = \min\left[1, e^{-\beta W}\right], \tag{1}$$

where $\beta \equiv 1/k_{\mathrm{B}}T$ is the inverse temperature and W is the work done during the nonequilibrium switch.[6] In addition, alterations to the sign of the atomic momenta must be introduced to ensure microscopic detailed balance.[4] For the sake of simplicity, the following analysis is restricted to a linear coupling between two endpoint Hamiltonians

$$\begin{aligned} H(\boldsymbol{x};\lambda) &= \lambda H_1(\boldsymbol{x}) + (1-\lambda)H_0(\boldsymbol{x}) \\ &\equiv H_0(\boldsymbol{x}) - \lambda A(\boldsymbol{x}), \end{aligned} \tag{2}$$

where $A(\boldsymbol{x}) \equiv -\partial H(\boldsymbol{x};\lambda)/\partial\lambda$ is the coupling "force." In the present context, the Hamiltonian $H(\boldsymbol{x};\lambda)$ is purely a construction meant to couple the two physical endpoints, H_1 and H_0, and thus, the linear coupling assumption should not present any significant limitations in practice. For example, in the context of constant-pH neMD/MC simulations, the physical endpoints would correspond to the unprotonated and protonated states. More generally, the analysis is also valid if the endpoints correspond to artificially biased states intended to enhance sampling in configuration space.[12]

During a nonequilibrium simulation, λ is treated as a continuous protocol function, $\lambda(t)$, varying within the range $[0,1]$ during the time interval $[0,\tau]$. Outside this time interval, λ is strictly fixed at a value of either 0 or 1, and sampling is performed in an equilibrium phase space of coordinates and momenta. The mean expected work $\overline{W(\tau)}$ from this protocol can be computed by integrating the average power $\overline{\mathcal{P}(t)} = \overline{\partial H/\partial t} = \overline{\partial H/\partial\lambda}\dot{\lambda}(t) = -\dot{\lambda}(t)\overline{A(t)}$ (an overline bar indicates an average over the repeated application of a nonequilibrium protocol to equilibrium starting conditions). Introducing the response relative to equilibrium $\delta A(t) \equiv A(t) - \langle A\rangle$ allows a straightforward decomposition into reversible (free energy ΔF) and excess quantities [power $\overline{\mathcal{P}_{\mathrm{ex}}(t)}$ and work $\overline{W_{\mathrm{ex}}(t)}$],[8]

$$\begin{aligned} \overline{W(\tau)} &= -\int_0^\tau dt\, \dot{\lambda}(t)\left(\langle A\rangle_{\lambda(t)} + \overline{\delta A(t)}\right) \\ &= \int_0^1 d\lambda' \left\langle \frac{\partial H(\boldsymbol{x};\lambda)}{\partial\lambda}\right\rangle_{\lambda'} + \int_0^\tau dt\, \overline{\mathcal{P}_{\mathrm{ex}}(t)} \\ &= \Delta F + \overline{W_{\mathrm{ex}}(\tau)}. \end{aligned} \tag{3}$$

The first term in the second line formally corresponds to the work resulting from a switch applied infinitely slowly, such that the equilibrium average changes with time.[13] Equation (3) yields the nonequilibrium "protocol" work which is formally only one component of the quantity W in Eq. (1). In practice, an additional contribution arising from the fact that the MD propagator relies on a discrete time step Δt should also be taken into account in the Metropolis acceptance criterion to rigorously maintain microscopic detailed balance.[6,14,15] This small quantity $\sim O(\tau\Delta t^4)$ is referred to as the "shadow" work.[14] For the sake of simplicity, the small contribution from shadow work will be ignored in the present analysis (see the supplementary material). Aspects of this problem were recently addressed by Sivak and Crooks,[8] and their analysis is extended in the following for a Metropolis neMD/MC framework.

To make progress, it is assumed that the work distribution is predominantly unimodal (i.e., the mean and modal behavior are similar). In this scenario, the mean acceptance probability $\overline{P_{\mathrm{acc}}(\tau)}$ is a reasonable proxy for the most probable acceptance behavior. In the linear coupling case, the value of $\overline{P_{\mathrm{acc}}(\tau)}$ is dictated by the forward ($0 \to 1$) and backward ($1 \to 0$) work distributions $\rho_{\mathrm{f}}(W;\tau)$ and $\rho_{\mathrm{b}}(W;\tau)$. Each contribution must be weighted by the equilibrium probability of its starting point, P_0 and P_1, respectively, with $P_0 + P_1 = 1$ and $P_1/P_0 \equiv \exp(-\beta\Delta F)$. This leads to the following expression for the mean acceptance probability between the two states:

$$\overline{P_{\mathrm{acc}}(\tau)} = \int_{-\infty}^{\infty} dW\, P_{\mathrm{acc}}(W)[P_0\,\rho_{\mathrm{f}}(W;\tau) + P_1\,\rho_{\mathrm{b}}(W;\tau)]. \tag{4}$$

So long as the MC scheme only considers pairwise state transitions, this is equally applicable when considering more than two states.

In seeking to study the behavior of Eq. (4), it is necessary to know the form of $\rho_{\mathrm{f}}(W;\tau)$ and $\rho_{\mathrm{b}}(W;\tau)$. A reasonable assumption is to use Gaussian distributions (this shifts the time dependence of the distributions into the mean and variance parameters). This assertion follows, for example, from fluctuation-dissipation arguments[16] or by a cumulant expansion of Jarzynski's equality.[13] Although this is not exact except in very particular instances, only a semiquantitative idea of the behavior of Eq. (4) is being sought. This approximation has been considered extensively in the literature and, in conjunction with Crooks's fluctuation theorem, leads to an expression for the excess work in *both directions* in terms of a single variance $\sigma^2(\tau)$ but different means $\overline{W_{\mathrm{f}}(\tau)}$ and $\overline{W_{\mathrm{b}}(\tau)}$,[17–19]

$$\overline{W_{\mathrm{ex}}(\tau)} = \frac{\beta\sigma^2(\tau)}{2} = -\Delta F + \overline{W_{\mathrm{f}}(\tau)} = \Delta F + \overline{W_{\mathrm{b}}(\tau)}. \tag{5}$$

Equation (5) can be used to give an approximate, but explicit, form to Eq. (4) (see the supplementary material),

$$\begin{aligned} \overline{P_{\mathrm{acc}}(\tau)} = P_0\,\mathrm{erfc}&\left(\frac{\beta\sigma^2(\tau) + 2\Delta F}{2\sqrt{2\sigma^2(\tau)}}\right) \\ &+ P_1\,\mathrm{erfc}\left(\frac{\beta\sigma^2(\tau) - 2\Delta F}{2\sqrt{2\sigma^2(\tau)}}\right). \end{aligned} \tag{6}$$

In the case that $\Delta F = 0$, this reduces to

$$\overline{P_{\mathrm{acc}}(\tau)} = \mathrm{erfc}\left(\frac{\beta}{2}\sqrt{\frac{\sigma^2(\tau)}{2}}\right). \tag{7}$$

This is especially relevant in neMD/MC simulations where a splitting of the transition operator can be performed such that the free energy is effectively set to zero (see Chen and Roux[5]). The actual deviations from zero are subsumed into a proposal probability that is strictly independent of the switch process. The efficacy of this approach can be inferred from a graph of Eq. (6) which shows that the highest mean acceptance probability is obtained for small $|\Delta F|$ and $\sigma^2(\tau)$ (Fig. 1). A more rigorous proof is given by Chen and Roux.[5]

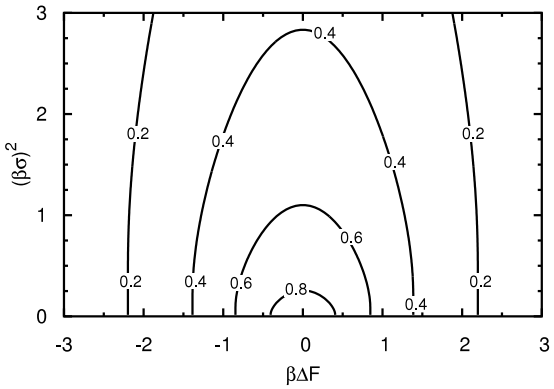

FIG. 1. Assuming Gaussian work distributions, the mean acceptance probability decreases monotonically as either a function of absolute relative free energy or variance. See Eq. (6).

In seeking an expression for $\sigma^2(\tau)$, it is noted that Eq. (2) is a standard departure point in nonequilibrium linear response theory.[20–22] Specifically, a mechanical observable $B(\boldsymbol{x})$ at equilibrium up to a perturbation at time $t = 0$ will have a mean response

$$\overline{\delta B(t)} \approx -\beta \int_0^t dt'\, \dot{\lambda}(t')\, C_{AB}\,(t - t') \tag{8}$$

to first order in perturbation of the phase space density. The effect of the coupling force enters through the definition of the equilibrium correlation function $C_{AB}(t) \equiv \langle \delta A(0) \delta B(t) \rangle$. The excess power from linear response is obtained by the choice $\overline{\delta B(t)} = -\dot{\lambda}(t) \overline{\delta A(t)}$,

$$\overline{\mathcal{P}_{\rm ex}(t)} = \beta \dot{\lambda}(t) \int_0^t dt'\, \dot{\lambda}(t')\, C_{AA}\,(t - t'). \tag{9}$$

Equation (9) was previously formulated by Sivak and Crooks[8] (Eq. (9) therein). The excess work follows

$$\overline{W_{\rm ex}(\tau)} = \beta \int_0^\tau dt\, \dot{\lambda}(t) \int_0^t dt'\, \dot{\lambda}(t')\, C_{AA}\,(t - t'). \tag{10}$$

The validity of this expression is apparent when taking the limit in which the perturbation is instantaneous. In this case $\dot{\lambda}(t)$ is a Dirac delta function and $\overline{W_{\rm ex}(\tau)} = \beta \langle \delta A(0)^2 \rangle / 2 = \beta \sigma(0)^2 / 2$. This recovers Eq. (5) for $\tau = 0$ and clarifies the connection between the Gaussian approximation and linear response.[13,16]

For any pair of states, the opposing protocols have formally different work distributions and correlation functions. However, linear response lifts this asymmetry if $\lambda(t)$ is assumed to be symmetric in time [i.e., $\lambda(t) = 1 - \lambda(\tau - t)$], since then Eq. (2) is invariant to swapping $H_0(\boldsymbol{x})$ and $H_1(\boldsymbol{x})$ (the signs on $A(t)$, $\dot{\lambda}(t)$, and ΔF are just inverted). The derivation for the backward protocol thus follows in a nearly identical way. The correspondence between Eqs. (5) and (10) now becomes apparent such that the work variance is a simple functional of a correlation function and the work protocol

$$\sigma^2(\tau) = 2 \int_0^\tau dt\, \dot{\lambda}(t) \int_0^t dt'\, \dot{\lambda}(t')\, C_{AA}\,(t - t'). \tag{11}$$

Since there is a single variance parameter these approximations also require that the correlation functions at both endpoints be *identical*. A simple and natural approximation is an exponentially decaying function, $C_{AA}(t) = \langle \delta A(0)^2 \rangle e^{-t/\tau_m}$, which introduces a molecular relaxation time τ_m. With this approximation and a choice of $\lambda(t)$, it is possible to solve Eq. (11) directly. The solutions can then be inserted into Eq. (7) to yield an expression for the mean acceptance probability in terms of τ only.

Three main classes of protocols are considered here: (1) step functions, (2) sigmoid-like functions, and (3) inverse sigmoid-like functions. The first class only approximately satisfies the symmetry requirements, although this appears to be no worse than other approximations employed here. In any event, this is primarily intended for illustrative purposes. The other protocols are introduced as Chebyshev expansions (3rd, 5th, or 9th order) of $\tanh(t/\tau)$ and $\arctanh(t/\tau)$, respectively. This is convenient, since a first order expansion reduces to a linear protocol. The solutions for a step function and arbitrary polynomial are given in the supplementary material. The linear solution is particularly simple

$$\sigma^2(\tau) = 2 \langle \delta A(0)^2 \rangle \frac{\tau_m^2}{\tau^2} \left(e^{-\tau/\tau_m} - 1 + \frac{\tau}{\tau_m} \right). \tag{12}$$

It is easily verified that $\lim_{\tau \to 0} \sigma^2(\tau) = \langle \delta A(0)^2 \rangle$ and $\lim_{\tau \to \infty} \sigma^2(\tau) = 0$, as expected.

III. COMPUTATIONAL METHODS

The above expressions were numerically tested by constant-pH neMD/MC simulations run with a developmental version of NAMD 2.11.[23] Propionic acid was modeled with the CGenFF force field and solvated in a ~24 Å cube containing 464 rigid mTIP3P water molecules.[24] Alchemical coupling between the protonated and deprotonated forms of propionic acid was accomplished via a "dual-topology" paradigm with additional, zero-length bonds between equivalent but otherwise non-interacting atoms (force constant 100 kcal/mol Å^2).[25] Langevin dynamics was employed at 300 K with a friction coefficient of 5 ps^{-1} and a 2 fs time step; all bonds with hydrogen were kept rigid. Periodic boundary conditions were applied using particle mesh Ewald (30 grid points per axis) and force switching of the Lennard-Jones interactions between 6 and 8 Å, after which an isotropic long-range approximation was used.

In order to minimize the statistical variance between different protocols and for convenient parallelization, a constant-pH-like MD simulation was run as follows. The system was first minimized, equilibrated (100 ps), and simulated (10 ns) at both non-alchemical endpoints; during production coordinate/velocity snapshots were saved at 10 ps intervals (2000 snaphots total). These snapshots were "alchemified" by adding atoms for the opposite alchemical endpoint via direct sampling (the phase space distributions are simple multivariate Gaussians). All nonequilibrium protocols were evaluated on this set of initial conditions (but with different pseudo-random seeds). Each set of trajectories generated in this fashion is statistically equivalent to a neMD/MC constant-pH simulation with the pH equal to the

reference pK_a value (since the protonated and deprotonated forms are equally populated).[5] Since no energy shift constants are required *a priori* in this context, these were determined via post-processing using the Bennett acceptance ratio on the trajectories generated with the longest switching time (40 ps).[26] All reported nonequilibrium work calculations were adjusted by this value (−67.33 kcal/mol) such that the effective free energy is zero.

A single work variance and distribution was estimated for each set of trajectories produced by the same protocol and switching time. This is because the forward and reverse work values can be pooled as a single sample within the Gaussian approximation (see the supplementary material). Accordingly, variances (and parametric density estimates) were computed using a modified maximum likelihood estimator. Non-parametric density estimates were made using a Gaussian kernel density estimator with the bandwidth selected according to Silverman's method.[27] Error bars were estimated based on the standard deviation of 100 bootstrap trials.

IV. RESULTS AND DISCUSSION

The observed work distributions for a linear work protocol with multiple switching times are plotted in Fig. 2. The distributions from other protocols display essentially the same pattern (not shown). The similarity between nonparametric and Gaussian parametric estimates qualitatively and quantitatively increases with switching time. This also coincides with a decrease in the difference between free energy estimates from the Bennett acceptance ratio method[26] and the Gaussian form of Crooks's fluctuation theorem.[19]

Equation (11) is tested directly by inserting each protocol, deriving a variance expression, and then performing nonlinear least squares regression on the observed variances (Fig. 3). It is apparent that a linear protocol always leads to the lowest variance amongst all protocols tested here. This was implicitly predicted (but not stated) by Sivak and Crooks[8] for the case that $\tau_m \ll \tau$ and $C_{AA}(t)$ is not a function of λ (as is consistent with linear response).

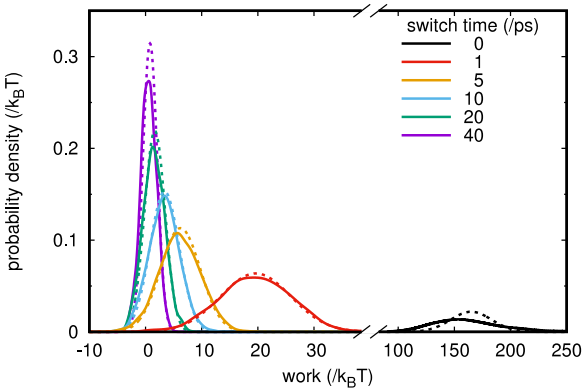

FIG. 2. Kernel density (solid lines) and Gaussian estimates for the work distribution (dashed lines) show that increasing the switch time for a linear protocol corresponds to both decreased peak width and modal work value. This also correlates with increasing similarity between the two estimates.

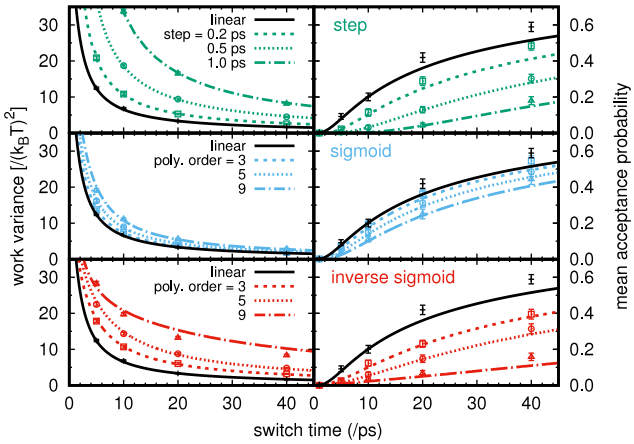

FIG. 3. The observed work variances and mean acceptance probabilities match well with those predicted from linear response. Least squares fitting is performed for each individual protocol for the variances only and then transformed via Eq. (7).

This result thus empirically addresses discussions in the literature regarding the efficacy of nonlinear protocols.[28–30] Both fit parameters (i.e., the variance and relaxation time) are, in principle, equilibrium constants and therefore the data sets can be fit globally. However, independently fitting the data leads to a broad range of values and superior residuals (see the supplementary material). All plots presented here show the independent fits. Rather than fitting Eq. (7) directly, the same fit parameters are used to compare with the mean acceptance probabilities obtained by bootstrap sampling (Fig. 3). This indirect fitting scheme seems to slightly underestimate the observed probability at longer switch times.

The accuracy of the theoretical framework outlined here is only semiquantitative (as evidenced by the poor global fit). However, treating $\langle \delta A(0)^2 \rangle$ and τ_m as adjustable parameters yields a predictive model that can be used to optimize neMD/MC efficiency by seeking to maximize the effective state-to-state transition rate during a simulation. Let us assume that n switches are to be attempted over a total simulation time τ_{tot}. Since only the time spent on switches directly impacts the transition rate, it is assumed, for simplicity, that all of this simulation time is spent on switches such that the switch length is $\tau = \tau_{tot}/n$. This is somewhat of an arbitrary construction, since the present framework does not offer any clear answers on how to divide computational resources between switching and nonswitching trajectories. With this in mind, the expected transition rate between states is $k(\tau) = n\overline{P_{acc}(\tau)}/\tau_{tot} = \overline{P_{acc}(\tau)}/\tau$. Maximizing $k(\tau)$ seeks to balance the cost of increasing the acceptance probability against decreasing the rate at which moves are attempted.

As seen in Fig. 4 extracted from constant-pH simulations of propionic acid in explicit solvent, this expression displays a distinct peak, implying that an optimal switching time τ_{opt} achieving maximum efficiency can be found. In this case, the most efficient neMD/MC scheme is obtained using an optimal switching time of ∼11 ps, which yields an effective transition rate of ∼20 ns^{-1}. Knowledge of $k_{opt} = k(\tau_{opt})$ can be used to estimate the minimal simulation time necessary in order

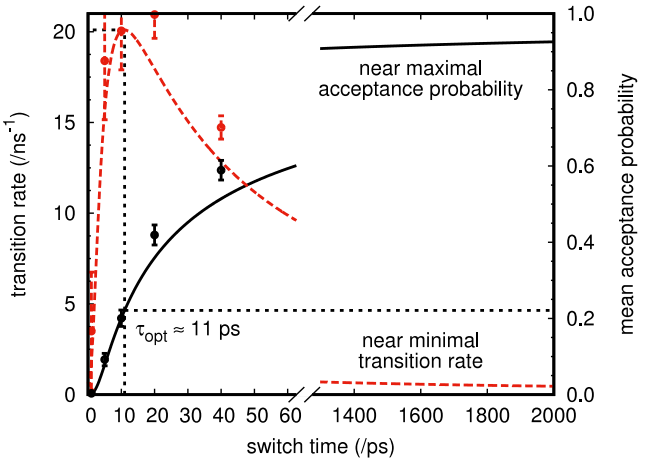

FIG. 4. The maximum of the expected transition rate (dashed line) does not coincide with that of the mean acceptance probability (solid line). Here the most efficient neMD/MC scheme is obtained using a linear work protocol with switch length ~11 ps, yielding an acceptance probability of ~22% and an effective transition rate of ~20 ns^{-1} (i.e., roughly one accepted transition every 50 ps).

to obtain a desired accuracy in the state populations. The total number of state-to-state transitions expected during a given simulation is on the order of $N_t = k_{opt}\tau_{tot}$. For example, assuming for the sake of argument that $N_t \approx 100$ transitions are necessary to obtain reasonable convergence, the optimal transition rate determined here suggests that a total simulation time of about 5 ns should be adequate to determine the relative population of the protonated and deprotonated states (neglecting the simulation time between switches). Because neMD/MC does not require an explicit enumeration of all states *a priori*, this argument may be extended to systems with a larger number of titratable sites (i.e., the algorithm is size extensive). Thus, for a system comprising N distinguishable titratable sites that are not strongly coupled to one another, a total simulation time proportional to NN_t/k_{opt} would be expected. However, in practice, the sites in a complex system could be strongly correlated and each transition is likely to have a different value of k_{opt}. For these reasons, the estimated simulation time is best regarded as a lower bound.

In principle, on-the-fly tuning of τ could be performed by gradually converging estimates for $\langle \delta A(0)^2 \rangle$ and τ_m. Unfortunately, gathering statistics from instantaneous switches alone is not a reliable approach since the approximations fail in this regime. A more plausible approach is based on the observation that fits involving very short trajectories are relatively consistent with longer trajectories. As such, statistics can be inexpensively obtained by many such short trajectories. Since these trajectories are unlikely to be accepted, it would even be possible to run them redundantly with multiple switch times and no intention of using them as MC candidates. This approach was tested by running 1 ns trajectories at both fixed protonation states. Work variances were accumulated by running 200 and 400 fs trajectories every 10 ps. The fit parameters from these two data points and Eq. (11) were then used to maximize the expected transition rate. The resulting estimates were ~6–7 ps, depending on the endpoint. This underestimates the global fit of ~11 ps. This test assumed perfect knowledge of the free energy; adding Gaussian error (mean and standard deviation 1 k_BT) to the work changed the predictions by <1 ps. The underestimation seems to be most strongly correlated with simultaneous underestimation of τ_m (~0.15 ps compared to 0.50). This trend in error is to be preferred in an adaptive scheme, since increasing τ too aggressively could waste time and cause instability due to inconsistent statistics.

The above discussion was focused on maximizing $k(\tau)$ in order to balance the cost of increasing $\overline{P_{acc}(\tau)}$ against decreasing the rate at which moves are attempted. Alternatively, if only enhanced sampling in configuration space is desired, another possible metric is the efficiency gain proposed by Nilmeier *et al.*[2] (Eq. (23) therein). Interestingly, although this quantity is quite different from $k(\tau)$, it also seems to predict a single maximum as a function of τ. However, the form of the efficiency gain is largely empirical and depends on dynamical parameters from both standard and neMD/MC simulations.

In closing this discussion, it should be mentioned that the present analysis considered only the protocol component of the work W in the Metropolis acceptance criterion [Eq. (1)] for the sake simplicity. Further analysis indicates that the present conclusions remain unchanged when the shadow work is taken into consideration. For instance, the contrast between $\overline{P_{acc}(\tau)}$ and $k(\tau)$ for large τ is likely only theoretical since, in that regime shadow work will likely dominate and both quantities become vanishingly small.[6,14] However, this caveat only reinforces the conclusion here that τ_{opt} be relatively small.

V. CONCLUSION

In conclusion, neMD/MC algorithms can be powerful tools for improving conventional simulations as well as for producing novel methods with an extended phase space. However, clear improvements can only be realized if well-defined metrics for efficiency are developed. Such a framework can be built around a generalized Metropolis criterion and linear response. Within this ansatz, the problem reduces to an expression for the work variance as a function of the work protocol and switching time. A substantial, but nonexhaustive, assessment of several protocols shows that a linear protocol is likely optimal; this is empirically validated by constant-pH MD. Although the effects of the switching time are more complex, they are well captured by two system dependent constants that derive from clear physical concepts, but are treated empirically in practice. The final results are theoretical expressions for the mean acceptance probability of MC moves and a related formula for the expected transition rate between states. The latter is demonstrated as a viable means for evaluating sampling and can be used in tandem with an adaptive procedure to estimate the necessary constants.

SUPPLEMENTARY MATERIAL

See supplementary material for discussion of shadow work, detailed derivations of specific results, and an extended description of simulation protocols.

ACKNOWLEDGMENTS

This research was supported under the Theta Early Science Program and used resources of the Argonne Leadership Computing Facility, which is a DOE Oce of Science User Facility supported under Contract DE-AC02-06CH11357. This work was also supported in part by the National Institutes of Health Grant No. U54-GM087519 (to B.R.) and by resources provided by the University of Chicago Research Computing Center. The authors are grateful to Yunjie Chen and John Chodera (Memorial Sloan-Kettering Cancer Center) for many helpful discussions.

[1]H. A. Stern, "Molecular simulation with variable protonation states at constant *pH*," J. Chem. Phys. **126**, 164112 (2007).
[2]J. P. Nilmeier, G. E. Crooks, D. D. L. Minh, and J. D. Chodera, "Nonequilibrium candidate Monte Carlo is an efficient tool for equilibrium simulation," Proc. Natl. Acad. Sci. U. S. A. **108**, E1009–E1018 (2011).
[3]A. J. Ballard and C. Jarzynski, "Replica exchange with nonequilibrium switches: Enhancing equilibrium sampling by increasing replica overlap," J. Chem. Phys. **136**, 194101 (2012).
[4]Y. Chen and B. Roux, "Efficient hybrid non-equilibrium molecular dynamics—Monte Carlo simulations with symmetric momentum reversal," J. Chem. Phys. **141**, 114107 (2014).
[5]Y. Chen and B. Roux, "Constant–pH hybrid nonequilibrium molecular dynamics–Monte Carlo simulation method," J. Chem. Theory Comput. **11**, 3919–3931 (2015).
[6]Y. Chen and B. Roux, "Generalized metropolis acceptance criterion for hybrid non-equilibrium molecular dynamics–Monte Carlo simulations," J. Chem. Phys. **142**, 024101 (2015).
[7]Y. Chen and B. Roux, "Enhanced sampling of an atomic model with hybrid nonequilibrium molecular dynamics–Monte Carlo simulations guided by a coarse-grained model," J. Chem. Theory Comput. **11**, 3572–3583 (2015).
[8]D. A. Sivak and G. E. Crooks, "Thermodynamic metrics and optimal paths," Phys. Rev. Lett. **108**, 190602 (2012).
[9]M. de Koning, "Optimizing the driving function for nonequilibrium free-energy calculations in the linear regime: A variational approach," J. Chem. Phys. **122**, 104106 (2005).
[10]T. Schmiedl and U. Seifert, "Optimal finite-time processes in stochastic thermodynamics," Phys. Rev. Lett. **98**, 108301 (2007).
[11]P. Geiger and C. Dellago, "Optimum protocol for fast-switching free-energy calculations," Phys. Rev. E **81**, 021127 (2010).
[12]G. M. Torrie and J. P. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling," J. Comput. Phys. **23**, 187–199 (1977).
[13]C. Jarzynski, "Nonequilibrium equality for free energy differences," Phys. Rev. Lett. **78**, 2690–2693 (1997).
[14]D. A. Sivak, J. D. Chodera, and G. E. Crooks, "Using nonequilibrium fluctuation theorems to understand and correct errors in equilibrium and nonequilibrium simulations of discrete Langevin dynamics," Phys. Rev. X **3**, 011007 (2013).
[15]D. A. Sivak, J. D. Chodera, and G. E. Crooks, "Time step rescaling recovers continuous-time dynamical properties for discrete-time Langevin integration of nonequilibrium systems," J. Phys. Chem. B **118**, 6466–6474 (2015).
[16]J. Hermans, "Simple analysis of noise and hysteresis in (slow-growth) free energy simulations," J. Phys. Chem. **95**, 9029–9032 (1991).
[17]G. E. Crooks, "Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems," J. Stat. Phys. **90**, 1481–1487 (1998).
[18]G. E. Crooks, "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences," Phys. Rev. E **60**, 2721–2726 (1999).
[19]J. Gore, F. Ritort, and C. Bustamante, "Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements," Proc. Natl. Acad. Sci. U. S. A. **100**, 12564–12569 (2003).
[20]R. Kubo, "The fluctuation–dissipation theorem," Rep. Prog. Phys. **29**, 255–284 (1966).
[21]D. A. McQuarrie, *Statistical Mechanics* (University Science Books, Mill Valley, CA, 1973).
[22]D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, New York, 1987).
[23]J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten, "Scalable molecular dynamics with NAMD," J. Comput. Chem. **26**, 1781–1802 (2005).
[24]K. Vanommeslaegh and A. D. MacKerell, Jr., "Automation of the CHARMM general force field (CGenFF) I: Bond perception and atom typing," J. Chem. Inf. Model **52**, 3144–3154 (2012); K. Vanommeslaegh, E. P. Raman, and A. D. MacKerell, Jr., "Automation of the CHARMM general force field (CGenFF) II: Assignment of bonded parameters and partial atomic charges," *ibid.* **52**, 3155–3168 (2012); A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," J. Phys. Chem. B **102**, 3586–3616 (1998).
[25]P. H. Axelsen and D. Li, "Improved convergence in dual-topology free energy calculations through use of harmonic restraints," J. Comput. Chem. **19**, 1278–1283 (1998).
[26]M. R. Shirts, E. Bair, G. Hooker, and V. S. Pande, "Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods," Phys. Rev. Lett. **91**, 140601 (2003).
[27]B. W. Silverman, *Density Estimation for Statistics and Data Analysis* (Chapman & Hall/CRC, London, 1998).
[28]L.-W. Tsao, S.-Y. Sheu, and C.-Y. Mou, "Absolute entropy of simple point charge model water by adiabatic switching processes," J. Chem. Phys. **101**, 2302–2308 (1994).
[29]M. de Koning and A. Antonelli, "Einstein crystal as a reference system in free energy estimation using adiabatic switching," Phys. Rev. E **53**, 465–474 (1996).
[30]M. de Koning and A. Antonelli, "Adiabatic switching applied to realistic crystalline solids: Vacancy-formation free energy in copper," Phys. Rev. B **55**, 735–744 (1997).