Sampling rare event energy landscapes via birth-death augmented dynamics

Benjamin Pampel [ORCID],[1] Simon Holbach [ORCID],[2,*] Lisa Hartung [ORCID],[2,†] and Omar Valsson [ORCID][1,3,‡]

[1]*Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany*
[2]*Institut für Mathematik, Johannes Gutenberg-Universität Mainz, Staudingerweg 9, 55099 Mainz, Germany*
[3]*Department of Chemistry, University of North Texas, Denton, Texas, USA*

A common problem that affects simulations of complex systems within the computational physics and chemistry communities is the so-called sampling problem or rare event problem where proper sampling of energy landscapes is impeded by the presences of high kinetic barriers that hinder transitions between metastable states on typical simulation time scales. Many enhanced sampling methods have been developed to address this sampling problem and more efficiently sample rare event systems. An interesting idea, coming from the field of statistics, was introduced in a recent work [Lu, Lu, and Nolen, Accelerating Langevin sampling with birth-death, arXiv:1905.09863] in the form of a novel sampling algorithm that augments overdamped Langevin dynamics with a birth-death process. In this work, we expand on this idea and show that this birth-death sampling scheme can efficiently sample prototypical rare event energy landscapes, and that the speed of equilibration is independent of the barrier height. We amend a crucial shortcoming of the original algorithm that leads to incorrect sampling of barrier regions by introducing an alternative approximation of the birth-death term. We establish important theoretical properties of the modified algorithm and prove mathematically that the relevant convergence results still hold. We investigate via numerical simulations the effect of various parameters, and we investigate ways to reduce the computational effort of the sampling scheme. We show that the birth-death mechanism can be used to accelerate sampling in the more general case of underdamped Langevin dynamics that is more commonly used in simulating physical systems. Our results show that this birth-death scheme is a promising method for sampling rare event energy landscapes.

I. INTRODUCTION

A common task in statistics, Bayesian inference, and machine learning is to sample a probability distribution π using sampling algorithms such as Monte Carlo [1,2], or Langevin dynamics (LD) [3,4]. However, it can be challenging to efficiently sample the probability distribution if it is multimodal and exhibits metastability. Then, the transition time to go between different high-probability modes is long compared to the simulation times that one can employ. In other words, a transition between modes is a rare event. This sampling problem has led to the development of a wide range of advanced sampling algorithms to more efficiently sample probability distributions [5–8].

A similar sampling or rare event problem is well known in the computational physics and chemistry communities [9–11].

There, one is interested in using atomistic molecular dynamics [12] simulations to understand the behavior of physical systems. For example, this could be the formation and growth of a crystal [13], the folding of a protein [14], unbinding of a ligand from a protein complex [15], and so forth. The rare event problem is viewed in terms of an energy landscape, given by the negative logarithm of the corresponding probability distribution, which is characterized by metastable states separated by high kinetic barriers that hinder transitions between states on typical simulation time scales. This energy landscape is called a free energy landscape if one considers the low-dimensional description of the system and its dynamics in the terms of so-called collective variables that capture the slow modes of the physical process. This rare event sampling problem has lead to the development of a wide range of so-called enhanced sampling methods within the molecular simulation field [9–11,16–27].

A common sampling strategy is to consider multiple independent simulations that are started from different initial conditions. Each simulation then explores a different area of the energy landscape and by pooling the simulations together, one can obtain improved sampling statistics and results [28,29]. However, for rare event systems, each independent simulation will still suffer from the same sampling issues due to a lack of transitions between metastable states, which will skew the sampling statistics and lead to incorrect results when the simulations are pooled together.

*s.holbach@uni-mainz.de
†lhartung@uni-mainz.de
‡omar.valsson@unt.edu

We can consider each independent simulation as a walker or a particle exploring an energy landscape. Thus, we can then view multiple independent simulations as an ensemble of independent particles that explore an energy landscape. To overcome the sampling problem, we can introduce some kind of interaction between the particles, for example by considering population dynamics for the particles, as has been done in various ways in different fields [5–8,30–40]. Multiple walkers are also routinely combined with other enhanced sampling methods to accelerate convergence [39–42].

In Ref. [7], the authors introduce an interesting algorithm for sampling multimodal probability distributions that augments overdamped Langevin dynamics with a birth-death process. Theoretically, this sampling scheme is formulated in terms of a Fokker-Planck birth-death equation that adds a birth-death term to the conventional Fokker-Planck equation. In practice, the scheme is formulated in terms of a set of particles, each diffusing on an energy landscape according to overdamped Langevin dynamics, but also interacting with each other via nonlocal moves that are determined by an approximation of the birth-death term. It is shown that this scheme greatly improves the sampling and leads to considerably faster convergence to the equilibrium probability distribution in comparison to overdamped Langevin dynamics without a birth-death process.

In this work, we explore the potential of this birth-death sampling scheme and show that it can efficiently sample prototypical rare event energy landscapes. We show that the original algorithm in Ref. [7] suffers from a deficiency that leads to incorrect sampling of barrier regions, and we amend this shortcoming by introducing an alternative approximation of the birth-death term. We establish important theoretical properties of the associated interacting particle systems and prove mathematically that the relevant convergence results still hold with our approximate birth-death term. Furthermore, we show that the birth-death mechanism can be used to accelerate sampling in the general case of underdamped Langevin dynamics that is more commonly used in simulating physical systems.

In Sec. II, we introduce the fundamental idea behind the method. Section III provides the theory behind the birth-death scheme and our alternative approximation of the birth-death term along with mathematical proofs. In Sec. IV, we present the algorithm and details on the implementation. In Sec. V, we show applications to prototypical rare event energy landscapes and investigate the effect of various parameters of the algorithm. Finally, in Sec. VI, we end with a few concluding remarks.

II. OVERVIEW OF THE METHOD

Before giving a formal theoretical description of the birth-death sampling scheme in the following section, we will provide here a simplified discussion of the basic mechanism of the method.

A. Motivation and context

The general starting point is a high-dimensional dynamical system whose dynamics we can describe precisely by propagating the system in time using molecular dynamics or Langevin dynamics simulations (or even Monte Carlo simulations). In order to better understand its metastable states and the transitions between them, it can, however, be much more insightful to study them in terms of only a few degrees of freedom of the system that capture the essential features one is interested in (so-called collective variables, see, e.g., Refs. [10,11,43–45]). In an atomistic simulation, for example, this might be the distances between selected atoms, or some dihedral angles, but also possibly more sophisticated quantities. Unfortunately, one typically cannot calculate the low-dimensional dynamics analytically from the full high-dimensional dynamics. However, it is possible to simulate the full system, track the low-dimensional degrees of freedom that are of interest, and then estimate features of the low-dimensional dynamics on this basis.

In this work, we will assume that the low-dimensional dynamics can be described by (overdamped) Langevin dynamics with respect to an energy landscape U. This is justified in practice if there is timescale separation between the slow degrees of freedom that define the energy landscape and the system's other degrees of freedom, in other words, they are adiabatically separated [46].

We will also make the (artificial) assumption that U is known *a priori* and therefore perform all simulations directly with respect to the low-dimensional space. This is typically not the case in practice (as there is no need to estimate U when it is known), and our work should be viewed as a first step in which we check how accurately and how fast our proposed sampling algorithm can estimate prototypical reference energy landscapes. Adjusting our algorithm in such a way that it can handle real applications will be the content of future work.

B. Description of the method

We address the sampling problem in a setup where we have multiple independent simulations, each of which we interpret as the trajectory of a particle. We consider an ensemble of N particles where each of them diffuses independently on the energy landscape U. Corresponding to U is an equilibrium distribution π that denotes the particle density in equilibrium. The energy landscape and the equilibrium distribution are related via the Boltzmann factor [see Eq. (2) below], so knowing (or estimating) U is equivalent to knowing (or sampling) π. In order to sample π (and, equivalently, estimate U) from the simulated data, we can consider two different averages as well as their combination: the time average and the particle average.

For time averaging, we consider an accumulated histogram, i.e., for a given particle we count how often it is observed at each position. In the limit of infinitely long simulation times this converges to the equilibrium distribution π. However, in rare event systems where energy barriers between metastable states are so high that transitions between them rarely occur on simulation time scales, a single particle will likely only explore one of the metastable states (depending on its initial position). Alternatively, we can for a given point in time average over the particles, i.e., we look at the current particle distribution in order to approximate π. By considering
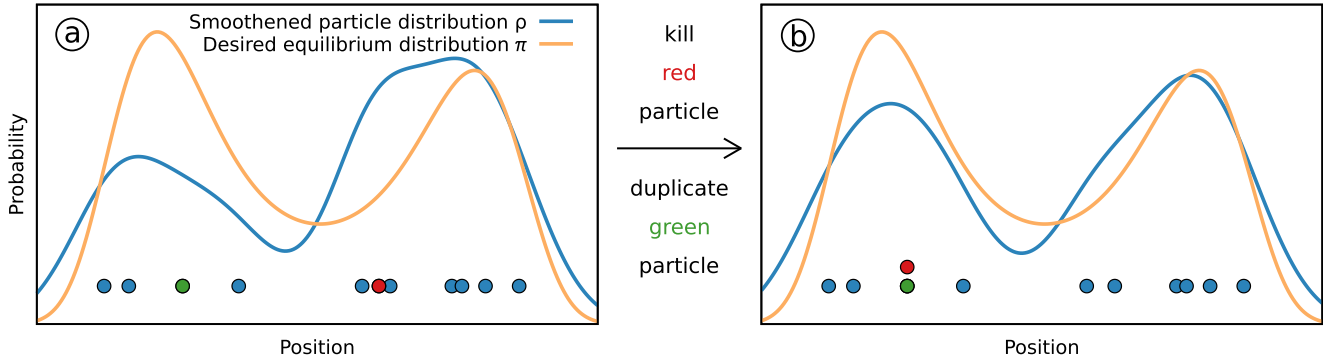
FIG. 1. Sketch of the method: (a) positions of particles (height irrelevant) and a smooth approximation of their ensemble particle distribution $\rho(x)$ (blue line) together with the desired equilibrium distribution $\pi(x)$ (orange line). At the position of the green particle the current ensemble particle distribution is significantly lower than the desired equilibrium distribution, while it is higher at the position of the red particle. (b) New ensemble particle distribution after the red particle of (a) has been killed and the green particle been duplicated. Effectively, this means that the red particle has performed a nonlocal move to the position of the green particle.

the whole ensemble of independent simulations, and combining the two averages into the ensemble average (i.e., averaging over the particles and time), we would hope to obtain better sampling statistics and results, but a lack of transitions between metastable states will typically still result in poor estimates of the barrier regions and the energy differences between different metastable states.

The birth-death sampling scheme that we present in this work aims to improve this situation by introducing non local particle moves. The basic idea is that if there is a significant difference between the particle distribution and the equilibrium one at some given time particles are killed and duplicated in such a way that this spatial redistribution reduces the deviation instantaneously.

To understand this idea a little better, let us consider the one-dimensional example shown in Fig. 1. Here, the energy landscape U is a double-well potential, so the equilibrium distribution π (orange line) features two local maxima with a local minimum in between. To be able to compare the discrete distribution of the particles to the continuous equilibrium density π more easily, we first smoothen it by placing centered Gaussian kernels at the particles' positions (similar to a kernel density estimation [47]) to arrive at the smoothened ensemble particle density ρ (blue line). In Fig. 1(a), we can see a significant difference between the equilibrium distribution that we want to sample and the current ensemble particle distribution: the region to the left is undersampled (i.e., $\rho < \pi$) and the region to the right is oversampled (i.e., $\rho > \pi$).

To bring the two distributions in better agreement, we now propose to kill and duplicate particles. While in the actual algorithm this decision would be made in a stochastic manner, we here consider an exemplary move based on intuition. We want to move some mass of the particle density from the oversampled area on the right to the undersampled region on the left. To achieve this, we decide to kill the red particle and duplicate the green particle. By doing this, the red particle has basically performed a nonlocal move from its original position to the position of the green particle. Figure 1(b) shows the resulting ensemble particle distribution where we can see that the two distributions now are in much better agreement. We have therefore managed to obtain a momentary particle

distribution that is closer to the one in equilibrium. Thus, by performing such nonlocal birth-death moves from time to time while in between the particles diffuse independently, we should quickly obtain an ensemble particle distribution that is in good agreement with the equilibrium distribution.

The crucial problem is now to perform these birth-death moves in such a way that we preserve global sampling statistics. Clearly, our manual killing and duplication of particles does not achieve this. We will therefore use a stochastic approach, where birth-death events occur at random times that depend on a slightly more involved comparison of the equilibrium distribution and the current local particle density [compare Eq. (10) below]. Analytically, this combination of Langevin dynamics and birth-death events can be expressed by a Fokker-Planck equation with an additional birth-death term, as will be introduced in the following section.

Recall that in practice π is typically not known *a priori*, making it more difficult to determine birth-death events based on how much the current particle distribution differs from it. Adjusting the birth-death mechanism in such a way that it still works with an on-the-fly estimation of π via enhanced sampling methods will be the content of future research.

III. THEORY

A. Overdamped Langevin dynamics with a birth-death process

The inertia-free motion of a particle with initial position $x(0) \in \mathbb{R}^d$ in the smooth potential (i.e., energy landscape) $U : \mathbb{R}^d \to \mathbb{R}$ is described by the overdamped Langevin equation

$$dx(t) = -D\beta \nabla U[x(t)] \, dt + \sqrt{2D} \, dW(t) \qquad (1)$$

where $D > 0$ is the diffusion coefficient, $\beta = 1/k_{\mathrm{B}}T > 0$ is the inverse thermal energy at temperature T with the Boltzmann constant k_{B}, and W is a standard Brownian motion on $\mathbb{R}^d$. Its solution $X = [x(t)]_{t \geqslant 0}$ is a Markov process that has a unique stationary distribution, the density of which is given by

$$\pi(x) = Z^{-1} \cdot e^{-\beta U(x)}, \qquad (2)$$

where Z is a normalizing constant (generally called the partition function in statistical physics). A constant shift of the potential does not change the dynamics described by Eq. (1). Furthermore, exact knowledge of the normalizing constant is not required for practical applications of the methodology described in this paper. Therefore, we may assume without loss of generality that $Z = 1$. Then

$$U(x) = -\beta^{-1} \log \pi(x) \quad (3)$$

and we will write everything in terms of $\pi(x)$ in the sequel, omitting the dependence on β. For functions of the space variable $x \in \mathbb{R}^d$, we will usually omit the argument unless it is specifically needed. We would like to stress that by Eqs. (2) and (3), knowing the energy landscape and knowing the equilibrium distribution is the same thing.

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a smooth function. The differential operator L that is the generator of X is given by

$$Lf = D\Delta f + D\boldsymbol{\nabla} \log \pi \cdot \boldsymbol{\nabla} f. \quad (4)$$

We denote its formal adjoint by $L^\dagger$, and it is given by

$$L^\dagger f = D\Delta f - D\boldsymbol{\nabla} \cdot (f\boldsymbol{\nabla} \log \pi) = D\boldsymbol{\nabla} \cdot \left(f\boldsymbol{\nabla} \log \frac{f}{\pi} \right). \quad (5)$$

The transition densities $\rho_t : \mathbb{R}^d \to [0, \infty)$ of the process X satisfy the corresponding (linear) Fokker-Planck equation

$$\partial_t \rho_t = L^\dagger \rho_t, \quad (6)$$

to which π is a stationary solution, i.e.,

$$L^\dagger \pi = 0. \quad (7)$$

If $f : \mathbb{R}^d \to \mathbb{R}$ is a probability density function and $g : \mathbb{R}^d \to \mathbb{R}$ is a formal probability density function (i.e., a convex combination of a probability density function in the classical sense and δ distributions), we define the unitless function $\alpha_\pi(f, g) : \mathbb{R}^d \to \mathbb{R}$ with

$$\alpha_\pi(f, g)(x) = \log \frac{f(x)}{\pi(x)} - \int \log \left(\frac{f(y)}{\pi(y)} \right) g(y) \, dy. \quad (8)$$

We then define the (nonlinear and nonlocal) Fokker-Planck birth-death equation as

$$\partial_t \rho_t = L^\dagger \rho_t - \tau_\alpha \alpha_\pi(\rho_t) \rho_t, \quad (9)$$

where

$$\alpha_\pi(\rho_t) = \alpha_\pi(\rho_t, \rho_t) \quad (10)$$

is called the birth-death term and $\tau_\alpha > 0$ is a rate factor that has units of 1/time. Note that the rate factor was not included in the Fokker-Planck birth-death equation in Ref. [7], in other words, the authors assumed that $\tau_\alpha = 1$. We have that

$$\alpha_\pi(\pi) = 0 \quad (11)$$

and hence π is also a stationary solution to the Fokker-Planck birth-death equation [Eq. (9)]. In other words, adding the birth-death term to the Fokker-Planck equation in this way does not change the equilibrium.

Note that the normalization constant of $\pi(x)$ cancels out in Eq. (8), so one can apply the birth-death term without knowing the normalization of the stationary distribution.

B. An interacting particle approach

The overdamped Langevin equation [Eq. (1)] can be thought of as the probabilistic counterpart of the Fokker-Planck equation [Eq. (6)]: the latter is solved by the transition densities ρ_t of the stochastic process X that solves the former. In a similar sense, we want to establish a probabilistic counterpart of the Fokker-Planck birth-death equation [Eq. (9)]. To achieve this, we introduce the following interacting particle system that is also described in Ref. [7].

We assume that there are N particles with positions $x_1(t), \dots, x_N(t) \in \mathbb{R}^d$ at time $t \geqslant 0$. By

$$\mu_t^N = \frac{1}{N} \sum_{k=1}^N \delta_{x_k(t)} \quad (12)$$

we denote the empirical measure of this N particle system, i.e., μ_t^N puts a mass of $1/N$ at each of the N particles' current positions. Since this is a singular measure and since we would like to be able to plug it into the birth-death term, we need to replace $\alpha_\pi(\cdot)$ with a smoothened approximation $\Lambda(\cdot)$, giving rise to an approximated Fokker-Planck birth-death equation

$$\partial_t \rho_t = L^\dagger \rho_t - \tau_\alpha \Lambda(\rho_t) \rho_t. \quad (13)$$

Now, we assume that each particle diffuses independently according to the overdamped Langevin dynamics defined in Eq. (1). On top of these independent dynamics, the particles interact via the following mechanism. Each particle has an independent exponential clock that strikes with the configuration-dependent birth-death rate

$$\tau_\alpha \left| \Lambda\left(\mu_t^N\right)[x_i(t)] \right|. \quad (14)$$

If the exponential clock for the ith particle strikes at time t, then one of two things happens:

(1) If $\Lambda(\mu_t^N)[x_i(t)] > 0$, then the ith particle is killed and a particle chosen uniformly at random from the others is duplicated.

(2) If $\Lambda(\mu_t^N)[x_i(t)] < 0$, then the ith particle is duplicated and a particle chosen uniformly at random from the others is killed.

Thus, the total number of particles is preserved. Alternatively, we can interpret this mechanism in the following way: in the first of the two cases above, the ith particle jumps to the current position of a random other particle, and in the second case a random other particle jumps to the current position of the ith particle.

Since Λ is a smoothened version of α_π, the term $\Lambda(\mu_t^N)[x_i(t)]$ approximately takes the logarithmic difference between the current particle density at $x_i(t)$ and the equilibrium density $\pi[x_i(t)]$, and then subtracts the average of the same quantity over all current particle positions $x_1(t), \dots, x_N(t)$. Therefore, the birth-death mechanism has a tendency to kill particles in space regions that are currently very crowded relative to the energy, and to duplicate particles in the opposite situation. Hence, we can expect this birth-death mechanism to help distribute the particles according to π and thus speed up the convergence of μ_t^N to π in comparison to a system without birth-death events.

We will now discuss three different choices of the approximation Λ, all of which feature a centered Gaussian kernel

$$K_{\Sigma}(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{x^{\top}\Sigma^{-1}x}{2}\right), \quad x \in \mathbb{R}^d, \quad (15)$$

where $\Sigma \in \mathbb{R}^{d \times d}$ is a positive definite invertible covariance matrix and $|\Sigma|$ denotes its determinant. Since Σ is usually fixed, we will often just write $K = K_{\Sigma}$. For our simulations, we are mostly interested in a diagonal covariance matrix, $\Sigma_{ij} = \delta_{ij}\sigma_i^2$ with $\sigma_i > 0$, where δ_{ij} is the Kronecker delta. In this case, $\sigma = (\sigma_1, \ldots, \sigma_d) \in \mathbb{R}^d$ will be referred to as the (vector of) bandwidths of the smoothing kernel. We denote the convolution of K with a (generalized) function f by $K * f$, and it is given by

$$K * f(x) = \int K(x - y)f(y)\,dy. \quad (16)$$

In particular, for the empirical distribution μ_t^N from Eq. (12), $K * \mu_t^N$ can be thought of as a kernel density estimation [47].

In Ref. [7], the authors use the approximation

$$\begin{aligned}\Lambda^0(f) &= \alpha_\pi(K * f, f) \\ &= \log \frac{K * f(x)}{\pi(x)} - \int \log\left(\frac{K * f(y)}{\pi(y)}\right)f(y)\,dy. \quad (17)\end{aligned}$$

This choice has one crucial shortcoming, as $\Lambda^0(\pi) \neq 0$ and hence π is not a stationary solution to the corresponding approximation in Eq. (13) of the Fokker-Planck birth-death equation. In other words, this approximation changes the equilibrium and hence using it as the basis for a sampling algorithm leads to sampling the wrong probability distribution.

One way to solve this problem is offered by the alternate approximation

$$\Lambda^{\mathrm{ad}}(f) = \Lambda^0(f) - \Lambda^0(\pi) \quad (18)$$

that adds a correction term to Λ^0. Clearly, $\Lambda^{\mathrm{ad}}(\pi) = 0$. Unfortunately, however, this additive correction is not particularly convenient for the mathematical analysis (see, e.g., the comments after Theorem 1). Hence, we consider the alternative multiplicative correction

$$\begin{aligned}\Lambda^{\mathrm{mu}}(f) &= \alpha_{K*\pi}(K * f, f) \\ &= \log \frac{K * f(x)}{K * \pi(x)} - \int \log\left(\frac{K * f(y)}{K * \pi(y)}\right)f(y)\,dy, \quad (19)\end{aligned}$$

where the equilibrium distribution π is also convoluted with a Gaussian kernel. Here, again, $\Lambda^{\mathrm{mu}}(\pi) = 0$. Unless explicitly stated otherwise, we will usually work with $\Lambda = \Lambda^{\mathrm{mu}}$.

Note that if we formally set $\Sigma = 0$ and interpret K_0 as the dirac delta δ_0, then all of the approximations $\Lambda^0, \Lambda^{\mathrm{mu}}, \Lambda^{\mathrm{ad}}$ coincide with the original birth-death term α_π. We establish two important theoretical properties of these interacting particle systems that explain and complement our findings in the subsequent sections on practical applications. First, we show that for fixed times $t > 0$, the empirical measure μ_t^N converges weakly to the solution ρ_t of the approximation of the Fokker-Planck birth-death equation [Eq. (13)] if the number N of particles tends to infinity (Theorem 1). In particular, this gives proper meaning to the idea that this interacting particle system is the probabilistic counterpart of the Fokker-Planck

birth-death equation. Second, we present reasonable assumptions under which ρ_t converges (exponentially fast) to π, as time t goes to infinity (Theorem 2). This rough summary is enough to understand the applications below, so the following section may be skipped on first reading.

Remark. In a recent talk [48], one of the authors of Ref. [7] presented the approximation

$$\Lambda^{0+}(f) = \log K * \frac{f}{\pi}(x) - \int \log\left(K * \frac{f}{\pi}(y)\right)f(y)\,dy, \quad (20)$$

which solves the main issue of Λ^0, as $\Lambda^{0+}(\pi) = 0$. Furthermore, it still satisfies Eq. (30), so Theorem 1 also holds with $\Lambda = \Lambda^{0+}$.

C. Theoretical results

Theorem 1. Let ρ_t be the solution to Eq. (13) with $\Lambda \in \{\Lambda^0, \Lambda^{\mathrm{mu}}\}$ and assume that μ_0^N converges weakly to the probability measure with density ρ_0 for $N \to \infty$. Then for all $t > 0$, the empirical measure μ_t^N of the interacting particle system converges weakly to the probability measure with density ρ_t for $N \to \infty$.

Nonrigorous proof. For the sake of notational simplicity and without loss of generality, we assume that $D = \tau_\alpha = 1$. Let $(x_n)_{n\in\mathbb{N}} \subset \mathbb{R}^d$. For any $N \in \mathbb{N}$, write

$$\mu_x^N = \frac{1}{N} \sum_{k=1}^{N} \delta_{x_k} \quad (21)$$

and let μ_x denote the formal limit of μ_x^N for the number N of particles going to infinity. The main idea of this proof is to show that for any smooth functional Ψ mapping a probability measure on $\mathbb{R}^d$ to an element of $\mathbb{R}^d$, we have convergence of $(\mathcal{L}_N\Psi)(\mu_x^N)$ to $(\mathcal{L}\Psi)(\mu_x)$, where $\mathcal{L}_N$ is the generator of the measure valued Markov process given by the empirical measure μ_t of the interacting particle system [see Eq. (12)], and $\mathcal{L}$ corresponds to how the right-hand side of Eq. (13) acts on the functional Ψ.

If μ_x^N is the current configuration of the system and the clock strikes for the ith particle, an index $j \in \{1, \ldots, N\} \setminus \{i\}$ is chosen uniformly at random, and the configuration changes to

$$\mu_x^N(x_i \to x_j) := \mu_x^N + \frac{1}{N}\mathrm{sgn}\big(\Lambda(\mu_x^N)(x_i)\big)(\delta_{x_j} - \delta_{x_i}). \quad (22)$$

This happens at rate $\Lambda(\mu_x^N)(x_i)$, and in between these birth-death events, each particle diffuses independently according to Eq. (1). Therefore, the infinitesimal generator $\mathcal{L}_N$ is given by

$$\begin{aligned}&(\mathcal{L}_N\Psi)\big(\mu_x^N\big) \\ &= \frac{1}{N} \sum_{i=1}^{N} \big(\Delta\Psi'_{\mu_x^N}(x_i) + \nabla\Psi'_{\mu_x^N}(x_i) \cdot \nabla\log\pi(x_i)\big) \\ &\quad + \frac{1}{N} \sum_{i,j=1}^{N} \big|\Lambda\big(\mu_x^N\big)(x_i)\big|\big(\Psi\big[\mu_x^N(x_i \to x_j)\big] - \Psi\big(\mu_x^N\big)\big)\end{aligned}$$

$$(23)$$

for any smooth functional Ψ, where Ψ' denotes its functional derivative, i.e.,

$$\int \Psi'_{\nu_0}(x)\nu(dx) = \lim_{\varepsilon \to 0} \frac{\Psi(\nu_0 + \varepsilon\nu) - \Psi(\nu_0)}{\varepsilon} \qquad (24)$$

for any probability measure ν_0 and any measure ν with $\int x\nu(dx) = 0$ (i.e., ν is centered). With the help of Eq. (21), we can rewrite Eq. (23) as

$$(\mathcal{L}_N\Psi)(\mu_x^N)$$
$$= \int \left[\Delta\Psi'_{\mu_x^N}(y) + \nabla\Psi'_{\mu_x^N}(y) \cdot \nabla\log\pi(y) \right]\mu_x^N(dy)$$
$$+ N \iint \left| \Lambda(\mu_x^N)(y) \right| \left\{ \Psi[\mu_x^N(y \to z)] - \Psi(\mu_x^N) \right\}$$
$$\times \mu_x^N(dy)\mu_x^N(dz). \qquad (25)$$

The formal limit of the first summand is

$$\int \left[\Delta\Psi'_{\mu_x}(y) + \nabla\Psi'_{\mu_x}(y) \cdot \nabla\log\pi(y) \right]\mu_x(dy). \qquad (26)$$

For the second summand, we note that by Eq. (24) with $\varepsilon = \frac{1}{N}$ and $\nu = \mathrm{sgn}[\Lambda(\mu_x^N)(y)](\delta_y - \delta_z)$, we have

$$\Psi[\mu_x^N(y \to z)] - \Psi(\mu_x^N)$$
$$\approx \frac{1}{N}\int \Psi'_{\mu_x^N}(u)\mathrm{sgn}[\Lambda(\mu_x^N)(y)](\delta_z - \delta_y)(du) \qquad (27)$$

for $N \to \infty$. Therefore, the second summand formally converges to

$$\iiint \Lambda(\mu_x)(y)\Psi'_{\mu_x}(u)(\delta_z - \delta_y)(du)\mu_x(dy)\mu_x(dz)$$
$$= \int \left(\int \Lambda(\mu_x)(z)\mu_x(dz) - \Lambda(\mu_x)(y) \right)\Psi'_{\mu_x}(y)\mu_x(dy), \qquad (28)$$

and hence $\mathcal{L}_N\Psi(\mu_x^N)$ converges to

$$(\mathcal{L}\Psi)(\mu_x)$$
$$= \int \left[\Delta\Psi'_{\mu_x}(y) + \nabla\Psi'_{\mu_x}(y) \cdot \nabla\log\pi(y) \right]\mu_x(dy)$$
$$+ \int \left(\int \Lambda(\mu_x)(z)\mu_x(dz) - \Lambda(\mu_x)(y) \right)$$
$$\times \Psi'_{\mu_x}(y)\mu_x(dy). \qquad (29)$$

Since for $\Lambda \in \{\Lambda^0, \Lambda^{\mathrm{mu}}\}$ we have

$$\int \Lambda(\mu_x)(z)\mu_x(dz) = 0, \qquad (30)$$

the formal limit generator $(\mathcal{L}\Psi)(\mu_x)$ from Eq. (29) corresponds to Eq. (13) with the respective choice of $\Lambda \in \{\Lambda^0, \Lambda^{\mathrm{mu}}\}$. ∎

Our proof uses the same arguments as the proof of Proposition 5.1 in Ref. [7] where the case $\Lambda = \Lambda^0$ was already treated. Note that up until Eq. (29), the proof works for any choice of Λ. However, if we use $\Lambda = \Lambda^{\mathrm{ad}}$, we have to be a little more careful, since then $\int \Lambda^{\mathrm{ad}}(\mu_x)(y)\mu_x(dy)$ does not vanish in general. The limit generator $(\mathcal{L}\Psi)(\mu_x)$ from Eq. (29)

then corresponds to Eq. (13) with

$$\Lambda(f) = \Lambda^{\mathrm{ad}}(f) - \int \Lambda^{\mathrm{ad}}(f)f\,dx. \qquad (31)$$

Unfortunately, it is currently unclear to us how one could say anything about the corresponding stationary solution.

In order to quantify the distance between ρ_t and π, we will use the Kullback-Leibler divergence

$$D_{\mathrm{KL}}(\rho_t|\pi) = \int \log\left(\frac{\rho_t}{\pi}\right)\rho_t\,dx, \qquad (32)$$

even though it is not a metric in the mathematical sense (as it is not symmetric and also violates the triangle inequality). However, it can be related to an actual metric, as Pinsker's inequality shows that the property $D_{\mathrm{KL}}(\rho_t|\pi) \to 0$ is stronger than convergence of ρ_t to π with respect to the total variation distance. We will also use the relative Fisher information

$$\mathcal{I}(\rho_t|\pi) = \int \left| \nabla\log\frac{\rho_t}{\pi} \right|^2 \rho_t\,dx. \qquad (33)$$

The following Theorem contains Theorem 3.2 of Ref. [7] as a special case, as for $K = \delta_0$ (i.e., $\Lambda = \alpha_\pi$), the second assumption trivially holds with $\lambda' = 0$.

Theorem 2. Let ρ_t the solution to Eq. (13) with $\Lambda = \Lambda^{\mathrm{mu}}$ and assume that the following conditions hold.

(1) There is a $\lambda > 0$ such that the log-Sobolev inequality

$$D_{\mathrm{KL}}(f|\pi) \leqslant \frac{1}{\lambda}\mathcal{I}(f|\pi) \qquad (34)$$

holds for all probability densities f on $\mathbb{R}^d$.

(2) There is a $\lambda' > -D\lambda/\tau_\alpha$ such that

$$\mathrm{Cov}_{\rho_t}\left(\log\frac{\rho_t}{\pi}, \log\frac{K*\rho_t}{K*\pi} \right) \geqslant \lambda' D_{\mathrm{KL}}(\rho_t|\pi) \qquad (35)$$

for all $t > 0$, where

$$\mathrm{Cov}_\rho(f, g) = \int fg\rho\,dx - \int f\rho\,dx \int g\rho\,dx. \qquad (36)$$

Then

$$D_{\mathrm{KL}}(\rho_t|\pi) \leqslant D_{\mathrm{KL}}(\rho_0|\pi)e^{-t(D\lambda+\tau_\alpha\lambda')} \xrightarrow{t\to\infty} 0. \qquad (37)$$

Proof. Since ρ_t is a smooth probability density, we have

$$\int (\partial_t\rho_t)\,dx = \partial_t \int \rho_t\,dx = 0 \qquad (38)$$

and hence

$$\partial_t D_{\mathrm{KL}}(\rho_t|\pi) = \int \partial_t\left(\log\left(\frac{\rho_t}{\pi}\right)\rho_t \right)dx$$
$$= \int (\partial_t\rho_t)\left(\log\frac{\rho_t}{\pi} + 1 \right)dx$$
$$= \int (\partial_t\rho_t)\log\frac{\rho_t}{\pi}\,dx. \qquad (39)$$

Plugging in Eq. (13) and using integration by parts yields

$$\partial_t D_{\mathrm{KL}}(\rho_t|\pi)$$

$$= \int \left(D\nabla \cdot \left(\rho_t \nabla \log \frac{\rho_t}{\pi} \right) - \tau_\alpha \Lambda^{\mathrm{mu}}(\rho_t)\rho_t \right) \log \frac{\rho_t}{\pi} \, dx$$

$$= -D\mathcal{I}(\rho_t|\pi) - \tau_\alpha \mathrm{Cov}_{\rho_t}\left(\log \frac{\rho_t}{\pi}, \log \frac{K * \rho_t}{K * \pi} \right). \quad (40)$$

Equation (40) together with our assumptions Eq. (34) and Eq. (35) implies

$$\partial_t D_{\mathrm{KL}}(\rho_t|\pi) \leqslant -(D\lambda + \tau_\alpha \lambda') D_{\mathrm{KL}}(\rho_t|\pi). \quad (41)$$

The claim now follows from Gronwall's Lemma. ∎

Note that the parameter $\lambda > 0$ in the log-Sobolev inequality Eq. (34) is present explicitly in the convergence rate in Eq. (37) and, of course, depends crucially on the potential U. This dependence can be described via the Eyring-Kramers formula for log-Sobolev inequalities (Corollary 2.17 in Ref. [49]). For a double-well potential, λ decreases exponentially with respect to the height of the energy barrier (compare Corollary 2.18 in Ref. [49]).

The covariance condition in Eq. (35) may be difficult to establish in practice, but let us present a rough idea why it is plausible. If ρ_t and π are sufficiently smooth, the amount to which they change after applying the smoothing kernel K is bounded uniformly in time and space by some constant times $|\boldsymbol{\Sigma}|$. If we pretend that π and ρ_t simply vanish entirely in very high-energy regions, we can then argue that

$$\int \left(\log \frac{\rho_t}{\pi} \right) \left(\log \frac{K * \rho_t}{K * \pi} \right) \rho_t \, dx$$

$$\geqslant \int \left(\log \frac{\rho_t}{\pi} \right) \left(\log \frac{K * \rho_t}{K * \pi} - \log \frac{\rho_t}{\pi} \right) \rho_t \, dx$$

$$\approx \varepsilon(\boldsymbol{\Sigma}) \cdot D_{\mathrm{KL}}(\rho_t|\pi), \quad (42)$$

where $|\varepsilon(\boldsymbol{\Sigma})|$ is small for $\boldsymbol{\Sigma} \to 0$. If our initial condition ρ_0 is not too far off, it is also reasonable to expect that $\rho_t \leqslant C\pi$ for some $C \in (1, \infty)$. If this holds, we can also estimate

$$\int \log \frac{\rho_t}{\pi} \rho_t \, dx \int \log \frac{K * \rho_t}{K * \pi} \rho_t \, dx \leqslant (\log C) D_{\mathrm{KL}}(\rho_t|\pi). \quad (43)$$

If $\boldsymbol{\Sigma}$ and C can be chosen suitably, combining Eq. (42) and Eq. (43) yields a version of the covariance condition in Eq. (35).

In Sec. S-I of the Supplemental Material (SM) [50], we show empirically that the covariance condition in Eq. (35) is satisfied for an exemplary simulation. There we find λ' to have a non-negative lower bound, so we can plausibly assume Eq. (35) to hold with $\lambda' = 0$, as is also the case when no smoothing kernel is applied at all.

The following Lemma gives meaning to the notion that increasing the bandwidth of the smoothing kernel corresponds to turning off the birth-death mechanism.

Lemma 1. If f is a probability density function on $\mathbb{R}^d$, then

$$\frac{K_{\boldsymbol{\Sigma}} * f}{K_{\boldsymbol{\Sigma}}} \xrightarrow{|\boldsymbol{\Sigma}| \to \infty} 1 \quad (44)$$

pointwise on $\mathbb{R}^d$. In particular, if

$$C^{-1}\pi \leqslant f \leqslant C\pi \quad (45)$$

for some $C > 1$, then

$$\Lambda^{\mathrm{mu}}(f) \xrightarrow{|\boldsymbol{\Sigma}| \to \infty} 0 \quad (46)$$

pointwise on $\mathbb{R}^d$.

Proof. First, Eq. (44) can be shown by a straightforward calculation and using dominated convergence. Then,

$$\frac{K_{\boldsymbol{\Sigma}} * f}{K_{\boldsymbol{\Sigma}} * \pi} = \frac{K_{\boldsymbol{\Sigma}} * f}{K_{\boldsymbol{\Sigma}}} \frac{K_{\boldsymbol{\Sigma}}}{K_{\boldsymbol{\Sigma}} * \pi} \xrightarrow{|\boldsymbol{\Sigma}| \to \infty} 1, \quad (47)$$

so that Eq. (46) follows from Eq. (45) and another application of dominated convergence. ∎

Note that

$$\Lambda_\pi^{\mathrm{ad}}(f) = \log \frac{K_{\boldsymbol{\Sigma}} * f}{K_{\boldsymbol{\Sigma}} * \pi} - \int \left(\log \frac{K_{\boldsymbol{\Sigma}} * f}{\pi} \right) f \, dx$$

$$+ \int \left(\log \frac{K_{\boldsymbol{\Sigma}} * \pi}{\pi} \right) \pi \, dx, \quad (48)$$

where Eq. (47) implies that the first term goes to zero for $|\boldsymbol{\Sigma}| \to \infty$, while the remaining terms do not depend on the position. Hence, even though the birth-death mechanism may not be turned off entirely in the limit, it no longer distinguishes between the different particles.

D. Underdamped Langevin dynamics case

Examining the behavior of the birth-death term for overdamped Langevin dynamics makes it feasible to mathematically prove convergence to the right distribution. However, for physical and chemical systems we often employ more general dynamics that take inertia into account and thus have to track not only the particle's position $x(t)$, but also its momentum $p(t)$. This is described by the underdamped Langevin equations

$$dx(t) = \frac{p(t)}{m} \, dt, \quad (49)$$

$$dp(t) = -\nabla U(x(t)) \, dt - \gamma p(t) \, dt + \sqrt{\frac{2m\gamma}{\beta}} \, dW(t), \quad (50)$$

where m denotes the particle mass, γ is a friction constant, and $\beta^{-1} = k_{\mathrm{B}}T$ is the thermal energy as before. Note that setting $dp(t) = 0$ in Eq. (50), plugging in Eq. (49), and rearranging the terms yields Eq. (1) with $D = (m\gamma\beta)^{-1}$.

The solution $(X, P) = [x(t), p(t)]_{t \geqslant 0}$ of Eqs. (49) and (50) is a $2d$-dimensional Markov process. It possesses a unique invariant distribution whose marginal with respect to the position x coincides with π, since in equilibrium, position and momentum become independent. In order to introduce a birth-death mechanism to an ensemble of N particles diffusing according to Eqs. (49) and (50), we follow the same approach as in Sec. III B and still use the same birth-death term $\Lambda(\mu_t^N)[x_i(t)]$ that depends only on the positions and ignores the momenta. When the ith particle is killed (or duplicated), the entire tuple (x_i, p_i) is killed (or duplicated).

While we do not present any theory for the resulting interacting particle system, we investigate it empirically in simulations in Sec. V C and verify that our algorithm can also successfully be used to sample π in the underdamped Langevin case. Additionally we provide an analysis of the

momentum distribution and correlation for an exemplary simulation in Sec. S-II of the SM. We conclude that the chosen approach does not result in deviation of the momentum distribution from the desired Maxwell-Boltzmann distribution [51] or deviation of the equilibration and is therefore justified for the presented system.

IV. IMPLEMENTATION

In the particle-based view, we can explicitly give the formula for the birth-death term. We can rewrite Eq. (19) as

$$
\Lambda^{\mathrm{mu}}(f)(x) = \log K * f(x) - \log K * \pi(x)
$$
$$
- \int \left(\log K * f(y) - \log K * \pi(y) \right) f(y) \, dy
$$
$$(51)$$

for all $x \in \mathbb{R}^d$. Since

$$
K * \delta_y(x) = \int K(x - z) \delta_y(z) \, dz = K(x - y) \quad (52)
$$

for any $x, y \in \mathbb{R}^d$ and since convolution is a linear operation, we can easily plug the empirical measure

$$
\mu_t^N = \frac{1}{N} \sum_{k=1}^{N} \delta_{x_k(t)} \quad (53)
$$

into Eq. (51). Dropping the time dependence of the particle positions $x_1(t), \ldots, x_N(t)$ for notational convenience, we find that

$$
\Lambda^{\mathrm{mu}}(\mu_t^N)(x_i) = \log \left(\frac{1}{N} \sum_{j=1}^{N} K(x_i - x_j) \right) - \log K * \pi(x_i)
$$
$$
- \frac{1}{N} \sum_{k=1}^{N} \left[\log \left(\frac{1}{N} \sum_{j=1}^{N} K(x_k - x_j) \right) \right.
$$
$$
\left. - \log K * \pi(x_k) \right]. \quad (54)
$$

In the sequel, we will simply write

$$
\Lambda_i := \Lambda^{\mathrm{mu}}(\mu_t^N)(x_i). \quad (55)
$$

We will always present results with $\Lambda = \Lambda^{\mathrm{mu}}$ in the following, although we show in the Supplemental Material [50] that similar results are obtained with $\Lambda = \Lambda^{\mathrm{ad}}$.

In our simulations we choose the covariance matrices of the Gaussian kernel as diagonal, $\Sigma_{ij} = \delta_{ij} \sigma_i^2$ with the bandwidths $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_d)$ where $\sigma_i > 0$. Then Eq. (15) turns into

$$
K(x) = K_{\boldsymbol{\sigma}}(x) = \frac{1}{(2\pi)^{d/2} \prod_{i=1}^{d} \sigma_i} \exp \left(- \sum_{i=1}^{d} \left(\frac{x^{(i)}}{\sqrt{2}\sigma_i} \right)^2 \right),
$$
$$(56)$$

where the sum goes over the d spatial dimensions of the state $x = (x^{(1)}, \ldots, x^{(d)}) \in \mathbb{R}^d$ of an individual particle.

Our algorithm mostly follows Algorithm 1 of Ref. [7] but with modifications to reduce the computational effort and the previously mentioned changes to the calculation of the approximate birth-death term Λ. First, while the original algorithm proposed to attempt birth-death events after every Langevin step, we do so only every M steps. This results in calculating the Λ values, which require the computationally involving density estimate ρ_t, less often. This has to be taken into account for the exponential clock: if q_i denotes the probability that the clock of the ith particle strikes after M steps, then these birth-death probabilities become

$$
q_i = 1 - \exp \left(-\tau_\alpha |\Lambda_i| M\theta \right), \quad (57)
$$

where θ denotes the Langevin time step. We will test if this results in deviation in the sampling. As in Ref. [7], we will use $\tau_\alpha = 1$ for the following applications but present a short discussion about the parameter in Sec. S-III in the SM [50].

Second, the original algorithm of Ref. [7] iterates over the particles, where it individually calculates the birth-death probability and executes accepted events immediately. Therefore, the values Λ_i have to be calculated for each particle individually, or at least recalculated from the new positions after each accepted birth-death event. For efficiency, we instead choose to calculate all birth-death rates $\tau_\alpha \Lambda_i$ from the positions only once before the birth-death step. Only the order in which the birth-death events are applied is randomized. No disadvantages could be found from this approach as long as the probabilities of birth-death events remain low, as will be investigated further in the following.

The algorithm was implemented in a custom PYTHON code together with Langevin solvers (i.e., integrators), and has been made available to the community [52]. Version v0.3.1 was used for all calculations in the following. We employ the Euler-Maruyama scheme [53] for overdamped Langevin dynamics and the Bussi-Parinello scheme [54] for the underdamped Langevin dynamics case. The input files and data supporting the results of this paper are openly available at Zenodo [55].

V. APPLICATIONS

As test cases, we choose to simulate the movement of sets of particles in artificial potentials that emulate prototypical energy landscapes. The focus is on rare event systems, where the energy landscapes are characterized by metastable states separated by high kinetic barriers (i.e., much higher than the thermal energy $k_{\mathrm{B}}T$) that hinder transitions between metastable states. Such rare event energy landscapes are common in the physical sciences so we foresee many applications that can benefit from this birth-death method.

A. Comparison of approximations Λ

To be able to show the effects of the different approximations Λ and other parameters, we start with a system with a moderate barrier height such that transitions are also observed within moderate simulation time by pure Langevin dynamics. We choose a one-dimensional double-well energy landscape that is described by the mathematical expression

$$
U(x) = x^4 - 4x^2 + 0.2x + C, \quad (58)
$$

where C is a constant so that $\min U(x) = 0$ to align it for easier comparison. A plot of this potential can be seen as a black

Algorithm 1. Birth-death augmented Langevin dynamics

Input:

- Potential U (and temperature T) corresponding to the equilibrium distribution π
- Langevin solver $L(X, P, U, \theta)$ with corresponding parameters
- Calculation rule for smoothed birth-death term Λ using Gaussian kernel K with bandwidths $\boldsymbol{\sigma}$
- Rate factor τ_α
- Langevin time step θ
- Number of Langevin steps J
- Number of Langevin steps between birth-death attempts M
- N particles with initial positions $X = \{x_i\}_{i=1}^N$ and momenta $P = \{p_i\}_{i=1}^N$

Output:

- Set of particles whose empirical measure approximates π

for $t \leftarrow 1$ **to** J **do**
    update X and P by Langevin solver $L(X, P, U, \theta)$
    if $(t \mod M) = 0$ **then**
        Calculate Λ for all particles
        Draw N independent random numbers $\{r_i\}_{i=1}^N$ uniformly from $[0, 1)$
        Make list ζ of indices i for which $r_i \leq q_i = 1 - \exp\left(-\tau_\alpha |\Lambda_i| M\theta\right)$
        Shuffle ζ randomly
        foreach $i \in \zeta$ [a] **do**
            Select particle j uniformly from all other particles
            if $\Lambda_i > 0$ **then**
                $x_i \leftarrow x_j;\ p_i \leftarrow p_j$
            else if $\Lambda_i < 0$ **then**
                $x_j \leftarrow x_i;\ p_j \leftarrow p_i$
            end if
        end foreach
    end if
end for

---

[a] In the foreach loop, we skip over all particles that were already killed randomly by a previous duplication event during the same birth-death step. This avoids duplicating the new position of a killed particle that was not actually considered for the event probability.

---

line in Fig. 2(a). It features two metastable states that have local minima at $x_L \approx -1.4$ and $x_R \approx 1.4$, and are separated by a barrier of $4.285\,k_B T$ that is centered at the local maximum at $x_0 \approx 0$. The metastable state at x_L on the left is lower in energy and thus has higher probability. In equilibrium, the probability of a particle being in the basin of attraction $B_L = (-\infty, x_0)$ for the left metastable state is $\pi(B_L) = \int_{-\infty}^{x_0} \pi(x)\,dx \approx 0.63$. Similarly, the probability for the higher-energy metastable state at x_R on the right is $\pi(B_R) = \pi[(x_0, \infty)] \approx 0.37$.

First, we test the influence of the present proposed birth-death term Λ^{mu} on the sampling compared to the previously proposed birth-death term Λ^0 from Ref. [7]. We use $N = 100$

particles and choose an initial distribution far from equilibrium: only ten particles are placed in the more likely left state at x_L, while the remaining 90 particles start in the less likely right state at x_R. We use the overdamped Langevin solver with a time step of $\theta = 0.001$ and set $D = 1$ (see Sec. S-IX of the SM [50] for a short discussion on the choice of the Langevin time step). We run the simulations for 2×10^6 steps at $T = 1$ so that the thermal energy is $k_B T = \beta^{-1} = 1$ (we use natural units such that $k_B = 1$). The number of steps between birth-death attempts is fixed to $M = 100$, while we investigate both Λ^0 and Λ^{mu} with different kernel bandwidths σ.

We assess the correctness of the sampling by obtaining estimates of the energy landscape from the simulations via histogramming. We bin all particle positions into a suitable histogram H and at the end of the simulations we calculate the estimated energy landscape $\tilde{U}(x)$ via

$$\tilde{U}(x) = -\beta^{-1} \log H(x) + \tilde{C}, \qquad (59)$$

where we choose the constant $\tilde{C}$ such that $\min \tilde{U}(x) = 0$. When constructing the histograms, we always omit the first 10^5 steps. We show results for the different kernel bandwidths σ in Figs. 2(a) and 2(b). We can observe that all simulations sample the basins and the lower regions of the energy landscapes correctly as indicated by the good agreement with the reference. However, we observe deviation from the reference energy landscape in the barrier region for some simulations, as can be seen in the insets in Figs. 2(a) and 2(b). In particular, the deviation is larger for the birth-death term Λ^0 from Ref. [7]. To quantify the deviation, we calculate from the estimated energy landscapes the height of the barrier going from the left minimum to the right minimum. The estimated barrier heights for both birth-death terms are shown in Fig. 2(c) as a function of the kernel width σ together with the reference value. We observe that for the original birth-death term Λ^0, the barrier height is always overestimated as compared to the reference value. In other words, the original birth-death term Λ^0 leads to an undersampling of the barrier region. On the other hand, the results for our birth-death term Λ^{mu} are much better and we only observe an undersampling if the kernel bandwidth is very small. The reason for this effect is likely the kernel density estimate in Eq. (54), as a too small bandwidth results in a very spiky density estimate. The lowest value of σ that results in correct sampling, σ_{crit}, depends on the system and the number of particles. This is investigated further in Sec. S-IV in the SM [50].

We further examine the performance of the different approximations by evaluating how quickly the birth-death algorithm manages to distribute the particles between the two states in the correct ratio according to the equilibrium distribution. In Fig. 2(d), we show the fraction of particles in left state, N_{left}/N, where N_{left} is the number of particles in the left state as defined by the basin of attraction $B_L = (-\infty, x_0)$. Note that we only show the initial 8000 steps of the simulations (i.e., the first 0.4% of the total simulation). Additionally, we show results obtained without birth-death events, that is, a pure overdamped Langevin dynamics simulation with the same parameters and number of particles but with the particles moving totally independent, so that the same amount of statistics are used to estimate the energy landscapes. The reference equilibrium value is $\pi(B_L) \approx 0.63$, in other words,
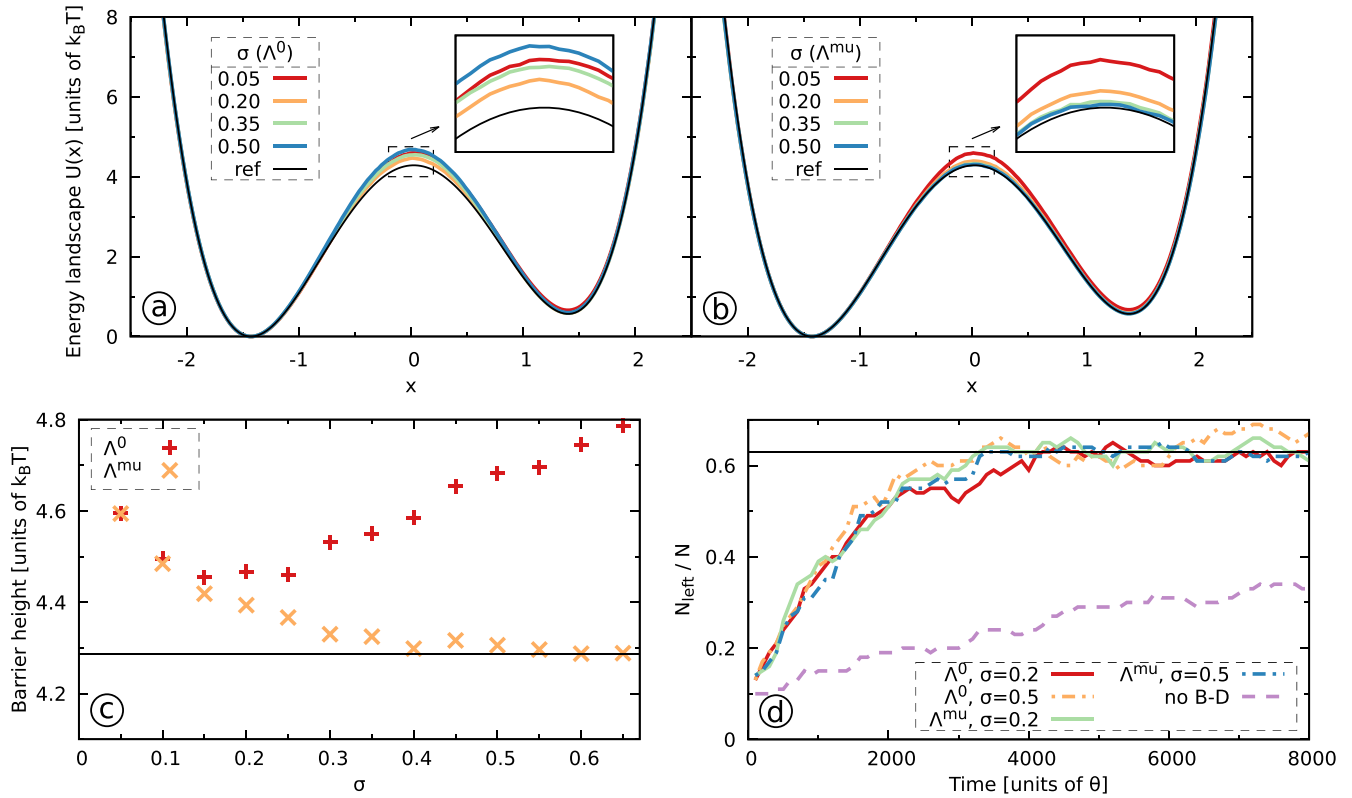
FIG. 2. Estimates of the energy landscape for the potential given in Eq. (58) obtained from sampling using histogramming for the different approximations to the birth-death term: (a) The original proposal Λ^0. (b) The present proposal Λ^{mu}. Both show results for different values of the kernel width σ (colored) as well as the reference (black). The small inset shows a magnification of the barrier region. (c) The height of the barrier going from the left minimum to the right minimum estimated from the energy landscape as a function of the kernel width. The reference value is given as a black horizontal line. (d) Fraction of particles in the left state as a function of simulation time. The black horizontal line is the expected equilibrium value. Shown are results from two exemplary simulations with different kernel widths for both approximations ($\sigma = 0.2$ as solid line, $\sigma = 0.5$ as lines with dashes and dots), as well as from a simulation without birth-death events (dashed line). Only the first 8000 steps of the simulations are displayed.

there should be around 63 particles in the left state in the current case of $N = 100$. As mentioned above, we start with a particle distribution far from equilibrium as initially there are only ten particles in the left state. Without the birth-death process, the simulation only slowly tends towards the equilibrium value and has not reached it within the time frame shown in Fig. 2(d). However, at longer times, the pure Langevin dynamics simulation reaches the correct equilibrium distribution. This is due to the fact that the moderate barrier height of the system allows for transitions from the Langevin dynamics alone within the simulation time, although the respective time scale of transition is long. In contrast, all simulations that employ the birth-death scheme quickly approach the correct equilibrium particle distribution and reach the reference equilibrium value $\pi(B_L) \approx 0.63$ with a few thousand Langevin steps.

We can see in Fig. 2(d) that similar results are obtain with our birth-death term Λ^{mu} and the original birth-death term Λ^0. Therefore, the incorrectness of the original birth-death term Λ^0 is mainly exhibited in the sampling of the barrier region while equilibrium properties seem to be less affected. Furthermore, we can see in Fig. S5 in the SM [50], that if we view the results in terms of the probability distribution, the issue with the undersampling of the barrier region with the original

birth-death term Λ^0 is barely noticeable. This can explain why this issue was not noticed in Ref. [7], as there the authors only viewed the results in terms of probability distributions. In this context it should be mentioned that obtaining accurate estimates of barriers is an important problem in computational physics and chemistry [56], so it is important that the birth-death scheme correctly samples the barrier region. As can be seen in Fig. S6 in the SM [50], we obtain overall similarly good results with the additive birth-death term Λ^{ad} in Eq. (18) as with the present multiplicative birth-death term Λ^{mu}.

As noted in Lemma 1, increasing the kernel bandwidth will gradually decreases the effect of the birth-death process. However, for the current case, we only observed a significant slowdown of the equilibration for very large kernel widths such as $\sigma = 5$, as shown in Sec. S-VI in the SM [50]. In practice, we should therefore focus on choosing σ large enough to get a smooth density estimate and correct sampling while keeping in mind that too large bandwidth values reduce the effectiveness of the method.

B. Influence of birth-death stride M

We next examine the effect of the birth-death stride M that determines how many Langevin dynamics steps are performed
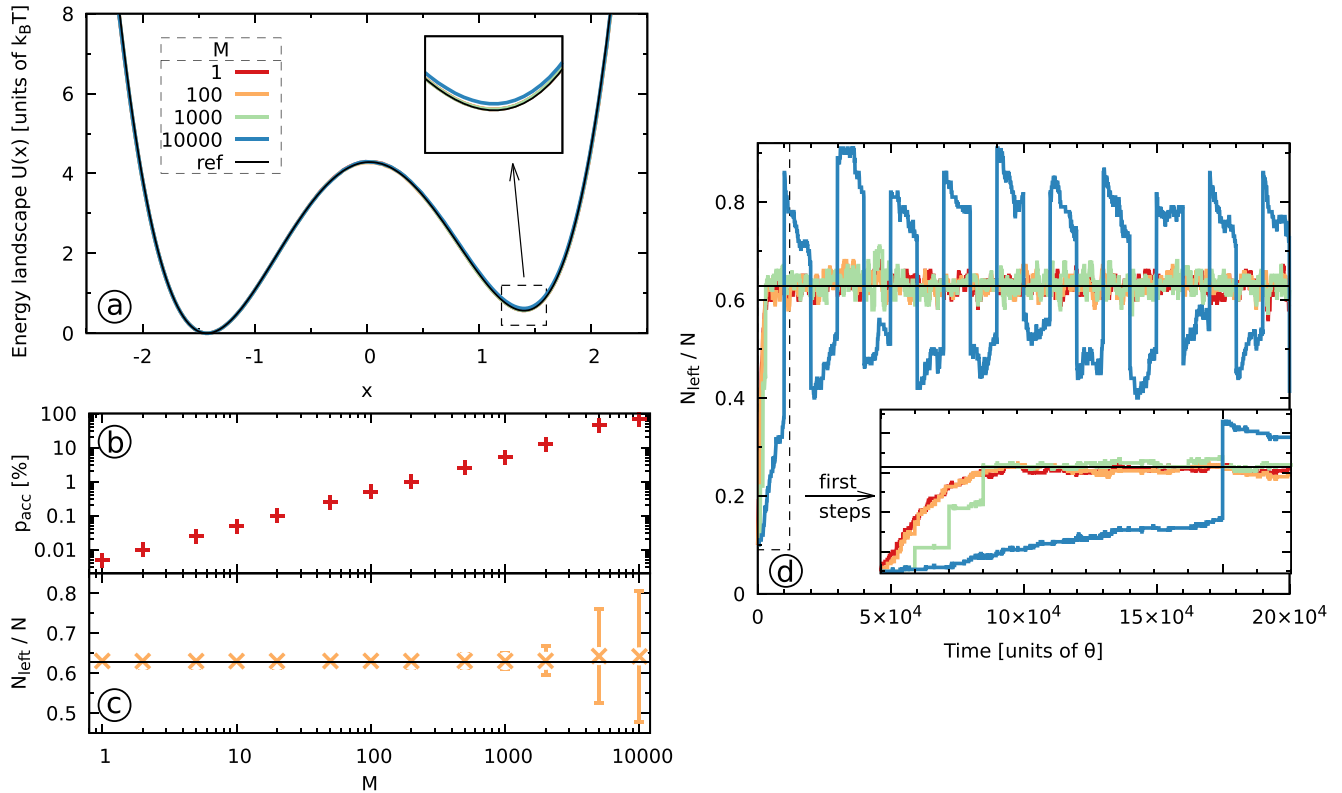
FIG. 3. (a) Estimates of the energy landscape for the potential given in Eq. (58) obtained from sampling using histogramming for different values M of Langevin dynamics steps between attempted birth-death events. All but the data for $M = 10000$ cannot be distinguished from the reference. (b) Percentage of accepted birth-death events of the total number of birth-death attempts. (c) Average fraction of particles in the left state. The error bars denote the standard deviation and the black horizontal line is the expected equilibrium value. The first 50000 steps were omitted when calculating the values. (d) Fraction of particles in the left state as a function of simulation time. The colored lines are from the same simulations as in (a), the black horizontal line is the expected equilibrium value. The inset shows a magnification of the first 12000 steps.

between birth-death attempts. To this end, we perform a set of simulations with the same double-well energy landscape as before and vary M while keeping the number of particles and the kernel width fixed at $N = 100$ and $\sigma = 0.4$ respectively. All other parameters are the same as in the previous section. As before, we start with an particle distribution far from equilibrium with 10% of the particles in the more probable left state and the rest in the right state. We use the same protocol as before and obtain estimates of the energy landscape using histogramming.

The results of the simulations are shown in Fig. 3. In Fig. 3(a), we present the estimated energy landscapes. In Fig. 3(b), we show the percentage of accepted birth-death events p_{acc} given by the number of executed birth-death moves divided by the total number of attempted ones. This can be understood as an estimate of the average birth-death probability of Eq. (57) during the simulations. In Fig. 3(d), we show the time evolution of the particle distribution for the two states obtained in the same way as in the previous section by considering the fraction of particles in the left state, N_{left}/N. Additionally, in Fig. 3(c), we show the mean and standard deviation of the fraction of particles in the left state, calculated by omitting the first 50000 steps of the equilibration phase.

We can see in Fig. 3(a) that all simulations yield a good estimate of the energy landscape, although the one with $M = 10000$ shows a slight deviation at the minimum of the right

state as can be seen in the inset. Looking at the time evolution of the particle distribution in Fig. 3(d), we see that for all simulations with $M < 10000$ the correct equilibrium distribution is reached within the first 4000 steps and there are only small fluctuations around the reference value afterwards. Smaller M values result in slightly faster equilibration, although we find this effect to be rather small.

For the simulation with $M = 10000$, the birth-death events result in overshooting, such that the number of particles in the left state becomes either too small or too large directly afterward. Between the birth-death events, a slow equilibration process due to the Langevin dynamics can be observed, as the moderate barrier height makes transitions only rare but not completely unlikely. The overshooting happens because we calculate the birth-death probabilities for all particles at once and then perform the respective events simultaneously. The time between birth-death calculations enters exponentially in the event probabilities in Eq. (57). For large values of M, the event probabilities thus become very large, and around 70% of the particles are killed or duplicated each time. While a per-particle approach with recalculation of the probabilities after each accepted event would solve the problem, this would also result in a lot more computational effort. An equivalent simulation with the recalculation of the birth-death probabilities after each birth-death event is shown in Sec. S-VII in the SM [50], where we can see that this solves the problem.

TABLE I. Coefficients for potentials according to Eq. (60) with increasing barrier height while keeping the equilibrium distribution of particles between the two states fixed so that $\pi(B_L) \approx 0.63$.

$a\ [k_BT]$	$b\ [k_BT]$	barrier height $[k_BT]$
1	0.2	4.285
2	0.1918	8.272
4	0.1889	16.267
8	0.1877	32.262

The reason for not calculating the probabilities at every Langevin step is to lower the computational effort. Therefore, we conclude that, as long as the birth-death events remain relatively rare, performing multiple Langevin steps between birth-death attempts helps to speed up simulations without negative side effects. To quantify this for the given system, we observe significant changes in the behavior only for $M > 1000$ in Fig. 3(c), which corresponds to $p_{acc} > 5\%$.

C. Underdamped Langevin dynamics and the effect of barrier height on the speed of equilibration

After investigating the influence of the parameters of the birth-death algorithm, we evaluate the behavior for the underdamped Langevin case that was introduced in Sec. III D. To simultaneously assess the speed of convergence for different barrier heights, we generalize the double-well energy landscape given in Eq. (58) with two parameters a and b:

$$U(x) = ax^4 - 4ax^2 + bx + C, \tag{60}$$

where, as before, C is a constant such that $\min U(x) = 0$. In Table I, we give sets of coefficients where we have systematically increased a and then set b such that the equilibrium distribution of the particles in the two states remains fixed at about $\pi(B_L) \approx 0.63$ to $\pi(B_R) \approx 0.37$ as it was in the previous sections, while the barrier height is increased.

For each set of coefficients, we perform simulations with the Langevin solver from Ref. [54]. We set $\beta = 1$, $m = 1$ and $\gamma = 10$ and simulate $N = 100$ particles for 2×10^6 steps with a time step of $\theta = 0.005$, where we again start with only ten particles in the left state and 90 in the right state. We perform birth-death events performed every $M = 100$ Langevin steps. We use the birth-death term Λ^{mu} and kernel width $\sigma = 0.5$. For comparison, we additionally perform pure Langevin dynamics simulations without birth-death events but otherwise the same parameters and numbers of particles. We perform analog analysis of the simulations as in the previous sections by estimating the energy landscapes using histogramming.

We show the results of the simulations in Fig. 4. For the simulations with birth-death events, we can see that the estimated energy landscape agrees well with the reference one in all cases. Note that for the higher barrier heights, the barrier region is not sampled due to low probability and insufficient simulation time, as can be expected.

On the contrary, we observe that the pure Langevin dynamics simulations without birth-death events are only sampling the system correctly if the barrier heights are low. Already for the system with a barrier height of $8\,k_BT$, there is a visible difference for the right state that is estimated to be lower in energy than the reference. For higher barriers, the estimates are completely off. They give the right state as lower in energy than the left one by several k_BT in comparison to the reference energy landscape.

We can see the reason for these results for the pure Langevin dynamics by looking at the distribution of the particles in the two states shown in Fig. 4(b). With increased barrier height, crossings between the states by pure Langevin dynamics become rarer. While the probability of crossing the barrier is not zero for the higher barrier heights, transitions are too rare to equilibrate the particles across the two states within the simulation time. In fact, we could not observe a single transition in the pure Langevin dynamics simulation for the system with a barrier height of $32\,k_BT$. Therefore, a pure Langevin dynamics simulation is unable to sample the energy landscape correctly, but this can be expected due to high barrier heights.

On the other hand, the simulations with birth-death events reach the equilibrium distribution of the particles very quickly, that is, within the first 1000 steps of the simulation. It is clearly
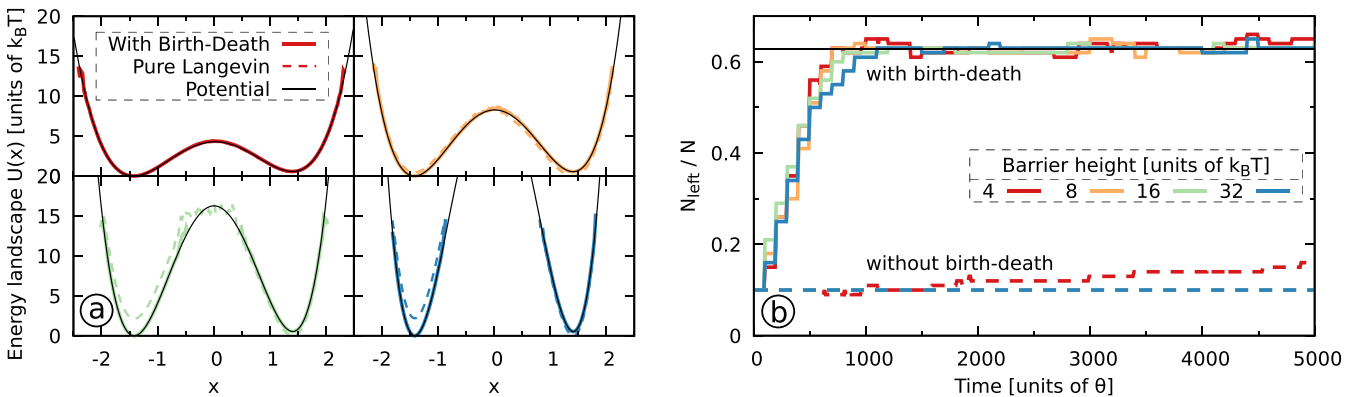

FIG. 4. (a) Estimates of the energy landscape obtained from sampling using histogramming for the potential given in Eq. (60) with different coefficients given in Table I. The solid lines are from simulations with birth-death events, the dashed lines from Langevin dynamics simulations without birth-death events (i.e., independent particles), and the thin black lines are the reference from the potential. (b) Number of particles in the left state as a function of simulation time for the different potentials. Shown are only the first 5000 steps. The solid black line is the equilibrium value calculated from the potentials (same for all potentials).
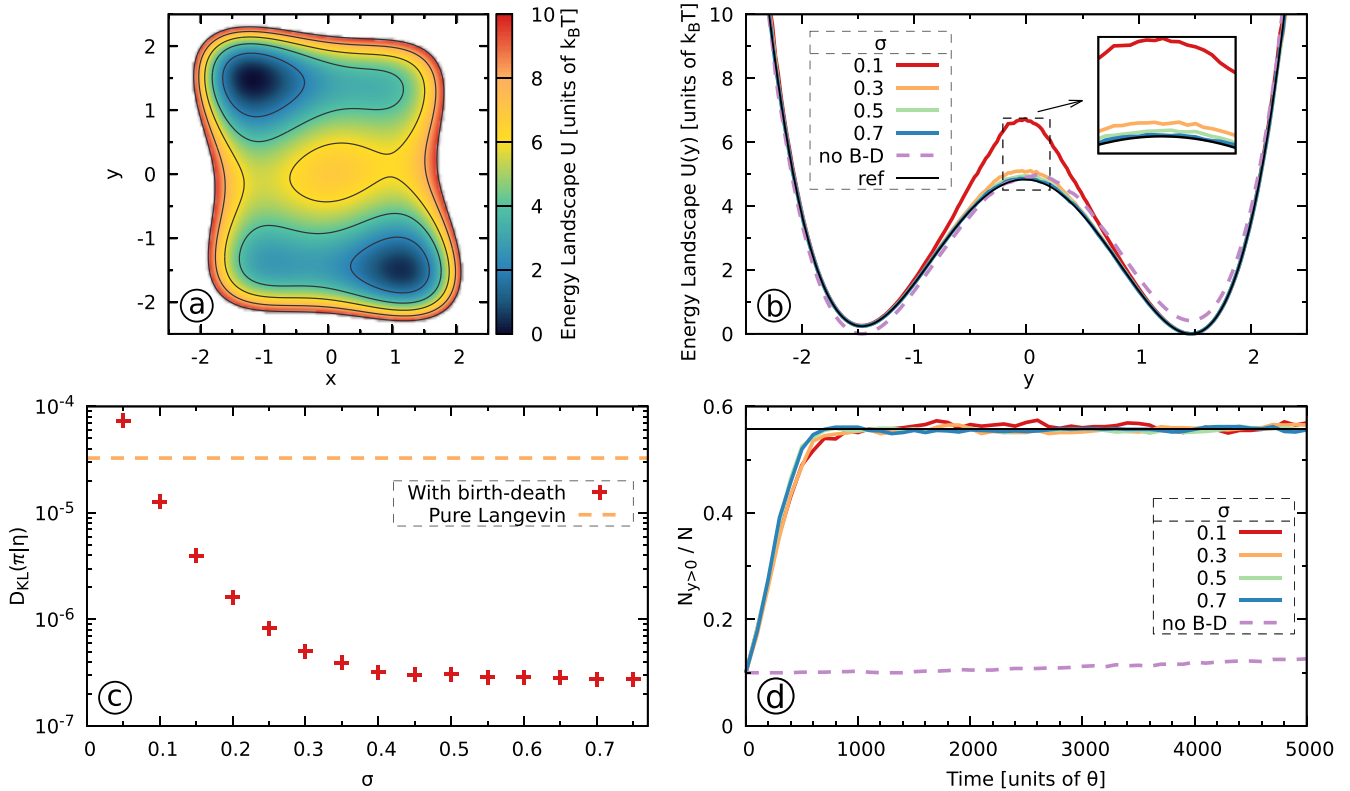
FIG. 5. (a) The reference energy landscape of the Wolfe-Quapp potential [Eq. (61)]. (b) The energy landscapes estimated from sampling using histogramming projected on the y-direction. Colored solid lines are from simulations with birth-death events using different kernel widths σ. We note that the lines for $\sigma = 0.55$ and $\sigma = 0.75$ are hardly distinguishable because they basically are on top of each other. The dashed line is from a Langevin simulation without birth-death events but the same number of particles. For clarity, the dashed line is omitted in the inset. The black line is the reference energy landscape calculated from the potential. (c) Kullback-Leibler divergences from the estimated probability distribution to the equilibrium distribution for simulations with different kernel widths σ. For comparison, the dashed horizontal line is from a Langevin simulation without the birth-death algorithm. (d) Fraction of particles in the state with $y > 0$ as a function of simulation time. The black horizontal line is the reference equilibrium value. The different lines represent the same simulations as in (b).

visible that the barrier height has only a negligible influence on the speed of equilibration, which is in accordance with a similar theoretic result for an overdamped system with no smoothing kernel (see Theorem 3.3 of Ref. [7]). Here, this theoretical derivation is found to be also true when using an approximation (in this case Λ^{mu}) to the birth-death term.

D. Higher dimensions: Two-dimensional Wolfe-Quapp potential

All of the previously presented simulations were performed for a system with only one spatial dimension. As the theory from Sec. III holds for higher dimensions, we also test the performance on the two-dimensional Wolfe-Quapp potential [57,58] given by

$$U(x, y) = x^4 + y^4 - 2x^2 - 4y^2 + xy + 0.3x + 0.1y + C,$$
$$(61)$$

where, as before, C is a constant such that $\min U(x, y) = 0$. This energy landscape can be seen in Fig. 5(a). Transitions between the states in y direction are rare events while the mobility in x direction is high, though the two coordinates are highly coupled.

We run simulations using $N = 1000$ particles for 200000 steps with the underdamped Langevin solver, but otherwise

use the same parameters as for the one-dimensional systems in Sec. V C. The initial distribution is again chosen to be far from equilibrium: we place 100 particles in the metastable state in the top left corner with a minimum at $(-1.17, 1.48)$ and 900 particles in the state in the bottom right corner with a minimum at $(1.12, -1.49)$. The bandwidths of the Gaussian kernel are chosen to be the same in each direction, $\sigma = \sigma_x = \sigma_y$, because the low-energy region of the potential has roughly the same size in both dimensions. We note that this is not a requirement and asymmetric kernels can be employed just as well. The kernel bandwidths are varied in the range $\sigma \in [0.05, 0.75]$ in steps of 0.05. Additionally, for comparison, we perform a pure Langevin dynamics simulation without the birth-death events but otherwise the same simulation protocol.

As before, we estimate the energy landscape by histogramming the simulations where we omit the first 10000 steps. Additionally, we assess the correctness of the sampling by using the Kullback-Leibler divergence [as defined in Eq. (32)] between the equilibrium probability distribution π and the estimated distributions η obtained from normalizing the histograms H of the simulations. A lower value of the Kullback-Leibler divergence indicates a better agreement of the estimated energy landscapes with the reference ones.

In Fig. 5, we show results of the simulations. As a visual inspection of two-dimensional (2D) energy landscapes is difficult, we choose to consider the projection onto the y direction that is shown in Fig. 5(b). We can observe similar results as for the 1D potential: above a certain kernel width σ_{crit}, the estimate energy landscapes are very close to the reference. This can also be seen in Kullback-Leibler divergence in Fig. 5(c). However, for too narrow kernels, the barrier regions are overestimated and we obtain a higher Kullback-Leibler divergence value. In Fig. 5(d), we can observe, that as before, the birth-death algorithm leads to a very swift equilibration of the particles between the two metastable states.

We can see that the pure Langevin simulation without birth-death event does not correctly sample the system and significantly deviates not only in the barrier region but also in the relative height of the two basins. This corresponds to the Kullback-Leibler divergence value that is more than one order of magnitude higher than results obtained with the birth-death simulations.

We also performed simulations with a lower number of particles, $N = 100$, compared to $N = 1000$ in Fig. 5, see results in Sec. S-VIII.1 in the SM [50]. We obtain similarly good results as for the case with more particles presented here, though the results are slightly more noisy.

Furthermore, we performed simulations with a scaled version of the Wolfe-Quapp potential, such that the barrier heights are increased, see results in Sec. S-VIII.2 in the SM [50]. Again, we observe there that the birth-death algorithm is able to obtain a correct sampling of the energy landscape, while pure Langevin sampling is unable to obtain good results. We can also see that the speed of equilibration to the correct particle distribution is independent of the barrier height, as observed in the previous section.

VI. SUMMARY AND OUTLOOK

In this paper, we have explored the usage of the birth-death scheme from Ref. [7] to sample rare event energy landscapes. We amend a deficiency of the original algorithm by introducing an approximate birth-death term that has the right mathematical limits and empirically leads to correct sampling of barrier regions between metastable states. We show empirically that the birth-death scheme can very efficiently sample prototypical rare event energy landscapes, both for overdamped and underdamped Langevin dynamics, and that the speed of equilibration is independent of the barrier height. We also show that the computational effort can be reduced by applying birth-death steps less frequently without negatively

affecting the quality of the sampling. This is an important point for future applications where we would like to reduce the communication between the different simulations.

Overall, our results show that this birth-death scheme is a promising sampling method that could extend not only Langevin dynamics but also other sampling schemes. We provide empirical evidence that the algorithm can also be used with more general sampling schemes, which motivates testing its applicability further, for example, to molecular dynamics or Monte Carlo simulations. However, to be able to apply it to simulations of high-dimensional systems, such as physical and chemical systems, will require considerable future work. To obtain smooth estimates of the particle distribution without increasing the kernel width (and therefore making the algorithm less efficient), the number of particles has to be increased with the number of dimensions, which makes the algorithm only applicable to a few degrees of freedom. A future extension could therefore modify the algorithm to perform the birth-death moves only in some relevant subspace, that is, in a low-dimensional space of a few collective variables. In other words, the birth-death algorithm would treat the particles that represent the simulations as if they were moving on the free energy landscape corresponding to these collective variables instead of the high-dimensional space. In this case, the free energy landscape, and thus the corresponding probability distribution that enters the birth-death term, is unknown *a priori*. One would need to estimate the probability distribution on the fly during the simulation, which could be done, for example, by combining the method with collective variable-based enhanced sampling methods. This is a subject for future work.

ACKNOWLEDGMENTS

[1] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. **21**, 1087 (1953).

[2] W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika **57**, 97 (1970).

[3] P. Langevin, Sur la théorie du mouvement brownien, C.R. Acad. Sci. **146**, 530 (1908).

[4] P. Turq, F. Lantelme, and H. L. Friedman, Brownian dynamics: Its application to ionic solutions, J. Chem. Phys. **66**, 3039 (1977).

[5] A. S. Sherman and C. S. Peskin, A Monte Carlo method for scalar reaction diffusion equations, SIAM J. Sci. Stat. Comput. **7**, 1360 (1986).

[6] Q. Liu and D. Wang, Stein variational gradient descent: A general purpose bayesian inference algorithm, in *Advances in Neural Information Processing Systems*, edited by D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (Curran Associates, Redhook, 2016), Vol. 29.

[7] Y. Lu, J. Lu, and J. Nolen, Accelerating Langevin sampling with birth-death, arXiv:1905.09863.

[8] M. Lindsey, J. Weare, and A. Zhang, Ensemble Markov Chain Monte Carlo with teleporting walkers, SIAM/ASA J. Uncert. Quant. **10**, 860 (2022).

[9] A. S. Kamenik, S. M. Linker, and S. Riniker, Enhanced sampling without borders: On global biasing functions and how to reweight them, Phys. Chem. Chem. Phys. **24**, 1225 (2022).

[10] O. Valsson, P. Tiwary, and M. Parrinello, Enhancing important fluctuations: Rare events and metadynamics from a conceptual viewpoint, Annu. Rev. Phys. Chem. **67**, 159 (2016).

[11] J. Hénin, T. Lelièvre, M. R. Shirts, O. Valsson, and L. Delemotte, Enhanced sampling methods for molecular dynamics simulations [article v1.0], LiveCoMS **4**, 1583 (2022).

[12] B. J. Alder and T. E. Wainwright, Studies in molecular dynamics. I. General method, J. Chem. Phys. **31**, 459 (1959).

[13] G. C. Sosso, J. Chen, S. J. Cox, M. Fitzner, P. Pedevilla, A. Zen, and A. Michaelides, Crystal nucleation in liquids: Open questions and future challenges in molecular dynamics simulations, Chem. Rev. **116**, 7078 (2016).

[14] R. B. Best, Atomistic molecular simulations of protein folding, Curr. Opin. Struct. Biol. **22**, 52 (2012).

[15] K. Ahmad, A. Rizzi, R. Capelli, D. Mandelli, W. Lyu, and P. Carloni, Enhanced-sampling simulations for the estimation of ligand binding kinetics: Current status and perspective, Front. Mol. Biosci. **9**, 899805 (2022).

[16] O. Valsson and M. Parrinello, Variational Approach to Enhanced Sampling and Free Energy Calculations, Phys. Rev. Lett. **113**, 090601 (2014).

[17] G. M. Torrie and J. P. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling, J. Comput. Phys. **23**, 187 (1977).

[18] T. Huber, A. E. Torda, and W. F. van Gunsteren, Local elevation: A method for improving the searching properties of molecular dynamics simulation, J. Comput.-Aided Mol. Des. **8**, 695 (1994).

[19] E. Darve and A. Pohorille, Calculating free energies using average force, J. Chem. Phys. **115**, 9169 (2001).

[20] U. H. E. Hansmann and L. T. Wille, Global Optimization by Energy Landscape Paving, Phys. Rev. Lett. **88**, 068105 (2002).

[21] J. Kästner, Umbrella sampling, WIREs Comput. Mol. Sci. **1**, 932 (2011).

[22] P. Maragakis, A. van der Vaart, and M. Karplus, Gaussian-mixture umbrella sampling, J. Phys. Chem. B **113**, 4664 (2009).

[23] A. Laio and M. Parrinello, Escaping free-energy minima, Proc. Natl. Acad. Sci. USA **99**, 12562 (2002).

[24] A. Barducci, G. Bussi, and M. Parrinello, Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method, Phys. Rev. Lett. **100**, 020603 (2008).

[25] J. K. Whitmer, C.-C. Chiu, A. A. Joshi, and J. J. de Pablo, Basis Function Sampling: A New Paradigm for Material Property Computation, Phys. Rev. Lett. **113**, 190602 (2014).

[26] M. Invernizzi and M. Parrinello, Rethinking metadynamics: From bias potentials to probability distributions, J. Phys. Chem. Lett. **11**, 2731 (2020).

[27] F. Giberti, G. A. Tribello, and M. Ceriotti, Global free-energy landscapes as a smoothly joined collection of local maps, J. Chem. Theory Comput. **17**, 3292 (2021).

[28] P. V. Coveney and S. Wan, On the calculation of equilibrium thermodynamic properties from molecular dynamics, Phys. Chem. Chem. Phys. **18**, 30236 (2016).

[29] A. Grossfield, P. N. Patrone, D. R. Roe, A. J. Schultz, D. Siderius, and D. M. Zuckerman, Best practices for quantification of uncertainty and sampling quality in molecular simulations, LiveCoMS **1**, 5067 (2019).

[30] J. B. Anderson, A random-walk simulation of the Schrödinger equation: H_3^+, J. Chem. Phys. **63**, 1499 (1975).

[31] W. R. Gilks, G. O. Roberts, and E. I. George, Adaptive direction sampling, The Statistician **43**, 179 (1994).

[32] D. Aldous and U. Vazirani, "Go with the winners" algorithms, in *Proceedings of the 35th Annual Symposium on Foundations of Computer Science* (IEEE Comput. Soc. Press, Santa Fe, 1994), pp. 492–501.

[33] G. A. Huber and S. Kim, Weighted-ensemble Brownian dynamics simulations for protein association reactions, Biophys. J. **70**, 97 (1996).

[34] P. Grassberger, Go with the winners: A general Monte Carlo strategy, Comput. Phys. Commun. **147**, 64 (2002).

[35] D. Bhatt, B. W. Zhang, and D. M. Zuckerman, Steady-state simulations using weighted ensemble path sampling, J. Chem. Phys. **133**, 014110 (2010).

[36] B. M. Austin, D. Y. Zubarev, and W. A. Lester, Quantum Monte Carlo and related approaches, Chem. Rev. **112**, 263 (2012).

[37] D. M. Zuckerman and L. T. Chong, Weighted ensemble simulation: Review of methodology, applications, and software, Annu. Rev. Biophys. **46**, 1 (2016).

[38] G. Rotskoff, S. Jelassi, J. Bruna, and E. Vanden-Eijnden, Global convergence of neuron birth-death dynamics, in *Proceedings of the 36th International Conference on Machine Learning*, edited by K. Chaudhuri and R. Salakhutdinov, Proceedings of Machine Learning Research Vol. 97 (PMLR, 2019), pp. 5508–5517.

[39] T. Lelièvre, M. Rousset, and G. Stoltz, Computation of free energy profiles with parallel adaptive dynamics, J. Chem. Phys. **126**, 134111 (2007).

[40] K. Minoukadeh, C. Chipot, and T. Lelièvre, Potential of mean force calculations: A multiple-walker adaptive biasing force approach, J. Chem. Theory Comput. **6**, 1008 (2010).

[41] P. Raiteri, A. Laio, F. L. Gervasio, C. Micheletti, and M. Parrinello, Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics, J. Phys. Chem. B **110**, 3533 (2006).

[42] O. Valsson and M. Parrinello, Well-tempered variational approach to enhanced sampling, J. Chem. Theory Comput. **11**, 1996 (2015).

[43] G. Fiorin, M. L. Klein, and J. Hénin, Using collective variables to drive molecular dynamics simulations, Mol. Phys. **111**, 3345 (2013).

[44] F. Pietrucci, Strategies for the exploration of free energy landscapes: Unity in diversity and challenges ahead, Rev. Phys. **2**, 32 (2017).

[45] F. Noé and C. Clementi, Collective variables for the study of long-time kinetics from molecular trajectories: theory and methods, Curr. Opin. Struct. Biol. **43**, 141 (2017).

[46] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, Oxford, 2001).

[47] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability No. 26 (Chapman & Hall/CRC, Boca Raton, 1998).

[48] Y. Lu, Sampling via birth-death dynamics (The Simons Institute for the Theory of Computing, 2021), https://simons.berkeley.edu/talks/sampling-birth-death-dynamics, accessed December 16, 2022.

[49] G. Menz and A. Schlichting, Poincaré and logarithmic Sobolev inequalities by decomposition of the energy landscape, Ann. Prob. **42**, 1809 (2014).

[50] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.107.024141 for further results and discussions, including; test of the assumptions of Theorem 2; momentum equilibrium in the general Langevin case; effect of varying the birth-death rate factor τ_α; influence of the number of particles N on the critical bandwidth σ_{crit}; additional figures for the comparison between approximations Λ; behavior for large kernel bandwidths; recalculation of birth-death probabilities for $M = 10000$; additional simulations with the 2D Wolfe-Quapp potential; influence of the choice of Langevin time step θ.

[51] D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, New York, 1987).

[52] B. Pampel, Birth-death augmented langevin dynamics sampler (version v0.3.1), Available at https://zenodo.org/record/5874773, 2022.

[53] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, corr. 3rd print ed., Applications of Mathematics No. 23 (Springer, Berlin, 1999).

[54] G. Bussi and M. Parrinello, Accurate sampling using Langevin dynamics, Phys. Rev. E **75**, 056707 (2007).

[55] B. Pampel, S. Holbach, L. Hartung, and O. Valsson, Dataset for "Sampling Rare-Event Energy Landscapes via a Birth-Death Process", Available at https://zenodo.org/record/5873264, 2022.

[56] K. M. Bal, S. Fukuhara, Y. Shibuta, and E. C. Neyts, Free energy barriers from biased molecular dynamics simulations, J. Chem. Phys. **153**, 114118 (2020).

[57] S. Wolfe, H. B. Schlegel, I. G. Csizmadia, and F. Bernardi, Chemical dynamics of symmetric and asymmetric reaction coordinates, J. Am. Chem. Soc. **97**, 2020 (1975).

[58] W. Quapp, A growing string method for the reaction pathway defined by a Newton trajectory, J. Chem. Phys. **122**, 174106 (2005).