D. Aristoff *et al.* (2014) "ParRep for Simulating Markov Chains,"
Applied Mathematics Research eXpress, Vol. 2014, No. 2, pp. 332–352
Advance Access publication August 12, 2014
doi:10.1093/amrx/abu005

The Parallel Replica Method for Simulating Long Trajectories of Markov Chains

David Aristoff[1], Tony Lelièvre[2], and Gideon Simpson[3]

[1]Department of Mathematics, Colorado State University, Fort Collins, CO 80523, USA, [2]Université Paris-Est, CERMICS (ENPC), INRIA, 6-8 Avenue Blaise Pascal, Cité Descartes, F-77455 Marne-la-Vallée, and [3]Department of Mathematics, Drexel University, Philadelphia, PA 19104, USA

Correspondence to be sent to: e-mail: daristof@umn.edu

The parallel replica dynamics, originally developed by A. F. Voter, efficiently simulates very long trajectories of metastable Langevin dynamics. We present an analogous algorithm for discrete time Markov processes. Such Markov processes naturally arise, for example, from the time discretization of a continuous time stochastic dynamics. Appealing to properties of quasistationary distributions, we show that our algorithm reproduces exactly (in some limiting regime) the law of the original trajectory, coarsened over the metastable states.

1 Introduction

We consider the problem of efficiently simulating time homogeneous Markov chains with *metastable states*: subsets of state space in which the Markov chain remains for a long time before leaving. By a Markov chain, we mean a *discrete time* stochastic process satisfying the Markov property. Heuristically, a set S is metastable for a given Markov chain if the Markov chain reaches local equilibrium in S much faster than it leaves S. We will define local equilibrium precisely below, using *quasistationary distributions* (QSDs). The simulation of an exit event from a metastable state using a naive integration technique can be very time consuming.

Received January 18, 2014; Revised May 27, 2014; Accepted June 23, 2014

Metastable Markov chains arise in many contexts. The dynamics of physical systems are often modeled by memoryless stochastic processes, including Markov chains, with widespread applications in physics, chemistry, and biology. In computational statistical physics (which is the main application field we have in mind), such models are used to understand macroscopic properties of matter, starting from an atomistic description. The models can be discrete or continuous in time. The discrete in time case has particular importance: even when the underlying model is continuous in time, what is simulated in practice is a Markov chain obtained by time discretization. In the context of computational statistical physics, a widely used continuous time model is the Langevin dynamics [17], while a popular class of discrete time models are the Markov State Models [7, 24]. For details, see [17, 25]. For examples of discrete time models not obtained from an underlying continuous time dynamics, see [5, 26]. In this article, we propose an efficient algorithm for simulating metastable Markov chains over very long time scales. Even though one of our motivations is to treat time discretized versions of continuous time models, we do not discuss errors in exit events due to time discretization; we refer the reader, for example, to [4] and references therein for an analysis of this error.

In the physical applications above, metastability arises from the fact that the microscopic time scale (i.e., the physical time between two steps of the Markov chain) is much smaller than the macroscopic time scale of interest (i.e., the physical time to observe a transition between metastable states). Both energetic and entropic barriers can contribute to metastability. Energetic barriers correspond to high energy saddle points between metastable states in the potential energy landscape, while entropic barriers are associated with narrow pathways between metastable states; see Figure 1.

Many algorithms exist for simulating metastable stochastic processes over long time scales. One of the most versatile such algorithms is the *parallel replica dynamics* (ParRep) developed by Voter [29] and Voter *et al.* [30]. ParRep can be used with both energetic and entropic barriers, and it requires no assumptions about temperature, barrier heights, or reversibility. The algorithm was developed to efficiently compute transitions between metastable states of Langevin dynamics. For a mathematical analysis of ParRep in its original continuous time setting, see [16, 27]. In this article, we present an algorithm which is an adaptation of ParRep to the discrete time setting. It applies to any Markov chain.

ParRep uses many replicas of the process, simulated in parallel asynchronously, to rapidly find transition pathways out of metastable states. The gain in efficiency over

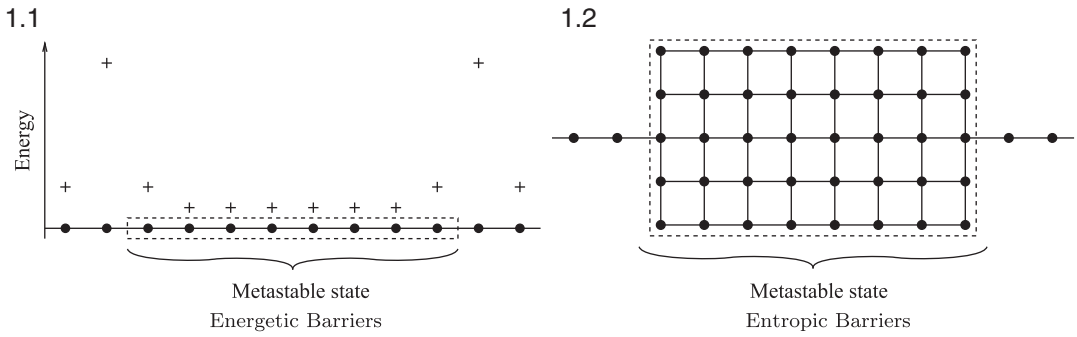

Fig. 1. (1.1) Energetic and (1.2) entropic metastable states of a discrete configuration space Markov chain. The chain jumps from one point to another according to the following Metropolis dynamics. If $X_n = x$, a direction (in (1.1), left or right; in (1.2), up, down, left, or right) is selected uniformly at random. If there is a point y which neighbors x in this direction, then with probability $\min\{1, e^{V(x)-V(y)}\}$ we take $X_{n+1} = y$; otherwise $X_{n+1} = x$. Here, V is a given potential energy function. On the left, each point has only two neighbors, and the potential energy is represented on the y-axis. On the right, each point has the same potential energy and between two and four neighbors.

direct simulation comes from distributing the computational effort across many processors, parallelizing the problem in time. The cost is that the trajectory becomes coarse-grained, evolving in the set of metastable states instead of the original state space. The continuous time version of ParRep has been successfully used in a number of problems in materials science (see, e.g., [1, 13, 15, 18, 22, 23, 28]), allowing for atomistic resolution while also reaching extended time scales of microseconds, 10^{-6} s. For reference, the microscopic time scale—typically the period of vibration of bond lengths—is about 10^{-15} s.

In the continuous time case, consistency of the algorithm relies on the fact that first exit times from metastable states are exponentially distributed. Thus, if N independent identically distributed (i.i.d.) replicas have first exit times T_i, $i = 1, \ldots, N$, then $N \min(T_1, \ldots, T_N)$ has the same law as T_1. Now if $K = \arg\min(T_1, \ldots, T_N)$ is the first replica which leaves the metastable state amongst all the replicas, then the simulation clock is advanced by NT_K, and this time agrees in law with the original process. In contrast, in the discrete time case, the exit times from metastable states are geometrically distributed. Thus, if τ_i are now the geometrically distributed first exit times, then $N \min(\tau_1, \ldots, \tau_N)$ does not agree in law with τ_1. A different function of the τ_i must be found instead. This is our achievement with Algorithm 3.1 and

Proposition 4.5. Our algorithm is based on the observation that $N[\min(\tau_1,\dots,\tau_N)-1]+\min[i\in\{1,\dots,N\},\ \tau_i=\min(\tau_1,\dots,\tau_N)]$ agrees in law with τ_1.

This article is organized as follows. In Section 2, we formalize the notion of local equilibrium using QSDs. In Section 3, we present our discrete time ParRep algorithm, and in Section 4 we study its consistency. Examples and a discussion follow in Section 5.

2 Quasistationary Distributions

Throughout this work, $(X_n)_{n\geq 0}$ will be a time homogeneous Markov chain with values in a probability space $(\Omega,\mathcal{F},\mathbb{P})$. For a random variable X and probability measure μ, we write $X\sim\mu$ to indicate X is distributed according to μ. For random variables X and Y, we write $X\sim Y$ when Y is a random variable with the same law as X. We write $\mathbb{P}^\mu(X_n\in A)$ and $\mathbb{E}^\mu[f(X_n)]$ to denote probabilities and expectations for the Markov chain $(X_n)_{n\geq 0}$ starting from the indicated initial distribution: $X_0\sim\mu$. In the case that $X_0=x$, we write $\mathbb{P}^x(X_n\in A)$ and $\mathbb{E}^x[f(X_n)]$ to denote probabilities and expectations for the Markov chain starting from x.

To formulate and apply ParRep, we first need to define the metastable subsets of Ω, which we will simply call *states*. The states will be used to coarse-grain the dynamics.

Definition 2.1. Let $\mathcal{S}$ be the collection of states, which we assume are disjoint bounded measurable subsets of Ω. We write S for a generic element of $\mathcal{S}$, and $\Pi:\Omega\to\Omega/\mathcal{S}$ for the quotient map identifying the states. □

As we will be concerned with when the chain exits states, we define the first exit time from S,

$$\tau:=\min\{n\geq 0:\ X_n\notin S\}.$$

Much of the algorithm and analysis depends on the properties of the QSD, which we now define.

Definition 2.2. A probability measure ν with support in S is a QSD if for all measurable $A\subset S$ and all $n\in\mathbb{N}$,

$$\nu(A)=\mathbb{P}^\nu(X_n\in A\,|\,\tau>n). \tag{1}$$

□

Of course, both τ and ν depend on S, but for ease of notation, we do not make this explicit. The QSD can be seen as a local equilibrium reached by the Markov chain,

conditioned on the event that it remains in the state. Indeed, it is easy to check that if ν is a measure with support in S such that,

$$\text{for any measurable } A \subset S \text{ and any } \mu \text{ with support in } S, \quad \nu(A) = \lim_{n\to\infty} \mathbb{P}^{\mu}(X_n \in A \,|\, \tau > n), \tag{2}$$

then ν is the QSD, which is then unique. In Section 4.1, we give sufficient conditions for existence and uniqueness of the QSD and for the convergence (2) to occur (see Theorem 4.2). We refer the reader to [6, 8, 9, 16, 20, 21] for additional properties of the QSD.

3 The Discrete Time ParRep Algorithm

Using the notation of the previous section, the aim of the ParRep algorithm is to efficiently generate a trajectory $(\hat{X}_n)_{n\geq 0}$ evolving in $\Omega/\mathcal{S}$ which has, approximately, the same law as the reference coarse-grained trajectory $(\Pi(X_n))_{n\geq 0}$. Two of the parameters in the algorithm—$T_{\text{corr}} = T_{\text{corr}}(S)$ and $T_{\text{phase}} = T_{\text{phase}}(S)$, called the *decorrelation* and *dephasing times*—depend on the current state S, but for ease of notation we do not indicate this explicitly. See the remarks below Algorithm 3.1.

Algorithm 3.1. Initialize a reference trajectory $X_0^{\text{ref}} \in \Omega$. Let N be a fixed number of replicas and T_{poll} be a fixed polling time at which the replicas resynchronize. Set the simulation clock to zero: $T_{\text{sim}} = 0$. A coarse-grained trajectory $(\hat{X}_n)_{n\geq 0}$ evolving in $\Omega/\mathcal{S}$ is obtained by iterating the following:

Decorrelation Step: Evolve the reference trajectory $(X_n^{\text{ref}})_{n\geq 0}$ until it spends T_{corr} consecutive time steps in some state $S \in \mathcal{S}$. Then proceed to the dephasing step. Throughout this step, the simulation clock T_{sim} is running and the coarse-grained trajectory is given by

$$\hat{X}_{T_{\text{sim}}} = \Pi(X^{\text{ref}}_{T_{\text{sim}}}). \tag{3}$$

Dephasing Step: The simulation clock T_{sim} is now stopped and the reference and coarse-grained trajectories do not evolve. Evolve N independent replicas $\{X_n^j\}_{j=1}^N$ starting at some initial distribution with support in S, such that whenever a replica leaves S it is restarted at the initial distribution. When a replica spends T_{phase} consecutive time steps in S, stop it and store its end position. When all the replicas have stopped, reset each replica's clock to $n = 0$ and proceed to the parallel step.

Parallel Step: Set $M = 1$ and iterate the following:

1. Evolve all N replicas $\{X_n^j\}_{j=1}^N$ from time $n = (M-1)T_{\text{poll}}$ to time $n = MT_{\text{poll}}$. The simulation clock T_{sim} is not advanced in this step.
2. If none of the replicas leaves S during this time, update $M = M + 1$ and return to 1, above.

   Otherwise, let K be the smallest number j such that X_n^j leaves S during this time, let τ^K be the corresponding (first) exit time, and set

$$X_{\text{acc}} = X_{\tau^K}^K, \quad T_{\text{acc}} = (N-1)(M-1)T_{\text{poll}} + (K-1)T_{\text{poll}} + \tau^K. \tag{4}$$

   Update the coarse-grained trajectory by

$$\hat{X}_n = \Pi(S) \quad \text{for} \quad n \in [T_{\text{sim}}, T_{\text{sim}} + T_{\text{acc}} - 1], \tag{5}$$

   and the simulation clock by $T_{\text{sim}} = T_{\text{sim}} + T_{\text{acc}}$. Set $X_{T_{\text{sim}}}^{\text{ref}} = X_{\text{acc}}$, and return to the decorrelation step.

□

The idea of the parallel step is to compute the exit time from S as the sum of the times spent by the replicas up to the first exit observed among the replicas. More precisely, if we imagine the replicas being ordered by their indices (1 through N), this sum is over all N replicas up to the last polling time, and then over the first K replicas in the last interval between polling times, K being the smallest index of the replicas which are the first to exit. Note that M and τ^K are such that $\tau^K \in [(M-1)T_{\text{poll}} + 1, MT_{\text{poll}}]$. See Figure 2 for a schematic of the Parallel Step. We comment that the formula for updating the simulation time in the parallel step of the original ParRep algorithm is simply $T_{\text{acc}} = N\tau^K$.

A few remarks are in order (see [16, 27] for additional comments on the continuous time algorithm):

The Decorrelation Step. In this step, the reference trajectory is allowed to evolve until it spends a sufficiently long time in a single state. At the termination of the decorrelation step, the distribution of the reference trajectory should be, according to (2), close to that of the QSD (see Theorem 4.2 in Section 4.1).

The evolution of the reference trajectory is *exact* in the decorrelation step, and so the coarse-grained trajectory is also exact in the decorrelation step.

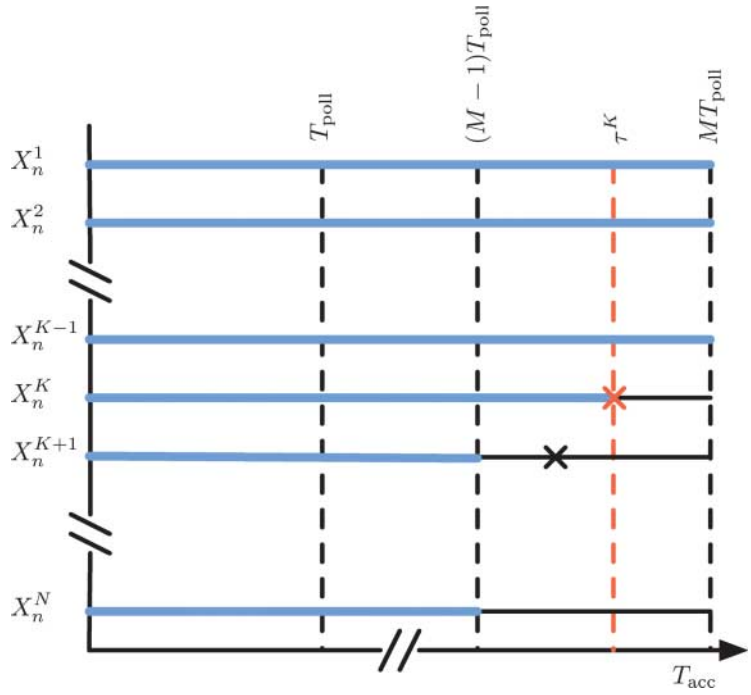

Fig. 2. A schematic of the parallel step. The horizontal lines represent the trajectories of replicas $1, \ldots, N$ while the crosses correspond to exit events. Index K is as defined as in Algorithm 3.1. Here, M cycles internal to the parallel step have taken place. The thicker lines correspond to the portions of the chains contributing to T_{acc}.

The Dephasing Step. The purpose of the dephasing step is to generate N i.i.d. samples from the QSD. While we have described a simple rejection sampling algorithm, there is another technique [3] based on a branching and interacting particle process sometimes called the Fleming–Viot particle process [11]. See [2, 9, 12, 19, 21] for studies of this process, and [3] for a discussion of how the Fleming–Viot particle process may be used in ParRep.

In our rejection sampling, we have flexibility on where to initialize the replicas. One could use the position of the reference chain at the end of the decorrelation step, or any other point in S.

The Decorrelation and Dephasing Times. T_{corr} and T_{phase} must be sufficiently large so that the distributions of both the reference process and the

replicas are as close as possible to the QSD, without exhausting computational resources. T_{phase} and T_{corr} play similar roles, and they both depend on the initial distribution of the processes in S.

Choosing good values of these parameters is nontrivial, as they determine the accuracy of the algorithm. In [3], the Fleming–Viot particle process together with convergence diagnostics are used to determine these parameters on the fly in each state. They can also be postulated from some a priori knowledge (e.g., barrier height between states), if available.

The Polling Time. The purpose of the polling time T_{poll} is to permit for periods of asynchronous computation of the replicas in a distributed computing environment. For the accelerated time to be correct, it is essential that all replicas have run for at least as long as replica K. Ensuring this requires resynchronization, which occurs at the polling time.

If communication amongst the replicas is cheap or there is little loss of synchronization per time step, one can take $T_{\text{poll}} = 1$. In this case, $M = \min\{n : \exists j \in \{1, \dots, N\} \text{ s.t. } X_n^j \notin S\}$ is the first exit time observed among the N replicas, $K = \min\{j : X_M^j \notin S\}$ (so $M = \tau^K$) and $T_{\text{acc}} = N(\tau^K - 1) + K$.

Efficiency of the Algorithm. For the algorithm to be efficient, the states must be truly metastable: within each state, the typical time to reach the QSD (T_{corr} and T_{phase}) should be small relative to the typical exit time. If most states are not metastable, then the exit times will be typically smaller than the decorrelation times, and the algorithm will rarely proceed to the dephasing and parallel steps.

The algorithm is consistent even if some or all the states are not metastable. Indeed, the states can be *any* collection of disjoint sets. However, if these sets are not reasonably defined, it will be difficult to obtain any gain in efficiency with ParRep. Defining the states requires some a priori knowledge about the system.

4 Mathematical Analysis of Discrete Time ParRep

The main result of this section, Proposition 4.5, shows that the coarse-grained trajectory simulated in ParRep is *exact* if the QSD has been exactly reached in the decorrelation and dephasing steps; see Equation (7).

4.1 Properties of QSDs

Before examining ParRep, we give a condition for existence and uniqueness of the QSD. We also state important properties of the exit law starting from the QSD. Many of these results can be found in [8, 9]. We assume the following, which is sufficient to ensure existence and uniqueness of the QSD.

Assumption 4.1. Let $S \in \mathcal{S}$ be any state.

(1) For any $x \in S$, $\mathbb{P}^x(X_1 \in S) > 0$.
(2) There exist $m \geq 1$ and $\delta \in (0,1)$, such that for all $x, y \in S$ and all bounded nonnegative measurable functions $f : S \to \mathbb{R}$, $\mathbb{E}^x[f(X_m)\, 1_{\{\tau > m\}}] \geq \delta \mathbb{E}^y[f(X_m) 1_{\{\tau > m\}}]$. □

With this condition, the following holds (see [10, Theorem 1]).

Theorem 4.2. Under Assumption 4.1, there exists a unique QSD ν in S. Furthermore, for any probability measure μ with support in S and any bounded measurable function $f : S \to \mathbb{R}$,

$$\left| \mathbb{E}^\mu[f(X_n) \mid \tau > n] - \int_S f(x)\, \nu(\mathrm{d}x) \right| \leq \|f\|_\infty 4\delta^{-1}(1-\delta^2)^{\lfloor n/m \rfloor}. \tag{6}$$

□

Theorem 4.2 shows that the law of $(X_n)_{n \geq 0}$, conditioned on not exiting S, converges in total variation norm to the QSD ν as $n \to \infty$. Thus, at the end of the decorrelation and dephasing steps, if T_{corr} and T_{phase} are sufficiently large, then the law of the reference process and replicas will be close to that of the QSD. Note that Theorem 4.2 provides an explicit error bound in total variation norm.

Next we state properties of the exit law starting from the QSD which are essential to our analysis. While these results are well known (see, for instance, [8, 9]), we give brief proofs for completeness.

Theorem 4.3. If $X_0 \sim \nu$, with ν the QSD in S, then τ and X_τ are independent, and τ is geometrically distributed with parameter $p = \mathbb{P}^\nu(X_1 \notin S)$. □

Proof. Let $k(x, \mathrm{d}y)$ denote the transition kernel of $(X_n)_{n\geq 0}$. We compute

$$\begin{aligned}\mathbb{E}^\nu[f(X_\tau) \mid \tau = n] &= \frac{\mathbb{E}^\nu[f(X_n)\, 1_{\{\tau=n\}}]}{\mathbb{E}^\nu[1_{\{\tau=n\}}]} = \frac{\mathbb{E}^\nu[1_{\{\tau>n-1\}} \int_{\Omega\setminus S} f(y)k(X_{n-1}, \mathrm{d}y)]}{\mathbb{E}^\nu[1_{\{\tau>n-1\}} \int_{\Omega\setminus S} k(X_{n-1}, \mathrm{d}y)]} \\ &= \frac{\mathbb{E}^\nu[\int_{\Omega\setminus S} f(y)k(X_{n-1}, \mathrm{d}y) \mid \tau > n-1]}{\mathbb{E}^\nu[\int_{\Omega\setminus S} k(X_{n-1}, \mathrm{d}y) \mid \tau > n-1]} \\ &= \frac{\int_S (\int_{\Omega\setminus S} f(y)k(x, \mathrm{d}y))\nu(\mathrm{d}x)}{\int_S (\int_{\Omega\setminus S} k(x, \mathrm{d}y))\nu(\mathrm{d}x)} \\ &= \mathbb{E}^\nu[f(X_\tau) \mid \tau = 1].\end{aligned}$$

The second to last equality is an application of (1). As $\mathbb{E}^\nu[f(X_\tau) \mid \tau = 1]$ is independent of n, this establishes independence of τ and X_τ.

Concerning the distribution of τ, we first calculate

$$\mathbb{P}^\nu(\tau > n) = \mathbb{P}^\nu(\tau > n | \tau > n-1)\mathbb{P}^\nu(\tau > n-1),$$

and then again use (1):

$$\begin{aligned}\mathbb{P}^\nu(\tau > n | \tau > n-1) &= \frac{\mathbb{E}^\nu[1_{\{\tau>n\}}]}{\mathbb{P}^\nu(\tau > n-1)} = \frac{\mathbb{E}^\nu[1_{\{\tau>n-1\}} \int_S k(X_{n-1}, \mathrm{d}y)]}{\mathbb{P}^\nu(\tau > n-1)} \\ &= \mathbb{E}^\nu\left[\int_S k(X_{n-1}, \mathrm{d}y) \mid \tau > n-1\right] \\ &= \int_S \left(\int_S k(x, \mathrm{d}y)\right) \nu(\mathrm{d}x) = \mathbb{P}^\nu(X_1 \in S).\end{aligned}$$

Thus, $\mathbb{P}(\tau^\nu > n) = \mathbb{P}(X_1^\nu \in S)\mathbb{P}(\tau^\nu > n-1)$ and by induction, $\mathbb{P}^\nu(\tau > n) = [\mathbb{P}^\nu(X_1 \in S)]^n = (1-p)^n$. ■

4.2 Analysis of the exit event

We can now state and prove our main result. We make the following idealizing assumption, which allows us to focus on the parallel step in Algorithm 3.1, neglecting the errors due to imperfect sampling of the QSD.

Idealization 4.4. Assume that

(A1) After spending T_{corr} consecutive time steps in S, the process $(X_n)_{n\geq 0}$ is *exactly* distributed according to the QSD ν in S. In particular, at the end of the decorrelation step, $X^{\mathrm{ref}}_{T_{\mathrm{sim}}} \sim \nu$.

(A2) At the end of the dephasing step, all N replicas are i.i.d. with law *exactly* given by ν. □

Idealization 4.4 is introduced in view of Theorem 4.2, which ensures that the QSD sampling error from the dephasing and decorrelation steps vanishes as T_{corr} and T_{phase} become large. Of course, for finite T_{corr} and T_{phase}, there is a nonzero error; this error will indeed propagate in time, but it can be controlled in terms of these two parameters. For a detailed analysis in the continuous time case, see [16, 27]. Although the arguments in [16, 27] could be adapted to our time discrete setting, we do not go in this direction; instead we focus on showing consistency of the parallel step.

Under Idealization 4.4, we show that ParRep is *exact*. That is, the trajectory generated by ParRep has the same probability law as the true coarse-grained chain:

$$(\hat{X}_n)_{n\geq 0} \sim (\Pi(X_n))_{n\geq 0}. \tag{7}$$

The evolution of the ParRep coarse-grained trajectory is exact in the decorrelation step. Together with Idealization 4.4, this means (7) holds if the parallel step is consistent (i.e., exact, if all replicas start at i.i.d. samples of the QSD). This is the content of the following proposition.

Proposition 4.5. Assume that the N replicas at the beginning of the parallel step are i.i.d. with law *exactly* given by the QSD ν in S (this is Idealization 4.4(A2)). Then the parallel step of Algorithm 3.1 is exact:

$$(X_{\text{acc}}, T_{\text{acc}}) \sim (X_\tau, \tau),$$

where $(X_{\text{acc}}, T_{\text{acc}})$ is defined as in Algorithm 3.1, while (X_τ, τ) is defined for $(X_n)_{n\geq 0}$ starting at $X_0 \sim \nu$. □

To prove Proposition 4.5, we need the following lemma.

Lemma 4.6. Let $\tau^1, \tau^2, \ldots, \tau^N$ be i.i.d. geometric random variables with parameter p: for $t \in \mathbb{N} \cup \{0\}$,

$$\mathbb{P}(\tau^j > t) = (1-p)^t.$$

Define

$$\begin{aligned} M &= \min\{m \geq 1 : \exists\, j \in \{1, \ldots, N\} \text{ s.t. } \tau^j \leq m T_{\text{poll}}\}, \\ K &= \min\{j \in \{1, \ldots, N\} : \tau^j \leq M T_{\text{poll}}\}, \\ \xi &= (N-1)(M-1)T_{\text{poll}} + (K-1)T_{\text{poll}} + \tau^K. \end{aligned}$$

Then ξ has the same law as τ^1. □

Proof. Note that ξ can be rewritten as

$$\xi = N(M-1)T_{\text{poll}} + (K-1)T_{\text{poll}} + [\tau^K - (M-1)T_{\text{poll}}].$$

Indeed, any natural number z can be uniquely expressed as $z = N(m-1)T_{\text{poll}} + (k-1)T_{\text{poll}} + t$ where $m \in \mathbb{N} \setminus \{0\}$, $k \in \{1, \ldots, N\}$, and $t \in \{1, 2, \ldots, T_{\text{poll}}\}$. For such m, k, and t, we compute

$$\begin{aligned}
&\mathbb{P}(\xi = N(m-1)T_{\text{poll}} + (k-1)T_{\text{poll}} + t) \\
&\quad = \mathbb{P}(M = m,\ K = k,\ \tau^K - (M-1)T_{\text{poll}} = t) \\
&\quad = \mathbb{P}(\tau^1 > mT_{\text{poll}}, \ldots, \tau^{k-1} > mT_{\text{poll}}, \tau^k = (m-1)T_{\text{poll}} \\
&\quad + t, \tau^{k+1} > (m-1)T_{\text{poll}}, \ldots, \tau^N > (m-1)T_{\text{poll}}) \\
&\quad = \mathbb{P}(\tau^1 > mT_{\text{poll}})^{k-1}\mathbb{P}(\tau^k = (m-1)T_{\text{poll}} + t)[\mathbb{P}(\tau^{k+1} > (m-1)T_{\text{poll}})]^{N-k} \\
&\quad = (1-p)^{(k-1)mT_{\text{poll}}}\, p(1-p)^{(m-1)T_{\text{poll}}+t-1}(1-p)^{(N-k)(m-1)T_{\text{poll}}} \\
&\quad = p(1-p)^{N(m-1)T_{\text{poll}}+(k-1)T_{\text{poll}}+t-1} = \mathbb{P}(\tau^1 = N(m-1)T_{\text{poll}} + (k-1)T_{\text{poll}} + t).
\end{aligned}$$

■

We can now proceed to the proof of Proposition 4.5.

Proof. In light of Theorem 4.3, it suffices to prove:

(i) T_{acc} is a geometric random variable with parameter $p = \mathbb{P}^\nu(X_1 \notin S)$,
(ii) X_{acc} and X_τ have the same law: $X_{\text{acc}} \sim X_\tau$, and
(iii) T_{acc} is independent of X_{acc},

where $(X_n)_{n \geq 0}$ is the process starting at the $X_0 \sim \nu$.

We first prove (i). For $j \in \{1, 2, \ldots, N\}$, let τ^j be a random variable representing the first exit time from S of the jth replica in the parallel step of ParRep, if the replica were allowed to keep evolving indefinitely. By (A2), $\tau^1, \ldots, \tau^N$ are independent and all have the same distribution as τ. Now by Theorem 4.3, $\tau^1, \ldots, \tau^N$ are i.i.d. geometric random variables with parameter p, so by Lemma 4.6, T_{acc} is also a geometric random variable with parameter p.

Now we turn to (ii) and (iii). Note that $K = k$ if and only if $X_{\text{acc}} = X^k_{\tau^k}$ and there exists $m \in \mathbb{N}$ such that $\tau^1 > mT_{\text{poll}}, \ldots, \tau^{k-1} > mT_{\text{poll}}$, $(m-1)T_{\text{poll}} < \tau^k \leq mT_{\text{poll}}$, and $\tau^{k+1} > (m-1)T_{\text{poll}}, \ldots, \tau^N > (m-1)T_{\text{poll}}$. From Theorem 4.3 and (A2), $X^k_{\tau^k}$ is independent of $\tau^1, \ldots, \tau^N$, so X_{acc} must be independent of K. From this and (A2), it follows that $X_{\text{acc}} \sim X_\tau$. To see that X_{acc} is independent of T_{acc}, let $\sigma(K, \tau^K)$ be the sigma algebra generated by K and τ^K. Knowing the value of K and τ^K is enough to deduce the value of T_{acc};

that is, T_{acc} is $\sigma(K, \tau^K)$-measurable. Also, by the preceding analysis and Theorem 4.3, $X_{\text{acc}} = X^K_{\tau^K}$ is independent of $\sigma(K, \tau^K)$. To conclude that T_{acc} and X_{acc} are independent, we compute for suitable test functions f and g:

$$\begin{aligned}\mathbb{E}[f(T_{\text{acc}})g(X_{\text{acc}})] &= \mathbb{E}[\mathbb{E}[f(T_{\text{acc}})g(X_{\text{acc}}) \mid \sigma(K, \tau^K)]] \\ &= \mathbb{E}[f(T_{\text{acc}})\mathbb{E}[g(X_{\text{acc}}) \mid \sigma(K, \tau^K)]] = \mathbb{E}[f(T_{\text{acc}})]\,\mathbb{E}[g(X_{\text{acc}})].\end{aligned}$$

■

5 Numerical Examples

In this section, we consider two examples. The first illustrates numerically the fact that the parallel step in Algorithm 3.1 is consistent. The second shows typical errors resulting from a naive application of the original ParRep algorithm to a time discretization of Langevin dynamics. These are simple illustrative numerical examples. For a more advanced application, we refer the reader to the paper [3], where our Algorithm 3.1 was used to study the 2D Lennard–Jones cluster of seven atoms.

5.1 One-dimensional random walk

Consider a random walk on $\mathbb{Z}$ with transition probabilities $p(i, j)$ defined as follows:

$$p(i,j) = \begin{cases} \frac{3}{4}, & i < 0 \text{ and } j = i+1, \\ \frac{1}{4}, & i < 0 \text{ and } j = i-1, \\ \frac{1}{3}, & i = 0 \text{ and } |j| \le 1, \\ \frac{1}{4}, & i > 0 \text{ and } j = i+1, \\ \frac{3}{4}, & i > 0 \text{ and } j = i-1, \\ 0, & \text{otherwise.} \end{cases}$$

We use ParRep to simulate the first exit time τ of the random walk from $S = [-5, 5]$, starting from the QSD ν in S. At each point except 0, steps toward 0 are more likely than steps toward the boundaries -5 or 5.

We perform this simulation by using the dephasing and parallel steps of Algorithm 3.1; for sufficiently large T_{phase}, the accelerated time T_{acc} should have the same law as τ. In this simple example, we can analytically compute the distribution of τ. We perform 10^6 independent ParRep simulations to obtain statistics on the distribution of T_{acc} and the gain in "wall clock time", defined below. We find that T_{acc} and τ have very close probability mass functions when $T_{\text{phase}} = 25$; see Figure 3. To measure

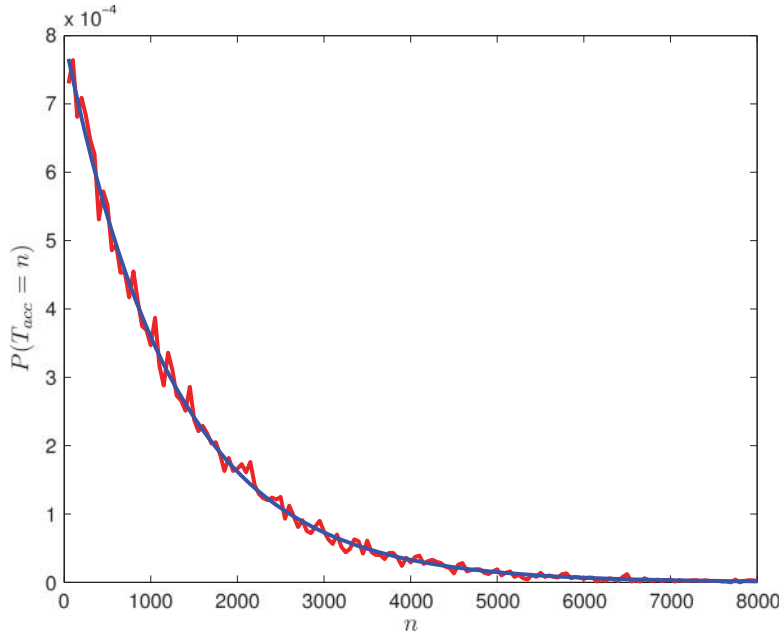

Fig. 3. Probability mass function of T_{acc}, estimated by 10^6 ParRep simulations with $N = 10$ replicas and $T_{\text{phase}} = T_{\text{corr}} = 25$, versus exact distribution of τ (smooth curve).

the gain in wall clock efficiency using ParRep, we introduce the parallel time T_{par}—defined, using the notation of Algorithm 3.1, by $T_{\text{par}} = MT_{\text{poll}}$, where we recall M is such that $\tau^K \in [(M-1)T_{\text{poll}} + 1, MT_{\text{poll}}]$. Thus, the wall clock time of the parallel step is $C_0 T_{\text{par}}$, with C_0 the computational cost of a single time step of the Markov chain for one replica. Note in Figure 4 the significant parallel time speedup in ParRep compared with the direct sampling time. The speedup is approximately linear in N.

5.2 Discretized diffusions

Consider the overdamped Langevin stochastic process in $\mathbb{R}^d$,

$$\mathrm{d}\tilde{X}_t = -\nabla V(\tilde{X}_t)\,\mathrm{d}t + \sqrt{2\beta^{-1}}\,\mathrm{d}W_t. \tag{8}$$

The associated Euler–Maruyama discretization is

$$X_{n+1} = X_n - \nabla V(X_n)\Delta t + \sqrt{2\beta^{-1}\Delta t}\xi_n, \tag{9}$$

where $\xi_n \sim N(0, I)$ are d-dimensional i.i.d. random variables. It is well known [14] that $(X_n)_{n\geq 0}$ is then an approximation of $(\tilde{X}_{n\Delta t})_{n\geq 0}$.

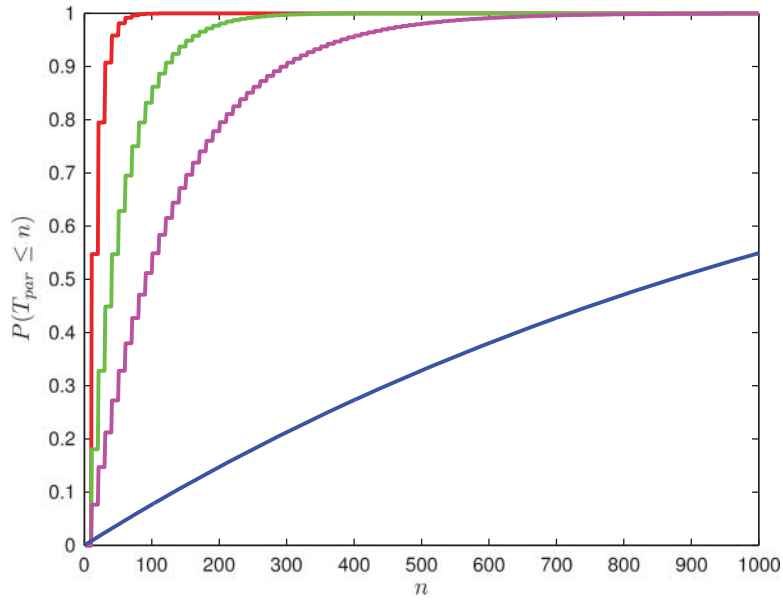

Fig. 4. Cumulative distribution function of parallel time required for ParRep sampling with $T_{\text{poll}} = 10$ and, from top: $N = 100, 25, 10$. The bottom curve is the (analytic) cumulative distribution function of τ (corresponding to $N = 1$).

5.2.1 *Existence and uniqueness of the QSD*

We first show that the conditions in Assumption 4.1 hold (see [10] for a similar example in 1D).

Proposition 5.1. Assume that $S \subset \mathbb{R}^d$ is bounded and ∇V is bounded on S. Then (9) satisfies Assumption 4.1. □

Proof. First, for any $x \in S$,

$$\mathbb{P}^x(X_1 \in S) = \mathbb{E}^x[1_S(X_1)] = (4\pi\beta^{-1}\Delta t)^{-d/2} \int_{\mathbb{R}^d} 1_S(y) \exp\left\{-\frac{|y - x + \nabla V(x)\Delta t|^2}{4\beta^{-1}\Delta t}\right\} \mathrm{d}y$$
$$\geq |S|(4\pi\beta^{-1}\Delta t)^{-d/2} \min_{y \in S}\left\{\exp\left\{-\frac{|y - x + \nabla V(x)\Delta t|^2}{4\beta^{-1}\Delta t}\right\}\right\} > 0. \tag{10}$$

Next, for any $x, y \in S$,

$$\begin{aligned}
\mathbb{E}^x[f(X_1)1_{\{\tau>1\}}] &= (4\pi\beta^{-1}\Delta t)^{-d/2}\int_S f(z)\exp\left\{-\frac{|z-x+\nabla V(x)\Delta t|^2}{4\beta^{-1}\Delta t}\right\}\mathrm{d}z \\
&= (4\pi\beta^{-1}\Delta t)^{-d/2}\int_S f(z)\exp\left\{-\frac{|z-y+\nabla V(y)\Delta t|^2}{4\beta^{-1}\Delta t}\right\} \\
&\quad\times\exp\left\{-\frac{|z-x+\nabla V(x)\Delta t|^2-|z-y+\nabla V(y)\Delta t|^2}{4\beta^{-1}\Delta t}\right\}\mathrm{d}z \\
&\geq C(4\pi\beta^{-1}\Delta t)^{-d/2}\int_S f(z)\exp\left\{-\frac{|z-y+\nabla V(y)\Delta t|^2}{4\beta^{-1}\Delta t}\right\}\mathrm{d}z \\
&= C(4\pi\beta^{-1}\Delta t)^{-d/2}\mathbb{E}^y[f(X_1)1_{\{\tau>1\}}], \qquad (11)
\end{aligned}$$

where

$$C = \min_{x,y,z\in S}\exp\left\{-\frac{|z-x+\nabla V(x)\Delta t|^2-|z-y+\nabla V(y)\Delta t|^2}{4\beta^{-1}\Delta t}\right\}.$$

Since S is bounded and terms in the brackets are bounded, $C>0$. In Assumption 4.1, we can then take $m=1$ and $\delta = C(4\pi\beta^{-1}\Delta t)^{-d/2}$. ■

Theorem 4.2 ensures that $(X_n)_{n\geq 0}$ converges to a unique QSD in S, with a precise error estimate in terms of the parameters m and δ obtained in the proof of Proposition 5.1. This error estimate is certainly not sharp; better estimates can be obtained by studying the spectral properties of the Markov kernel. We refer the reader to [16] for such convergence results in the continuous time case (8).

5.2.2 *Numerical example*

Here we consider the 1D process

$$\mathrm{d}\tilde{X}_t = -2\pi\sin(\pi\tilde{X}_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t, \qquad (12)$$

discretized with $\Delta t = 10^{-2}$. We compute the first exit time from $S=(-1,1)$, starting at $\tilde{X}_0 = \frac{1}{2}$. We use Algorithm 3.1 with $T_{\mathrm{corr}} = T_{\mathrm{phase}} = 100$, corresponding to the physical time scale $T_{\mathrm{corr}}\Delta t = T_{\mathrm{phase}}\Delta t = 1$, and $N = 1,000$ replicas.

Consider a direct implementation of the continuous time ParRep algorithm into the time discretized process. In that algorithm, the accelerated time is (in units of physical time instead of time steps)

$$T_{\mathrm{acc}}^{\mathrm{continuous}} = N\tau^K\Delta t, \qquad (13)$$

with τ^K the same as in Algorithm 3.1. As $T_{\mathrm{acc}}^{\mathrm{continuous}}$ is by construction a multiple of $N\Delta t = 10$, a staircasing effect can be seen in the exit time distribution; see Figure 5.

This staggering worsens as the number of replicas increases. In our Algorithm 3.1, we use the accelerated time formula (again in units of physical time)

$$T_{\text{acc}}^{\text{corrected}} = T_{\text{acc}} \Delta t.$$

We find excellent agreement between the serial data—that is, the data obtained from direct numerical simulation—and the data obtained from Algorithm 3.1; See Figure 5. (The agreement is perfect in the decorrelation step; see Figure 6.) We comment further on this in the next section.

5.2.3 *Discussion*

In light of the discretization example, one may ask what kind of errors were introduced in previous numerical studies which used ParRep with (13). Taking $T_{\text{poll}} = 1$ for simplicity, we calculate

$$\begin{aligned}\mathbb{E}[|T_{\text{acc}}^{\text{corrected}} - T_{\text{acc}}^{\text{continuous}}|] &= \mathbb{E}[|(N(\tau^K - 1) + K)\Delta t - N\tau^K \Delta t|] \\ &= \Delta t\, \mathbb{E}[|N - K|] = \Delta t \sum_{k=1}^{N} (N-k)\mathbb{P}(K=k).\end{aligned}$$

Using calculations analogous to those used to study T_{acc}, it can be shown that

$$\mathbb{P}(K=k) = \frac{(1-p)^{k-1}p}{1-(1-p)^N}.$$

Therefore, the error in the number of time steps per parallel step is

$$\text{Absolute Error} = \frac{N\Delta t}{1-(1-p)^N} - \frac{\Delta t}{p}, \quad \text{Relative Error} = \frac{pN}{1-(1-p)^N} - 1. \tag{14}$$

Consider the relative error, writing it as

$$pN\left[\frac{1}{1-r^N} - \frac{1}{(1-r)N}\right] \quad \text{where } r = 1-p.$$

We claim the quantity in the brackets,

$$f(r,N) := \frac{1}{1-r^N} - \frac{1}{(1-r)N} = \frac{r^N - Nr + N - 1}{Nr^{N+1} - Nr^N - Nr + N} \tag{15}$$

is bounded from above by 1. Indeed, for any $0 < r < 1$, we immediately see that $f(r,N)$ is 0 at $N = 1$ and 1 as $N \to \infty$. Let us reason by contradiction and assume that $\sup_{r\in(0,1),N>0} f(r,N) > 1$. Since f is continuous in $N > 0$ and $0 < r < 1$, there is then a point

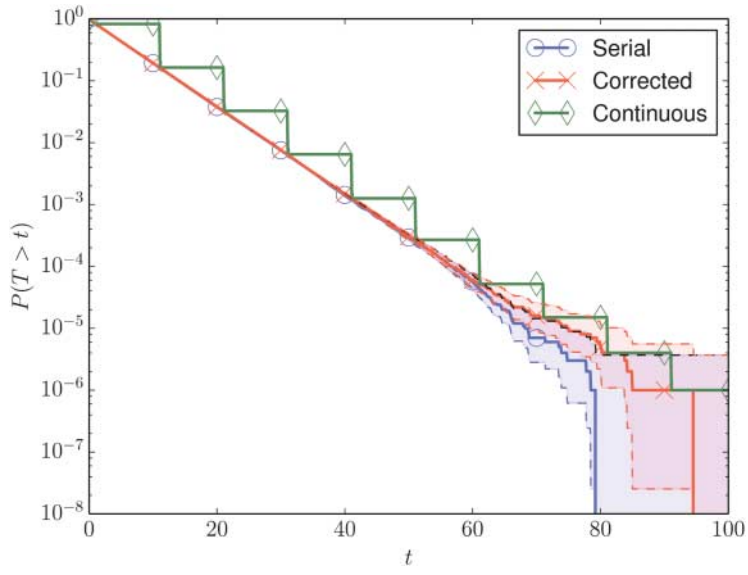

Fig. 5. Exit time distributions for the Euler–Maruyama discretization of (12). Here T represents the first exit time from $S = (-1, 1)$, starting at $\frac{1}{2}$. There is excellent agreement between the serial, unaccelerated simulation data ($T = \tau^{\nu} \Delta t$), and our ParRep algorithm ($T = T_{\text{acc}}^{\text{corrected}}$), while the original ParRep formula ($T = T_{\text{acc}}^{\text{continuous}}$) deviates significantly. Dotted lines represent 95% Clopper–Pearson confidence intervals obtained from 10^6 independent simulations; confidence interval widths increase in t as fewer samples are available.

(r, N) such that $f(r, N) = 1$; thus

$$g_N(r) = 0 \quad \text{where } g_N(r) := Nr^{N+1} - (N+1)r^N + 1.$$

Note that $g_N(0) = 1$ and $g_N(1) = 0$ for all values of N. Computing the derivative with respect to r, we observe

$$g_N'(r) = -N(N+1)(1-r)r^{N-1} < 0.$$

Therefore, $g_N(r)$ is decreasing, from 1 at $r = 0$ to 0 at $r = 1$, in the interval $(0, 1)$. Hence, $g_N(r) = 0$ has no solution, contradiction. We conclude that (15) is bounded from above by 1.

Consequently, we are assured

$$\text{Absolute Error} \leq N\Delta t, \quad \text{Relative Error} \leq pN. \tag{16}$$

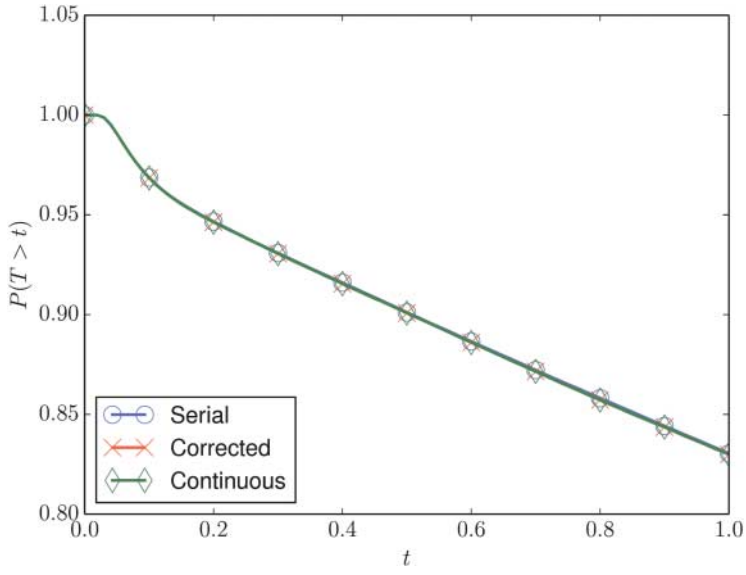

Fig. 6. A zoomed-in version of Figure 5, highlighting the decorrelation step (recall $T_{\text{corr}}\Delta t = 1$). Serial simulation, our ParRep algorithm, and the original ParRep algorithm all produce identical data. This comes from the fact that serial and ParRep simulations are identical in law during the decorrelation step.

Thus, so long as $pN \ll 1$, the relative error using the accelerated time $T_{\text{acc}}^{\text{continuous}}$ will be modest, especially for very metastable states where $p \ll 1$. If also $N\Delta t \ll 1$, then the absolute error will be small.

The above calculations are generic. Although our discretized diffusion example in Section 5.2.2 is a simple 1D problem, the errors displayed in Figure 5 are expected whenever the continuous time ParRep rule (13) is used for a time discretized process. Although this error (as we showed above) will be small provided $Np \ll 1$ and $N\Delta t \ll 1$, our Algorithm 3.1 has the advantage of being consistent for any Δt, including relatively large values of $N\Delta t$.

Acknowledgements

We would like to thank the anonymous referees for their many constructive remarks.

Funding

The work of D. Aristoff and G. Simpson was supported in part by DOE Award DE-SC0002085. G. Simpson was also supported by the NST PIRE Grant OISE-0967140. The work of T. Lelièvre

is supported by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement number 614492.

References

[1] Baker, K. L. and D. H. Warner. "Extended timescale atomistic modeling of crack tip behavior in aluminum." *Modelling and Simulation in Materials Science And Engineering* 20, no. 6 (2012): 065005.

[2] Bieniek, M., K. Burdzy, and S. Finch. "Non-extinction of a Fleming–Viot particle model." *Probability Theory and Related Fields* 153, nos 1–2 (2012): 293–332.

[3] Binder, A. J., T. Lelièvre, and G. Simpson. "A generalized Parallel Replica dynamics." (2014): preprint arXiv:1404.6191. http://arxiv.org/abs/1404.6191.

[4] Bouchard, B., S. Geiss, and E. Gobet. "First time to exit of a continuous Itô process: general moment estimates and L_1-convergence rate for discrete time approximations." (2013): preprint http://hal.archives-ouvertes.fr/hal-00844887.

[5] Bovier, A., M. Eckhoff, V. Gayrard, and M. Klein. "Metastability and low lying spectra in reversible Markov chains." *Communications in Mathematical Physics* 228, no. 2 (2002): 219–55.

[6] Cattiaux, P., P. Collet, A. Lambert, S. Martínez, S. Méléard, and J. San Martín. "Quasi-stationary distributions and diffusion models in population dynamics." *The Annals of Probability* 37, no. 5 (2009): 1926–69.

[7] Chodera, J. D., N. Singhal, V. S. Pande, K. A. Dill, and W. C. Swope. "Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics." *The Journal of Chemical Physics* 126, no. 15 (2007): 155101.

[8] Collet, P., S. Martinez, and J. San Martin. *Quasi-Stationary Distributions*. Berlin: Springer, 2013.

[9] Del Moral, P. *Feynman–Kac Formulae*. Berlin: Springer, 2004.

[10] Del Moral, P. and A. Doucet. "Particle motions in absorbing medium with hard and soft obstacles." *Stochastic Analysis and Applications* 22, no. 5 (2004): 1175–207.

[11] Ferrari, P. A. and N. Maric. "Quasi stationary distributions and Fleming–Viot processes in countable spaces." *Electronic Journal of Probability* 12, no. 24 (2007): 684–702.

[12] Grigorescu, I. and M. Kang. "Hydrodynamic limit for a Fleming–Viot type system." *Stochastic Processes and their Applications* 110, no. 1 (2004): 111–43.

[13] Joshi, K. L., S. Raman, and A. C. T. van Duin. "Connectivity-based parallel replica dynamics for chemically reactive systems: from femtoseconds to microseconds." *The Journal of Physical Chemistry Letters* 4, no. 21 (2013): 3792–7.

[14] Kloeden, P. E. and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Berlin: Springer, 1992.

[15] Komanduri, R., N. Chandrasekaran, and L. Raff. "Molecular dynamics simulation of atomic-scale friction." *Physical Review B* 61, no. 20 (2000): 14007–19.

[16] Le Bris, C., T. Lelièvre, M. Luskin, and D. Perez. "A mathematical formalization of the parallel replica dynamics." *Monte Carlo Methods and Applications*, 18, no. 2 (2012): 119–46.

[17] Lelièvre, T., G. Stoltz, and M. Rousset. *Free Energy Computations: A Mathematical Perspective*. London: Imperial College Press, 2010.

[18] Lu, C.-Y., A. F. Voter, and D. Perez. "Extending atomistic simulation timescale in solid/liquid systems: crystal growth from solution by a parallel-replica dynamics and continuum hybrid method." *The Journal of Chemical Physics* 140, no. 4 (2014): 044116.

[19] Mariusz, B., B. Krzysztof, and P. Soumik. "Extinction of Fleming–Viot-type particle systems with strong drift." *Electronic Journal of Probability* 17, no. 11 (2012): 1–15.

[20] Martínez, S. and J. San Martín. "Quasi-stationary distributions for a Brownian motion with drift and associated limit laws." *Journal of Applied Probability* 31, no. 4 (1994): 911–20.

[21] Méléard, S. and D. Villemonais. "Quasi-stationary distributions and population processes." *Probability Surveys* 9 (2012): 340–410.

[22] Perez, D., Y. Dong, A. Martini, and A. F. Voter. "Rate theory description of atomic stick-slip friction." *Physical Review B* 81, no. 24 (2010): 245415.

[23] Perez, D., S.-N. Luo, A. F. Voter, and T. C. Germann. "Entropic stabilization of nanoscale voids in materials under tension." *Physical Review Letters* 110, no. 20 (2013): 206001.

[24] Prinz, J.-H., H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé. "Markov models of molecular kinetics: generation and validation." *The Journal of Chemical Physics* 134, no. 17 (2011): 174105.

[25] Schütte, C. and M. Sarich. *Metastability and Markov State Models in Molecular Dynamics*. Providence, RI: American Mathematical Society, 2013.

[26] Scoppola, E. "Metastability for Markov chains: a general procedure based on renormalization group ideas." In *Probability and Phase Transition*, edited by G. Grimmett, 303–22. NATO ASI Series 420. Netherlands: Springer, 1994.

[27] Simpson, G. and M. Luskin. "Numerical analysis of parallel replica dynamics." *M2AN* 47, no. 5 (2013): 1287–314.

[28] Uberuaga, B. P., S. J. Stuart, and A. F. Voter. "Parallel replica dynamics for driven systems: derivation and application to strained nanotubes." *Physical Review B* 75, no. 1 (2007): 014301.

[29] Voter, A. F. "Parallel replica method for dynamics of infrequent events." *Physical Review B* 57, no. 22 (1998): 13985–8.

[30] Voter, A. F., F. Montalenti, and T. C. Germann. "Extending the time scale in atomistic simulation of materials." *Annual Review of Materials Science* 32 (2002): 321–46.