Sampling rare events in stochastic reaction-diffusion systems within trajectory looping

Pawel J. Zuk

Department of Biosystems and Soft Matter, Institute of Fundamental Technological Research,
Polish Academy of Sciences, 02-106 Warsaw, Poland and
Department of Mechanical and Aerospace Engineering,
Princeton University, Princeton, New Jersey 08544, USA

Marek Kochariczyk and Tomasz Lipniacki∗

Department of Biosystems and Soft Matter, Institute of Fundamental Technological Research,
Polish Academy of Sciences, 02-106 Warsaw, Poland

In bistable reaction–diffusion systems, transitions between stable states typically occur on timescales orders of magnitude longer than the chemical equilibration time. Estimation of transition rates within explicit Brownian dynamics simulations is computationally prohibitively costly. We present a method that exploits a single trajectory, generated by a prior simulation of diffusive motions of molecules, to sample chemical kinetic processes on timescales several orders of magnitude longer than the duration of the diffusive trajectory. In this approach, we “loop” the diffusive trajectory by transferring chemical states of the molecules from the last to the first time step of the trajectory. Trajectory looping can be applied to enhance sampling of rare events in biochemical systems in which the number of reacting molecules is constant, as in cellular signal transduction pathways. As an example, we consider a bistable system of autophosphorylating kinases, for which we calculate state-to-state transition rates and traveling wave velocities. We provide an open-source implementation of the method.

This is an author-created version of an article published in Physical Review E 98, 022401 (2018).

I. INTRODUCTION

Processes involving constituents that modify each other upon contact and diffuse in space arise in a multitude of fields as diverse as, e.g., geomorphology, population ecology, developmental biology and, most naturally, (bio)chemistry. Such reaction–diffusion systems are often expressed and analyzed in terms of partial differential equations (PDEs) [1]. When molecular noise has a significant impact on the process of interest, a deterministic description in terms of PDEs becomes inappropriate and more fine-grained descriptions are necessary, such as stochastic simulations with a single-molecule resolution. However, molecular simulations of spatially extended systems are computationally demanding. When, in the chemical context, individual reacting molecules perform Brownian random walk in continuous space, the most compute-intensive part is finding and resolving their collisions. If hydrodynamic interactions are additionally taken into account, simulation of diffusive dynamics is particularly expensive [2]. Since typically molecules perform multiple random diffusive steps between collisions and only a fraction of collisions result in chemical reactions, the numerical cost of spatial stochastic simulations can be prohibitively high.

Computational analysis of oscillatory or multistable biochemical systems requires long simulation times to properly sample the phase space or collect multiple state-to-state transitions. Simulations of multistable reaction systems are demanding even in the well-mixed limit, when the efficient Gillespie algorithm [3] can be applied. To enhance sampling, reweighting-based [4] and importance sampling-based [5] approaches were developed in addition to metamethods, such as transition path sampling [6] or forward flux sampling [7], in which the Gillespie algorithm simulations can be framed. Such auxiliary rare event sampling schemes often turn out to be inevitable for the estimation of state-to-state transition rates and, in this way, for determining the relative stability of the steady states. Several approaches were devised to speed up spatial stochastic simulations by splitting the reaction and diffusion events, and treating the latter implicitly. For example, in the next subvolume (NSV) method [8], the reactor is decomposed into subreactors that are assumed to be well mixed whereas diffusion is included in the form of random transport of molecules between neighboring subreactors. Another approach that circumvents explicit simulation of Brownian motion, yet works at the single-molecule level, is the event-driven Green’s function reaction dynamics (enhanced GFRD, eGFRD) [9, 10]. In this approach, the spatial propagator displaces molecules directly to the locations where they react. One may accommodate chemical reactions within the direct simulation Monte Carlo method (DSMC), originally developed to simulate rarefied gases [11]. In DSMC it is assumed that each simulated molecule may represent a large number of actual molecules and that molecular motions (treated as ballistic, modeled deterministically) and intermolecular collisions (modeled stochastically) are uncoupled over small time intervals. The framework can be extended to account for disparate diffusivities of reacting molecules to study, e.g., formation of Turing patterns [12].

* tlipnia@ippt.pan.pl
Here we present “trajectory looping,” a method that speeds up spatial stochastic simulations while still utilizing explicit simulations of Brownian motion of individual molecules. In the method we generate a trajectory of diffusive motions of the molecules and use it multiple times for sampling chemical processes on time scales that are several orders of magnitude longer than the simulated time of the original trajectory. The generated trajectory should be long enough to allow each molecule to form contacts with molecules other than its initial neighbors.

The proposed method is suitable for simulation of biochemical reactions in which substrate molecules are reversibly modified by enzymes, as often found in cellular signal transduction systems [13, 14]. Crucially for the method, we assume that in the enzymatic reaction scheme: enzyme + substrate → enzyme–substrate complex → enzyme + product, the enzyme–substrate complexes are short-lived and thus the substrate can be modified instantaneously when in contact with the enzyme, according to the simplified reaction scheme: enzyme + substrate → enzyme + product. In such systems, the number of reacting molecules is constant.

We explain the algorithm, explore limitations, and test the accuracy of this approach applied to two systems of coupled biochemical reactions: a monostable phosphorylation–dephosphorylation cycle and a bistable system of coupled biochemical reactions: a monostable and a bistable phosphorylation–dephosphorylation cycle and a bistable system of coupled biochemical reactions: a monostable and a bistable system of coupled biochemical reactions.

II. RESULTS

A. Algorithm

A generic simulation workflow in which trajectory looping is embedded is presented in Fig. 1.

a. Base mechanical trajectory. Initially, one has to obtain a trajectory of \( N \) molecules represented by identical spheres of radius \( a \), recorded at \( M \) equidistant time points separated by \( \Delta t \). This trajectory, referred to as the base mechanical trajectory, should be obtained according to a scheme that generates molecule positions in a Markovian manner (that is, in a manner that is devoid of inertial effects). It is required that in each simulation step the positions of molecules are accordant with the equilibrium spatial distribution, guaranteeing that the neighborhoods of all molecules are statistically indistinguishable. For reasons that are explained further, the simulated system is expected to be translationally invariant which is achieved by imposing periodic boundary conditions in at least one spatial dimension.

The molecule-scale Brownian time \( \tau_B \) that sets a natural time scale for the base mechanical trajectory is defined as the expected time required for a molecule to diffuse over its diameter \( \tau_B = (2a)^2/D_0 \), where \( D_0 \) is the diffusion coefficient of a molecule (in the absence of other molecules). All base mechanical trajectories subjected to looping and analyzed in this article were generated by performing standard Brownian dynamics simulations at a given volume fraction \( \phi \), as described in the Methods section.

b. Base contacts sequence. The only information contained in the base mechanical trajectory that is relevant to chemistry is that about molecule contacts. Two molecules are in contact when their center-to-center distance is smaller than the sum of their reaction radii. If reaction radii are identical for all molecules, being all equal \( a + \delta/2 \), then a significant simplification may be introduced: for each time step of the base mechanical trajectory a list of molecules of pairwise center-to-center distances smaller than the sum of reaction radii, \( d_{ij} < 2a + \delta \), can be derived. Such a list of pairs of molecule indices, termed the base contacts sequence, is used as an input to the stochastic simulation of chemical events instead of the base mechanical trajectory.

c. Assignments. The base mechanical trajectory of duration \( \theta = (M-1) \times \Delta t \) is used to create a mechanical trajectory of length \( \Theta > \theta \) by means of a (recursively applied) procedure that consists in joining the end of the base mechanical trajectory with its beginning and a subsequent molecule indices reassignment. To perform a join, one has to find an assignment in the form of a complete molecule-to-molecule index map that is used to transfer molecule identities (associated with chemical states) between molecules in the last time step and molecules in the first time step of the base mechanical trajectory. An optimal assignment that minimizes mean square displacements in pairs of corresponding molecules can be found in polynomial time, \( O(N^3) \), according to the Kuhn–Munkres algorithm [15, 16] (also known as the Hungarian algorithm; in this article, we use an implementation from DLIB [17]). When performing consecutive joins, to prevent finding identical molecule-to-molecule
assignments, one can make use of the translational invariance of the simulated system and, prior to finding pairs of corresponding molecules, apply a common random shift to all molecule coordinates along the direction(s) in which the system has periodic boundary conditions. For an exemplary assignment see Fig. 2.

d. Looped contacts sequence. Consecutive assignments are applied recursively to the base contacts sequence, yielding the looped contacts sequence. The looped contacts sequence contains replicas of the base contacts sequence with molecules re-indexed at each join. In this way, computationally expensive search of molecule contacts is performed once (for the base mechanical trajectory).

e. Chemical initial condition. Chemical states are assigned to every molecule in the first time step of the simulation; this assignment is called the chemical initial condition. It is assumed that all molecules have the same diffusivity and that their chemical state have no influence on their diffusion properties.

f. Looped chemical trajectory. In the course of the simulation, chemical states of molecules can be altered as a result of chemical reactions that are either unimolecular (i.e., first order) or bimolecular (i.e., second order). A unimolecular reaction may occur independently of the molecule neighborhood, whereas a bimolecular reaction may occur only when two molecules are in contact, i.e., when their center-to-center distance is smaller than the reaction distance, \( d_{ij} < 2a + \delta \). No reaction is allowed to change the number of molecules in the system, \( N \). In every time step within the looped contacts sequence, a complete list of possible reaction events is created. Time to a next reaction is drawn at random from exponential distribution with the expected value of the inverse of the sum of chemical rates of all events; reaction events are selected randomly from the complete list with probabilities that are proportional to respective rates. After a substrate is modified, new reaction events that are possible between the modified substrate and other molecules within reaction distance are added to the list. Stochastic simulation of reaction kinetics in a given time step of the looped contacts sequence proceeds as long as times to a next event lie within \( \Delta t \).

The described event scheduling scheme constitutes an exact stochastic simulation method that for a sufficiently fast diffusion is consistent with the Gillespie algorithm [3] in which possible reactants are limited to molecules within the reaction distance. Since configurations of molecules in the base mechanical trajectory are in equilibrium, one can use the radial distribution function \( g(\sigma|\phi) \), determined from the analytical solution of the Percus–Yevick integral equation for hard spheres [18], to count the expected number of molecules within the reaction distance:

\[
\chi = \frac{N}{V} \times \int_{2a}^{2a+\delta} \left[ g(\sigma|\phi) \times 4\pi \sigma^{2} \right] \, d\sigma.
\]

To compare kinetics of our simulations with those obtained in the Gillespie simulations, bimolecular reaction rate constants for the latter method are multiplied by \( \chi/N \).

After \( K \) joins, a simulated record of chemical states of \( N \) molecules in \((K+1) \times M \) time steps will be referred to as the looped chemical trajectory. Trajectory looping performed multiple times enables recurrent trajectory reuse and simulation of chemical processes on timescales several orders of magnitude longer than the original diffusion trajectory.

Simulations of stochastic chemical kinetics with trajectory looping can be performed in our open-source software tool, LOOPER [19].
B. Limitations

Looped trajectories are distorted due to (i) displacements of molecules in end-to-begin joins and (ii) limited repertoire of single-molecule trajectories contained in the base mechanical trajectory.

a. Displacements of molecules in end-to-begin joins.

In every end-to-begin join, each source molecule in the last time step of the base mechanical trajectory is assigned uniquely to a target molecule in the initial time step. A subsequent transfer of the "identity" of a source molecule to a target molecule causes an instantaneous molecule movement from the source to the target location [see kinks that appear every \( n \theta = n 10 \tau_B \) in Fig. 3(a)]. The mean-square displacement (MSD) that results from such a join is 

\[
\langle r^2 \rangle_{\phi}^{\text{join}} = \sum_{k=1}^{K} \left( \frac{1}{K} \sum_{i} | \tilde{R}_\sigma(k,i)(0) - \tilde{R}_\sigma(k-1,i)(\theta) |^2 \right) / N, 
\]

where \( \sigma(k,i) \) denotes an index of an \( i \)th molecule in the base mechanical trajectory after \( k \) end-to-begin joins (initially \( \sigma(0,i) := i \)). Although the Hungarian algorithm is applied to minimize \( \langle r^2 \rangle_{\phi}^{\text{join}(k)} \), these displacements are still usually larger than the Brownian molecule displacements in \( \Delta t \). The impact of distortions introduced by end-to-begin joins on the looped mechanical trajectory can be made relatively small when \( \langle r^2 \rangle_{\phi}^{\text{join}} \ll \langle r^2(\tau = \theta) \rangle_{\phi} \), where \( \langle r^2 \rangle_{\phi}^{\text{join}} = \sum_{k=1}^{K} \langle r^2 \rangle_{\phi}^{\text{join}(k)} / K \), that is, for a sufficiently long base mechanical trajectory.

Molecule displacements in an end-to-begin join can significantly alter molecular neighborhoods. New molecule–molecule contacts can be formed at the expense of breaking existing contacts, which affects the looped contacts trajectory. To quantify this effect, we calculated a distribution of new encounter times in a trajectory of \( \theta = 20 \tau_B \) (without looping) and then a distribution of new encounter times after a first half of this trajectory was looped once [see Fig. 3(b)]. Encounters that are an artifact of looping are registered as an excess of the new encounters that peaks just after \( 10 \tau_B \). At the assumed \( \theta \), the excess of the new encounters appears to be lower than 10% for volume fractions \( \phi \) between 5% and 40% [see Fig. 3(c)].

b. Limited repertoire of single-molecule trajectories.

If the optimal assignments were used to loop the base mechanical trajectory with explicit tracking of absolute positions of molecules, then one would obtain a looped mechanical trajectory. The base mechanical trajectory is just a single stochastic realization of diffusion of a finite set of molecules. A property that is specific to each base mechanical trajectory, and affects the apparent long-time MSD of molecules in the looped mechanical trajectory, is the displacement of the base mechan-
ical trajectory center of the mass (COM) \( \vec{\delta}_{\text{COM}}(\theta) = \sum_{N} [\tilde{R}(\theta) - \tilde{R}_0]/N \). Generically, \( \vec{\delta}_{\text{COM}} \) is a nonzero vector that due to looping adds to diffusive displacements of all molecules a contribution of direction and magnitude consistent among all end-to-begin trajectory joins. The expected value of \( \langle r^2(\theta) \rangle_\phi \) is proportional to \( \theta \) and inversely proportional to \( N \). If we define \( \langle r^2(\theta) \rangle_\phi \) as the MSD in the base mechanical trajectory with COM displacement-corrected end time coordinates, \( \langle r^2(\theta) \rangle_\phi = \langle r^2(\theta) \rangle_\phi - \delta^2_{\text{COM}}(\theta) \), then the MSD in the looped mechanical trajectory can be approximated as:

\[
\langle r^2(t) \rangle_\phi \approx \frac{t}{\theta} \langle r^2(\theta) \rangle_\phi + \frac{t}{\theta} \times \langle r^2 \rangle_{\text{join}} + \left( \frac{t}{\theta} \right)^2 \times \delta^2_{\text{COM}}(\theta).
\]

After multiple joins, on timescales that are much longer than the reactor-scale Brownian time, the influence of the COM drift on MSD may be significant [see Fig. 3(d), solid lines]. If obtained MSDs one subtracts the respective expected contribution of the COM drift (calculated separately for each base mechanical trajectory), a linear time dependence is recovered [Fig. 3(d), dashed lines]. It should be emphasized that a super-linear increase of MSD in not indicative of accelerating diffusion trajectories, a linear time dependence is recovered [Fig. 3(d), dashed lines].

\[
\phi = 0.2, \quad \langle r^2 \rangle_{\text{join}} = 0.3375, \quad \epsilon = 0.05, \quad \delta = 0.1, \quad q_0 = \lambda/\tau_B, \quad k_0 = 0.01 q_0.
\]

C. Reaction kinetics in a monostable system

We consider a simple phosphorylation-dephosphorylation reaction cycle,

\[
K + S_u \xrightarrow{k_0} K + S_p, \quad P + S_p \xrightarrow{q_0} P + S_u,
\]

where \( K \) and \( P \) denote enzymes, kinase and phosphatase, and \( S_u, S_p \) denote dephosphorylated and phosphorylated substrate molecules, respectively. This monostable reaction system has been studied in, e.g., Ref. [20] where it is demonstrated that the steady-state fraction of phosphorylated substrates can increase or decrease with diffusivity depending on relative concentrations of both enzymes. In Ref. [20], we focus on the case when \( K \) are more abundant than \( P \) \((K = 10 P)\) and \( K \) are catalytically much less active than \( P \) \((q_0 = 100 k_0)\). In this case, the fraction of phosphorylated \( S \) markedly depends on the diffusion coefficient: at low diffusivity, \( S \) remain mostly in the phosphorylated state; at high diffusivity, \( S \) are mostly dephosphorylated.

The diffusion coefficient in the base mechanical trajectory cannot be changed; however, by scaling uniformly the rate constants of all chemical reactions within trajectory looping one can simulate the stochastic kinetics in either reaction-controlled or diffusion-controlled regime. When the average distance traveled by a phosphorylated substrate until dephosphorylation (Kuramoto correlation length), \( \sqrt{D_0/q_0} \), is larger than the diameter of the reactor, \( L \), then the chemical reactor shall be considered well mixed. Henceforth, we assume that the dephosphorylation reaction rate \( q_0 = \lambda/\tau_B \) sets the time scale of chemical processes, that is, all other chemical rates are defined relative to \( q_0 \). Since propensities of reactions are proportional to \( \lambda \), decreasing \( \lambda \) is equivalent to simultaneously increasing diffusivity and lengthening the time scale.

In Fig. 4(a) we demonstrate that the equilibration kinetics of the monostable system in Eq. (1) simulated within trajectory looping in the reaction-controlled regime (in the well-mixed reactor) matches very well the equilibration kinetics simulated according to the Gillespie algorithm with both bimolecular reaction rates multiplied by \( \chi/N \). For parameter values assumed in Fig. 4, the reactor becomes well mixed for \( \lambda \) of order of \( 10^{-2} \). For this \( \lambda \), to reach the end time of 300 \( q_0^{-1} \) in Fig. 4(a), the base contacts sequence of duration \( \theta = 3 \tau_B \) had to be used \( 10^3 \) times (for \( \theta = 10 \tau_B \)).

\[
\begin{align*}
\text{Gillespie} & \quad \text{SRR} \\
\text{TLoop, } \theta = 3 \tau_B & \quad \text{TLoop, } \theta = 10 \tau_B
\end{align*}
\]

FIG. 4. Fraction of phosphorylated substrates, \( S_p/(S_u + S_p) \), in the monostable reaction system (1). (a) Fraction of phosphorylated substrates in time. Initially, \( S_p = S \ (S_u = 0) \). Two methods are used to simulate system equilibration: Gillespie algorithm with scaled reaction rates (GillespieSRR) and trajectory looping (“TLoop”) with \( \lambda = 10^{-2} \) applied to base mechanical trajectories of two different lengths, \( \theta \). Each curve results from averaging over 3 stochastic simulations. (b) Fraction of phosphorylated substrates in the steady state as a function of \( \lambda \) for base mechanical trajectories of different lengths, \( \theta \). In the fast-diffusion regime, the fraction tends to 1/11, whereas in the slow-diffusion regime, the fraction tends to 10/11 (both limits are marked with dashed lines). In both panels: \( N = 3375, \ K = [0.3 N], \ P = [0.03 N], \ S = N - (K + P), \phi = 0.2 \ (\chi \approx 0.855), \epsilon = 0.05, \delta = 0.1, \ q_0 = \lambda/\tau_B, \ k_0 = 0.01 q_0 \).

\[
\begin{align*}
\text{Gillespie} & \quad \text{SRR} \\
\text{TLoop, } \theta = 3 \tau_B & \quad \text{TLoop, } \theta = 10 \tau_B
\end{align*}
\]

The intensity of reactions increases with \( \lambda \) and the system exhibits a gradual transition from the reaction-limited to the diffusion-limited regime, as demonstrated in Fig. 4(b). In the fast-diffusion regime, the fraction tends to 1/11, whereas in the slow-diffusion regime, the fraction tends to 10/11, in agreement with the results of [20]. For intermediate \( \lambda = 10 \) and \( 10^2 \), the fraction of phosphorylated substrates in Fig. 4(b) is slightly lower for \( \theta = 3 \tau_B \) than for \( \theta = 10 \tau_B \), which results from the fact that joins that are more frequent than for \( \theta = 10 \tau_B \) introduce additional mixing [as analyzed in Fig. 3(b)].
D. State-to-state transition rates in a bistable system

Trajectory looping can be applied to enhance sampling of rare events in stochastic spatially extended systems. In this subsection we analyze events of switching between metastable states in a bistable reaction–diffusion system that has been analyzed previously by means of on-lattice kinetic Monte Carlo simulations in Ref. [21]. The system contains autophosphorylating bisphosphorylatable kinases $K$ and phosphatases $P$, that react as follows:

\[ K_{u} + K_{p} \xrightleftharpoons{k_{c}} K_{u} + K_{p}, \quad K_{u} + K_{pp} \xrightleftharpoons{k_{c}} K_{u} + K_{pp}, \quad (2a) \]

\[ K_{p} + K_{u} \xrightleftharpoons{k_{c}} K_{p} + K_{u}, \quad K_{p} + K_{pp} \xrightleftharpoons{k_{c}} K_{p} + K_{pp}, \quad (2b) \]

\[ K_{pp} + K_{u} \xrightleftharpoons{k_{c}} K_{pp} + K_{u}, \quad K_{pp} + K_{p} \xrightleftharpoons{k_{c}} K_{pp} + K_{p}, \quad (2c) \]

\[ P + K_{pp} \xrightarrow{k_{p}} P + K_{p}, \quad P + K_{p} \xrightarrow{k_{p}} P + K_{u}. \quad (2d) \]

Stochastic kinetics of the system exhibits fast fluctuations in the vicinity of each of the metastable steady states and relatively infrequent switches between basins of attraction of these states, as shown in Fig. 5(a).

Of note, the timescale associated with switching can be much longer than $\theta$ or the reactor-scale Brownian time. Trajectories of the system simulated with trajectory looping with a sufficiently small $\lambda$ exhibit characteristic bimodal distribution of states, depicted in Fig. 5(b) (for $\lambda = 10^{-1}$ the system is effectively monostable). Stationary state distribution is sensitive to $\lambda$. For $\lambda = 1 \times 10^{-3}$, when the system is well mixed, the distribution matches perfectly the distribution obtained from the Gillespie algorithm-based simulations with scaled reaction rates. The residence times or, equivalently, the mean first-passage times that correspond to distributions shown in Fig. 5(b), obtained each from at least 1000 switches, are contained in Fig. 5(c). Since in cellular biochemistry kinase phosphorylation is often associated with its activation, the state of low (high) number of $K_{u}$ is referred to as the active (inactive) state. The ratio $\tau_{\text{active}}/\tau_{\text{inactive}}$ is the same as the probability mass contained in the first peak of a respective bimodal distribution in Fig. 5(b). When in the Gillespie algorithm-based simulations the phosphorylation rates ($c_{1}$, $c_{2}$, and $c_{3}$) are either decreased or increased by just one per-mille (0.1%), $\tau_{\text{active}}$ and $\tau_{\text{inactive}}$ are noticeably different [Fig. 5(c)]. This, in addition to the match of probability distributions shown in Fig. 5(b), indicates that the well-mixed limit parameters can be reproduced by trajectory looping to a very high accuracy.

E. Traveling waves in a bistable system

In spatially extended domains, the stochastic bistable reaction–diffusion system, Eq. (2), admits traveling wave solutions. In the mean-field limit, the evolution of the normalized kinase species concentrations $k_{u}$, $k_{p}$, and $k_{pp}$ (where $k_{u} + k_{p} + k_{pp} = 1$) is governed by the system of partial differential equations (PDEs):

\[
\frac{\partial k_{u}}{\partial t} = D\nabla^{2}k_{u} + Q_{0}k_{p} - 2(C_{1}k_{u} + C_{2}k_{p} + C_{3}k_{pp})k_{u}, \quad (3a)
\]

\[
\frac{\partial k_{p}}{\partial t} = D\nabla^{2}k_{p} + 2(C_{1}k_{u} + C_{2}k_{p} + C_{3}k_{pp})k_{u} + 2Q_{0}k_{pp} - (C_{1}k_{u} + C_{2}k_{p} + C_{3}k_{pp})k_{p} - Q_{0}k_{p}, \quad (3b)
\]

\[
\frac{\partial k_{pp}}{\partial t} = D\nabla^{2}k_{pp} + (C_{1}k_{u} + C_{2}k_{p} + C_{3}k_{pp})k_{p} - 2Q_{0}k_{pp}, \quad (3c)
\]

where the diffusion coefficient $D$ is equal to that of the spatial stochastic system for the appropriate volume fraction. Reaction rate coefficients $C_{i}$ and $Q_{0}$ are related to the original coefficients $c_{i}$ and $q_{0}$ as follows: $C_{i} = (K/N)c_{i}$ and $Q_{0} = (P/N)q_{0}$ where $N$ is the
total number of molecules, $K$ is the number of kinases, and $P$ is the number of phosphatases. To obtain traveling wave solutions to Eqs. (3), we assume that the functions $k_u, k_p,$ and $k_{pp}$ depend only on time and one spatial variable $z$, and start from initial condition in which for $z < 0$ the system assumes one of two stable states, whereas for $z > 0$ it assumes the other.

We performed simulations of the spatial stochastic model within trajectory looping using the base mechanical trajectory of length $θ = 10τ_0$ in an elongated domain that contained $N = 5000$ molecules at the volume fraction $φ = 0.2$. The simulations started from the initial condition [Fig. 6(a)] in which the whole domain was in the inactive stationary state (high fraction of $K_u$) except for a stretch of length of 10% of the domain, which is in the active state (low fraction of $K_u$). Propagation of (stochastic) activating traveling wave implies that the area occupied by the active state increases (approximately) linearly with time as $2ν \times t$, where $ν$ is the propagation velocity; coefficient 2 arises from the fact that the traveling wave propagates in both directions. As a result, the fraction of $K_u$ averaged over the whole domain, $fu(t)$, decreases linearly with time [Fig. 6(b)]:

$$\frac{df_u}{dt} = 2ν(f_u^{\text{active}} - f_u^{\text{inactive}})/\ell,$$

where $f_u^{\text{active}}$ and $f_u^{\text{inactive}}$ denote the value of $f_u$ in the active state and in the inactive state, respectively, and $\ell$ is the length of the domain. Based on the above formula we can calculate the traveling wave velocity as

$$ν = \frac{df_u}{dt} \frac{\ell}{2(f_u^{\text{active}} - f_u^{\text{inactive}})}.$$  

To estimate $ν(λ)$ numerically, for each $λ$ we performed 20 simulations and based on the profiles of $f_u^{\text{inactive}}(t)$ we estimated $df_u^{\text{inactive}}/dt$. In Fig. 6(c) we show that these estimates are in agreement with the values obtained from numerical solutions of Eqs. (3) in MATLAB (The MathWorks, Inc., USA).

III. DISCUSSION

Systems of coupled biochemical reactions which underlie cellular regulatory processes often have two or more steady states associated with cell physiological outcomes [22–24]. Noise—inherent in biochemical systems—allows for transitions between these states [25, 26]. The transitions can however be notoriously rare: biochemically implemented genetic switch in phage $λ$ requires millions of generations to switch on the lytic pathway [27]. Since transition times are typically much longer than relaxation times of monostable biochemical systems, sampling of transition events in spatially extended systems by means of direct simulations is computationally very demanding.

Chemical kinetics in spatially extended stochastic systems strongly depends on the diffusivity of substrates. While in well-mixed systems time to a stochastic transition grows exponentially with the size of the system [28], spatially extended systems at a small diffusivity have their transition rates very different from those of the corresponding systems in the well-mixed limit. This is because for slow diffusion the state-to-state transitions can be achieved by means of a local transition that initi-
ates a traveling wave [1] which then drives the whole rea-
tor towards the “more stable” steady state [29]. Even
though the expected time for nucleation of the wave is
much shorter than the time necessary for a simultane-
ous switch of the whole domain, both timescales may be
orders of magnitude longer than time required by a
molecule to cover a distance comparable to the reactor
diameter.

Trajectory looping is capable of reaching simulated
times orders of magnitude longer than the simulated
length of base mechanical trajectory, significantly en-
hancing the sampling of rare events. Assuming that $\tau_B$
of a protein molecule in a signaling system is of order of
$10^{-6}$ s of real time (as can be estimated based on,
e.g., Ref. [30]), simulated time required to estimate tran-
sition times in the analyzed bistable system for $\lambda = 10^{-2}$
should be of order of an hour. The trajectory looping al-
gorithm is thus suitable for simulating reaction–diffusion
systems on the physiologically relevant timescales at the
single-molecule resolution. The method is independent of
the numerical scheme used to obtain the underlying base
mechanical trajectory; it can be applied equally well to
mechanical trajectories with or without hydrodynamic
interactions.

We verified the accuracy of the proposed method in
three cases: (i) in the limits of fast and slow diffusion, by
comparing simulated steady state values of monostable
system with an analytical prediction; (ii) comparing simulated steady state values of monostable
systems with those obtained using Gillespie algorithm;
(iii) fast diffusion, by comparing trajectory looping simula-
tions with or without hydrodynamic
mechanical trajectories. It can be applied equally well to
the numerical scheme used to obtain the underlying base
single-molecule resolution. The method is independent of
the numerical scheme used to obtain the underlying base
mechanical trajectory; it can be applied equally well to
mechanical trajectories with or without hydrodynamic
interactions.

In addition to the problems analyzed in this article, we
expect that trajectory looping can be applicable to cap-
ture and characterize other behaviors and properties that
can be hardly observable in spatial stochastic simulations
such as the range of a homoclinic traveling wave, time to
extinction of an oscillating system, or synchronization of
oscillators (both in systems with a limit cycle). Exem-
plary binolecular reaction systems that are extensions of
system Eq. (2) and allow for either a homoclinic traveling
wave or a limit-cycle oscillations are provided in the
Appendix.

IV. METHODS

A. Brownian dynamics simulation

The algorithm of trajectory looping is described in de-
tail in the first subsection of the Results section. The
base mechanical trajectory, which is an input to trajec-
tory looping, is obtained independently. Below we de-
scribe the method for obtaining the base mechanical tra-
djectories used in simulations reported in this article.

Base mechanical trajectories were generated using
standard Brownian dynamics simulations of $N$ identical
molecules. The volume $V$ of the simulation box was set to
match the required volume fraction $\phi = N \times v/V$, where
$v = (4/3)\pi a^3$ is the single-molecule volume. In both cub-
ic and elongated cuboidal boxes, periodic boundary con-
ditions were assumed at all boundaries. Average distance
traveled by a molecule in a single simulation time step $\Delta t$
is of order of $\sqrt{\Delta t} \times D_0$. To set the time step we com-
pare this distance with characteristic surface-to-surface
distance between molecules, $d_{ss} = (V/N)^{1/3} - 2a$, which
is the distance between surfaces of neighboring molecules
arranged in a cubic lattice at a given $\phi$. To this end, we
introduce parameter $\epsilon$ and set $\Delta t = (\epsilon \times d_{ss})^2/D_0$. In all simulations $\epsilon \times d_{ss} < \delta$, which makes the probability of omitting a reaction event low.

Molecules interact with a hard-sphere potential ensur-
ing perfectly elastic collisions [31, 32]. In every time step, potential random displacements of all molecules are calculated with zero mean and the variance equal
$\langle r^2(\tau = \Delta t) \rangle_0 = 6D_0 \times \Delta t$. Based on initial molecule po-
sitions and velocities calculated from the displacements,
a list of collisions is created. If there are no collisions,
molecules are moved directly to their target locations.
If the list of colliding molecules is nonempty, molecules
are moved with constant velocities until a first collision.
Upon a collision, positions and velocities of the colliding
molecules and the list of all further collisions are updated.
Due to the presence of other molecules in the simula-
tion box, the mean square displacement of a molecule in
the base mechanical trajectory of duration $\theta$, $\langle r^2(\tau = \theta) \rangle_0$
$= \sum_{i=1}^{N} [\vec{R}_i(\theta) - \vec{R}_i(0)]^2/N$, is smaller than $\langle r^2(\tau = \theta) \rangle_0$
$= (M - 1) \times \langle r^2(\tau = \Delta t) \rangle_0$.

B. Simulation times

A detailed digest of computational cost of the proce-
dures used to generate the key data structures present
in Fig. 1 is provided in Table I. Generation of the base
mechanical trajectory that was used to estimate prob-
ability density distribution in the bistable reaction sys-
tem Eq. (2), presented in Fig. 5(b), required about 3 h of
a single-core processor. At $\lambda = 10^{-2}$, simulation
of the base mechanical trajectory for $\theta = 1/\langle q_{\text{in}}^{-1} \rangle$
whereas simulation time to observe 1000 switches that
were used to estimate both $\tau_{\text{active}}$ and $\tau_{\text{inactive}}$ (Fig. 5(c))
is $1000 \times (8.5 + 7.4) \times 10^3 \approx 1.6 \times 10^7 / \langle q_{\text{in}}^{-1} \rangle$, which means
that within trajectory looping the base mechanical tra-
djectory was reused $K = 1.6 \times 10^7$ times. Single-core
processor time to obtain the looped chemical trajectory
(see Table I for processor times of individual procedures)
was $3 \text{ h} + 1 \text{ min} + K \times (10^{-2} \text{ s} + 0.5 \text{ s}) \approx 3$ months (in
practice, the simulations were performed in parallel on a
computer cluster, so this is an aggregate real time). In
a direct, naive approach without trajectory looping, in
which diffusive steps are interleaved with simulation of chemical reactions, single-core processor time to obtain equally long chemical trajectory could be estimated (see Table I) as $K \times (3 \text{ h} + 1 \text{ min} + 0.5 \text{ s})$, which amounts to more than 5000 years.

We should notice that simulation of the trajectory used to estimate probability density distribution for the same reaction system (2) [dotted line in Fig. 5(b)], but without spatial resolution (i.e., under the assumption that the system is perfectly mixed) using Gillespie algorithm takes only 5 min.

**ACKNOWLEDGMENTS**

Numerical codes used to generate base mechanical trajectories were written by late Prof. Eligiusz Wajnryb. Simulations were performed using computational resources of ACK Cyfronet AGH cluster Prometheus (within the PL-grid infrastructure) and the Della cluster of the Princeton University. P.J.Z. was supported by the Polish Ministry of Science and Higher Education (“Mobility Plus” Project No. 1294/MOB/IV/2015/0). T.L. acknowledges support from the National Science Centre (NCN), Poland (Grant No. 2014/14/M/NZ6/00537). P.J.Z and M.K. contributed equally to this work.

**APPENDIX**

Extended bimolecular reaction system

When the system Eq. (2) is amended with an additional mechanism of inactivation of kinases $K$, mediated by a $K$-phosphorylatable inhibitor $I$ that shares phosphatase $P$ with kinases $K$:

\[
K_{pp} + I_u \xrightarrow{c_4} K_{pp} + I_u,
\]

\[
K_{pp} + I_p \xrightarrow{c_3} K_{pp} + I_p,
\]

\[
I_{pp} + K_{pp} \xrightarrow{c_5} I_{pp} + K_u,
\]

\[
I_{pp} + K_p \xrightarrow{c_5} I_{pp} + K_u,
\]

then the corresponding system of PDEs:

\[
\frac{dk_u}{dt} = D\nabla^2 k_u + (Q_0 + C_5i_{pp})k_p - 2(C_1k_u + C_2k_p + C_3k_{pp})k_u,
\]

\[
\frac{dk_p}{dt} = D\nabla^2 k_p + 2(C_1k_u + C_2k_p + C_3k_{pp})k_u + 2(Q_0 + C_5i_{pp})k_{pp} - (C_1k_u + C_2k_p + C_3k_{pp})k_p - (Q_0 + C_5i_{pp})k_p,
\]

\[
\frac{dk_{pp}}{dt} = D\nabla^2 k_{pp} + (C_1k_u + C_2k_p + C_3k_{pp})k_p - 2(Q_0 + C_5i_{pp})k_{pp},
\]

\[
\frac{di_u}{dt} = D\nabla^2 i_u - 2C_4k_{pp}i_u + Q_1i_p,
\]

\[
\frac{di_p}{dt} = D\nabla^2 i_p + 2C_4k_{pp}i_u + 2Q_1i_{pp} - (Q_1 + C_4k_{pp})i_p,
\]

\[
\frac{di_{pp}}{dt} = D\nabla^2 i_{pp} - 2Q_1i_{pp} + K_4k_{pp}i_p
\]

exhibits excitable (traveling pulse) dynamics for parameter values:

\[
Q_0 = 1,
\]

\[
C_1 = 0.02,
\]

\[
C_2 = 0.2,
\]

\[
C_3 = 4,
\]

\[
C_4 = 0.03,
\]

\[
C_5 = 1,
\]

\[
Q_1 = 0.003,
\]

\[
D = 0.1
\]

and limit-cycle oscillations for parameter values:

\[
Q_0 = 1,
\]

\[
C_1 = 3 \times 0.02,
\]

\[
C_2 = 3 \times 0.2,
\]

\[
C_3 = 3 \times 4,
\]

\[
C_4 = 10 \times 0.03,
\]

\[
C_5 = 10 \times 1,
\]

\[
Q_1 = 10 \times 0.003,
\]

\[
D = 0.1
\]

\[\text{Table I.} \text{ Computational cost associated with key procedures used to perform simulations with and without looping. Computational complexity of obtaining the base mechanical trajectory depends on the underlying method, and may vary from } \mathcal{O}(N) \text{ for cell list-based approaches to } \mathcal{O}(N^3) \text{ in the case of simulations that include, e.g., hydrodynamic interactions. The last column contains single-core processor times and the number of calls to a given procedure in the case of a trajectory looping-based simulation.}\]

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Number of calls</th>
<th>Complexity</th>
<th>CPU time for Fig. 5 (λ = 10^{-2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base mechanical trajectory (of length θ)</td>
<td>$K$</td>
<td>1</td>
<td>$\mathcal{O}(N)$ to $\mathcal{O}(N^3)$</td>
</tr>
<tr>
<td>Base contacts sequence</td>
<td>$K$</td>
<td>1</td>
<td>$\mathcal{O}(N^2)$</td>
</tr>
<tr>
<td>Assignment (Hungarian algorithm)</td>
<td>0</td>
<td>$K$</td>
<td>$\mathcal{O}(N^3)$</td>
</tr>
<tr>
<td>Chemical trajectory (within θ)</td>
<td>$K$</td>
<td>$K$</td>
<td>$\mathcal{O}(N)$</td>
</tr>
</tbody>
</table>


