The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

Slack reactants: A state-space truncation framework to estimate quantitative behavior of the chemical master equation

Cite as: J. Chem. Phys. 153, 054117 (2020); doi: 10.1063/5.0013457
Submitted: 11 May 2020 • Accepted: 15 July 2020 •
Published Online: 5 August 2020

Jinsu Kim,[1,a] Jason Dark,[1,2] German Enciso,[1,3,b] and Suzanne Sindi[2,c]

AFFILIATIONS

[1] Department of Mathematics, University of California, Irvine, California 92697, USA
[2] Department of Applied Mathematics, University of California, Merced, California 95343, USA
[3] Department of Developmental and Cell Biology, University of California, Irvine, California 92617, USA

[a] Author to whom correspondence should be addressed: jinsu.kim@uci.edu
[b] Electronic mail: enciso@uci.edu
[c] Electronic mail: ssindi@ucmerced.edu

ABSTRACT

State space truncation methods are widely used to approximate solutions of the chemical master equation. While most methods of this kind focus on truncating the state space directly, in this work, we propose modifying the underlying chemical reaction network by introducing *slack reactants* that indirectly truncate the state space. More specifically, slack reactants introduce an expanded chemical reaction network and impose a truncation scheme based on desired mass conservation laws. This network structure also allows us to prove inheritance of special properties of the original model, such as irreducibility and complex balancing. We use the network structure imposed by slack reactants to prove the convergence of the stationary distribution and first arrival times. We then provide examples comparing our method with the stationary finite state projection and finite buffer methods. Our slack reactant system appears to be more robust than some competing methods with respect to calculating first arrival times.

I. INTRODUCTION

Chemical reaction networks (CRNs) are a fundamental tool in the modeling of biological systems, providing a concise representation of known chemical or biological dynamics. A CRN is defined by a family of chemical reactions of the form

$$\sum_{i=1}^{n} \alpha_i X_i \to \sum_{i=1}^{n} \beta_i X_i,$$

where α_i and β_i are the number of species X_i consumed and produced in this reaction, respectively. The classical approach to modeling this CRN is to consider the concentrations $c(t) = (c_1(t), c_2(t), \ldots, c_n(t))^\top$, where $c_i(t)$ is the concentration of species X_i at time t, and to use a system of nonlinear differential equations to describe the evolution of the concentrations.

Suppose we are interested in studying the typical enzyme–substrate system[1] given by the CRN

$$S + E \rightleftharpoons C \to P + E, \tag{1}$$

where the species S, E, C, and P stand for a substrate, an enzyme, an enzyme–substrate complex, and a product, respectively. The substrate and enzyme bind reversibly to form a substrate–enzyme complex, and the enzyme then acts upon the substrate forming the product. Given an initial condition where the molecular numbers of both C and P are 0, a natural question to ask is how long the system takes to produce the *first* copy of P. Similarly, there exists a time in the future where S and C are fully depleted, resulting in a chemically inactive system, and one can ask when this occurs. These are both quantities that the classical deterministic modeling leaves unanswered—by considering only continuous concentrations, there

is no well-defined way to address modeling questions at the level of single molecules as the model assumes that all the reactions simultaneously occur within infinitesimally small time intervals.

Instead, by explicitly considering each individual copy of a molecule, we may formulate a continuous-time Markov chain. This stochastic modeling is especially important when the system consists of low copies of species, in which case the influence of intrinsic noise is magnified.[2–4] Rather than deterministic concentrations $c(t)$, we consider the continuous-time Markov chain $X(t) = (X_1(t), X_2(t), \ldots, X_d(t))^\top$ describing the molecular number of each species X_i at time t.

The questions regarding the enzyme–substrate system (1), such as the time of the first production of P, simply correspond to the *first passage times* of various combinations of states. For a continuous-time Markov process X, the first passage time to visit a set K of system states is formally defined as $\tau = \inf\{t \geq 0 : X(t) \in K\}$.

One can directly compute $E(\tau)$ by using the transition rate matrix A. With few exceptions (such as CRNs with only zero- or first-order reactions[5]), most chemical reaction networks of any notable complexity will have an intractably large or infinite state-space, i.e., they exhibit the curse of dimensionality. This direct approach can, therefore, suffer from the high dimensionality of the transition rate matrix.

An alternative approach is to estimate the mean first passage time by generating sample trajectories with stochastic simulation algorithms such as the Gillespie algorithm.[6] This overcomes the curse of dimensionality since a single trajectory needs only to keep track of its current population numbers. Nevertheless, there still remain circumstances under which it is more efficient to numerically evaluate the exact solution of the chemical master equation (CME)—in particular, when the Markov process rarely visits K so that significantly long simulations may be required to sample enough trajectories to estimate the mean first passage time.[7]

Fortunately, a broad class of state space reduction methods have recently been developed that allow for direct treatment of the transition rate matrix. These methods are based on the truncation-and-augmentation of the state space,[8–11] the finite buffer method,[12,13] and linear programming.[14,15] A recent review summarizes truncation-based methods.[16] The *stationary finite state projection (sFSP)*[11] and the *finite buffer method*[12,13] are examples of such truncation-based methods, which we will describe in detail below. They all satisfy provable error estimates on the probability distributions when compared with the original distribution. On the other hand, each of these methods has potential limitations for estimating mean first passage times and other quantitative features. For instance, using sFSP depends on the choice of a designated state, which can significantly alter the estimate for first passage times.

In this paper, we provide a new algorithm of state space reduction, the *slack reactant method*, for stochastically modeled CRNs. In this algorithm, we generate a new CRN from an existing CRN by adding one or multiple new species so that the associated stochastic system satisfies mass conservation relations and is confined to a finite number of states. For instance, we convert a simple birth and death model $\emptyset \rightleftharpoons X$ admitting an infinite state space to $Y \rightleftharpoons X$ with the "slack reactant" Y to confine it on a finite state space. In order to ensure equivalent dynamics to the original system, we define a mild form of non-mass action kinetics for the new system. Since the state space reduction is implemented using a fully determined CRN, we can study the CRN using well-known tools of CRN theory such as deficiency zero and Lyapunov functions as long as they extend to this form of kinetics.

In Sec. III B, we provide an algorithm to produce a slack variable network, given a desired set of mass conservation relations. In addition to its theoretical uses, this algorithm allows us to implement existing software packages such as CERENA,[17] StochDynTools,[18] and FEEDME[19] for chemical reaction networks to generate quantities such as the moment dynamics of the associated stochastic processes using the network structures as input.

We employ classical truncation Markov chain approaches to prove convergence theorems for slack networks. For fixed time t, if a probability density of each slack system under conservation quantity N converges to its stationary distribution uniformly in N, then the stationary distribution of the slack system converges to the original stationary distribution as N tends to infinity. We further prove that under a uniform tail condition of first passage times, the mean first passage time of the original system can be approximated with slack systems confined on a sufficiently large truncated state space. Finally, we show that the existence of the Lyapunov function for the original system guarantees that all the sufficient conditions for the first passage time convergence are satisfied. We also show that this truncation method is natural in the sense that a slack system admits the same stationary distribution up to a constant multiplication as the stationary distribution of the original system if the original system is complex balanced.

This paper is outlined as follows: In Sec. III, we introduce the slack reactant method and include several illustrative examples. In Sec. IV, we demonstrate that the slack method compares favorably with other state space truncation methods (sFSP and finite buffer method) when calculating mean first passage times. We prove convergence in the mean first passage time, and other properties, in Sec. V. In Sec. VII, we use slack reactants to estimate the mean first passage times for practical CRN models such as a Lotka–Volterra population model and a system of protein synthesis with a slow toggle switch.

II. STOCHASTIC CHEMICAL REACTION NETWORKS

A chemical reaction network (CRN) is a graph that describes the evolution of a biochemical system governed by a number of *species* ($\mathscr{S}$) and *reactions* ($\mathscr{R}$). Each node in the graph represents a possible state of the system, and nodes are connected by directed edges when a single reaction transforms one state into another. Each reaction consists of *complexes* and a *reaction rate constant*. For example, the reaction representing the transformation of complex ν to complex ν' at rate κ is written as follows:

$$\nu \xrightarrow{\kappa} \nu'. \tag{2}$$

A complex, such as ν, is defined as a number of each species $S_i \in \mathscr{S}$. That is, $\nu = (\nu_1, \nu_2, \ldots, \nu_d)$ representing a complex $\sum_{i=1}^{d} \nu_i S_i$, where $\nu_i \geq 0$ are the *stoichiometric coefficients* indicating how many copies of each species $S_i \in \mathcal{S}$ belong in complex ν. The full CRN is thus defined by $(\mathscr{S}, \mathscr{C}, \mathscr{R}, \Lambda)$, where $\mathscr{C}$ and Λ represent the set of complexes and reaction propensities, respectively.

When intrinsic noise plays a significant role for system dynamics, we use a continuous-time Markov chain $\mathbf{X}(t) = (X_1(t), X(2), \ldots, X_d(t))$ to model the copy numbers of species S_i of a reaction network. The stochastic system treats individual reactions as discrete transitions between integer-valued states of the system. The probability density for $X(t)$ is denoted as

$$p_{\mathbf{x}_0}(\mathbf{x}, t) = P(\mathbf{X}(t) = \mathbf{x} \mid \mathbf{X}(0) = \mathbf{x}_0),$$

where $\mathbf{X}(0)$ is the initial state. We occasionally denote by $p(\mathbf{x}, t)$ or $P(X(t) = \mathbf{x})$ the probability omitting the initial state $\mathbf{x}_0$ when the contexts allow. For each state $\mathbf{x}$, the probability density $p(\mathbf{x})$ obeys the chemical master equation (CME), which gives the time-evolution of $p(t)$ with a linear system of ordinary differential equations (ODEs),[20]

$$\frac{d}{dt}\mathbf{p}^\top(t) = \mathbf{p}^\top A. \tag{3}$$

Here, the entry A_{ij} $(i \neq j)$ is the transition rate at which the ith state transitions to the jth state. Letting $\mathbf{x}$ and $\mathbf{x}'$ be the ith and jth states, the transition rate from $\mathbf{x}$ to $\mathbf{x}'$ is

$$A_{ij} = \sum_{\substack{\nu \to \nu' \\ \mathbf{x}+\nu'-\nu=\mathbf{x}'}} \lambda_{\nu\to\nu'}(\mathbf{x}),$$

where $\lambda_{\nu\to\nu'}$ is the *reaction intensity* for a reaction $\nu \to \nu'$.

The diagonal elements of A are defined as

$$A_{jj} = -\sum_{i\neq j} A_{ij}.$$

Regarding the reaction intensities, we will assume that

$$\lambda_{\nu\to\nu'}(\mathbf{x}) > 0 \quad \text{if and only if} \quad x_i \geq \nu_i \quad \text{for each } i. \tag{4}$$

This is a slightly stronger condition than the so-called *stoichiometric compatibility*[21] of the intensity function, namely, that $\lambda_{\nu\to\nu'}(\mathbf{x})$ can only be positive if $x_i \geq \nu_i$ for each i.

A typical choice for $\lambda_{\nu\to\nu'}$ is *mass-action*, which defines for $\mathbf{x} = (x_1, x_2, \ldots, x_d)$ and $\nu = (\alpha_1, \alpha_2, \ldots, \alpha_d)$,

$$\lambda_{\nu\to\nu'}(\mathbf{x}) = \kappa_{\nu\to\nu'} \prod_{k=1}^{d} x_i^{(\alpha_i)},$$

where $\kappa_{\nu\to\nu'}$ is the rate constant. Here, we used the notation $m^{(n)} = m(m-1)(m-2)\cdots(m-n+1)\mathbb{1}_{\{n\geq m\}}$ for positive integers m and n.

A state $\mathbf{x}$ is called an *absorbing state* if $\lambda_{\nu\to\nu'} = 0$ for all reactions $\nu \to \nu'$. This means that if the Markov chain enters this state, it will remain there for all time. A Markov chain $\mathbf{X}(t)$ with $\mathbf{X}(0) = \mathbf{x}_0$ is *accessible* to a state $\mathbf{x}_T$ if $P(\mathbf{X}(t) = \mathbf{x}_T \text{ for some } t \mid \mathbf{X}(0) = \mathbf{x}_0) = 1$.

III. CONSTRUCTION OF SLACK NETWORKS

In this section, we introduce the method of slack reactants, which adds new species to an existing CRN so that the state space of the associated stochastic process is truncated to a finite subset of states. This model reduction accurately approximates the original system as the size of the truncated state space increases. We begin with a simple example to demonstrate the main idea of the method.

A. Slack reactants for a simple birth-death model

Consider a mass-action birth-death model of a single species,

$$\emptyset \underset{\kappa_2}{\overset{\kappa_1}{\rightleftharpoons}} X. \tag{5}$$

For the associated stochastic process X, the mass-action assumption defines reaction intensities as $\lambda_{\emptyset\to X}(x) = \kappa_1$ and $\lambda_{X\to\emptyset}(x) = \kappa_2 x$ for each reaction in (5). Note that the count of species X could be any positive integer value as the birth reaction $\emptyset \to X$ can occur unlimited times. Therefore, the state space for this system consists of infinitely many states. Consider instead the CRN

$$Y \underset{\kappa_2}{\overset{\kappa_1}{\rightleftharpoons}} X, \tag{6}$$

where we have introduced the *slack reactant* Y. This new network admits $X + Y = N$ as a *conservation law* for some N since for each reaction, either one species is degraded by one, while the other is produced by one.

Since the purpose of this new network is to approximate the qualitative behavior of the original system (5), we minimize the contribution of the slack reactant Y for modeling the associated stochastic system. Hence, we assign Y a special reaction intensity—instead of $\lambda_{Y\to X}(x, y) = \kappa_1 y$ using mass-action, we choose

$$\lambda_{Y\to X}(x, y) = \kappa_1 \mathbb{1}_{\{y\geq 0\}}, \tag{7}$$

and we use the same intensity $\lambda_{X\to Y}(x, y) = \kappa_2 x$ for the reaction $X \to Y$. By forcing Y to have "zero-order kinetics," we ensure that the computed rates remain the same throughout the chemical network except for on the imposed boundary. This choice of reaction intensities not only preserves the conservation law $\mathbf{X}(t) + \mathbf{Y}(t) = \mathbf{X}(0) + \mathbf{Y}(0)$ but also prevents the slack reactant Y from having negative counts with the characteristic term $\mathbb{1}_{\{y\geq 0\}}$.

B. Algorithm for constructing a network with slack reactants

In general, by using slack reactants, *any* conservation law can be deduced in a similar fashion to encode the desired truncation directly in the CRN. We have found this perspective to be advantageous with respect to studying complex CRNs. Rather than thinking about the software implementation of the truncation, it is often easier to design the truncation in terms of slack reactants and then implement the already-truncated system exactly.

We now provide an algorithm to automatically determine the slack reactants required for a specified truncation. It is often the case that a "natural" formulation arises (typically by replacing zero-order complexes with slack reactants), but when that is not the case, one can still systematically find a slack network by following our algorithm.

Consider any CRN $(\mathscr{S}, \mathscr{C}, \mathscr{R}, \Lambda)$ such that $\mathscr{S} = \{X_1, X_2, \ldots, X_d\}$ and $\mathscr{C} = \{\nu^j \mid j = 1, 2, \ldots, |\mathscr{C}|\}$. Then, we define matrices associated with the given CRN.

1. Let S be the $|\mathscr{C}| \times |\mathscr{R}|$ *connectivity matrix* such that for each r,

$$-S_{ir} = S_{jr} = \begin{cases} 1 & \text{if } r\text{th reaction is } \nu^i \to \nu^j \\ 0 & \text{otherwise.} \end{cases}$$

2. Let C be the $|\mathscr{S}| \times |\mathscr{C}|$ *complex matrix* such that $C_{ij} = v_i^j$.
   Suppose we wish to apply a number of *conservation bounds* to reduce the state-space of the associated chemical master equation, e.g., many equations of the form
   $$w_{i,1}X_1 + w_{i,2}X_2 + \cdots + w_{i,d}X_d \le N_i, \tag{8}$$
   for $i = 1, 2, \ldots, m$. Then, we define additional matrices associated with the conservation bounds (8).
3. Let W be the $m \times d$ matrix of the conservation bounds in (8) such that $W_{ij} = w_{i,j}$, and let $\mathbf{N} = (N_1, N_2, \ldots, N_m)^\top$.
4. For arbitrary positive integers u_i, let U be an $m \times |C|$ matrix with each row of the form $u_i \mathbf{1}^\top = u_i(1, 1, \ldots, 1)^\top$.
5. Define $m \times |\mathscr{C}|$ matrix $D = U - WC$.
   Finally, we use these matrices to define the slack network $(\widetilde{\mathscr{S}}, \widetilde{\mathscr{C}}, \widetilde{\mathscr{R}})$.
6. Let $\widetilde{\mathscr{S}} = \{X_1, \ldots, X_d, Y_1, \ldots, Y_m\}$.
7. Let $\widetilde{\mathscr{C}} = \{\tilde{v}^j \mid j = 1, 2, \ldots, |\mathscr{C}|\}$, where $\tilde{v}^j$ is the jth column of $\widetilde{C} = \begin{pmatrix} C \\ D \end{pmatrix}$.
8. Let $\widetilde{\mathscr{R}} = \{\tilde{v}^i \to \tilde{v}^j \mid -S_{ir} = S_{ij} = 1 \text{ for some } r\}$.

We next verify that the newly generated network $(\widetilde{\mathscr{S}}, \widetilde{\mathscr{C}}, \widetilde{\mathscr{R}})$ with the slack reactants Y_i admits conservation laws

$$w_{i,1}X_1 + w_{i,2}X_2 + \cdots + w_{i,d}X_d + Y_i = N_i \tag{9}$$

for each $i = 1, 2, \ldots, m$ so that the state space is truncated onto a finite set and the conservation bounds (8) hold. Note that for a given CRN with the matrices S and C, conservation laws $\mathbf{r}^i \cdot X = N_i$ hold with conservation vectors $\mathbf{r}^1, \mathbf{r}^2, \ldots, \mathbf{r}^m$ if and only if each reaction vector is orthogonal to $\mathbf{r}^j$. This is equivalent to $R\Gamma = 0$, where R is a matrix with rows ${\mathbf{r}^j}^\top$ and $\Gamma = CS$ is the *stoichiometric matrix* whose columns indicate the reaction vector $v' - v$ associated with a reaction $v \to v' \in \mathscr{R}$.

Note that the network $(\widetilde{\mathscr{S}}, \widetilde{\mathscr{C}}, \widetilde{\mathscr{R}})$ is generated with the connectivity matrix S and the new complex matrix $\widetilde{C}$. Thus, the stoichiometric matrix of $(\widetilde{\mathscr{S}}, \widetilde{\mathscr{C}}, \widetilde{\mathscr{R}})$ is $\widetilde{\Gamma} = \widetilde{C}S$. Then, since $(1, 1, \ldots, 1)S = \mathbf{0}$, we have

$$(W\ I)\widetilde{\Gamma} = (W\ I)\widetilde{C}S = \mathbf{0}.$$

Therefore, (9) holds.

We model the stochastic system associated with the slack network by $\mathbf{X}^{\mathbf{N}}(t) = (X_1^{\mathbf{N}}(t), X_2^{\mathbf{N}}(t), \ldots, X_d^{\mathbf{N}}(t))$, where each of the entries represents the count of species in the new network. Note that the count of each slack reactant Y_i is fully determined by species counts $X_i^{\mathbf{N}}$'s because of the conservation law(s). As such, we do not explicitly model Y_i's.

The intensity function $\lambda^{\mathbf{N}}_{\tilde{v}\to\tilde{v}}$ of $X^{\mathbf{N}}$ for a reaction $\tilde{v} \to \tilde{v}'$ is defined as

$$\lambda^{\mathbf{N}}_{\tilde{v}\to\tilde{v}'}(\mathbf{x}) = \lambda_{v\to v'}(\mathbf{x}) \prod_{i=1}^{m} \mathbb{1}_{\{y_i \ge \tilde{v}_{d+1}\}}, \tag{10}$$

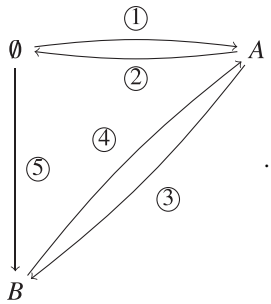
where $y_i = N_i - (w_{i1}x_1 + w_{i1}x_1 + \cdots + w_{id}x_d)$ and $v \to v'$ is the reaction in $\mathscr{R}$. Then, we denote by $(\widetilde{\mathscr{S}}, \widetilde{\mathscr{C}}, \widetilde{\mathscr{R}}, \Lambda^{\mathbf{N}})$ a new system with slack reactants obtained from the algorithm, where $\mathscr{K}_N$ is the collection of kinetic rates $\{\lambda^N_{\tilde{v}\to\tilde{v}'} : \tilde{v} \to \tilde{v}' \in \widetilde{\mathscr{R}}\}$. We refer this system with slack reactants to a *slack system*.

Here, we highlight that the connectivity matrix S and the complex matrix C of the original network are preserved for a new network with slack reactants. Thus, the original network and the new network obtained from the algorithm have the same connectivity. This identical connectivity allows the qualitative behavior of the original network, which solely depends on S and C, to be inherited to the new network with slack reactants. We particularly exploit the inheritance of accessibility and the inheritance of Poissonian stationary distribution in Sec. VI.

Remark III.1. A single conservation relation, such as $\mathbf{w} \cdot X + y = N$ *with a non-negative vector* $\mathbf{w}$*, is sufficient to truncate the given state space into a finite state space. Hence, in this manuscript, we mainly consider a slack network that is obtained with a single conservation vector* $\mathbf{w}$ *and a single slack reactant* Y.

Remark III.2. Although we primarily think about bounding our species counts from above, we could also bound species counts from below by choosing negative integers for $w_{i,j}$. *For instance, suppose that the stochastic process* X *associated with the birth-death model (5) satisfies* $X(0) = 100$. *If we are interested in computing the first hitting time to state 50, then by adding a slack reactant* Y*, we can truncate the state space of* X *from below with a conservation law* $-X(t) + Y(t) = -X(0) + Y(0) = -40$ *so that* $X(t) \ge 40$ *for each* t. *In this case, a slack network is* $X + Y \rightleftharpoons 0$ *with* $Y(0) = 60$.

Example III.1. We illustrate our slack algorithm with an example CRN consisting of two species and five reactions indicated by edge labels on the following network:



. (11)

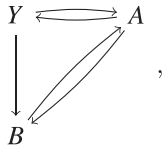
We enumerate the complexes in the order of $\emptyset$, A, and B. We order the reactions according to their labels on the network. Thus, the connectivity matrix S and complex matrix C are defined as follows:

$$S = \begin{bmatrix} -1 & 1 & 0 & 0 & -1 \\ 1 & -1 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 1 \end{bmatrix},$$

$$C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Suppose we set a conservation bound $A + B \le N$ for some $N > 0$. Then, the matrix $W = \begin{bmatrix} 1 & 1 \end{bmatrix}$ and $D = \begin{bmatrix} u & u & u \end{bmatrix} - \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}$ for an arbitrary positive integer u.

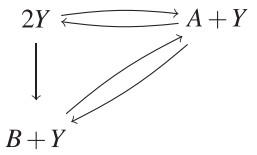
When $u = 1$, the network with the slack reactant Y is



(12)

where we have the conservation relation $A + B + Y = N$, where $N = A(0) + B(0) + Y(0)$.

When $u = 2$, the network with the slack reactant Y is



(13)

where we have the same conservation relation $A + B + Y = N$.

Here, we explain why network (12) is preferred to (13) because it is less intrusive. Let $\mathbf{X} = (X_A, X_B, X_Y)$ and $\mathbf{X}' = (X'_A, X'_B, X'_Y)$ be the stochastic process associated with networks (12) and (13), respectively. Suppose that the initial state of both $\mathbf{X}$ and $\mathbf{X}'$ is $(0, 0, N)$. $\mathbf{X}$ can reach the state $(0, N, 0)$ by transitioning N times with the reaction $Y \to B$. This state corresponds to the state $(0, N)$ in the original network (11). On the other hand, $\mathbf{X}'$ cannot reach the state $(0, N, 0)$. This is because the states $(0, 0, N)$ and $(0, N - 1, 1)$ are the only states from which $\mathbf{X}'$ jumps to $(0, 0, N)$. However, no reaction in (13) can be fired at the states since no species Y presents at those states.

Consequently, one state, which is accessible in the original network (11), is lost in the system associated with network (13). However, it can be shown that the stochastic process associated with network (12) preserves all the states of the original network. This occurs mainly because the matrix D for network (12) is sparser than the matrix D for network (13). We discuss how to minimize the effect of slack reactants in Sec. III C.

C. Optimal slack CRNs for effective approximation of the original systems

The algorithm we introduced to construct a network with slack reactants is valid and unique up to any user-defined conservation bounds (8), and the outcome is the matrix D that indicates the stoichiometric coefficient of slack reactants at each complex.

As we showed in Example III.1, to minimize the "intrusiveness" of a slack network, we can simplify a slack network by setting as many $D_{ij} = 0$ as possible. To do that, we choose the entries of $\mathbf{u}$ such that u_i is the maximum entry of the ith row of AC. We further optimize the effect of the slack reactants by removing the "redundant" stoichiometric coefficient of slack reactants. For example, for a CRN,

$$\emptyset \rightleftharpoons A \to 2A, \tag{14}$$

the algorithm generates the following new CRN with a single slack reactant Y:

$$2Y \rightleftharpoons A + Y \to 2A. \tag{15}$$

However, by breaking up the connectivity, we can also generate another network

$$Y \rightleftharpoons A, \quad A + Y \to 2A. \tag{16}$$

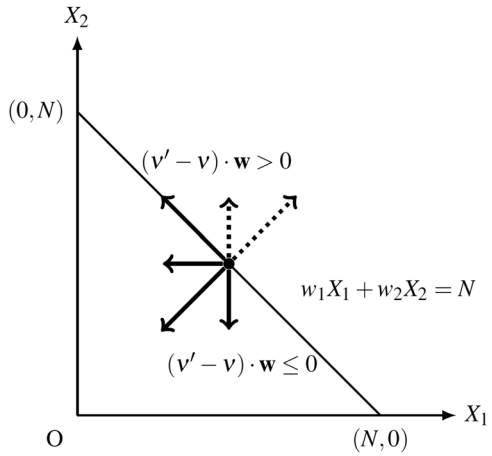


FIG. 1. The dotted arrows correspond to the reaction vectors that are turned off (i.e., the associated reaction intensity is zero) at the boundary of the state space for a slack system, as described in Sec. III C.

The network in (15) is more intrusive than the network in (16) in the sense of accessibility. At any state where $Y = 0$, the system associated with (15) will remain unchanged because no reaction can take place. However, the reaction $A \to Y$ in (16) can always occur despite $Y = 0$. Hence, (16) preserves the accessibility of the original system associated with (14) as any state for A is accessible from any other state in the original reaction system (14). We refer such a system with slack reactants generated by canceling redundant slack reactants to an *optimized slack system*. In Sec. VI, we explore the accessibility of an optimized reaction network with slack reactants in a more general setting.

Finally, we can make a network with slack reactants admit a better approximation of a given CRN by choosing an optimized conservation relation in (8). First, we assume that only a single conservation law and a single slack reactant are added to a given CRN. For the purpose of state space truncation onto finitely many states, a single conservation law is enough as all species could be bounded by N, as shown in (8). Let this single conservation law be

$$w_1X_1 + w_2X_2 + \cdots w_dX_d + Y = N.$$

Then, the matrix W defined in Sec. III B is a vector $(w_1, w_2, \ldots, w_d)^\top$, and we denote this by $\mathbf{w}$. By the definition of the intensities (10) for a network with slack reactants, some reactions are turned off when $Y = 0$, i.e., $w_1X_1 + \cdots + w_dX_d = N$. Geometrically, a reaction outgoing from the hyperplane $w_1X_1 + \cdots + w_dX_d = N$ is turned off (Fig. 1).

Hence, we optimize the estimation with slack reactants by minimizing such intrusiveness of turned off reactions. To do that, we choose $\mathbf{v}$, which minimizes the number of the reactions in $\{\nu \to \nu' \in \mathscr{R} : (\nu' - \nu) \cdot \mathbf{w} > 0\}$.

IV. COMPARISON TO OTHER TRUNCATION-BASED METHODS

In this section, we demonstrate that our method can potentially resolve limitations in calculating mean first passage times observed

in other methods of state space truncation, namely, sFSP and the finite-buffer method. Both methods require the user to make decisions about the state-space truncation that may introduce variability in the results. While all methods will converge to the true result as the size of the state space increases, we show that our method is less dependent on user-defined quantities. This minimizes additional decision-making on the part of the user that can lead to suboptimal results, especially in a context where the solution of the original system is not known.

A. Comparison to the sFSP method

A well-known state truncation algorithm is known as the *Finite State Projection* (FSP) method.[10] For a given continuous-time Markov chain, the associated FSP model is restricted to a finite state space. If the process escapes this truncated space, the state is sent to a designated absorbing state [see Fig. 2(b)]. For a fixed time t, the probability density function of the original system can be approximated by using the associated FSP model with sufficiently many states. The long-term dynamics of the original system, however, is not well approximated because the probability flow of the FSP model leaks to the designated absorbing state in the long run.

To fix this limitation of FSP, Gupta *et al.* proposed the *stationary Finite State Projection* (sFSP) method.[11] This method also projects the original state space onto a finite state space as the FSP method intended to. However, sFSP does not create a designated absorbing state as all outgoing transitions from the projected finite state space are merged to a single state x^* "inside" the finite state space [Fig. 2(c)]. The sFSP has been frequently used to estimate the long-term distribution of discrete stochastic models. However, if the size of the truncated state space is not sufficiently large, this method could fail to provide accurate estimation for the first passage time. To demonstrate this case, we consider the following simple 2-dimensional model. In the network shown in Fig. 3(a), two X_1 proteins are dimerized into protein X_2, while X_1 is being produced at a relatively high rate. The state space of the original model is the full 2-dimensional positive integer grid. We estimate the time until the system reaches one of the two states indicated in red in Fig. 3(c), and we use alternative methods to do this.

For the sFSP, we project the original state space onto the rectangle by restricting $X_1 \le N$ and $X_2 \le N$ for some $N > 0$, and we fix the origin (0, 0) as the designated state x^* [Fig. 3(c)]. If the process associated with the sFSP model escapes the rectangle, it transports to the designated state immediately. On the other hand, we also consider a slack network shown in Fig. 3(b), where we introduce the conservation law $X_1 + 2X_2 + Y = N$ for some $N > 0$. Let $\tau = \inf\{t \ge 0 : X_1 = 1$ and $X_2 \in \{1, 2\}\}$ be the first passage time we want to estimate.

For a Markov chain defined on a finite state space, the mean first passage time is computable with the inverse of an absorbing transition matrix, as detailed in Ref. 22 (see Appendix A for details). Hence, by using the inverse absorbing matrix for each truncation method, we obtain the mean of τ by using different values of N. We also obtain an "almost" true mean of τ by using 10^5 Gillespie simulations of the original process. As shown in Fig. 3(b), the slack network model provides a more accurate mean first passage time estimation for the size of truncation in between 100 and 400 if the designated return state of sFSP is (0, 0).

The inaccurate estimate from the sFSP is due to the choice of a return state. The sFSP model escapes the confined space often because the production rate of X_1 is relatively high. When it returns back to the origin, it is more likely to visit the target states $\{X_1 = 1$ and $X_2 \in \{1, 2\}\}$ than the original process.

Figure 3 shows that the mean first passage time of this system using sFSP depends significantly on the location of the chosen designated state. One of the two states is a particularly poor choice for sFSP, but it illustrates the idea that without previous knowledge of the system, it can be difficult to know which states will perform well.

We display the behavior of individual solutions of the original model, the slack network, and the sFSP model in Fig. 3(d). The trajectory plots show that within the time interval [0, 500], almost half of the 100 samples from both the original model and the slack network model stay far away from the target states, while all 100 sample trajectories from the sFSP model stay close to the target states. We also illustrate this point in Fig. 3(e) with heat maps of the three models at $t = 500$. Note that only in the case of the sFSP, the probability densities are concentrated at the target states.

B. Comparison to the finite buffer method

The finite buffer method was proposed to estimate the stationary probability landscape with state space truncation and a novel

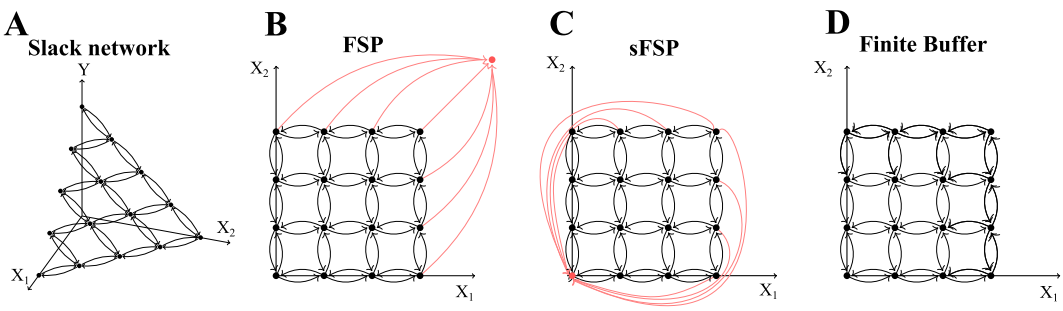


FIG. 2. Schematic images of various state-space truncation methods for a CRN with two species: (a) slack network (this paper), (b) Finite State Projection (FSP),[10] (c) stationary Finite State Projection (sFSP),[11] and (d) finite buffer method.[12,13]

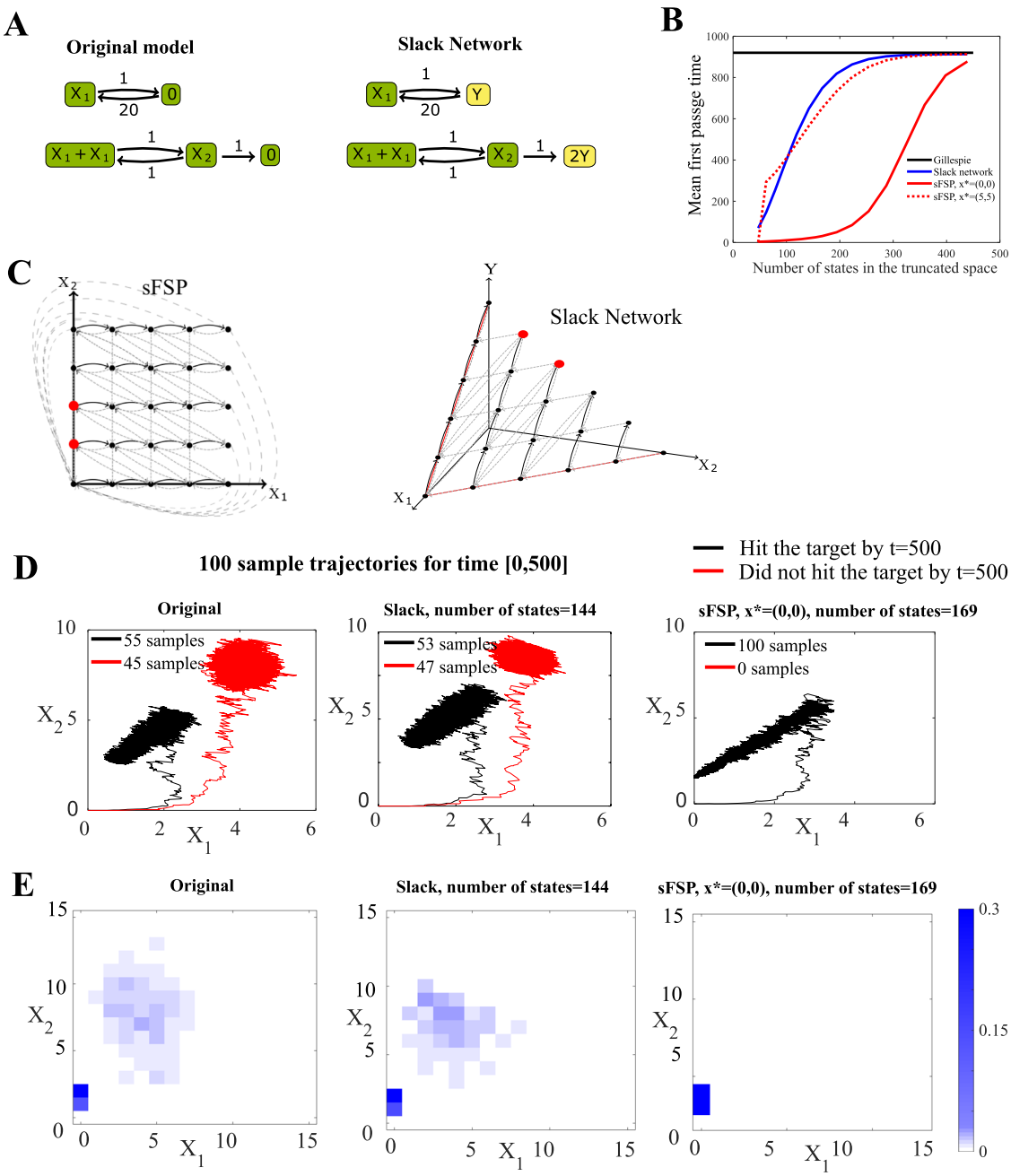


FIG. 3. Comparison between an sFSP model and a slack network for the same reaction system. (a) Original 2-dimensional system and its slack network. (b) Mean first passage time as a function of the truncated state size, comparing sFSP and the slack reactant method. (c) The state space truncation corresponding to the sFSP method and the slack method for this model. Target states for the first passage time are indicated in red. (d) Mean of 100 sample trajectories obtained by Gillespie simulations for the original system, the slack system, and the sFSP system. The red lines indicate the mean of the trajectories that have touched the target states within [0, 500], while the black lines are the mean of the trajectories that did not touch the target space during this time. (e) Probability density heat maps of the stochastic processes modeled under the original system, the slack system, and the sFSP model, respectively, at t = 500 (see Sec. IV A for more information).

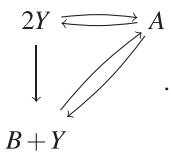
state space enumeration algorithm.[12,13] For a given stochastic model associated with a CRN, the finite buffer method sets inequalities among species such as (8), so-called buffer capacities. Then, at each state $\mathbf{x}$, the transition rate of a reaction $\nu \to \nu'$ is set to be zero if at least one of the inequalities does not hold at $\mathbf{x} + \nu' - \nu$. We note that the algorithm, described in Sec. III B, for generating a slack network uses the same inequalities. Thus, the finite buffer method and the slack reactant method truncate the state space in the same way.

We have shown, in Sec. III C, that this type of truncation can create "additional" absorbing states. These additional absorbing states change the accessibility between the states, which means the mean first passage times cannot be accurately estimated. However, the regular slack systems preserve the network structure of the original network. Hence, we are able to prove, as we already noted, that regular slack networks inherit the accessibility of the original network as long as the original network is "weakly reversible," as we will define below.

We demonstrate this disparity between the finite buffer and slack methods with the following network. Consider the mass-action system (11) with a fixed initial state $\mathbf{x}_0 = (a_0, b_0)$. We are interested in estimating the mean first passage time to a target state $\mathbf{x}_T = (10, 10)$. Note that the state space is irreducible (i.e., every state is accessible from any other state) as the network consists of unions of cycles. This condition, the union of cycles, is precisely what is meant by weakly reversible.[23,24] Thus, the original stochastic system has no absorbing state and is accessible to the target state $\mathbf{x}_T$. Therefore, it is critical for a state space truncation method not to create an additional absorbing state so that the reduced system still can reach the target state with probability 1.

To use the finite buffer method on this network, we set $2X_A + X_B \leq N$ as the buffer capacity, where $\mathbf{X} = (X_A, X_B)$ is the associated stochastic process. (Here, we choose $N > 30$, so the state space contains the target state $\mathbf{x}_T$.) Hence, when X satisfies $2X_A + X_B = N$, the reactions $\emptyset \to A$, $B \to A$, and $\emptyset \to B$ cannot be fired as $2X_A + X_B$ exceeds the buffer capacity. We now demonstrate that the system has a new absorbing state. By first using reaction $A \to B$, to deplete all A, and then $\emptyset \to B$, every state can reach the state $(0, N)$ in finite time with positive probability. The state $(0, N)$ is the absorbing state because no other reactions can occur. Reactions $A \to \emptyset$ and $A \to B$ require at least one A species, and any other reactions lead to states exceeding the buffer capacity. Therefore, the finite buffer method has introduced a new absorbing state not present in the original model so that the system can be trapped at $(0, N)$ with positive probability, and in turn, it is not accessible to $\mathbf{x}_T$. Therefore, the mean first passage time to $\mathbf{x}_T$ is infinite under the finite buffer method.

Now, we show that the explicit network structure of our slack network formulation will preserve the accessibility of the original system. We consider the same inequality $2X_A + X_B \leq N$ as above with $N > 30$. We generate the slack network by using the algorithm shown in Sec. III B,



(17)

Note that the associated stochastic process $\mathbf{X} = (X_A, X_B, Y)$ admits the conservation relation $2X_A + X_B + Y = N$, implying that $2X_A + X_B \leq N$. The state $(0, N, 0)$ cannot be reached as the only state that is accessible to $(0, N, 0)$ is $(0, N-1, 2)$, but it violates the conservation law.

As we highlighted in Sec. III C, slack networks preserve the connectivity of the original network (11); hence, network (17) is also weakly reversible. Thus, the state space of the stochastic process associated with the slack network is irreducible by Corollary VI.1. This implies that there is no absorbing state and the system is accessible to $\mathbf{x}_T$, unlike the stochastic process associated with the finite buffer relation. See Sec. VI for more details about the accessibility of slack systems.

V. CONVERGENCE THEOREMS FOR SLACK NETWORKS

In this section, we establish theoretical results on the convergence of properties of a slack network to the original network. (Proofs of the theorems below are provided in Appendix B.) Many of these results rely on theorems from Hart and Tweedie[25] who studied when the probability density function of a truncated Markov process converges to that of the original Markov process. We employ the same idea of their proof to show the convergence of a slack network to the original network.

By assuming "uniform mixing," we show the convergence of the stationary distribution of the slack system to the stationary distribution of the original system as the conservation quantity N grows. Furthermore, we show the convergence of mean first passage times for the slack network to the true mean first passage times. In particular, all these conditions hold when there is a Lyapunov function for the original system.

In this section, assume that a given CRN $(\mathscr{S}, \mathscr{C}, \mathscr{R}, \Lambda)$ is well defined for all time t and let $(\widetilde{\mathscr{S}}, \widetilde{\mathscr{C}}, \widetilde{\mathscr{R}}, \Lambda^N)$ be an associated slack network obtained with a single conservation quantity $\mathbf{w} \cdot \mathbf{X} \leq N$. We denote by X and $\mathbf{X}^N$ the associated stochastic processes of the original CRN and the slack network, respectively. We fix the initial state for both systems, i.e., $\mathbf{X}(0) = \mathbf{X}^N(0) = \mathbf{x}_0$ for some $\mathbf{x}_0$ and for each N. [This means that we can only consider slack systems where N is large enough so that $\mathbf{w} \cdot \mathbf{x}(0) < N$.] Assume that both the original and slack systems are irreducible, and we denote by $\mathbb{S}$ and $\mathbb{S}_N$ the state spaces for each, respectively. (In Sec. VI, we prove accessibility properties that the slack system can inherit from the original system.)

Note that every state in $\mathbb{S}_N$ satisfies our conservation inequality. That is, for every $\mathbf{x} \in \mathbb{S}_N$, we have $\mathbf{w} \cdot \mathbf{x} \leq N$. It is possible that $\mathbb{S}_N = \mathbb{S}_{N+1}$ for some N. For simplicity, we assume that the truncated state space is always enlarged with respect to the conservation quantity N, that is, $\mathbb{S}_N \subset \mathbb{S}_{N+1}$ for each N. (For the general case, we could simply consider a subsequence N_k such that $N_k < N_{k+1}$ and $\mathbb{S}_{N_k} \subset \mathbb{S}_{N_{k+1}}$ for each k.)

As defined in Sec. II, $\lambda_{\nu \to \nu'} \in \Lambda$ is the intensity of a reaction $\nu \to \nu'$ for the associated stochastic system X. We also denote by $\lambda^N_{\tilde{\nu} \to \tilde{\nu}'} \in \Lambda^N$ the intensity of a reaction $\tilde{\nu} \to \tilde{\nu}'$ for the associated stochastic system $\mathbf{X}^N$. Finally, we let $p(\mathbf{x}, t)$ and $p_N(\mathbf{x}, t)$ be the probability density function of X and $\mathbf{X}^N$, respectively. We begin with the convergence of the probability density functions of the slack network to the original network with increasing N.

Theorem V.1. *For any $\mathbf{x} \in \mathbb{S}_N$ and $T \geq 0$, we have*

$$\lim_{N \to \infty} \sup_{t \in [0,T]} |p(\mathbf{x}, t) - p_N(\mathbf{x}, t)| = 0.$$

A Markov process defined on a finite state space admits a stationary distribution. Hence, $\mathbf{X}^N$ admits a stationary distribution π_N. If the slack system satisfies the condition of "uniform mixing, that is,

the convergence rate of $\|p_N(\mathbf{x}, t) - \pi_N(\mathbf{x})\|_1$ is uniform in N, then we have the following result.

Theorem V.2. *Suppose that* $\mathbf{X}$ *admits a stationary distribution* π. *Suppose further that there exists a positive function* $h(t)$, *which is independent of* N, *such that* $\|p_N(\cdot, t) - \pi_N\|_1 \le h(t)$ *and* $\lim_{t\to\infty} h(t) = 0$. *Then,*

$$\lim_{N\to\infty} \|\pi - \pi_N\|_1 = 0.$$

We now consider the convergence of the mean first passage time of $\mathbf{X}^N$. Recall that we assumed that both stochastic processes have the same initial state $\mathbf{X}(0) = \mathbf{X}^N(0) = \mathbf{x}_0$ and both state spaces $\mathbb{S}$ and $\mathbb{S}_N$ are irreducible. Hence, for any $K \subseteq \mathbb{S}$, each state in $\mathbb{S}_N$ is accessible to K for sufficiently large N.

Theorem V.3. *For a subset* K *of the state space of* $\mathbf{X}$, *let* τ *and* τ_N *be the first passage times to* K *for* $\mathbf{X}$ *and to* $K \cap \mathbb{S}_N$ *for* $\mathbf{X}^N$, *respectively. Assume the following conditions:*

1. $\mathbf{X}$ *admits a stationary distribution* π.
2. $\lim_{N\to\infty} \|\pi - \pi_N\|_1 = 0$.

Then, for any $T \ge 0$,

$$\lim_{N\to\infty} \sup_{t\in[0,T]} |P(\tau > t) - P(\tau_N > t)| = 0.$$

If we further assume that

3. $E(\tau) < \infty$.
4. *There exists* $g(t)$ *such that* $P(\tau_N > t) \le g(t)$ *for all* N *and* $\int_0^\infty g(t)dt < \infty$.

Then,

$$\lim_{N\to\infty} |E(\tau) - E(\tau_N)| = 0.$$

Remark V.1. To obtain the convergence of higher moments of the first passage time, we need only replace conditions $E(\tau_K) < \infty$ *and* $\int_0^\infty g(t)dt < \infty$ *with*

$$E(\tau_K^n) < \infty \quad \text{and} \quad \int_0^\infty g\left(t^{\frac{1}{n}}\right)dt < \infty, \tag{18}$$

respectively.

We now show that if a Lyapunov function exists for the original system, the conditions in Theorem V.2 and Theorem V.3 hold. The Lyapunov function approach was proposed by Meyn and Tweedie,[26] and it has been used to study long-term dynamics of Markov processes,[11,27–29] especially exponential ergodicity. Gupta *et al.*[11] used a linear Lyapunov function to show that the stationary distribution of an sFSP model converges to a stationary distribution of the original stochastic model and used the Lyapunov function to explicitly compute the convergence rate. In particular, we show Lyapunov functions exist for the examples we consider in Sec. VII.

Theorem V.4. *Suppose that there exists a function* V *and positive constants* C *and* D *such that for all* $\mathbf{x}$,

1. $V(\mathbf{x}) \ge 1$ *for all* $\mathbf{x} \in \mathbb{S}$,
2. V *is an increasing function in the sense that*

$$V(\mathbf{x}_{N+1}) \ge V(\mathbf{x}_N)$$

for each $\mathbf{x}_{N+1} \in \mathbb{S}_{N+1}\backslash\mathbb{S}_N$ *and* $\mathbf{x}_N \in \mathbb{S}_N$, *and*

3. $$\sum_{\nu\to\nu'\in\mathscr{R}} \lambda_{\nu\to\nu'}(\mathbf{x})(V(\mathbf{x} + \nu' - \nu) - V(\mathbf{x})) \le -CV(\mathbf{x}) + D.$$

Then, the conditions in Theorem V.3 hold.

Remark V.2. Conditions (18) hold if a Lyapunov function satisfying the conditions in Theorem V.4 exists. Thus, the convergence of the higher moments of the first passage time also follows.

VI. INHERITANCE OF SLACK NETWORKS

As we showed in Sec. IV B, not all state space truncations preserve accessibility of states in the original system. (For the example in Sec. IV B, the truncation created a new absorbing state.) Thus, it is desirable to obtain reduced models that are guaranteed to maintain the accessibility of the original system to predetermined target states. In this section, we show that under mild conditions, both a regular slack system and an optimized slack system preserve the original accessibility. The proofs of the theorems introduced in this section are in Appendix C. The key to these results is the condition of weak reversibility.

Definition VI.1. A reaction network is ***weakly reversible*** *if each connected component of the network is strongly connected. That is, if there is a path of reactions from a complex* ν *to* ν', *then there is a path of reactions* ν' *to* ν.

We note that the weakly reversible condition applies to the network graph of the CRN. The network graph consists of complexes (nodes) and reactions (edges). It is a sufficient condition for irreducibility of the associated mass-action stochastic process. Indeed, the sufficiency of weak reversibility holds even under general kinetics as long as condition (4) is satisfied.[30] Hence, irreducibility of a regular slack network follows since it preserves weak reversibility of the original network, and the kinetics modeling the regular slack system satisfies (4).

Corollary VI.1. Let $(\mathscr{S}, \mathscr{C}, \mathscr{R}, \Lambda)$ *be a weakly reversible CRN with intensity functions* $\Lambda = \{\lambda_{\nu\to\nu'}\}$ *satisfying (4). Then, the state space of the associated stochastic process with a regular slack network* $(\widetilde{\mathscr{S}}, \widetilde{\mathscr{C}}, \widetilde{\mathscr{R}}, \Lambda^N)$ *is a union of closed communication classes for any* N.

In case the original network is not weakly reversible, we can still guarantee that optimized slack systems have the same accessibility as the original system, provided that all species have a degradation reaction $(S_i \to \emptyset)$.

Theorem VI.1. *Let* $(\mathscr{S}, \mathscr{C}, \mathscr{R}, \Lambda)$ *be a reaction network such that* $\{S_i \to \emptyset : S_i \in \mathscr{S}\} \subset \mathscr{R}$. *Suppose that the stochastic process* X *associated with* $(\mathscr{S}, \mathscr{C}, \mathscr{R}, \Lambda)$ *and beginning at the point* $\mathbf{x}_0$ *is irreducible. Let* $\mathbf{X}^N$ *be the stochastic process associated with an optimized slack network* $(\widetilde{\mathscr{S}}, \widetilde{\mathscr{C}}, \widetilde{\mathscr{R}}, \Lambda^N)$ *such that* $\mathbf{X}^N(0) = \mathbf{x}_0$ *for every* N *large*

enough. Then, for any subset K of the state space of $\mathbf{X}$*, there exists* N_0 *such that* $\mathbf{X}^N$ *reaches K almost surely for* $N \geq N_0$.

This theorem follows from the fact that a slack system only differs from the original system when it runs out of slack reactants. However, in an optimized slack system, degradation reactions are allowable with no slack reactants. Hence, our proof of Theorem VI.1 relies on the presence of all degradation reactions.

A slack network may also inherit its stationary distribution from the original reaction system. When the original system admits a stationary distribution of a product form of Poisson distributions under the *complex balance* condition, a slack system inherits the same form of the stationary distribution as well. A reaction system is complex balanced if the associated deterministic mass-action system admits a steady state c^* such that

$$\sum_{\substack{v\in\mathscr{C}\\ v\to v'\in\mathscr{R}}} f_v(\mathbf{c}^*) = \sum_{\substack{v'\in\mathscr{C}\\ v\to v'\in\mathscr{R}}} f_{v'}(\mathbf{c}^*),$$

where $f_v(x) = \kappa_{v\to v'} x_1^{v_1}\cdots x_d^{v_d}$ is the deterministic mass-action rate[31] with a rate constant $\kappa_{v\to v'}$. If a reaction system is complex balanced, then its associated stochastic mass-action system admits a stationary distribution corresponding to a product of Poisson distributions centered at the complex balance steady state.[32] The following lemma shows that the complex balancing of the original network is inherited by a regular slack network.

Lemma VI.1. Suppose that $(\mathscr{S},\mathscr{C},\mathscr{R},\Lambda)$ *is a reaction network whose mass-action deterministic model admits a complex balanced steady state* c^**. Then, any regular slack network* $(\widetilde{\mathscr{S}},\widetilde{\mathscr{C}},\widetilde{\mathscr{R}},\Lambda^N)$ *with slack reactants* $Y_1, \ldots, Y_m$ *also admits a complex balanced steady state at* $\tilde{c} = (c^*, 1, 1, \ldots, 1)^\top$.

Remark VI.1. Note that a regular slack network also preserves the deficiency of the original network. Deficiency δ *of a reaction network is an index such that*

$$\delta = n - \ell - s,$$

where n is the number of the complexes, ℓ *is the number of connected components, and s is the rank of the stoichiometric matrix of the reaction network. Deficiency characterizes the connectivity of the network structure, and surprisingly, it can also determine the long-term behavior of the system dynamics regardless of the system parameters.*[31–33] *A regular slack network and original network have the same number of complexes n and the same connectivity matrix S, which implies they have the same number of connected components* ℓ*. Furthermore, using the notation from* Sec. III B*, the stoichiometric matrices are* $\Gamma = CS$ *for the original network and*

$$\widetilde{\Gamma} = \begin{pmatrix} C \\ U - AC \end{pmatrix} S = \begin{pmatrix} CS \\ -WCS \end{pmatrix}$$

for a slack network, which means that they have the same rank s. Together, these imply that the original network and its regular slack network have the same deficiency.

Since the complex balancing is inherited with the same steady state values for X_i, we have the following stochastic analog of inheritance of the Poissonian stationary distribution for regular slack systems.

Theorem VI.2. *Let* $\mathbf{X}$ *be the stochastic mass-action system associated with a complex balanced* $(\mathscr{S},\mathscr{C},\mathscr{R},\Lambda)$ *with an initial condition* $\mathbf{X}(0) = \mathbf{x}_0$*. Let* $\mathbf{X}^N$ *be the stochastic system associated with a regular slack system* $(\widetilde{\mathscr{S}},\widetilde{\mathscr{C}},\widetilde{\mathscr{R}},\Lambda^N)$ *with* $\mathbf{X}^N(0) = \mathbf{x}_0$*. Then, for the state space* $\mathbb{S}_N$ *of* $\mathbf{X}^N$*, there exists a constant* $M_N > 0$ *such that*

$$\pi_N(\mathbf{x}) = M_N \pi(\mathbf{x}) \quad \text{for } \mathbf{x} \in \mathbb{S}_N,$$

where π *and* π_N *are the stationary distributions of* $\mathbf{X}$ *and* $\mathbf{X}^N$*, respectively.*

We demonstrate Lemma VI.1 and Theorem VI.2 with a simple example.

Example VI.1. Consider two networks,

$$X \underset{2}{\overset{1}{\rightleftharpoons}} 0, \tag{19}$$

$$X \underset{2}{\overset{1}{\rightleftharpoons}} Y. \tag{20}$$

Let X and X^N be systems (19) and (20), respectively, where N is the conservation quantity $X^N(t) \leq N$. Under mass-action kinetics, the complex balance steady state of (19) is $c^* = 2$. Under mass-action kinetics, $\tilde{c} = (2, 1)$ is a complex balance steady state of (20).

Now, let π and π_N be the stationary distribution of X and X^N, respectively. By Theorem 6.4, Ref. 32, π is a product form of Poisson distributions such that

$$\pi(x) = e^{-2}\frac{c^{*x}}{x!} \quad \text{for each state } x.$$

By plugging π into the chemical master equation (3) of X^N and showing that for each $\mathbf{x}$

$$\begin{aligned}
&\lambda^N_{X\to Y}(\mathbf{x}+1)\pi(\mathbf{x}+1) + \lambda^N_{Y\to X}(\mathbf{x}-1)\pi(\mathbf{x}-1)\\
&\quad = (x+1)\mathbb{1}_{\{N-x-1\geq 0\}} e^{-2}\frac{2^{x+1}}{(x+1)!} + 2\mathbb{1}_{\{N-x+1\geq 1\}} e^{-2}\frac{2^{x-1}}{(x-1)!}\\
&\quad = x\mathbb{1}_{\{N-x\geq 0\}} e^{-2}\frac{2^x}{x!} + 2\mathbb{1}_{\{N-x\geq 1\}} e^{-2}\frac{2^x}{x!}\\
&\quad = \lambda^N_{X\to Y}(\mathbf{x})\pi(\mathbf{x}) + \lambda^N_{Y\to X}(\mathbf{x})\pi(\mathbf{x}),
\end{aligned}$$

we can verify that π is a stationary solution of the chemical master equation (3) of X^N. Since the state space of X^N is $\{x \in \mathbb{Z}_{\geq 0} : x \leq N\}$, we choose a constant M_N such that

$$\sum_{0\leq x\leq N} M_N \pi(x) = 1.$$

Then, $\pi_N = M_N\pi$ is the stationary distribution of X^N.

VII. APPLICATIONS OF SLACK NETWORKS

In this section, we demonstrate the utility of slack reactants in computing mean first-passage times for two biological examples.

For both examples, we compute the mean first passage time via the matrix inversion approach, as shown in Appendix A.

A. A Lotka–Volterra model with migration

Consider a Lotka–Volterra model with migration shown in Fig. 4(a). In this model, species A is the prey, and species B is the predator. Clearly, the state space of this model is infinite (A, B) such that $A \geq 0$, $B \geq 0$. We will use slack reactants to determine the expected time to extinction of either species. More specifically, let $K = \{(A, B) : A = 0 \text{ or } B = 0\}$. We will calculate the mean first arrival time to K from an initial condition $[A(0), B(0)]$. [In our simulations in Fig. 4, we chose $(A(0), B(0)) = (3, 3)$.]

To generate our slack network, we choose a conservation bound $w \cdot (A, B)^\top \leq N$ with $w = (1, 1)$. As we discussed in Sec. III C, this w minimizes the intrusiveness of slack reactants because the number of reactions $\nu \to \nu'$ such that $(\nu' - \nu) \cdot w > 0$ is minimized. By using the algorithm introduced in Sec. III B, we generate a regular slack network (21) with a slack reactant Y,

$$\begin{aligned} &B + Y \underset{\kappa_2}{\overset{\kappa_1}{\rightleftharpoons}} 2Y \underset{\kappa_4}{\overset{\kappa_3}{\rightleftharpoons}} A + Y \xrightarrow{\kappa_5} 2A, \\ &A + B \xrightarrow{\kappa_6} 2B. \end{aligned} \tag{21}$$

As the slack reactant Y in reactions $B + Y \rightleftharpoons 2Y \rightleftharpoons A + Y$ can be canceled, we further generate the optimized slack network shown in Fig. 4(a). We let $A(0) + B(0) + Y(0) = N$, which is the conservation quantity of the new network.

Let τ be the first passage time from our initial condition to K. First, we examine the accessibility of the set K. Because our reaction network contains creation and destruction of all species (i.e., $B \rightleftharpoons \emptyset \rightleftharpoons A$), the original model is irreducible and any state is accessible to K. Furthermore, Theorem VI.1 guarantees that the stochastic model associated with the optimized slack network is also accessible to K from any state.

Next, by showing there exists a Lyapunov function satisfying the condition of Theorem V.4 for the original model, we are guaranteed that the first passage times from our slack network will converge to the true first passage times (see Appendix E 1 for more details). Therefore, as the plot shows in Fig. 4(b), the mean first extinction time of the slack network converges to that of the original model as N increases. The mean first passage time of the original model was obtained by averaging 10^9 sample trajectories. These trajectories were computed serially on a commodity machine and took 4.6 h to run. In contrast, the mean first passage times of the slack systems were computed analytically on the same computer and took at most 13 s. Figure 4 also shows that using only 10^3 samples is misleading as the simulation average has not yet converged to the true mean first passage time. Finally, as expected from Theorem V.1, the probability density of the slack network converges to that of the original network [see Fig. 4(c)].

B. Protein synthesis with a slow toggle switch

We now consider a protein synthesis model with a toggle switch [see Fig. 5(a)]. Protein species X and Z may be created but only when

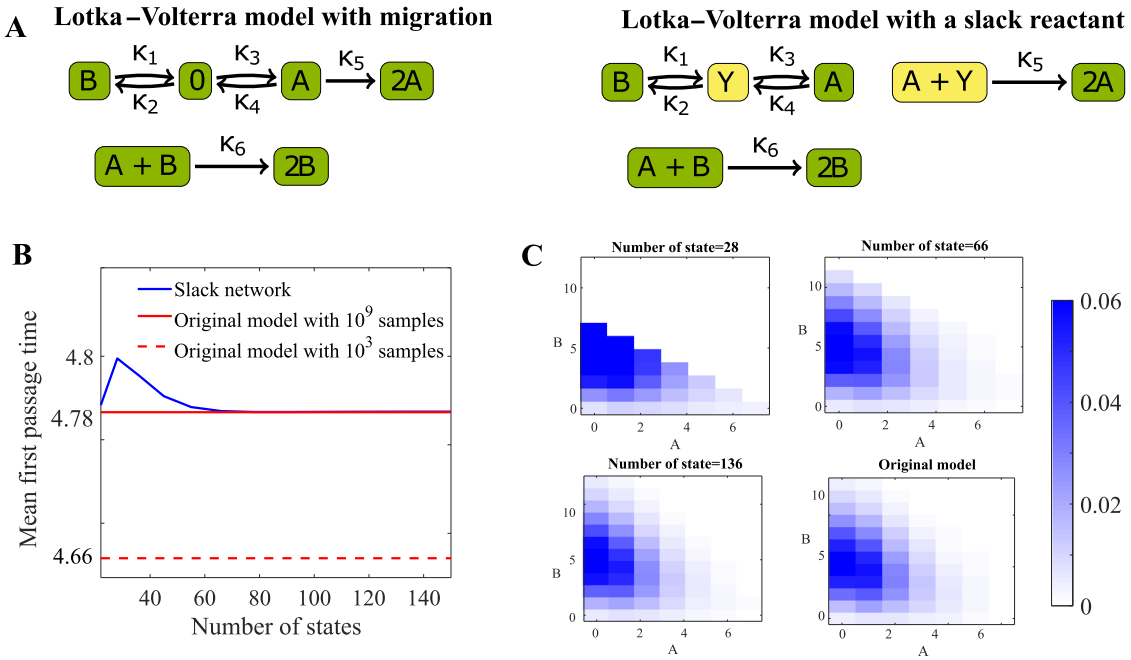


FIG. 4. Calculating mean time to extinction of a Lotka–Volterra model with migration using slack reactants. (a) The reaction network for the Lotka–Volterra model with migration (left). The Lotka–Volterra model with a slack reactant Y (right). The parameters are $\kappa_1 = 0.1$, $\kappa_2 = 0.1$, $\kappa_3 = 0.2$, $\kappa_4 = 0.6$, $\kappa_5 = 0.2$, and $\kappa_6 = 0.2$. (b) Convergence of the mean time to extinction of the slack network (blue) to the true network (solid red). (c) Heatmaps of the probability density at time 1000 of the original model and the slack system with various truncation size (see Sec. VII A for details).

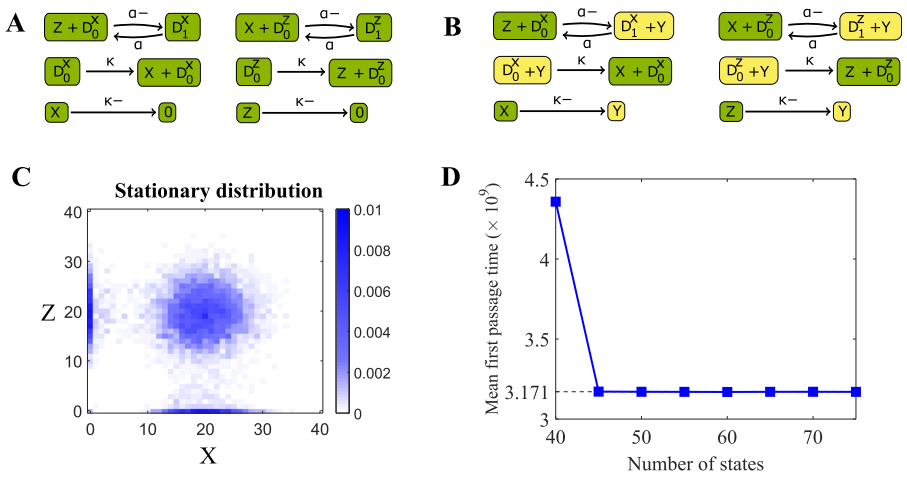


FIG. 5. Calculating mean first passage time of a slow toggle switch using slack reactants. (a) Protein synthesis with slow toggle switch. (b) The toggle switch model with a slack reactant Y. Parameters are $\alpha- = 0.1$, $\alpha = 10$, $\kappa = 2000$, and $\kappa- = 100$. (c) Multimodal long-term probability distribution of the original model. (d) Convergence of the first passage time of the slack system with various truncation sizes (see Sec. VII B for details).

their respective genes D^X or D^Z are in the active (unoccupied) state, D_0^X and D_0^Z. Each protein acts as a repressor for the other by binding at the promoter of the opposite gene and forcing it into the inactive (occupied) state (D_1^X and D_1^Z). In this system, we consider only one copy of each gene so that $D_0^X + D_1^X = D_0^Z + D_1^Z = 1$ for all time. Thus, we focus primarily on the state space of protein numbers only (X, Z).

The deterministic form of such systems is often referred to as a "bi-stable switch" as it is characterized by steady states $(X^*, 0)$ (X "on" and Z "off") and $(0, Z^*)$ (X "off" and Z "on"). This stochastic form of toggle switch has been shown to exhibit a multi-modal long-term probability density due to switches between these two deterministic states due to rapid significant changes in the numbers of proteins X and Z by synthesis or degradation (depending on the state of promoters).[34] Figure 5(c) shows that the associated stochastic system admits a *tri-modal* long-term probability density. Thus, the system transitions from one mode to other modes and, for the kinetic parameters chosen in Fig. 5, rarely leaves the region $R = \{(X, Z)|0 \le X \le 30 \text{ or } 0 \le Z \le 30\}$. Significant departures from a stable region of a genetic switch may be associated with irregular and diseased cell fates. As such, the first passage time of this system outside of R may indicate the appearance of an unexpected phenotype in a population. Because this event is rare, estimating first passage times with direct stochastic simulations, such as with the Gillespie algorithm,[6] will be complicated by the long time taken to exit the region.

As in the previous example, slack systems provide a valuable tool for direct calculation of mean first passage times. In this example, we consider the time a trajectory beginning at state $(X, Z, D_0^X, D_0^Z) = (0, 0, 1, 1)$ enters the target set $K = \{(X, Z)|X > 30$ and $Z > 30\} = R^c$ and compute τ, the first passage time to K.

Since the species corresponding to the status of promoters $(D_0^X, D_1^X, D_0^Z$ and $D_1^Z)$ are already bounded, we use the conservation bound $X + Z \le N$ to generate a regular slack network in Fig. 5(b) with the algorithm introduced in Sec. III B. The original toggle switch model is irreducible (because of the degradation $X \to 0$, $Z \to 0$ and protein synthesis $Z + D_0^X \leftarrow D_1^X$, $X + D_0^Z \leftarrow D_1^Z$ reactions). Moreover, by Theorem VI.1, the degradation reactions guarantee that the slack system is also accessible to K from any state.

As shown in Fig. 5(d), the mean first passage time of the slack system appears to be converging to approximately 3.171×10^9. To prove that the limit of the slack system is actually the original mean first passage time, we construct a Lyapunov function satisfying the conditions of Theorem V.4. See Appendix E 2 for more details about the construction of the Lyapunov function.

C. The exclusive mutual inhibition, self-activation model

The exclusive Mutual Inhibition, Self-Activation (ExMISA) model is a cell-fate decision model that contains two genes A and B, each of which produces transcription factors X and Z.[35,36] The transcription factors each bind to their own gene, promoting their own production, as well as binding to the other gene, inhibit the production of the other transcription factor. We consider the ExMISA model described by the following reaction network:

$$\begin{aligned}
&A_{00} + 2X \underset{f_a}{\overset{h_a}{\rightleftharpoons}} A_{10}, \quad A_{00} + 2Z \underset{f_r}{\overset{h_r}{\rightleftharpoons}} A_{01},\\
&B_{00} + 2Z \underset{f_a}{\overset{h_a}{\rightleftharpoons}} B_{10}, \quad B_{00} + 2X \underset{f_r}{\overset{h_r}{\rightleftharpoons}} B_{01},\\
&A_{00} \xrightarrow{g_0} A_{00} + X, \quad B_{00} \xrightarrow{g_0} B_{00} + Z, \quad X \xrightarrow{k} \emptyset, \qquad (22)\\
&A_{01} \xrightarrow{g_0} A_{01} + X, \quad B_{01} \xrightarrow{g_0} B_{00} + Z, \quad Z \xrightarrow{k} \emptyset,\\
&A_{10} \xrightarrow{g_1} A_{10} + X, \quad B_{10} \xrightarrow{g_1} B_{10} + Z,
\end{aligned}$$

where $g_0 < g_1$ so that the rate of production of X and Z is highest when genes A and B are in the "on-state," A_{10} and B_{10}, respectively. The transcription factor binding/unbinding rates f_a, h_a, f_r, and h_r are much smaller than the rates of protein production/degradation. (In our simulations, we follow the previous paper[36] and set $f_a = 10^{-5}$, $f_r = 1$, $h_a = 10^{-5}$, $h_r = 10^{-1}$, $g_0 = 4$, $g_1 = 18$, and $k = 1$.) Hence, the gene state switching reactions are operating in the slow-time scale,

while the protein production/degradation reactions are operating in the fast-time scale.

Similar to the toggle switch model we introduced in Sec. VII B,[35] these distinct scales create multiple modes in the long-term system behavior. However, because X and Z can still be produced when A and B are in the "off-state," the behavior of the ExMISA model is more complicated. The system admits four modes in the long-term probability landscape, each of which corresponds to a phenotype of cell [see Fig. 6(b)]. Once the system settles down at one mode, it is extremely rare to transition to other mode because of the slow-time scale.[36,37]

Margaret *et al.*[36] used a rare-event based simulation method to study metastable gene regulatory systems, including ExMISA, and approximated the long-time probability landscape and time between phenotype-transitions.

In this section, we study the ExMISA model with slack reactants. First, we construct the slack network in (23) with the slack reactant Y. We then use the new network to estimate long-term probability distribution and approximate the mean first transition time between two phenotype modes of the original ExMISA model. Note that the probability density can be explicitly derived for a Markov chain with a finite-state space as $p(t) = \mu e^{Qt}$, where Q is the transition rate matrix and μ is the initial distribution. Hence, we compute the long-time probability density function for the slack network (23) [Fig. 6(b)]. The four peaks in the probability density correspond to high/high, high/low, low/high, and low/low statuses of X/Z. Then, we estimate the mean transition time from the low/high status to high/low status and show the convergence of the mean transition time in Fig. 6(c). (More specifically, we begin a trajectory at the state $X = 5$, $Z = 15$, $A_{00} = 0$, $A_{01} = 1$, $A_{10} = 0$, $B_{00} = 0$, $B_{01} = 0$, $B_{10} = 1$ and calculate the time it reaches the unit ball centered at $X = 15$, $Z = 5$.) The mean first passage time of the original ExMISA was estimated by averaging 5×10^3 sample trajectories obtained using the Gillespie algorithm. A commodity machine was used to simulate these trajectories in parallel (`parfor` in Matlab with six workers) and took 3.52 h to run. In contrast, on the same computer, our slack system (23) took only 6 min to compute the mean first passage time for each conservation quantity N,

$$
\begin{aligned}
&A_{00}+2X \underset{f_a}{\overset{h_a}{\rightleftharpoons}} A_{10}+2Y, \quad A_{00}+2Z \underset{f_r}{\overset{h_r}{\rightleftharpoons}} A_{01}+2Y,\\
&B_{00}+2Z \underset{f_a}{\overset{h_a}{\rightleftharpoons}} B_{10}+2Y, \quad B_{00}+2X \underset{f_r}{\overset{h_r}{\rightleftharpoons}} B_{01}+2Y,\\
&A_{00}+Y \xrightarrow{g_0} A_{00}+X, \quad B_{00}+Y \xrightarrow{g_0} B_{00}+Z, \quad X \xrightarrow{k} Y, \qquad (23)\\
&A_{01}+Y \xrightarrow{g_0} A_{01}+X, \quad B_{01}+Y \xrightarrow{g_0} B_{00}+Z, \quad Z \xrightarrow{k} Y,\\
&A_{10}+Y \xrightarrow{g_1} A_{10}+X, \quad B_{10}+Y \xrightarrow{g_1} B_{10}+Z.
\end{aligned}
$$

VIII. DISCUSSION AND CONCLUSIONS

We propose a new state space truncation method for stochastic reaction networks (see Sec. III). In contrast to other methods, such as FSP, sFSP, and finite buffer methods, we truncate the state space indirectly by expanding the original chemical reaction network to include slack reactants. The truncation is imposed through user-defined conservation relations among species. The explicit network structure of slack reactants facilitates proofs of convergence (Sec. V) and allows the use of existing software packages to study the slack network itself.[17] Indeed, any user-defined choices for conservation laws, conservation amounts, and stoichiometric structure can be used to construct a slack network with our algorithm. We provide guidelines for optimal user choices that can increase the similarity between the slack system and the original model (see Sec. III C).

Slack systems can be used to estimate the dynamical behavior of the original stochastic model. In Sec. IV, we used a simple example to show that the slack method can lead to a better approximation for the mean first passage time than the sFSP method and the finite buffer method. In particular, in Sec. V, we provide theorems that show that the slack system approximates the probability density and the mean first passage time of the original system. Because slack networks preserve network properties, such as weak reversibility, the slack system is also likely to have the same accessibility to a target state as the original model (see Sec. VI). In particular, we note

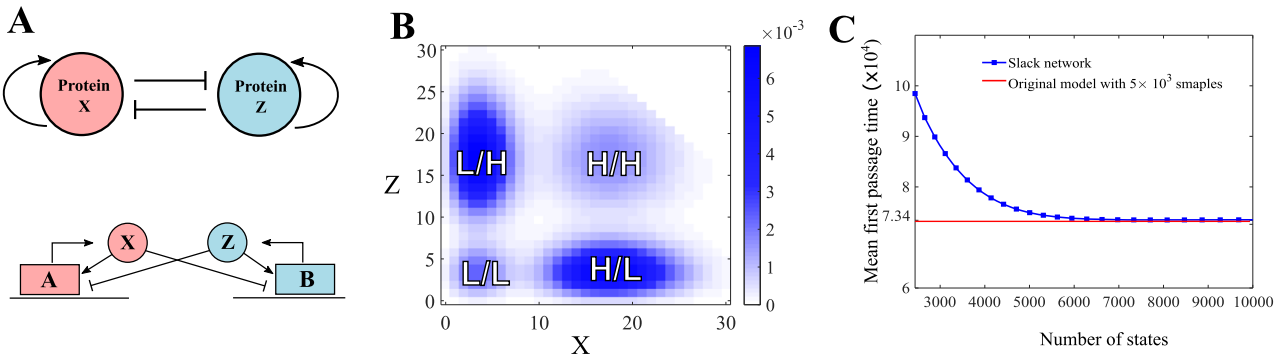


FIG. 6. Calculating mean first passage time of ExMISA. (a) Schematic description of ExMISA. (b) Multimodal long-term probability distribution of the ExMISA model with the slack network (23). The modes are labeled by status (H = high, L = low). (c) Approximating the mean transition time from the low/high mode to the high/low mode under a slack system with different truncation size (blue curve) and by time averages from 5000 direct Gillespie simulations (red line) (see Sec. VII C for more details).

that weak reversibility guarantees that our slack truncation does not introduce absorbing states.

In Sec. VI, we show that this truncation method is natural in the sense that the stationary distributions of the original and slack systems are identical up to multiplication by a constant when the original system is complex balanced. Finally, in Sec. VII, we use slack networks to calculate first passage times for two biological examples. Our method can be useful to study various biological systems for which we estimate rare event probabilities and mean first passage times. We expect that this new theoretical framework for state space truncation will be useful in the study of other biologically motivated stochastic chemical reaction systems.

AUTHORS' CONTRIBUTIONS

G.E. and S.S. contributed equally to this work.

ACKNOWLEDGMENTS

This work was partially supported by the NSF (Grant Nos. DMS1616233, DMS1763272, and BCS1344279), by the Simons Foundation (Grant No. 594598), and then by the Joint DMS/NIGMS Initiative to Support Research at the Interface of the Biological and Mathematical Sciences (Grant No. R01-GM126548).

APPENDIX A: FIRST PASSAGE TIME FOR MARKOV PROCESSES WITH FINITELY MANY STATES

The Markov chain associated with a slack network has always a finite state space. There are many different methods to analytically derive the mean first passage time of a Markov chain with a finite state space.[22,38,39] In this paper, we use the method of Laplace transform, which is also used in Ref. 39.

For a continuous time Markov process defined on a finite state space $\mathbb{S}$, let Q be the transition rate matrix, i.e., Q_{ij} is the transition rate from state i to state j if $i \neq j$ and $Q_{ii} = -\sum_{j\neq i} Q_{ij}$.

For a subset $K = \{i_1, i_2, \ldots, i_k\} \subset \mathbb{S}$, we define an *absorbing transition matrix* Q_K that is obtained by removing the i_jth row and column from Q for $j = 1, 2, \ldots, k$. Then, the mean first passage time to the set K starting from the ith state is the ith entry of $-Q_K^{-1}\mathbf{1}$, where $\mathbf{1}$ is a column vector with each entry 1.

APPENDIX B: PROOFS OF CONVERGENCE THEOREMS

In this section, we provide the proofs of the theorems introduced in Sec. V. We use the same notations and the same assumptions as we used in Sec. V. We also use a projection function q such that

$$q(x_1, \ldots, x_d, y_1, \ldots, y_r) = (x_1, \ldots, x_d). \tag{B1}$$

This linear function q projects a complex in a slack network onto a complex in the original network as $q(\tilde{\nu}) = \nu$ if $\tilde{\nu}$ is obtained by adding slack reactants to ν. In the same sense, $q(\tilde{\nu}' - \tilde{\nu}) = \nu' - \nu$ if a reaction $\tilde{\nu} \to \tilde{\nu}'$ is obtained from a reaction $\nu \to \nu'$ by adding slack reactants.

Proof of Theorem V.1. We employ the main idea shown in the proof of Theorem 2.1 in Ref. 25. Let a state $\mathbf{x}$ and time t be fixed. We consider large enough N so that $\mathbf{x} \in \mathbb{S}_{N-1}$.

We use an FSP model on $\mathbb{S}_{N-1}$ of the original system X with the designated absorbing state $\mathbf{x}^* \in \mathbb{S}_{N-1}^c$. Let p_{N-1}^{FSP} be the probability density function of this FSP model.

Let T_N be the first time for $\mathbf{X}^N$ to hit $\mathbb{S}_N \backslash \mathbb{S}_{N-1}$. We generate a coupling of $\mathbf{X}^N$ and the FSP model restricted on $\mathbb{S}_{N-1}$ as they move together by T_N and they move independently after T_N. Then, $p_{N-1}^{\mathrm{FSP}}(\mathbf{x}, t) = P(\mathbf{X}^N(t) = \mathbf{x}, T_N > t)$ for $\mathbf{x} \in \mathbb{S}_{N-1}$ because the FSP model has stayed in $\mathbb{S}_{N-1}$ if and only if $\mathbf{X}^N$ has never touched $\mathbb{S}_N \backslash \mathbb{S}_{N-1}$. Thus,

$$\begin{aligned} p_N(\mathbf{x}, t) &= P(\mathbf{X}^N(t) = \mathbf{x}, t < T_N) + P(\mathbf{X}^N(t) = \mathbf{x}, t \geq T_N) \\ &= p_{N-1}^{\mathrm{FSP}}(\mathbf{x}, t) + P(\mathbf{X}^N(t) = \mathbf{x}, t \geq T_N) \\ &\geq p_{N-1}^{\mathrm{FSP}}(\mathbf{x}, t). \end{aligned} \tag{B2}$$

Furthermore,

$$\begin{aligned} P\big(\mathbf{X}^N(t) = \mathbf{x}, t \geq T_N\big) &\leq P(t \geq T_N) \\ &= p_{N-1}^{\mathrm{FSP}}(\mathbf{x}^*, t) = 1 - p_{N-1}^{\mathrm{FSP}}(\mathbb{S}_{N-1}, t), \end{aligned}$$

where we used the fact that after T_N, the FSP process is absorbed at $\mathbf{x}^*$. Thus,

$$p_{N-1}^{\mathrm{FSP}}(\mathbf{x}, t) \leq p_N(\mathbf{x}, t) \leq p_{N-1}^{\mathrm{FSP}}(\mathbf{x}, t) + 1 - p_{N-1}^{\mathrm{FSP}}(\mathbb{S}_{N-1}, t),$$

and hence,

$$|p_N(\mathbf{x}, t) - p_{N-1}^{\mathrm{FSP}}(\mathbf{x}, t)| \leq 1 - p_{N-1}^{\mathrm{FSP}}(\mathbb{S}_{N-1}, t). \tag{B3}$$

Note that

$$p_{N-1}^{\mathrm{FSP}}(\mathbf{x}, t) = P(\mathbf{X}(t) = \mathbf{x}, t < T_N) \leq p(\mathbf{x}, t).$$

Since T_N increases to ∞ almost surely as N increases, $p_{N-1}^{\mathrm{FSP}}(\mathbf{x}, t)$ monotonically increases in N and converges to $p(\mathbf{x}, t)$ as $N \to \infty$ for each $\mathbf{x} \in \mathbb{S}_N$. Then, by using the monotone convergence theorem, the term

$$p_{N-1}^{\mathrm{FSP}}(\mathbb{S}_{N-1}, t) = \sum_{\mathbf{x}\in\mathbb{S}} p_{N-1}^{\mathrm{FSP}}(\mathbf{x}, t) \mathbb{1}_{\{x\in\mathbb{S}_{N-1}\}}$$

converges to $p(\mathbb{S}, t) = 1$. Since for any $t \in [0, T]$

$$\begin{aligned} p_{N-1}^{\mathrm{FSP}}(\mathbb{S}_{N-1}, T) &= P(\mathbf{X}^N(t) \in \mathbb{S}_{N-1}, T < T_N) \\ &\leq P(\mathbf{X}^N(t) \in \mathbb{S}_{N-1}, t < T_N) = p_{N-1}^{\mathrm{FSP}}(\mathbb{S}_{N-1}, t), \end{aligned}$$

therefore, by taking $\lim_{N\to\infty} \sup_{t\in[0,T]}$ in both sides of (B3), the result follows. □

Proof of Theorem V.2. This proof is a slight generalization of the proof of Theorem 3.3 of Ref. 25. Since the convergence of $\|p_N(\cdot, t) - \pi_N\|_1$ is independent of N, for any $\varepsilon > 0$, we choose sufficiently large t_0 such that

$$\|p(\cdot, t_0) - \pi(t_0)\|_1 \leq \varepsilon \quad \text{and} \quad \|p_N(\cdot, t_0) - \pi_N(t_0)\|_1 \leq \varepsilon$$

for all N. Then, by using the triangle inequalities,

$$\begin{aligned}\|\pi-\pi_N\|_1 &\le \|p(\cdot,t_0)-\pi\|_1+\|p_N(\cdot,t_0)-\pi_N\|_1+\|p(\cdot,t_0)-p_N(\cdot,t_0)\|_1\\ &\le \|p(\cdot,t_0)-\pi\|_1+\|p_N(\cdot,t_0)-\pi_N\|_1\\ &\quad+\|p(\cdot,t_0)-p^{\mathrm{FSP}}_{N-1}(\cdot,t_0)\|_1+\|p^{\mathrm{FSP}}_{N-1}(\cdot,t_0)-p_N(\cdot,t_0)\|_1\\ &\le 2\varepsilon+\sum_{\mathbf{x}\in\mathbb{S}}|p(\mathbf{x},t_0)-p^{\mathrm{FSP}}_{N-1}(\mathbf{x},t_0)|\\ &\quad+\sum_{\mathbf{x}\in\mathbb{S}}|p^{\mathrm{FSP}}_{N-1}(\mathbf{x},t_0)-p_N(\mathbf{x},t_0)|. \end{aligned}\tag{B4}$$

Note that as we mentioned in the proof of Theorem V.1, we have monotone convergence of $p^{\mathrm{FSP}}_{N-1}(\mathbf{x},t_0)$ to $p(\mathbf{x},t_0)$ for each $\mathbf{x}\in\mathbb{S}_{N-1}$ as $N\to\infty$. Hence, by the monotone convergence theorem, the first summation in (B4) goes to zero as $N\to\infty$. Note further that from (B2), we have

$$|p^{\mathrm{FSP}}_{N-1}(\mathbf{x},t_0)-p_N(\mathbf{x},t_0)|=P(\mathbf{X}^N(t)=\mathbf{x},T_N\le t_0).$$

Hence, the second summation in (B4) satisfies

$$\begin{aligned}\sum_{\mathbf{x}\in\mathbb{S}}|p^{\mathrm{FSP}}_{N-1}(\mathbf{x},t_0)-p_N(\mathbf{x},t_0)| &\le \sum_{\mathbf{x}\in\mathbb{S}_{N-1}}|p^{\mathrm{FSP}}_{N-1}(\mathbf{x},t_0)-p_N(\mathbf{x},t_0)|\\ &\quad+p^{\mathrm{FSP}}_{N-1}(\mathbf{x}^*,t_0)\\ &\quad+P(\mathbf{X}^N(t)\in\mathbb{S}_N\backslash\mathbb{S}_{N-1}).\end{aligned}$$

Note that by (B2),

$$\begin{aligned}&\sum_{\mathbf{x}\in\mathbb{S}_{N-1}}|p^{\mathrm{FSP}}_{N-1}(\mathbf{x},t_0)-p_N(\mathbf{x},t_0)|\\ &\quad=\sum_{\mathbf{x}\in\mathbb{S}_{N-1}}P(\mathbf{X}^N(t_0)=\mathbf{x},t_0\ge T_N)=P(t_0\ge T_N).\end{aligned}$$

Furthermore, $p^{\mathrm{FSP}}_{N-1}(\mathbf{x}^*,t_0)=P(t_0\ge T_N)$ and $P(\mathbf{X}^N(t)\in\mathbb{S}_N\backslash\mathbb{S}_{N-1})=P(T_N<t_0)$. Hence, $\sum_{\mathbf{x}\in\mathbb{S}}|p^{\mathrm{FSP}}_{N-1}(\mathbf{x},t_0)-p_N(\mathbf{x},t_0)|\to 0$ as $N\to\infty$ because $T_N\to\infty$ almost surely as $N\to\infty$.

Consequently, we have

$$\lim_{N\to\infty}\|\pi-\pi_N\|_1\le 2\varepsilon.$$

Since we choose an arbitrary ε, this completes the proof.

In order to prove Theorem V.3, we consider an "absorbing" Markov process associated with X and $\mathbf{X}^N$. Let $\bar{\mathbf{X}}$ and $\bar{\mathbf{X}}^N$ be Markov processes such that

$$\bar{\mathbf{X}}(t)=\begin{cases}\mathbf{X}(t) & \text{if } t<\tau\\ \mathbf{X}(\tau) & \text{if } t\ge\tau,\end{cases}$$

$$\bar{\mathbf{X}}^N(t)=\begin{cases}\mathbf{X}^N(t) & \text{if } t<\tau_N\\ \mathbf{X}^N(\tau_N) & \text{if } t\ge\tau_N.\end{cases}$$

That is, $\bar{\mathbf{X}}$ and $\bar{\mathbf{X}}^N$ are coupled processes to $\mathbf{X}$ and $\mathbf{X}^N$, respectively. Furthermore, they are absorbed to K once the coupled process ($\mathbf{X}$ for $\bar{X}$ and $\mathbf{X}^N$ for $\bar{\mathbf{X}}^N$) visits the set K. These coupled processes have the following relation.

Lemma B.1. Let $t\ge 0$ be fixed and $\varepsilon>0$ be arbitrary. Suppose X admits a stationary distribution π. Suppose further that $\lim_{N\to\infty}\|\pi-\pi_N\|_1=0$. Then, there exists a finite subset $\bar{K}$ such that $P(\bar{X}(t)\in\bar{K}^c)<\varepsilon$ and $P(\bar{\mathbf{X}}^N(t)\in\bar{K}^c)<\varepsilon$ for sufficiently large N.

Proof. Since the probabilities of $\bar{\mathbf{X}}$ and $\bar{\mathbf{X}}^N$ are leaking to K, we have for each t and for each $\mathbf{x}\in K^c$

$$P(\bar{\mathbf{X}}(t)=\mathbf{x})\le P(\mathbf{X}(t)=\mathbf{x}) \text{ and } P(\bar{\mathbf{X}}^N(t)=\mathbf{x})<P(\mathbf{X}^N(t)=\mathbf{x}).\tag{B5}$$

This can be formally proved as for each $x\in K^c$,

$$\begin{aligned}P(X(t)=\mathbf{x}) &= P(X(t)=\mathbf{x},\tau>t)+P(\mathbf{X}(t)=\mathbf{x},\tau\le t)\\ &= P(\bar{\mathbf{X}}(t)=\mathbf{x})+P(X(t)=\mathbf{x},\tau^K\le t)\\ &\ge P(\bar{\mathbf{X}}(t)=\mathbf{x}).\end{aligned}$$

In the same way, we can prove that $P(\bar{\mathbf{X}}^N(t)=\mathbf{x})<P(\mathbf{X}^N(t)=\mathbf{x})$ for each $\mathbf{x}\in K^c$.

Let $\varepsilon'>0$ be arbitrary. Then, $\lim_{N\to\infty}\|\pi-\pi_N\|_1=0$ implies that for any subset U, we have

$$\pi(U)-\varepsilon'\le\pi_N(U)\le\pi(U)+\varepsilon'\tag{B6}$$

for sufficiently large N. We use this property and (B5) combined with the stationarity of the systems.

First, note that we are assuming $\mathbf{X}(0)=\mathbf{X}^N(0)=\mathbf{x}_0$ for a fixed $\mathbf{x}_0$. Then, by the definition of the stationary distribution π,

$$P_{\mathbf{x}_0}(\mathbf{X}(t)=\mathbf{x})\pi(\mathbf{x}_0)\le\sum_{\mathbf{z}\in\mathbb{S}}P(\mathbf{X}(t)=\mathbf{x}|\mathbf{X}(t)=\mathbf{z})\pi(\mathbf{z})=\pi(\mathbf{x})$$

for each $t\ge 0$. Then, it follows that

$$P_{\mathbf{x}_0}(\mathbf{X}(t)=\mathbf{x})\le\frac{\pi(\mathbf{x})}{\pi(\mathbf{x}_0)}=\gamma_1\pi(\mathbf{x})\quad\text{for any }\mathbf{x}\text{ and }t,\tag{B7}$$

where $\gamma_1=1/\pi(\mathbf{x}_0)$. In the same way, for any N, it follows that

$$P_{\mathbf{x}_0}(\mathbf{X}^N(t)=\mathbf{x})\le\frac{\pi_N(\mathbf{x})}{\pi_N(\mathbf{x}_0)}\le\gamma_2\pi_N(\mathbf{x})\quad\text{for any }\mathbf{x}\text{ and }t,\tag{B8}$$

where $\gamma_2=1/(\pi(\mathbf{x}_0)-\varepsilon')\ge 1/\pi_N(\mathbf{x}_0)$ by (B6) with sufficiently small ε'.

Second, there exists a finite set $\bar{K}$ such that $\pi(\bar{K}^c)<\varepsilon'$ because π is a probability distribution. Furthermore, by (B6), $\pi_N(\bar{K}^c)\le\pi(\bar{K}^c)+\varepsilon'<2\varepsilon'$ for any sufficiently large N.

Finally, we choose $\varepsilon'=\frac{\varepsilon}{2\gamma}$, where $\gamma=\max\{\gamma_1,\gamma_2\}$. Then, by summing up (B5), (B7), and (B8) over $\mathbf{x}\in\bar{K}^c$, the result follows.

We prove Theorem V.3 by using the "absorbing" Markov processes $\bar{\mathbf{X}}$ and $\bar{\mathbf{X}}^N$ coupled to $\mathbf{X}$ and $\mathbf{X}^N$, respectively.

Proof of Theorem V.3. We first break the term $P(\tau>t)-P(\tau_N>t)$ to show that

$$\lim_{N\to\infty}\sup_{t\in[0,T]}|P(\tau>t)-P(\tau_N>t)|=0.$$

Let $\varepsilon > 0$ be an arbitrarily small number. Let $\tilde{K}$ be a finite set we found in Lemma B.1. Then, by using triangular inequalities,

$$\begin{aligned}|P(\tau > t) - P(\tau_N > t)| &= |P(\tilde{\mathbf{X}}(t) \in K^c) - P(\tilde{\mathbf{X}}^N(t) \in K^c)| \\ &\le \sum_{\mathbf{x} \in K^c \cap \tilde{K}} |P(\tilde{\mathbf{X}}(t) = \mathbf{x}) - P(\tilde{\mathbf{X}}^N(t) = \mathbf{x})| \\ &\quad + \sum_{\mathbf{x} \in K^c \cap \tilde{K}^c} |P(\tilde{\mathbf{X}}(t) = \mathbf{x}) - P(\tilde{\mathbf{X}}^N(t) = \mathbf{x})| \\ &\le \sum_{\mathbf{x} \in K^c \cap \tilde{K}} |P(\tilde{\mathbf{X}}(t) = \mathbf{x}) - P(\tilde{\mathbf{X}}^N(t) = \mathbf{x})| \\ &\quad + P(\tilde{\mathbf{X}}(t) \in \tilde{K}^c) + P(\tilde{\mathbf{X}}^N(t) \in \tilde{K}^c) \\ &\le \sum_{\mathbf{x} \in K^c \cap \tilde{K}} |P(\tilde{\mathbf{X}}(t) = \mathbf{x}) - P(\tilde{\mathbf{X}}^N(t) = \mathbf{x})| + 2\varepsilon. \end{aligned} \tag{B9}$$

Note that by the same proof of Theorem V.1, we have the convergence

$$\lim_{N \to \infty} \sup_{t \in [0,T]} |P(\tilde{\mathbf{X}}(t) = \mathbf{x}) - P(\tilde{\mathbf{X}}^N(t) = \mathbf{x})| = 0 \quad \text{for each } \mathbf{x} \in \mathbb{S}.$$

Since the summation $\sum_{\mathbf{x} \in K^c \cap \tilde{K}}$ is finite, we have that by taking $N \to \infty$ in (B9),

$$\lim_{N \to \infty} \sup_{t \in [0,T]} |P(\tau^K > t) - P(\tau_N^K > t)| = 0 + 2\varepsilon. \tag{B10}$$

Now, to show the convergence of the mean first passage times, note that

$$|E(\tau) - E(\tau_N)| \le \int_0^\infty |P(\tau > t) - P(\tau_N > t)| dt,$$

where the integrand is bounded by $P(\tau_K > t) + g(t)$.

Condition (iii) in Theorem V.3 implies that $E(\tau_k) < \infty$ and $\int_0^\infty g(t)dt < \infty$. Hence, the dominant convergence theorem and (B10) imply that

$$\lim_{N \to \infty} |E(\tau) - E(\tau_N)| = 0.$$

□

The existence of a special Lyapunov function ensures that the conditions in Theorem V.3 hold. In order to use the absorbing Markov processes, as we did in the previous proof, we define a Lyapunov function for $\tilde{\mathbf{X}}$ based on a given Lyapunov function V. Let $\tilde{\lambda}_{v \to v'}$ and $\tilde{\lambda}^N_{v \to v'}$ denote the intensity of a reaction $v \to v'$ for $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{X}}^N$, respectively. For a given function V such that $V(\mathbf{x}) \ge 1$ for any $\mathbf{x} \in \mathbb{S}$, we define $\overline{V}$ such that

$$\overline{V}(\mathbf{x}) = \begin{cases} V(\mathbf{x}) & \text{if } \mathbf{x} \in K^c \\ 1 & \text{if } \mathbf{x} \in K \end{cases}$$

so that $V(\mathbf{x}) \ge \overline{V}(\mathbf{x})$ for any $\mathbf{x}$.

Note that $\tilde{\lambda}_{v \to v'}(\mathbf{x}) = \lambda_{v \to v'}(\mathbf{x})$ if $\mathbf{x} \in K^c$. Hence, for each $\mathbf{x} \in K^c$,

$$\begin{aligned}&\sum_{v \to v' \in \mathscr{R}} \tilde{\lambda}_{v \to v'}(\mathbf{x})(\overline{V}(\mathbf{x} + v' - v) - \overline{V}(\mathbf{x})) \\ &\quad \le \sum_{v \to v' \in \mathscr{R}} \lambda_{v \to v'}(\mathbf{x})(V(\mathbf{x} + v' - v) - V(\mathbf{x})) \end{aligned} \tag{B11}$$

because $V(\mathbf{x}) = \overline{V}(\mathbf{x})$ and $V(\mathbf{x} + v' - v) \ge \overline{V}(\mathbf{x} + v' - v)$ for every reaction $\mathbf{x} + v' - v$. Moreover, $\tilde{\lambda}_{v \to v'}(\mathbf{x}) = 0 \le \lambda_{v \to v'}(\mathbf{x})$ if $\mathbf{x} \in K$ by the definition of $\tilde{\mathbf{X}}$ for each reaction $v \to v'$. Hence, for each $\mathbf{x} \in K$,

$$\sum_{v \to v' \in \mathscr{R}} \tilde{\lambda}_{v \to v'}(\mathbf{x})(\overline{V}(\mathbf{x} + v' - v) - \overline{V}(\mathbf{x})) = 0. \tag{B12}$$

From (B11) and (B12), therefore, we conclude that for any $\mathbf{x}$,

$$\begin{aligned}&\sum_{v \to v' \in \mathscr{R}} \tilde{\lambda}_{v \to v'}(\mathbf{x})(\overline{V}(\mathbf{x} + v' - v) - \overline{V}(\mathbf{x})) \\ &\quad \le \begin{cases} -CV(\mathbf{x}) + D & \text{if } \mathbf{x} \in K^c \\ 0 & \text{if } \mathbf{x} \in K \end{cases} \\ &\quad \le -C'\overline{V}(\mathbf{x}) + D', \end{aligned} \tag{B13}$$

where $C' = C$ and $D' = \max\{C + 1, D\}$.

Proof of Theorem V.4. In this proof, we show that all the conditions in Theorem V.3 are met by using the given function V. First, we show that X admits a stationary distribution. This follows straightforwardly by Theorem 3.2 of Ref. 25 because condition 3 in Theorem V.4 means that V is a Lyapunov function for $\mathbf{X}$.

The condition basically means that every outward reaction $v \to v'$ (i.e., $\mathbf{x} + v' - v \in \mathbb{S}_N^c$ and $\mathbf{x} \in \mathbb{S}_N$) in $\mathscr{R}$ gives a non-negative drift as $V(\mathbf{x} + v' - v) - V(\mathbf{x}) \ge 0$. We use condition 2 in Theorem V.4 to show that V is also a Lyapunov function for $\mathbf{X}^N$ for any N.

We denote by $\lambda^N_{\tilde{v} \to \tilde{v}'}$ the intensity of a reaction $\tilde{v} \to \tilde{v}'$ in $(\widetilde{\mathscr{S}}, \widetilde{\mathscr{C}}, \widetilde{\mathscr{R}}, \Lambda^N)$. Suppose that a reaction $\tilde{v} \to \tilde{v}' \in \widetilde{\mathscr{R}}$ is obtained from $v \to v' \in \mathscr{R}$ by adding a slack reactant. That is, $q(\tilde{v}' - \tilde{v}) = v' - v$, where q is the projection function defined at (B1). Then, by the definition (10) of the intensity in a slack network, we have $\lambda^N_{\tilde{v} \to \tilde{v}'}(\mathbf{x}) = 0$ when $\mathbf{x} + q(\tilde{v}' - \tilde{v}) \notin \mathbb{S}_N$ because the $\mathbf{X}^N$ is confined in $\mathbb{S}_N$. Furthermore, by condition 2 in Theorem V.4, $V(\mathbf{x} + v' - v) \ge V(x)$ when $\mathbf{x} + q(\tilde{v}' - \tilde{v}) \notin \mathbb{S}_N$ because this means that $x + v' - v \in \mathbb{S}_M$ for some $M > N$. This implies that if $\mathbf{x} + v' - v \notin \mathbb{S}_N$,

$$\begin{aligned} 0 &= \lambda^N_{\tilde{v} \to \tilde{v}'}(\mathbf{x})(V(\mathbf{x} + v' - v) - V(x)) \\ &\le \lambda_{v \to v'}(\mathbf{x})(V(\mathbf{x} + v' - v) - V(x)). \end{aligned} \tag{B14}$$

In case $\mathbf{x} + q(\tilde{v}' - \tilde{v}) \in \mathbb{S}_N \subseteq \mathbb{S}$, we have that $\lambda^N_{\tilde{v} \to \tilde{v}'}(\mathbf{x}) = \lambda_{v \to v'}(\mathbf{x})$ by definition (10). Hence, by condition 3 and (B14), we have for any $\mathbf{x}$

$$\begin{aligned}&\sum_{\tilde{v} \to \tilde{v}' \in \mathscr{R}} \lambda^N_{\tilde{v} \to \tilde{v}'}(\mathbf{x})(V(\mathbf{x} + v' - v) - V(\mathbf{x})) \\ &\quad \le \sum_{v \to v' \in \mathscr{R}} \lambda_{v \to v'}(\mathbf{x})(V(\mathbf{x} + v' - v) - V(\mathbf{x})) \\ &\quad \le -CV(x) + D. \end{aligned} \tag{B15}$$

Thus, V is also a Lyapunov function for $\mathbf{X}^N$. Hence, Theorem 6.1 (the Foster–Lyapunov criterion for exponential ergodicity) in Ref. 26 implies that for each N, there exist $\beta > 0$ and $\eta > 0$, which are only dependent of C and D, such that

$$\|p_N(\cdot, t) - \pi_N\|_1 \le \beta V(\mathbf{x}_0) e^{-\eta t}.$$

This guarantees that the condition in Theorem V.2 holds with $h(t) = \beta V(\mathbf{x}_0) e^{-\eta t}$. Hence, we have $\lim_{N \to \infty} \|\pi - \pi_N\|_1 = 0$ by Theorem V.2.

Now, to show that the first passage time τ has the finite mean, we apply the Foster–Lyapunov criterion to $\bar{X}$. Since (B13) meets the conditions of Theorem 6.1 of Ref. 26, the probability of $\bar{X}$ converges in time to its stationary distribution exponentially fast. That is, for any subset U, there exist $\tilde{\beta} > 0$ and $\tilde{\eta} > 0$ such that

$$\|P(\bar{\mathbf{X}}(t) \in U) - \tilde{\pi}(U)\| \le \beta V(\mathbf{x}_0) e^{-\eta t},$$

where $\tilde{\pi}$ is the stationary distribution of $\overline{X}$. This, in turn, implies that

$$\begin{aligned} E(\tau_K) &= \int_0^\infty P(\tau > t) dt \\ &= \int_0^\infty P(\bar{\mathbf{X}}(t) \in K^c) dt \\ &\le \int_0^\infty |P(\bar{\mathbf{X}}(t) \in K^c) - \tilde{\pi}(K^c)| + \tilde{\pi}(K^c) dt \\ &\le \int_0^\infty \beta V(\mathbf{x}_0) e^{-\eta t} dt < \infty, \end{aligned} \tag{B16}$$

where the second inequality follows as $\tilde{\pi}(K^C) = 0$, which is because $\bar{\mathbf{X}}$ is eventually absorbed in K as the original process $\mathbf{X}$ is irreducible and closed.

Finally, we show that for any N, there exists $g(t)$ such that $P(\tau_N < t) < g(t)$ and $\int_0^\infty g(t)dt < \infty$ for the first passage time $\tau_{K,N}$. By the same reasoning we used to derive (B13), we also derive by using (B15) that

$$\sum_{\tilde{\nu} \to \tilde{\nu}' \in \widetilde{\mathscr{R}}} \tilde{\lambda}^N_{\tilde{\nu} \to \tilde{\nu}'}(\mathbf{x}, y)(\overline{V}(\mathbf{x} + \nu' - \nu) - \overline{V}(\mathbf{x})) \le -C' \overline{V}(\mathbf{x}) + D'. \tag{B17}$$

Hence, in the same way as used for (B16), we have the exponential ergodicity of $\bar{\mathbf{X}}^N$ by Theorem 6.1 of Ref. 26, and then, we derive that

$$\begin{aligned} P(\tau_N > t) &= P(\bar{\mathbf{X}}^N(t) \in K^c) \\ &\le |P(\bar{\mathbf{X}}(t) \in K^c) - \tilde{\pi}_N(K^c)| + \tilde{\pi}_N(K^c) \\ &\le \beta V(\mathbf{x}_0) e^{-\eta t}, \end{aligned}$$

where $\tilde{\pi}_N$ is the stationary distribution of $\bar{\mathbf{X}}^N$, which also has zero probability in K^C as $\bar{\mathbf{X}}^N$ is eventually absorbed in K. Since β and η only depend on C' and D', we let $g(t) = \beta V(\mathbf{x}_0) e^{-\eta t}$ that satisfies that $P(\tau_{K,N} > t) \le g(t)$ for any N and $\int_0^\infty g(t)dt < \infty$. □

APPENDIX C: PROOFS OF ACCESSIBILITY THEOREMS

In this section, we prove Theorem VI.1. We begin with a necessary lemma. In the following lemma, for a given reaction system $(\mathscr{S}, \mathscr{C}, \mathscr{R}, \Lambda)$, we generate a slack system $(\widetilde{\mathscr{S}}, \widetilde{\mathscr{C}}, \widetilde{\mathscr{R}}, \Lambda^N)$ admitting a single slack reactant Y. We denote by $\mathbf{w}$ the vector for which the slack system admits the conservation bound $\mathbf{w} \cdot \mathbf{x} \le N$ for each state x of the slack system. We also let c be the maximum stoichiometric coefficient of the slack reactant Y in the slack network. Finally, note that, as shown in Sec. III B, $\widetilde{\mathscr{R}}$ and $\mathscr{R}$ have the one-to-one correspondence as every reaction in $\widetilde{\mathscr{R}}$ is obtained by adding the slack reactant to a reaction in $\mathscr{R}$. For the proof of Theorem V.4, we use the projection function q defined at (B1).

Lemma C.1. For a reaction system $(\mathscr{S}, \mathscr{C}, \mathscr{R}, \Lambda)$, let $(\widetilde{\mathscr{S}}, \widetilde{\mathscr{C}}, \widetilde{\mathscr{R}}, \Lambda^N)$ be a slack system with a single slack reactant Y and a conservation vector $\mathbf{w}$. Let c be the maximum stoichiometric coefficient of the slack reactant Y in $(\widetilde{\mathscr{S}}, \widetilde{\mathscr{C}}, \widetilde{\mathscr{R}}, \Lambda^N)$. Then, if $c \le N - (\mathbf{w} \cdot \mathbf{x})$ for a state $\mathbf{x}$, then

$$\lambda^N_{\tilde{\nu} \to \tilde{\nu}'}(\mathbf{x}) > 0 \quad \text{if only if} \quad \lambda_{\nu \to \nu'}(\mathbf{x}) > 0,$$

where $\tilde{\nu} \to \tilde{\nu}' \in \widetilde{\mathscr{R}}$ and $\nu \to \nu' \in \mathscr{R}$ are reactions such that $q(\tilde{\nu}) = \nu$ and $q(\tilde{\nu}') = \nu'$.

Proof. Let $\tilde{\nu} \in \widetilde{\mathscr{C}}$ such that $\tilde{\nu} = (\nu, u_\nu)$ with the stoichiometric coefficient u_ν of Y. By the definition of $\lambda^N_{\tilde{\nu} \to \tilde{\nu}'}$ shown in (10),

$$\lambda^N_{\tilde{\nu} \to \tilde{\nu}'}(\mathbf{x}) > 0 \quad \text{if only if} \quad \lambda_{\nu \to \nu'}(\mathbf{x}) > 0,$$

so long as $y = N - \mathbf{w} \cdot \mathbf{x} \ge u_\nu$. Since $c = \max_{\nu \in \mathscr{C}}\{u_\nu\}$, the result follows.

Proof of Theorem VI.1. We denote by $\mathbb{S}$ the irreducible and closed state space of $\mathbf{X}$ such that $\mathbf{X}(0) = \mathbf{x}_0$. We also denote by $\mathbb{S}_N$ the communication class of $\mathbf{X}^N$ such that $x_0 \in \mathbb{S}_N$ for each N. Then, $\mathbb{S}_N \subseteq \mathbb{S}$, and the irreducibility of $\mathbb{S}$ guarantees that $\cup_{i=1}^\infty \mathbb{S}_i = \mathbb{S}$. Therefore, there exists N_1 such that $\mathbb{S}_{N,\mathbf{x}_0} \cap K \neq \emptyset$ if $N \ge N_1$. This implies that it is enough to show that $\mathbb{S}_{N,\mathbf{x}_0}$ is closed for N large enough because $\mathbb{S}_{N,\mathbf{x}_0}$ is a finite subset. To prove this, we claim that there exists N_0 such that if $N \ge N_0$; then, $\mathbb{S}_N$ admits no out-going flow.

To prove the claim by contradiction, we assume that for a sequence N_k such that $\lim_{k\to\infty} N_k = \infty$, there is an out-going flow from $\mathbb{S}_{N_k}$ to a state $\mathbf{x}_k \in \mathbb{S}^c_{N_k}$. We find a sequence of reactions with which X_{N_k} returns to $\mathbb{S}_{N_k}$ from $\mathbf{x}_k$ for all large enough k.

First of all, note $\{X_i \to Y\} \subset \widetilde{\mathscr{R}}$ because each reaction $X_i \to \emptyset \in \mathscr{R}$ is converted to $X_i \to Y$ in any optimized slack network by its definition of an optimized slack network (see Sec. III C). Hence, for each k, by firing the reactions $\{X_i \to Y\}$, $\mathbf{X}_{N_k}$ can reach $0 = (0, 0, \ldots, 0)$ from $\mathbf{x}_k$.

The next aim is to show that for any fixed k large enough, there exist a sequence of reactions $\{\tilde{\nu}_1 \to \tilde{\nu}'_1, \ldots \tilde{\nu}_n \to \tilde{\nu}'_n\} \subset \widetilde{\mathscr{R}}$ such that we have (i) $\lambda^{N_k}_{\tilde{\nu}_i \to \tilde{\nu}'_i}(\mathbf{x}_k + \sum_{j=1}^{i-1}(\tilde{\nu}'_j - \tilde{\nu}_j)) > 0$ for each i and (ii) $\mathbf{x}_k + \sum_{j=1}^n (q(\tilde{\nu}'_j - \tilde{\nu}_j)) \in \mathbb{S}_{N_k}$. This means that X_{N_k} can reach $\mathbb{S}_{N_k}$ from $\mathbf{x}_k$ along the reactions $\tilde{\nu}_i \to \tilde{\nu}_i$ in that order. For this aim, we make use of the irreducibility of $\mathbb{S}$ so that there exists a sequence of reactions $\{\nu_1 \to \nu'_1, \ldots, \nu_n \to \nu'_n\} \subset \mathscr{R}$ such that (i) $\lambda_{\nu_i \to \nu'_i}(\mathbf{0} + \sum_{j=1}^{i-1}(\nu'_j - \nu_j)) > 0$ for each i and (ii) $\mathbf{0} + \sum_{j=1}^n (\nu'_j - \nu_j) \in \mathbb{S}_{N_1}$. By making $N_0 \ge N_1$ large enough, we have that if $N_k \ge N_0$, then

$$\mathbf{w} \cdot \left(\mathbf{0} + \sum_{j=1}^i (\nu'_j - \nu_j) \right) < N - c \quad \text{for each } i.$$

Hence, by Lemma C.1, for each reaction $\tilde{\nu}_i \to \tilde{\nu}'_i$ such that $q(\tilde{\nu}_i) = \nu_i$ and $q(\tilde{\nu}'_i) = \nu'_i$, we have

$$\lambda^N_{\tilde{\nu}_i \to \tilde{\nu}'_i}\left(\mathbf{0} + \sum_{j=1}^{i-1} (q(\tilde{\nu}'_j - \tilde{\nu}_j)) \right) > 0 \quad \text{for each } i. \tag{C1}$$

Finally, since $q(\tilde{\nu}_i) = \nu_i$ and $q(\tilde{\nu}'_i) = \nu'_i$ for each i, we have

$$\mathbf{0} + \sum_{j=1}^{i-1} (q(\tilde{\nu}'_j - \tilde{\nu}_j)) \in \mathbb{S}_{N_0} \subseteq \mathbb{S}_N. \tag{C2}$$

Hence, (C1) and (C2) imply that if $N_k \ge N_0$, then X_{N_k} can re-enter $\mathbb{S}_{N_k}$ from $\mathbf{0}$ with positive probability.

Consequently, we constructed a sequence of reactions along which $\mathbf{X}_{N_k}$ can re-enter $\mathbb{S}_{N_k}$ from $\mathbf{x}_k$ for any k such that $N_k \geq N_0$. Since each $\mathbb{S}_{N_k}$ is a communication class, $\mathbf{x}_k \in \mathbb{S}_{N_k}$, and in turn, this contradicts to the assumption that $\mathbf{x}_k \in \mathbb{S}_{N_k}$. Hence, the claim holds so that $\mathbb{S}_N$ is closed for any N large enough. □

APPENDIX D: PROOFS FOR LEMMA VI.1 AND THEOREM VI.2

Proof of Lemma VI.1. In the deterministic model $\tilde{\mathbf{x}}(t) = (\tilde{x}_1(t), \ldots, \tilde{x}_d(t), y_1(t), \ldots, y_m(t))^\top$ associated with $(\widetilde{\mathscr{S}}, \widetilde{\mathscr{C}}, \widetilde{\mathscr{R}}, \Lambda^N)$, if we fix $y_i(t) \equiv 1$ for all $i = 1, 2, \ldots, m$, then $\tilde{\mathbf{x}}(t)$ follows the same ODE system as the deterministic model $\mathbf{x}(t)$ associated with the original network $(\mathscr{S}, \mathscr{C}, \mathscr{R}, \Lambda)$ does. Therefore, $\tilde{c} = (c^*, 1, 1, \ldots, 1)^\top$ is a complex balanced steady state for $\tilde{\mathbf{x}}(t)$. It has been shown that the existence of a single positive complex balanced steady state implies that all other positive steady states are complex balanced,[31] therefore completing the proof. □

Proof of Theorem VI.2. Let $\mathbf{w}_i$ and N_i be the conservation vector and the conservation quantity of the slack network $(\widetilde{\mathscr{S}}, \widetilde{\mathscr{C}}, \widetilde{\mathscr{R}}, \Lambda^N)$, respectively. Then, π is a stationary solution of the chemical master equation (3) of X,

$$\sum_{\nu \to \nu' \in \mathscr{R}} \lambda_{\nu \to \nu'}(\mathbf{x} - \nu' + \nu)\pi(\mathbf{x} - \nu' + \nu) = \sum_{\nu \to \nu' \in \mathscr{R}} \lambda_{\nu \to \nu'}(\mathbf{x})\pi(\mathbf{x}).$$

Especially, as shown in Theorem 6.4, Ref. 32, π satisfies the *stochastic complex balance* for X: for each complex $\nu^* \in \mathscr{R}$ and a state $\mathbf{x}$,

$$\sum_{\substack{\nu \to \nu' \in \mathscr{R} \\ \nu' = \nu^*}} \lambda_{\nu \to \nu'}(\mathbf{x} - \nu' + \nu)\pi(\mathbf{x} - \nu' + \nu) = \sum_{\substack{\nu \to \nu' \in \mathscr{R} \\ \nu = \nu^*}} \lambda_{\nu \to \nu'}(\mathbf{x})\pi(\mathbf{x}). \quad \text{(D1)}$$

Let q be the projection function defined at (B1). Note that each reaction $\tilde{\nu} \to \tilde{\nu}' \in \widetilde{\mathscr{R}}$ is defined as it satisfies the conservation law

$$\mathbf{w}_i \cdot (\nu' - \nu) + \tilde{\nu}'_{d+i} - \tilde{\nu}_{d+i} = 0, \quad \text{(D2)}$$

where $q(\tilde{\nu}) = \nu$ and $q(\tilde{\nu}') = \nu'$. Then, π also satisfies the stochastic complex balance for $\mathbf{X}^N$ because (D1) and (D2) imply that for each complex $\tilde{\nu}^* \in \widetilde{\mathscr{R}}$ and a state $\mathbf{x}$,

$$\begin{aligned}
&\sum_{\substack{\tilde{\nu} \to \tilde{\nu}' \in \widetilde{\mathscr{R}} \\ \tilde{\nu}' = \tilde{\nu}^*}} \lambda^N_{\tilde{\nu} \to \tilde{\nu}'}(\mathbf{x} - q(\nu' - \nu))\pi(\mathbf{x} - q(\nu' - \nu)) \\
&= \sum_{\substack{\nu \to \nu' \in \mathscr{R} \\ \tilde{\nu}' = \tilde{\nu}^*}} \lambda_{\tilde{\nu} \to \tilde{\nu}'}(\mathbf{x} - \nu' + \nu)\pi(\mathbf{x} - \nu' + \nu) \prod_{i=1}^{r} \mathbb{1}_{\{N_i - \mathbf{w}_i \cdot (\mathbf{x} - \nu' + \nu) \geq \tilde{\nu}_{d+i}\}} \\
&= \sum_{\substack{\nu \to \nu' \in \mathscr{R} \\ \tilde{\nu}' = \tilde{\nu}^*}} \lambda_{\tilde{\nu} \to \tilde{\nu}'}(\mathbf{x} - \nu' + \nu)\pi(\mathbf{x} - \nu' + \nu) \prod_{i=1}^{r} \mathbb{1}_{\{N_i - \mathbf{w}_i \cdot \mathbf{x} \geq \tilde{\nu}^*_{d+i}\}} \\
&= \sum_{\substack{\nu \to \nu' \in \mathscr{R} \\ \nu = \nu^*}} \lambda_{\nu \to \nu'}(\mathbf{x})\pi(\mathbf{x}) \mathbb{1}_{\{N_i - \mathbf{w}_i \cdot \mathbf{x} \geq \tilde{\nu}^*_{d+i}\}} \\
&= \sum_{\substack{\tilde{\nu} \to \tilde{\nu}' \in \widetilde{\mathscr{R}} \\ \tilde{\nu} = \tilde{\nu}^*}} \lambda^N_{\tilde{\nu} \to \tilde{\nu}'}(\mathbf{x})\pi(\mathbf{x}).
\end{aligned}$$

Then, by summing up for each $\tilde{\nu}^* \in \widetilde{\mathscr{R}}$,

$$\begin{aligned}
&\sum_{\tilde{\nu} \to \tilde{\nu}'; \in \widetilde{\mathscr{R}}} \lambda^N_{\tilde{\nu} \to \tilde{\nu}'}(\mathbf{x} - q(\nu' - \nu))\pi(\mathbf{x} - q(\nu' - \nu)) \\
&\quad = \sum_{\tilde{\nu} \to \tilde{\nu}' \in \widetilde{\mathscr{R}}} \lambda^N_{\tilde{\nu} \to \tilde{\nu}'}(\mathbf{x})\pi(\mathbf{x}),
\end{aligned}$$

and in turn, π_N is a stationary solution of the chemical master equation of $\mathbf{X}^N$. Since the state space of $\mathbf{X}$ and $\mathbf{X}^N$ differ, M_N is a constant such that the sum of $M_N\pi(\mathbf{x})$ over the state space of $\mathbf{X}^N$ is one. Then, $\pi_N = M_N\pi$.

APPENDIX E: LYAPUNOV FUNCTIONS FOR EXAMPLE SYSTEMS

1. Lotka–Volterra with migration

We construct a Lyapunov function satisfying the conditions of Theorem V.4 for the Lotka–Volterra model with migration [Fig. 4(a)]. First, for both the original model and the slack network, we assume $A(0) = a_0$, $B(0) = b_0$ that do not depend on N.

Let $V(\mathbf{x}) = e^{w \cdot \mathbf{x}}$ with $w = (1, 1)^\top$. The condition in Theorem V.4 holds. Note that the slack network of Fig. 4(a) admits a conservation law $A + B + Y = N$ and the state space of the system $\mathbb{S}_N$ is $\{(a, b) | a + b \leq N\}$. Then, condition 2 in Theorem V.4 also follows by the definition of V.

Now, to show condition 3 in Theorem V.4, we first note that for $\mathbf{x} = (a, b)$,

$$\begin{aligned}
&\sum_{\nu \to \nu' \in \mathscr{R}} \lambda_{\nu \to \nu'}(\mathbf{x})(V(\mathbf{x} + \nu' - \nu) - V(\mathbf{x})) \\
&\quad = e^{\mathbf{w} \cdot \mathbf{x}} \sum_{\nu \to \nu' \in \mathscr{R}} \lambda_{\nu \to \nu'}(\mathbf{x})(e^{\mathbf{w} \cdot (\nu' - \nu)} - 1) \\
&\quad = V(\mathbf{x})\big(\kappa_1 b(e^{-1} - 1) + \kappa_2(e - 1) + \kappa_3(e^{-1} - 1) \\
&\qquad + \kappa_4 a(e^{-1} - 1) + \kappa_5 a(e - 1)\big). \quad \text{(E1)}
\end{aligned}$$

Since we assumed $\frac{\kappa_4}{\kappa_5} = 3$ (see the caption of Fig. 4), each coefficient of a and b is negative. Thus, there exists $M > 0$ such that for each $\mathbf{x} \in \{(a, b) \mid a > M \text{ or } b \geq M\}$, the left-hand side of (E1) satisfies

$$\sum_{\nu \to \nu' \in \mathscr{R}} \lambda_{\nu \to \nu'}(\mathbf{x})(V(\mathbf{x} + \nu' - \nu) - V(\mathbf{x})) \leq -CV(\mathbf{x})$$

for some $C > 0$. Letting

$$D = (C + 1) \max_{a \leq M \text{ and } b \leq M} \sum_{\nu \to \nu' \in \mathscr{R}} \lambda_{\nu \to \nu'}(\mathbf{x})(V(\mathbf{x} + \nu' - \nu) - V(\mathbf{x})),$$

condition 3 in Theorem V.4 holds with C and D.

2. Protein synthesis with a slow toggle switch

We also make use of the Lyapunov approach shown in Theorem V.4 to prove that the first passage time of the slack system in Fig. 5(b) converges to the original first passage time as the truncation size N goes to infinity. Let $\mathbf{X}^N = (X, Z, D_0^X, D_1^X, D_0^Z, D_1^Z)$ be the stochastic system associated with the slack system. Recall that the slack network admits a conservation relation $\mathbf{w} \cdot \mathbf{X}^N \leq N$ with $\mathbf{w} = (1, 1, 0, 0, 0, 0)$. Hence, we define a Lyapunov function $V(\mathbf{x}) = e^{x+z}$, where $\mathbf{x} = (x, z, dx_0, dx_1, dz_0, dz_1)$. By the definition of V, it is obvious that condition 1 in Theorem V.4 holds. Furthermore,

since $\mathbb{S}_N = \{\mathbf{x} \mid x + z \le N\}$, condition 2 in Theorem V.4 also holds. So, we show that condition 3 in Theorem holds.

For a reaction $\nu \to \nu' \in \mathscr{R}_I := \{Z + D_0^X \to D_1^X, X \to 0, X + D_0^Z \to D_1^Z, Z \to 0\}$, it is clear that the term $V(\mathbf{x} + \nu' - \nu) - V(\mathbf{x})$ is negative. For each reaction $\nu \to \nu'$ in $\mathscr{R}_I^c$, it is also clear that the term $V(\mathbf{x} + \nu' - \nu) - V(\mathbf{x})$ is positive. However, the reaction intensity for a reaction in $\mathscr{R}_I$ is linear in either x or z, while the reaction intensity for a reaction in $\mathscr{R}_I^c$ is constant. Therefore, there exists a constant $M > 0$ such that if $x > M$ or $z > M$, then

$$\begin{aligned}\sum_{\nu\to\nu'\in\mathscr{R}} \lambda_{\nu\to\nu'}(\mathbf{x})(V(\mathbf{x}+\nu'-\nu)-V(\mathbf{x})) &\le C' \sum_{\nu\to\nu'\in\mathscr{R}_I} \lambda_{\nu\to\nu'}(\mathbf{x})(V(\mathbf{x}+\nu'-\nu)-V(\mathbf{x}))\\ &= -CV(\mathbf{x})\end{aligned}$$

for some constants $C' > 0$ and $C > 0$. Hence, letting

$$D = (C+1) \max_{x\le M \text{ and } z\le M} \sum_{\nu\to\nu'\in\mathscr{R}} \lambda_{\nu\to\nu'}(\mathbf{x})(V(\mathbf{x}+\nu'-\nu)-V(\mathbf{x})),$$

condition 3 in Theorem V.4 holds with C and D.

3. The exclusive mutual inhibition, self-activation model

We show that $V(\mathbf{x}) = e^{x+z}$ is a Lyapunov function satisfying the conditions in V.4 for each $\mathbf{x}$ in the state space $\mathbb{S} = \{\mathbf{x} = (x, z, a00, a01, a10, b00, b01, b10) \in \mathbb{Z}_{\ge 0}^8 \mid a00 + a01 + a10 = 1, b00 + b01 + b10 = 1\}$ of ExMISA model (22). Note that every reaction producing either X or Z in (22) has a constant order reaction intensity because the counts of genes A and B cannot exceed 1. On the contrary, each reaction consuming either X or Z has either linear or quadratic reaction intensity. Let $\mathscr{R}_I$ be a collection of reactions in the ExMISA model that degrade either X or Z such as $A_{00} + 2X \to A_{01}$. Then, in the same way we showed in Appendix E 2, there exists a constant $M > 0$ such that for each $\mathbf{x} \in \mathbb{S}$, if $x > M$ or $z > M$, then

$$\begin{aligned}\sum_{\nu\to\nu'\in\mathscr{R}} \lambda_{\nu\to\nu'}(\mathbf{x})(V(\mathbf{x}+\nu'-\nu)-V(\mathbf{x})) &\le C' \sum_{\nu\to\nu'\in\mathscr{R}_I} \lambda_{\nu\to\nu'}(\mathbf{x})(V(\mathbf{x}+\nu'-\nu)-V(\mathbf{x}))\\ &= -CV(\mathbf{x})\end{aligned}$$

for some constant $C' > 0$ and $C > 0$. Hence, letting

$$D = (C+1) \max_{x\le M \text{ and } z\le M} \sum_{\nu\to\nu'\in\mathscr{R}} \lambda_{\nu\to\nu'}(\mathbf{x})(V(\mathbf{x}+\nu'-\nu)-V(\mathbf{x})),$$

condition 3 in Theorem V.4 holds with C and D. Finally, note that $\mathbb{S}_N = \{\mathbf{x} \mid x + z \le N\}$ since the slack system admits a conservation bound

$$\mathbf{w} \cdot \mathbf{x} \le N \quad \text{with } \mathbf{w} = (1, 1, 0, 0, 0, 0, 0, 0)^\top.$$

Hence, conditions 1 and 2 in Theorem V.4 hold.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] K. A. Johnson and R. S. Goody, "The original Michaelis constant: Translation of the 1913 Michaelis-Menten paper," Biochem. **50**(39), 8264–8269 (2011).
[2] J. Paulsson, "Summing up the noise in gene networks," Nature **427**, 415–418 (2004).
[3] G. Enciso and J. Kim, "Embracing noise in chemical reaction networks," Bull. Math. Biol. **81**, 1261–1267 (2019).
[4] D. F. Anderson and D. Cappelletti, "Discrepancies between extinction events and boundary equilibria in reaction networks," J. Math. Biol. **79**, 1253–1277 (2019).
[5] F. López-Caamal and T. T. Marquez-Lago, "Exact probability distributions of selected species in stochastic chemical reaction networks," Bull. Math. Biol. **76**, 2334–2361 (2014).
[6] D. T. Gillespie, "Stochastic simulation of chemical kinetics," Annu. Rev. Phys. Chem. **58**, 35–55 (2007).
[7] S. MacNamara, K. Burrage, and R. B. Sidje, "Multiscale modeling of chemical kinetics via the master equation," Multiscale Model. Simul. **6**, 1146–1168 (2008).
[8] E. Seneta, "Finite approximations to infinite non-negative matrices," Math. Proc. Cambridge Philos. Soc. **63**, 983–992 (1967).
[9] R. L. Tweedie, "Truncation procedures for non-negative matrices," J. Appl. Probab. **8**, 311–320 (1971).
[10] B. Munsky and M. Khammash, "The finite state projection algorithm for the solution of the chemical master equation," J. Chem. Phys. **124**, 044104 (2006).
[11] A. Gupta, J. Mikelson, and M. Khammash, "A finite state projection algorithm for the stationary solution of the chemical master equation," J. Chem. Phys. **147**, 154101 (2017).
[12] Y. Cao and J. Liang, "Optimal enumeration of state space of finitely buffered stochastic molecular networks and exact computation of steady state landscape probability," BMC Syst. Biol. **2**, 30 (2008).
[13] Y. Cao, A. Terebus, and J. Liang, "Accurate chemical master equation solution using multi-finite buffers," Multiscale Model. Simul. **14**, 923–963 (2016).
[14] J. Kuntz, P. Thomas, G.-B. Stan, and M. Barahona, "Approximations of countably-infinite linear programs over bounded measure spaces," arXiv:1810.03658 (2018).
[15] J. Kuntz, P. Thomas, G.-B. Stan, and M. Barahona, "Bounding the stationary distributions of the chemical master equation via mathematical programming," J. Chem. Phys. **151**, 034109 (2019).
[16] J. Kuntz, P. Thomas, G.-B. Stan, and M. Barahona, "Stationary distributions of continuous-time Markov chains: A review of theory and truncation-based approximations," SIAM Rev. (to be published); arXiv:1909.05794 (2019).
[17] A. Kazeroonian, F. Fröhlich, A. Raue, F. J. Theis, and J. Hasenauer, "CERENA: Chemical reaction network analyzer—A toolbox for the simulation and analysis of stochastic chemical kinetics," PLoS One **11**, e0146732 (2016).
[18] J. P. Hespanha, "`StochDynTools`—A MATLAB toolbox to compute moment dynamics for stochastic networks of bio-chemical reactions," available at http://www.ece.ucsb.edu/~hespanha, 2007.
[19] J. K. Kim and E. Sontag, "`FEEDME`—A MATLAB codes to calculate stationary moments of feed forward network and complex balanced network," available at http://www.ece.ucsb.edu/~hespanha, 2007.
[20] D. T. Gillespie, "A rigorous derivation of the chemical master equation," Physica A **188**, 404–425 (1992).
[21] D. F. Anderson and T. G. Kurtz, *Stochastic Analysis of Biochemical Systems* (Springer, 2015).
[22] T. Chou and M. D'orsogna, "First passage problems in biology," World Sci. Rev. **2014**, 306–345.
[23] M. Feinberg, "Chemical reaction network structure and the stability of complex isothermal reactors—I. The deficiency zero and deficiency one theorems," Chem. Eng. Sci. **42**, 2229–2268 (1987).
[24] E. August and M. Barahona, "Solutions of weakly reversible chemical reaction networks are bounded and persistent," IFAC Proc. Vol. **43**, 42–47 (2010).
[25] A. G. Hart and R. L. Tweedie, "Convergence of invariant measures of truncation approximations to Markov processes," Appl. Math. **3**, 2205–2215 (2012).
[26] S. P. Meyn and R. L. Tweedie, "Stability of Markovian processes III: Foster–Lyapunov criteria for continuous-time processes," Adv. Appl. Probab. **25**, 518–548 (1993).

[27]C. Briat, A. Gupta, and M. Khammash, "Antithetic integral feedback ensures robust perfect adaptation in noisy biomolecular networks," Cell Syst. **2**, 15–26 (2016).
[28]D. F. Anderson and J. Kim, "Some network conditions for positive recurrence of stochastically modeled reaction networks," SIAM J. Appl. Math. **78**, 2692–2713 (2018).
[29]M. C. Hansen and W. Carsten, "Existence of a unique quasi-stationary distribution in stochastic reaction networks," Electron. J. Probab. **25**, 45 (2020).
[30]L. Paulevé, G. Craciun, and H. Koeppl, "Dynamical properties of discrete reaction networks," J. Math. Biol. **69**, 55–72 (2014).
[31]M. Feinberg, "Complex balancing in general kinetic systems," Arch. Ration. Mech. Anal. **49**, 187–194 (1972).
[32]D. F. Anderson, G. Craciun, and T. G. Kurtz, "Product-form stationary distributions for deficiency zero chemical reaction networks," Bull. Math. Biol. **72**, 1947–1970 (2010).
[33]F. Horn, "Necessary and sufficient conditions for complex balancing in chemical kinetics," Arch. Ration. Mech. Anal. **49**, 172–186 (1972).
[34]M. A. Al-Radhawi, D. Del Vecchio, and E. D. Sontag, "Multi-modality in gene regulatory networks with slow promoter kinetics," PLoS Comput. Biol. **15**, e1006784 (2019).
[35]B. K. Chu, J. T. Margaret, R. R. Sato, and E. L. Read, "Markov state models of gene regulatory networks," BMC Syst. Biol. **11**, 14 (2017).
[36]J. T. Margaret, B. K. Chu, C. P. Gallivan, and E. L. Read, "Rare-event sampling of epigenetic landscapes and phenotype transitions," PLoS Comput. Biol. **14**, e1006336 (2018).
[37]M. Acar, J. T. Mettetal, and A. Van Oudenaarden, "Stochastic switching as a survival strategy in fluctuating environments," Nat. Genet. **40**, 471–475 (2008).
[38]J. J. Hunter, "The computation of the mean first passage times for Markov chains," Linear Algebra Appl. **549**, 100–122 (2018).
[39]R. Yvinec, M. R. D'Orsogna, and T. Chou, "First passage times in homogeneous nucleation and self-assembly," J. Chem. Phys. **137**, 244107 (2012).