

Efficient stochastic simulation of metastable Markov chains

Andreas Miliadis-Argeitis and John Lygeros

Abstract—We address the problem of metastable Markov chain simulation, a class of systems characterized by the existence of two or more “pseudo-equilibrium” states and very slow convergence towards global equilibrium [1]. For such systems, approximation of the stationary distribution by direct application of the Stochastic Simulation Algorithm (SSA) [2] would be very inefficient. In this paper we propose a new method for steady-state simulation of metastable chains that is centered around the concept of stochastic complementation [3]. The use of this mathematical device along with SSA results in an algorithm with much better convergence properties, that facilitates the analysis of rarely switching stochastic biochemical systems.

I. INTRODUCTION

In this work we consider continuous-time Markov chains with countable state space and assume they satisfy the necessary and sufficient conditions for having a unique invariant distribution. Except for special cases, the invariant distribution of these chains cannot be calculated explicitly and the main approximation technique used is state-space truncation. Following this approach, one must verify that the truncated chain “approaches” the original in some sense as the truncation size grows, so that their invariant distributions also approach each other. However, the necessary truncation size is difficult to determine and the resulting truncations are usually too big for the invariant distribution to be computed.

On the other extreme, stochastic simulation is always possible and, given the ergodicity assumption, is guaranteed to give correct results as the simulation length increases. Often, however, the computation time is too long, as systems may evolve slowly towards equilibrium. One cause for the slow convergence is the coexistence of two or more different timescales in the Markov chain transitions. The way out of this situation is the use of quasi-steady-state approximation simulation techniques that treat the two timescales separately [2].

Not all slowly converging systems involve different timescales, however. In some cases the Markov chain may evolve quickly towards a state of “pseudo-equilibrium” (which may be different depending on the initial condition), in which it remains trapped for a large amount of time. Viewed at a longer timescale, though, the system is not globally in equilibrium because it makes rare transitions from one pseudo-equilibrium to another. This behavior is translated into the existence of two or more modes in the

invariant distribution, separated by areas of very low probability. Such systems are called *metastable* [1], because they spend a large amount of time around each mode, making rare fluctuation-driven transitions between the mode regions. This is the case of bistable switch-like systems, for example, that are frequently encountered in the systems biology literature. Convergence to equilibrium of metastable systems can be extremely slow, and rare event simulation techniques have to be used for efficient simulation [4], [5], which require much more effort and tuning than SSA to be implemented.

In this paper we propose a simpler method for calculating stationary distributions of metastable systems, that involves “off-line” computations (like the truncation techniques) as well as use of SSA, and is based on the theory of stochastic complement [3]. With this device, one can simulate the Markov chain on a given subset of the state space, by “masking out” transitions to states outside this set so that the true invariant distribution over this subset is preserved. Our main idea is to first divide the state space in different parts, namely the metastable and the transition regions. Two tasks have to be accomplished then: the approximation of the invariant distribution of the chain, conditioned on it being inside a certain region, and the processing of the results from individual regions to obtain the full stationary distribution.

II. STOCHASTIC COMPLEMENT THEORY

Assumptions and notation: Consider an ergodic Markov chain $\{X_t, t \geq 0\}$ with generator matrix Q over a countable set $\mathcal{S} \subseteq \mathbb{Z}_{\geq 0}^n$, with a (unique) invariant distribution denoted by π . All vectors are assumed to be row vectors (so that $\pi Q = 0$ for example) and e_i denotes the unit vector with 1 at the i^{th} position. $\|\cdot\|$ denotes the norm of $\ell^1(\mathcal{S})$. Finally, vector $\mathbf{1}$ consists all of ones. The main references for this Section are the seminal paper of Meyer [3], as well as books [6], [7] and articles [8], [9].

Definition 2.1: Consider generator matrix Q with a k -level partition:

$$Q = \begin{pmatrix} Q_{11} & Q_{12} & \dots & Q_{1k} \\ Q_{21} & Q_{22} & \dots & Q_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{k1} & Q_{k2} & \dots & Q_{kk} \end{pmatrix}$$

For a given index i , let Q_i denote the principal block submatrix of Q obtained by deleting the i -th row and i -th column of blocks, and let

$$Q_{i*} = (Q_{i1} \quad Q_{i2} \quad \dots \quad Q_{i,i-1} \quad Q_{i,i+1} \quad \dots \quad Q_{i,k}) \text{ and} \\ Q_{*i} = (Q_{1i}^T \quad Q_{2i}^T \quad \dots \quad Q_{i-1,i}^T \quad Q_{i+1,i}^T \quad \dots \quad Q_{k,i}^T)^T.$$

Andreas Miliadis-Argeitis and John Lygeros are with the Automatic Control Laboratory, Department of Information Technology and Electrical Engineering, ETH Zurich, Switzerland miliadis,lygeros@control.ee.ethz.ch

The stochastic complement of Q_{ii} in Q is then defined as [3], [8], [6]

$$R_{ii} = Q_{ii} - Q_{i*}Q_i^{-1}Q_{*i}. \quad (1)$$

The partition of Q naturally induces a partition of \mathcal{S} into k disjoint subsets $\{\mathcal{B}_i, i = 1, \dots, k\}$, such that $\cup_{i=1}^k \mathcal{B}_i = \mathcal{S}$. Furthermore, each R_{ii} is itself a generator matrix of a reduced Markov chain on \mathcal{B}_i , and is irreducible if Q is [3]. One can also easily see that the reduced chain on \mathcal{B}_i is derived by observing the full chain on \mathcal{S} only when it is in \mathcal{B}_i . This means that transitions from \mathcal{B}_i to \mathcal{B}_i^c are “masked out” and a path $b_i^1 \rightarrow b_i^2$ in \mathcal{B}_i can correspond either to a direct path $b_i^1 \rightarrow b_i^2$ or a detour $b_i^1 \rightarrow \mathcal{B}_i^c \rightarrow b_i^2$ in the full chain.

The next two very important facts about stochastic complements from [3] form the basis of our approach:

Theorem 2.1: Let π denote the (unique) stationary distribution of the full Markov chain and write $\pi = (\pi^{(1)} \ \pi^{(2)} \ \dots \ \pi^{(k)})$, according to the state space k -level partition. Then the vector

$$\nu_i = \frac{\pi^{(i)}}{\pi^{(i)} \cdot \mathbf{1}^T} \quad (2)$$

is the unique stationary distribution of the reduced chain on \mathcal{B}_i (i.e. $\nu_i R_{ii} = 0$).

Theorem 2.2: The stationary distribution of Q is given by

$$\pi = (\xi_1 \nu_1 \quad \xi_2 \nu_2 \quad \dots \quad \xi_k \nu_k), \quad (3)$$

where the vector $\xi = (\xi_1 \quad \xi_2 \quad \dots \quad \xi_k)$ is the invariant distribution of the $k \times k$ generator matrix C with entries given by

$$c_{ij} = \nu_i Q_{ij} \mathbf{1}^T. \quad (4)$$

Theorem 2.1 says that ν_i assigns to each state b_i in \mathcal{B}_i a mass equal to the probability that the full chain is at b_i , conditioned on the fact that it is in \mathcal{B}_i . The result is of great practical importance, since it states that the invariant distributions of the stochastic complements, ν_1, \dots, ν_k , are scaled versions of the corresponding blocks of π . The scaling constants are called *coupling factors* and are given by Theorem 2.2.

III. SIMULATION ALGORITHM

The following is a key observation from (1): if each state in \mathcal{S} leads to finitely many states and \mathcal{B}_i^c is finite (which implies that \mathcal{B}_i infinite), then R_{ii} will be obtained by modifying Q_{ii} at *finitely many* points. This means that we can practically calculate R_{ii} and use it for simulation purposes. Based on this fact, we will next present our simulation algorithm for the simple case of a Markov chain with two metastable sets. Generalization to more than two sets will then be straightforward.

A. Set-up

Consider a metastable Markov chain with two finite and disjoint metastable sets, \mathcal{B}_1 and \mathcal{B}_2 . Let also $\mathcal{B}_3 = (\mathcal{B}_1 \cup \mathcal{B}_2)^c$ and assume that no transition from \mathcal{B}_1 to \mathcal{B}_2 is possible in

one step. By appropriately enumerating the states of \mathcal{S} , the generator of this Markov chain can then be written as:

$$Q = \begin{bmatrix} Q_{11} & Q_{12} & Q_{13} \\ Q_{21} & Q_{22} & Q_{23} \\ Q_{31} & Q_{32} & Q_{33} \end{bmatrix} = \begin{bmatrix} Q_{11} & 0 & Q_{13} \\ 0 & Q_{22} & Q_{23} \\ Q_{31} & Q_{32} & Q_{33} \end{bmatrix},$$

and the stochastic complement R_{33} is then given by (1):

$$\begin{aligned} R_{33} &= Q_{33} - [Q_{31} \quad Q_{32}] \begin{bmatrix} Q_{11}^{-1} & 0 \\ 0 & Q_{22}^{-1} \end{bmatrix} \begin{bmatrix} Q_{13} \\ Q_{23} \end{bmatrix} \\ &= Q_{33} - Q_{31}Q_{11}^{-1}Q_{13} - Q_{32}Q_{22}^{-1}Q_{23} \end{aligned} \quad (5)$$

We also introduce a few more subsets of \mathcal{S} : Let $\mathcal{B}_3^{3 \rightarrow 1}$ ($\mathcal{B}_3^{3 \rightarrow 2}$) denote the subset \mathcal{B}_3 from which a transition to \mathcal{B}_1 (\mathcal{B}_2) is possible. We also consider $\mathcal{B}_1^{1 \rightarrow 3} \subset \mathcal{B}_1$ and $\mathcal{B}_2^{2 \rightarrow 3} \subset \mathcal{B}_2$, the sets of points that are accessible from \mathcal{B}_3 as well as $\mathcal{B}_1^{1 \rightarrow 3} \subset \mathcal{B}_1$ and $\mathcal{B}_2^{2 \rightarrow 3} \subset \mathcal{B}_2$, the subsets from which \mathcal{B}_3 can be accessed.

B. First step

The structure of R_{33} suggests a simple way to simulate a Markov chain with generator R_{33} : The reduced Markov chain on \mathcal{B}_3 evolves exactly like the full Markov chain in $\mathcal{B}_3 \setminus (\mathcal{B}_3^{3 \rightarrow 1} \cup \mathcal{B}_3^{3 \rightarrow 2})$, until it reaches $\mathcal{B}_3^{3 \rightarrow 1}$ or $\mathcal{B}_3^{3 \rightarrow 2}$. Since it is not allowed to exit the set, the transition rates for these states are modified according to the second or the third term in (5), depending on the current state of the chain. Note that after a transition out of such a state, the reduced Markov chain might take a “leap” into a state that the system dynamics would not normally allow. This happens because every time an exiting state is hit, the full chain can exit \mathcal{B}_3 with a certain probability and return to it through another state.

This jump in space observed in the reduced space, is actually caused by an underlying *jump in time*, which is exactly the benefit gained from using the stochastic complement. Since the full chain spends most of its time in a metastable set, by “fast-forwarding” its transitions to the moment it returns to \mathcal{B}_3 , we can make large leaps forward in time. Moreover, we know that ν_3 , differs from $\pi^{(3)}$ only by a (yet unknown) factor.

Thus, the first step of our simulation algorithm simply requires the offline calculation of the stochastic complement R_{33} and the simulation of the reduced Markov chain to sample its stationary distribution. This chain will in general converge much faster than the full chain.

C. Second step

The second step is the approximation of ν_1 and ν_2 . This can be accomplished in two ways: by simulation or computation.

1) *Simulation method:* From the previous step, every state in \mathcal{B}_3 has a weight proportional to its invariant mass. This gives information about the probability flux entering \mathcal{B}_1 and \mathcal{B}_2 at stationarity, namely $f_{31} = \pi^{(3)}Q_{31}$ and $f_{32} = \pi^{(3)}Q_{32}$, which are known up to a constant multiple. These vectors assign a positive weight to each state in $\mathcal{B}_1^{1 \rightarrow 3}$ and $\mathcal{B}_2^{2 \rightarrow 3}$. We can now estimate ν_i , $i = 1, 2$ by picking an initial state in $\mathcal{B}_i^{3 \rightarrow i}$ according to the incoming flux distribution $f_{3i}/\|f_{3i}\|$

and simulating a path of the full chain until it exits \mathcal{B}_i . Since the trajectory segments in \mathcal{B}_i , $i = 1, 2$ are started according to the stationary flux distribution, they are parts of stationary trajectories that cross \mathcal{B}_i . Thus, we can easily approximate ν_i , $i = 1, 2$ by collecting statistics over the generated paths.

2) *Computational method*: The method suggested above may fail due to the extremely stable behavior of the chain in some \mathcal{B}_i , i.e. it is possible that the exit time from a metastable set is too long. In principle, this is not detrimental for the approximation of ν_1 or ν_2 , since a single long trajectory would be enough in this case. However, not all probabilities are calculated with the same accuracy. This means that, while the masses of high-probability states in \mathcal{B}_1 and \mathcal{B}_2 may be very well approximated, the states in $\mathcal{B}_1^{1 \rightarrow 3}$ or $\mathcal{B}_2^{2 \rightarrow 3}$ may be very badly sampled and their estimated masses (albeit very small) may be completely off their real values. In turn, this may have a detrimental effect in the next step of the algorithm, as we shall see below.

In this case it is more effective to consider the very stable set \mathcal{B}_i to be practically closed and modify the matrix Q_{ii} so that it becomes a generator matrix (i.e. it has row sum equal to zero). We thus have to make a redistribution of the exit rate from \mathcal{B}_i to the states within \mathcal{B}_i . This leads to a new matrix Q_i^{cl} , from which we can compute an approximation to ν_i by singular value decomposition, or any other method.

D. Third step

The last step consists of the coupling factors calculation, which will weigh the distributions computed in Steps 1 and 2 correctly against each other, to produce the unconditional stationary distribution over the whole space. Again, there are two ways to carry out this computation:

1) *Using the coupling factors definition*: Equation (4) suggests that the coupling matrix can be calculated from the data obtained so far. The condition $\xi C = 0$ in Theorem 2.2 is nothing more than a systematic way of writing the equilibrium condition: the probability flux into each of the three sets is equal to the probability flux leaving them. In our case, the unknowns in these equations are the constants that multiply the ν_i and make the condition hold, i.e. precisely the coupling factors.

We also observe that by its definition (4), C involves only the states in each \mathcal{B}_i from which a transition to \mathcal{B}_j can be made. This justifies our claim in the previous subsection, that a wrong approximation of their masses could severely throw off the present calculation. However, as the examples will show, this method works and is quite robust against approximation errors.

2) *Approximating the coupling factors*: Combination of (3) with (2) shows that $\xi_i = \pi^{(i)} \mathbf{1}^T$, which is proportional to the fraction of time the chain spends in \mathcal{B}_i . Thus, we could arrive at the coupling factors if we could tell how much time the full chain spent in each of the three sets over the course of a simulation run. Since we are simulating the reduced chain, we cannot get this answer directly.

To proceed with this approach, we need to define $\mathcal{B}'_3 = \mathcal{B}_3 \cup \mathcal{B}_1^{3 \rightarrow 1} \cup \mathcal{B}_2^{3 \rightarrow 2}$ and compute R'_{33} . We are also going to

use the fact that $-Q_{ii}^{-1}(m, n)$ is equal to the average time the chain spends at state $b_i^m \in \mathcal{B}_i$, starting from state $b_i^m \in \mathcal{B}_i$ before hitting \mathcal{B}_i^c [7]. Thus, the average total time spent in \mathcal{B}_i before hitting its complement, starting from state b_i^m , is $-e_m Q_{ii}^{-1} \mathbf{1}^T$.

To approximate the coupling factors, we start by initializing two counters t_1 and t_2 to 0 and simulating a trajectory of length T in \mathcal{B}'_3 . We then trace back this trajectory and locate transitions $\mathcal{B}_3 \rightarrow \mathcal{B}_1^{3 \rightarrow 1}$ and $\mathcal{B}_3 \rightarrow \mathcal{B}_2^{3 \rightarrow 2}$. For each $\mathcal{B}_3 \rightarrow \mathcal{B}_i^{3 \rightarrow i}$ ($i = 1, 2$) transition we locate the target state -say b_i^m -, increment t_i by $-e_m Q_{ii}^{-1} \mathbf{1}^T$ and subtract the time spent at b_i^m from T .

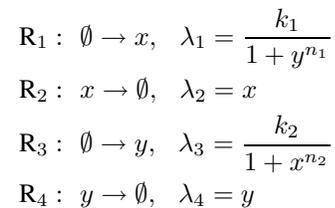
Note that the actual distribution of the hitting time whose average we consider here, is very hard (or impossible) to compute in general, but the use of averages gives very precise results, provided we record many transitions over the course of the simulation.

The above calculations give us t_1 , t_2 and t_3 (the total time spent in \mathcal{B}_3) and the coupling factors can now be computed as $\xi_i = t_i / (t_1 + t_2 + t_3)$.

At first glance, the second method seems less efficient than the first from a computational cost viewpoint, since it involves the post-processing of the generated trajectory. Its advantage, however, is that it does not demand any specific accuracy in the calculation of ν_1 and ν_2 , since their values are not used anywhere. On the other hand, when matrix inversions are computationally expensive (i.e. \mathcal{B}_i are big sets), the second method may not be applicable.

IV. EXAMPLES

To demonstrate its effectiveness, we present the application of our algorithm to two different bistable systems. Both are 2-dimensional, in order to facilitate visual representation of the results, and are based on the genetic toggle switch model of Gardner [10]. Both consist of two mutually repressing genes, X and Y, producing proteins x and y respectively. The chemical reactions and their corresponding propensities are given below:



Based on these equations, we define two models with the following characteristics:

$$\text{Model 1: } k_1 = 50, k_2 = 16, n_1 = 2.5, n_2 = 1$$

$$\text{Model 2: } k_1 = 60, k_2 = 30, n_1 = 3, n_2 = 2$$

Despite its small dimension, Model 1 is already quite hard to simulate to stationarity with SSA, as we shall show below. On the other hand, the parameters of Model 2 were picked in a way that the system switches extremely rarely and is almost impossible to simulate to stationarity.

A. Simulation of Model 1

This system fluctuates most of the time around two metastable regions of the $x - y$ space, which we shall call \mathcal{B}_1 and \mathcal{B}_2 . In \mathcal{B}_1 , x is much higher than y (i.e. gene X is on and Y is off), while in \mathcal{B}_2 the opposite holds. Random fluctuations can drive the system from one region to the other, but the switching is not frequent, as the sample trajectory in Figure 1 suggests:

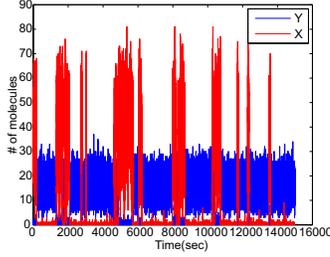


Fig. 1. A sample trajectory of the first toggle switch. After 15000 seconds the system has made only a dozen transitions

Starting from this generated trajectory, we can now locate the boundaries of the two regions. Our goal is to isolate the regions in which most of the simulation time is spent, but of course there is no clear distinction between the metastable and transition regions. For our subsequent calculations we used the following:

$$\mathcal{B}_1 = \{x, y : 0 \leq x \leq 2 \text{ and } 6 \leq y \leq 30\}$$

$$\mathcal{B}_2 = \{x, y : 4 \leq x \leq 70 \text{ and } 0 \leq y \leq 2\}$$

To assess the accuracy of our results, we also calculated a good approximation to the invariant distribution of the system using a large truncation of the state space (130×130 molecules). We arrived at this truncation size starting from a smaller truncated state space and gradually increasing its size, until the resulting invariant distributions ceased to change. Excursions of the system outside this truncation box are extremely rare, so we can safely assume that the invariant distribution we calculated is, for all practical purposes, the true one (Fig. 2). Our simulation approximations will be compared to this distribution.

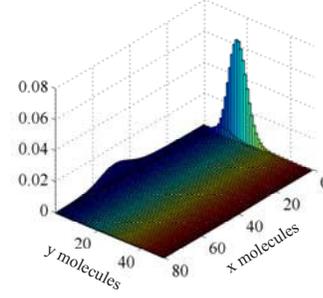
SSA simulation: As mentioned above, Model 1 is still tractable by SSA. Table I contains the approximation error and computation time needed to simulate a trajectory of length T . The numbers reported are the means obtained from 5 runs of the algorithm for each value of T .

T (sec)	$\ \pi - \hat{\pi}\ $	comp.time (sec)
1000	0.3916	16.2
10000	0.1306	2097
50000	0.0665	47620 (13.2h)
80000	0.0491	116800 (32.4h)

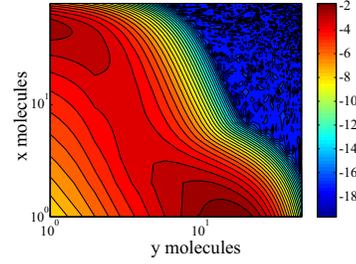
TABLE I

APPROXIMATION RESULTS FROM SSA SIMULATION

We next performed two numerical tests, with the following settings:



(a)



(b)

Fig. 2. (a): invariant distribution of the system computed from a large finite truncation. (b): \log_{10} of the same distribution shown in contour plot in logarithmic scale. The blue region of the plot contains very small elements of the invariant distribution, that reach all the way to the accuracy of Matlab.

- 1) Simulation of ν_1 and ν_2 , calculation of the coupling factors from $\xi \cdot C = 0$
- 2) Simulation of ν_1 and ν_2 , approximation of the coupling factors

All calculations were done using Matlab running on a 2.66 GHz CPU with 4GB of RAM.

1st test: We first varied the length of the trajectory in \mathcal{B}_3 (denoted by t_3), keeping the number of trajectories for estimating ν_1 and ν_2 ($N_{1,2}$) fixed at the (relatively large) value of 10000. Next, we kept the trajectory in \mathcal{B}_3 fixed and examined the influence of the number of iterations performed for the approximation of ν_1 and ν_2 .

The results are listed in Table II. Total computation time (T_C) is reported as the sum of the \mathcal{B}_3 trajectory generation time and the time for computation of $N_{1,2}$ trajectories in each of \mathcal{B}_1 and \mathcal{B}_2 (mean values from 3-5 runs). π represents the truncation-derived stationary distribution and $\hat{\pi}$ the simulation-based approximation.

The results of Table II imply that a \mathcal{B}_3 trajectory 10000 seconds long and 2500 trajectories in each metastable set are enough to give an approximation accuracy comparable to that of our longest SSA runs (the user is of course free to choose the iterations in each metastable set independently, but we kept them equal in this test for simplicity). Indeed, as we verified by simulating with these parameters, the approximation error in this case was $\|\pi - \hat{\pi}\| = 0.0380$ for a total simulation time of $180 + 57 = 237sec$. Including the inverse matrix and stochastic complements calculation does not increase the computation time by more than a couple

1 st test			
t_3 (sec)	$N_{1,2}$	$\ \pi - \hat{\pi}\ $	T_C (sec)
2500	10000	0.1287	9+240=249
5000	10000	0.0392	40+240=280
10000	10000	0.0159	180+240=420
20000	10000	0.0195	600+240=840
30000	10000	0.0160	1600+240=1840
40000	10000	0.0173	3000+240=3240
2 nd test			
t_3 (sec)	$N_{1,2}$	$\ \pi - \hat{\pi}\ $	T_C (sec)
30000	600	0.0802	1600+14=1614
30000	1200	0.0410	1600+27=1627
30000	2500	0.0128	1600+57=1657
30000	5000	0.0198	1600+120=1720
30000	10000	0.0160	1600+240=1840

TABLE II
SIMULATION PARAMETERS AND RESULTS FOR THE FIRST TEST

of seconds in this example. Figure 3 shows a comparison of the resulting approximation with the calculated invariant distribution.

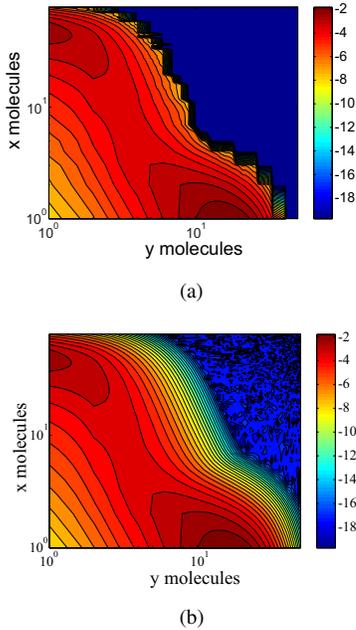


Fig. 3. (a): invariant distribution approximation ($\log_{10}(\hat{\pi})$). (b): calculated invariant distribution ($\log_{10}(\pi)$). The values of $\hat{\pi}$ on the yellow/green/blue region of π are zero, hence the abrupt transition in (a). The invariant mass of those points is so small that it would be computationally impossible to estimate it in a reasonable amount of time. However, the metastable and transition regions, where the bulk of the invariant mass lies, are very well approximated

2nd test: For this test we used the second approach outlined in Section III-D to estimate the average amount of time spent each metastable set. The distributions ν_1 and ν_2 were simulated just like in the first test. Table III shows the accuracy of the results obtained with a few combinations of t_3 and $N_{1,2}$, as well as the average time (t_1 and t_2) that the full process spent in each metastable set over the course of the \mathcal{B}_3 simulation.

t_3 (sec)	$N_{1,2}$	t_1 (sec)	t_2 (sec)	$\ \pi - \hat{\pi}\ $	T_C (sec)
5000	2500	180800	65200	0.0447	40+57=97
10000	2500	335600	139200	0.0175	180+57=237
10000	5000	335600	139200	0.0144	180+120=300
20000	2500	686500	279000	0.0124	600+57=657
20000	5000	686500	279000	0.0087	600+120=720

TABLE III
SIMULATION PARAMETERS AND RESULTS FOR THE SECOND TEST.

The achievable accuracy with this method is better than that of the first test and we were able to use smaller t_3 and $N_{1,2}$ to obtain good results, which implies a decrease in computation time. Moreover, there seems to be more space for improvement than in the first test, where the approximation error seemed to saturate (most probably due to the sensitivity issues we described in Section III-D).

B. Simulation of Model 2

Despite its slow convergence, the system of the previous example could be treated with the SSA, given enough computer time. The toggle switch variant presented in this example is most probably impossible to simulate with SSA, since it switches extremely rarely. Its only difference from the previous one is that the transition rates involve much steeper Hill functions.

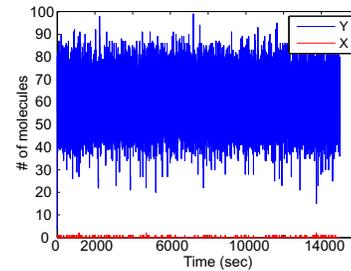


Fig. 4. A sample trajectory of the second toggle switch. After 15000 seconds, no switching has occurred yet. The simulation time for this trajectory was 22000 sec.

Based on trajectories started from different points in the state space, we could identify the two metastable regions:

$$\mathcal{B}_1 = \{x, y : 0 \leq x \leq 2 \text{ and } 5 \leq y \leq 60\}$$

$$\mathcal{B}_2 = \{x, y : 14 \leq x \leq 105 \text{ and } 0 \leq y \leq 2\}$$

The only way to look at the invariant distribution of this system is by state space truncation (since SSA is out of the question). The result of our calculation is displayed in Figure 5

Since this chain can be considered Nearly Completely Decomposable [3], we chose to calculate ν_1 and ν_2 as outlined in Section III-D and performed the following two tests:

- 1) Calculation of the coupling factors from $\xi \cdot C = 0$
- 2) Approximation of the coupling factors

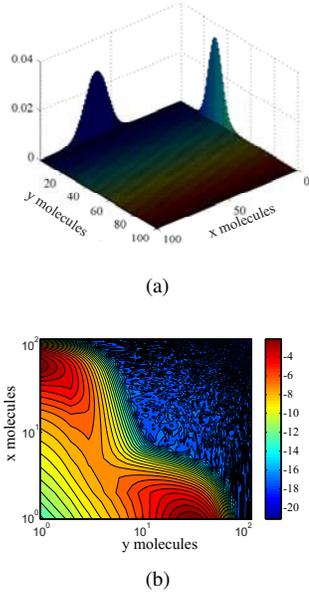


Fig. 5. (a): invariant distribution of the system computed from a large finite truncation. (b): \log_{10} of the same distribution shown in contour plot in logarithmic scale (\log_{10}), to bring out the details in the metastable and transition regions. Clearly, the transition states carry several orders of magnitude less mass compared to the previous example, which implies that transitions between the metastable regions are extremely rare. The transition region is also much narrower in this example

t_3 (sec)	$\ \pi - \hat{\pi}\ $	T_C (sec)
2500	0.0673	23
5000	0.0568	95
10000	0.0215	290
20000	0.0346	1400

TABLE IV
MODEL 2: SIMULATION RESULTS

1st test: In this experiment we varied the length of simulation in \mathcal{B}_3 . The results are displayed on Table IV:

Since the distributions ν_1 and ν_2 are obtained from calculation in this example, a total computation time of around 290 seconds is enough to provide a good approximation to the invariant distribution of a system that is practically impossible to simulate with SSA.

2nd test: The main reason we performed this test was to calculate, based on the approach of Section III-D, the average amount of time spent each metastable set by the full chain. This gives us a rough estimate of the stability of each metastable region, which is indicative of the difficulty to obtain a good SSA approximation of the stationary distribution of the system. The results are displayed in the next table:

t_3 (sec)	t_1 (sec)	t_2 (sec)	$\ \pi - \hat{\pi}\ $	T_C (sec)
2500	1.609×10^8	1.429×10^8	0.0921	23
5000	3.100×10^8	3.000×10^8	0.0688	95
10000	5.633×10^8	6.203×10^8	0.0064	290
20000	1.096×10^9	1.244×10^9	0.0088	1400

As expected, the accuracy of the coupling factors estimates

improves as the \mathcal{B}_3 trajectory gets longer. It seems that less than 5000 seconds are not enough to obtain the right estimate, but the accuracy improves significantly beyond 10000 seconds. The times t_1 and t_2 also demonstrate the extreme stability of this system.

V. DISCUSSION

The variants of the simulation algorithm for metastable Markov chains presented above are based on the theory of stochastic complements and absorbing Markov chains. Its main characteristic is the ability to make large “leaps” in time, using the observed Markov chain on the set of transition states. With this divide-and-conquer approach, a hard simulation task is broken up into a few much easier steps, which take much less time to complete.

This could in turn facilitate the study of multistable biological systems in several ways: for example, stochastic bifurcation analysis can be carried out more efficiently having a fast simulation method. Also, one can calculate the rates of switching among the metastable sets, which is the subject of our future work.

Our methodology can find applications to various other simulation problems as well. One obvious use of it could be rare event simulation. In this case, one just needs to “mask out” transitions in the set where the bulk of the invariant mass of the system lies and concentrate on sampling the rest of the state space. This would require far less computation time than using SSA on the full Markov chain. Moreover, the definitions of coupling factors and the aggregated Markov chain (Theorem 2.2) suggest ways of approximate aggregation of the state space, which have to be explored.

Finally, we should note that the steady-state analysis of metastable systems can be seen as a complement to the transient analysis via the Finite State Projection method [11].

REFERENCES

- [1] A. Bovier, *Metastability*, pp. 177–221. Methods of contemporary mathematical statistical physics, Springer, 2009.
- [2] D. T. Gillespie, “Stochastic simulation of chemical kinetics,” *Annual Review of Physical Chemistry*, vol. 58, no. 1, pp. 35–55, 2007.
- [3] C. D. Meyer, “Stochastic complementation, uncoupling markov chains, and the theory of nearly reducible systems,” *SIAM Review*, vol. 31, no. 2, pp. 240–272, 1989.
- [4] R. J. Allen, P. B. Warren, and P. R. ten Wolde, “Sampling rare switching events in biochemical networks,” *Phys. Rev. Lett.*, vol. 94, no. 1, p. 018104, 2005.
- [5] A. Warmflash, P. Bhimalapuram, and A. Dinner, “Umbrella sampling for nonequilibrium processes,” *The Journal of chemical physics*, vol. 127, p. 154112, 2007.
- [6] D. Freedman, *Approximating countable Markov chains*. Holden-Day, 1972.
- [7] P. Brémaud, *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*. Springer Verlag, 1999.
- [8] Y. Q. Zhao and D. Liu, “The censored markov chain and the best augmentation,” *Journal of Applied Probability*, vol. 33, no. 3, pp. 623–629, 1996.
- [9] S. Tavaré, “A Note on Finite Homogeneous Continuous-Time Markov Chains,” *Biometrics*, vol. 35, pp. 831–834, 1979.
- [10] T. Gardner, C. Cantor, and J. Collins, “Construction of a genetic toggle switch in *Escherichia coli*,” *Nature*, vol. 403, no. 6767, pp. 339–342, 2000.
- [11] B. Munsky and M. Khammash, “Transient analysis of stochastic switches and trajectories with applications to gene regulatory networks,” *IET Systems Biology*, vol. 2, no. 5, pp. 323–333, 2008.