Error Bound for Hill-Function Approximations in a Class of Stochastic Transcriptional Network Models (Extended Version)

Dylan Hirsch[1], Theodore W. Grunberg[2], Domitilla Del Vecchio[3]

***Abstract*— Hill functions are often used in stochastic models of gene regulation to approximate the dependence of gene activity on the concentration of the transcription factor which regulates the gene. It is incompletely known, however, how much error one may incur from this approximation. We investigate this question in the context of transcriptional networks (TN). In particular, under the assumption of rapid binding and unbinding of transcription factors with their gene targets, we bound the approximation error associated with Hill functions for TNs in which each transcription factor regulates a gene in a one-to-one fashion and each regulated gene produces a single transcription factor. We also assume that transcription factors do not homodimerize or heterodimerize and that each gene only has a single transcription factor binding site. These results are pertinent for the modeling of TNs and may also carry relevance for more general biological processes.**

I. INTRODUCTION

A transcriptional network (TN) is a type of chemical reaction network (CRN) which consists of transcription factors (TFs) and genes (Figure 1a), where the genes produce TFs and the TFs regulate the genes' activities, that is the rate at which a gene produces a TF, via reversible binding [1].

In deterministic models of gene regulation, one often exploits for model reduction the separation between the timescale of reversible binding of TFs with genes and the timescale of production and decay of TFs. Per singular perturbation theory, this separation justifies approximation of the TF-gene complex concentration by its steady state value, an approximation referred to as the quasi-steady state approximation (QSSA) [2]–[4]. One can use this QSSA to determine how the activity of a gene depends on the total concentration of the TF that regulates it. When the binding strength between TFs and genes is sufficiently weak, this dependence takes the form of a Hill function [3].

Hill functions are also often used to approximate this same dependence in the stochastic setting. Previous work, however, has demonstrated that such an approximation can result in significant errors in the number of TFs produced by a gene, even when there is a separation of timescales between reversible binding and TF production and decay [5]–[7].

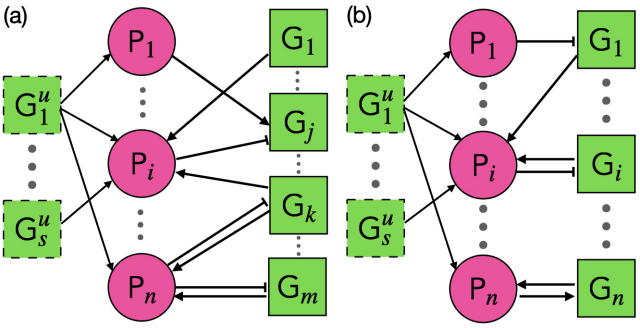

Fig. 1. Visualization of most general TNs via bipartite directed graphs. (a) A graph of a general TN. An edge from gene G to TF P indicates that G produces P. An edge from a TF P to a gene G indicates that P regulates G by changing the rate at which G produces TFs. An "→" arrow from P to G indicates that G produces more TF (positive regulation) when bound than unbound, and a "⊣" arrow indicates the opposite (negative regulation). Unregulated genes, those that are not regulated by any TF, are shown with dashed borders. (b) A graph of the special class of TNs that we consider in this work. Unlike Figure 1a, each TF regulates at most one gene, each regulated gene produces only one TF, and no two different TFs regulate the same gene.

Another widely used approach for reducing stochastic models of CRNs with slow and fast reactions is to apply a stochastic analogue of the QSSA, called the stochastic QSSA (sQSSA) [8]–[12]. The sQSSA is the approximation that immediately after a slow reaction of the CRN occurs, the fast reactions of the CRN fire enough times to make the probability distribution of the CRN close to the stationary distribution of a "virtual" CRN which only includes these fast reactions [8]. For some CRNs (such as those with a finite number of states), one can use singular perturbation theory or related techniques to prove that the sQSSA produces an accurate reduced model for sufficiently rapid fast reactions [13]–[18]. However, the propensity functions of the reduced model are typically complicated in form [19]–[21]. As a result, for situations such as CRN design, working with the Hill function models may be more convenient thanks to their simpler propensity functions.

In [5,6,22], the authors show that in the limit of weak binding, models of various CRNs using the Hill function (or in [22] a modified Michaelis-Menten function) are accurate when timescale separation between reversible binding and protein production is sufficiently large. It is still not clear, however, how much error one may incur using Hill functions in TN models that do not assume weak binding of TF to the target genes.

This material is based upon work supported in part by the Air Force Office of Scientific Research Grants No. FA9550-22-1-0316 and FA9550-14-1-0060, and the NSF Graduate Research Fellowship under Grant No. 1745302

[1] Dylan Hirsch is with the Department of Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA dylanhir@mit.edu

[2] Theodore W. Grunberg is with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA tgrunber@mit.edu

[3] Domitilla Del Vecchio is with the Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA ddv@mit.edu

In this work, we provide for a restricted class of TNs an explicit bound on the error (measured in total variation distance) between the distributions of the full model and a reduced model using Hill functions, in the limit of rapid binding and unbinding of TFs with genes, without assuming weak binding. The class of TNs we investigate are those of the form in Figure 1b, namely those in which each TF can only regulate the activity of one gene, each regulated gene can only produce a single TF, and no two TFs regulate the same gene. We also assume the TFs do not homodimerize or heterodimerize and that each gene has only one TF binding site. To obtain our results, we first show the sQSSA is indeed accurate for this class of TNs, and thereafter we apply facts about the moment dynamics of CRNs [23] and results on the finite state projection algorithm [24,25] to the reduced model as to derive our error bound.

II. MATHEMATICAL BACKGROUND

We use $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Z}_+$, and $\mathbb{R}$ to respectively denote the set of natural numbers (excluding 0), integers, non-negative integers, and real numbers. We use $\mathbb{Z}_+^n$ ($\mathbb{Z}^n$) to denote row vectors with n non-negative integer (integer) entries. We define $\mathbb{1}_S(x)$ to equal 1 if $x \in S$ and 0 otherwise.

A. Continuous Time Markov Chains

A (minimal) Continuous Time Markov Chain (CTMC) is a random process $Y = \{Y(t)\}_{t\geq 0}$ whose sample paths at each time $t > 0$ take a value in some countable set $\mathcal{Y} \cup \{\infty\}$ with the property that the probability the process will be in a given state at a given future time is entirely determined by its present state [26]. $\mathcal{Y}$ is called the *state space* of Y.

The jump chain $\tilde{Y}_0, \tilde{Y}_1, \ldots$ of Y is the sequence of states which Y visits, with $\tilde{Y}_0$ being the initial state. The amount of time Y spends in state $\tilde{Y}^{k-1}$ between the $(k-1)^{\text{th}}$ and k^{th} state transition is the k^{th} holding time, S_k. The *explosion time* of Y is defined by $T_Y^\infty := \sum_{k=1}^\infty S_k$. Following its explosion time (and only following this time), Y is defined to be in the special state ∞ (this is the meaning of *minimal*). Y is *nonexplosive* if $T_Y^\infty = \infty$ almost surely.

In agreement with the assumptions of the references [25]–[27] we invoke in our proofs, all CTMCs in this work are assumed to be minimal and have right-continuous trajectories with finite left-limits (again see [26] for details).

B. Infinitesimal Generators and Stationary Distributions

An *infinitesimal generator* is a function $Q : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ such that for each $y, y' \in \mathcal{Y}$, $0 \leq -Q(y,y) < \infty$, $y \neq y'$ implies $Q(y,y') \geq 0$, and $\sum_{y'' \in \mathcal{Y}} Q(y,y'') = 0$. Each minimal CTMC Y with state space $\mathcal{Y}$ is associated with an infinitesimal generator $Q : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ that completely specifies the likelihood the CTMC will have a given state at a given future time [26].

Two states y, y' in $\mathcal{Y}$ *communicate* (with respect to Q) if there is a nonzero probability that a CTMC with infinitesimal generator Q and initial state y will have state y' at some point in the future and *vice versa*. Communication is an equivalence relation on the set Y, partitioning it into *communicating classes* (of Q). A communicating class C is *closed* if the restriction of Q to the domain $C \times C$ is itself an infinitesimal generator (i.e. if once in C the CTMC cannot leave this set). A *stationary distribution* of Q is a probability distribution $\pi : \mathcal{Y} \to [0,1]$ such that for all $y \in \mathcal{Y}$, $\sum_{y' \in \mathcal{Y}} \pi(y') Q(y', y) = 0$.

C. Stochastic CRNs

We define a *stochastic CRN* (SCRN) $\mathcal{S}$ to be an (ordered) set of chemical species $\mathrm{S}_1, \ldots, \mathrm{S}_n$, a state space $\mathcal{Y} \subset \mathbb{Z}_+^n$, and a finite collection of reactions (indexed by r) of the form $\sum_{i=1}^n a_{r,i}\mathrm{S}_i \to \sum_{i=1}^n b_{r,i}\mathrm{S}_i$ ($a_{r,i}, b_{r,i} \in \mathbb{Z}_+$), each with a *propensity function* $\upsilon_r : \mathcal{Y} \to [0,\infty)$ and *stoichiometry vector* $u_r := (b_{r,1} - a_{r,1}, \ldots, b_{r,n} - a_{r,n})$, such that for each r, $u_r \neq 0$, and for each $y \in \mathcal{Y}$, if $y + u_r \notin \mathcal{Y}$ then $\upsilon_r(x) = 0$.

For each $y_0 \in \mathcal{Y}$, $\mathcal{S}$ has a naturally associated CTMC Y with state space $\mathcal{Y}$, initial state y_0, and infinitesimal generator $Q : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ given by

$$Q(y,y') = \sum\nolimits_r \upsilon_r(y) \mathbb{1}_{\{y'\}}(y + u_r),$$

for $y \neq y'$. We define this infinitesimal generator to be the one *determined by* the set of stoichiometry vector-propensity pairs $\{(u_r, \upsilon_r)\}_r$. We call Q the infinitesimal generator of $\mathcal{S}$ and refer to the communicating classes and stationary distributions of $\mathcal{S}$ as those inherited from Q.

D. Total Variation Distance

Given two random variables Y and Y', both taking values in $\mathcal{Y}$, we define their total variation distance (TVD) to be $d_{TV}(Y,Y') = \frac{1}{2}\sum_{y \in \mathcal{Y}} |\mathbb{P}(Y = y) - \mathbb{P}(Y' = y)|$.

III. PROBLEM SETTING

We first describe the "full" SCRN corresponding to the system that we wish to approximate, specifically a TN of the form visualized in Figure 1b. We next define the "reduced" SCRN corresponding to the reduction of the full SCRN via the sQSSA. Finally, we introduce the "Hill" SCRN which has the same reactions and chemical species as the reduced SCRN, but uses Hill functions in the propensities. Our goal is to bound the TVD between the CTMCs associated with the full SCRN and the Hill SCRN, under the assumption of rapid binding and unbinding of genes and TFs. We do so by first showing the TVD between the CTMCs associated with the full and reduced SCRNs vanishes in the limit of rapid reversible binding. We then bound the TVD between the CTMCs corresponding to the reduced and Hill SCRNs.

A. Full SCRN

Fix $n \in \mathbb{N}$, and for each $i \in \{1, \ldots, n\}$ fix some $\bar{g}_i \in \mathbb{Z}_+$ representing the total number of copies of gene i. Define $\mathcal{X} := \{(p_1, \ldots, p_n, c_1, \ldots, c_n, g_1, \ldots, g_n) \in \mathbb{Z}_+^{3n} : \forall i \; g_i + c_i = \bar{g}_i\}$. Define $\bar{\mathcal{X}} := \mathbb{Z}_+^n$.

For convenience, throughout the rest of this text, the variable x always represents a vector of variables $(p_1, \ldots, p_n, g_1, \ldots, g_n, c_1, \ldots, c_n)$ which belongs to $\mathcal{X}$, and the variable $\bar{x}$ always represents a vector of variables $(\bar{p}_1, \ldots, \bar{p}_n)$ which belongs to $\bar{\mathcal{X}}$. With this notation in mind,

let $\mathbf{T} : \mathcal{X} \to \bar{\mathcal{X}}$ be defined by $\mathbf{T}(x) \mapsto (p_1 + c_1, \ldots, p_n + c_n)$. Conceptually, $\mathbf{T}(x)$ represents the total (free + bound) number of TFs of each species associated with state x.

For each $i \in \{1, \ldots, n\}$, let f_i and b_i be positive scalars, and let α_i, β_i, κ_i, and γ_i be non-negative. Let $\Omega = V_C N_A > 0$, where V_C represents the cell volume and N_A is the Avogadro constant. Fix $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$, which represents the map from each gene G_i to the TF $\mathrm{P}_{\sigma(i)}$ which G_i produces.

Now, for each $\epsilon > 0$, consider the SCRN $\mathcal{S}_\epsilon$ with chemical species $\mathrm{P}_1, \ldots, \mathrm{P}_n, \mathrm{G}_1, \ldots, \mathrm{G}_n, \mathrm{C}_1, \ldots, \mathrm{C}_n$, state space $\mathcal{X}$, and the following reaction-propensity pairs described by mass-action kinetics for all $i \in \{1, \ldots, n\}$:

$$\mathrm{P}_i + \mathrm{G}_i \to \mathrm{C}_i,\ \upsilon_{f,i,\epsilon}(x) = \frac{1}{\epsilon}\frac{f_i}{\Omega} p_i g_i; \tag{1a}$$

$$\mathrm{C}_i \to \mathrm{P}_i + \mathrm{G}_i,\ \upsilon_{b,i,\epsilon}(x) = \frac{1}{\epsilon} b_i c_i; \tag{1b}$$

$$\mathrm{C}_i \to \mathrm{C}_i + \mathrm{P}_{\sigma(i)},\ \upsilon_{\alpha,i}(x) = \alpha_i c_i; \tag{1c}$$

$$\mathrm{G}_i \to \mathrm{G}_i + \mathrm{P}_{\sigma(i)},\ \upsilon_{\beta,i}(x) = \beta_i g_i; \tag{1d}$$

$$\emptyset \to \mathrm{P}_i,\ \upsilon_{\kappa,i}(x) = \Omega\kappa_i; \tag{1e}$$

$$\mathrm{P}_i \to \emptyset,\ \upsilon_{\gamma,i}(x) = \gamma_i p_i; \tag{1f}$$

$$\mathrm{C}_i \to \mathrm{G}_i,\ \upsilon_{\gamma',i}(x) = \gamma_i c_i. \tag{1g}$$

Note that for notational simplicity, we do not explicitly include the unregulated genes in the above model. We instead use (1e) to model the production of TF P_i from the set of unregulated genes. With reference to Figure 1(b), if we denote the unregulated genes in the TN by $\mathrm{G}_1^u, \ldots, \mathrm{G}_s^u$ with respective copy numbers $\bar{g}_1^u, \ldots, \bar{g}_n^u$, then $\kappa_i = \sum_{r=1}^{s} \kappa_{i,r} \bar{g}_r^u$, where $\kappa_{i,r}$ is the molar production rate of P_i per copy of gene G_r^u. Note also that ϵ is inversely related to the speed of the binding and unbinding reactions, so we will ultimately be interested in the behavior of this SCRN as ϵ approaches 0.

Fix $x_0 \in \mathcal{X}$. For each $\epsilon > 0$, let X_ϵ be the CTMC associated with $\mathcal{S}$ having initial state x_0.

B. Reduced SCRN

We first define a "fast subsystem" SCRN $\mathcal{R}$, with chemical species $\mathrm{P}_1, \ldots, \mathrm{P}_n, \mathrm{G}_1, \ldots, \mathrm{G}_n, \mathrm{C}_1, \ldots, \mathrm{C}_n$, state space $\mathcal{X}$, and the following reactions and propensities for all $i \in \{1, \ldots, n\}$:

$$\mathrm{P}_i + \mathrm{G}_i \to \mathrm{C}_i,\ \upsilon_{f,i}(x) = \frac{f_i}{\Omega} p_i g_i;$$

$$\mathrm{C}_i \to \mathrm{P}_i + \mathrm{G}_i,\ \upsilon_{b,i}(x) = b_i c_i.$$

Thus $\mathcal{R}$ consists of only the binding and unbinding reactions of the full network.

For each $\bar{x} \in \bar{\mathcal{X}}$, let $E_{\bar{x}} = \{x \in \mathcal{X} : \mathbf{T}x = \bar{x}\}$. In other words, if we write $\bar{x}$ as $(\bar{p}_1, \ldots, \bar{p}_n)$, $E_{\bar{x}}$ is the set of states in $\mathcal{X}$ where the total (free + bound) count of each TF species i is $\bar{p}_i$. $\mathcal{E} := \{E_{\bar{x}} : \bar{x} \in \bar{\mathcal{X}}\}$ is the collection of communicating classes of $\mathcal{R}$. Since each such communicating class is finite and closed (binding and unbinding does not change total TF counts), for each $\bar{x}$ there is a unique stationary distribution $\pi_{\bar{x}} : \mathcal{X} \to [0, 1]$ of $\mathcal{R}$ which is supported on $E_{\bar{x}}$ (see Theorem 3.5.2 in [26]).

We define the reduced SCRN $\bar{\mathcal{S}}$ as having chemical species $\bar{\mathrm{P}}_1, \ldots, \bar{\mathrm{P}}_n$, state space $\bar{\mathcal{X}}$, and the following reaction-propensity pairs (for all $i \in \{1, \ldots, n\}$):

$$\bar{\mathrm{P}}_i \to \bar{\mathrm{P}}_i + \bar{\mathrm{P}}_{\sigma(i)},\ \bar{\upsilon}_{+,i}(\bar{x}) = \alpha_i \mathbb{E}_{W \sim \pi_{\bar{x}}}[C_i] + \beta_i \mathbb{E}_{W \sim \pi_{\bar{x}}}[G_i]; \tag{2a}$$

$$\emptyset \to \bar{\mathrm{P}}_i,\ \bar{\upsilon}_{+',i}(\bar{x}) = \Omega\kappa_i; \tag{2b}$$

$$\bar{\mathrm{P}}_i \to \emptyset,\ \bar{\upsilon}_{-,i}(\bar{x}) = \gamma_i \bar{p}_i, \tag{2c}$$

where $W = (P_1, \ldots, P_n, G_1, \ldots, G_n, C_1, \ldots, C_n)$ is a vector-valued random variable. Note that this SCRN only involves the TF species, $\bar{\mathrm{P}}_i$, without distinction between whether they are free or bound (we use the overbar to emphasize this difference from P_i, which refers to the free TF).

Remark 1. *Note that the propensity functions $\bar{\upsilon}_{+,i}(\bar{x})$ are weighted averages of $\upsilon_{\alpha,i}(x) + \upsilon_{\beta,i}(x)$ over $x \in E_{\bar{x}}$, with the weights corresponding to the stationary distribution of the fast subsystem $\mathcal{R}$ that is supported on $E_{\bar{x}}$. Similarly, $\bar{\upsilon}_{+',i}(\bar{x})$ is the weighted average of $\upsilon_{\kappa,i}(x)$, and $\bar{\upsilon}_{-,i}(\bar{x})$ is the weighted average of $\upsilon_{\gamma,i}(x) + \upsilon_{\gamma',i}(x)$. In light of these facts, this SCRN may be viewed as the reduced model for the full SCRN via application of the sQSSA (see [8] for more details), where we make a change of coordinates so that we only keep track of total TF (a "slow" species).*

Let $\bar{X}$ be the CTMC associated with $\bar{\mathcal{S}}$ having initial state $\bar{x}_0 := \mathbf{T}x_0$.

C. Hill SCRN

We define another SCRN $\bar{\mathcal{S}}_H$ with chemical species $\bar{\mathrm{P}}_1, \ldots, \bar{\mathrm{P}}_n$, state space $\bar{\mathcal{X}}$, and the following reaction-propensity pairs (for all $i \in \{1, \ldots, n\}$):

$$\bar{\mathrm{P}}_i \to \bar{\mathrm{P}}_i + \bar{\mathrm{P}}_{\sigma(i)},\ \bar{\upsilon}^H_{+,i}(\bar{x}) = \frac{\alpha_i \bar{g}_i \bar{p}_i/\Omega}{\bar{p}_i/\Omega + K_i} + \frac{\beta_i \bar{g}_i K_i}{\bar{p}_i/\Omega + K_i}; \tag{3a}$$

$$\emptyset \to \bar{\mathrm{P}}_i,\ \bar{\upsilon}^H_{+',i}(\bar{x}) = \Omega\kappa_i; \tag{3b}$$

$$\bar{\mathrm{P}}_i \to \emptyset,\ \bar{\upsilon}^H_{-,i}(\bar{x}) = \gamma_i \bar{p}_i, \tag{3c}$$

where for each i, $K_i := {}^{b_i}/_{f_i}$. The expressions $\alpha_i \bar{g}_i \frac{\bar{p}_i/\Omega}{\bar{p}_i/\Omega + K_i}$ and $\beta_i \bar{g}_i \frac{K_i}{\bar{p}_i/\Omega + K_i}$ are the Hill functions for the TF-bound and TF-unbound gene activites, respectively.

Let $\bar{X}_H$ be the CTMC associated with $\bar{\mathcal{S}}_H$ having initial state $\bar{x}_0$.

IV. MAIN RESULTS

Our first lemma states that our CTMCs are nonexplosive. The next lemma states that the joint probability distributions of total TF counts of each species at a given time are identical between the full SCRN and the reduced SCRN in the limit of rapid binding and unbinding between TFs and genes. In other words, the sQSSA indeed produces an accurate reduced model when ϵ is small.

Lemma 1. *For each $\epsilon > 0$, X_ϵ, $\bar{X}$, and $\bar{X}_H$ are nonexplosive.*

Proof. See Appendix A. □

Remark 2. *To interpret the following results, it is helpful to write* $X_\epsilon(t)$ *as* $(P_{\epsilon,1}(t),\dots,P_{\epsilon,n}(t), G_{\epsilon,1}(t),\dots,G_{\epsilon,n}(t), C_{\epsilon,1}(t),\dots,C_{\epsilon,n}(t))$, $\bar{X}(t)$ *as* $(\bar{P}_1(t),\dots,\bar{P}_n(t))$, *and* $\bar{X}_H(t)$ *as* $(\bar{P}_1^H(t),\dots,\bar{P}_n^H(t))$. *Then* $\mathbf{T}X_\epsilon(t) = (P_{\epsilon,1}(t) + C_{\epsilon,1}(t),\dots,P_{\epsilon,n}(t) + C_{\epsilon,n}(t))$ *represents the total counts of each TF in the full system. Similarly,* $\bar{X}(t)$ *and* $\bar{X}_H(t)$ *represent the total TF counts in the reduced SCRN and Hill SCRN, respectively.*

Lemma 2. *For all* $t \geq 0$,

$$\lim_{\epsilon^+\to 0} d_{TV}(\mathbf{T}X_\epsilon(t), \bar{X}(t)) = 0.$$

Proof. See Appendix B. □

The following theorem, which is our main result, bounds the error between the full SCRN and the Hill approximation-based SCRN in the same rapid binding and unbinding limit.

Theorem 1. *For all* $t \geq 0$,

$$\lim_{\epsilon^+\to 0} d_{TV}(\mathbf{T}X_\epsilon(t), \bar{X}_H(t)) \leq (n - n_{01})t \max_i \bar{g}_i|\alpha_i - \beta_i| \frac{\bar{g}_i/\Omega}{\bar{g}_i/\Omega + K_i}, \quad (4)$$

where n_{01} *is the number of* i *such that* $\bar{g}_i \in \{0, 1\}$.

Proof. See Appendix C. □

Our last result states that if each gene only has a single copy, the Hill approximation error is in fact zero. This result may be seen either as a corollary of the above, or as an application of the results of [19] (section 4.1.3) to our system.

Corollary 1. *Suppose that for each* i, $\bar{g}_i \in \{0, 1\}$. *Then, for all* $t \geq 0$,

$$\lim_{\epsilon^+\to 0} d_{TV}(\mathbf{T}X_\epsilon(t), \bar{X}_H(t)) = 0.$$

V. EXAMPLES

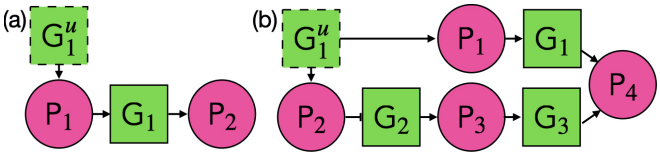

Fig. 2. Example TNs. (a) Simple regulation network. (b) Incoherent Feedforward network.

A. Example 1: Simple positive regulation

For our first example, we investigate a simple TN in which the constituitively expressed TF P_1 positively regulates production of P_2 (Figure 2a). We define the full SCRN by (1a)-(1g), with $n = 2$ and $\sigma(1) = 2$. Since P_2 does not regulate a gene, $\bar{g}_2 = 0$. The reduced SCRN and the Hill SCRN are respectively given by (2a)-(2c) and (3a)-(3c). The probability distributions for each SCRN were computed and compared in TVD (Figure 2a). We observe that the TVD bound (4) is conservative by approximately an order of magnitude. In line with the findings of [5,6] the marginal probability distribution of the downstream TF can be significantly different between the full and Hill approximations when $\bar{g}_1$ is not much smaller than ΩK.

Remark 3. *At the cost of simplicity, the TVD bound* (4) *can be made less conservative by computing the supremal error between the propensity functions in the reduced system and corresponding propensities in the Hill approximation (see Proof of Theorem 1).*

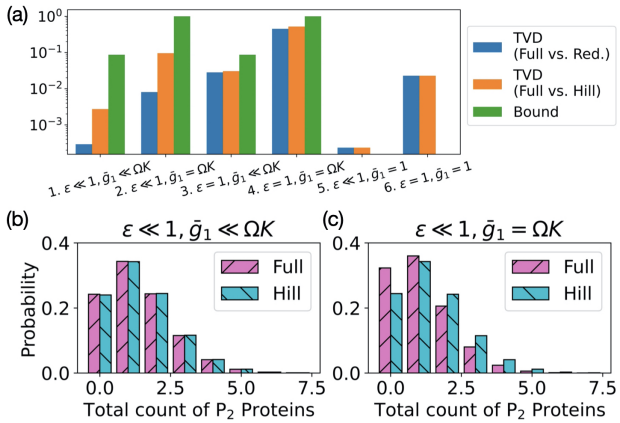

Fig. 3. (a) Simulation results for simple positive regulation for six cases: 1. $\bar{g}_1 = 3$, $\epsilon = 0.01$, $b_1 = 250$/hr; 2. $\bar{g}_1 = 3$, $\epsilon = 0.01$, $b_1 = 3$/hr; 3. $\bar{g}_1 = 3$, $\epsilon = 1$, $b_1 = 100$/hr; 4. $\bar{g}_1 = 3$, $\epsilon = 1$, $b_1 = 3$/hr; 5. $\bar{g}_1 = 1$, $\epsilon = 0.01$, $b_1 = 100$/hr ; 6. $\bar{g}_1 = 1$, $\epsilon = 1$, $b_1 = 250$/hr. In each case, $t = .1$hr, $\Omega = N_A \cdot 1\mu\text{m}^3$, $f_1 = 1\Omega$/hr, $\gamma_1 = \gamma_2 = 1$/hr, $\kappa_2 = 0$/hr/Ω, $\alpha_1 = 10$/hr, $\beta_1 = 0$/hr, $\bar{g}_2 = 0$, and $\kappa_1 = K$, where $K := f_1/b_1/\Omega$. We let the initial state be $x_0 := (p_{1,0}, p_{2,0}) = (\Omega K, 0)$. TVD is compared between the full SCRN ("Full") and both the reduced SCRN ("Red.") and the Hill SCRN ("Hill"). The TVD error bound ("Bound") from (4) is also shown. Note this bound is only guaranteed to hold in the limit of small ϵ. Probabilities are approximated via the Finite State Projection Algorithm [24], with a guaranteed absolute computation error of less than 10^{-4}. (b) Marginal distributions of P_2 count in the full and Hill SCRNs for case 1. (c) Marginal distributions of P_2 count in full and Hill SCRNs for case 2.

B. Example 2: Incoherent Feedforward-Like Network

In the next example, we investigate the Hill approximation for a network modeling an Incoherent Feedforward Loop (IFFL). Unlike a standard IFFL, however, in this TN the output TF P_4 is produced by two different genes, each with its own regulator, rather than a single gene that has two interacting regulators. In particular, the network has direct positive regulation from P_1 to P_4 along with indirect negative regulation from P_2 to P_4 via P_3 (Figure 2b). The full, reduced, and Hill SCRNs for this circuit are defined respectively by (1a)-(1g), (2a)-(2c), (3a)-(3c), with $n = 4$ and σ given by $\sigma(1) = 4$, $\sigma(2) = 3$, $\sigma(3) = 4$.

When properly parameterized, this TN displays robustness of steady-state mean levels of P_4 to disturbances in the activity of upstream gene $\bar{G}_1^u$ (Figure 4a). The transient period during which the adaptation occurs is shown in Figure 4b. One can use Theorem 1 to investigate whether the Hill function model reliably reproduces the probability distribution of the more complicated reduced model during this transient period. In particular, the TVD error bound at $t = 1$ hr is below 10% (Figure 4c).

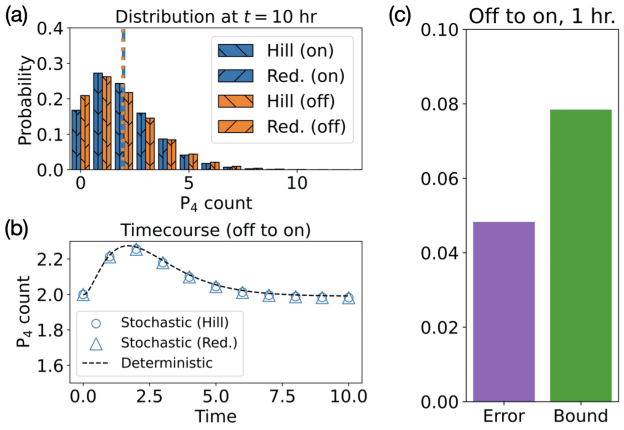

Fig. 4. Results from simulations of the system in Figure 2b, in the "on" state (P_1 and P_2 are produced by $\bar{G}_1^u$) and "off" state (P_1 and P_2 are not produced). In both cases, the system is initialized in the state $(p_1, p_2, p_3, p_4) = (0, 0, 2, 2)$. For the on state, $\kappa_1 = \kappa_2 = 100/\Omega/hr$, and for the off state, $\kappa_1 = \kappa_2 = 0/\Omega/hr$. The following other parameters were used: $\bar{g}_1 = \bar{g}_2 = 2$, $\bar{g}_3 = 1$, $\bar{g}_4 = 0$, $b_1 = b_2 = b_3 = 100/hr$, $f_1 = f_2 = f_3 = 1\Omega/hr$, $\alpha_1 = 1/hr$, $\beta_2 = 1/hr$, $\alpha_3 = 100/hr$, $\gamma_1 = \gamma_2 = 1/hr$, $\Omega = 1\mu m^3$, with all other parameters set to 0. (a) Probability distributions of the downstream TF P_4 at 10 hours, for the on and off states, simulated using for the reduced SCRN ("Red.") and Hill SCRN. For both the on and off states, the reduced and Hill SCRNs are almost identical. The means of the distributions are shown via dotted lines (they are very close). (b) Timecourse of mean counts of P_4 in the "on" state, with initialization in the (deterministic) "off" steady state, for both the reduced system and Hill approximation. The deterministic Hill approximation trajectory is shown for reference. (c) Simulated TVD for the on-state approximation at 1 hr. (from 10^7 simulations performed according to [28]; the numerical result likely overestimates because TVD is sensitive to noise in the empirical distributions), along with the error bound (4).

VI. CONCLUSION

In this work, we derived a bound on the TVD error incurred when using a Hill approximation for stochastic modeling of a class of TNs in the limit of rapid reversible binding of TFs to their target genes. This bound is non-asymptotic in all model parameters other than ϵ, which scales the reversible binding speeds. These results can be applied to understand how different results derived using reduced models of TNs that use Hill functions may be from the full model when reversible binding is indeed sufficiently fast. Possible further work in this pursuit includes deriving similar bounds when the same TF may bind to multiple different genes, different TFs to compete for the same gene, and TFs may homodimerize and heterodimerize.

APPENDIX

In the theorem proofs we use the following notation. Given a CTMC Y with state space $\mathcal{Y}$, for all $S \subset \mathcal{Y}$ we let $\tau_{Y,S} = \inf\{0 \leq t < T_Y^\infty : Y(t) \notin S\}$ represent the exit time of Y from the subset S, where $\inf \emptyset = +\infty$. We also define $\rho_Y : \mathcal{Y} \times [0, \infty) \to [0, 1]$ by $\rho_Y(y, t) = \mathbb{P}(Y(t) = y, t < T_Y^\infty)$. Given $t \in [0, \infty)$, we let $\rho_Y(\cdot, t) : \mathcal{Y} \to \mathbb{R}$ be the map that sends y to $\rho_Y(y, t)$. If Y is nonexplosive, $\rho_Y(\cdot, t)$ is the probability density of Y at time t (ignoring the special state ∞, which is irrelevant since Y is nonexplosive). Given a function $f : U \to \mathbb{R}$ and $V \subset U$, we let $f|_V : V \to \mathbb{R}$ be the restriction of f to the domain V. We also let $\|f\|_1 = \sum_{u \in U} |f(u)|$ be the 1-norm of f and if $F : U \times U \to \mathbb{R}$, we let $\|F\|_1 = \sup_{u \in U} \sum_{u' \in U} |F(u, u')|$ be the induced 1-norm of F.

A. Proof of Lemma 1

Proposition 1. *Let Y be a CTMC with state space $\mathcal{Y}$, infinitesimal generator Q, and explosion time T_Y^∞. Let $\tilde{Y}_0, \tilde{Y}_1, \ldots$ be the jump chain of Y. Suppose that there exist $c_1, c_2 > 0$ such that for all $k \in \mathbb{Z}_+$ and each $y_0, \ldots, y_k \in \mathcal{Y}$ such that $\mathbb{P}(\tilde{Y}_0 = y_0, \ldots, \tilde{Y}_k = y_k) \neq 0$, $|Q(y_k, y_k)| \leq c_1 k + c_2$. Then Y is nonexplosive.*

Proof. Fix $t \geq 0$. Denote by $S_1, S_2, \ldots$ the holding times of Y. Since $T_Y^\infty = \sum_{i=1}^\infty S_i$ by definition, for each $k \in \mathbb{Z}_+$,

$$\begin{aligned}\mathbb{P}(T_Y^\infty \leq t) &\leq \mathbb{P}(\sum_{i=1}^{k+1} S_i \leq t) \\ &= \sum_{y_0, \ldots y_k \in \mathcal{Y}} \mathbb{P}(\sum_{i=1}^{k+1} S_i \leq t | \tilde{Y}_0 = y_0, \ldots, \tilde{Y}_k = y_k) \\ &\qquad \times \mathbb{P}(\tilde{Y}_0 = y_0, \ldots, \tilde{Y}_k = y_k). \qquad (5)\end{aligned}$$

Let $E_0^*, E_1^*, E_2^* \ldots$ be independent exponentially distributed random variables, with E_i^* having rate parameter $c_1 i + c_2$. Choose $k \in \mathbb{Z}_+$ and $y_0, \ldots, y_k \in \mathcal{Y}$ such that $\mathbb{P}(\tilde{Y}_0 = y_0, \ldots, \tilde{Y}_k = y_k) \neq 0$. By the Markov property, given $\tilde{Y}_0 = y_0, \ldots, \tilde{Y}_k = y_k$, the collection of random variables $S_1, \ldots, S_{k+1}$ are independent exponentially distributed random variables with respective rate parameters $|Q(y_0, y_0)|, \ldots, |Q(y_k, y_k)|$ (See Chapter 2.6 in [26]). Thus

$$\mathbb{P}(\sum_{i=1}^{k+1} S_i \leq t | \tilde{Y}_0 = y_0, \ldots, \tilde{Y}_k = y_k) = \mathbb{P}(\sum_{i=0}^{k} E_i \leq t), \quad (6)$$

where the E_i are a collection of independent exponential random variables, each with rate parameter $|Q(y_i, y_i)|$.

Since for each $i \in \{0, \ldots, k\}$, $|Q(y_i, y_i)| \leq c_1 i + c_2$ by assumption,

$$\mathbb{P}(\sum_{i=0}^{k} E_i \leq t) \leq \mathbb{P}(\sum_{i=0}^{k} E_i^* \leq t). \qquad (7)$$

Putting (5), (6), and (7) together then gives that for each k, $\mathbb{P}(T_Y^\infty \leq t) \leq \mathbb{P}(\sum_{i=0}^k E_i^* \leq t)$.

However, by Theorem 2.3.2(ii) in [26], $\mathbb{P}(\sum_{i=0}^\infty E_i^* \leq t) = 0$. Since $\lim_{k \to \infty} \mathbb{P}(\sum_{i=0}^k E_i^* \leq t) = \mathbb{P}(\sum_{i=0}^\infty E_i^* \leq t) = 0$, we have that $\mathbb{P}(T_Y^\infty \leq t) = 0$. As $t \geq 0$ was arbitrary,

$$\begin{aligned}\mathbb{P}(T_Y^\infty < \infty) &= \mathbb{P}(\cup_{t=1}^\infty \{T_Y^\infty \leq t\}) \\ &\leq \sum_{t=1}^\infty \mathbb{P}(T_Y^\infty \leq t) = 0.\end{aligned}$$

□

To prove Lemma 1, we apply Proposition 1 to each CTMC:

Proof of Lemma 1.

- Fix $\epsilon > 0$, let $x_0 := (p_{1,0}, \ldots, p_{n,0}, g_{1,0}, \ldots, g_{n,0}, c_{1,0}, \ldots, c_{n,0})$, and let Q_ϵ be the infinitesimal generator of X_ϵ. By inspection, for any state y that X_ϵ can access within k transitions, $|Q(y,y)| \leq n(\gamma_*(p_{*,0} + c_{*,0} + k) + \kappa_* + (\frac{b_*}{\epsilon} + \frac{f_*}{\epsilon}(p_{*,0} + c_{*,0} + k))\bar{g}_*) + n^2(\alpha_* + \beta_*)\bar{g}_*$, where the subscript $*$ refers to the maximal value of the relevant variable over all $i \in \{1, \ldots, n\}$.
- Let $\bar{x}_0 = (\bar{p}_{1,0}, \ldots, \bar{p}_{n,0})$ and let $\bar{Q}$ be the infinitesimal generator of $\bar{X}$. From inspection, for any state y that $\bar{X}$ can access within k transitions, $|\bar{Q}(y,y)| \leq n(\gamma_*(\bar{p}_{*,0} + k) + \kappa_*) + n^2(\alpha_* + \beta_*)\bar{g}_*$, where the quantities with subscript $*$ are again the maximal ones.
- The same bound used above for $|\bar{Q}(y,y)|$ also applies to $|\bar{Q}_H(y,y)|$, so that $\bar{X}_H$ is nonexplosive.

□

B. Proof of Lemma 2

Our approach to this proof uses the finite state projection (see [24,25]) to "project" our original problem onto one with a finite number of finite classes at which point we use the results from [13,14]. To then return to the problem with infinite states, we perform an interchange of the limits in the finite state projection and singular perturbation. In [15], the authors apply a finite state projection to an SCRN prior to applying singular perturbation, but they are only concerned with analyzing the projection, so they do not perform this exchange of limits to justify the QSSA for the full system. We note that in [16], the authors offer another way to perform model reduction on infinite systems via analysis of stochastic equations. We choose to instead prove this proposition via an approach more similar to the formalism of [13,14], leveraging niceties of our system of interest (namely that each fast subsystem has only a finite number of states).

Before proving Lemma 2 we prove two necessary propositions.

Let $\mathcal{Y}$ be a countable set and let $A, B : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ be infinitesimal generators. Let $\mathcal{C}$ be the set of communicating classes of A, and suppose that each $C \in \mathcal{C}$ is finite and closed, so that A has a unique stationary distribution $\pi_C : \mathcal{C} \to \mathbb{R}$ supported on C. For each $\epsilon > 0$, let $Q_\epsilon = \frac{1}{\epsilon}A + B$, and let Y_ϵ be a CTMC with state space $\mathcal{Y}$, initial state $y_0 \in \mathcal{Y}$, and infinitesimal generator Q_ϵ. Let C_0 be the element of $\mathcal{C}$ to which y_0 belongs, and let Y be a CTMC with state space $\mathcal{C}$, initial state C_0, and infinitesimal generator $Q : \mathcal{C} \times \mathcal{C} \to \mathbb{R}$ given by

$$Q(C, C') = \sum_{y \in C} \sum_{y' \in C'} \pi_C(y) B(y, y'),$$

for $C \neq C'$.

Proposition 2. *Fix $t \geq 0$. If there exists a nonexplosive CTMC Z with state space $\mathcal{Y}$ and a sequence of sets $S_1 \subset S_2 \subset \ldots$ each of which is finite and whose union is $\mathcal{Y}$ such that*

$$\mathbb{P}(\tau_{Y_\epsilon, S_k} \leq t) \leq \mathbb{P}(\tau_{Z, S_k} \leq t) \tag{8}$$

for each $\epsilon > 0$ and $k \in \mathbb{N}$, then

$$\lim_{\epsilon \to 0^+} \sum_{C \in \mathcal{C}} \left| \sum_{y \in C} \rho_{Y_\epsilon}(y,t) - \rho_Y(C,t) \right| = 0.$$

Proof. It is in this proposition that we perform singular perturbation on countable Markov Chains by combining the finite state projection with singular perturbation for finite Markov Chains via a limit exchange.

Fix $t \geq 0$. Throughout this proof, we let k be the generic element of $\mathbb{N}$. Given a function $f : \mathcal{Y} \to \mathbb{R}$ with $\|f\|_1 < \infty$, let $Tf : \mathcal{C} \to \mathbb{R}$ be defined by $Tf(C) = \sum_{y \in C} f(y)$. Note that $\|Tf\|_1 \leq \|f\|_1$.

We first use the finite state projection (see [24,25]) to turn our singular perturbation problem into one that is finite-dimensional. Enumerate the communicating classes of $\mathcal{C}$ as $C_1, C_2, \ldots$, and for each k, let $D_k = \cup_{i=1}^k C_i$ and $\mathcal{D}_k = \{C_1, \ldots, C_k\}$. For each k and $\epsilon > 0$, define $\theta_{\epsilon,k} : \mathcal{Y} \times [0, \infty) \to [0,1]$ and $\eta_k : \mathcal{C} \times [0, \infty) \to [0,1]$ to be the solutions to the respective initial value problems

$$\frac{d}{ds}\theta_{\epsilon,k}(y,s) = \sum_{y' \in D_k} \theta_{\epsilon,k}(y',s) Q_\epsilon(y',y);\ \theta_{\epsilon,k}(y,0) = \delta_{y,y_0}, \tag{9}$$

$$\frac{d}{ds}\eta_k(C,s) = \sum_{C' \in \mathcal{D}_k} \eta_k(C',s) Q(C',C);\ \eta_k(C,0) = \delta_{C,C_0}, \tag{10}$$

for all $y \in D_k$ and $C \in \mathcal{D}_k$, where for $y \notin D_k$, we set $\theta_{\epsilon,k}(y,t) = 0$ and for $C \notin \mathcal{D}_k$, we set $\eta_k(C,t) = 0$.

$T\theta_{\epsilon,k}(\cdot,t)$ and $\eta_k(\cdot,t)$ are the respective finite state projections of $T\rho_{Y_\epsilon}(\cdot,t)$ and $\rho_Y(\cdot,t)$. We wish to show that these projections are equal in the limit of $\epsilon \to 0^+$.

Note that the linear system of ODEs (9) and (10) are finite dimensional (this is the point of using the finite state projection). Also note that $Q_\epsilon|_{D_k \times D_k} = \frac{1}{\epsilon}A|_{D_k \times D_k} + B|_{D_k \times D_k}$, where $A|_{D_k \times D_k}$ is "block-diagonal" in the sense that for each $j \in \{1, \ldots, k\}$, each $A|_{C_j \times C_j}$ is itself an infinitesimal generator and $A_{D_k \times D_k}(y,y') = 0$ when y and y' do not belong to the same communicating class. Based on these facts, via a standard change of basis and application of singular perturbation theory for finite linear systems (see [13,14] for details) we can conclude that

$$\lim_{\epsilon \to 0^+} \|T\theta_{\epsilon,k}(\cdot,t) - \eta_k(\cdot,t)\|_1 = 0. \tag{11}$$

Now that we know the projections are equal in the desired limit, we wish to show that the probability densities $T\rho_{Y_\epsilon}(\cdot,t)$ and $\rho_Y(\cdot,t)$ are equal in this same limit.

Theorem 2.5 (iv) in [25] then tells us that

$$\lim_{k \to \infty} \|\rho_Y(\cdot,t) - \eta_k(\cdot,t)\|_1 = 0. \tag{12}$$

We also have that

$$\begin{aligned} \|T\rho_{Y_\epsilon}(\cdot,t) - T\theta_{\epsilon,k}(\cdot,t)\|_1 &\leq \|\rho_{Y_\epsilon}(\cdot,t) - \theta_{\epsilon,k}(\cdot,t)\|_1 \\ &\leq \mathbb{P}(\tau_{Y_\epsilon, D_k} \leq t), \end{aligned} \tag{13}$$

where the second inequality comes from combining Theorem 2.5 parts (ii) and (iii) in [25] (note that (iii) is an extension of results from [24]).

Let $S_0 = \emptyset$ and choose $r_1 \leq r_2 \leq \cdots \in \mathbb{Z}_+$ such that $r_k \to \infty$ and $S_{r_k} \subset D_k$ for each k. From the definition of the exit time, for each k, $\mathbb{P}(\tau_{Y_\epsilon, D_k} \leq t) \leq \mathbb{P}(\tau_{Y_\epsilon, S_{r_k}} \leq t)$. By assumption, for each k, $\sup_{\epsilon>0} \mathbb{P}(\tau_{Y_\epsilon, S_{r_k}} \leq t) \leq \mathbb{P}(\tau_{Z, S_{r_k}} \leq t)$. Since Z is assumed nonexplosive, $\mathbb{P}(\tau_{Z,S_{r_k}} \leq t) \to 0$ as $k \to \infty$ by the Theorem 2.1 in [25], so that

$$\lim_{k\to\infty} \sup_{\epsilon>0} \mathbb{P}(\tau_{Y_\epsilon, D_k} \leq t) = 0.$$

Combining the above with (13) gives

$$\lim_{k\to\infty} \sup_{\epsilon>0} \|T\rho_{Y_\epsilon}(\cdot, t) - T\theta_{\epsilon,k}(\cdot, t)\|_1 = 0. \tag{14}$$

Equation (14) states that $T\theta_{\epsilon,k}(\cdot, t) \to T\rho_{Y_\epsilon}(\cdot, t)$ as $k \to \infty$ uniformly in $\epsilon > 0$, and (11) states that $T\theta_{\epsilon,k}(\cdot, t) \to \eta_k(\cdot, t)$ as $\epsilon \to 0^+$ point-wise in k, so that a standard result in analysis (see [Tao, Proposition 3.3.3]) guarantees we can exchange these limits:

$$\lim_{\epsilon\to 0^+} T\rho_{Y_\epsilon}(\cdot, t) = \lim_{k\to\infty} \eta_k(\cdot, t).$$

The above result together with (12) gives

$$\lim_{\epsilon\to 0^+} \|T\rho_{Y_\epsilon}(\cdot, t) - \rho_Y(\cdot, t)\|_1 = 0,$$

which is equivalent to our desired result. □

In order to apply Proposition 2, one must somehow construct the CTMC Z which in some sense bounds the CTMCs Y_ϵ uniformly as in (8). The upcoming proposition provides the machinery for constructing such a CTMC.

Definition 1 (Increasing set)**.** *Let $n, m \in \mathbb{N}$ and let $\mathcal{Y} \subset \mathbb{Z}_+^n$ be nonempty. We say a set $\Gamma \subset \mathcal{Y}$ is increasing in $\mathcal{Y}$ with respect to a matrix $L \in \mathbb{R}^{n\times m}$ with nonzero columns if for each $y \in \Gamma$ and $y' \in \mathcal{Y}$, $(y' - y)L \geq \mathbf{0}$ implies that $y \in \Gamma$.*

Proposition 3. *Let $n, m \in \mathbb{N}$, $\mathcal{Y} \subset \mathbb{Z}_+^n$ be nonempty, and $L \in \mathbb{R}^{n\times m}$ have nonzero columns. Let $\mathcal{S}$ and $\breve{\mathcal{S}}$ be two SCRNs with the same chemical species, state space $\mathcal{Y} \subset \mathbb{Z}_+^n$, and stoichiometry vectors $u_1, \ldots, u_R$. Let $\upsilon_1, \ldots, \upsilon_R$ and $\breve{\upsilon}_1, \ldots, \breve{\upsilon}_R$ respectively be the propensity functions of $\mathcal{S}$ and $\breve{\mathcal{S}}$ associated with each reaction.*

Fix $y_0 \in \mathcal{Y}$, choose $S \subset \mathcal{Y}$, and let Y and $\breve{Y}$ be the CTMCs with initial state y_0 associated with $\mathcal{S}$ and $\breve{\mathcal{S}}$, respectively. Suppose that Y and $\breve{Y}$ are nonexplosive and that for each $r \in \{1, \ldots, R\}$ the following conditions hold

- *$\mathcal{Y} - S$ is increasing in $\mathcal{Y}$ with respect to L.*
- *if $u_r L$ has at least one negative entry, then $sup_{y\in\mathcal{Y}} \breve{\upsilon}_r(y) \leq \inf_{y\in\mathcal{Y}} \upsilon_r(y)$*
- *if $u_r L$ has at least one positive entry, then $sup_{y\in\mathcal{Y}} \upsilon_r(y) \leq \inf_{y\in\mathcal{Y}} \breve{\upsilon}_r(y)$.*

Then for all $t \geq 0$,

$$\mathbb{P}(\tau_{Y,S} \leq t) \leq \mathbb{P}(\tau_{\breve{Y},S} \leq t).$$

Proof. This is an direct consequence of Theorem 3.4 of [27]. □

Proof of Lemma 2. For convenience, define $\Theta = \{\alpha, \beta, \kappa, \gamma, \gamma'\}$, which represents the set of "slow" reaction types in $\mathcal{S}$, define $\Phi = \{f, b\}$, which represents the set of "fast" reaction types in $\mathcal{S}$, and define $\Xi = \{+, +', -\}$, which represents the set of reaction types in $\bar{\mathcal{S}}$. Throughout the rest of this proof, let θ be the generic element of Θ, ϕ the generic element of Φ, ξ the generic element of Ξ, and i the generic element of $\{1, \ldots, n\}$. We also let $x := (p_1, \ldots, p_n, g_1, \ldots, g_n, c_1, \ldots, c_n)$ and $x' := (p'_1, \ldots, p'_n, g'_1, \ldots, g'_n, c'_1, \ldots, c'_n)$ represent generic elements of $\mathcal{X}$, and we let $\bar{x} := (\bar{p}_1, \ldots, \bar{p}_n)$ and $\bar{x}' := (\bar{p}'_1, \ldots, \bar{p}'_n)$ represent generic elements of $\bar{\mathcal{X}}$.

We first explicitly write the infinitesimal generators of X_ϵ and $\bar{X}$. For each i and ϕ, let $u_{\phi,i}$ be the stoichiometry vector of the reaction in $\mathcal{R}$ associated with propensity $\upsilon_{\phi,i}$. Let $A : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be the infinitesimal generator determined by $\{(u_{\phi,i}, \upsilon_{\phi,i})\}_{\phi,i}$ (i.e. the one associated with $\mathcal{R}$). Recall that for each $\bar{x} \in \bar{\mathcal{X}}$, $\pi_{\bar{x}}$ is defined to be the stationary distribution of $\mathcal{R}$ (and thus also of A) supported on $E_{\bar{x}} := \{x \in \mathcal{X} : \forall_i\ p_i + c_i = \bar{p}_i\}$, with $\mathcal{E} := \{E_{\bar{x}} : \bar{x} \in \bar{\mathcal{X}}\}$ being the set of communicating classes of $\mathcal{R}$ (and thus also of A), each of which is finite and closed.

For each i and θ, let $u_{\theta,i}$ be the stoichiometry vector of the reaction in each $\mathcal{S}_\epsilon$ that is associated with the propensity $\upsilon_{\theta,i}$. Let $B : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be the infinitesimal generator determined by $\{(u_{\theta,i}, \upsilon_{\theta,i})\}_{\theta,i}$. Then by definition, for each $\epsilon > 0$, X_ϵ is a CTMC with state space $\mathcal{X}$, initial state x_0, and infinitesimal generator $Q_\epsilon := \frac{1}{\epsilon}A + B$.

For each i and ξ, let $\bar{u}_{\xi,i}$ be the stoichiometry vector of the reaction in $\bar{\mathcal{S}}$ associated with propensity $\bar{\upsilon}_{\xi,i}$. By definition, $\bar{X}$ is a CTMC with state space $\bar{\mathcal{X}}$, initial state $\bar{x}_0$, and infinitesimal generator $\bar{Q} : \bar{\mathcal{X}} \times \bar{\mathcal{X}} \to \mathbb{R}$ given by

$$\bar{Q}(\bar{x}, \bar{x}') = \sum_{\xi,i} \bar{\upsilon}_{\xi,i}(\bar{x}) \mathbb{1}_{\{\bar{x}'\}}(\bar{x} + \bar{u}_{\xi,i}), \tag{15}$$

for $\bar{x} \neq \bar{x}'$.

We next introduce an intermediate CTMC X which will turn out to be equivalent to $\bar{X}$, albeit defined to have a different state space. In particular, let X be a CTMC with state space $\mathcal{E}$, initial state $E_{\bar{x}_0}$, and infinitesimal generator $Q : \mathcal{E} \times \mathcal{E} \to \mathbb{R}$ given by

$$\begin{aligned} Q(E_{\bar{x}}, E_{\bar{x}'}) :&= \sum_{x\in E_{\bar{x}}} \sum_{x'\in E_{\bar{x}'}} \pi_{\bar{x}}(x) B(x, x') \\ &= \sum_{x\in E_{\bar{x}}} \sum_{x'\in E_{\bar{x}'}} \pi_{\bar{x}}(x) \sum_{\theta,i} \upsilon_{\theta,i}(x) \mathbb{1}_{\{x'\}}(x + u_{\theta,i}) \\ &= \sum_{\theta,i} \sum_{x\in E_{\bar{x}}} \pi_{\bar{x}}(x) \upsilon_{\theta,i}(x) \mathbb{1}_{E_{\bar{x}'}}(x + u_{\theta,i}) \\ &= \sum_{\theta,i} \mathbb{E}_{W\sim\pi_{\bar{x}}}[\upsilon_{\theta,i}(W)] \mathbb{1}_{\{\bar{x}'\}}(\bar{x} + \Delta\mathbf{T}(u_{\theta,i})) \end{aligned} \tag{16}$$

for $\bar{x} \neq \bar{x}'$, where $\Delta\mathbf{T} : \mathbb{Z}^{3n} \to \mathbb{Z}^n$ maps $(\Delta p_1, \ldots, \Delta p_n, \Delta g_1, \ldots, \Delta g_n, \Delta c_1, \ldots, \Delta c_n)$ to $(\Delta p_1 + \Delta c_1, \ldots, \Delta p_n +$

Δc_n). In other words, $\Delta\mathbf{T}(u)$ gives the change in total counts of each protein species associated with the stoichiometry vector u.

Our the proof will consist of first showing

$$\lim_{\epsilon\to 0^+}\sum_{\bar{x}\in\mathcal{X}}\left|\sum_{x\in E_{\bar{x}}}\rho_{X_\epsilon}(x,t)-\rho_X(E_{\bar{x}},t)\right|=0, \tag{17}$$

and subsequently showing that for each $\bar{x},\bar{x}'\in\bar{\mathcal{X}}$,

$$\rho_X(E_{\bar{x}},t)=\rho_{\bar{X}}(\bar{x},t). \tag{18}$$

Combining (17) and (18) then gives our desired result.

Towards showing (17), for each $k\in\mathbb{N}$, let $S_k=\{x\in\mathcal{X}:\forall i\ p_i+c_i\le k\}$. In light of Proposition 2, (17) follows if we can find a nonexplosive CTMC Z with state space $\mathcal{X}$ such that for all $\epsilon>0$ and $k\in\mathbb{N}$,

$$\mathbb{P}(\tau_{X_\epsilon,S_k}\le t)\le\mathbb{P}(\tau_{Z,S_k}\le t). \tag{19}$$

Define the SCRN $\breve{\mathcal{S}}$ with species $\mathrm{P}_1,\dots,\mathrm{P}_n,\mathrm{G}_1,\dots,\mathrm{G}_n,\mathrm{C}_1,\dots,\mathrm{C}_n$, state space $\mathcal{X}$, and the following reaction-propensity pairs for each $i\in\{1,\dots,n\}$:

$$\begin{aligned}
\mathrm{P}_i+\mathrm{G}_i\to\mathrm{C}_i,\ &\breve{\upsilon}_{f,i}(x)=0;\\
\mathrm{C}_i\to\mathrm{P}_i+\mathrm{G}_i,\ &\breve{\upsilon}_{b,i}(x)=0;\\
\mathrm{C}_i\to\mathrm{C}_i+\mathrm{P}_{\sigma(i)},\ &\breve{\upsilon}_{\alpha,i}(x)=\alpha_i\bar{g}_i;\\
\mathrm{G}_i\to\mathrm{G}_i+\mathrm{P}_{\sigma(i)},\ &\breve{\upsilon}_{\beta,i}(x)=\beta_i\bar{g}_i;\\
\emptyset\to\mathrm{P}_i,\ &\breve{\upsilon}_{\kappa,i}(x)=\Omega\kappa_i;\\
\mathrm{P}_i\to\emptyset,\ &\breve{\upsilon}_{\gamma,i}(x)=0;\\
\mathrm{C}_i\to\mathrm{G}_i,\ &\breve{\upsilon}_{\gamma',i}(x)=0.
\end{aligned}$$

Define $L=\begin{bmatrix}\mathbf{e}_{P,1}+\mathbf{e}_{C,1} & \dots & \mathbf{e}_{P,n}+\mathbf{e}_{C,n}\end{bmatrix}$, where $\mathbf{e}_{P,1},\dots,\mathbf{e}_{P,n},\mathbf{e}_{G,1},\dots,\mathbf{e}_{G,n},\mathbf{e}_{C,1},\dots,\mathbf{e}_{C,n}$ are the standard unit vectors of $\mathbb{Z}_+^{3n}$. Fix $\epsilon>0$ and $k\in\mathbb{N}$.

Let Z be the CTMC with initial state x_0 that is associated with $\breve{\mathcal{S}}$. We check that the hypotheses of Proposition 3 are satisfied (with $\mathcal{X}$, $\mathcal{S}_\epsilon$, S_k, X_ϵ, and Z respectively in place of $\mathcal{Y}$, $\mathcal{S}$, S, Y, and $\breve{Y}$): X_ϵ is nonexplosive by Proposition 1 and Z is nonexplosive because its propensity functions are bounded (see Theorem 2.7.1(ii) in [26]). Furthermore:

- Suppose $x\in\mathcal{X}-S_k$, $x'\in\mathcal{X}$, and $(x'-x)L\ge\mathbf{0}$. Then for some j, $p_j+c_j>k$. But since $(x'-x)L\ge\mathbf{0}$, $p'_j+c'_j\ge p_j+c_j>k$, so that $x'\in\mathcal{X}-S_k$.
- For each i and $\mu\in\{\alpha,\beta,\kappa\}$, $u_{\mu,i}L$ has non-negative entries, and for $\lambda\in\{\gamma,\gamma',f,b\}$, $\breve{\upsilon}_{\lambda,i}=0\le\upsilon_{\lambda,i}$.
- For each i and $\mu\in\{\alpha,\beta,\kappa\}$, we have that $\breve{\upsilon}_{\mu,i}\ge\upsilon_{\mu,i}$, and for $\lambda\in\{\gamma,\gamma',f,b\}$, $u_{\lambda,i}L$ has non-positive entries.

Thus by Lemma 3, Z satisfies (19). Since $\epsilon>0$ and $k\in\mathbb{N}$ were chosen arbitrarily and Z is nonexplosive, (17) is shown.

Towards showing (18), note that there is a natural one-to-one correspondence between $\bar{\mathcal{X}}$ and $\mathcal{E}$ given by $\bar{x}\leftrightarrow E_{\bar{x}}$. Since $\bar{X}$ is a CTMC on state space $\bar{\mathcal{X}}$ with infinitesimal generator $\bar{Q}(\bar{x},\bar{x}')$ and initial state $\bar{x}_0$ and X is a CTMC on state space $\mathcal{E}$ with infinitesimal generator $Q(E_{\bar{x}},E_{\bar{x}'})$ and initial state $E_{\bar{x}_0}$, if we can show that

$$Q(E_{\bar{x}},E_{\bar{x}'})=\bar{Q}(\bar{x},\bar{x}'), \tag{20}$$

then we will have demonstrated (18).

Note that (16) says that Q is determined by the propensity functions $\hat{\upsilon}_{\theta,i}(\bar{x}):=\mathbb{E}_{W\sim E_{\bar{x}}}[\upsilon_{\theta,i}(W)]$ and reaction vectors $\hat{u}_{\theta,i}:=\Delta\mathbf{T}(u_{\theta,i})$ (with $\theta\in\Theta$, $i\in\{1,\dots,n\}$). But $\hat{u}_{\alpha,i}=\hat{u}_{\beta,i}=\bar{u}_{+,i}$, $\hat{u}_{\kappa,i}=\bar{u}_{+',i}$, and $\hat{u}_{\gamma,i}=\hat{u}_{\gamma',i}=\bar{u}_{-,i}$. Thus if we can show $\hat{\upsilon}_{\alpha,i}+\hat{\upsilon}_{\beta,i}=\bar{\upsilon}_{+,i}$, $\hat{\upsilon}_{\kappa,i}=\bar{\upsilon}_{+',i}$, and $\hat{\upsilon}_{\gamma,i}+\hat{\upsilon}_{\gamma',i}=\bar{\upsilon}_{-,i}$ (i.e. the sum of propensities associated with a given reaction vector are the same for Q and $\bar{Q}$), then by comparing (15) and (16) we are done. We check these conditions one by one:

- $\mathbb{E}_{W\sim\pi_{\bar{x}}}[\upsilon_{\alpha,i}(W)]+\mathbb{E}_{W\sim\pi_{\bar{x}}}[\upsilon_{\beta,i}(W)]=\alpha_i\mathbb{E}_{W\sim\pi_{\bar{x}}}[C_i]+\beta_i\mathbb{E}_{W\sim\pi_{\bar{x}}}[G_i]=\bar{\upsilon}_{+,i}(\bar{x})$,
- $\mathbb{E}_{W\sim\pi_{\bar{x}}}[\upsilon_{\kappa,i}(W)]=\Omega\kappa_i=\bar{\upsilon}_{+',i}(\bar{x})$,
- $\mathbb{E}_{W\sim\pi_{\bar{x}}}[\upsilon_{\gamma,i}(W)]+\mathbb{E}_{W\sim\pi_{\bar{x}}}[\upsilon_{\gamma',i}(W)]=\gamma_i\mathbb{E}_{W\sim\pi_{\bar{x}}}[P_i]+\gamma_i\mathbb{E}_{W\sim\pi_{\bar{x}}}[C_i]=\gamma_i\mathbb{E}_{W\sim\pi_{\bar{x}}}[P_i+C_i]=\gamma_i\bar{p}_i=\bar{\upsilon}_{-,i}(\bar{x})$,

where $W=(P_1,\dots,P_n,G_1,\dots,G_n,C_1,\dots,C_n)$. Hence we have shown (17) and (18), so the proof is complete. □

C. Proof of Theorem 1

Proof of Theorem 1. We first show that for all $i\in\{1,\dots,n\}$ and all $\bar{x}\in\bar{\mathcal{X}}$,

$$|\bar{\upsilon}^H_{+,i}(\bar{x})-\bar{\upsilon}_{+,i}(\bar{x})|\le|\alpha_i-\beta_i|\frac{\bar{g}_i^2}{\bar{g}_i+\Omega K_i}. \tag{21}$$

To show the above, fix $i\in\{1,\dots,n\}$ and let $\bar{x}=(\bar{p}_1,\dots,\bar{p}_n)$ and let $W=(P_1,\dots,P_n,G_1,\dots,G_n,C_1,\dots,C_n)$ be a random variable with distribution $\pi_{\bar{x}}$. Given that $\pi_{\bar{x}}$ has finite support, from the moment dynamics equations (see [23]) corresponding to $\mathcal{R}$,

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}[C_i]&=\frac{f_i}{\Omega}\mathbb{E}[P_iG_i]-b_i\mathbb{E}[C_i]\\
&=\frac{f_i}{\Omega}\mathbb{E}[(\bar{g}_i-C_i)(\bar{p}_i-C_i)]-b_i\mathbb{E}[C_i]\\
&=\frac{f_i}{\Omega}\mathbb{E}[C_i^2]-\left(b_i+\frac{f_i}{\Omega}(\bar{p}_i+\bar{g}_i)\right)\mathbb{E}[C_i]+\frac{f_i}{\Omega}\bar{p}_i\bar{g}_i,
\end{aligned}$$

where $P_i=\bar{p}_i-C_i$ because W is by definition supported in $E_{\bar{x}}$. But since $\pi_{\bar{x}}$ is a stationary distribution of $\bar{X}$, $\frac{d}{dt}\mathbb{E}[C_i]=0$, so that

$$\mathbb{E}[C_i^2]-(\Omega K_i+\bar{p}_i+\bar{g}_i)\mathbb{E}[C_i]+\bar{p}_i\bar{g}_i=0. \tag{22}$$

Since $0\le C_i\le\bar{g}_i$, we have that $0\le\mathbb{E}[C_i^2]\le\bar{g}_i\mathbb{E}[C_i]$. Using this fact and the above equation, it follows that

$$\frac{\bar{p}_i\bar{g}_i}{\bar{p}_i+\bar{g}_i+\Omega K_i}\le\mathbb{E}[C_i]\le\frac{\bar{p}_i\bar{g}_i}{\bar{p}_i+\Omega K_i},$$

which implies

$$\left|\mathbb{E}[C_i]-\frac{\bar{p}_i\bar{g}_i}{\bar{p}_i+\Omega K_i}\right|\le\frac{\bar{g}_i^2}{\bar{g}_i+\Omega K_i}.$$

Together with the relation $\mathbb{E}[C_i]+\mathbb{E}[G_i]=\bar{g}_i$, the above immediately implies (21).

We now continue to the main proof. We use k as the generic element of $\mathbb{N}$ and $\bar{x}$ as the generic element of $\bar{\mathcal{X}}$. Let $\bar{Q}$ and $\hat{Q}$ be the respective infinitesimal generators of $\bar{X}$

and $\bar{X}_H$. Let $S_1 \subset S_2 \subset \dots$ be a sequence of sets, each of which is finite and whose union is $\bar{\mathcal{X}}$.

We use the finite state projection to "project" $\bar{X}$ and $\bar{X}_H$ onto the S_k and find a bound on the TV error between the projections which applies uniformly for each k. To begin, for each k, define $\bar{\rho}_k : \bar{\mathcal{X}} \times [0,\infty) \to [0,1]$ and $\hat{\rho}_k : \bar{\mathcal{X}} \times [0,\infty) \to [0,1]$ to be the solutions to the initial value problems

$$\frac{d}{ds}\bar{\rho}_k(\bar{x},s) = \sum_{\bar{x}' \in S_k} \bar{\rho}_k(\bar{x}',s)\bar{Q}(\bar{x}',\bar{x});\ \bar{\rho}_k(\bar{x},0) = \delta_{\bar{x},\bar{x}_0};$$
$$\frac{d}{ds}\hat{\rho}_k(\bar{x},s) = \sum_{\bar{x}' \in S_k} \hat{\rho}_k(\bar{x}',s)\hat{Q}(\bar{x}',\bar{x});\ \hat{\rho}_k(\bar{x},0) = \delta_{\bar{x},\bar{x}_0},$$

for all $\bar{x} \in S_k$, and for all $\bar{x} \in \mathcal{X} - S_k$, set $\bar{\rho}_k(\bar{x},t) = \hat{\rho}_k(\bar{x},t) = 0$.

Define $\Delta\rho_k = \bar{\rho}_k - \hat{\rho}_k$. Theorem 2.5 (iv) from [25] guarantees that $\lim_{k\to\infty} \|\rho_{\bar{X}}(\cdot,t) - \bar{\rho}_k(\cdot,t)\|_1 = 0$ and $\lim_{k\to\infty} \|\rho_{\hat{X}}(\cdot,t) - \hat{\rho}_k(\cdot,t)\|_1 = 0$, so that

$$\begin{aligned} d_{TV}(\bar{X}(t),\bar{X}_H(t)) &= \frac{1}{2}\|\rho_{\bar{X}}(\cdot,t) - \rho_{\hat{X}}(\cdot,t)\|_1 \\ &\le \lim_{k\to\infty} \frac{1}{2}\|\Delta\rho_k(\cdot,t)\|_1. \end{aligned} \tag{23}$$

Define $\Delta Q = \bar{Q} - \hat{Q}$. Then for each k and $\bar{x} \in S_k$,

$$\begin{aligned} \frac{d}{ds}\Delta\rho_k(\bar{x},s) = &\sum_{\bar{x}' \in S_k} \Delta\rho_k(\bar{x}',t)\bar{Q}(\bar{x}',\bar{x}) \\ &+ \sum_{\bar{x}' \in S_k} \hat{\rho}_k(\bar{x}',t)\Delta Q(\bar{x}',\bar{x}). \end{aligned}$$

The solution to the above finite system of linear ODEs equation, given the zero initial condition $\Delta\rho_k(\cdot,0) = 0$, is given by the following convolution:

$$\begin{aligned} \Delta\rho_k(\bar{x},t) = \int_0^t \sum_{\bar{x}',\bar{x}'' \in S_k} &\hat{\rho}_k(\bar{x}'',t-\tau)\Delta Q(\bar{x}'',\bar{x}') \\ &\times e^{\bar{Q}_k\tau}(\bar{x}',\bar{x})d\tau, \end{aligned}$$

where $\bar{Q}_k = \bar{Q}|_{S_k \times S_k}$. Thus

$$\begin{aligned} \|\Delta\rho_k(\cdot,t)\|_1 &\le \int_0^t \|\hat{\rho}_k(\cdot,t-\tau)\|_1 \|\Delta Q\|_1 \left\|e^{\bar{Q}_k\tau}\right\|_1 d\tau \\ &\le \int_0^t \|\Delta Q\|_1 d\tau, \end{aligned} \tag{24}$$

where $\|e^{\bar{Q}_k\tau}\|_1 \le 1$ because $\bar{Q}$ is an infinitesimal generator. It follows from [19] (section 4.1.3) that if $\bar{g}_i = 1$, $\bar{v}^H_{+,i}(\bar{x}) = \bar{v}_{+,i}(\bar{x})$. Using this fact in combination with (21), we have that

$$\begin{aligned} \sum_{\bar{x}' \in \bar{\mathcal{X}}} |\Delta Q(\bar{x}'',\bar{x}')| &= 2\,|\Delta Q(\bar{x}'',\bar{x}'')| \\ &= 2\sum_{\xi,i} |\bar{v}^H_{\xi,i}(\bar{x}'') - \bar{v}_{\xi,i}(\bar{x}'')| \\ &= 2\sum_{i=1}^{n} |\bar{v}^H_{+,i}(\bar{x}'') - \bar{v}_{+,i}(\bar{x}'')| \\ &\le 2(n-n_{01}) \max_{i=1,\dots,n} |\alpha_i - \beta_i| \frac{\bar{g}_i^2}{\bar{g}_i + \Omega K_i}, \end{aligned}$$

where $\xi \in \{+,+',-\}$. Thus $\|\Delta Q\|_1 \le 2(n - n_{01}) \max_{i=1,\dots,n} |\alpha_i - \beta_i| \bar{g}_i^2/(\bar{g}_i + \Omega K_i)$. Together with (23) and (24), this gives our desired result. □

REFERENCES

[1] U. Alon, *An introduction to systems biology: design principles of biological circuits*. Chapman and Hall/CRC, 2006.

[2] L. A. Segel and M. Slemrod, "The quasi-steady-state assumption: a case study in perturbation," *SIAM Rev.*, vol. 31, no. 3, pp. 446–477, 1989.

[3] D. Del Vecchio and R. M. Murray, *Biomolecular feedback systems*. Princeton University Press Princeton, NJ, 2015.

[4] J. D. Murray, *Mathematical biology: I. An introduction*. Springer, 2002.

[5] J. K. Kim, K. Josić, and M. R. Bennett, "The validity of quasi-steady-state approximations in discrete stochastic simulations," *Biophys. J.*, vol. 107, no. 3, pp. 783–793, 2014.

[6] J. Holehouse and R. Grima, "Revisiting the reduction of stochastic models of genetic feedback loops with fast promoter switching," *Biophys. J.*, vol. 117, no. 7, pp. 1311–1330, 2019.

[7] J. K. Kim, K. Josić, and M. R. Bennett, "The relationship between stochastic and deterministic quasi-steady state approximations," *BMC Syst. Bio.*, vol. 9, pp. 1–13, 2015.

[8] Y. Cao, D. T. Gillespie, and L. R. Petzold, "The slow-scale stochastic simulation algorithm," *J. Chem. Phys.*, vol. 122, no. 1, p. 014116, 2005.

[9] C. V. Rao and A. P. Arkin, "Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the gillespie algorithm," *J. Chem. Phys.*, vol. 118, no. 11, pp. 4999–5010, 2003.

[10] K. R. Sanft, D. T. Gillespie, and L. R. Petzold, "Legitimacy of the stochastic michaelis–menten approximation," *IET Systems Biology*, vol. 5, no. 1, pp. 58–69, 2011.

[11] J. Goutsias, "Quasiequilibrium approximation of fast reaction kinetics in stochastic biochemical systems," *J. Chem. Phys.*, vol. 122, no. 18, p. 184102, 2005.

[12] L. D. EW and E. Vanden-Eijnden, "Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates," *J. Chem. Phys.*, vol. 123, no. 19, p. 194107, 2005.

[13] R. Phillips and P. Kokotovic, "A singular perturbation approach to modeling and control of markov chains," *IEEE Trans. Autom. Control*, vol. 26, no. 5, pp. 1087–1094, 1981.

[14] S. Peleš, B. Munsky, and M. Khammash, "Reduction and solution of the chemical master equation using time scale separation and finite state projection," *J. Chem. Phys.*, vol. 125, no. 20, p. 204104, 2006.

[15] B. Munsky, S. Peles, and M. Khammash, "Stochastic analysis of gene regulatory networks using finite state projections and singular perturbation," in *2007 Am. Control Conf.* IEEE, 2007, pp. 1323–1328.

[16] H.-W. Kang and T. G. Kurtz, "Separation of time-scales and model reduction for stochastic reaction networks," *The Annals of Applied Probability*, vol. 23, no. 2, pp. 529–583, 2013.

[17] A. Crudu, A. Debussche, A. Muller, and O. Radulescu, "Convergence of stochastic gene networks to hybrid piecewise deterministic processes," *The Annals of Applied Probability*, vol. 22, no. 5, pp. 1822–1859, 2012.

[18] R. Dessalles, V. Fromion, and P. Robert, "A stochastic analysis of autoregulation of gene expression," *J. Math. Bio.*, vol. 75, no. 5, pp. 1253–1283, 2017.

[19] B. Mélykúti, J. P. Hespanha, and M. Khammash, "Equilibrium distributions of simple biochemical reaction systems for time-scale separation in stochastic reaction networks," *J. R. Soc., Interface*, vol. 11, no. 97, p. 20140054, 2014.
[20] J. K. Kim and E. D. Sontag, "Reduction of multiscale stochastic biochemical reaction networks using exact moment derivation," *PLoS Comput. Bio.*, vol. 13, no. 6, p. e1005571, 2017.
[21] U. Kwon, M. Naghnaeian, and D. Del Vecchio, "Approximation of the chemical master equation using conditional moment closure and time-scale separation," in *2019 Am. Control Conf.* IEEE, 2019, pp. 585–592.
[22] E. Levine and T. Hwa, "Stochastic fluctuations in metabolic pathways," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 22, pp. 9224–9229, 2007.
[23] S. Engblom, "Computing the moments of high dimensional solutions of the master equation," *Applied Mathematics and Computation*, vol. 180, no. 2, pp. 498–515, 2006.
[24] B. Munsky and M. Khammash, "The finite state projection algorithm for the solution of the chemical master equation," *J. Chem. Phys.*, vol. 124, no. 4, p. 044104, 2006.
[25] J. Kuntz *et al.*, "The exit time finite state projection scheme: bounding exit distributions and occupation measures of continuous-time Markov chains." *SIAM J. Sci. Comput.*, vol. 41, no. 2, pp. A748–A769, 2019.
[26] J. R. Norris, *Markov chains*. Cambridge university press, 1998, no. 2.
[27] F. A. Campos, S. Bruno, Y. Fu, D. Del Vecchio, and R. J. Williams, "Comparison theorems for stochastic chemical reaction networks," *Bull. Math. Biol.*, vol. 85, no. 5, p. 39, 2023.
[28] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *J. Phys. Chem.*, vol. 81, no. 25, pp. 2340–2361, 1977.