2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Error Bound for Hill-Function Approximations in a Class of Stochastic Transcriptional Network Models

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 (TF) that regulates the gene. However, it is generally unknown how much error one may incur from this approximation. We investigate this question in the context of transcriptional networks (TNs). Under the assumption of rapid binding and unbinding of TFs with their gene targets, we bound the approximation error (in terms of the total variation distance) between a mass-action stochastic model and a corresponding model with Hill function propensities. To do so, we use a combination of singular perturbation theory and moment analysis for stochastic chemical reaction networks. We assume throughout that TFs regulate genes in a one-to-one fashion, each regulated gene produces a single TF, TFs do not multimerize, and each gene only has a single TF 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). In TNs, genes produce TFs, and TFs regulate gene activities, i.e. the rates at which genes are transcribed, via reversible TF-gene binding [1].

One often reduces deterministic TN models by exploiting the separation between the timescale of binding and unbinding of TFs with genes and the timescale of production and decay of TFs. Per singular perturbation theory, this separation justifies the "quasi-steady-state" approximation (QSSA) that the time-derivative of the TF-gene complex concentration is approximately zero after a short transient [2]–[4]. One can use this approximation 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 [5,6]. In [7,8], the authors show that in the limit of weak TF-gene binding, stochastic models of various TNs using Hill function propensities are accurate when the relevant timescales are sufficiently separated. In [9], the authors showed an analogous result for the similar case of the Michaelis-Menten approximation in stochastic enzyme kinetics.

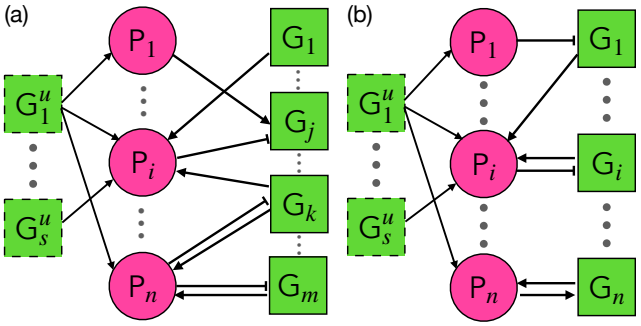

Fig. 1. Visualization of TNs via bipartite directed graphs. (a) A graph of a general TN. An edge from gene G to TF P indicates that G produces P. An edge from a TF P to a gene G indicates that P regulates G by changing the rate at which G produces TFs. An "$\rightarrow$" arrow from P to G indicates that G produces more TF (positive regulation) when bound to P than when unbound, and a "$\dashv$" 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 a TN from the restricted class that we consider in this work. Unlike Figure 1a, in this TN, each TF regulates at most one gene, each regulated gene produces only one TF, and no two different TFs regulate the same gene.

The stochastic QSSA (sQSSA) is an analogue of the QSSA that produces reduced stochastic models of two-timescale CRNs [10]–[14]. 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 [10]. For some CRNs, one can use singular perturbation theory or related techniques to prove that the sQSSA produces an accurate reduced model for sufficiently large timescale separation [15]–[20]. Furthermore, for CRNs with certain structures, expressions for the propensity functions of the reduced model can be found explicitly [21,22]. In related work, Kwon et al. gave an algorithm that exploits timescale separation and the moment dynamics of the fast virtual process to approximately, but efficiently, simulate a CRN and compute error bounds [23].

In this work, we consider a restricted class of TNs in the limit of rapid binding and unbinding of TFs with genes. In this setting, we provide an explicit bound on the error (measured in total variation distance) between a stochastic model

This material is based upon work supported in part by the Air Force Office of Scientific Research under grant FA9550-22-1-0316, the National Institutes of Health under grant R01-HD105947-03, and the National Science Foundation Graduate Research Fellowship Program 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

of a TN using mass-action propensities and a simplified stochastic model using Hill function propensities. Critically, we do not assume weak binding between TFs and genes. To obtain our results, we first justify the sQSSA for this class of TNs using singular perturbation theory, and thereafter we use the moment dynamics of the virtual fast CRN to bound the error between the propensity functions of the resulting reduced model and the Hill propensity functions. 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 multimerize (e.g. do not homodimerize or heterodimerize) and that each gene has only one TF binding site. However, we do not assume that each gene appears in a single copy. For simplicity, we ignore mRNA dynamics.

II. MATHEMATICAL BACKGROUND

We use $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Z}_+$, and $\mathbb{R}$ to denote the set of natural numbers (excluding 0), integers, non-negative integers, and real numbers, respectively. 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 $t \geq 0$ 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 [24]. $\mathcal{Y}$ 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 *non-explosive* if $T_Y^\infty = \infty$ almost surely.

In agreement with the assumptions of the references [24]–[26] 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 [24] 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 [24].

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 $\mathcal{Y}$, 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 $u_r \neq 0$ for each r, and for each $y \in \mathcal{Y}$, if $y + u_r \notin \mathcal{Y}$ then $\upsilon_r(y) = 0$.

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 $d_{TV}(Y, Y') = \frac{1}{2}\sum_{y \in \mathcal{Y}} |\mathbb{P}(Y = y) - \mathbb{P}(Y' = y)|$.

Conceptually, the total variation distance between two random variables quantifies the degree to which their probability distributions do not overlap. The TVD will be our primary metric of dissimilarity between random variables corresponding to different stochastic models.

III. PROBLEM SETTING

We first describe the "full" SCRN, which is a model of a TN of the form shown in Figure 1b that uses mass-action propensities. We next define the "reduced" SCRN, which is in essence 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 function propensities. Our ultimate 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. The reduced SCRN will serve as an intermediary by which to achieve this goal.

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 $\mathcal{X} := \{(p_1, \ldots, p_n, g_1, \ldots, g_n, c_1, \ldots, c_n) \in \mathbb{Z}_+^{3n} : \forall i \; g_i + c_i = \bar{g}_i\}$. Define $\tilde{\mathcal{X}} := \mathbb{Z}_+^n$.

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 $\mathbf{T}x = (p_1 + c_1, \ldots, p_n + c_n)$. Conceptually, $\mathbf{T}x$ represents the total (unbound + 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 $i \in \{1, \ldots, n\}$:

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

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

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

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

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

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

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

Physically the P_i's represent unbound TFs, the G_i's represent unbound genes, and the C_i's represent TF-gene complexes. 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 1b, 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 "virtual" 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 fast 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 $\mathcal{X}$ for which the total (unbound + 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 [24]).

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 unbound or bound (we use the overbar to emphasize this difference from P_i, which refers to the unbound 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 $\upsilon_{\gamma,i}(x) + \upsilon_{\gamma',i}(x)$. Thus, this SCRN may be viewed as the reduced model for the full SCRN via application of the sQSSA (see [10] for more details), where we make a change of coordinates so that we only keep track of total TF.*

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. Note that due to the assumption that TFs do not form multimers, the Hill coefficient in the above expression is equal to 1.

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 of interest are non-explosive. 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 non-explosive.*

Proof. See Appendix A. □

Remark 2. *To interpret the following results, it is helpful to write $X_\epsilon(t)$ as $(P_{\epsilon,1}(t), \ldots, P_{\epsilon,n}(t), G_{\epsilon,1}(t), \ldots, G_{\epsilon,n}(t), C_{\epsilon,1}(t), \ldots, C_{\epsilon,n}(t))$, $\bar{X}(t)$ as $(\bar{P}_1(t), \ldots, \bar{P}_n(t))$, and $\bar{X}_H(t)$ as $(\bar{P}_1^H(t), \ldots, \bar{P}_n^H(t))$. Then $\mathbf{T}X_\epsilon(t) = (P_{\epsilon,1}(t) + C_{\epsilon,1}(t), \ldots, P_{\epsilon,n}(t) + C_{\epsilon,n}(t))$ 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 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 [21] (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

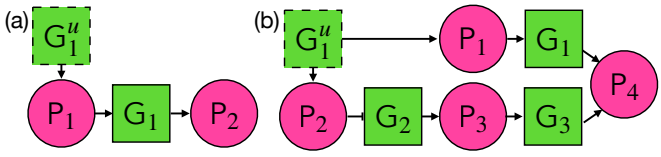

Fig. 2. Example TNs. (a) Simple regulation network. (b) Incoherent feedforward network.

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). The full, reduced, and Hill SCRNs are respectively given by (1a)-(1g), (2a)-(2c), and (3a)-(3c), with $n = 2$ and $\sigma(1) = 2$. Since P_2 does not regulate a gene, $\bar{g}_2 = 0$.

The probability distributions for each SCRN were computed and compared in TVD. For sufficiently large separation of timescales ($\epsilon \ll 1$), we expect based on Theorem 1 that the TVD between the full and Hill SCRNs will be smaller than or approximately equal to the right-hand side of (4), as in cases 1-3 of Figure 3a. Note that if the binding strength is too large, the bound becomes trivial (case 3). In line with Corollary 1, if $\bar{g}_1 = 1$, then the TVD can be made small by making ϵ small, even when TF-gene binding is not weak (case 4). If ϵ is not sufficiently small, the errors can substantially exceed the bound (cases 5-6). As expected based on (4), when there is both a large separation of timescales and weak binding ($\bar{g} \ll \Omega K$), all three systems produce approximately the same marginal distributions of downstream TF (Figure 3b). When binding is not sufficiently weak, there can indeed be discrepancies between the full and Hill systems (Figure 3c). These last two results agree with findings in [7,8].

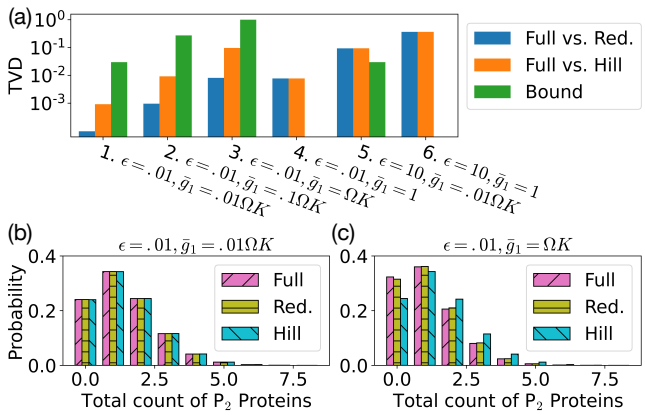

Fig. 3. (a) Simulation results for simple positive regulation for six cases: 1. $\bar{g}_1 = 3$, $\epsilon = 0.01$, $b_1 = 300$/hr; 2. $\bar{g}_1 = 3$, $\epsilon = 0.01$, $b_1 = 30$/hr; 3. $\bar{g}_1 = 3$, $\epsilon = 0.01$, $b_1 = 3$/hr; 4. $\bar{g}_1 = 1$, $\epsilon = 0.01$, $b_1 = 3$/hr; 5. $\bar{g}_1 = 3$, $\epsilon = 10$, $b_1 = 300$/hr ; 6. $\bar{g}_1 = 1$, $\epsilon = 10$, $b_1 = 3$/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, $\kappa_2 = 0/\Omega$/hr, $\alpha_1 = 10$/hr, $\beta_1 = 0$/hr, $\bar{g}_2 = 0$, and $\kappa_1 = K\gamma_1$, where $K := b_1/f_1$. We let the initial state be $x_0 := (p_{1,0}, p_{2,0}, g_{1,0}, g_{2,0}, c_{1,0}, c_{2,0}) = (\Omega K, 0, \bar{g}_1, 0, 0, 0)$. 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 [27], 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 3.

B. Example 2: incoherent feedforward network

In the next example, we investigate the Hill approximation for a TN with an incoherent feedforward loop (IFFL). However, unlike a standard IFFL, 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 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 simpler Hill function model reliably reproduces the probability distribution of the more complicated reduced model during this transient period. Indeed, the TVD error bound at $t = 1$ hr is below 10% (Figure 4c). Note that simulations of the full system were not included due to computational limitations, but Proposition 2 guarantees that for sufficiently small ϵ, the reduced SCRN well approximates the full SCRN.

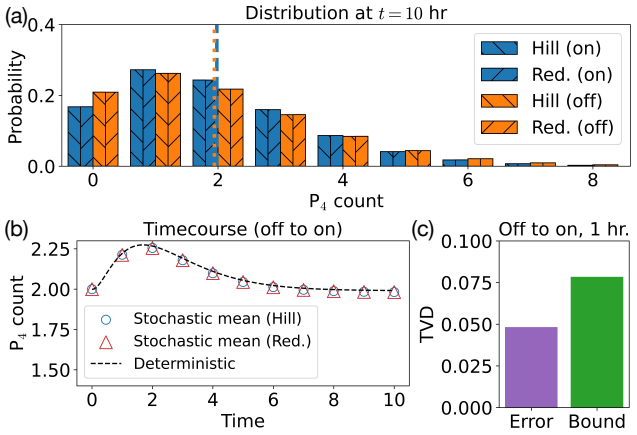

Fig. 4. Results from simulations of the system in Figure 2b, in the "on" state (P_1 and P_2 are produced by $\bar{G}_1^u$) and "off" state (P_1 and P_2 are not produced). In both cases, the system is initialized in the state $x_0 := (p_1, p_2, p_3, p_4, g_1, g_2, g_3, g_4, c_1, c_2, c_3, c_4) = (0, 0, 2, 2, 2, 2, 1, 0, 0, 0, 0, 0)$. 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: $\bar{g}_1 = \bar{g}_2 = 2$, $\bar{g}_3 = 1$, $\bar{g}_4 = 0$, $b_1 = b_2 = b_3 = 100/hr$, $f_1 = f_2 = f_3 = 1\Omega/hr$, $\alpha_1 = 1/hr$, $\beta_2 = 1/hr$, $\alpha_3 = 100/hr$, $\gamma_1 = \gamma_2 = 1/hr$, $\Omega = N_A \cdot 1\mu m^3$, with all other parameters set to 0. (a) Probability distributions of the downstream TF P_4 in the reduced ("Red.") and Hill SCRNs at 10 hours, for the on and off states. For both of these states, the reduced and Hill SCRNs are almost identical. The means of the on/off distributions are shown with dashed/dotted lines. (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 corresponding deterministic Hill approximation trajectory is shown for reference. (c) Empirical TVD ("Error") for the on-state approximation at 1 hr. (from 10^7 simulations performed according to [28]; sampling noise likely causes the empirical TVD to overestimate the true TVD), 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 extend previous work in which the reliability of the Hill function approximation for stochastic TN models was investigated under the additional assumption of weak TF-gene binding. We demonstrated the utility of our result in evaluating the potential error associated with the Hill approximation in models of two TNs. Possible further work includes deriving similar bounds when the same TF may bind to multiple different genes, different TFs compete for the same gene, and TFs may homodimerize and heterodimerize.

APPENDIX

In the proofs we use the following notation. Given a CTMC Y with state space $\mathcal{Y}$, for all $S \subset \mathcal{Y}$ we let $\tau_{Y,S} = \inf\{0 \le t < T_Y^\infty : Y(t) \notin S\}$ 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 $\rho_Y(y,t)$. 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 $\|F\|_1 = \sup_{u \in U} \sum_{u' \in U} |F(u,u')|$.

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 $\mathbb{P}(\tilde{Y}_0 = y_0, \ldots, \tilde{Y}_k = y_k) \neq 0$, $|Q(y_k, y_k)| \le c_1 k + c_2$. Then Y is non-explosive.*

Proof. Fix $t \ge 0$. 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 $k \in \mathbb{Z}_+$,

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

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 $\mathbb{P}(\tilde{Y}_0 = y_0, \ldots, \tilde{Y}_k = y_k) \neq 0$. 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 [24]). Since for each $i \in \{0, \ldots, k\}$, $|Q(y_i, y_i)| \le c_1 i + c_2$ by assumption, we have that $\mathbb{P}(\sum_{i=1}^{k+1} S_i \le t | \tilde{Y}_0 = y_0, \ldots, \tilde{Y}_k = y_k) \le \mathbb{P}(\sum_{i=0}^{k} E_i^* \le t)$. This inequality together with (5) gives that for each k, $\mathbb{P}(T_Y^\infty \le t) \le \mathbb{P}(\sum_{i=0}^{k} E_i^* \le t)$. However, by Theorem 2.3.2(ii) in [24], $\mathbb{P}(\sum_{i=0}^{\infty} E_i^* \le t) = 0$. Since $\lim_{k \to \infty} \mathbb{P}(\sum_{i=0}^{k} E_i^* \le t) = \mathbb{P}(\sum_{i=0}^{\infty} E_i^* \le t) = 0$, then $\mathbb{P}(T_Y^\infty \le t) = 0$. As t was arbitrary, $\mathbb{P}(T_Y^\infty < \infty) = \mathbb{P}(\cup_{t=1}^\infty \{T_Y^\infty \le t\}) \le \sum_{t=1}^\infty \mathbb{P}(T_Y^\infty \le t) = 0$. □

Proof of Lemma 1. We apply Proposition 1 to each CTMC: X_ϵ: 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. For any state y that X_ϵ can access within k transitions, $|Q(y,y)| \le n(\gamma_*(p_{*,0} + c_{*,0} + k) + \kappa_* + (\frac{b_*}{\epsilon} + \frac{f_*}{\epsilon}(p_{*,0} + c_{*,0} + k))\bar{g}_*) + n^2(\alpha_* + \beta_*)\bar{g}_*$, where the subscript $*$ refers to the maximal value of the relevant variable over all i.

$\bar{X}$: Let $\bar{x}_0 = (\bar{p}_{1,0}, \ldots, \bar{p}_{n,0})$ and let $\bar{Q}$ be the infinitesimal generator of $\bar{X}$. For any state y that $\bar{X}$ can access within k

transitions, $|\bar{Q}(y,y)| \leq n(\gamma_*(\bar{p}_{*,0}+k)+\kappa_*)+n^2(\alpha_*+\beta_*)\bar{g}_*$. An identical argument applies to $\bar{X}_H$. □

B. Proof of Lemma 2

Our approach to this proof uses the finite state projection (see [26,27]) to "project" our countable-state CTMC onto one with a finite number of states, at which point we use the results from [15,16]. To then return to the problem with infinite states, we perform an interchange of the limits in the finite state projection and singular perturbation. We note that in [18,19], the authors offer another way to perform model reduction via analysis of stochastic equations. We instead prove this proposition via an approach similar to the formalism of [15,16], leveraging niceties of our system of interest (namely that each fast subsystem has only a finite number of states). We first prove two 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 $\mathcal{C}$ containing y_0, 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 non-explosive CTMC Z with state space $\mathcal{Y}$ and a sequence of finite sets $S_1 \subset S_2 \subset \dots$ whose union is $\mathcal{Y}$, such that for each $\epsilon > 0$ and $k \in \mathbb{N}$*

$$\mathbb{P}(\tau_{Y_\epsilon,S_k} \leq t) \leq \mathbb{P}(\tau_{Z,S_k} \leq t), \tag{6}$$

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. Fix $t \geq 0$. 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 $Tf(C) = \sum_{y\in C} f(y)$.

Enumerate the communicating classes of $\mathcal{C}$ as $C_1, C_2, \dots$, and for each k, let $D_k = \cup_{i=1}^k C_i$ and $\mathcal{D}_k = \{C_1, \dots, C_k\}$. 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

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

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

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 $\eta_k(C,t) = 0$.

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 $j \in \{1,\dots,k\}$, 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. Applying the results for singular perturbation of finite CTMCs in [15,16],

$$\lim_{\epsilon\to 0^+}\|T\theta_{\epsilon,k}(\cdot,t) - \eta_k(\cdot,t)\|_1 = 0. \tag{7}$$

Theorem 2.5 (v) in [26] gives

$$\lim_{k\to\infty}\|\rho_Y(\cdot,t) - \eta_k(\cdot,t)\|_1 = 0, \tag{8}$$

and theorem 2.5 parts (ii) and (iii) in [26] gives

$$\begin{aligned}\|T\rho_{Y_\epsilon}(\cdot,t) - T\theta_{\epsilon,k}(\cdot,t)\|_1 &\leq \|\rho_{Y_\epsilon}(\cdot,t) - \theta_{\epsilon,k}(\cdot,t)\|_1 \\ &\leq \mathbb{P}(\tau_{Y_\epsilon,D_k} \leq t).\end{aligned} \tag{9}$$

Let $S_0 = \emptyset$, and choose $r_1 \leq r_2 \leq \dots \in \mathbb{Z}_+$ such that $r_k \to \infty$ and $S_{r_k} \subset D_k$ for each k. From the definition of the exit time, $\mathbb{P}(\tau_{Y_\epsilon,D_k} \leq t) \leq \mathbb{P}(\tau_{Y_\epsilon,S_{r_k}} \leq t)$. 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 non-explosive, $\mathbb{P}(\tau_{Z,S_{r_k}} \leq t) \to 0$ as $k \to \infty$ by the Lemma 2.1 in [26], so that $\lim_{k\to\infty}\sup_{\epsilon>0}\mathbb{P}(\tau_{Y_\epsilon,D_k} \leq t) = 0$. Combining the above with (9) gives

$$\lim_{k\to\infty}\sup_{\epsilon>0}\|T\rho_{Y_\epsilon}(\cdot,t) - T\theta_{\epsilon,k}(\cdot,t)\|_1 = 0. \tag{10}$$

Given (10) and (7), Proposition 3.3.3 in [29], permits exchanging these limits (with $\|\cdot\|_1$ the underlying norm):

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

The above result together with (8) 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 (6). The following 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 $\mathcal{Y}$, and stoichiometry vectors $u_1, \dots, u_R$. Let $\upsilon_1, \dots, \upsilon_R$ and $\breve{\upsilon}_1, \dots, \breve{\upsilon}_R$ 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 non-explosive and that for each $r \in \{1,\dots,R\}$ 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 a direct result of Theorem 3.4 of [25]. □

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, ϕ, ξ be the respective generic elements of Θ, Φ, and Ξ, and let be 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 $\bar{\mathcal{X}}$.

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 $\mathcal{S}_\epsilon$ 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 $\{(u_{\theta,i},\upsilon_{\theta,i})\}_{\theta,i}$. Let X be a CTMC with state space $\mathcal{E}$, initial state $E_{\bar{x}_0}$, and generator $Q : \mathcal{E}\times\mathcal{E} \to \mathbb{R}$ given by

$$\begin{aligned}
&Q(E_{\bar{x}},E_{\bar{x}'}) := \sum\nolimits_{x\in E_{\bar{x}}}\sum\nolimits_{x'\in E_{\bar{x}'}} \pi_{\bar{x}}(x)B(x,x')\\
&= \sum\nolimits_{x\in E_{\bar{x}}}\sum\nolimits_{x'\in E_{\bar{x}'}} \pi_{\bar{x}}(x)\sum\nolimits_{\theta,i}\upsilon_{\theta,i}(x)\mathbb{1}_{\{x'\}}(x+u_{\theta,i})\\
&= \sum\nolimits_{\theta,i}\sum\nolimits_{x\in E_{\bar{x}}} \pi_{\bar{x}}(x)\upsilon_{\theta,i}(x)\mathbb{1}_{E_{\bar{x}'}}(x+u_{\theta,i})\\
&= \sum_{\theta,i}\mathbb{E}_{W\sim\pi_{\bar{x}}}[\upsilon_{\theta,i}(W)]\mathbb{1}_{\{\bar{x}'\}}(\bar{x}+\Delta\mathbf{T}u_{\theta,i}) \qquad (11)
\end{aligned}$$

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 $(\Delta p_1+\Delta c_1,\ldots,\Delta p_n+\Delta c_n)$.

The proof will consist of first showing

$$\lim_{\epsilon\to 0^+}\sum\nolimits_{\bar{x}\in\mathcal{X}}\left|\sum\nolimits_{x\in E_{\bar{x}}}\rho_{X_\epsilon}(x,t)-\rho_X(E_{\bar{x}},t)\right| = 0, \quad (12)$$

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

$$\rho_X(E_{\bar{x}},t) = \rho_{\bar{X}}(\bar{x},t). \qquad (13)$$

Combining (12) and (13) then gives our desired result.

Towards showing (12), for each $k \in \mathbb{N}$, let $S_k = \{x \in \mathcal{X} : \forall i\ p_i + c_i \leq k\}$. In light of Proposition 2, (12) follows if we can find a non-explosive CTMC Z with state space $\mathcal{X}$ such that for all $\epsilon > 0$ and $k \in \mathbb{N}$,

$$\mathbb{P}(\tau_{X_\epsilon,S_k} \leq t) \leq \mathbb{P}(\tau_{Z,S_k} \leq t). \qquad (14)$$

Define the SCRN $\check{\mathcal{S}}$ with 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 for each $i \in \{1,\ldots,n\}$:

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

Define $L = \begin{bmatrix}\mathbf{e}_{P,1}+\mathbf{e}_{C,1} & \ldots & \mathbf{e}_{P,n}+\mathbf{e}_{C,n}\end{bmatrix}$, where $\mathbf{e}_{P,1},\ldots,\mathbf{e}_{P,n},\mathbf{e}_{G,1},\ldots,\mathbf{e}_{G,n},\mathbf{e}_{C,1},\ldots,\mathbf{e}_{C,n}$ 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 $\check{\mathcal{S}}$. X_ϵ is non-explosive by Lemma 1, and Z is non-explosive because its corresponding propensity functions are bounded (Theorem 2.7.1(ii) in [24]). One can check straightforwardly that the other conditions 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 $\check{Y}$. Thus by Proposition 3, Z satisfies (14). Since $\epsilon > 0$ and $k \in \mathbb{N}$ were chosen arbitrarily, (12) is indeed shown by Proposition 2.

Towards showing (13), note that, 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

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

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

Next, note that there is a natural one-to-one correspondence between $\bar{\mathcal{X}}$ and $\mathcal{E}$ given by $\bar{x} \leftrightarrow E_{\bar{x}}$. Equation (11) 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 $\theta \in \Theta$, $i \in \{1,\ldots,n\}$). But $\hat{u}_{\alpha,i} = \hat{u}_{\beta,i} = \bar{u}_{+,i}$, $\hat{u}_{\kappa,i} = \bar{u}_{+',i}$, and $\hat{u}_{\gamma,i} = \hat{u}_{\gamma',i} = \bar{u}_{-,i}$, and moreover $\hat{\upsilon}_{\alpha,i} + \hat{\upsilon}_{\beta,i} = \bar{\upsilon}_{+,i}$, $\hat{\upsilon}_{\kappa,i} = \bar{\upsilon}_{+',i}$, 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 $\bar{Q}$). Comparing (15) and (11), $Q(E_{\bar{x}},E_{\bar{x}'}) = \bar{Q}(\bar{x},\bar{x}')$, which implies (13). □

C. Proof of Theorem 1

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

$$|\bar{\upsilon}^H_{+,i}(\bar{x}) - \bar{\upsilon}_{+,i}(\bar{x})| \leq |\alpha_i-\beta_i|\frac{\bar{g}_i^2}{\bar{g}_i+\Omega K_i}. \qquad (16)$$

To show the above, choose $i \in \{1,\ldots,n\}$, let $\bar{x} = (\bar{p}_1,\ldots,\bar{p}_n) \in \bar{\mathcal{X}}$, and let $W = (P_1,\ldots,P_n,G_1,\ldots,G_n,C_1,\ldots,C_n)$ be a random variable with distribution $\pi_{\bar{x}}$. From the moment dynamics (see [30]) of $\mathcal{R}$, $\frac{d}{dt}\mathbb{E}[C_i] = \frac{f_i}{\Omega}\mathbb{E}[P_iG_i] - b_i\mathbb{E}[C_i] = \frac{f_i}{\Omega}\mathbb{E}[C_i^2] - \left(b_i + \frac{f_i}{\Omega}(\bar{p}_i+\bar{g}_i)\right)\mathbb{E}[C_i] + \frac{f_i}{\Omega}\bar{p}_i\bar{g}_i$, where $P_i = \bar{p}_i - C_i$ and $G_i = \bar{g}_i - C_i$ because $\pi_{\bar{x}}$ is supported on $E_{\bar{x}}$. But since $\pi_{\bar{x}}$ is a stationary distribution of $\bar{X}$,

$$\mathbb{E}[C_i^2] - (\Omega K_i + \bar{p}_i + \bar{g}_i)\mathbb{E}[C_i] + \bar{p}_i\bar{g}_i = 0. \qquad (17)$$

Since $0 \leq C_i \leq \bar{g}_i$, we have that $0 \leq \mathbb{E}[C_i^2] \leq \bar{g}_i\mathbb{E}[C_i]$. Using this fact and the above equation, it follows that $\frac{\bar{p}_i\bar{g}_i}{\bar{p}_i+\bar{g}_i+\Omega K_i} \leq \mathbb{E}[C_i] \leq \frac{\bar{p}_i\bar{g}_i}{\bar{p}_i+\Omega K_i}$, which implies $\left|\mathbb{E}[C_i] - \frac{\bar{p}_i\bar{g}_i}{\bar{p}_i+\Omega K_i}\right| \leq \frac{\bar{g}_i^2}{\bar{g}_i+\Omega K_i}$. Together with the relation $\mathbb{E}[C_i] + \mathbb{E}[G_i] = \bar{g}_i$, the above implies (16).

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 $S_1 \subset S_2 \subset \ldots$ be a sequence of sets, each of which is finite and whose union is $\bar{\mathcal{X}}$. We will use the

finite state projection to "project" $\bar{X}$ and $\bar{X}_H$ onto each S_k and find a bound on the TVD between the projections which applies uniformly in k.

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

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

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 (v) from [26] guarantees that $\lim_{k\to\infty} \|\rho_{\bar{X}}(\cdot,t) - \bar{\rho}_k(\cdot,t)\|_1 = 0$ and $\lim_{k\to\infty} \|\rho_{\bar{X}_H}(\cdot,t) - \hat{\rho}_k(\cdot,t)\|_1 = 0$, so that

$$d_{TV}(\bar{X}(t),\bar{X}_H(t)) = \frac{1}{2}\|\rho_{\bar{X}}(\cdot,t) - \rho_{\bar{X}_H}(\cdot,t)\|_1 \leq \lim_{k\to\infty}\frac{1}{2}\|\Delta\rho_k(\cdot,t)\|_1. \tag{18}$$

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

$$\frac{d\Delta\rho_k}{ds}(\bar{x},s) = \sum_{\bar{x}'\in S_k} \Delta\rho_k(\bar{x}',s)\bar{Q}(\bar{x}',\bar{x}) + \hat{\rho}_k(\bar{x}',s)\Delta Q(\bar{x}',\bar{x}).$$

The solution to the above finite system of linear ODEs, given the zero initial condition $\Delta\rho_k(\cdot,0) = 0$, is given by the convolution $\Delta\rho_k(\bar{x},t) = \int_0^t \sum_{\bar{x}',\bar{x}''\in S_k} \hat{\rho}_k(\bar{x}'',t-\tau)\Delta Q(\bar{x}'',\bar{x}')e^{\bar{Q}_k\tau}(\bar{x}',\bar{x})d\tau$, where $\bar{Q}_k = \bar{Q}|_{S_k\times S_k}$. Thus

$$\|\Delta\rho_k(\cdot,t)\|_1 \leq \int_0^t \|\hat{\rho}_k(\cdot,t-\tau)\|_1 \|\Delta Q\|_1 \left\|e^{\bar{Q}_k\tau}\right\|_1 d\tau \leq \int_0^t \|\Delta Q\|_1 d\tau, \tag{19}$$

where $\|e^{\bar{Q}_k\tau}\|_1 \leq 1$ because $\bar{Q}$ is an infinitesimal generator. It follows from [21] (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 (16), we have

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

where $\xi \in \{+,+',-\}$. Thus $\|\Delta Q\|_1 \leq 2(n - n_{01}) \max_{i=1,\ldots,n} |\alpha_i - \beta_i|\bar{g}_i^2/(\bar{g}_i + \Omega K_i)$. This inequality, with (18), (19), and Lemma 2, gives the 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] A. Roy and S. Klumpp, "Simulating genetic circuits in bacterial populations with growth heterogeneity." *Biophys. J.*, vol. 114, no. 2, pp. 484–492, 2018.

[6] T. Tian and K. Burrage, "Stochastic models for regulatory networks of the genetic toggle switch," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 22, pp. 8372–8377, 2006.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

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

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] J. R. Norris, *Markov chains*. Cambridge university press, 1998, no. 2.

[25] 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.

[26] 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.

[27] 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.

[28] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *J. Phys. Chem.*, vol. 81, no. 25, pp. 2340–2361, 1977.

[29] T. Terence, *Analysis II*, 3rd ed. Springer, 2016.

[30] 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.