

# Multi-modality in gene regulatory networks with slow promoter kinetics

M. Ali Al-Radhawi<sup>1</sup>, D. Del Vecchio<sup>2</sup>, E. D. Sontag<sup>3\*</sup>,

**1** Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, USA.

**2** Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139-4307, USA.

**3** Departments of Bioengineering and of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, USA. Harvard Program in Therapeutic Science, Harvard Medical School, Boston, MA 02115, USA.

\*sontag@sontaglab.org

## Abstract

Phenotypical variability in the absence of genetic variation often reflects complex energetic landscapes associated with underlying gene regulatory networks (GRNs). In this view, different phenotypes are associated with alternative states of complex nonlinear systems: stable attractors in deterministic models or modes of stationary distributions in stochastic descriptions. We provide theoretical and practical characterizations of these landscapes, specifically focusing on stochastic Slow Promoter Kinetics (SPK), a time scale relevant when transcription factor binding and unbinding are affected by epigenetic processes like DNA methylation and chromatin remodeling. In this case, largely unexplored except for numerical simulations, adiabatic approximations of promoter kinetics are not appropriate. In contrast to the existing literature, we provide rigorous analytic characterizations of multiple modes. A general formal approach gives insight into the influence of parameters and the prediction of how changes in GRN wiring, for example through mutations or artificial interventions, impact the possible number, location, and likelihood of alternative states. We adapt tools from the mathematical field of singular perturbation theory to represent stationary distributions of Chemical Master Equations for GRNs as mixtures of Poisson distributions and obtain explicit formulas for the locations and probabilities of metastable states as a function of the parameters describing the system. As illustrations, the theory is used to tease out the role of cooperative binding in stochastic models in comparison to deterministic models, and applications are given to various model systems, such as toggle switches in isolation or in communicating populations, a synthetic oscillator, and a trans-differentiation network.

## Author summary

Regulatory mechanisms of slow gene activation and deactivation play a role in triggering and sustaining phenotypically heterogeneous, yet genetically identical (clonal), cellular populations in a wide variety of biological processes. These range from embryonic development and hematopoietic cell differentiation to the emergence of tumor heterogeneity and consequent resistance to therapy. In contrast to previously

1  
2  
3  
4  
5

reported numerical simulations, we introduce in this paper a theoretical and computational approach to the characterization of the multi-attractor dynamic landscape of gene networks with slow promoter kinetics. We obtain precise formulas that are then illustrated through applications to several systems biology models including a trans-differentiation network and a communicating population of synthetic toggle switches.

## Introduction

A gene regulatory network (GRN) consists of a collection of genes that transcriptionally regulate each other through their expressed proteins. Through these interactions, including positive and negative feedback loops, GRNs play a central role in the overall control of cellular life [1–3]. The behavior of such networks is stochastic due to the random nature of transcription, translation, and post-translational protein modification processes, as well as the varying availability of cellular components that are required for gene expression. Stochasticity in GRNs is a source of phenotypic variation among genetically identical (clonal) populations of cells or even organisms [4], and is considered to be one of the mechanisms facilitating cell differentiation and organism development [5]. This phenotypic variation may also confer a population an advantage when facing fluctuating environments [6, 7]. Stochasticity due to randomness in cellular components and transcriptional and translational processes have been thoroughly researched [8, 9].

The fast equilibration of random processes sometimes allows stochastic behavior to be “averaged out” through the statistics of large numbers at an observational time-scale, especially when genes and proteins are found in large copy numbers. In those cases, an entire GRN, or portions of it, might be adequately described by a deterministic model. Stochastic effects that occur at a slower time scale, however, may render a deterministic analysis inappropriate and might alter the steady-state behavior of the system. This paper addresses a central question about GRNs: how many different “stable steady states” can such a system potentially settle upon, and how does stochasticity, or lack thereof, affect the answer? To answer this question, it is necessary to understand the possibly different predictions that follow from stochastic versus deterministic models of gene expression. Indeed, qualitative conclusions regarding the steady-state behavior of gene expression levels in a GRN are critically dependent on whether a deterministic or stochastic model is used (see [10] for a recent review). It follows that the mathematical characterization of phenomena such as non-genetic phenotype heterogeneity, switching behavior in response to environmental conditions, and lineage conversion in cells, will depend on the choice of the model.

In order to make the discussion precise, we must clarify the meaning of the term “stable steady state” in both the deterministic and stochastic frameworks. Deterministic models are employed when molecular concentrations are large, or if stochastic effects can be averaged out. They consist of systems of ordinary differential equations describing averaged-out approximations of the interactions between the various molecular species in the GRN under study. For these systems, steady states are the zeroes of the vector field defining the dynamics, and “stable” states are those that are locally asymptotically stable. The number of such stable states quantifies the degree of “multi-stability” of the system. Stochastic models of GRNs, in contrast, are based upon continuous-time Markov chains which describe the random evolution of discrete molecular count numbers. Their long-term behavior is characterized by a stationary Probability Mass Function (PMF) that describes the gene activity configurations and the protein numbers recurrently visited. Under weak ergodicity assumptions, this stationary PMF is unique [11], so multi-stability in the sense of multiple steady states of the Markov chain

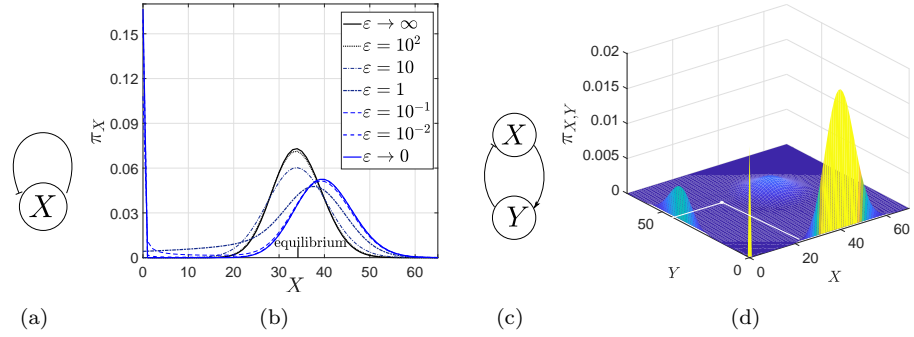
(MC) is not an interesting notion. A biologically meaningful notion of “multi-stability” in this context, and the one that we employ in our study, is “multi-modality,” meaning the existence of multiple modes (local maxima) of stationary PMFs.

Intuitively, given a multi-stable deterministic system, adding noise may help to “shake” states, dislodging them from one basin of attraction of one stable state, and sending them into the basin of attraction of another stable state. Therefore, in the long run, we are bound to see the various deterministic stable steady states with higher probability, that is to say, we expect that they will appear as modes in the stationary PMF of the MC of the associated stochastic model. This is indeed a typical way in which modes can be interpreted as corresponding to stable states, with stochasticity responsible for the transitions between multiple stable states [12]. However, new modes could arise in the stationary PMF of a stochastic system besides those associated with stable states of the deterministic model, and this can occur even if the deterministic model had just a single stable state. This phenomenon of “stochastic multi-stability” has attracted considerable attention lately, both in theoretical and experimental work [8, 9, 13–15]. Stochastic multi-stability has been linked to behaviors such as transcriptional bursting/pulsing [16, 17] and GRN’s binary response [18]. Furthermore, multi-state gene transcription [4] has been used to propose explanations for phenotypic heterogeneity in isogenic populations.

A common assumption in gene regulation models is that transcription factor (TF) to gene binding/unbinding is significantly faster than the rate of protein production and decay [1]. However, it has been proposed [9, 19] that the emergence of new modes in stochastic systems in addition to those that arise from the deterministic model might be due to low gene copy numbers and *Slow Promoter Kinetics* (SPK), which means that the process of binding and unbinding of TFs to promoters is slow. Thus, the emergence of multi-modality may be due to the slow TF-gene binding and unbinding. Already in prokaryotic cells, where DNA is more accessible to TF binding than in eukaryotic cells, some transcription factors can take several minutes to find their targets, comparable or even higher than the time required for gene expression [20], [21], [22]. This is more relevant in eukaryotic cells, in which transcriptional regulation is often mediated by an additional regulation layer dictated by DNA methylation and histone modifications, commonly referred to as chromatin dynamics. For example, the presence of nucleosomes makes binding sites less accessible to TFs and therefore TF-gene binding/unbinding is modulated by the process of chromatin opening [23], [9, 24–26]. DNA methylation, in particular, has also been reported to slow down TF-gene binding/unbinding [27]. Several experiments have consolidated the role of the aforementioned complex transcription processes in SPK [17, 27–29].

In summary, new modes may appear in the stationary PMF that do not correspond to stable states in the deterministic model. Conversely, multiple steady states in the deterministic model may collapse, being “averaged out” by noise, with a single mode representing their mean. It is a well-established fact that, in general, multi-stability of the deterministic description of a biochemical network and multi-modality of the associated stochastic model do not follow from each other [30]. This is especially true in low copy number regimes with SPK. Figure 1 gives two examples for the emergence of new modes due to SPK, and it shows that equilibria derived from the corresponding deterministic model do not provide relevant information on the number and locations of the modes.

Here, we pursue a mathematical analysis of the role of SPK in producing multi-modality in GRNs and show analytically how the shape of the stationary PMF is dictated by key biochemical parameters. Previous studies of the Chemical Master Equation (CME) for single genes have already observed bimodality emerging with slow TF-gene binding/unbinding [19, 31–34]. This phenomenon was also studied by taking



**Fig 1. Emergence of multi-modality due to SPK.** (a) A diagram of a self-repressing gene, where  $\epsilon$  is a parameter that multiplies the kinetic rates of all gene reactions (b) The stationary PMF for different  $\epsilon$  which is showing transition from fast promoter kinetics, i.e.,  $\epsilon \rightarrow \infty$ , to SPK, i.e.,  $\epsilon \rightarrow 0$ , in a non-cooperative self-repressing gene. The stationary PMF is bimodal for small  $\epsilon$  and unimodal for large  $\epsilon$ . The deterministic equilibrium coincides with the fast kinetics mode. Refer to SI-§6.1. (c) A diagram of a repression-activation two-node network. (d) SPK gives rise to four modes while the deterministic model admits a unique stable equilibrium which is marked as a white point, refer to SI-§6.2. The surface is plotted using (10).

the limit of SPK using the linear noise approximation [35], linear mapping approximation [36], or hybrid stochastic models of gene expression [37, 38]. The unregulated gene (SI-§3.1) has been validated for transcriptional bursting [17]. However, and despite its application relevance, mathematical analysis of the CME for multi-gene networks with SPK has been missing, and only numerical solutions have been reported [39, 40].

In this work, an underlying theoretical contribution is the partitioning of the state space into weakly-coupled ergodic classes [11] which, in the limit of slow binding/unbinding, results in the reduction of the infinite-dimensional MC into a finite-dimensional MC whose states correspond to “promoter states”. In this limit, the stationary PMF of the network can be expressed as a mixture of Poisson distributions, each corresponding to conditioning the MC on a certain promoter configuration. The framework proposed here enables us to analytically determine how the number of modes, their locations, and weights depend on the biophysical parameters. Hence, the proposed framework can be applied to GRNs to predict the different phenotypes that the network can exhibit with low gene copy numbers and SPK.

The results are derived by introducing a new formalism to model GRNs with arbitrary numbers of genes, based on continuous-time MCs. Then, we analyze the stationary solution of the associated CME through a systematic application of the method of singular perturbations [41]. Specifically, we study the SPK limit by letting the ratio of kinetic rate constants of the TF-gene binding/unbinding reactions with respect to protein reactions approach zero. The stationary solution is computed for the singularly-perturbed CME.

In order to illustrate the practical significance of our results, we work out several examples, some of which have not been studied before in the literature. As a first application, we discover that, with SPK, a self-regulating gene can exhibit bimodality even with non-cooperative binding to the promoter site. We then investigate the role of cooperativity. In contrast to deterministic systems, we find that cooperativity does not change the number of modes. Nevertheless, cooperativity adds extra degrees of freedom by allowing the network to tune the relative weight of each mode without changing its

location.

As a second application, we revisit the classical toggle switch, under slow TF-gene binding/unbinding. It has been reported before that, with fast TF-gene binding/unbinding, the toggle switch with single-gene copies can be “bistable” without cooperative binding [42]. We show that this can also happen with SPK, and, moreover, that a new mode having both proteins at high copy numbers can emerge. We provide a method to calculate the weight of each mode and show that the third mode is suppressed for sufficiently high kinetic rates for the dimerization reactions.

A third application that we consider is a simplified model of synchronization of communicating toggle switches. In bacterial populations, quorum sensing has been proposed [43] as a way for bacterial cells to broadcast their internal states to other cells in order to facilitate synchronization. Quorum sensing communication has been adopted also as a tool in synthetic biology [44, 45]. Mathematical analysis of coupled toggle switches designs usually employs deterministic models [46]. We study a simplified stochastic model of coupled toggle switches with SPK and compare the resulting number of modes with deterministic equilibria.

Our final, and potentially most significant, application is motivated by cellular differentiation. A well-known metaphor for cell lineage specification arose from the 1957 work of Waddington [47], who imagined an “epigenetic landscape” with a series of branching valleys and ridges depicting stable cellular states. In that context, the emergence of new modes in cell fate circuits is often interpreted as the creation of new valleys in the epigenetic landscape, and (deterministic) multi-stability is employed to explain cellular differentiation [5]. However, an increasing number of studies have suggested stochastic heterogeneous gene expression as a mechanism for differentiation [13, 48, 49]. Numerical analysis of the CME for the canonical cell-fate circuit have shown the emergence of new modes due to SPK in such models [39, 50]. This general category of cell-fate circuits includes pairs such as PU.1:GATA1, Pax5:C/EBP $\alpha$  and GATA3:T-bet [51]. Cell fate circuits are characterized by TF cross-antagonism. However, their behavior is affected by the promoter configurations available for binding, the cooperativity index of the TFs, and the relative ratio of production rates. Hence, we study two models that differ in the aforementioned aspects and we highlight the differences between our findings and the behavior predicted by the corresponding deterministic model. The first model employs independent cooperative binding. We show that such a network can exhibit more than four modes. In contrast, the deterministic model predicts up to four modes only with cooperativity [52]. The second network is a PU.1/GATA.1 network which employs non-cooperative binding and a restricted set of promoter configurations. The deterministic model is monostable, while the parameters of the stochastic model can be chosen to have additional modes including the cases of bistability and tristability.

## The Reaction Network Structure

In this paper, a GRN is formally defined as a set of nodes (genes) that are connected with each other through regulatory interactions via the proteins that the genes express. The regulatory proteins are called *transcription factors* (TFs). A TF regulates the expression of a gene by reversibly binding to the gene’s promoter and by either enhancing expression or repressing it.

The formalism we employ in order to describe GRNs at the elementary level is that of Chemical Reaction Networks (CRNs) [53]. A CRN consists of *species* and *reactions*, which we describe below.

**Species:** The species in our context consist of promoter configurations for the various genes participating in the network, together with the respective TFs expressed from these genes and some of their multimers. A configuration of a promoter is characterized by the possible locations and number of TFs bound to the promoter at a given time. If a promoter is expressed constitutively, then there are two configurations specifying the expression activity state, active or inactive. A multimer is a compound consisting of a protein binding to itself several times. For instance, dimers and trimers are 2-mers and 3-mers, respectively. If a protein forms an  $n^{\text{th}}$ -order multimer then we say that it has a cooperativity index of  $n$ . If species is denoted by  $X$ , then its copy number is denoted by  $X$ . The set of all species is  $\mathcal{S}$ .

For simplicity we assume the following:

- (A1) Each promoter can have up to two TFs binding to it.;
- (A2) Each TF is a single protein that has a fixed cooperativity index, i.e, it cannot act as a TF with two different cooperativity indices;
- (A3) Each gene is present with only a *single copy*.

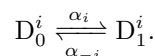
All the above assumptions can be relaxed. SI-§4,5 contain generalizations of the results to heterogeneous TFs, and arbitrary numbers of gene copy numbers.

Consider the  $i^{\text{th}}$  promoter. The expression rate of a gene is dependent on the current configuration of its promoter. We call the set of all possible such configurations the *binding-site set*  $B_i$ . Each member of  $B_i$  corresponds to a configuration that translates into a specific species  $D_j^i, j \in B_i$ . If a promoter has just one or no regulatory binding sites, then we let  $B_i = \{0, 1\}$ . Hence, the promoter configuration can be represented by *two species*: the unbound species  $D_0^i$  and the bound species  $D_1^i$ . If the promoter has no binding sites then the promoter configuration species are interpreted as the inactive and active configurations, respectively. On the other hand, if the promoter has *two binding sites* then  $B_i = \{00, 01, 10, 11\}$ <sup>1</sup>. The first digit in a member of  $B_i$  specifies whether the first binding site is occupied, and the second digit specifies the occupancy of the second binding site. Hence, the promoter configuration can be represented by four species  $D_{00}^i, D_{10}^i, D_{01}^i, D_{11}^i$ . Note that in general we need to define  $2^\kappa$  species for a promoter with  $\kappa$  binding sites.

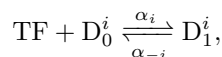
The species that denotes the protein produced by the  $i^{\text{th}}$  gene is  $X_i$ . A protein's multimer is denoted by  $X_{ic}$ . If protein  $X_i$  does not form a multimer then  $X_{ic} := X_i$ .

**Reactions:** In our context, the reactions consist of TFs binding and unbinding with promoters and the respective protein expression (with transcription and translation combined in one step), decay, and  $n$ -merization. For each gene, we define a *gene expression block*. Each block consists of a set of *gene reactions* and a set of *protein reactions* as shown in Figure 2.

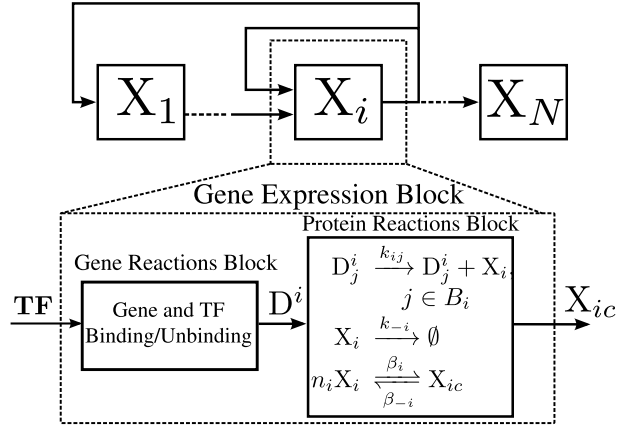
If the promoter is constitutive, i.e. it switches between two configurations autonomously without an explicitly modeled TF-promoter binding, then  $B_i = \{0, 1\}$  and the gene reactions block consists of:



We refer to  $D_0^i$  and  $D_1^i$  as the *inactive* and *active* configurations, respectively. If the promoter has one binding site, then also  $B_i = \{0, 1\}$  and the gene reactions block consists of just two reactions:

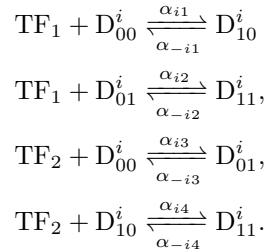


<sup>1</sup>We interpret the elements of the binding set as integers in binary representation.



**Fig 2. A Gene Expression Block.** A GRN that consists of gene expression blocks. A block consists of a gene reactions block and a protein reactions block. The gene reactions are described in the text.  $\mathbf{TF}$  is a vector of TFs which can be monomers, dimers, or higher order multimers.  $D^i$  is a vector whose components consist of the  $D_j^i$ 's. The dimension of  $\mathbf{TF}$  is equal to the number of binding sites of the gene.

where  $D_0^i$  and  $D_1^i$  are the promoter configurations when *unbound* and *bound* to the TF, respectively. Note that we did not designate a specific species as the active one since it depends on whether the TF is an activator or a repressor. Specifically, when TF is an activator,  $D_1^i$  will be the active configuration and  $D_0^i$  will be the inactive configuration, and vice versa when TF is a repressor. Finally, if the promoter has two TFs binding to it, then they can bind *independently*, *competitively*, or *cooperatively*. Cooperative binding is discussed in SI-§4.3.1. If they bind independently, then the promoter has two binding sites. Hence,  $B_i = \{00, 01, 10, 11\}$  and the gene block contains the following *gene reactions*:

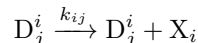


The activity of each configuration species is dependent on whether the TFs are 229  
 activators or repressors, and on how they behave jointly. This can be characterized fully 230  
 by assigning a production rate for each configuration as will be explained below. In the 231  
 case of competitive binding, two different TFs compete to bind to the same location. 232  
 This can be modeled similarly to the previous case except that the transitions to  $D_{11}^i$ , i.e. 233  
 the configuration where both TFs are bound, are not allowed. Hence, the gene reactions 234  
 block will have only the first and third, and the binding set reduces to  $B_i = \{00, 01, 10\}$ . 235

Our binding/unbinding reaction models account for the stoichiometric change in 236  
 proteins when binding to the promoters, and the bound proteins are accounted for by 237  
 designating a gene state for each promoter configuration (defined by the number and 238  
 location of bound proteins). The bound protein molecules are only governed by slow 239  
 gene reactions until they are released. 240

We assume that RNA polymerase and ribosomes are available in high copy numbers, 241  
 and that we can lump transcription and translation into one simplified “production” 242

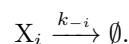
reaction. The latter assumption is a reasonable approximation since the turnover of mRNA is typically faster than that of protein. The rate of production is dependent on the promoter's configuration. So for each configuration  $D_j^i, j \in B_i$  the *production* reaction is:



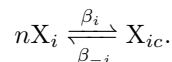
, where the kinetic constant  $k_{ij}$  is a non-negative number. The case  $k_{ij} = 0$  means that when the promoter configuration is  $D_j^i$  there is no protein production, and hence  $D_j^i$  is an inactive configuration. The promoter configuration can be ranked from the most active to the least active by ranking the corresponding production kinetic rate constants.

Consequently, the character of a TF is manifested as follows: if the maximal protein production occurs at a configuration with the TF being bound we say that the TF is *activating*, and if the reverse holds it is *repressing*. And, if the production is maximal with multiple configurations such that the TF is bound in some of them and unbound in others then the TF is neither repressing nor activating.

We model decay and/or dilution as a single reaction:



The expressed proteins can act as TFs. They may combine to form dimers or higher order multimers before acting as TFs. The numbers of copies of the TF needed to form a multi-mer is called the *the cooperativity index* and we denote it by  $n$ . Hence, we model the cooperativity reactions as given in Figure 2 as follows (called the *n-imerization* reactions):



If the cooperativity index of  $X_i$  is 1, then the species  $X_{ic} := X_i$ , and the multimerization reaction becomes empty. Higher order multi-merization processes can be modelled as multi-step or sequential reactions [54]. We discuss how our theory includes this case in SI-§4.3.3, by showing how an equivalent one-step model can be formulated.

**Kinetics:** In order to keep track of molecule counts, each species  $Z_i \in \mathcal{S}$  is associated with a copy number  $z_i \in \mathbb{Z}_{\geq 0}$ . To each reaction  $R_j$  one associates a propensity function  $R_j$ . We use the well-known *Mass-Action Kinetics*, which are reviewed in SI-§1.1.

**A gene regulatory network:** Consider a set of  $N$  genes, binding sets  $\{B_i\}_{i=1}^N$ , and kinetic constants  $k_j$ 's. A *gene expression block*, as shown in Figure 2, is a set of gene reactions and protein reactions as defined above. Each gene block has an output that is either the protein or its  $n$ -mer, and it is designated by  $X_{ic}$ . The input to each gene expression block is a subset of the set of the outputs of all blocks. Then, a GRN is an arbitrary interconnection of gene expression blocks (Figure 2). SI-§4 defines a more general class of network that we can study. A directed graph can be associated with a GRN as follows. Each vertex corresponds to a gene expression block. There is a directed edge from vertex A to vertex B if the output of A is an input to B. For simplicity, we assume the following: **(A4)** The graph of gene expression blocks is connected. Note that if A4 is violated, our analysis can be applied to each connected component.

**Time-Scale Separation:** As mentioned in the introduction, we assume that the gene reactions are considerably slower than the protein reactions. In order to model this assumption, we write the kinetic rates of gene reactions in the form  $\varepsilon k_j$ , where  $0 < \varepsilon \ll 1$  and assume that all other kinetic rates (for protein production, decay and multi-merization) are  $\varepsilon^{-1}$ -times faster.



Events in biological cells usually take place at different time-scales [1], and hence singular perturbation techniques are widely used in deterministic settings in order to reduce models for analysis. On the other hand, model-order reduction by time-scale separation in stochastic processes has been mainly used in the literature for computational purposes, for example to accelerate the stochastic simulation algorithm [55, 56], or to compute finite-space-projection solutions to the CME [57]. In this work, we use a singular perturbation approach for the analytical purpose of characterizing the form of the stationary PMF in the regimes of slow gene-TF binding/unbinding.

In the case of a finite MC, the CME is a finite-dimensional linear ODE, and reduction methods for linear systems can be used [41] and applied to MCs [58, 59]. For continuous-time MCs on a countable space, as needed when analyzing gene networks, there are difficult and open technical issues. Exponential stochastic stability [60] needs to be established for the stationary solution in order to guarantee the existence of the asymptotic expansion in  $\varepsilon$  [61]. Although it has been shown for a class of networks [62], the general problem needs further research. In this paper, we will not delve into technical issues of stochastic stability ; we assume that these expansions exist and that the solutions converge to a unique equilibrium solution.

## Dynamics and the CME

The dynamics of the network refers to the manner in which the *state* evolves in time, where the state  $Z(t) \in \mathbf{Z} \subset \mathbb{Z}_{\geq 0}^{|\mathcal{S}|}$  is the vector of copy numbers of the species of the network at time  $t$ . The standard stochastic model for a CRN is that of a continuous-time MC. Let the state be  $Z(t) = z \in \mathbf{Z}$ , where  $\mathbf{Z}$  is the state-space. The relevant background is reviewed in SI-§1.1.

Let  $p_z(t) = \Pr[Z(t) = z | Z(0) = z_0]$  be the stationary PMF for any given initial condition  $z_0$ . Its time evolution is given by the CME.

Since our species are either gene species or protein species, we split the stochastic process  $Z(t)$  into two subprocesses: *the gene process*  $D(t)$  and *the protein process*  $X(t)$ , as explained below.

For each gene we define one process  $D_i$  such that  $D_i(t) \in B_i$ .  $D_i(t) = j$  if and only if the promoter configuration is encoded by  $j \in B_i$ . Collecting these into a vector, define the gene process  $D(t) := [D_1(t), \dots, D_N(t)]^T$  where  $D(t) \in \prod_{i=1}^N B_i$ . The  $i^{\text{th}}$  gene can be represented by  $|B_i|$  states, so  $L := \prod_{i=1}^N |B_i|$  is the total number of promoter configurations in the GRN. With abuse of notation, we write also  $D(t) \in \{0, \dots, L-1\}$  in the sense of the bijection between  $\{0, \dots, L-1\}$  and  $\prod_{i=1}^N B_i$  defined by interpreting  $D_1 \dots D_N$  as a binary representation of an integer. Hence,  $d \in \{0, \dots, L-1\}$  corresponds to  $(d_1, \dots, d_N) \in B_1 \times \dots \times B_N$  and we write  $d = (d_1, \dots, d_N)$ .

Since each gene expresses a corresponding protein, we define  $X_{i1}(t) \in \mathbb{Z}_{\geq 0}$ ,  $i = 1, \dots, N$  protein processes. If the multimerized version of the  $i^{\text{th}}$  protein participates in the network as an activator or repressor then we define  $X_{ic}(t)$  as the corresponding multimerized protein process, and we denote  $X_i(t) := [X_{i1}(t), X_{ic}(t)]^T$ . If there is no multimerization reaction then we define  $X_i(t) := X_{i1}(t)$ . Since not all proteins are necessarily multimerized, the total number of protein processes is  $N \leq M \leq 2N$ . Hence, the *protein process* is  $X(t) = [X_1^T(t), \dots, X_N^T(t)]^T \in \mathbb{Z}_{\geq 0}^M$  and  $\mathbf{Z} = \mathbb{Z}_{\geq 0}^M \times \prod_{i=1}^N B_i$ .

# Results

## Decomposition of the CME

It is crucial to our analysis to represent the linear system of differential equations given by the CME as an interconnection of weakly coupled linear systems. To this end, we present the appropriate notation in this subsection.

Consider the joint PMF:  $p_{d,x}(t) = \Pr[X(t) = x, D(t) = d]$ , which represents the probability at time  $t$  that the protein process  $X$  takes the value  $x \in \mathbb{Z}_+^M$  and the gene process  $D$  takes the value  $d \in \{0, \dots, L-1\}$ . Recall that  $x$  is a vector of copy numbers for the protein processes while  $d$  encodes the configuration of each promoter in the network. Then, we can define for each fixed  $d$ :  $p_d(t) := [p_{dx_0}(t), p_{dx_1}(t), \dots]^T$ , representing the vector enumerating the joint probabilities for all values of  $x$  and for a fixed  $d$ , where  $x_0, x_1, \dots$  is an indexing of  $\mathbb{Z}_{\geq 0}^M$ . Note that  $p_d(t)$  can be thought of as an infinite vector with respect to the aforementioned indexing. Finally, let

$$p(t) := [p_0(t)^T, \dots, p_{L-1}^T(t)]^T, \quad (1)$$

representing a concatenation of the vectors  $p_{d,x}(t)$  for  $d = 0, \dots, L-1$ . Note that  $p(t)$  is a finite concatenation of infinite vectors. The joint stationary PMF  $\bar{\pi}$  is defined as the following limit, which we assume to exist and is independent of the initial PMF:

$\bar{\pi} = \lim_{t \rightarrow \infty} p(t)$ . Note that  $\bar{\pi}$  depends on  $\varepsilon$ .

Consider a given GRN. The CME is defined over a countable state space  $\mathbf{Z}$ . Hence, the CME can be interpreted as an infinite system of differential equations with an infinite infinitesimal generator matrix  $\Lambda$  which contains the reaction rates (see SI-§1.1).

Consider partitioning the PMF vector as in (1). Recall that reactions have been divided into two sets: slow gene reactions and fast protein reactions. This allows us to write  $\Lambda$  as a sum of a slow matrix  $\varepsilon \hat{\Lambda}$  and a fast matrix  $\tilde{\Lambda}$ , which we call a fast-slow decomposition. Furthermore,  $\tilde{\Lambda}$  can be written as a block diagonal matrix with  $L$  diagonal blocks which correspond to conditioning the MC on a specific gene state  $d$ . This is stated in the following basic proposition (see SI-§2.1 for the proof):

**Proposition 1.** *Given a GRN. Its CME can be written as*

$$\dot{p}(t) = \Lambda_\varepsilon p(t) = (\tilde{\Lambda} + \varepsilon \hat{\Lambda}) p(t), \quad (2)$$

where  $p(t) = [p_0^T(t), \dots, p_{L-1}^T(t)]^T$ , and

$$\tilde{\Lambda} = \text{diag}[\Lambda_0, \dots, \Lambda_{L-1}] \quad (3)$$

where  $\tilde{\Lambda}$  is the fast matrix,  $\hat{\Lambda}$  is the slow matrix, and  $\Lambda_0, \dots, \Lambda_{L-1}$  are stochastic matrices.

## Conditional MCs

For each  $d$ , consider modifying the MC  $Z(t)$  defined in the previous section by replacing the stochastic process  $D(t)$  by a deterministic constant process  $D(t) = d$ . This means that the resulting MC does not describe the gene process dynamics, it only describes the protein process dynamics *conditioned on  $d$* . Henceforth, we refer to the resulting MC as the *MC conditioned on  $d$* . The infinitesimal generator of a MC conditioned on  $d$  is denoted by  $\Lambda_d$ , and is identical to the corresponding block on the diagonal of  $\tilde{\Lambda}$  as given in (3). In other words, fixing  $D(t) = d \in \{0, \dots, L-1\}$ , the dynamics of the network can be described by a CME:

$$\dot{p}_{X|d} = \Lambda_d p_{X|d}, \quad (4)$$

$\dot{p}_{X|d} = \Lambda_d p_{X|d}$ , where  $p_{X|d}$  is a vector that enumerates the conditional probabilities  $p_{x|d} = \Pr[X(t) = x | D(t) = d]$  for a given  $d$ . The conditional stationary PMF is denoted by:  $\pi_{X|d}^{(J)} = \lim_{t \rightarrow \infty} p_{X|d}(t)$ , where  $(J)$  refers to the fact that it is joint in the protein and multimerized protein processes. Note that  $\pi_{X|d}^{(J)}$  is independent of  $\varepsilon$ . This notion of a conditional MC is useful since, at the SPK limit,  $D(t)$  stays constant. It can be noted from (3) that when  $\varepsilon = 0$  the dynamics of  $p_d$  decouples and becomes independent of  $p_{\tilde{d}}, \tilde{d} = 0, \dots, L-1, \tilde{d} \neq d$ .

We show below that each conditional MC has a simple structure. Fixing the promoter configuration  $D(t) = d = (d_1, \dots, d_N)$ , the network consists of *uncoupled* birth-death processes. So for each  $d_i$ , the protein reactions of production and dimerization corresponding to the  $i^{\text{th}}$  promoter can be written as follows without multimerization:  $\emptyset \xrightleftharpoons[k_{-i}]{k_{id_i}} X_i$ , where the subscript  $id_i$  refers to the production kinetic constant corresponding to the configuration species  $D_{d_i}^i$ , or, if there is a multimerization reaction, it takes the form:  $\emptyset \xrightleftharpoons[k_{-i}]{k_{id_i}} X_i, n_i X_i \xrightleftharpoons[\beta_{-i}]{\beta_i} X_{ic}$ . Note that the stochastic processes  $X_i(t), i = 1, \dots, N$  conditioned on  $D(t) = d$  are independent of each other. Hence, the conditional stationary PMF  $\pi_{X|d}^{(J)}$  can be written as a product of stationary PMFs and the individual stationary PMFs have Poisson expressions. The following proposition gives the analytic expression of the conditional stationary PMFs: (see SI-§2.2 for proof)

**Proposition 2.** Fix  $d \in \{0, \dots, L-1\}$ . Consider (4), then there exists a conditional stationary PMF  $\pi_{X|d}^{(J)}$  and it is given by

$$\pi_{X|d}^{(J)}(x) = \prod_{i=1}^N \pi_{X|d_i}(x_i), \quad (5)$$

where

$$\pi_{X|d_i}^{(J)}(x_i) = \begin{cases} \mathbf{P}\left(x_{i1}, x_{i2}; \frac{k_{id_i}}{k_{-i}}, \frac{k_{id_i}^{n_i} \beta_i}{n_i! k_{-i}^{n_i} \beta_{-i}}\right) : n_i > 1 \\ \mathbf{P}\left(x_i; \frac{k_{id_i}}{k_{-i}}\right) : n_i = 1, \end{cases} \quad (6)$$

where  $(J)$  refers to the joint PMF in multimerized and non-multimerized processes,  $x_{i1}$  refers to the copy number of  $X_i$ , while  $x_{i2}$  refers to the copy number of  $X_{ic}$ ,

$$\mathbf{P}(x; a) := \frac{a^x}{x!} e^{-a}, \mathbf{P}(x_1, x_2; a_1, a_2) := \frac{a_1^{x_1}}{x_1!} \frac{a_2^{x_2}}{x_2!} e^{-a_1 - a_2}.$$

**Remark 1.** The conditional PMF in (5) is a joint PMF in the protein and multimerized protein processes. If we want to compute a marginal stationary PMF for the protein process only, then we average over the multimerized protein processes  $X_{ic}, i = 1, \dots, N$  to get a joint Poisson in  $N$  variables. Hence, the formulae (5)-(6) can be replaced by:

$$\pi_{X|d}(x) := \sum_{i=1}^{M-N} \sum_{x_{i2}=0}^{\infty} \pi_{X|d}^{(J)}(x) = \prod_{i=1}^N \mathbf{P}\left(x_i; \frac{k_{id_i}}{k_{-i}}\right), \quad (7)$$

where  $M - N$  is the number of  $n$ -merized protein processes, and  $\pi_{X|d}$  is the marginal stationary PMF for the protein process.

## Decomposition of The Stationary Distribution

Recall the slow-fast decomposition of the CME in (2) and the joint stationary PMF  $\bar{\pi}$ . In order to emphasize the dependence on  $\varepsilon$  we denote  $\bar{\pi}^\varepsilon := \bar{\pi}(\varepsilon)$ . Hence,  $\bar{\pi}^\varepsilon$  is the

unique stationary PMF that satisfies  $\Lambda_\varepsilon \bar{\pi}^\varepsilon = 0$ ,  $\bar{\pi}^\varepsilon > 0$ , and  $\sum_z \bar{\pi}_z^\varepsilon = 1$ , where the subscript denotes the value of the stationary PMF at  $z$ .

Our aim is to characterize the stationary PMF as  $\varepsilon \rightarrow 0$ . Writing  $\bar{\pi}_\varepsilon$  as an asymptotic expansion to first order in terms of  $\varepsilon$ , we have

$$\bar{\pi}^\varepsilon = \bar{\pi}^{(0)} + \bar{\pi}^{(1)}\varepsilon + o(\varepsilon). \quad (8)$$

Our aim is to find  $\bar{\pi}^{(0)}$ . We use singular perturbations techniques to derive the following theorem (see SI-§2.3):

**Theorem 3.** Consider a given GRN with  $L$  promoter states with the CME (2).

Writing (8), then the joint stationary PMF  $\bar{\pi} := \lim_{\varepsilon \rightarrow 0^+} \bar{\pi}^\varepsilon$  can be written as:

$\bar{\pi}(x, d) = \sum_{d=0}^{L-1} \lambda_d \bar{\pi}_{X|d}(x, d)$ , where  $\lambda = [\lambda_0, \dots, \lambda_{L-1}]^T$  is the principal normalized eigenvector of:

$$\Lambda_r := \text{diag}[\mathbf{1}^T, \dots, \mathbf{1}^T] \hat{\Lambda} [\bar{\pi}_{X|0} \ \bar{\pi}_{X|1} \ \dots \ \bar{\pi}_{X|L-1}], \quad (9)$$

where  $\bar{\pi}_{X|0}, \dots, \bar{\pi}_{X|L-1}$  are the extended conditional stationary PMFs defined as:

$$\bar{\pi}_{X|d}(x, d) = \pi_{X|d}(x), \quad \bar{\pi}_{X|d}(x, d') = 0 \text{ when } d' \neq d.$$

The result characterizes the stationary solution of (2) which is a joint PMF in  $X$  and  $D$ . However, we are particularly interested in the marginal stationary PMF of the protein process  $X$  and the marginal stationary PMF of the non-multimerized protein process, since these PMFs are typically experimentally observable. Therefore, we can use Remark 1 to write the stationary PMF as mixture of  $L$  Poisson distributions with weights  $\{\lambda_d\}_{d=0}^{L-1}$ :

**Corollary 4.** Consider a given GRN with  $L$  genes with the CME (2). Writing (8), let  $\pi_{X|0}, \dots, \pi_{X|L-1}$  be the conditional stationary PMFs of  $\Lambda_0, \dots, \Lambda_{L-1}$ , where explicit expressions are given in (5). Then, we can write the following

$$\pi^{(J)}(x) := \lim_{\varepsilon \rightarrow 0^+} \lim_{t \rightarrow \infty} \Pr[X(t) = x] = \sum_{d=0}^{L-1} \lambda_d \pi_{X|d}^{(J)}(x), \text{ where } \lambda = [\lambda_0, \dots, \lambda_{L-1}]^T$$

is as given Theorem 3. Furthermore, the marginal stationary PMF of the non-multimerized protein process can be written as:

$$\pi(x) := \sum_{d=0}^{L-1} \lambda_d \pi_{X|d}(x) = \sum_{d=0}^{L-1} \lambda_d \prod_{i=1}^N \mathbf{P}\left(x_i; \frac{k_i d_i}{k-i}\right). \quad (10)$$

**Remark 2.** In the remainder of the Results section, when we refer to the “stationary PMF” we mean the marginal stationary PMF of the non-multimerized protein process given in (10).

**Remark 3.** If a mode is defined as a local maximum of a stationary PMF, then this does not necessarily imply that the stationary PMF has  $L$  modes since the peak values of two Poisson distributions can be very close to each other. In the remainder of the paper we will call each Poisson distribution in the mixture as a “mode” in the sense that it represents a component in the mixture PMF. The number of local maxima of a PMF can be found easily given the expression (10).

## The Reduced-Order Finite MC

The computation of the weighting vector  $\lambda$  in Theorem 3 requires computing the  $L \times L$  matrix  $\Lambda_r$  in (9) which can be interpreted as the infinitesimal generator of an  $L$ -dimensional MC. The expression in (9) involves evaluating the product of infinite dimensional matrices. Since the structure of the GRN and the form of the conditional PMF in (5) are known, an easier algorithm to compute  $\Lambda_r$  for our GRNs is given in

Proposition SI-2. The algorithm provides an intuitive way to interpret Theorem 3 and can be informally described as follows.

Let  $D(t) = d_i$ , the algorithm implies that each binding reaction of the form:  $\text{TF} + D_{d_i}^i \xrightarrow{\alpha} D_{d_i'}^i$ , gives the rate  $\alpha \mathbb{E}[\text{TF}|D = d]$ , where  $\mathbb{E}$  denotes mathematical expectation. Hence it corresponds to a reaction of the following form in the reduced-order MC:



Using Proposition 2, we can write: (See SI-§2.4)

$$\mathbb{E}[\text{TF}|D = d] = \frac{\alpha}{n_i!} \frac{\beta_i}{\beta_{-i}} \left( \frac{k_{id_i}}{k_{-i}} \right)^{n_i}. \quad (12)$$

## Generalization

Theorem 3 and Corollary 4 have been stated for GRNs that have gene expression blocks of the form given in Figure 2. Nevertheless, the same results can also be stated for a larger class of networks. The generalized class consists of GRNs with weakly reversible deficiency zero conditional Markov chains. The stationary PMF for networks in this class can also be expressed as a mixture of Poisson PMFs. This enables us to include networks with hetero-dimerization, diffusion and multi-step multi-merization in our study. The full details are in SI-§4, and a diffusion-based interconnection of toggle switches will be studied as an example.

## The Gene Bursting Model

The simplest network is the unregulated gene which is used for transcriptional bursting [17] and studied using time-scale separation in [32, 63]. Consider:

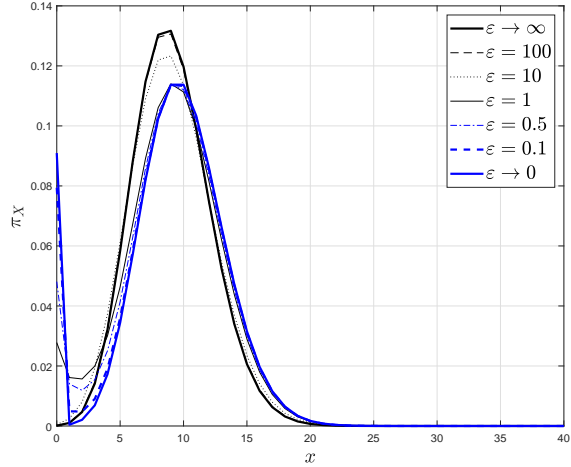


Referring to Figure 2, we identify a single gene block with two states. Using (5), the conditional stationary PMFs are two Poissons at 0 and  $k/k_-$ , and the stationary PMF is a bimodal mixture of them with weights  $\alpha_-/(\alpha + \alpha_-)$ ,  $\alpha/(\alpha + \alpha_-)$ , respectively (See SI-§3.1). In the case of fast promoter kinetics, the resulting stationary PMF is a Poisson with mean  $\frac{\alpha}{\alpha + \alpha_-} \frac{k}{k_-}$  which coincides with the deterministic equilibrium. Although both stochastic models share the mean, the stationary PMFs and their variances differ drastically.

Figure 3 shows the transition from fast to slow promoter kinetics using the exact solution [64] and compares it to the predicted mixture of two Poissons.

## The Role of Cooperativity

A TF is said to be cooperative if it acts only after it forms a dimer or a higher-order  $n$ -mer that binds to the gene's promoter [54]. In standard deterministic modelling, a cooperative activation changes the form of the quasi-steady state activation rate from a Michaelis-Menten function into a Hill function. Cooperativity is often necessary for a network to have multiple equilibria in some kinetic parameter ranges. For example, a



**Fig 3.** The stationary probability distribution for different  $\varepsilon$  which shows the transition from fast promoter kinetics, i.e.,  $\varepsilon \rightarrow \infty$ , to slow promoter kinetics, i.e.,  $\varepsilon \rightarrow 0$ , in a single unregulated gene. The stationary distribution is bimodal for small  $\varepsilon$ , i.e.  $\varepsilon \leq 1$ , and unimodal for large  $\varepsilon$ . The deterministic equilibrium coincides with the fast kinetics mode at  $\frac{\alpha}{\alpha + \alpha_-} \frac{k}{k_-}$ . The slow kinetic limit is calculated via Corollary 4, the fast kinetics limit is a Poisson centered at the deterministic equilibrium, while the remaining curves are computed by evaluating the exact solution given in [64]. The parameters are  $\alpha_- = 0.1, \alpha = 1, k_- = 2, k = 20$ .

non-cooperative self-activating gene can only be mono-stable, while its cooperative counterpart can be multi-stable for some parameters. 473

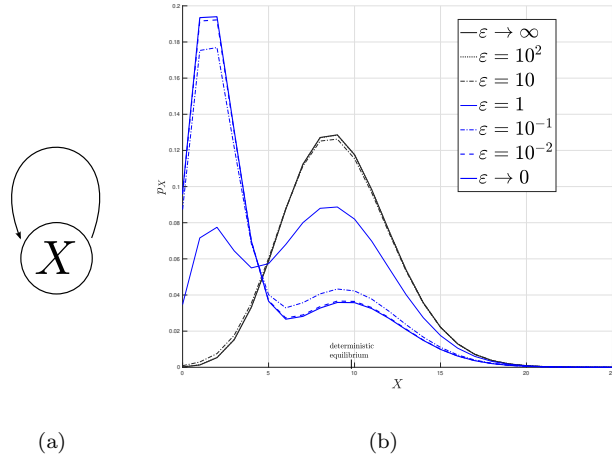
Corollary 4 and (12) show that cooperativity plays in the context of SPK a role that is very different from the deterministic setting. This is since the stationary PMF is a mixture of  $L$  Poisson processes (7) which are independent of the TFs' cooperativity indices and ratios. In the non-cooperative case, a certain mode can be made more probable only by changing either the location of the mode or the dissociation ratio (the ratio of the binding to unbinding kinetic constants). On the other hand, a multimerized TF gives extra tuning parameters, namely the multimerization ratio and the cooperativity index. Hence, a certain mode can be made more or less probable by modifying either of them without changing the location of the peaks or the dissociation ratio. In order to illustrate the above idea, we analyze a self-regulating gene with SPK with and without cooperativity. 474 475 476 477 478 479 480 481 482 483 484 485

## A Self-Regulating Gene 486

Consider a non-cooperative self-regulating gene. The unbound and bound gene states are  $D_0, D_1$  with  $k_0, k_1$  production rates, respectively. The network is activating if  $k_1 > k_0$ , and repressing otherwise. 487 488 489

Similar to the previous example, the stationary PMF is a mixture of two Poissons centered at  $k_0/k_-, k_1/k_-$  with weights  $\alpha_-/(\alpha\rho + \alpha_-), \alpha\rho/(\alpha\rho + \alpha_-)$ , respectively, where  $\rho = \mathbb{E}[X|D = 0] = k_0/k_-$ . (Refer to (12) and SI-§3.2) 490 491 492

Next, consider the cooperative counterpart with dimerization rates  $\beta, \beta_-$ . The stationary PMF stays the same except for  $\rho = \mathbb{E}[X_2|D = 0] = k_0^2\beta/(2k_-^2\beta_-)$ . Hence, in both cases, the PMF has modes at  $k_1/k_-, k_0/k_-$ , where the weight of the first mode is proportional to  $\rho$  which can be used in order to tune the weights freely while keeping 493 494 495 496



**Fig 4. More modes emerge due to SPK in cooperative self-activating gene**  
 (a) A self-activating gene. (b) The stationary PMF for different  $\varepsilon$  which shows the transition from fast promoter kinetics to SPK in a leaky cooperative self-activation of a gene with cooperativity index 2. The slow kinetic limit is calculated via (10), while the remaining curves are computed by a finite projection solution [57] of the CME. . The parameters are  $\alpha = \alpha_- = \varepsilon$ ,  $k_0 = 20$ ,  $k_1 = 100$ ,  $k_- = 10$ ,  $\beta = 10$ ,  $\beta_- = 50$ .]

	Non-Cooperative		Cooperative	
	Leaky	Non-Leaky	Leaky	Non-Leaky
Stochastic (Slow Promoter Kinetics)	2	1 (at 0)	2	1 (at 0)
Deterministic	1	1	1-2	1-2

**Table 1.** Comparing the number of stable equilibria/modes for a self-activating gene between stochastic with SPK and deterministic modelling frameworks. Details are given in SI-§6.3, where leakiness means that the unbound state has a non-zero expression rate.

the modes and the dissociation ratio unchanged. For instance, the PMF can be made effectively unimodal with a sufficiently high  $\rho$ .

**Comparison with the deterministic model:** Table 1 compares the number of stable equilibria in the deterministic model with the number of modes in the stochastic model in the case of a single gene copy. There is no apparent correlation between the numbers of deterministic equilibria and stochastic modes. Figure 4 depicts the transition from a unique mode with fast promoter kinetics to multiple modes with SPK with a cooperative leaky self-activating gene.

## The Toggle Switch

A toggle switch is a basic GRN that exhibits deterministic multi-stability. It has two stable steady states and can switch between them with an external input or via noise. The basic design is a pair of two mutually repressing genes as in Figure 5-a. The ideal behavior is that only one gene is “on” at any moment in time. The network consists of two *identical* genes whose expressed proteins X, Y act as TFs for each other (The general toggle switch is discussed in SI-§3.3). Each gene has dissociation ratio  $\alpha/\alpha_-$ , production ratio  $k_0/k_-$  for the unbound state only, multimerization ratio  $\beta/\beta_-$ , and

cooperativity index  $n$ . Using the algorithm of Proposition SI-2, the reduced-order Markov chain infinitesimal generator is:

$$\Lambda_r = \begin{bmatrix} -\alpha\rho - \alpha\rho & \alpha_- & \alpha_- & 0 \\ \alpha\rho & -\alpha_- & 0 & \alpha_- \\ \alpha\rho_2 & 0 & -\alpha_- & \alpha_- \\ 0 & 0 & 0 & -\alpha_- - \alpha_- \end{bmatrix}, \quad (14)$$

where  $\rho = \frac{k_0^n}{k_-^n} \frac{\beta}{n! \beta_-}$ . Notice immediately that the transition rates towards the configuration (1,1) are zero, which implies that the weight of the mode corresponding to (1,1) is zero. Hence, we have three modes only. The weights corresponding to the modes can be found as the principal eigenvector of  $\Lambda_r$  as given in Corollary 4. Hence, the stationary distribution for  $X, Y$  is:

$$\pi(x, y) = \frac{1}{2 \frac{\alpha}{\alpha_-} \rho + 1} \left( \mathbf{P}(y; \frac{k_0}{k_-}) \mathbf{P}(x; \frac{k_0}{k_-}) + \frac{\alpha}{\alpha_-} \rho \mathbf{P}(y; \frac{k_0}{k_-}) \delta(x) + \frac{\alpha}{\alpha_-} \rho \mathbf{P}(x; \frac{k_0}{k_-}) \delta(y) \right).$$

Hence, we get that the PMF has three modes only at  $(0, \frac{k_0}{k_-}), (\frac{k_0}{k_-}, 0), (\frac{k_0}{k_-}, \frac{k_0}{k_-})$  with relative weights  $\frac{\alpha}{\alpha_-} \rho, \frac{\alpha}{\alpha_-} \rho, 1$ , respectively. Since the stationary PMF has three modes, it deviates from the ideal behavior of a switch where at most two stable steady states, under appropriate parameter conditions, are possible. Nevertheless, a bimodal PMF can be achieved by minimizing the weight of the first mode at  $(\frac{k_0}{k_-}, \frac{k_0}{k_-})$ . If we fix  $\alpha/\alpha_-$ , then this can be satisfied by tuning  $n, \beta/\beta_-$  to maximize  $\rho$ . Choosing higher cooperativity indices, subject to  $n < k_0/k_-$ , achieves this also. For instance, a standard asymmetric design [65] uses cooperativity indices 2, 3. Figure 5 depicts the effect of cooperativity on achieving the desired behavior with the same dissociation constant and production ratios. Notice that cooperativity allows us to minimize or maximize the weight of the (high,high) mode by tuning the dimerization ratio.

The toggle switch has three modes regardless of the cooperativity index. This is unlike the deterministic model where only one positive stable state is realizable with non-cooperative binding, and two stable steady states are realizable with cooperative binding. SI-§3.3 contains further Monte-Carlo simulations that show that the predicted third mode appears with a two-to-one time scale separation between the slow gene reactions and the fast protein reactions. Experimentally, a recent study has reported that the CRI-Cro toggle switch exhibits the third (high,high) mode and the authors proposed SPK as a contributing mechanism [66], a behavior predicted by our results.

## Interconnected toggle switches

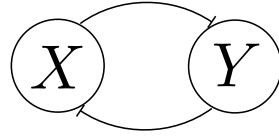
Consider  $N$  copies of the toggle switch defined in the previous section (we consider switches with identical genes for simplicity). They express proteins  $X_i, Y_i, i = 1, \dots, N$ . Let us assume that the switches are interconnected via the diffusion of the proteins among cells, modeled with a diffusion coefficient  $\Omega$  as:

$X_i \xrightleftharpoons[\Omega]{\Omega} X_j, \quad Y_i \xrightleftharpoons[\Omega]{\Omega} Y_j, \quad i \neq j, i, j = 1, \dots, N$ . We view this model as a simplification of a more complex quorum sensing communication mechanism, in which orthogonal AHL molecules are produced by cells and act as activators of TFs in receiving cells, as analyzed for example in [46]. Figure 6a depicts a block diagram of such a network.

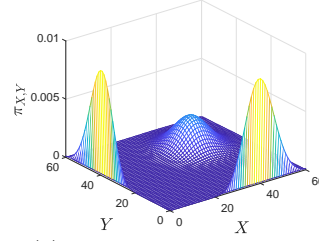
For a deterministic model, there exists a parameter range for which all toggle switches synchronize into bistability for sufficiently high diffusion coefficient [46]. This implies each switch in the network behaves as a bistable switch, and it converges with all the other switches to the same steady-states.

Our aim is to analyze the stochastic model at the limit of SPK and compare it to the deterministic model. This network is not in the form of the class of networks in

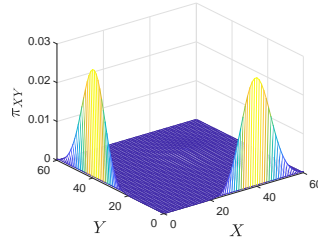




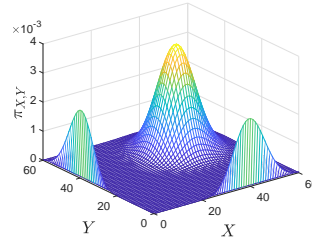
(a) Diagram of the toggle switch.



(b) Non-cooperative,  $n = 1$ .



(c) Cooperative,  $\beta/\beta_- = 1$ .



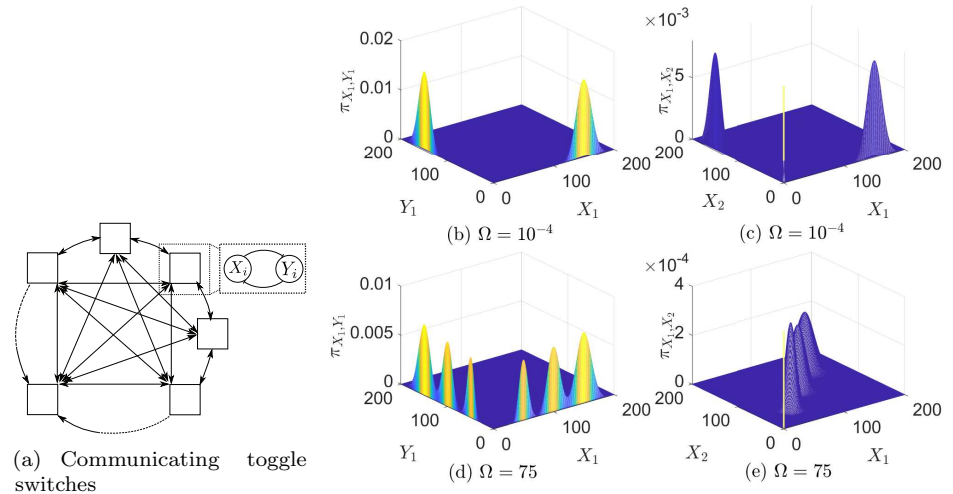
(d) Cooperative,  $\beta/\beta_- = 0.01$ .

**Fig 5. Cooperativity enables tuning of modes' weights.** Comparison of the stationary PMF between non-cooperative and cooperative binding. For all cases:  $\alpha/\alpha_- = 1/200, k_0/k_- = 40$ . (a) Diagram of the toggle switch. (b) The stationary PMF for the non-cooperative case. (c) The stationary PMF for the cooperative case with  $n = 2, \beta/\beta_- = 1$ . (d) The stationary PMF for the cooperative case  $\beta/\beta_- = 0.01$ . All surfaces are plotted using (10).

Figure 1. Nevertheless, we show in SI-§4 that our results can be generalized to networks with weakly reversible deficiency zero conditional MCs. There are  $4^N$  conditional MCs, and using Theorem 3, the stationary PMF is a mixture of  $4^N - 1$  Poissons.

Consider now the case of a high diffusion coefficient. We show (see SI-§3.5) that as  $\Omega \rightarrow \infty$ ,  $X_1, \dots, X_N$  synchronize in the sense that the joint PMF of  $X_1, \dots, X_N$  is symmetric with respect to all permutations of the random variables. This implies that the marginal stationary PMFs  $p_{X_i}, i = 1, \dots, N$  are identical. Hence, for sufficiently large  $\Omega$ , the probability mass is concentrated around the region for which  $X_1, \dots, X_N$  are close to each other. Consequently, for large  $\Omega$  we can replace the population of toggle switches with a *single toggle switch* with the *synchronized protein processes*  $X(t), Y(t)$ , which are defined, for the sake of convenience, as  $X(t) := X_1(t), Y(t) := Y_1(t)$ . Next, we describe the stationary PMF of  $X(t), Y(t)$ .

The state of synchronized toggle switches does not depend on individual promoter configurations, and it depends only on the total number of unbound promoter sites in the network. Hence, the number of modes will drop from  $4^N - 1$  to  $(N + 1)^2 - 1$ . Note that similar to the single toggle switch, there are modes which have both  $X, Y$  with non-zero copy number. On the other hand, there are many additional modes. Recall that in the case of a single toggle switch, we have tuned the cooperativity ratios such that the modes in which both genes are ON are suppressed. Similarly, the undesired modes can be suppressed by tuning the cooperativity ratio which can be achieved by choosing  $\rho_{d_i}^X, \rho_{d_i}^Y, d = 0, \dots, 4^N - 1$  sufficiently large. In particular, letting the multi-merization ratio  $\beta/\beta_- \rightarrow \infty$ , the weights of modes in the interior of the positive orthant  $\mathbb{R}_+^2$  approach zero. In conclusion, for sufficiently high  $\Omega$  and sufficiently high  $\beta/\beta_-$  the population behaves as a *multimodal switch*, which means that the whole network can have either the gene X ON, or the gene Y ON. And every gene can take



**Fig 6. SPK lead to the emergence of a multi- modal toggle switch** (a) A diagram of population of toggle switches. Arrows between blocks represent reversible diffusion reactions. Each block contains a toggle switch. The remaining subfigures show stationary PMFs for a population of three identical cooperative toggle switches. Due to the symmetries we plot joint PMFs of  $X_1, Y_1$  and  $X_1, X_2$  only. Subplots (b),(c) depict the uncoupled toggle switches. Note that  $X_1$  and  $X_2$  are not synchronized. Subplots (d), (e) depict a high diffusion case. The toggle switches synchronize into a multi-modal toggle switch. More details are given SI-§6.4

$2N$  modes which are:  $\{(ik_0/(Nk_-), 0), (0, ik_0/(Nk_-)) : i = 1, \dots, N\}$ . Comparing to the low diffusion case, the network will have up to  $2^N - 1$  modes with sufficiently high multimerization ratio.

In order to illustrate the previous results, consider a population of three toggle switches ( $N = 3$ ) and cooperativity  $n = 2$ . For large  $\Omega$ , the deterministic system bifurcates into bistability. This means that all toggle switches converge to the same exact equilibria if  $\Omega$  is greater than a threshold. In contrast, the modes in the stochastic model of the toggle switches converge *asymptotically* to each other. Hence, we need to choose a threshold for  $\Omega$  that constitutes “sufficient” synchronization. We define this as the protein processes synchronizing within one copy number. In other words, we require the maximum distance between the modes to be less than 1. It can be shown (see SI-§3.5) that  $\Omega$  has to satisfy:  $\Omega \geq \frac{1}{N}(k - k_-)$ . In this example, the minimal  $\Omega$  is 75. The stationary PMF is depicted in Figure 6d. The network has 15 modes, nine of which are in the interior and are suppressed due to cooperativity. In contrast, the deterministic model bifurcates into synchronization for  $\Omega > 0.5$ . The stable synchronized equilibria are  $(149.98, 0.02), (0.02, 149.98)$ .

The stochastic model with SPK adds four additional modes at  $(0, 100), (100, 0), (50, 0), (0, 50)$ . To interpret this, note that the protein processes synchronize while the promoter configurations do not. The high states  $(150, 0), (0, 150)$  correspond to the case when all the binding sites are empty. In the case when one binding site is empty, the first gene is active while the second and the third are not. Due to diffusion, the first gene “shares” its expressed protein with the other two genes, which implies that each gene will receive a *third* of the total protein copy numbers produced in the network. A similar situation arises when two binding sites are empty.

## The Repressilator

A very different example is provided by a well-studied synthetic oscillator, the repressilator. The repressilator is a synthetic biological circuit that implements a ring oscillator [67], and it has been simulated with slow-promoter kinetics [21]. It is a canonical example of a GRN that exhibits a limit cycle, i.e. sustained oscillation. For simplicity, we study a network consisting of three *identical* genes whose expressed proteins are X, Y, Z. The protein X represses Y, Y represses Z, and Z represses X as shown in Figure 7. Each gene has dissociation ratio  $\alpha/\alpha_-$ , production ratio  $k/k_-$  for the unbound state only, multimerization ratio  $\beta/\beta_-$ , and cooperativity index  $n$ .

Deterministic analysis of the repressilator [3] reveals that it does not oscillate with non-cooperative binding. Applying our techniques for the stochastic case with slow promoter kinetics, we are able to find the values of the parameters so that the probability is concentrated in three modes  $(K, 0, 0)$ ,  $(0, K, 0)$ ,  $(0, 0, K)$  if  $w := 2(\alpha/\alpha_-)(k/k_-)^n(\beta/(n!\beta_-)) \gg 1$ , where  $K = k/k_-$  (see SI-§3.4). The obtained tri-modal stationary distribution is consistent with the classical oscillations of the repressilator, and this is independent of the cooperativity index. Note that this condition is analogous to the oscillation condition in the deterministic model [3] (but with cooperativity only) which also requires “large” production ratios.

In order to study whether the network oscillates, we need to define a notion of limit cycle for a stochastic system. Due to randomness, the time-series can not be periodic. Nevertheless, since the stationary distribution is tri-modal, we say that the network oscillates if sample paths (time histories of trajectories) typically jump between the modes in the same order.

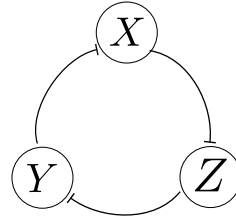
Assume  $w \gg 1$ . Let  $d_x, d_y, d_z$  be the three dominant modes. We show that if the reduced-order Markov chain is at mode  $d_x$  then it is much more likely to transition to  $d_y$  rather than to  $d_z$ . Similar arguments apply if we start from  $d_y, d_z$ . In particular, let  $Q(t) = e^{t\Lambda_r}$  be the probability transition matrix. We are interested in comparing the probabilities of transiting from  $r_x$  to  $r_y, r_z$ . Hence we study small  $t \ll 1$ . We give expressions for  $Q_{d_x d_x}(t), Q_{d_y d_x}(t), Q_{d_z d_x}(t)$  in SI-§3.4 which show that if the Markov chain is at  $d_x$  then it is most likely to stay there. The transition is much more likely to happen to  $d_y$  rather than  $d_z$ . Hence, we expect to see “long” periods of protein X being expressed, and then it jumps to express protein Y, and then protein Z. Since the finite Markov chain is ergodic, the pattern repeats.

Note that the analysis above predicts that both the cooperative and the noncooperative repressilator are capable of oscillation with slow-promoter kinetics when  $w \gg 1$ . The average “period” increases with the production ratio and the transient behavior of the network follows the analysis in SI-§5.

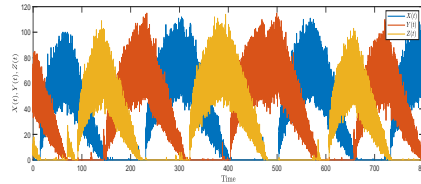
We performed Monte-Carlo simulations via the Gillespie algorithm for both fast and slow kinetics. The results are shown in Figure 8. We observe that the network always oscillates with cooperative binding. With non-cooperative binding, only the network with slow kinetics oscillates, as predicted. The network with fast kinetics does not oscillate. Recall that the deterministic model with non-cooperative binding does not oscillate [3].

## Trans-Differentiation Network

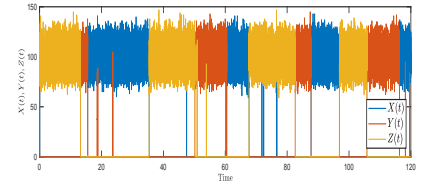
We consider two networks for TF cross-antagonism in cell fate decision in this section. Both networks consist of two self-activating genes repressing each other as depicted in Figure 9-a [5]. The first network [39] has independent cooperative binding of the TFs to the promoters. The genes states are  $D_{00}^X, D_{01}^X, D_{10}^X, D_{11}^X$  for gene X, and vice versa for gene Y. In order for the genes to be cross-inhibiting and self-activating we let:  $D_{01}^X, D_{10}^Y$  have zero production rates. Also, the maximal production rates for genes X, Y occur at



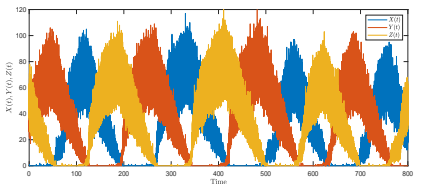
**Fig 7.** A diagram of the repressilator.



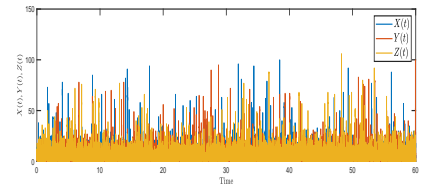
(a) Cooperative, Slow-Promoter Kinetics.



(b) Noncooperative, Slow-Promoter Kinetics.



(c) Cooperative, Fast-Promoter Kinetics.



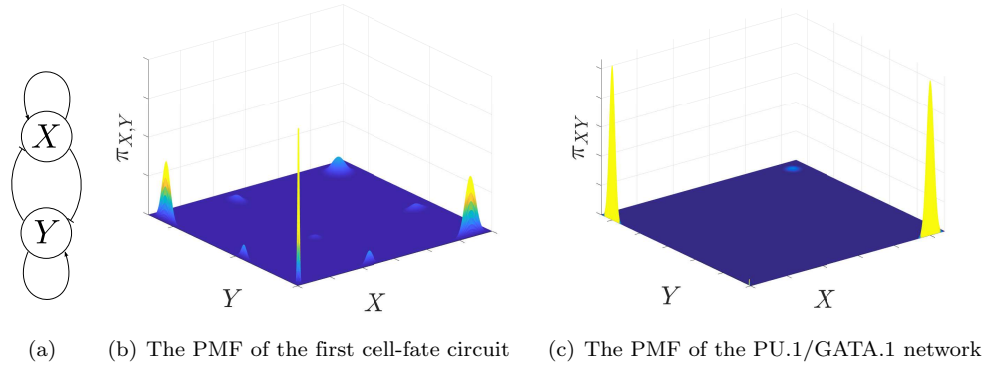
(d) Noncooperative, Fast-Promoter Kinetics.

**Fig 8. The noncooperative repressilator oscillates.** (a) A time-series for the cooperative repressilator with cooperativity index 2, and slow promoter kinetics. (b) A time-series for the cooperative repressilator with cooperativity index 2, and fast promoter kinetics. (c) A time-series for the noncooperative repressilator, and slow promoter kinetics. (d) A time-series for the noncooperative repressilator, and fast promoter kinetics. The plots were generated by stochastic simulation via the Gillespie algorithm. For all the figures, the parameters are:

$\alpha = 5\varepsilon, \alpha_- = 1\varepsilon, k = 2000, k_- = 20, \beta_{\pm} = 1$ , where  $\varepsilon = 0.1$  for slow kinetics, and  $\varepsilon = 1000$  for fast kinetics.

gene states  $D_{10}^X, D_{01}^Y$ , respectively. (See SI-§6.5) The network can be analyzed with the proposed framework, as it consists of two genes each with two binding sites. Hence it can theoretically admit up to 16 modes according to (10). The PMF is depicted in Figure 9-b for an example parameter set. Note that despite the fact that we have 16 modes, only eight of them contribute to most of the stationary PMF. This is to be contrasted with a deterministic model, which cannot produce more than 4 stable equilibria [52].

The second network that we study is a model of the PU.1/GATA.1 network, which is a lineage determinant in hematopoietic stem cells [68]. Diagrammatically, it can also be presented by Figure 9-a. However, it differs from the first network presented above in several ways. First, PU.1 needs GATA.1 to bind to the promoter of GATA.1 [69], and vice versa [70]. In our modelling framework this means that the promoter configurations  $D_{01}^X, D_{10}^Y$  do not exist, where X stands for PU.1 and Y stands for GATA.1. Hence, the network has 9 gene states. Second, there is no evidence that PU.1 and GATA.1 form dimers to activate their own promoters cooperatively. In fact, it has been shown that self-activation for GATA-1 occurs primarily through monomeric binding [71]. Further discussion of the model is included in SI-§6.5, and is further discussed in [72].



**Fig 9. The cell-fate decision network with SPK has more modes than what a deterministic model predicts.** (a) A diagram of a generic cell-fate circuit that can describe the networks considered, (b) The PMF of the first cell-fate circuit computed using Theorem 3. (c) The PMF of the PU.1/GATA.1 circuit, where X denotes PU.1 and Y denotes GATA.1. Three modes can be seen. Details are given in SI-§6.5.

With lack of cooperativity, the deterministic model is only monostable and cannot explain the emergence of bistability for the above network. However, using our framework, up to nine modes can be realized. In order to simplify the landscape, we group the nine into four modes. This is possible since the states  $D_{11}^X, D_{00}^X, D_{11}^Y, D_{00}^Y$  have very low production rates. This gives a total of *four* modes which are (low,low),(high,low),(low,high),(high,high). Using our model, we choose the parameters to realize bistability and tristability. Figure 9-c depicts the stationary PMF for a set of parameters that satisfies the assumptions and give rise to a tristable PMF.

## Methods

**Numerical Simulation Software** Calculations were performed using MATLAB 9. Bertini 1.5 was used for the computation of deterministic solutions of the “quorum sensing” numerical example.

## Discussion

Phenotypical variability in the absence of genetic variation is a phenomenon of great interest in current biological and translational research, as it plays an important role in processes as diverse as embryonic development [73], hematopoietic cell differentiation [74], and cancer heterogeneity [75]. A conceptual, and often proposed, unifying framework to explain non-genetic variability is to think of distinct phenotypes as multiple “metastable states” or “modes” in the complex energetic landscape associated to an underlying GRN. Following this point of view, we studied a general but simplified mathematical model of gene regulation. Our focus has been on stochastic SPK, the time scale relevant when transcription factor binding and unbinding are affected by epigenetic processes such as DNA methylation and chromatin remodeling. In that regime, adiabatic approximations of promoter kinetics are not appropriate. In contrast to the existing literature, which largely confines itself to numerical simulations, in this work we provided a rigorous analytic characterization of multiple modes.

The general formal approach that we developed provides insight into the relative influence of model parameters on system behavior. It also allows making theoretical

693 predictions of how changes in wiring of a GRN, be it through natural mutations or  
694 through artificial interventions, impact the possible number, location, and likelihood, of  
695 alternative states. We were able to tease out the role of cooperative binding in  
696 stochastic models in comparison to deterministic models, which is a question of great  
697 interest in both the analysis of natural systems and in synthetic biology engineering.  
698 Specifically, we found that, unlike deterministic systems, the number of modes is  
699 independent of whether the TF-promoter binding is cooperative or not; on the other  
700 hand, cooperative binding gives extra degrees of freedom for assigning weights to the  
701 different modes. Emergence of bimodality in noncooperative single gene networks in  
702 different contexts has been reported in [76], which studies exogenous TF and fast  
703 promoter kinetics, and [77], which studies the effect of temperature fluctuations. The  
704 intermediate promoter kinetic domain has been studied in [78]. Switching behavior in a  
705 single gene driven by a bursty exogenous input has been studied in [79]. More generally,  
706 we characterized the stationary PMFs of CMEs for our GRNs as mixtures of Poisson  
707 PMFs, which enabled us to obtain explicit formulas for the locations and probabilities  
708 of metastable states as a function of the parameters describing the system.

709 Although we formulate our study in terms of stationary PMFs, one may equally well  
710 view our results as describing the typical dynamic behavior of realizations of the  
711 stochastic process. These recapitulate the form of the stationary PMFs: modes are  
712 reflected in metastable states along sample paths, states in which the system will stay  
713 for prolonged periods until switching to other states corresponding to alternative modes.  
714 In the SI, we provide Monte-Carlo simulations showing such metastable behavior along  
715 sample paths. We do so for the toggle switch as well as for a version of a well-studied  
716 genetic circuit [67] which exhibits oscillatory behavior along sample paths even though  
717 the corresponding deterministic model cannot admit oscillations. One application of our  
718 mathematical results was to models of single or communicating “toggle switches” in  
719 bacteria, where we showed that, for suitable parameters, there are a very large number  
720 of metastable attractors.

721 This work was in fact motivated by our interest in hematopoietic cell differentiation,  
722 and in this paper we discussed two possible models of trans-differentiation networks in  
723 mammalian cells. In a first model, based on previous publications, we uncovered more  
724 modes than had been predicted with different analyses of the same model. This implies  
725 that in practice there could be unknown “intermediate” phenotypes that result from the  
726 network’s dynamics, which may be acquired by cells during the natural differentiation  
727 process or which one might be able to induce through artificial stimulation. The second  
728 model included only binding reactions that have been experimentally documented, and  
729 as such might be more biologically realistic than the first model. For this second model,  
730 a deterministic analysis predicts monostability, which is inconsistent with the fact that  
731 the network should control a switch between two stable phenotypes (erythroid and  
732 myeloid). This suggests that stochasticity, likely due to low copy numbers and/or SPK,  
733 might be responsible for the multiple attractors (phenotypes) that are possible in cell  
734 differentiation GRNs. Our mathematical results, being quite generic, should also be  
735 useful in the analysis of networks that have been proposed for understanding aspects of  
736 cancer biology. For example, non-genetic heterogeneity has been recently recognized as  
737 an important factor in cancer development and resistance to therapy, with stochastic  
738 multistability in gene expression dynamics acting as a generator of phenotype  
739 heterogeneity, setting a balance between mesenchymal, epithelial, and cancer  
740 stem-cell-like states [80] [81] [82] [83], and nongenetic variability due to multistability  
741 arising from mutually repressing gene networks has been proposed to explain metastatic  
742 progression [84].

743 Application of the results to practical problems entails deciding whether  $\varepsilon$  is small  
744 enough. Since singular perturbations rely on a first-order approximation of the

stationary PMF (8), the exact determination of the range of  $\varepsilon$  requires determining coefficients of higher-order terms, which can be estimated by computing the asymptotic expansion for a finite state projection of the specific problem at hand. Nevertheless, we provide a simple intuitive rule. The predicted behavior is expected to emerge when the largest rate in the reduced matrix  $\varepsilon\Lambda_r$  is slower than the decay rate of proteins. Recall that the elements of  $\Lambda_r$  depend on the association and dissociation constants and the conditional expectation of protein copy numbers given in (12) as seen in (11). Our numerical examples depicted in Figures 1, 3, 4, SII examine how the approximation fares with multiple levels of time scale separation and agree with the rule. In particular, the latter figure provides Monte-Carlo simulations depicting the third mode of the toggle switch with a two-to-one scale separation per the definition above. Furthermore, since (12) can be tuned by the multimerization ratio, we note that the ratio can make the network more “robust” or vulnerable with respect to the emergence of modes predicted in the SPK regime. In practice, it may be difficult to estimate experimentally the average time that a TF of interest takes to find its binding targets. Hence, we suggest that our results should be considered if there is a very low number of gene copies (i.e., 1-5) and it is suspected that TF-gene binding kinetics are slower than protein kinetics, which may happen particularly in Eukaryotic cells as discussed in the introduction. Our approach can be seen as an addition to the toolbox for analysis of the spectrum of possible behaviors in GRNs, and it can explain apparent multi-modality when the deterministic model can’t. As an example, the recent experimental work on the toggle switch [66] which validated the observation of a third mode and proposed slow promoter kinetics as a mechanism, is consistent with our results.

## Acknowledgements

This work has been carried out while first author was with the Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA. We thank Nithin S. Kumar for discussions regarding the PU.1/GATA.1 network and Cameron McBride for proofreading the manuscript.

## References

1. Alon U. An Introduction to Systems Biology. CRC press; 2006.
2. Karlebach G, Shamir R. Modelling and analysis of gene regulatory networks. Nature Reviews Molecular Cell Biology. 2008;9(10):770–780.
3. Del Vecchio D, Murray RM. Biomolecular Feedback Systems. Princeton Univ Press; 2015.
4. Munsky B, Neuert G, van Oudenaarden A. Using gene expression noise to understand gene regulation. Science. 2012;336(6078):183–187.
5. Zhou JX, Huang S. Understanding gene circuits at cell-fate branch points for rational cell reprogramming. Trends in Genetics. 2011;27(2):55–62.
6. Arkin A, Ross J, McAdams HH. Stochastic kinetic analysis of developmental pathway bifurcation in phage  $\lambda$ -infected Escherichia coli cells. Genetics. 1998;149(4):1633–1648.
7. Stamatakis M, Mantzaris NV. Comparison of deterministic and stochastic models of the lac operon genetic network. Biophysical journal. 2009;96(3):887–906.

8. Kepler TB, Elston TC. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophysical journal*. 2001;81(6):3116–3136. 788  
789  
790
9. Kærn M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: from theories to phenotypes. *Nature Reviews Genetics*. 2005;6(6):451–464. 791  
792
10. Hahl SK, Kremling A. A Comparison of Deterministic and Stochastic Modeling Approaches for Biochemical Reaction Systems: On Fixed Points, Means, and Modes. *Frontiers in Genetics*. 2016;7:157. 793  
794  
795
11. Norris JR. *Markov Chains*. Cambridge University Press; 1998. 796
12. Gardiner CW, et al. *Handbook of stochastic methods*. Springer Berlin; 1985. 797
13. Eldar A, Elowitz MB. Functional roles for noise in genetic circuits. *Nature*. 2010;467(7312):167–173. 798  
799
14. Symmons O, Raj A. What's Luck Got to Do with It: Single Cells, Multiple Fates, and Biological Nondeterminism. *Molecular cell*. 2016;62(5):788–802. 800  
801
15. Bressloff PC. Stochastic switching in biology: from genotype to phenotype. *Journal of Physics A: Mathematical and Theoretical*. 2017;50(13):133001. 802  
803
16. Pirone JR, Elston TC. Fluctuations in transcription factor binding can explain the graded and binary responses observed in inducible gene expression. *Journal of Theoretical Biology*. 2004;226(1):111–121. 804  
805  
806
17. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol*. 2006;4(10):e309. 807  
808
18. Karmakar R, Bose I. Graded and binary responses in stochastic gene expression. *Physical biology*. 2004;1(4):197. 809  
810
19. Shahrezaei V, Swain PS. Analytical distributions for stochastic gene expression. *Proceedings of the National Academy of Sciences*. 2008;105(45):17256–17261. 811  
812
20. Tabaka M, Kalwarczyk T, Hołyst R. Quantitative influence of macromolecular crowding on gene regulation kinetics. *Nucleic acids research*. 2013;42(2):727–738. 813  
814
21. Potapov I, Zhurov B, Volkov E. Multi-stable dynamics of the non-adiabatic repressilator. *Journal of The Royal Society Interface*. 2015;12(104):20141315. 815  
816
22. Choi PJ, Cai L, Frieda K, Xie XS. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science*. 2008;322(5900):442–446. 817  
818
23. Briegel K, Bartunek P, Stengl G, Lim KC, Beug H, Engel JD, et al. Regulation and function of transcription factor GATA-1 during red blood cell differentiation. *Development*. 1996;122(12):3839–3850. 819  
820  
821
24. Paldi A. Stochastic gene expression during cell differentiation: order from disorder? *Cellular and molecular life sciences*. 2003;60(9):1775–1778. 822  
823
25. Miller-Jensen K, Dey SS, Schaffer DV, Arkin AP. Varying virulence: epigenetic control of expression noise and disease processes. *Trends in biotechnology*. 2011;29(10):517–525. 824  
825  
826
26. Voss TC, Hager GL. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nature Reviews Genetics*. 2014;15(2):69–81. 827  
828



27. Yuan L, Chan GC, Beeler D, Janes L, Spokes KC, Dharaneeswaran H, et al. A role of stochastic phenotype switching in generating mosaic endothelial cell heterogeneity. *Nature communications*. 2016;7:10160. 829  
830  
831
28. To TL, Maheshri N. Noise can induce bimodality in positive transcriptional feedback loops without bistability. *Science*. 2010;327(5969):1142–1145. 832  
833
29. Mariani L, Schulz EG, Lexberg MH, Helmstetter C, Radbruch A, Löhning M, et al. Short-term memory in gene induction reveals the regulatory principle behind stochastic IL-4 expression. *Molecular systems biology*. 2010;6(1):359. 834  
835  
836
30. Ebeling W, Schimansky-Geier L. Stochastic dynamics of a bistable reaction system. *Physica A: Statistical Mechanics and its Applications*. 1979;98(3):587–600. 837  
838  
839
31. Hornos J, Schultz D, Innocentini G, Wang J, Walczak A, Onuchic J, et al. Self-regulating gene: an exact solution. *Physical Review E*. 2005;72(5):051907. 840  
841
32. Qian H, Shi PZ, Xing J. Stochastic bifurcation, slow fluctuations, and bistability as an origin of biochemical complexity. *Physical Chemistry Chemical Physics*. 2009;11(24):4861–4870. 842  
843  
844
33. Iyer-Biswas S, Hayot F, Jayaprakash C. Stochasticity of gene products from transcriptional pulsing. *Physical Review E*. 2009;79(3):031911. 845  
846
34. Grima R, Schmidt DR, Newman TJ. Steady-state fluctuations of a genetic feedback loop: An exact solution. *The Journal of chemical physics*. 2012;137(3):035104. 847  
848  
849
35. Thomas P, Popović N, Grima R. Phenotypic switching in gene regulatory networks. *Proceedings of the National Academy of Sciences*. 2014;111(19):6994–6999. 850  
851  
852
36. Cao Z, Grima R. Linear mapping approximation of gene regulatory networks with stochastic dynamics. *Nature communications*. 2018;9(1):3305. 853  
854
37. Potoyan DA, Wolynes PG. Dichotomous noise models of gene switches. *The Journal of chemical physics*. 2015;143(19):195101. 855  
856
38. Liu P, Yuan Z, Huang L, Zhou T. Roles of factorial noise in inducing bimodal gene expression. *Physical Review E*. 2015;91(6):062706. 857  
858
39. Feng H, Wang J. A new mechanism of stem cell differentiation through slow binding/unbinding of regulators to genes. *Scientific reports*. 2012;2(2):550. 859  
860
40. Chen C, Wang J. A physical mechanism of cancer heterogeneity. *Scientific reports*. 2016;6:20679. 861  
862
41. Campbell SL, Rose NJ. Singular perturbation of autonomous linear systems. *SIAM Journal on Mathematical Analysis*. 1979;10(3):542–551. 863  
864
42. Lipshtat A, Loinger A, Balaban NQ, Biham O. Genetic toggle switch without cooperative binding. *Physical review letters*. 2006;96(18):188101. 865  
866
43. Miller MB, Bassler BL. Quorum sensing in bacteria. *Annual Reviews in Microbiology*. 2001;55(1):165–199. 867  
868

44. Garcia-Ojalvo J, Elowitz MB, Strogatz SH. Modeling a synthetic multicellular clock: repressilators coupled by quorum sensing. *Proceedings of the National Academy of Sciences of the United States of America*. 2004;101(30):10955–10960. 869–871
45. Kobayashi H, Kærn M, Araki M, Chung K, Gardner TS, Cantor CR, et al. Programmable cells: interfacing natural and engineered gene networks. *Proceedings of the National Academy of Sciences of the United States of America*. 2004;101(22):8414–8419. 872–875
46. Nikolaev EV, Sontag ED. Quorum-sensing synchronization of synthetic toggle switches: A design based on monotone dynamical systems theory. *PLoS Comput Biol*. 2016;12(4):e1004881. 876–878
47. Waddington CH. *The strategy of the genes*. George Allen and Unwin; 1957. 879
48. Balázsi G, van Oudenaarden A, Collins JJ. Cellular decision making and biological noise: from microbes to mammals. *Cell*. 2011;144(6):910–925. 880–881
49. Norman TM, Lord ND, Paulsson J, Losick R. Stochastic switching of cell fate in microbes. *Annual review of microbiology*. 2015;69:381–403. 882–883
50. Chu BK, Margaret JT, Sato RR, Read EL. Markov State Models of gene regulatory networks. *BMC systems biology*. 2017;11(1):14. 884–885
51. Graf T, Enver T. Forcing cells to change lineages. *Nature*. 2009;462(7273):587–594. 886–887
52. Wu F, Su RQ, Lai YC, Wang X. Engineering of a synthetic quadrastable gene network to approach Waddington landscape and cell fate determination. *eLife*. 2017;6:e23702. 888–890
53. Érdi P, Tóth J. *Mathematical models of chemical reactions: theory and applications of deterministic and stochastic models*. Manchester University Press; 1989. 891–893
54. Ferrell JE. Q&A: Cooperativity. *Journal of Biology*. 2009;8(6):53. 894
55. Rao CV, Arkin AP. Stochastic chemical kinetics and the quasi-steady-state assumption: application to the Gillespie algorithm. *The Journal of chemical physics*. 2003;118(11):4999–5010. 895–897
56. Kim JK, Sontag ED. Reduction of multiscale stochastic biochemical reaction networks using exact moment derivation. *PLOS Computational Biology*. 2017;13(6):e1005571. 898–900
57. Peleš S, Munsky B, Khammash M. Reduction and solution of the chemical master equation using time scale separation and finite state projection. *The Journal of Chemical Physics*. 2006;125(20):204104. 901–903
58. Phillips R, Kokotović P. A singular perturbation approach to modeling and control of Markov chains. *IEEE Transactions on Automatic Control*. 1981;26(5):1087–1094. 904–906
59. Khasminskii R, Yin G, Zhang Q. Constructing asymptotic series for probability distributions of Markov chains with weak and strong interactions. *Quarterly of Applied Mathematics*. 1997;55(1):177–200. 907–908

60. Meyn SP, Tweedie RL. Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes. *Advances in Applied Probability*. 1993; p. 518–548. 910–912
61. Altman E, Avrachenkov KE, Núñez-Queija R. Perturbation analysis for denumerable Markov chains with application to queueing models. *Advances in Applied Probability*. 2004; p. 839–853. 913–915
62. Gupta A, Briat C, Khammash M. A scalable computational framework for establishing long-term behavior of stochastic reaction networks. *PLoS computational biology*. 2014;10(6):e1003669. 916–918
63. Duncan A, Liao S, Vejchodský T, Erban R, Grima R. Noise-induced multistability in chemical systems: Discrete versus continuum modeling. *Physical Review E*. 2015;91(4):042111. 919–921
64. Wang S. Quantifying gene regulatory networks. Duke University; 2014. 922
65. Gardner TS, Cantor CR, Collins JJ. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*. 2000;403(6767):339–342. 923–924
66. Fang X, Liu Q, Bohrer C, Hensel Z, Han W, Wang J, et al. Cell fate potentials and switching kinetics uncovered in a classic bistable genetic switch. *Nature communications*. 2018;9(1):2787. 925–927
67. Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. *Nature*. 2000;403(6767):335. 928–929
68. Doré LC, Crispino JD. Transcription factor networks in erythroid cell and megakaryocyte development. *Blood*. 2011;118(2):231–239. 930–931
69. Zhang P, Zhang X, Iwama A, Yu C, Smith KA, Mueller BU, et al. PU. 1 inhibits GATA-1 function and erythroid differentiation by blocking GATA-1 DNA binding. *Blood*. 2000;96(8):2641–2648. 932–934
70. Burda P, Vargova J, Curik N, Salek C, Papadopoulos GL, Strouboulis J, et al. GATA-1 inhibits PU. 1 Gene via DNA and histone H3K9 methylation of its distal enhancer in erythroleukemia. *PloS one*. 2016;11(3):e0152234. 935–937
71. Crossley M, Merika M, Orkin S. Self-association of the erythroid transcription factor GATA-1 mediated by its zinc finger domains. *Molecular and Cellular Biology*. 1995;15(5):2448–2456. 938–940
72. Al-Radhawi MA, Kumar NS, Sontag ED, Del Vecchio D. Stochastic multistationarity in a model of the hematopoietic stem cell differentiation network. In: *Proceedings of 2018 IEEE conference on Decision and Control (CDC)*. IEEE; 2018. p. 1887–1892. 941–944
73. Calvanese V, Fraga MF. Epigenetics of embryonic stem cells. *Adv Exp Med Biol*. 2012;741:231–253. 945–946
74. Sharma S, Gurudutta G. Epigenetic Regulation of Hematopoietic Stem Cells. *Int J Stem Cells*. 2016;9(1):36–43. 947–948
75. Easwaran H, Tsai HC, Baylin SB. Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance. *Mol Cell*. 2014;54(5):716–727. 949–950

76. Ochab-Marcinek A, Tabaka M. Bimodal gene expression in noncooperative regulatory systems. *Proceedings of the National Academy of Sciences*. 2010;107(51):22096–22101. 951  
952  
953
77. Charlebois DA, Hauser K, Marshall S, Balázsi G. Multiscale effects of heating and cooling on genes and gene networks. *Proceedings of the National Academy of Sciences*. 2018;115(45):E10797–E10806. 954  
955  
956
78. Ge H, Qian H, Xie XS. Stochastic phenotype transition of a single cell in an intermediate region of gene state switching. *Physical review letters*. 2015;114(7):078101. 957  
958  
959
79. Kumar N, Platini T, Kulkarni RV. Exact distributions for stochastic gene expression models with bursting and feedback. *Physical review letters*. 2014;113(26):268105. 960  
961  
962
80. Dean M, Fojo T, Bates S. Tumour stem cells and drug resistance. *Nat Rev Cancer*. 2005;5(4):275–284. 963  
964
81. Huang S, Eichler G, Bar-Yam Y, Ingber DE, Ingber DE. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys Rev Lett*. 2005;94(12):128701. 965  
966  
967
82. Gupta PB, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C, et al. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell*. 2011;146(4):633–644. 968  
969  
970
83. Huang S. On the intrinsic inevitability of cancer: from foetal to fatal attraction. *Semin Cancer Biol*. 2011;21(3):183–199. 971  
972
84. Lee J, Lee J, Farquhar KS, Yun J, Frankenberger CA, Bevilacqua E, et al. Network of mutually repressive metastasis regulators can promote cell heterogeneity and metastatic transitions. *Proc Natl Acad Sci USA*. 2014;111(3):E364–373. 973  
974  
975  
976

## Supporting Information Legend 977

- S1 Text** Supporting information file with mathematical proofs, detailed analysis of examples, generalization of the results and additional simulations. (PDF) 978  
979