52nd IEEE Conference on Decision and Control
December 10-13, 2013. Florence, Italy

Modeling Environmental Disturbances with the Chemical Master Equation

Enoch Yeung, James L. Beck, and Richard M. Murray

I. Introduction

Abstract— **In this paper we consider the problem of representing a biological system and its environment using a stochastic modeling framework. We first introduce a decomposition of the global chemical reaction system into two systems: a system of interest and its environment. We then present and derive a decomposition of the chemical master equation to achieve a representation describing the dynamics of the system of interest, perturbed by an environmental disturbance. We use this decomposition to model examples of two types of environmental disturbances: the disturbance a system experiences through loading effects from limited resources and the disturbance a system experiences when perturbed by an antibiotic that modifies transcription or translation rates.**

Cell to cell variability in gene expression [1] is a property of small volume, small copy number biochemical systems. From a controls standpoint, this variability imposes fundamental constraints on feedback performance [2] and create a challenge in designing circuits that must function around a specific operating point. Classic studies of synthetic oscillators [3] reveal that variable gene expression leads to variable oscillator phases, desynchronization, variable amplitude etc. However, recent strategies using combinatorial promoter architectures provide hope that the design of a robust oscillator [4] is possible. However, when the same oscillator was exposed to loading effects in [5], oscillation disappeared entirely in some cells while other cells produced slow irregular oscillations. Despite using a stochastic model to account for cell-to-cell variability, the stochastic model used in [4] could not account for environmental disturbances.

Perhaps the most widely accepted stochastic model for biochemical systems is the chemical master equation, a special instance of the forward Chapman Kolmogorov equation [6]. In [7], the author shows that the chemical master equation is an exact model for a well-mixed, thermally equilibrated gas-phase system. Typically, when used to model biochemical systems in liquid phase, it is deemed a "mesoscopic description" of dynamics, as it is considered an intermediate representation between the microscopic representation of molecular dynamics and the macroscopic representation of a mass action kinetics model. Thus, in theory, the chemical master equation contains the necessary information to capture the randomness of molecules colliding and moving in a well-mixed volume as well as an appropriate level of abstraction to escape the analytical burden of simulating physical trajectories and collisions of individual molecules in the system.

However, finding an analytical solution for the chemical master equation is generally difficult, if not impossible, as it is typically infinite dimensional in the state-space [6]. Except in special instances where models are amenable to generating function approaches for exact solutions [8], [9] or where conservation laws enable finite bounds on the state-space [10], exact solutions for the master equation are difficult to obtain in closed form analytical expressions. Two alternatives exist to address this problem: 1) simulation using techniques such as τ-leaping, hybrid approaches, time scale separation approaches, or 2) reducing the model to a simpler or tractable form, e.g. using the finite state space projection algorithm [10], the sliding window abstraction approach [11], as well as spectral methods using basis functions to expand and approximate probability densities [12], [13], [14].

One of the outstanding challenges in modeling stochastic biochemical systems is the problem of accounting for system complexity—in a single cell there are millions of biomolecules present at any point in the cell cycle and many are often neglected in models but are critical to system function, e.g. ribosomes, RNAP, tRNA, σ-factors. Additionally, there is strong evidence to suggest that host, compositional, spatial and functional context, often ignored in synthetic and systems biology models, play a role in regulating gene expression [15], [16]. Global variables such as these are typically unaccounted for in stochastic models, yet they impact the dynamics of the biological systems studied.

In control theory, it is standard to include modeling terms that account for environmental disturbances [17] — often the disturbances are considered bounded and controllers are subsequently designed to be robust to disturbances contained within those bounds. Such a perspective may be valuable when complemented with recent advances in experimental techniques employing optogenetic *ex vivo* control based on *in silico* models to regulate *in vivo* gene expression in cells [18], [19]. Once a notion of environmental disturbance is formulated, we can begin to probe the robustness of a particular *ex vivo* controller with respect to environmental perturbations in a stochastic modeling framework.

Toward this end, in this work we develop an approach for capturing environmental disturbances using the chemical master equation. We view our efforts as supplementary to the model reduction results of [12], [13], [14], as their techniques can be applied in concert with our own approaches or in a stepwise approach. Our results are complementary to the results in [9], [20], where total output noise is decomposed into system-extrinsic noise and system-intrinsic noise. Here we consider decomposition at the level of system *dynamics* as opposed to system outputs, as ultimately our goal is a framework for designing synthetic systems with dynamics robust to bounded environmental disturbances. Additionally, we seek to develop a framework that enables exclusion of

any system variables that add unwanted model complexity but that do not substantially enrich the dynamical behavior of the system. Therefore, our aim is to develop models that account for environmental disturbances, but only those that substantially impact the dynamics of the system.

Our paper is organized as follows. In Section II we introduce notation, define the concept of a chemical reaction system, and the review the classical chemical master equation: a dynamical model for describing the reaction system dynamics. In Section III, we introduce the notion of a system-environment decomposition on a chemical reaction system preparatory for our main result. In Section IV A and IV B, we derive the main result, an additive decomposition of (plant) system dynamics into two terms: the first being a description of the evolving intrinsic uncertainty in the system and the second being a description of the disturbance that extrinsic uncertainty can have on the intrinsic state. We conclude in Section V with two simple examples of environmental disturbance, a model describing loading effects between two genes and a model describing antibiotic perturbation to transcription and translation rates.

II. Background: The Chemical Master Equation

In this section, we introduce the mathematical framework and notation for our analysis. We begin by reviewing the concept of a chemical reaction system. Since we ultimately seek to decompose this global system into a specific system and its environment, we will refer to it as the global chemical reaction system.

Definition 1: Define $\mathcal{C} = (\mathcal{S}, \mathcal{R})$ to be a *global chemical reaction system* with $\mathcal{S} = \{S_1, .., S_N\}$ being a set containing all N chemical species in the global chemical reaction system. Let $\mathcal{R} = \{R_1, ..., R_M\}$ be a set enumerating all M reactions in $\mathcal{C}$.

Remark 1: The elements of the set $\mathcal{R}$ are reactions. Mathematically, a reaction $R_j \in \mathcal{R}$ is defined based on a species set $\mathcal{S}$ and can be thought of as an ordered 4-tuple of sets $R_j = (\{c_1, .., c_k\}, \{S_1, .., S_k\}, \{d_1, .., d_n\}, \{P_1, ..., P_n\})$, where $c_1, .., c_k, d_1, .., d_n \in \mathbb{N}$, $S_1, .., S_k, P_1, .., P_n \in \mathcal{S}$. The first set of R specifies the stoichiometry of the reactants, the second set the list of reactants, the third set the stoichiometry of the products, and the fourth set the list of the products. Typically, we will follow convention and express the reaction R_j as $c_1S_1 + \cdots + c_kS_k \to d_1P_1 + \cdots + d_nP_n$, and not as an ordered 4-tuple. The above convention can be viewed as an implicit reference to the underlying mathematical object that defines the reaction R_j: an ordered 4-tuple of sets.

The global chemical reaction system is thus a list of all potential chemical species and chemical reactions occurring in a relevant biological chassis, e.g. a cell, an *in vitro* test tube, vesicle, etc. In principle, the size of $\mathcal{S}$ and $\mathcal{R}$ are very large, since it must include all possible partial products of transcription, i.e. aborted transcripts, background molecules critical for metabolism, intermediate metabolites, etc. Most biological models exclude the complexity found in the global chemical reaction system, as it contents are mostly unknown, in addition to being computationally and analytically intractable.

We restrict our attention to global chemical reaction systems whose contents are well stirred, in a fixed volume, and at a constant temperature. Under these conditions, we define $X(t)$ to be a vector of copy numbers, with $X_i(t)$ being the copy number of S_i at time $t, i = 1, .., N$. We suppose that for each reaction $R_j \in \mathcal{R}$ there exists a propensity function $w_j(X(t))$ that characterizes the probability of reaction R_j firing in time interval dt as $w_j(X(t))dt$ [7]. We note this is an assumption, rather than a consequence as in [7] since $\mathcal{C}$ is not necessarily a gas-phase system. We define the stoichiometric transition vector for each reaction R_j as $\xi_j = \begin{bmatrix}\nu_1 & \dots & \nu_N\end{bmatrix}^T$, where ν_k describes the stoichiometric change in X_k during reaction R_j. Thus, if $X = x_o$ before R_j fires, then $X = x_o + \xi_j$ after R_j has fired. Further, with some abuse of notation, we will suppose that if $X = (Y, Z)$, then $\xi_j[Y]$ denotes the subvector of ξ_j that records the stoichiometric change of Y. The chemical master equation of the system $\mathcal{C}$ is then given as

$$\frac{d}{dt}P(X(t)|X(t_o)) = \sum_{j=1}^{M} w_j(X(t) - \xi_j)P(X(t) - \xi_j|X(t_o)) - P(X(t)|X(t_o))\sum_{j=1}^{M} w_j(X(t)) \quad (1)$$

The chemical master equation specifies the evolution of the joint probability mass function of $X(t)$. Since $X(t)$ is a vector of species copy numbers, its entries take on nonnegative integer values. We refer to the set of values that $X(t)$ can take as the configuration space.

III. Decomposition of the Global Chemical Reaction System

Now that we have a way of describing the global chemical reaction system $\mathcal{C}$, we can consider its relationship to a system of interest. This system may coincide with all the measurable chemical species in the global chemical reaction system, a select set of genes under study and their associated transcriptional and translational products, or even a set of chemical species that are associated with a synthetic biocircuit. Our representation of this system should thus be flexible, as it may require the inclusion of specific reporter molecules and their precursor mRNA transcripts, or include only a single chemical species, corresponding to an inducible and measurable protein. The only constraint we impose is that all its chemical species are within $\mathcal{S}$.

Definition 2: Let $S_1, .., S_n \in \mathcal{S}$ be a list of relevant chemical species. We define the chemical reaction system

$$\mathcal{S}^p \equiv (S^p, R^p)$$

associated with this list of species and refer to this as our *system* or *plant*, where $S^p \equiv \{S_1, .., S_n\}$ and $R^p = \{R_j \in \mathcal{R} | \text{all products and reactants of } R_j \text{ are in } S^p\}$.

Notice in defining such a system in the global chemical reaction system, we assume knowledge of a pre-specified list

of chemical species $S_1, .., S_n$. This list of chemical species then determines the list of reactions intrinsic to this system, as they do not require the presence of chemical species outside the system to function. Alternatively, we could proceed by defining a list of relevant reactions and subsequently impose that all products and reactants associated with those reactions be the list of species for our system. However, a reaction set defined in that manner may not include all self-contained reactions of chemical species in the system, as there may be other chemical reactions that only involve elements of S^p. Finally, we use this particular approach as it is typical to think of biological systems first as a collection of chemical species and subsequently enumerate the list of relevant reactions. We define the environmental chemical reaction system as follows:

Definition 3: Define the chemical reaction system $\mathcal{S}^e = (S^e, R^e)$ as the *environmental system*, where

$$S^e \equiv \mathcal{S} \backslash S^p \ , \ R^e \equiv \mathcal{R} \backslash R^p$$

We will suppose X is ordered so that $X = (X^p, X^e)^T$, i.e. the first n elements specify the copy number of the species in S^p while the last $N - n$ elements specify the species in S^e. Viewing the chemical master equation as a state-space model with $P(X^p(t), X^e(t)|X^p(t_o), X^e(t_o))$, we will refer to $P(X^p(t))$ as the state of the system and $P(X^e(t))$ as the state of the environmental system. Finally, we denote the number of reactions in R^p and R^e as m_p and m_e respectively.

IV. An Additive Decomposition of the Chemical Master Equation

Our goal is achieve a representation of the chemical master equation that captures only state of the system $\mathcal{S}^p$, $P(X^p(t))$, how it evolves over time and how the environmental system impacts that evolution. Ideally, we would like to write a decomposition of the form

$$\frac{d}{dt}P(X^p(t)) = f(X^p(t)) + g(X^p(t), X^e(t)). \qquad (2)$$

Such a representation would allow us to include in $f(X^p(t))$ any dynamics that are relevant to the system, e.g. for design or parameter estimation purposes, while the environmental disturbance term $g(X^e(t))$ would act as a perturbation or disturbance to the nominal system's trajectory. As the derivation of the decomposition is long, we divide it into two parts: the first part evaluates the consequences of decomposing the species set $\mathcal{S}$ of the global chemical reaction system and the second part evaluates the consequences of decomposing the reaction set $\mathcal{R}$ of the global chemical reaction system.

A. Part I : Leveraging the Partition on Chemical Species

The primary obstacle to achieving a decomposition of the form (2) is that the chemical master equation (1) describes the evolution of the joint probability mass function $P(X^p(t), X^e(t)|X^p(t_o), X^e(t_o))$. Typically, to separate $P(X^p(t))$ from $P(X^p(t), X^e(t), X^p(t_o), X^e(t_o))$ requires an assumption of independence between the stochastic processes $X^p(t)$ and $X^e(t)$. This is a strong assumption, one that contradicts the very purpose of our analysis: to understand how the environmental state affects system dynamics.

Alternatively, we consider averaging out the effects of $X^e(t)$, i.e. marginalizing the joint probability mass to obtain the marginal in $X^p(t)$. Rather than laboriously analyzing the effect of individual sample trajectories of $X^e(t)$ on $X^p(t)$, this approach has the advantage of describing the average effect of the distribution of sample trajectories $X^e(t)$ on $X^p(t)$. First, we write the chemical master equation to include the decomposition of the global chemical species set $\mathcal{S}$:

$$\frac{d}{dt}P\left(\begin{bmatrix}X^p(t)\\X^e(t)\end{bmatrix}\right) = \sum_{j=1}^{M} w_j \left(\begin{bmatrix}X^p(t)\\X^e(t)\end{bmatrix} - \xi_j\right) P\left(\begin{bmatrix}X^p(t)\\X^e(t)\end{bmatrix} - \xi_j\right) - w_j \left(\begin{bmatrix}X^p(t)\\X^e(t)\end{bmatrix}\right) P\left(\begin{bmatrix}X^p(t)\\X^e(t)\end{bmatrix}\right).$$

Here we have suppressed the convention of carrying the initial condition as a conditioning argument in each probability mass function, as it will make the derivation easier to read. With some abuse of notation, we write the argument of the probability mass function as X^p or X^e, which will be an abbreviation for the probability mass function actually evaluated at a point (x, y) in the configuration space, i.e. $P(X^p(t) = x, X^e(t) = y)$. If we use $P(X^p(t), X^e(t))$ to refer to the *probability mass function*, we will explicitly say so. The same notation will hold true for conditional and marginal probability density functions. Let $\mathcal{S}(X^e)$ denote the set of values that X^e can assume in the configuration space. If we sum over $\mathcal{S}(X^e)$, the left hand side becomes

$$\begin{aligned}
&\sum_{\mathcal{S}(X^e)} \frac{d}{dt}P\left(\begin{bmatrix}X^p(t)\\X^e(t)\end{bmatrix}\right) = \frac{d}{dt}\sum_{\mathcal{S}(X^e)} P\left(\begin{bmatrix}X^p(t)\\X^e(t)\end{bmatrix}\right)\\
&= \frac{d}{dt}\sum_{\mathcal{S}(X^e)} P\left(X^e(t)|X^p(t)\right) P\left(X^p(t)\right)\\
&= \frac{d}{dt}P(X^p(t)) \sum_{\mathcal{S}(X^e)} P\left(X^e(t)|X^p(t)\right)\\
&= \frac{d}{dt}P\left(X^p(t)\right).
\end{aligned}$$

The first equality holds due to uniform convergence of the sum $\sum_{S(X^e)} P(X^e(t)|X^p(t), X^p(t_o), X^e(t_o))$. The second and third equality holds from the law of conditioning. In the last equality, we use the fact that the *probability mass function* $P(X^e(t)|X^p(t), X^p(t_o), X^e(t_o))$ when summed over all values of X^e in the configuration space is unity. We now address the right hand side of the chemical master equation. Summing over $\mathcal{S}(X^e)$ and conditioning on $X^p(t)$ gives

$$\begin{aligned}
&\sum_{\mathcal{S}(X^e)} \Big[\sum_{j=1}^{M} w_j \left(\begin{bmatrix}X^p(t)\\X^e(t)\end{bmatrix} - \xi_j\right) P\left(\begin{bmatrix}X^p(t)\\X^e(t)\end{bmatrix} - \xi_j\right)\\
&\quad - w_j \left(\begin{bmatrix}X^p(t)\\X^e(t)\end{bmatrix}\right) P\left(\begin{bmatrix}X^p(t)\\X^e(t)\end{bmatrix}\right)\Big]\\
&= \sum_{j=1}^{M} \sum_{\mathcal{S}(X^e)} w_j \left(\begin{bmatrix}X^p(t)\\X^e(t)\end{bmatrix} - \xi_j\right) f_c(X^p, X^e) f_m(X^p)\\
&\quad - \sum_{j=1}^{M} \sum_{\mathcal{S}(X^e)} w_j \left(\begin{bmatrix}X^p(t)\\X^e(t)\end{bmatrix}\right) P\left(X^e(t)|X^p(t)\right) P\left(X^p(t)\right)
\end{aligned}$$

where the conditional and marginal probability mass functions are written as

$$f_c(\cdot) = P\left(X^e(t) - \xi_j\left[X^e\right] | X^p(t) - \xi_j\left[X^p\right]\right)$$
$$f_m(\cdot) = P(X^p(t) - \xi_j\left[X^p\right]).$$

For each $j = 1, .., M$ we pull out the marginal of $X^p(t)$ and summing over $\mathcal{S}(X^e)$ gives

$$\frac{d}{dt}P(X^p(t)) = \sum_{j=1}^{M} P(X^p(t) - \xi_j\left[X^p\right])\alpha_j(X^p(t) - \xi_j\left[X^p\right]) - \sum_{j=1}^{M} P\left(X^p(t)\right)\alpha_j(X^p(t))$$

where

$$\alpha_j(X^p(t)) \equiv \sum_{\mathcal{S}(X^e)} w_j\left(\begin{bmatrix}X^p(t)\\ X^e(t)\end{bmatrix}\right) P\left(X^e(t)|X^p(t)\right).$$

In summary, the preceding equations marginalize the master equation of the joint probability mass function, leveraging the decomposition of X into X^p and X^e. We consider $\alpha_j(X^p(t))$ as the averaged propensity functions for the system $\mathcal{S}$, since they can also be expressed as

$$\alpha_j(X^p(t)) = \mathbb{E}_{X^e(t)|X^p(t)}\left[w_j(X^p(t), X^e(t))\right].$$

Writing out the marginalized master equation, we have

$$\frac{d}{dt}P(X^p(t)) = \sum_{j=1}^{M} \alpha_j(X^p(t) - \xi_j\left[X^p\right])P(X^p(t) - \xi_j\left[X^p\right]) - \sum_{j=1}^{M} \alpha_j(X^p(t))P\left(X^p(t)\right) \quad (3)$$

and note that the averaged propensity functions $\alpha_j(X^p(t))$ specify the probability that reaction R_j will happen in the time interval $[t, t + dt]$, $\alpha_j(X^p(t)dt$, averaged over all possible values of $X^e(t)$. The decomposition of the species set $\mathcal{S} = S^p \cup S^e$ thus produces a representation of the chemical master equation that describes only the evolution of the marginal density $P(X^p(t))$.

B. Part II: Leveraging the Partition on the Chemical Reactions

If we now incorporate the decomposition on the reaction set $\mathcal{R}$, we can also rewrite the propensity functions in terms of the m_p reactions that only involve chemical species in S^p and m_e reactions involving system or environmental species. Notice the term $\sum_{j=1}^{M} \alpha_j(X^p(t) - \xi_j\left[X^p\right])P(X^p(t) - \xi_j\left[X^p\right])$ can be written as

$$\begin{aligned}
&\sum_{j=1} \alpha_j(X^p(t) - \xi_j\left[X^p\right])P\left(X^p(t) - \xi_j\left[X^p\right]\right)\\
=&\sum_{j=1}^{m_p} \alpha_j(X^p(t) - \xi_j\left[X^p\right])P\left(X^p(t) - \xi_j\left[X^p\right]\right)\\
+&\sum_{j=1}^{m_e} \alpha_j(X^p(t) - \xi_j\left[X^p\right])P\left(X^p(t) - \xi_j\left[X^p\right]\right)
\end{aligned}$$

and since the first m_p reactions do not involve X^e, we can write for each reaction $j = 1, ..m_p$ the associated propensity function for those reactions as $w_j(X^p(t), X^e(t)) = w_j(X^p(t))$ and so we can further write $\sum_{j=1}^{M} \alpha_j(X^p(t) - \xi_j\left[X^p\right])P(X^p(t) - \xi_j\left[X^p\right])$ as

$$\begin{aligned}
&= \sum_{j=1}^{m_p} P\left(X^p(t) - \xi_j\left[X^p\right]\right) \times\\
&\sum_{\mathcal{S}(X^e)} w_j\left(X^p(t) - \xi_j\left[X^p\right]\right) P\left(X^e(t)|X^p(t)\right)\\
&+ \sum_{j=1}^{m_e} P\left(X^p(t) - \xi_j\left[X^p\right]\right) \sum_{\mathcal{S}(X^e)} \alpha_j\left(\begin{bmatrix}X^p(t)\\ X^e(t)\end{bmatrix} - \xi_j\right)\\
&= \sum_{j=1}^{m_p} w_j\left(X^p(t) - \xi_j\left[X^p\right]\right) \times\\
&P\left(X^p(t) - \xi_j\left[X^p\right]\right) \sum_{\mathcal{S}(X^e)} P\left(X^e(t)|X^p(t)\right)\\
&+ \sum_{j=1}^{m_e} P\left(X^p(t) - \xi_j\left[X^p\right]\right) \sum_{\mathcal{S}(X^e)} \alpha_j\left(\begin{bmatrix}X^p(t)\\ X^e(t)\end{bmatrix} - \xi_j\right)\\
&= \sum_{j=1}^{m_p} w_j\left(X^p(t) - \xi_j\left[X^p\right]\right) P\left(X^p(t) - \xi_j\left[X^p\right]\right) \quad (1)\\
&+ \sum_{j=1}^{m_e} \alpha_j(X^p(t) - \xi_j\left[X^p\right])P\left(X^p(t) - \xi_j\left[X^p\right]\right)
\end{aligned}$$

with a similar derivation holding for $\xi_j \equiv 0$, thus implying that the marginalized chemical master equation, or state-space model for $P(X^p(t))$, becomes:

$$\begin{aligned}
&\frac{d}{dt}P\left(X^p(t)\right)\\
&= \sum_{j=1}^{m_p} w_j\left(X^p(t) - \xi_j\left[X^p\right]\right) P\left(X^p(t) - \xi_j\left[X^p\right]\right)\\
&- \sum_{j=1}^{m_p} w_j\left(X^p(t)\right) P\left(X^p(t)\right)\\
&+ \sum_{j=1}^{m_e} \alpha_j\left(\begin{bmatrix}X^p(t)\\ X^e(t)\end{bmatrix} - \xi_j\right) P\left(X^p(t) - \xi_j\left[X^p\right]\right)\\
&- \sum_{j=1}^{m_e} \alpha_j\left(\begin{bmatrix}X^p(t)\\ X^e(t)\end{bmatrix}\right) P\left(X^p(t)\right)\\
&\equiv f^P(P(X^p(t))) + f^E(P(X^e(t)|X^p(t)), P(X^p(t))) \quad (4)
\end{aligned}$$

where '$\equiv$' indicates that we define f^P and f^E as

$$\begin{aligned}
f^P(\cdot) = &\sum_{j=1}^{m_p} w_j\left(X^p(t) - \xi_j\left[X^p\right]\right) P\left(X^p(t) - \xi_j\left[X^p\right]\right)\\
&- \sum_{j=1}^{m_p} w_j\left(X^p(t)\right) P\left(X^p(t)\right),\\
f^E(\cdot) = &\sum_{j=1}^{m_e} \alpha_j\left(\begin{bmatrix}X^p(t)\\ X^e(t)\end{bmatrix} - \xi_j\right) P\left(X^p(t) - \xi_j\left[X^p\right]\right)\\
&- \sum_{j=1}^{m_e} \alpha_j\left(\begin{bmatrix}X^p(t)\\ X^e(t)\end{bmatrix}\right) P\left(X^p(t)\right).
\end{aligned}$$

To summarize, we first imposed a decomposition on the chemical species of the global chemical reaction system to obtain two subsystems: the system of interest and its environment. Second, we marginalized the chemical master equation to obtain a master equation that described only the time-evolution of the state of the system $P(X^p(t))$ using averaged propensity functions $\alpha_j(X^p(t))$. Finally, we imposed knowledge about the dependencies of the reactions and this resulted in a simple additive decomposition of the marginalized dynamics.

$$\frac{d}{dt}P\left(X^p(t)\right) = f^P(P(X^p(t))) + f^E(P(X^e(t)|X^p(t)), P(X^p(t)))$$

Notice this decomposition depends on two functionals: f^P which depends only on the state of the system $P(X^p(t))$ and f^E which depends on the state of the environmental system $P(X^e|X^p(t))$ and the state of the system $P(X^p(t))$. Since our derivation began from the chemical master equation (1) of the state of the global chemical reaction system $\mathcal{S}$ and since we have not imposed any additional assumptions—only using the normalization property of a probability mass function and conditioning arguments—the decomposition is exactly consistent with the dynamics of the global chemical reaction system.

Also, notice that the term $f^P(P(X^p(t)))$ depends on the exact propensity functions $w_j(X^p(t))$ in its definition. Thus, if $S^e = \emptyset$ or $R^e = \emptyset$, the term $f^P(P(X^p(t)))$ is precisely the right half-side of the chemical master equation (1). This is important for several reasons: 1) the term $f^P(X^p(t))$ can be viewed as the *complete* dynamics for a simple model system involving only the system variables, see the supplementary information of [9], [20], [8] for examples, 2) recognizing that $f^P(X^p(t))$ describes a simple model system's dynamics, it may be possible to omit any species in S^p that make the simplified model $\dot{P}(X^p(t)) = f^P(X^p(t))$ intractable or to introduce additional species from S^e to ensure the presence of conservation laws, potentially making the configuration space of S^p finite.

V. Using the System-Environment Decomposition to Model Environmental Disturbances: Examples

A. Ribosomal loading between two genes

We now consider an example system to illustrate how our decomposition enables modeling of environmental disturbances. We suppose the system of interest consists of $n = 2$ chemical species, S_1 is an mRNA m which encodes the protein p. There are $m_p = 4$ system reactions:

$$\emptyset \xrightarrow{k_{TX}} m, \qquad m \xrightarrow{k_{TL}} m + p,$$
$$m \xrightarrow{\delta_m} \emptyset, \qquad p \xrightarrow{\delta_p} \emptyset$$

A diagram illustrating the structure of this simple system is shown in Figure 1. We will assume that basal expression of mRNA is very low in the absence of an environmental cue (e.g. transcriptional machinery is scarce) which results in a small basal transcription rate k_{TX}. Furthermore, we will

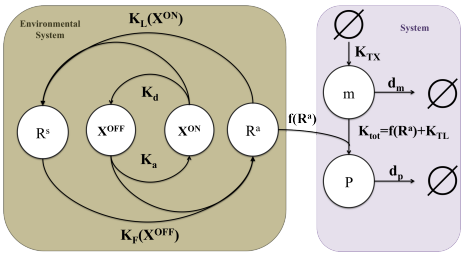

Fig. 1. A schematic illustrating the interactions between chemical species in the system (5). mRNA and protein are produced constitutively, while ribosome in the abundant state R^a are able to augment production. A competing gene X sequesters ribosome away from m, facilitating R's transition from the abundant state R^a to the scarce state R^s. A confluence of two arrows indicates that either k_L depends on X^{ON}, k_F depends on X^{OFF}, or K_{tot} depends on $f(R)$.

assume that in the absence of a separate environmental cue (e.g. ribosomal and translation machinery is scarce), the rate of translation k_{TL} is quite small. Finally, since the rates represent weak or basal expression, we suppose that k_{TX} and k_{TL} are zero-order rates that do not depend on the actual concentration of mRNA or protein (i.e. they are rate limited by RNAP and ribosome counts). We suppose that the degradation rates do depend on the copy number of m and p. We write the dynamics of the chemical master equation for the isolated (or toy) system as $\dot{P}(p, m, t) = f^P(P(X^p, t))$ where $f^P(P(X^p, t))$ equals

$$\begin{aligned} f^P(P(X^p,t)) &= k_{TL}P(p-1,m,t) + k_{TX}P(p,m-1,t) \\ &+ \delta_m(m+1)P(p,m+1,t) + \delta_p(p+1)P(p+1,m,t) \\ &- k_{TL}P(p,m,t) - k_{TX}P(p,m,t) - (\delta_m m + \delta_p p)P(p,m,t) \end{aligned} \tag{5}$$

The solution for this system is obtained by computing the probability generating function $F(z_1, z_2) = \sum_{m,p} z_1^p z_2^m P(p, m, t)$. Transforming the system (5) we obtain

$$\begin{aligned} \frac{\partial F}{\partial t} &= (k_{TL}z_1 + \delta_p - \delta_p)\frac{\partial F}{\partial z_1} + (k_{TX}z_2 + \delta_m - \delta_m z_2)\frac{\partial F}{\partial z_2} \\ &+ (-k_{TL} - k_{TX})F(z_1, z_2, t). \end{aligned}$$

By applying the method of characteristics, we obtain a closed form expression for the probability generating function $F(z_1, z_2, t)$:

$$\begin{aligned} &e^{-(k_{TX}+k_{TL})t} \sum_{m,p\in\mathbb{Z}\geq 0} \left((z_1+\delta_p)e^{(d_p-k_{TL})t} - \delta_p\right)^p \\ &\times \left((z_2+\delta_m)e^{(\delta_m-k_{TX})t} - \delta_m\right)^m P(p,m,0) \end{aligned}$$

from which we can calculate the probability mass function to be written as follows:

$$P(p=k_1,m=k_2,t)=e^{-(k_{TX}+k_{TL})t}\frac{1}{k_1!\,k_2!}\times \tag{6}$$
$$\sum_{m,p\in\mathbb{Z}\geq 0}\Bigg(\prod_{j=0,k_1\leq p}^{k_1-1}(p-j)f_1(k_1,t)\left(\delta_p e^{-(k_{TL}-\delta_p)t}-\delta_p\right)^{p-k_1}$$
$$\prod_{i=0,k_2\leq m}^{k_2-1}(m-i)f_2(k_2,t)\left(\delta_m e^{-(k_{TX}-\delta_m)t}-\delta_m\right)^{m-k_2}$$
$$\times P(p,m,0)\Bigg)$$

and $f_1(k_1,t)=e^{-(k_{TL}-\delta_p)k_1 t}, f_2(k_2,t)=e^{-(k_{TX}-\delta_m)k_2 t}$ As our system has been chosen to be relatively simple, i.e. reflecting simplified models that typically exclude environmental species from the list of chemical species, the solution to (5) is a closed form analytical solution. Notice that the configuration space is not necessarily finite, so we sum over the positive integers.

Our goal is to now modify the system dynamics using $f^E(P(X^e(t)))$ to explore the effect of ribosomal loading. Because the complexity of the system may introduce nonlinearities into the generation function, our approach will be to simulate the perturbed system and compare it to the isolated system. We suppose the environmental system contains the ribosomal species R necessary to increase translation rates, but that a species X has the ability to sequester ribosomes away from translating m to p. We will suppose that ribosomes can assume two states: either abundant or scarce and we denote them as R^a and R^s, respectively. When X is in the X^{ON} state, it facilitates the conversion of R^a to R^s (and vice versa when X is on the X^{OFF} state). We denote the environmental reactions as follows:

$$R^s \xrightarrow{k_{F(X)}} R^a, \qquad X^{OFF} \xrightarrow{k_a} X^{ON},$$
$$R^a \xrightarrow{k_L(X)} R^s, \qquad X^{ON} \xrightarrow{k_d} X^{OFF},$$
$$R^a + m \xrightarrow{k_{TLF}} R^a + p + m, \tag{7}$$

with the last reaction denoting enhanced translation rates of m due to the abundance of ribosomes.

We suppose that X is a gene regulated by some external input and switches randomly between its off and on states, independent of the current state of R. We assume its expression to be strong, so that the dynamics of m and p do not impact its transition rates. We then write the solution for X as $p(X,t)=e^{At}P(X,0)$ where

$$A=\begin{bmatrix}-k_d & k_a\\ k_d & -k_a\end{bmatrix},$$

We suppose that $P(X(0))=\begin{bmatrix}k_a & k_d\end{bmatrix}^T$ where $k_a+k_d=1$. Under these assumptions, we can write

$$P(X(t))=e^{At}\begin{bmatrix}k_a\\ k_d\end{bmatrix}=\begin{bmatrix}k_a\\ k_d\end{bmatrix}$$

Further, substituting and conditioning with $P(X(t))$ gives us the following linear equation.

$$\frac{d}{dt}\begin{bmatrix}P(R^a,t)\\ P(R^s,t)\end{bmatrix}=\begin{bmatrix}-k_a k_L & k_d k_F\\ k_a k_L & -k_d k_F\end{bmatrix} \tag{8}$$

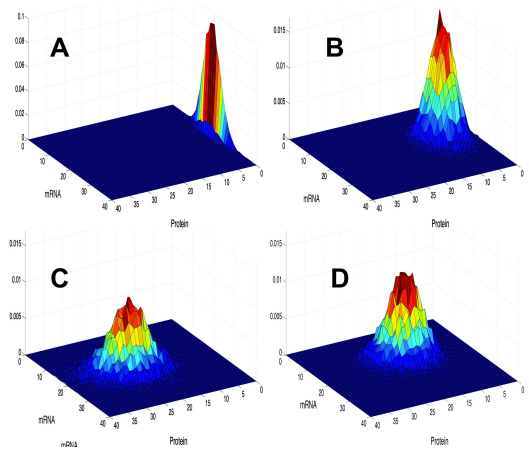

Fig. 2. (A) The unperturbed system (7) plotted at $t = 13$ minutes. Notice that weak basal expression of protein produces a distribution that reflects high mRNA average copy number but low average protein copy number. Here $k_a = k_d = 0$. (B) The system simulated with ribosomal loading effects and a gene X that is in X^{ON} state with high probability ($k_a = .9$) Notice the significant reduction in protein expression when compared against C or D. (C) The system simulated with X in the off state ($k_d = .9$) with high probability and in the on state with low probability ($k_a = 1$). Consequently, the system has significantly higher concentrations of protein than in plots A or B. (D) The system (7) when X has one half probability of being on and one half probability of being off. All simulations were performed in MATLAB using the Gillespie Stochastic Simulation Algorithm. Common parameters used for all four simulations were $k_F = 0.6$ /s , $k_L = 0.4$/s, $k_{TX} = 0.052$ /s , $k_{TLF} = 0.4$ /s, $\delta_m = 1.4 \times 10^{-3}$ s , $\delta_p = .015$ s and $m(0), p(0) \sim$(Poiss(7)))

The solution can be substituted into $f^E(P(X^e|p,m,t),P(X^p|t))$, allowing us to write it as

$$-k_{TLF}P(R^a,t)P(p,m,t)+k_{TLF}P(R^a,t)P(p-1,m,t).$$

A simulation of the system is plotted in Figure 2; we see that protein expression is the highest when the probability that gene X stays off is close to 1. The reason is that when X is on, the amount of free ribosomes decrease (sequestration of ribosomes by X) and the amount of p produced is less.

To summarize, in this example we have posed a simple approach for capturing the effects of enzyme sequestration or loading effects [21]. We showed that the state of the protein and mRNA of our system can be strongly influenced by the state of X, which is a chemical species that does not directly interact with m or p. Thus, ribosomal loading can lead to indirect interactions between chemical species, even in a chemical master equation modeling framework.

B. Stochastic switch with antibiotic attenuation

We now examine a particular approach for modeling the effect of antibiotics on a system. We suppose the system carries no resistance for two antibiotics and that these two antibiotics, when perturbing the system, can reduce the rate of transcription and translation respectively. The system is composed of an mRNA and a protein, whose expression is controlled by an upstream binary oscillator X.

We denote the two states of the binary oscillator as X^H and X^L. When in the high state, transcription and translation of m and p occurs using the same chemical

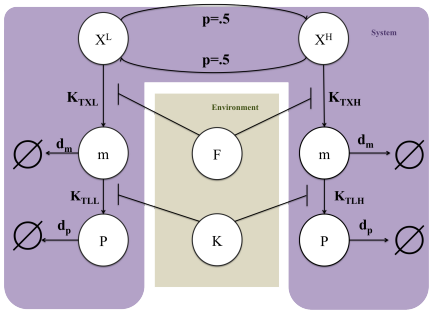

Fig. 3. A schematic illustrating the reaction channels and interactions between chemical species in Example 2. F and K are antibiotics that attenuate transcription and translation rates respectively. X is a binary oscillator switching back and forth between a low and a high state with no transition bias. In the high state, m and p are produced with faster transcription and translation rates. At the low state, m and p are produced at much slower rates.

reactions as in Example 1; however, we denote the high state propensity coefficient of transcription as k_{TXH} and the high state propensity coefficient of translation K_{TXL}. In the low state, the reaction structure is the same again, but this time with the low transcription and translation reaction propensity coefficients k_{TXL} and k_{TLL}. A diagram of the system is shown in Figure 3. Let $P_H(p,m,t)$ be the probability mass function for a system with $X = X^H$ and $P_L(p,m,t)$ be the probability mass function for a system with $X = X^L$. Notice that P_H (and P_L) can be obtained by a direct application of the solution generated using the method of probability generating functions from Example 1, evaluated with $k_{TX} = k_{TXH}$ and $k_{TL} = k_{TLH}$ ($k_{TX} = k_{TLH}$ and $k_{TL} = k_{TLL}$). Thus, we can calculate $P(p,m,t)$ as

$$\begin{aligned} P(p,m,t) &= P(p,m,t|X^H)P(X^H,t) + P(p,m,t|X^L)P(X^L,t) \\ &= P_H(p,m,t)P(X^H,t) + P_L(p,m,t)P(X^L,t) \end{aligned}$$

We suppose that $P(X,t) = P(X) = \frac{1}{2}$ is stationary and we model it as a Bernoulli random variable with parameter $p = .5$, i.e. the oscillator is unbiased. With these assumptions, we can calculate the solution of the system $P(p,m,x,t)$ and in particular, the marginal $P(p,m,t)$. Without any environmental disturbances, the distribution $P(p,m,t)$ has a bimodal distribution, see Figure 4A. However, let us now introduce an environmental system to add disturbance to the dynamics of the system. In particular, we suppose there are two types of antibiotic added to the system. The first, which we denote as K, can be thought of as an antibiotic that disrupts ribosomal activity (e.g. kanamycin, streptomycin, chloromphenicol). The second, which we denote as F, can be viewed as an antibiotic that disrupts the transcription process (e.g. rafimycin). Accordingly, we suppose their effect on transcriptional and translation reactions has an overall negative effect. In particular, we suppose that their reactions are of the following form:

$$F \xrightarrow{-\Delta_{TX}F} m, \qquad K + m \xrightarrow{-\Delta_{TL}K} P \tag{9}$$

That is, regardless if $X = X^H$ or $X = X^L$, the antibiotic K decreases the rate of translation as a function of K while the antibiotic F slows the rate of transcription as a function of F. Let us assume that K and F have independent distributions to describe their copy number. Besides this assumption, let us suppose that we do not know the distribution. We can write the environmental disturbance $f^E(P(X^e(t)|X^p(t))$ as

$$\begin{aligned} &\sum_{x=0}^{\infty} -\Delta_{TX} x P(F = x|p,m,t)P(p,m-1,t) \\ &+ \sum_{y=0}^{\infty} -\Delta_{TL} y P(K = y|p,m,t)P(p,m-1,t) \\ &- \left(\sum_{x=0}^{\infty} -\Delta_{TX} x P(F = x|p,m,t)P(p,m,t)\right) \\ &- \left(\sum_{y=0}^{\infty} -\Delta_{TL} y P(K = y|p,m,t)P(p,m,t)\right) \end{aligned}$$

In this scenario, we have an analytically tractable model for our system but no clear expression for the conditional distribution of the environment $P(X^E|p,m,t)$. Hence there is no way to compute or simulate $P(X^e|p,m,t)$. However, we can justify using a particular distribution by the principle of maximum entropy, which specifies the functional form of the distribution if there are constraints on the moments of $P(X^e|p,m,t)$. Certainly, we can assume that the mean value of K and F are both finite. If so, then from Theorem 5.7 in [22] we then can write

$$P(X^e = x|p,m,t) = Cr^x$$

where $C = \frac{1}{\mu_{X^e}}, r = \frac{\mu_{X^e}}{\mu_{X^e}+1}$ and $X^e = F$ or K. Further, if we suppose that μ_F and μ_K are given (or estimated using empirical measurements), then we get that

$$\begin{aligned} f^E(\cdot) = &-\Delta_{TX}\mathbb{E}_{F|p,m,t}[F]\,P(p,m-1,t) \\ &-\Delta_{TL}\mathbb{E}_{K|p,m,t}[K]\,P(p,m-1,t) \\ &+\Delta_{TX}\mathbb{E}_{F|p,m,t}[F]\,P(p,m,t) \\ &+\Delta_{TL}\mathbb{E}_{K|p,m,t}[K]\,P(p,m,t) \end{aligned}$$

Notice that this expression for f^E leads to a closed form solution of $P(X^p,t) = P(p,m,t)$ in this case. Writing down the expression for $P_H(p,m,t)$ and $P_L(p,m,t)$ the reader will see that f^E has the effect of perturbing the transcription and translation rates of the original system to be

$$\begin{aligned} k_{TX}^{pet} &= k_{TX} - \Delta_{TX}\mathbb{E}_{F|p,m,t}[F], \\ k_{TL}^{pet} &= k_{TL} - \Delta_{TL}\mathbb{E}_{K|p,m,t}[K], \end{aligned}$$

where k_{TX} and k_{TL} can be replaced with k_{TXH}, k_{TXL}, k_{TLH}, k_{TLL} respectively to obtain $P_H(p,m,t)$ and $P_L(p,m,t)$ as a function of the perturbed rates. The final solution is then calculated as before, as

$$\begin{aligned} P(p,m,t) &= P(p,m,t|X^H)P(X^H,t) + P(p,m,t|X^L)P(X^L,t) \\ &= P_H(p,m,t)P(X^H,t) + P_L(p,m,t)P(X^L,t) \end{aligned}$$

In Figure 4, we plot the results of our simulation. When we perturb just with antibiotic F, the mean of the mRNA decreases while the mean of the protein remains approximately the same (the second peak remains and bimodality is not

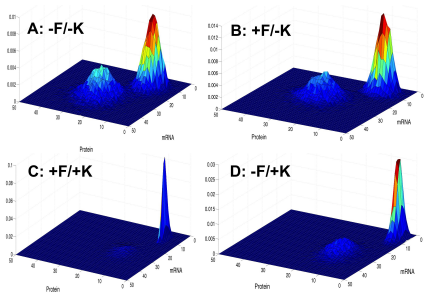

Fig. 4. A: The unperturbed system is bimodal, with a large peak at low mRNA and low protein count and a smaller peak at high protein and high mRNA count. B: Antibiotic F reduces transcription, thus reducing average mRNA count. (compare to A) C: Antibiotics F and K reduce transcription and translation and abolish bimodality of the system. D: Antibiotic K reduces translation, abolishing bimodality but leaves a strong peak at lower mRNA copy numbers.

abolished). When we perturb the system with just K, there is a decrease in translation rates and bimodality disappears. Finally, when we perturb with F and K at the same time, both protein and mRNA copy number decrease as expected and we lose bimodality.

Our example thus illustrates a simple way of modeling the effects of antibiotics on transcription and translation. It does not require complete knowledge about the distribution of the antibiotics but it does require some estimate on the parameters for μ_F, μ_K, and Δ_{TX}, Δ_{TL}.

VI. Conclusions and Future Work

In this work we derived a decomposition of the chemical master equation into an additive sum of two terms: the first describes the dynamics of a system of interest, the second has the interpretation of the averaged environmental disturbance or more precisely, averaged propensity functions for all reactions involving environmental species. We illustrated the use of this decomposition to model two types of environmental effects: 1) the effect of ribosomal loading from an orthogonal gene with high (or low) demand for the ribosomes in a cell, 2) the effect of antibiotics on a bimodal system with unknown environmental distribution. We approximated the latter environmental effect by using a maximum entropy distribution to show that antibiotics directly perturb the transcription and translation rates of the system, scaled by the mean of the antibiotic copy number distribution. Future work will involve experimental studies to ascertain the appropriate model classes for describing the various environmental disturbances, i.e. antibiotic stress, heat stress, oxidative stress, osmotic stress, nutritional stress, etc. We plan on developing novel system identification procedures, as well as levering existing techniques [23] to build a library of models that characterize the manner in which environmental disturbances impact both synthetic and natural biological processes.

VII. Acknowledgments

The authors thank Jongmin Kim, Ophelia Venturelli, Marcella Gomez and Matanya Horowitz for insightful discussions and acknowledge funding from a National Defense Science and Engineering Graduate Fellowship.

References

[1] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, "Stochastic gene expression in a single cell", *Science*, vol. 297, no. 5584, pp. 1183–1186, 2002.

[2] I. Lestas, J. Paulsson, N.E. Ross, and G. Vinnicombe, "Noise in gene regulatory networks", *Automatic Control, IEEE Transactions on*, vol. 53, no. Special Issue, pp. 189–200, Jan.

[3] M. B. Elowitz and S. Leibler, "A synthetic oscillatory network of transcriptional regulators", *Nature*, vol. 403, no. 6767, pp. 335–338, 2000.

[4] J. Stricker et al, "A fast, robust, and tunable synthetic gene oscillator", *Nature*, vol. 456, no. 7221, pp. 516–519, 2008.

[5] N. Cookson et al, "Queueing up for enzymatic processing: correlated signaling through coupled degradation", *Molecular Systems Biology*, vol. 7, no. 561, 2011.

[6] NG Van Kampen, *Stochastic processes in physics and chemistry*, North Holland, 2007.

[7] D. T. Gillespie, "A rigorous derivation of the chemical master equation", *Physica A: Statistical Mechanics and its Applications*, vol. 188, no. 13, pp. 404 – 425, 1992.

[8] D. R. Rigney, "Stochastic model of constitutive protein levels in growing and dividing bacterial cells", *Journal of Theoretical Biology*, vol. 76, no. 4, pp. 453 – 480, 1979.

[9] P. S. Swain, M. B. Elowitz, and E. D. Siggia, "Intrinsic and extrinsic contributions to stochasticity in gene expression", *Proceedings of the National Academy of Sciences*, vol. 99, no. 20, pp. 12795–12800, 2002.

[10] B. Munsky and M. Khammash, "The finite state projection algorithm for the solution of the chemical master equation", *The Journal of Chemical Physics*, vol. 124, pp. 044104, 2006.

[11] T. A. Henzinger, M. Mateescu, and V. Wolf, "Sliding window abstraction for infinite markov chains", in *In Proc. CAV, volume 5643 of LNCS*. 2009, pp. 337–352, Springer.

[12] M. Nip, J. P. Hespanha, and M. Khammash, "A spectral methods-based solution of the chemical master equation for gene regulatory networks", in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*. IEEE, 2012, pp. 5354–5360.

[13] S. Engblom, "Spectral approximation of solutions to the chemical master equation", *Journal of computational and applied mathematics*, vol. 229, no. 1, pp. 208–221, 2009.

[14] P. Deuflhard, W. Huisinga, T. Jahnke, and M. Wulkow, "Adaptive discrete galerkin methods applied to the chemical master equation", *SIAM Journal on Scientific Computing*, vol. 30, no. 6, pp. 2990–3011, 2008.

[15] S. Cardinale and A. Arkin, "Contextualizing context for synthetic biology–identifying causes of failure of synthetic biological systems", *Biotechnology Journal*, 2012.

[16] J. Pedraza and A. van Oudenaarden, "Noise propagation in gene networks", *Science Signalling*, vol. 307, no. 5717, pp. 1965, 2005.

[17] K. Zhou J. Doyle, K. Glover, *Robust and Optimal Control*, Prentice Hall, Englewood Cliffs, N.J., 1996.

[18] J. J. Tabor, A. Levskaya, and C. A. Voigt, "Multichromatic control of gene expression in escherichia coli", *Journal of Molecular Biology*, vol. 405, no. 2, pp. 315 – 324, 2011.

[19] A. Milias-Argeitis, S. Summers, J. Stewart-Ornstein, I. Zuleta, D. Pincus, H. El-Samad, M. Khammash, and J. Lygeros, "In silico feedback for in vivo regulation of a gene expression circuit", *Nature Biotechnology*, 2011.

[20] A. Hilfinger and J. Paulsson, "Separating intrinsic from extrinsic fluctuations in dynamic biological systems", *Proceedings of the National Academy of Sciences*, vol. 108, no. 29, pp. 12167–12172, 2011.

[21] D. Del Vecchio, A. J. Ninfa, and E. D. Sontag, "Modular cell biology : retroactivity and insulation", *Mol. Syst. Biol*, vol. 4, pp. 161, 2008.

[22] K. Conrad, "Probability distributions and maximum entropy", *Expository Paper*, vol. 6, 2004.

[23] G. Lillacci and M. Khammash, "Parameter estimation and model selection in computational biology", *PLoS Computational Biology*, vol. 6, no. 3, pp. e1000696, 2010.