Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Noise suppression in auto-regulatory gene networks

Abhyudai Singh and João Pedro Hespanha

Abstract— **Living cells are characterized by small populations of key molecular components that have large stochastic noise associated with them. Various gene network motifs exist within cells that help reduce these stochastic fluctuations. A common such motif is an auto-regulatory gene network where the protein expressed from the gene inhibits its own transcription. Here the transcription rate of the gene is given as some function of the number of protein molecules present in the cell. We refer to this function as the *transcriptional response* of the gene network.**

We develop analytical formulas that relate the stochastic fluctuations in protein numbers with the functional form of the transcriptional response. This is done by first approximating the transcriptional response by a polynomial and then using recently developed moment closure techniques to solve for the statistical moments of the protein population. We show that the protein noise level in these auto-regulatory gene networks is related to the stability of the network and increasing (decreasing) stability leads to attenuation (magnification) of protein noise. Using the above formulas we also investigate the transcriptional response of a specific gene network in lambda phage and show that this network is especially effective at reducing stochastic fluctuations in protein levels.

I. INTRODUCTION

Gene expression and regulation is inherently a noisy process. The origins of this stochasticity lie in the probabilistic nature of transcription and translation and low copy numbers of RNAs and proteins within cells, which can lead to large statistical fluctuations in molecule numbers. Recent work [1], [2], [3] has provided considerable experimental evidence for these stochastic fluctuations and may explain for the large amounts of cell to cell variation observed in genetically identical cells exposed to the same environmental conditions [4], [5]. Various gene network motifs within cells decrease/increase these stochastic fluctuations. A common such motif is an auto-regulatory gene network where the protein expressed from the gene inhibits/activates its own transcription [6]. Both theoretical and experimental studies have shown that negative feedback in these auto-regulatory gene networks reduces stochastic fluctuations in the protein population [7], [8], [9], [10], [11], [12] whereas positive feedback has the opposite effect [13], [14].

We consider a simple gene expression model for a protein X with molecular count $\mathbf{x}(t)$ at time t. The protein is transcribed at a rate $g(\mathbf{x})$ and we call the function g the transcriptional response of the network. Each transcription event leads to the formation of $\mathbf{N}$ protein molecules where $\mathbf{N}$ is a random variable with mean N and variance V^2. We assume that the protein is degraded at a constant rate d. In the stochastic formulation of this gene expression model, transcription and degradation are treated as probabilistic events and $\mathbf{x}(t)$ is a stochastic process. Details on the stochastic formulation are provided in Section II. We quantify the noise in $\mathbf{x}(t)$ by its *coefficient of variation* defined by

$$CV_X^2 := \frac{\mathbf{E}^*[\mathbf{x}^2] - \mathbf{E}^*[\mathbf{x}]^2}{\mathbf{E}^*[\mathbf{x}]^2} \tag{1}$$

where $\mathbf{E}^*[\mathbf{x}^k]$ denotes the steady-state value of the moment $\mathbf{E}[\mathbf{x}^k]$, $k \in \{1,2\}$. We show in Section II that when there is no auto-regulation, i.e., the transcriptional response $g(\mathbf{x})$ is independent of $\mathbf{x}$ and equal to a constant K, the noise in the protein numbers is given by

$$CV_X^2 = \frac{d(N^2+V^2+N)}{2KN^2}. \tag{2}$$

In Section III we consider what happens to the noise in the protein when the transcriptional response is dependent on $\mathbf{x}$. We first use a linear approximation for $g(\mathbf{x})$ given by

$$g(\mathbf{x}) \approx g(\mathbf{x}^*) + g'(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*) \tag{3}$$

where $\mathbf{x}^*$ is the steady-state protein count from the deterministic chemical rate equations. As we will see later, this is a valid approximation as long as the stochastic fluctuations in the protein numbers are sufficiently small. We quantify the stability of the equilibrium $\mathbf{x}^*$ by the response time of the gene network T_r, a quantity defined in terms of the deterministic chemical rate equations. In particular, the response time of the gene network is the time taken for any initial perturbation about $\mathbf{x}^*$ to decay by 50% of its initial value. We show in Section III-C that when the transcriptional response is given by (3), the stochastic noise in the protein is

$$CV_{Xlinear}^2 = \frac{T_r}{T_p}\frac{N^2+V^2+N}{2\mathbf{x}^*N} \tag{4}$$

where T_p is the protein half-life. Hence for a fixed $\mathbf{x}^*$, decreasing (increasing) the response time T_r of the gene network attenuates (magnifies) stochastic noise in the protein. We also investigate the effects of nonlinearities in the transcriptional response on the protein noise level in Section III-D. In particular, we consider transcriptional responses given by

$$g(\mathbf{x}) \approx g(\mathbf{x}^*) + g'(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*) + \frac{1}{2}g''(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)^2. \tag{5}$$

This material is based upon work supported by the Institute for Collaborative Biotechnologies through grant DAAD19-03-D-0004 from the U.S. Army Research Office and by the National Science Foundation under Grant No. CCR-0311084.

A.Singh and J.P.Hespanha are with the Center for Control Engineering and Computation University of California, Santa Barbara, CA 93101. abhi@engineering.ucsb.edu, hespanha@ece.ucsb.edu

We show that a transcriptional response that is convex (concave) at $\mathbf{x}^*$ causes the noise in the protein to be larger (smaller) then what would be predicted by equation (4) which assumed a linear $g(\mathbf{x})$ as in (3).

Finally in Section IV, we investigate the transcriptional response for a specific gene network: the lambda repressor gene network. The transcriptional response $g(\mathbf{x})$ of this network is an increasing function when $\mathbf{x}$ is small (i.e., protein activates itself when its numbers are small) and a decreasing function when $\mathbf{x}$ is large (i.e., protein inhibits itself when its numbers are large). Using the above developed formulas we show that this particular transcriptional response is especially effective at reducing noise levels in the lambda repressor protein.

II. Intrinsic noise in unregulated gene expression

We start by considering a simple model of gene expression where a gene expresses a protein X at a constant rate K. Each expression event leads to the formation of $\mathbf{N}$ molecules of the protein X. Recent work suggests that the burst of proteins from each mRNA transcript follows a geometric distribution [15]. Thus instead of assuming $\mathbf{N}$ to be a constant we assume it to be a random variable with mean N and variance V^2. We also assume that the protein decays at a constant rate d. Our model omits the mRNA dynamics. This is a valid approximation as long as the protein's life time is much longer than the mRNA's life time, which is generally the case in gene-protein networks [16]. Ignoring the mRNA dynamics leads to relatively simple expressions for the protein noise level, which help develop a qualitative understanding of how noise level changes in response to alterations of the gene network parameters

In the stochastic formulation, gene expression and protein degradation are treated as probabilistic events where the probabilities of them happening in the infinitesimal time interval $(t, t+dt]$ are given by

$$\Pr\{\mathbf{x}(t+dt) = x + \mathbf{N} \mid \mathbf{x}(t) = x\} = K dt \tag{6a}$$

$$\Pr\{\mathbf{x}(t+dt) = x - 1 \mid \mathbf{x}(t) = x\} = dx dt, \tag{6b}$$

respectively, where $\mathbf{x}(t)$ denotes the number of molecules of protein X at time t. A convenient way to model the time evolution of the number of molecules $\mathbf{x}$ is through a Stochastic Hybrid System (SHS) characterized by trivial continuous dynamics

$$\dot{\mathbf{x}} = 0, \tag{7}$$

two reset maps

$$\mathbf{x} \mapsto \phi_1(\mathbf{x}) = \mathbf{x} + \mathbf{N}, \quad \mathbf{x} \mapsto \phi_2(\mathbf{x}) = \mathbf{x} - 1 \tag{8}$$

with corresponding transition intensities given by

$$\lambda_1(\mathbf{x}) = K, \ \lambda_2(\mathbf{x}) = d\mathbf{x} \tag{9}$$

[19]. In order to gauge the noise level in the protein population, we determine the time evolution of the first and second order moments of $\mathbf{x}$, i.e., the expected values $\mathbf{E}[\mathbf{x}]$ and $\mathbf{E}[\mathbf{x}^2]$. The moment dynamics can be obtained using the Dynkin's formula for the above SHS, according to which, for every differentiable function $\psi(\mathbf{x})$ we have that

$$\frac{d\mathbf{E}[\psi(\mathbf{x})]}{dt} = \mathbf{E}\left[\sum_{i=1}^{2} (\psi(\phi_i(\mathbf{x})) - \psi(\mathbf{x}))\, \lambda_i(\mathbf{x})\right] \tag{10}$$

[17], [18]. Taking $\psi(\mathbf{x}) = \mathbf{x}$ and $\psi(\mathbf{x}) = \mathbf{x}^2$ in (10) we obtain the following moment dynamics

$$\frac{d\mathbf{E}[\mathbf{x}]}{dt} = NK - d\mathbf{E}[\mathbf{x}], \tag{11a}$$

$$\frac{d\mathbf{E}[\mathbf{x}^2]}{dt} = K(N^2 + V^2) + d\mathbf{E}[\mathbf{x}] + 2KN\mathbf{E}[\mathbf{x}] - 2d\mathbf{E}[\mathbf{x}^2]. \tag{11b}$$

The corresponding steady-state moments are given by

$$\mathbf{E}^*[\mathbf{x}] = \frac{NK}{d} \tag{12a}$$

$$\mathbf{E}^*[\mathbf{x}^2] = \frac{KdN + 2K^2N^2 + Kd(N^2 + V^2)}{2d^2}. \tag{12b}$$

where $\mathbf{E}^*$ denotes the steady-state value of the respective moment. Replacing the above steady-states in (1) we obtain

$$CV_X^2 = \frac{d(N^2 + V^2 + N)}{2KN^2} = \frac{(N^2 + V^2 + N)}{2\mathbf{E}^*[\mathbf{x}]N}, \tag{13}$$

which quantifies the noise in the protein X due to random gene expression and protein degradation, and is referred to as the *intrinsic noise* in the protein. Note that the noise in the protein increases as the variance V^2 in the number of protein molecules produced per mRNA transcript increases. A special case of (13) is obtained for $N = 1$ and $V = 0$, for which $\mathbf{x}(t)$ has a Poisson distribution and $CV_X^2 = 1/\mathbf{E}^*[\mathbf{x}]$. In the next section we examine what happens to this intrinsic noise when the gene expression rate is not a constant but a function of the number of molecules of the protein.

III. Intrinsic noise in auto-regulatory gene networks

Often the expressed protein binds to the promoter region of its own gene. In doing so it either recruits RNAP to the promoter (which leads to an increase in gene expression) or blocks RNAP from binding to the promoter (which causes a decrease in gene expression). Such gene expression with negative/positive feedbacks is referred to as an *auto-regulatory gene network*. We model this network by assuming that the rate of gene expression is a function $g(\mathbf{x})$ of the number of molecules $\mathbf{x}$ of the protein X. We refer to this function $g(\mathbf{x})$ as the *transcriptional response* of the network and is typically determined empirically from experiments. Monotonic decreasing and increasing functions $g(\mathbf{x})$ denote negative and positive feedback, respectively. However, as we will see later, it is also possible for the function $g(\mathbf{x})$ to be decreasing for some values of $\mathbf{x}$ and increasing for other values.

A. Deterministic model

We first construct a deterministic model of the auto-regulatory gene network. This is done by writing the chemical rate equations which provide a deterministic and continuous approximation $\mathbf{x}_D$ to the number of molecules of the protein X. According to mass-action kinetics, $\mathbf{x}_D$ evolves according to the differential equation

$$\frac{d\mathbf{x}_D}{dt} = Ng(\mathbf{x}_D) - d\mathbf{x}_D \tag{14}$$

and the equilibrium $\mathbf{x}^*$ of the above system satisfies the equation

$$Ng(\mathbf{x}^*) = d\mathbf{x}^*. \tag{15}$$

We assume that this equilibrium is stable with a negative eigenvalue

$$\lambda = Ng'(\mathbf{x}^*) - d < 0 \tag{16}$$

corresponding to the linearization of (14) about $\mathbf{x}^*$. In the sequel we use λ as a measure of the stability of the equilibrium, with more negative values of λ [which correspond to more negative values of $g'(\mathbf{x}^*)$] enhancing the stability of the equilibrium. The eigenvalue λ can also be related to the response time of the gene network T_r, a quantity that can be measured experimentally: Given a linearized system

$$\dot{\delta x} = \lambda \delta x, \quad \lambda < 0, \quad \delta x(0) = \delta_0 \in \mathbb{R} \tag{17}$$

its response time T_r is defined as the time taken for $\delta x(t)$ to decay by 50% of the initial condition, i.e., $\delta x(T_r) = \delta_0/2$ and is given by

$$T_r = -\frac{\ln(2)}{\lambda} > 0, \quad \lambda = Ng'(\mathbf{x}^*) - d < 0. \tag{18}$$

B. Stochastic model

We now consider a stochastic model where the probabilities of a gene expression and protein degradation event happening in the infinitesimal time interval $(t, t+dt]$ are given by

$$\Pr\{\mathbf{x}(t+dt) = x + \mathbf{N} \mid \mathbf{x}(t) = x\} = g(x)dt \tag{19a}$$

$$\Pr\{\mathbf{x}(t+dt) = x - 1 \mid \mathbf{x}(t) = x\} = dxdt. \tag{19b}$$

To write the moment dynamics of $\mathbf{x}$ we first approximate $g(\mathbf{x})$ as a polynomial in $\mathbf{x}$, which is done by expanding $g(\mathbf{x})$ as a Taylor series expansion about $\mathbf{x}^*$:

$$g(\mathbf{x}) = g(\mathbf{x}^*) + g'(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*) + \frac{1}{2}g''(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)^2 + \ldots.$$

C. Linear transcriptional response

We begin by ignoring quadratic and higher order terms in $\mathbf{x} - \mathbf{x}^*$ which results in a transcriptional response

$$g(\mathbf{x}) \approx g(\mathbf{x}^*) + g'(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*) \tag{20}$$

linear in $\mathbf{x}$. This approximation is valid as long as the stochastic fluctuations in the protein are sufficiently small around $\mathbf{x}^*$. As in Section II, we model the time evolution of $\mathbf{x}$ through a Stochastic Hybrid System (SHS) but now the transition intensities are given by

$$\lambda_1(\mathbf{x}) = g(\mathbf{x}^*) + g'(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*), \ \lambda_2(\mathbf{x}) = d\mathbf{x}. \tag{21}$$

Again using the Dynkin's formula for the SHS given by (7), (8) and (21) the time evolution of $\mathbf{E}[\mathbf{x}]$ and $\mathbf{E}[\mathbf{x}^2]$ are given by the following differential equations

$$\frac{d\mathbf{E}[\mathbf{x}]}{dt} = N[g(\mathbf{x}^*) - \mathbf{x}^* g'(\mathbf{x}^*)] - d\mathbf{E}[\mathbf{x}] + g'(\mathbf{x}^*)N\mathbf{E}[\mathbf{x}], \tag{22a}$$

$$\begin{aligned}\frac{d\mathbf{E}[\mathbf{x}^2]}{dt} &= [g(\mathbf{x}^*) - \mathbf{x}^* g'(\mathbf{x}^*)](N^2+V^2) + d\mathbf{E}[\mathbf{x}] \\ &+ 2[g(\mathbf{x}^*) - \mathbf{x}^* g'(\mathbf{x}^*)]N\mathbf{E}[\mathbf{x}] - 2d\mathbf{E}[\mathbf{x}^2] \\ &+ g'(\mathbf{x}^*)(N^2+V^2)\mathbf{E}[\mathbf{x}] + 2g'(\mathbf{x}^*)N\mathbf{E}[\mathbf{x}^2].\end{aligned} \tag{22b}$$

Performing a steady-state analysis of the above equations and using (15) we obtain the following steady-state mean and coefficient of variation

$$\mathbf{E}^*[\mathbf{x}] = \mathbf{x}^*, \quad CV^2_{Xlinear} = \frac{d(N^2+V^2+N)}{2IN^2}, \tag{23}$$

where $I = g(\mathbf{x}^*) - \mathbf{x}^* g'(\mathbf{x}^*)$ is the y-intercept of the tangent to the transcriptional response $g(\mathbf{x})$ at the point $(\mathbf{x}^*, g(\mathbf{x}^*))$ (see Figure 1). As we have assumed that the transcriptional response is linear we obtained a steady-state stochastic mean equal to the equilibrium $\mathbf{x}^*$ of the deterministic chemical rate equations. This will not be true in later sections where $g(\mathbf{x})$ is nonlinear. From (15) and (16) we see that I can be expressed

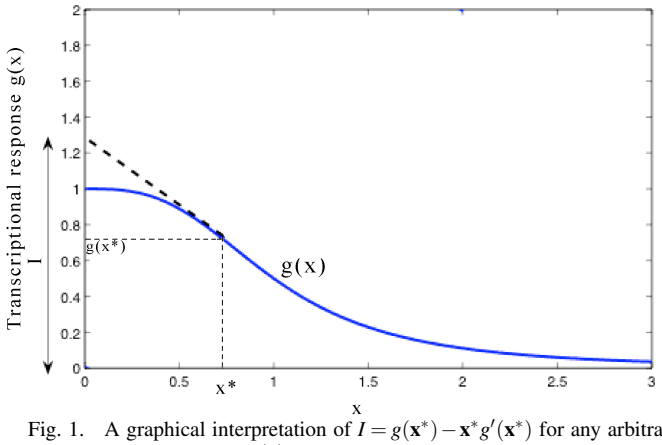

Fig. 1. A graphical interpretation of $I = g(\mathbf{x}^*) - \mathbf{x}^* g'(\mathbf{x}^*)$ for any arbitrary transcriptional response $g(\mathbf{x})$. I is the y-intercept of the tangent to the transcriptional response $g(\mathbf{x})$ at $(\mathbf{x}^*, g(\mathbf{x}^*))$

as

$$I = g(\mathbf{x}^*) - \mathbf{x}^* g'(\mathbf{x}^*) = \mathbf{x}^*[d/N - g'(\mathbf{x}^*)] = -\frac{\lambda \mathbf{x}^*}{N} > 0 \tag{24}$$

which, using (18), allows us to re-write (23) as

$$CV^2_{Xlinear} = -\frac{d(N^2+V^2+N)}{2\mathbf{x}^*\lambda N} = \frac{T_r}{T_p}\frac{N^2+V^2+N}{2\mathbf{x}^* N} \tag{25}$$

where T_r and $T_p = \ln(2)/d$ denote the protein's response time and half-life, respectively. Note that T_p will be the response time when there is no feedback in gene expression (i.e., $g'(\mathbf{x}^*) = 0$ and the transcription rate is a constant as in

Section II). From (18) and (25) one can see that for a fixed $\mathbf{x}^*$, making the slope $g'(\mathbf{x}^*)$ more negative (positive) causes a decrease (increase) in the response time, which attenuates (magnifies) stochastic noise in the protein population.

The above formulas allow one to compare intrinsic noise in protein counts for two different transcriptional responses $g_1(\mathbf{x})$ and $g_2(\mathbf{x})$. When the two different transcriptional responses result in different stochastic means then one can see from the middle expression in (25) that stochastic noise will be lower for the transcriptional response that has the higher value of $-\mathbf{x}^*\lambda$, a quantitiy determined by how large and how stable is the equilibrium population $\mathbf{x}^*$. Figure 2 plots two transcriptional response which lead to different stochastic means but the same intercept $I = -\lambda\mathbf{x}^*/N$, and hence from (23), have the same intrinsic noise in the protein. Note that in this case although $g_1(\mathbf{x})$ gives a larger equilibrium $\mathbf{x}_1^*$, this equilibrium is less stable than that of $g_2(\mathbf{x})$ as $g_2'(\mathbf{x}_2^*) < g'(\mathbf{x}_1^*) = 0$.

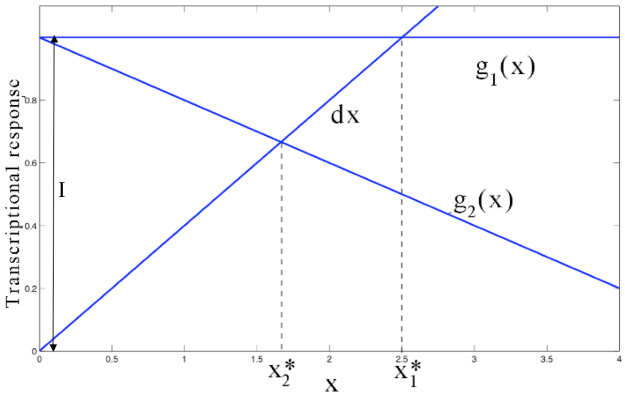

Fig. 2. Two different transcriptional responses $g_1(\mathbf{x})$ and $g_2(\mathbf{x})$ that lead to the same intrinsic noise in the protein. $\mathbf{x}_1^*$ and $\mathbf{x}_2^*$ represents the steady-state average number of protein molecules when the transcriptional response is given by $g_1(\mathbf{x})$ and $g_2(\mathbf{x})$, respectively.

An important feature of equation (25) is that it relates the noise in the protein to parameters that can be experimentally determined. In particular, $N = L/d_r$ where L is the translation rate of the mRNA and d_r is the mRNA degradation rate. As the number of proteins produced per mRNA generally follows a geometric distribution [15], the variance V^2 is equal to $N^2 - N$. Finally, the response times can be measured by tracking the time evolution of the number of molecules within the cell. For example, in [20] an auto-regulatory gene network was designed where the protein TetR represses its own transcription. The protein was fluorescently tagged which allowed one to compute the time evolution of the average number of protein molecules in the cell. Figure 3 plots this time evolution with and without the negative feedback in the gene. The promoter strength was appropriately adjusted such that the steady-state population of the protein was the same in both cases. Figure 3 shows that in the case of the negative feedback it takes about $T_r = .21$ time units for the TetR protein count to reach half of its steady-state protein count $\mathbf{x}^*$. The response time when there is no feedback is $T_p = 1$ time units, which is about five times larger than T_r. This implies from (25) that for this network, the presence of negative feedback reduces stochastic fluctuations in the protein levels by a factor of $\sqrt{5} \approx 2.24$.

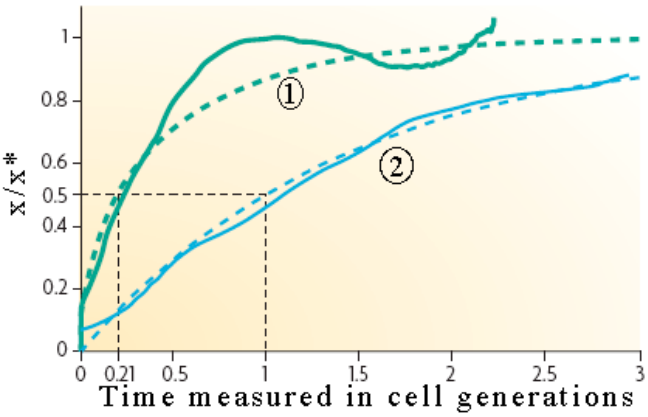

Fig. 3. Time evolution of the average number of TetR protein molecules. 1) represents the situation when these is negative feedback (i.e., protein TetR repressed its own transcription) and 2) represents the situation when these is there is simple gene expression with no negative feedback. Solid and dashed lines represent experimentally measured and fitted approximations to the time evolution of the average number of protein molecules, respectively. This figure was taken from [6].

D. Effect of nonlinearities

In this section we examine the effects of quadratic terms in $g(\mathbf{x})$. Towards that end we now approximate $g(\mathbf{x})$ as

$$g(\mathbf{x}) \approx g(\mathbf{x}^*) + g'(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*) + \frac{1}{2}g''(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)^2 \qquad (26)$$

and ignore cubic and higher order terms in $\mathbf{x} - \mathbf{x}^*$. As before, we can write the moment dynamics of $\mathbf{E}[\mathbf{x}]$, $\mathbf{E}[\mathbf{x}^2]$ corresponding to the transcriptional response given by (26). However, because of the presence of quadratic terms in $g(\mathbf{x})$ the time derivative of $\mathbf{E}[\mathbf{x}]$ and $\mathbf{E}[\mathbf{x}^2]$ now depend on $\mathbf{E}[\mathbf{x}]$, $\mathbf{E}[\mathbf{x}^2]$ and $\mathbf{E}[\mathbf{x}^3]$. More specifically, their time evolution can be written more compactly as

$$\begin{bmatrix} \frac{d\mathbf{E}[\mathbf{x}]}{dt} \\ \frac{d\mathbf{E}[\mathbf{x}^2]}{dt} \end{bmatrix} = \mathbf{a} + \mathbf{A}\begin{bmatrix} \mathbf{E}[\mathbf{x}] \\ \mathbf{E}[\mathbf{x}^2] \end{bmatrix} + \mathbf{B}\ \mathbf{E}[\mathbf{x}^3]. \qquad (27)$$

for an appropriately defined vector $\mathbf{a}$ and matrices $\mathbf{A}$, $\mathbf{B}$. One can see that the above moment equations are not closed in the sense that the time evolution of the lower order moments depends on higher order moments. For analysis purposes, we close the above system by approximating the third order moment $\mathbf{E}[\mathbf{x}^3]$ as a nonlinear function of $\mathbf{E}[\mathbf{x}]$ and $\mathbf{E}[\mathbf{x}^2]$. This procedure is commonly referred to as *moment closure*. We use the recently developed moment closure method in [21] to approximate the higher order moment $\mathbf{E}[\mathbf{x}^3]$ as

$$\mathbf{E}[\mathbf{x}^3] \approx \left(\frac{\mathbf{E}[\mathbf{x}^2]}{\mathbf{E}[\mathbf{x}]}\right)^3 \qquad (28)$$

which gives us the closed moment dynamics

$$\begin{bmatrix} \frac{d\mathbf{E}[\mathbf{x}]}{dt} \\ \frac{d\mathbf{E}[\mathbf{x}^2]}{dt} \end{bmatrix} \approx \mathbf{a} + \mathbf{A}\begin{bmatrix} \mathbf{E}[\mathbf{x}] \\ \mathbf{E}[\mathbf{x}^2] \end{bmatrix} + \mathbf{B}\left(\frac{\mathbf{E}[\mathbf{x}^2]}{\mathbf{E}[\mathbf{x}]}\right)^3. \qquad (29)$$

Our goal now is to compute the steady-state of the above closed system of differential equations. Analytically solving for these steady-state moments from (29) is not an easy task so we use perturbation methods to compute approximate steady-states. This done by writing $\mathbf{E}[\mathbf{x}]$ as a perturbation about $\mathbf{x}^*$ and $\mathbf{E}[\mathbf{x}^2]$ as a perturbation about $\mathbf{E}[\mathbf{x}]^2$, as follows

$$\mathbf{E}[\mathbf{x}] := \mathbf{x}^*(1+\varepsilon_1), \qquad \mathbf{E}[\mathbf{x}^2] := \mathbf{E}[\mathbf{x}]^2(1+\varepsilon_2). \tag{30}$$

Assuming that $|\varepsilon_1| << 1$ and $|\varepsilon_2| << 1$ we have

$$\mathbf{E}[\mathbf{x}^2] \approx \mathbf{x}^{*2}(1+2\varepsilon_1+\varepsilon_2) \tag{31a}$$

$$\left(\frac{\mathbf{E}[\mathbf{x}^2]}{\mathbf{E}[\mathbf{x}]}\right)^3 \approx \mathbf{x}^{*3}(1+3\varepsilon_1+3\varepsilon_2). \tag{31b}$$

Substituting (31) in (29) we obtain a linear system

$$\begin{bmatrix} \frac{d\varepsilon_1}{dt} \\ \frac{d\varepsilon_2}{dt} \end{bmatrix} = \hat{\mathbf{a}} + \hat{\mathbf{A}} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \tag{32}$$

for appropriately defined vector $\hat{\mathbf{a}}$ and matrix $\hat{\mathbf{A}}$. From (30) and (32) the steady-state mean and coefficient of variation are given by

$$\mathbf{E}^*[\mathbf{x}] = \mathbf{x}^*(1+\varepsilon_1^*), \quad CV^2_{Xquad} = \varepsilon_2^*, \tag{33a}$$

$$\varepsilon_1^* = \frac{1}{1+\frac{N\mathbf{x}^* g''(\mathbf{x}^*) CV^2_{Xlinear}}{2\lambda}} - 1, \tag{33b}$$

$$\varepsilon_2^* = \frac{CV^2_{Xlinear}}{1+\frac{N\mathbf{x}^* g''(\mathbf{x}^*) CV^2_{Xlinear}}{2\lambda}}, \tag{33c}$$

where ε_1^* and ε_2^* are the steady-state solutions of (32) and $CV^2_{Xlinear}$ is the noise in the protein when $g(\mathbf{x})$ is linear (as in Section III-C) and given by (25).

The above result shows three important points: Firstly, in this case $\mathbf{E}^*[\mathbf{x}] \neq \mathbf{x}^*$ and a convex (concave) $g(\mathbf{x})$ which corresponds to $g''(\mathbf{x}^*) > 0$ ($g''(\mathbf{x}^*) < 0$) makes the stochastic mean larger (smaller) than the equilibrium $\mathbf{x}^*$ from the deterministic model. Secondly, a transcriptional response which is convex (concave) at $\mathbf{x}^*$ results in larger (smaller) noise in the protein then as predicted by equation (25). Finally, as long as $CV^2_{Xlinear}$ is small enough such that

$$\left|\frac{N\mathbf{x}^* g''(\mathbf{x}^*) CV^2_{Xlinear}}{2\lambda}\right| \ll 1 \tag{34}$$

linearzing the transcriptional response to obtain the stochastic noise in the protein will yield a good approximation for the actual noise in the protein.

IV. Lambda Repressor Gene Network

We now use the results of the previous sections to investigate a well-known gene motif that arises in a gene associated with lambda phage, a virus that infects bacteria. The lambda phage has a gene which encodes for a protein called the lambda repressor that activates its own transcription. Large levels of this protein causes the virus to lysogenize (i.e., integrate its own chromosome into the bacteria DNA). For an auto-regulatory gene network with such positive feedback the transcriptional response is typically given by

$$g_1(\mathbf{x}) = g_0 + \frac{\alpha \mathbf{x}^M}{1+\beta \mathbf{x}^M} \tag{35}$$

where g_0, α, β are positive constants and $M \geq 1$ represents the Hill coefficient [22]. This function is generally sigmoidally shaped and monotonically increasing (see transcriptional response $g_1(\mathbf{x})$ in Figure 4). However, for the lambda repressor gene, the transcriptional response has been modified and the protein activates the gene only when the number of protein molecules is small. At larger protein populations the protein inhibits its own transcription [23]. As a consequence, the transcriptional response of this particular gene network is an increasing function when $\mathbf{x}$ is small and a decreasing function when $\mathbf{x}$ becomes large (see modified transcriptional response $g_2(\mathbf{x})$ in Figure 4).

As can be seen in Figure 4, this modified transcriptional response $g_2(\mathbf{x})$ has a larger intercept I_2 when compared to that of the original transcriptional response $g_1(\mathbf{x})$. Consequently, in view of (23) the modified transcriptional response leads to smaller levels of intrinsic noise in the protein compared to the original transcriptional response. Low stochastic fluctuations in the lambda repressor population ensure that its number do not become small just by random chance, which will cause the virus to come out of lysogeny and lyse the cell. In summary, the transcriptional response $g_2(\mathbf{x})$ allows the virus to have a more robust lysogeny.

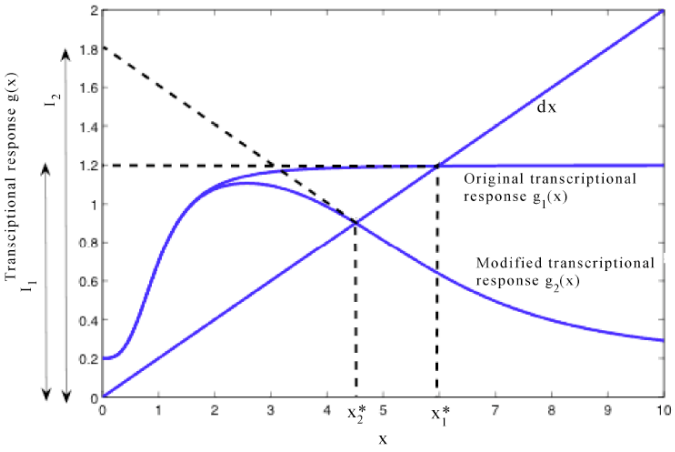

Fig. 4. $g_1(\mathbf{x})$ is the standard transcriptional response of a gene network with positive feedback while $g_2(\mathbf{x})$ is the observed transcriptional response in case of the gene in lambda phage encoding the protein lambda repressor. I_1 and I_2 is the y-intercept of the tangent to the corresponding transcriptional response $g(\mathbf{x})$ at $(\mathbf{x}^*, g(\mathbf{x}^*))$.

V. Conclusion and Future Work

Auto-regulatory gene networks where the protein inhibits/activates its own transcription are common motifs occurring within living cell. These networks are characterized by a transcriptional response that provides information on how the transcription rate of the gene varies as a function of the number of protein molecules present in the cell. We presented results relating the amount of stochastic fluctuations in

protein numbers to the functional form of the transcriptional response. Using a linear approximation for g(**x**), we showed that the noise levels are determined by $I = g(\mathbf{x}^*) - \mathbf{x}^* g'(\mathbf{x}^*)$ which is the y-intercept of the tangent to the transcriptional response at $\mathbf{x} = \mathbf{x}^*$ (as shown in Figure 1), with larger values of I leading to smaller levels of intrinsic noise. We also considered deviations from a linear transcriptional response and showed that concave responses have better noise suppression properties compared to convex responses.

Formulas relating the noise in the protein population to the protein's response time show that one mechanism to achieve low noise in the protein is by having a very small response time. This corresponds to a transcriptional response with $g'(\mathbf{x}^*) \ll 0$ and represents a strong negative feedback in the auto-regulatory gene network. However, this strategy of reducing noise by decreasing the response time will only work if the steady-state average number of protein molecules $\mathbf{x}^*$ is kept moderately large. This is because lowering $\mathbf{x}^*$ will increase the noise in the protein. As a result the basal transcription rate (the transcription rate at $\mathbf{x} = 0$) given by

$$g(\mathbf{x}^*) - g'(\mathbf{x}^*)\mathbf{x}^* \tag{36}$$

will have to be large and corresponds to more energy expenditure by the cell.

In this paper we assumed that the transcription rate is given by $g(\mathbf{x})$ and is only a function of the number of protein molecules. As various enzymes/signaling molecules are involved in the process of transcription, a more general form for the transcription rate would be $g(\mathbf{x}, \mathbf{z})$ where $\mathbf{z}$ represents a noisy exogenous signal. Fluctuations in $\mathbf{z}$ are often referred to as the extrinsic noise entering the gene network. We are currently investigating the effects of this extrinsic noise on the protein noise level and under what conditions the auto-regulatory gene network can attenuate or amplify this extrinsic noise.

REFERENCES

[1] M. C. Walters, S. Fiering, J. Eidemiller, W. Magis, M. Groudine, and D. I. K. Martin, "Enhancers increase the probability but not the level of gene expression," *Proceedings of the National Academy of Sciences U.S.A*, vol. 92, pp. 7125–7129, 1995.

[2] A. Arkin, J. Ross, and H. H. McAdams, "Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *Escherichia coli* cells," *Genetics*, vol. 149, pp. 1633–1648, 1998.

[3] W. J. Blake, M. Krn, C. R. Cantor, and J. J. Collins, "Noise in eukaryotic gene expression," *Nature*, vol. 422, pp. 633–637, 2003.

[4] J. L. Spudich and D. E. K. Jr, "Non-genetic individuality: chance in the single cell," *Nature (London)*, vol. 262, pp. 467–471, 1976.

[5] H. H. McAdams and A. P. Arkin, "Stochastic mechanisms in gene expression," *Proceedings of the National Academy of Sciences U.S.A*, vol. 94, pp. 814–819, 1997.

[6] U. Alon, "Network motifs: theory and experimental approaches," *Nature Reviews Genetics*, vol. 8, pp. 450–461, 2007.

[7] M. A. Savageau, "Comparison of classical and autogenous systems of regulation in inducible operons," *Nature*, vol. 252, pp. 546–549, 1974.

[8] D. Orrell and H. Bolouri, "Control of internal and external noise in genetic regulatory networks," *J. of Theoretical Biology*, vol. 230, pp. 301–312, 2004.

[9] Y. Tao, X. Zheng, and Y. Sun, "Effect of feedback regulation on stochastic gene expression," *J. of Theoretical Biology*, vol. 247, pp. 827–836, 2007.

[10] A. Becskei and L. Serrano, "Engineering stability in gene networks by autoregulation," *Nature*, vol. 405, pp. 590–593, 2000.

[11] R. Tomioka, H. Kimura, T. J. Kobayashi, and K. Aihara, "Multivariate analysis of noise in genetic regulatory networks," *J. of Theoretical Biology*, vol. 229, pp. 501–521, 2004.

[12] M. L. Simpson, C. D. Cox, and G. S. Sayler, "Frequency domain analysis of noise in autoregulated gene circuits," *PNAS*, vol. 100, pp. 4551–4556, 2003.

[13] J. Hasty, J. Pradines, M. Dolnik, and J. J. Collins, "Noise-based switches and amplifiers for gene expression," *PNAS*, vol. 97, pp. 2075–2080, 2000.

[14] O. Brandman, J. E. Ferrell, R. Li, and T. Meyer, "Interlinked fast and slow positive feedback loops drive reliable cell decisions," *Science*, vol. 310, pp. 496 – 498, 2005.

[15] D. Longo and J. Hasty, "Imaging gene expression: Tiny signals make a big noise," *Nature Chemical Biology*, vol. 2, pp. 181–182, 2006.

[16] J. Paulsson, "Model of stochastic gene expression," *Physics of Life Reviews*, vol. 2, pp. 157–175, 2005.

[17] M. H. A. Davis, *Markov models and Optimization*. Chapman and Hall, 1993.

[18] J. P. Hespanha, "Stochastic hybrid systems: Applications to communication networks," in *Hybrid Systems: Computation and Control*, ser. Lect. Notes in Comput. Science, R. Alur and G. J. Pappas, Eds. Berlin: Springer-Verlag, Mar. 2004, no. 2993, pp. 387–401.

[19] J. P. Hespanha and A. Singh, "Stochastic models for chemically reacting systems using polynomial stochastic hybrid systems," *Int. J. of Robust and Nonlinear Control*, vol. 15, pp. 669–689, 2005.

[20] N. Rosenfeld, M. B. Elowitz, and U. Alon., "Negative autoregulation speeds the response times of transcription networks." *J. Molecular Biology*, vol. 323, pp. 785–793, 2002.

[21] A. Singh and J. P. Hespanha, "Lognormal moment closures for biochemical reactions." in *Proc. of the 45th Conf. on Decision and Control, San Diego*, 2006.

[22] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman and Hall, 2006.

[23] M. Ptashne, *Genetic Switch: Phage Lambda Revisited*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2004.