

Stochastic Analysis of Gene Regulatory Networks Using Moment Closure

Abhyudai Singh and João Pedro Hespanha

Abstract—Random fluctuations in gene regulatory networks are inevitable due to the probabilistic nature of chemical reactions and the small populations of proteins, mRNAs present inside cells. These fluctuations are usually reported in terms of the first and second order statistical moments of the protein populations. If the birth-death rates of the mRNAs or the proteins are nonlinear, then the dynamics of these moments generally do not form a closed system of differential equations, in the sense that their time-derivatives depends on moments of order higher than two. Recent work has developed techniques to obtain the two lowest-order moments by closing their dynamics, which involves approximating the higher order moments as nonlinear functions of the two lowest ones. This paper uses these moment closure techniques to quantify noise in several gene regulatory networks.

In gene expression mechanisms in which a protein inhibits its own transcription, the resulting negative feedback reduces stochastic variations in the protein populations. Often the protein itself is not active and combines with itself to form an active multimer, which then inhibits the transcription. We demonstrate that this more sophisticated form of negative feedback (using multimerization) is more effective in suppressing noise. We also consider a two-gene cascade activation network in which the protein expressed by one gene activates another gene to express a second protein. Analysis shows that the stochastic fluctuations in the population of the activated protein increases with the degree of multimerization in the activating protein.

I. INTRODUCTION

Gene expression and regulation is inherently a noisy process. The origins of this stochasticity lies in the probabilistic nature of chemical reactions and small populations of reactants living inside cells, which can lead to large statistical fluctuations in molecule numbers. Recent work [1], [2] has provided considerable experimental evidence for these stochastic fluctuations and may account for the large amounts of cell to cell variation observed in genetically identical cells exposed to the same environment conditions [3], [4]. Although often considered deleterious for the cell, this stochasticity can also be advantageous (for example, by creating a heterogeneous population of cells that can adapt better in varying environments) and building models that capture it is becoming increasingly important.

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

To quantify stochastic fluctuations one uses the stochastic formulation of chemical kinetics, which treats the different reaction involved in a biochemical process as probabilistic events. The time evolution of the system, is then described by a single equation for a probability function, where time and species populations appear as independent variables, called the Master Equation [5], [6]. However, this equation can only be solved for relatively few, highly idealized cases and generally Monte Carlo simulation techniques are used to study stochasticity in bio-chemical reactions [4], [7], [8]. Using them, various authors have investigated stochastic gene expression from a single gene and groups of genes in which protein production is activated or halted by particular signals [9], [10]. Such groups of genes are known as *gene regulatory networks*. The stochastic fluctuation in the protein population is usually reported in terms of the ratio of its standard deviation and its mean, also known as the coefficient of variation. Much effort can be saved by applying approximate methods to produce these low-order statistical moments directly, without actually having to solve for the probability density function or running Monte Carlo simulations. Various such approximate methods have been applied, for example, expanding the Master equation [11]. In this paper we use *moment closure techniques* to estimate these low-order statistical moments and quantify stochastic fluctuations in different gene regulatory networks.

In Section II we consider a very simple model for gene expression where the mRNA is transcribed from the

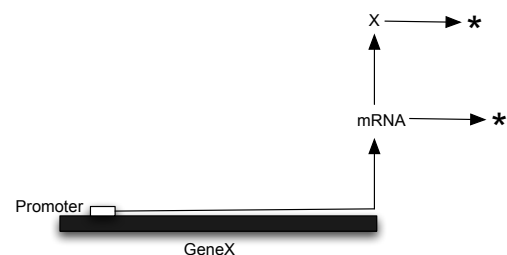


Fig. 1. A simple model for gene expression.

gene *GeneX* at a rate K_r and the protein X is translated from the mRNA at a rate K_p as shown in Figure 1. Both mRNA and the protein decay at constant rates d_r and d_p , respectively. As the ratio of the lifetimes of the protein and its mRNA is typically small we assume $d_p/d_r \ll 1$. To derive differential equations for the time evolution of moments of the populations, we model the gene expression process as a Stochastic Hybrid System (SHS) with state

$\mathbf{x} = [\mathbf{x}_r, \mathbf{x}_p]^T$, where \mathbf{x}_r and \mathbf{x}_p denote the number of molecules of the mRNA and the protein, respectively. In order to fit the framework of our problem, this SHS has trivial continuous dynamics $\dot{\mathbf{x}} = 0$, four reset maps representing births and deaths of the mRNA and the protein, with transitional intensities defined by their birth-death rates. The time evolution of the moments of \mathbf{x} is then obtained using results from the SHS literature [12]. A steady state analysis of these moment equations in Section II, gives the steady-state coefficient of variation for the protein population, which we denote by CV_X as

$$CV_X = \sqrt{\frac{1 + \frac{K_p}{d_r}}{\mathbf{E}[\mathbf{x}_p(\infty)]}}, \quad \mathbf{E}[\mathbf{x}_p(\infty)] = \frac{K_p K_r}{d_p d_r} \quad (1)$$

where $\mathbf{E}[\mathbf{x}_p(\infty)]$ denotes the steady-state mean protein population. We refer to this CV_X as the *inherent stochasticity in GeneX*.

In Section III we investigate negative feedback in gene expression caused by the protein X inhibiting its own transcription rate. This is done by simply taking the transcription rate of *GeneX* as $K_r = K_r^{max} - k_1 \mathbf{x}_p$ for some constants K_r^{max} and k_1 . Analysis of the corresponding moment equations show that CV_X becomes lower than the one in (1) and decreases with increasing k_1 for fixed $\mathbf{E}[\mathbf{x}_p(\infty)]$. Often the protein X itself is inactive but combines with itself, i.e. multimerizes, to form an active multimer which then inhibits the transcription rate of X . Assuming fast multimerizations, the population of the multimer is assumed to be proportional to \mathbf{x}_p^N where N denotes the degree of multimerization in X . The transcriptional rate then becomes $K_r = K_r^{max} - k_1 \mathbf{x}_p^N$. Due to the nonlinear transcription rate, the time derivative of the first and second order moments of \mathbf{x} is not closed, in the sense that, it now depends on moments of order higher than two. Using the recent developed techniques of moment closure [13], [14], [15], the moment dynamics is closed by approximating these higher order moments as nonlinear functions of the two lowest ones. Steady state analysis of the closed moment equations show that larger values of N lead to lower CV_X . Hence we conclude that this sophisticated form of negative feedback (using multimerizations) is more effective in suppressing noise in the protein.

A two-gene cascade activation network, as shown in Figure 2 is analyzed in Section IV, where a multimer of protein X activates another gene *GeneY* to make a second protein Y . This is done by assuming that the transcription rate of *GeneY* is given by $A \mathbf{x}_p^N$ where A is an activation constant. If there are no multimerizations, i.e. $N = 1$, we show that in such a network the steady-state coefficient of variation of protein Y is given by

$$CV_Y = \sqrt{\frac{1 + \frac{K_{\bar{p}}}{d_{\bar{r}}}}{\mathbf{E}[\mathbf{y}_p(\infty)]} + \frac{CV_X^2 d_{\bar{p}}}{d_{\bar{p}} + d_p}}, \quad \mathbf{E}[\mathbf{y}_p(\infty)] = \frac{AK_{\bar{p}}K_p K_r}{d_p d_r d_{\bar{p}} d_{\bar{r}}}, \quad (2)$$

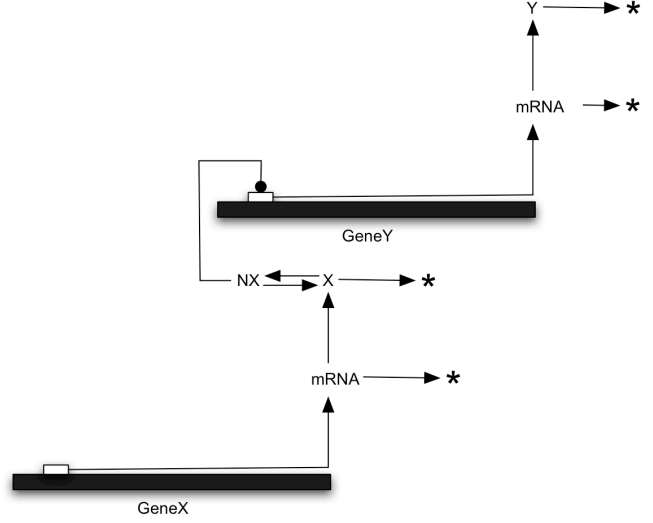


Fig. 2. A gene network where a gene *GeneX* makes the protein X . This protein or its multimer then activates gene *GeneY* to make protein Y

where $K_{\bar{p}}$, $d_{\bar{r}}$ and $d_{\bar{p}}$ denote the translation rate, mRNA and protein death rate for *GeneY*, respectively. Also we denote by \mathbf{y}_p and $\mathbf{E}[\mathbf{y}_p(\infty)]$ the number of molecules of protein Y and its steady state mean. For $\mathbf{E}[\mathbf{y}_p(\infty)] \approx \mathbf{E}[\mathbf{x}_p(\infty)]$ and other parameters being approximately the same for both genes, we have $CV_Y = 1.23 CV_X$. Note from the above formula for CV_Y that when $d_{\bar{p}} \ll d_p$ then the contribution in stochasticity from *GeneX*, $CV_X^2 d_{\bar{p}} / (d_{\bar{p}} + d_p)$ is negligible and CV_Y is just the inherent stochasticity in *GeneY*. We demonstrate using moment closure techniques that multimerizations in the activating protein X , i.e. $N \geq 2$, lead to a larger values of CV_Y than in (2) for fixed $\mathbf{E}[\mathbf{y}_p(\infty)]$

II. MODELING OF GENE EXPRESSION

A very simple model for gene expression is shown in Figure 1, where the mRNA is transcribed from the gene *GeneX* at a constant rate K_r and the protein X is translated from the mRNA at a constant rate K_p . Both mRNA and the protein decay at rates d_r and d_p respectively. Thus K_p/d_r denotes the number of proteins produced per mRNA. As both the processes of transcription and translation happen concurrently, this model will correspond to gene expression in a Prokaryotic cell. We denote by \mathbf{x}_p and \mathbf{x}_r the number of molecules of protein X and its mRNA, respectively. Continuous approximations \mathbf{x}_{rD} and \mathbf{x}_{pD} for \mathbf{x}_r and \mathbf{x}_p are obtained from the chemical rate equations

$$\dot{\mathbf{x}}_{rD} = K_r - d_r \mathbf{x}_{rD}, \quad \dot{\mathbf{x}}_{pD} = K_p \mathbf{x}_{rD} - d_p \mathbf{x}_{pD}$$

leading to a steady-state population of

$$\mathbf{x}_{rD}(\infty) = \frac{K_r}{d_r}, \quad \mathbf{x}_{pD}(\infty) = \frac{K_r K_p}{d_r d_p}. \quad (3)$$

As this model does not provide any information about the stochastic fluctuation in the protein, we turn to a stochastic formulation which treats births and deaths of the mRNA and

the protein as probabilistic events. Thus, assuming $\mathbf{x}_p(t) = x_p$ and $\mathbf{x}_r(t) = x_r$, the probability of any of the four reactions corresponding to births and deaths of the mRNA and the protein happening in the infinitesimal time interval $(t, t + dt]$ is given as

$$\Pr\{\mathbf{x}_p(t + dt) = x_p, \mathbf{x}_r(t + dt) = x_r + 1\} = K_r dt \quad (4a)$$

$$\Pr\{\mathbf{x}_p(t + dt) = x_p, \mathbf{x}_r(t + dt) = x_r - 1\} = d_r x_r dt \quad (4b)$$

$$\Pr\{\mathbf{x}_p(t + dt) = x_p + 1, \mathbf{x}_r(t + dt) = x_r\} = K_p x_r dt \quad (4c)$$

$$\Pr\{\mathbf{x}_p(t + dt) = x_p - 1, \mathbf{x}_r(t + dt) = x_r\} = d_p x_p dt. \quad (4d)$$

A convenient way to model the time evolution of the number of molecules \mathbf{x}_p and \mathbf{x}_r is through a Stochastic Hybrid System (SHS), the state of which is $\mathbf{x} = [\mathbf{x}_r, \mathbf{x}_p]^T$. This special class of SHS was introduced in [13] and to fit the framework of the problem, is characterized by trivial dynamics $\dot{\mathbf{x}} = 0$, a family of *reset maps* $\phi_i(\mathbf{x})$ and a corresponding family of *transitional intensities* $\lambda_i(\mathbf{x})$. For the stochastic gene expression model we have

$$\mathbf{x} \mapsto \phi_1(\mathbf{x}) = \begin{bmatrix} \mathbf{x}_r + 1 \\ \mathbf{x}_p \end{bmatrix}, \quad \mathbf{x} \mapsto \phi_2(\mathbf{x}) = \begin{bmatrix} \mathbf{x}_r - 1 \\ \mathbf{x}_p \end{bmatrix} \quad (5a)$$

$$\mathbf{x} \mapsto \phi_3(\mathbf{x}) = \begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_p + 1 \end{bmatrix}, \quad \mathbf{x} \mapsto \phi_4(\mathbf{x}) = \begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_p - 1 \end{bmatrix} \quad (5b)$$

with corresponding transition intensities given by

$$\lambda_1(\mathbf{x}) = K_r, \quad \lambda_2(\mathbf{x}) = d_r x_r, \quad \lambda_3(\mathbf{x}) = K_p x_r, \quad \lambda_4(\mathbf{x}) = d_p x_p. \quad (6)$$

In essence, if no reaction takes place, the state remains constant and whenever any of the reactions happens, the corresponding reset map $\phi_i(\mathbf{x})$ is “activated” and the state \mathbf{x} is reset, furthermore, the probability of the activation taking place in an “infinitesimal” time interval $(t, t + dt]$ is $\lambda_i(\mathbf{x})dt$.

In order to gauge the noise level in the protein population, we determine the time evolution of the first and second order moments of $\mathbf{x}(t)$. Towards that end we define for $n \in \mathbb{N}$ and $m \in \mathbb{N}$ the *un-centered* moment of $\mathbf{x}(t)$ as

$$\mu^{(n,m)}(t) := \mathbf{E}[\mathbf{x}_r^n(t) \mathbf{x}_p^m(t)] \quad (7)$$

where \mathbf{E} stands for the expected value. The time evolution of moments is then obtained from the following Theorem which is a straightforward application of Theorem 1 in [12] to the above SHS.

Theorem 1: Consider a SHS with trivial dynamics $\dot{\mathbf{x}} = 0$ and K reset maps $\phi_i(\mathbf{x})$ and K corresponding transition intensities $\lambda_i(\mathbf{x})$. Then for every differentiable function $\psi(\mathbf{x})$ we have that

$$\frac{d\mathbf{E}[\psi(\mathbf{x})]}{dt} = \mathbf{E}[(\mathbf{L}\psi)(\mathbf{x})], \quad (8)$$

where

$$(\mathbf{L}\psi)(\mathbf{x}) := \sum_{i=1}^K (\psi(\phi_i(\mathbf{x})) - \psi(\mathbf{x})) \lambda_i(\mathbf{x}). \quad (9)$$

Taking $\psi(\mathbf{x}) = \mathbf{x}_r^n(t) \mathbf{x}_p^m(t)$ in the above Theorem and using (5), (6), (7) one obtains

$$\dot{\mu}^{(1,0)} = K_r - d_r \mu^{(1,0)}, \quad \dot{\mu}^{(0,1)} = K_p \mu^{(1,0)} - d_p \mu^{(0,1)}$$

$$\dot{\mu}^{(2,0)} = K_r + d_r \mu^{(1,0)} + 2K_r \mu^{(1,0)} - 2d_r \mu^{(2,0)}$$

$$\dot{\mu}^{(0,2)} = K_p \mu^{(1,0)} + d_p \mu^{(0,1)} + 2K_p \mu^{(1,1)} - 2d_p \mu^{(0,2)}$$

$$\dot{\mu}^{(1,1)} = K_p \mu^{(2,0)} + K_r \mu^{(0,1)} - d_p \mu^{(1,1)} - d_r \mu^{(1,1)}.$$

The steady-state moments are obtained as

$$\mathbf{E}[\mathbf{x}_r(\infty)] = \frac{K_r}{d_r}, \quad \mathbf{E}[\mathbf{x}_p(\infty)] = \frac{K_p K_r}{d_p d_r}, \quad \mathbf{E}[\mathbf{x}_r^2(\infty)] = \frac{d_r K_r + K_r^2}{d_r^2} \quad (10a)$$

$$\mathbf{E}[\mathbf{x}_p^2(\infty)] = \frac{K_p K_r}{d_p d_r} + \frac{K_p (d_p d_r K_p K_r + d_p K_p K_r^2 + d_r K_p K_r^2)}{d_p^2 d_r^2 (d_p + d_r)} \quad (10b)$$

$$\mathbf{E}[\mathbf{x}_r(\infty) \mathbf{x}_p(\infty)] = \frac{d_p d_r K_p K_r + d_p K_p K_r^2 + d_r K_p K_r^2}{d_p d_r^2 (d_p + d_r)}. \quad (10c)$$

In this paper, we use the *coefficient of variation* defined by

$$CV_X := \frac{\sqrt{\mathbf{E}[\mathbf{x}_p^2] - (\mathbf{E}[\mathbf{x}_p])^2}}{\mathbf{E}[\mathbf{x}_p]} \quad (11)$$

to quantify noise strength in the protein population. Replacing the above steady-states in (11) we get

$$CV_X = \sqrt{\frac{1 + \frac{K_p}{d_r}}{\frac{K_p K_r}{d_p d_r} + 1}} \approx \sqrt{\frac{1 + \frac{K_p}{d_r}}{\mathbf{E}[\mathbf{x}_p(\infty)]}}. \quad (12)$$

Here $d_p/d_r \ll 1$ is the ratio of the lifetimes of the protein and the mRNA and is typically a very small number. In the sequel we refer to

$$\sqrt{\frac{1 + \frac{K_p}{d_r}}{\mathbf{E}[\mathbf{x}_p(\infty)]}}$$

as the *inherent stochasticity in GeneX*. Note from (3) and (10) that here

$$\mathbf{x}_{pD}(\infty) = \mathbf{E}[\mathbf{x}_p(\infty)]. \quad (13)$$

This is generally not true and holds only when birth and death rates are linear, as in (6).

III. GENE EXPRESSION WITH NEGATIVE FEEDBACK

In this section we consider negative feedback in gene expression caused by the protein inhibiting its own transcription. Such auto-regulatory networks are common means of stabilizing protein levels in biochemical pathways. Another sophisticated form of negative feedback is when multiple copies of a protein combines with itself to form a multimer and then instead of the protein this new multimer inhibits the transcription [4]. As the dynamics of multimerization is very fast, the population of the multimer rapidly comes in

equilibrium with the protein population and one can assume that the number of molecules of the multimer is proportional to \mathbf{x}_p^N where N denotes the degree of multimerization. The negative feedback is incorporated in the model by taking the transcription rate to be

$$K_r = K_r^{\max} - k_1 \mathbf{x}_p^N \quad (14)$$

for some constants K_r^{\max} and k_1 . The chemical rate equations corresponding to this model are then given by

$$\dot{\mathbf{x}}_{rD} = K_r^{\max} - k_1 \mathbf{x}_{pD}^N - d_r \mathbf{x}_{rD}, \quad \dot{\mathbf{x}}_{pD} = K_p \mathbf{x}_{rD} - d_p \mathbf{x}_{pD}. \quad (15)$$

A. No multimerization

We first analyze the case $N = 1$ which corresponds to no multimerization in the protein. The corresponding SHS for this model is exactly similar to (5)-(6) but with $\lambda_1(\mathbf{x}) = K_r^{\max} - k_1 \mathbf{x}_p$. Again using Theorem 1 the following moment dynamics are obtained

$$\begin{aligned} \dot{\mu}^{(1,0)} &= K_r^{\max} - d_r \mu^{(1,0)} - k_1 \mu^{(0,1)} \\ \dot{\mu}^{(0,1)} &= K_p \mu^{(1,0)} - d_p \mu^{(0,1)} \\ \dot{\mu}^{(2,0)} &= K_r^{\max} + d_r \mu^{(1,0)} + 2K_r^{\max} \mu^{(1,0)} - 2d_r \mu^{(2,0)} - k_1 \mu^{(0,1)} \\ &\quad - k_1 \mu^{(1,1)} \\ \dot{\mu}^{(0,2)} &= K_p \mu^{(1,0)} + d_p \mu^{(0,1)} + 2K_p \mu^{(1,1)} - 2d_p \mu^{(0,2)} \\ \dot{\mu}^{(1,1)} &= K_p \mu^{(2,0)} + K_r^{\max} \mu^{(0,1)} - d_p \mu^{(1,1)} - d_r \mu^{(1,1)} - k_1 \mu^{(0,2)} \end{aligned}$$

which using $T = d_p/d_r \ll 1$ leads to the following first and second order steady-state moments for the protein population

$$\begin{aligned} \mathbf{E}[\mathbf{x}_p(\infty)] &= \frac{K_p K_r^{\max}}{k_1 K_p + d_r d_p} \\ \mathbf{E}[\mathbf{x}_p^2(\infty)] &= \frac{K_p^2 (K_r^{\max})^2 + d_r K_p K_r^{\max} d_p + d_p K_p^2 K_r^{\max} + k_1 K_p^2 K_r^{\max} T}{(k_1 K_p + d_r d_p)^2}. \end{aligned}$$

As linearity in the birth and death rates is still maintained we have $\mathbf{x}_{pD}(\infty) = \mathbf{E}[\mathbf{x}_p(\infty)]$. For comparison purpose, we now vary parameters k_1 and K_r^{\max} so as to keep $\mathbf{x}_{pD}(\infty)$ fixed and see its effect on the coefficient of variation. Hence for a given K_r^{\max} , k_1 is always chosen such that

$$k_1 = \frac{K_r^{\max}}{\mathbf{x}_{pD}(\infty)} - \frac{d_r d_p}{K_p} \quad (16)$$

for some fixed $\mathbf{x}_{pD}(\infty)$. Thus larger values of $K_r^{\max} > d_r d_p \mathbf{x}_{pD}(\infty)/K_p$ correspond to more negative feedback. Using the above steady states one can show that

$$CV_X = \sqrt{\frac{1 + \frac{K_p}{d_r} + \frac{k_1 K_p}{d_r^2}}{\frac{K_r^{\max} K_p}{d_r d_p}}}.$$

This reduces to

$$CV_X = \sqrt{\frac{1 + \frac{K_p}{d_r}}{\frac{K_r^{\max} K_p}{d_r d_p}}}. \quad (17)$$

if $k_1 \ll d_r$. As CV_X is a decreasing function of K_r^{\max} , we conclude that negative feedback suppresses the stochastic variation in the protein.

B. Effects of Multimerization

We now investigate the effects of multimerization by taking $\lambda_1(\mathbf{x}) = K_r^{\max} - k_1 \mathbf{x}_p^N$ where N is an integer larger than 2. The dynamics of the vector μ , defined as

$$\mu = [\mu^{(1,0)}, \mu^{(0,1)}, \mu^{(2,0)}, \mu^{(0,2)}, \mu^{(1,1)}]^T,$$

can be written as

$$\dot{\mu} = \bar{\mathbf{A}} \mu + \mathbf{A} \mu + \mathbf{B} \bar{\mu}. \quad (18)$$

where

$$\bar{\mu} = [\mu^{(0,N)}, \mu^{(1,N)}, \mu^{(0,N+1)}]^T$$

for some matrices $\bar{\mathbf{A}}$, \mathbf{A} , \mathbf{B} . One can see that the above moment equations are not closed in the sense that the time evolution of the vector μ depends on $\bar{\mu}$ which is a vector containing higher order moments of \mathbf{x} . For analysis purposes, we close the above system by approximating each element of $\bar{\mu}$ as a nonlinear function $\varphi(\mu)$. This procedure is commonly referred to as *moment closure*. Referring the reader to [14] for further details we take the following approximations for elements of $\bar{\mu}$

$$\mu^{(0,N)} = \frac{[\mu^{(0,2)}]^{N(N-1)/2}}{[\mu^{(0,1)}]^{N^2-2N}}, \quad \mu^{(0,N+1)} = \frac{[\mu^{(0,2)}]^{N(N+1)/2}}{[\mu^{(0,1)}]^{N^2-1}} \quad (19a)$$

$$\mu^{(1,N)} = \frac{[\mu^{(0,2)}]^{N(N-1)/2}}{[\mu^{(1,0)}]^{N-1}} \left(\frac{[\mu^{(1,1)}]^N}{[\mu^{(0,1)}]^{N^2-N}} \right). \quad (19b)$$

As analytical solution of the closed moment equations are too complicated to be of any use we investigate them numerically by taking

$$K_p = 5 \text{ sec}^{-1}, \quad d_r = .1 \text{ sec}^{-1}, \quad d_p = .001 \text{ sec}^{-1}. \quad (20)$$

We now fix $\mathbf{x}_{pD}(\infty) = 500$. This implies from (15) that for a given K_r^{\max} and N , k_1 is chosen as

$$k_1 500^N = K_r^{\max} - .01 \quad (21)$$

The steady-state moments are then obtained from the approximate model (18)-(19) using *Mathematica* for varying K_r^{\max} . Table I lists CV_X obtained for different values of K_r^{\max} . CV_X values corresponding to $N = 1$ and given by (17) are also listed along with CV_X values obtained for $N = 2$ and 3. The case $K_r^{\max} = .01$ corresponds to $k_1 = 0$ and hence no negative feedback. As one can see, negative feedback through the multimer gives lower values of CV_X , and is a more effective mechanism to suppress noise in the protein.

IV. GENE ACTIVATION BY ANOTHER GENE

In this section we study a gene network where a gene *GeneX* makes the protein *X*. This protein *or* its multimer then activates gene *GeneY* to make protein *Y* as shown in Figure 2. This activation is incorporated in the model by taking the transcription rate of *GeneY* to be $A \mathbf{x}_p^N$ where A denotes a activation constant. We denote by \mathbf{y}_p , \mathbf{y}_r the number of molecules and $d_{\bar{y}}$, $d_{\bar{r}}$ the decay rates of the protein *Y* and its mRNA, respectively. The translation rate for *GeneY*

TABLE I

CV_X FOR DIFFERENT VALUES OF K_r^{max} AND N . PARAMETERS WERE TAKEN AS (20) WITH k_1 ALWAYS CHOSEN AS IN (21) TO KEEP $\mathbf{x}_{pD}(\infty) = 500$.

K_r^{max}	$CV_X(N=1)$	$CV_X(N=2)$	$CV_X(N=3)$
.01	.317	.317	.317
.02	.225	.182	.156
.05	.142	.105	.087
.1	.100	.070	.060

is $K_{\bar{p}}$. The deterministic chemical rate equations are then given by

$$\dot{\mathbf{x}}_{rD} = K_r - d_r \mathbf{x}_{rD}, \quad \dot{\mathbf{x}}_{pD} = K_p \mathbf{x}_{rD} - d_p \mathbf{x}_{pD} \quad (22a)$$

$$\dot{\mathbf{y}}_{rD} = A \mathbf{x}_{pD}^N - d_{\bar{r}} \mathbf{y}_{rD}, \quad \dot{\mathbf{y}}_{pD} = K_{\bar{p}} \mathbf{y}_{rD} - d_{\bar{p}} \mathbf{y}_{pD}. \quad (22b)$$

A. No multimerization

Assuming no multimerization, we set the transcription rate of gene *GeneY* to be $A \mathbf{x}_p$. The corresponding SHS with state $\mathbf{x} := [\mathbf{x}_r, \mathbf{x}_p, \mathbf{y}_r, \mathbf{y}_p]^T$ now has with 8 reset maps and transitional intensities corresponding to the birth-death of the two proteins ($\mathbf{x}_p, \mathbf{y}_p$) and two RNA's ($\mathbf{x}_r, \mathbf{y}_r$). Using it and Theorem 1 one can show that the time evolution of the vector

$$\boldsymbol{\mu} = [\boldsymbol{\mu}^{(1,0,0,0)}, \boldsymbol{\mu}^{(0,1,0,0)}, \boldsymbol{\mu}^{(0,0,1,0)}, \boldsymbol{\mu}^{(0,0,0,1)}, \boldsymbol{\mu}^{(2,0,0,0)}, \boldsymbol{\mu}^{(0,2,0,0)}, \boldsymbol{\mu}^{(0,0,2,0)}, \boldsymbol{\mu}^{(0,0,0,2)}, \boldsymbol{\mu}^{(1,1,0,0)}, \boldsymbol{\mu}^{(1,0,1,0)}, \boldsymbol{\mu}^{(1,0,0,1)}, \boldsymbol{\mu}^{(0,1,1,0)}, \boldsymbol{\mu}^{(0,1,0,1)}, \boldsymbol{\mu}^{(0,0,1,1)}]^T \quad (23)$$

is obtained as

$$\dot{\boldsymbol{\mu}} = \bar{\mathbf{A}} + \mathbf{A} \boldsymbol{\mu}$$

for some matrices $\bar{\mathbf{A}}$, \mathbf{A} and each moment in vector $\boldsymbol{\mu}$ is defined as

$$\boldsymbol{\mu}^{(n,m,q,s)}(t) := \mathbf{E}[\mathbf{x}_r^n(t) \mathbf{x}_p^m(t) \mathbf{y}_r^q(t) \mathbf{y}_p^s(t)].$$

Assuming the ratio of the life times of the protein and the mRNA is small for both the genes, the following steady states are calculated using *Mathematica*

$$\mathbf{E}[\mathbf{x}_p(\infty)] = \frac{K_p K_r}{d_p d_r}, \quad CV_X = \sqrt{\frac{1 + \frac{K_p}{d_r}}{\mathbf{E}[\mathbf{x}_p(\infty)]}} \quad (24a)$$

$$\mathbf{E}[\mathbf{y}_p(\infty)] = \frac{A K_{\bar{p}} K_p K_r}{d_p d_r d_{\bar{p}} d_{\bar{r}}}, \quad CV_Y = \sqrt{\frac{1 + \frac{K_{\bar{p}}}{d_{\bar{r}}}}{\mathbf{E}[\mathbf{y}_p(\infty)]} + \frac{CV_X^2 d_{\bar{p}}}{d_{\bar{p}} + d_p}}. \quad (24b)$$

Here CV_Y denotes the steady state coefficient of variation for protein *Y*. Note that

$$\sqrt{\frac{1 + \frac{K_{\bar{p}}}{d_{\bar{r}}}}{\mathbf{E}[\mathbf{y}_p(\infty)]}}$$

is the inherent stochasticity in *GeneY*. The other term in (24b) is the contribution from *GeneX* which will be negligible if $d_{\bar{p}} \ll d_p$. For $\mathbf{E}[\mathbf{y}_p(\infty)] \approx \mathbf{E}[\mathbf{x}_p(\infty)]$ and other parameters being approximately the same for both genes, we have $CV_Y = 1.23 CV_X$.

B. Effects of Multimerization

In this section we assume that the protein *X* forms a multimer which then activates *GeneY*. This corresponds to taking the transcription rate of *GeneY* as $A \mathbf{x}_p^N$. The dynamics of vector $\boldsymbol{\mu}$ defined as in (23) is now given by

$$\dot{\boldsymbol{\mu}} = \bar{\mathbf{A}} + \mathbf{A} \boldsymbol{\mu} + \mathbf{B} \bar{\boldsymbol{\mu}}.$$

for some appropriate matrices $\bar{\mathbf{A}}$, \mathbf{A} , \mathbf{B} and

$$\bar{\boldsymbol{\mu}} = [\boldsymbol{\mu}^{(0,N,0,0)}, \boldsymbol{\mu}^{(0,N+1,0,0)}, \boldsymbol{\mu}^{(1,N,0,0)}, \boldsymbol{\mu}^{(0,N,1,0)}, \boldsymbol{\mu}^{(0,N,0,1)}]^T.$$

These moment dynamics are closed using the following moment closure schemes

$$\begin{aligned} \boldsymbol{\mu}^{(0,N,0,0)} &= \frac{[\boldsymbol{\mu}^{(0,2,0,0)}]^{N(N-1)/2}}{[\boldsymbol{\mu}^{(0,1,0,0)}]^{N^2-2N}} \\ \boldsymbol{\mu}^{(0,N+1,0,0)} &= \frac{[\boldsymbol{\mu}^{(0,2,0,0)}]^{N(N+1)/2}}{[\boldsymbol{\mu}^{(0,1,0,0)}]^{N^2-1}} \\ \boldsymbol{\mu}^{(1,N,0,0)} &= \frac{[\boldsymbol{\mu}^{(0,2,0,0)}]^{N(N-1)/2}}{[\boldsymbol{\mu}^{(1,0,0,0)}]^{N-1}} \left(\frac{[\boldsymbol{\mu}^{(1,1,0,0)}]^{N}}{[\boldsymbol{\mu}^{(0,1,0,0)}]^{N^2-N}} \right) \\ \boldsymbol{\mu}^{(0,N,1,0)} &= \frac{[\boldsymbol{\mu}^{(0,2,0,0)}]^{N(N-1)/2}}{[\boldsymbol{\mu}^{(0,0,1,0)}]^{N-1}} \left(\frac{[\boldsymbol{\mu}^{(0,1,1,0)}]^{N}}{[\boldsymbol{\mu}^{(0,1,0,0)}]^{N^2-N}} \right) \\ \boldsymbol{\mu}^{(0,N,0,1)} &= \frac{[\boldsymbol{\mu}^{(0,2,0,0)}]^{N(N-1)/2}}{[\boldsymbol{\mu}^{(0,0,0,1)}]^{N-1}} \left(\frac{[\boldsymbol{\mu}^{(0,1,0,1)}]^{N}}{[\boldsymbol{\mu}^{(0,1,0,0)}]^{N^2-N}} \right). \end{aligned}$$

Its easy to show from the above moment dynamics that

$$\begin{aligned} \mathbf{E}[\mathbf{x}_p(\infty)] &= \mathbf{x}_{pD}(\infty) = \frac{K_p K_r}{d_p d_r} \\ \mathbf{y}_{pD}(\infty) &= \frac{A K_{\bar{p}} \mathbf{x}_{pD}^N(\infty)}{d_{\bar{p}} d_{\bar{r}}} \\ \mathbf{E}[\mathbf{y}_p(\infty)] &= \frac{A K_{\bar{p}} \mathbf{E}[\mathbf{x}_p^N(\infty)]}{d_{\bar{p}} d_{\bar{r}}} > \mathbf{y}_{pD}(\infty). \end{aligned}$$

Hence on an *average* the steady-state population of the protein *Y* is much larger than that predicted from the deterministic rate equations. This has been referred to in literature as stochastic focusing.

To investigate the effects of multimerization on CV_Y we take parameters for both genes as in (20), $K_r = .01$ we fix $\mathbf{x}_{pD}(\infty) = \mathbf{y}_{pD}(\infty) = 500$. This implies from (22) that $A = .01/500^N$. Table II lists $\mathbf{E}[\mathbf{y}_p(\infty)]$ and CV_Y for $N = 1, 2$ and 3. From the table we conclude that multimerization in the activating protein is an effective mechanism to create large stochastic variations in the activated protein. From these numerical results we conjecture that CV_Y increases quadratically with N and is approximately given by

$$CV_Y \approx \sqrt{\frac{1 + \frac{K_{\bar{p}}}{d_{\bar{r}}}}{\mathbf{E}[\mathbf{y}_p(\infty)]} + N^2 \frac{CV_X^2 d_{\bar{p}}}{d_{\bar{p}} + d_p}}, \quad (25)$$

however further work is needed to show it analytically. Compared with (24b), note that the contribution from *GeneX* is now multiplied by N^2 .

TABLE II

$E[y_p(\infty)]$ AND CV_Y FOR DIFFERENT VALUES OF N . PARAMETERS FOR BOTH GENES ARE TAKEN AS IN (20), $K_r = .01$ AND THE CONSTANT A WAS CHOSEN SO AS TO GIVE $y_{pD}(\infty) = 500$.

N	$E[y_p(\infty)]$	CV_Y
1	500	.389
2	550.5	.562
3	667	.743

V. CONCLUSION AND FUTURE WORK

Stochastic fluctuations in protein populations were quantified for expression from a single gene with and without negative feedback and a two-gene cascade activation network. This was done by modeling the populations of mRNAs and proteins in the gene network as the continuous state of a stochastic hybrid system. This enabled us to obtain the time evolution of their first and second order statistical moments. If the birth and death rates of the mRNAs and the proteins were linear, their steady-states statistical moments could be obtained directly from the moment dynamics. However for nonlinear rates the derivatives of these moments dependent on higher order moments and to perform a steady state analysis moment closure techniques were used to approximate their time derivative as functions of themselves.

From this analysis we concluded that negative feedback in gene expression, where a protein inhibits its own transcription, can decrease stochastic fluctuations in the protein. Indirect negative feedback through multimerization of the protein was found to be more effective in suppressing this noise. In a two-gene cascade activation network where the protein X expressed by one gene activates another gene, multimerizations in X increased the stochastic fluctuations in the activated protein Y .

As part of our current work we are developing techniques which will allow us to obtain approximate formulae for the steady-state coefficient of variation in terms of the parameters of the reactions. These involve linearizing the moment closure functions about $CV^2 = 0$ and then finding $CV^2(\infty)$ from the steady-state analysis of the moment equations. Another line of future work would be look at gene expression and regulation where the mRNA transcribed from the gene is not immediately accessible for translation. This would corresponds to gene expression in Eukaryotic cells where the mRNA is transported from the nucleus to the cytoplasm where translation takes place. This can be incorporated in our model by introducing another specie that corresponds to the inactive mRNA and is being converted into an active mRNA at some constant rate.

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

[4] 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.

[5] D. A. McQuarrie, "Stochastic approach to chemical kinetics," *J. of Applied Probability*, vol. 4, pp. 413–478, 1967.

[6] N. G. V. Kampen, *Stochastic Processes in Physics and Chemistry*. Amsterdam: North Holland, 1992.

[7] D. T. Gillespie, "Approximate accelerated stochastic simulation of chemically reacting systems," *J. of Chemical Physics*, vol. 115, no. 4, pp. 1716–1733, 2001.

[8] M. A. Gibson and J. Bruck, "Efficient exact stochastic simulation of chemical systems with many species and many channels," *J. of Physical Chemistry A*, vol. 104, pp. 1876–1889, 2000.

[9] T. B. Kepler and T. C. Elston, "Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations," *Biophysical Journal*, vol. 81, no. 6, pp. 3116–3136, 2001.

[10] M. Thattai and A. van Oudenaarden, "Intrinsic noise in gene regulatory networks," *Proceedings of the National Academy of Sciences U.S.A.*, vol. 98, no. 15, pp. 8614–8619, 2001.

[11] J. Paulsson, "Summing up the noise in gene networks," *Nature*, vol. 427, pp. 415–418, 2004.

[12] 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.

[13] 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.

[14] 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.

[15] —, "A derivative matching approach to moment closure for the stochastic logistic model," 2007, to appear in *Bull. of Math Biology*.