

Scaling of stochasticity in gene cascades

Abhyudai Singh and João Pedro Hespanha

Abstract—Stochastic fluctuations in protein levels are inevitable due to the probabilistic nature of gene expression. These fluctuations are attributed to two factors: intrinsic noise (noise associated with random transcription and translation events) and extrinsic noise (noise associated with fluctuations in the reaction rates due to correspondent fluctuations in cellular enzyme levels). We present results on how stochastic noise in proteins scales in a cascade of genes both in the presence and absence of extrinsic noise. In particular, we derive analytical formulas relating the noise in the proteins to the length, per stage magnification of the cascade and the amount of extrinsic noise present in the cell.

We show that when there is no extrinsic noise the gene cascade acts like a *noise attenuator* where downstream proteins exhibit reduced noise. Moreover, for two different cascades with the same average final protein level, the cascade with the *larger* number of stages will have *lower* stochastic noise in the final protein. We also show that adding extrinsic noise can change the qualitative behavior of the cascade from a noise attenuator to a *noise magnifier* where downstream protein have increased noise. Furthermore, there exists a critical level of extrinsic noise above which the cascade with *lower* number of stages will have *lesser* noise in the final protein.

I. INTRODUCTION

Gene expression is inherently a noisy process. The cause for this stochasticity lies in the probabilistic nature of the biochemical processes (such as transcription and translation) which make up gene expression and fluctuations in the cellular enzyme levels that carry out these processes. Recent work [1], [2], [3], [4] 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 environmental conditions [5], [6]. Gene cascades, where a protein expressed from one gene activates another gene to make a different protein, are common motifs occurring with in cells. In these cascades an initial signal (a protein with small number of molecules) can be amplified over a number of stages. The amplified signal (a protein with large number of molecules) can then be used to trigger some physiological response in the cell. We investigate how noise levels in the proteins of the gene cascade are effected by the length and per stage magnification of the cascade.

We first quantify stochastic noise in a single gene *GeneX* which expresses a protein *X*. We consider a simple model

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

of gene expression where the mRNA is transcribed from the gene *GeneX* at a rate T_x and the protein *X* is translated from the mRNA at a rate L_x . Both mRNA and the protein decay at constant rates a_x and d_x , respectively. In the stochastic formulation of this gene expression model, transcription, translation and degradation are treated as probabilistic events and $\mathbf{x}(t)$, the protein count at time t is a stochastic process. Details on the stochastic formulation are presented in Section II. We quantify the noise in $\mathbf{x}(t)$ by its *coefficient of variation* defined by

$$CV_X^2 := \frac{\mathbf{E}[\mathbf{x}^2(\infty)] - \mathbf{E}[\mathbf{x}(\infty)]^2}{\mathbf{E}[\mathbf{x}(\infty)]^2} \quad (1)$$

where $\mathbf{E}[\mathbf{x}^k(\infty)]$ denotes the steady-state value of the moment $\mathbf{E}[\mathbf{x}^k]$, $k \in [1, 2]$. We show in Section II that for this constitutively expressed gene the noise in the protein numbers is given by

$$CV_{intX}^2 = \frac{1 + \frac{L_x}{a_x}}{\mathbf{E}[\mathbf{x}(\infty)]}, \quad \mathbf{E}[\mathbf{x}(\infty)] = \frac{L_x T_x}{d_x a_x}. \quad (2)$$

The above expression for CV_{intX} is the noise associated with random transcription and translation events occurring in gene expression and we refer to it as the *intrinsic stochasticity* of the gene *GeneX*.

We next consider a scenario where the gene *GeneX* is not constitutively expressed but activated by a protein *Z*. We model this activation by assuming that the transcription rate of gene *GeneX* is $\alpha \mathbf{z}(t)$ where α is a constant and $\mathbf{z}(t)$ is the number of molecules of protein *Z* at time t . We further assume that protein *Z* is expressed from gene *GeneZ* whose intrinsic stochasticity is denoted by CV_{intZ} . For this two-gene cascade, we show in Section III that the steady-state average counts of the two proteins are related by

$$\mathbf{E}[\mathbf{x}(\infty)] = \frac{\alpha L_x}{d_x} \mathbf{E}[\mathbf{z}(\infty)] \quad (3)$$

and the noise in the protein *X* is given by

$$CV_X^2 = CV_{intX}^2 + \frac{d_x}{d_x + d_z} CV_{intZ}^2 \quad (4)$$

where $\mathbf{E}[\mathbf{z}(\infty)]$ is the steady-state average molecular count of protein *Z* and d_z is the degradation rate of protein *Z*. Equation (4) shows that the noise in protein *X* is the sum of the intrinsic stochasticity in *GeneX* plus the contribution from the activating protein *Z*.

In Section IV we generalize this to a cascade of N genes *GeneX*₁, *GeneX*₂, ..., *GeneX* _{N} where gene *GeneX* _{i} expresses protein X_i that activates gene *GeneX* _{$i+1$} , for all $i \in [1, \dots, N-1]$ (as shown in Figure 1). As before we model the activation

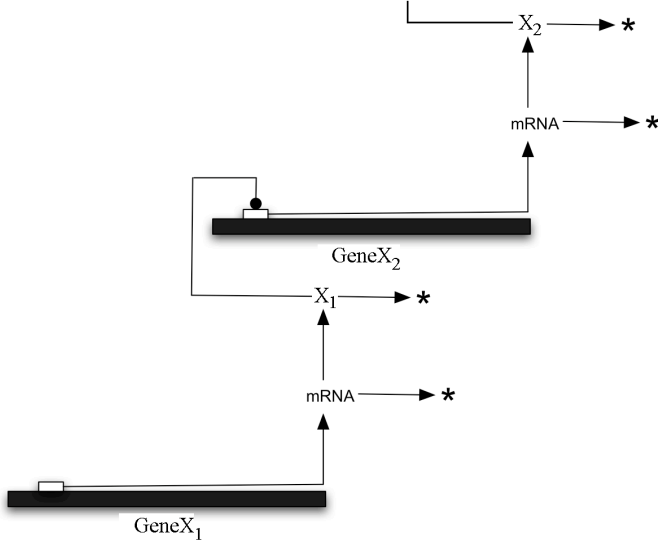


Fig. 1. A gene activation cascade where gene $GeneX_1$ expresses protein X_1 . This protein then activates gene $GeneX_2$ to make X_2 which then goes on to activate gene $GeneX_3$.

by assuming that the transcription rate of $GeneX_{i+1}$ is αx_i for some constant α and x_i denotes the number of molecules of protein X_i . For this cascade we have that

$$\mathbf{E}[x_{i+1}(\infty)] = M_{i+1} \mathbf{E}[x_i(\infty)], \quad \forall i \in [1, \dots, N-1] \quad (5)$$

where $\mathbf{E}[x_i(\infty)]$ denotes the steady-state mean number of molecules of protein X_i , M_{i+1} is the amount by which the population of protein X_i is magnified at the $i+1^{th}$ stage of the cascade and is assumed to be larger than one. For this network, the noise in protein X_{i+1} is a function of the noise in protein X_i according to

$$CV_{x_{i+1}}^2 = CV_{m_{i+1}}^2 + \frac{d_{i+1}}{d_i + d_{i+1}} CV_{x_i}^2 \quad (6)$$

where $CV_{m_{i+1}}$ represents the intrinsic stochasticity of the gene $GeneX_{i+1}$ and d_i is the degradation rate of protein X_i . From (6) we conclude that here CV_{x_i} monotonically decreases to zero as the number of stages i increases. Thus in this case the cascade acts like a *noise attenuator* where downstream proteins have reduced noise. We also compare the final protein noise in two different cascades of genes with different sizes but with the same average steady-state number of molecules of the final protein X_N . We show that the gene cascade that achieves the same total magnification in more number of stages but smaller magnification per stage has lesser stochastic fluctuations in the final protein compared to a cascade with lesser number of stages and a higher magnification per stage.

In Section V we investigate the same cascade of N genes, but we now consider the presence of extrinsic noise. We show that the presence of extrinsic noise can change the qualitative behaviour of the cascade. More specifically, for sufficiently large extrinsic noise, CV_{x_i} increases unboundedly with i , and hence, the cascade behaves as a *noise magnifier* instead of a noise attenuator. We also show that there exists

a critical level of extrinsic noise above which the cascade with the lower number of stages and higher magnification per stage has lesser noise in the final protein than a cascade with higher number of stages and lower magnification per stage.

II. SIMPLE GENE EXPRESSION AND INTRINSIC NOISE

A very simple model for gene expression is shown in Figure 2. This process is initiated when cellular enzymes known as *RNA polymerase* (RNAP) recognize and bind to specialized DNA sequences known as *promoters* present at the start of a gene. The RNAP then moves along the gene and synthesizes a *messenger RNA* (mRNA) which carries the information encoded by the gene. This process is known as

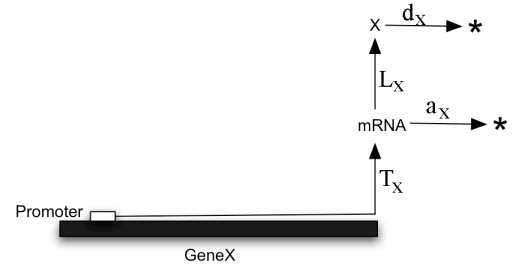


Fig. 2. A simple model for gene expression.

transcription. Ribosome's (small cellular components) attach to the nascent mRNA and start the *translation* process which consists of making a protein by assembling amino acids in an order that corresponds to the mRNA sequence. We assume that the mRNA is transcribed from the gene $GeneX$ at a constant rate T_x and the protein X is translated from the mRNA at a constant rate L_x . Both mRNA and the protein decay at rates a_x and d_x respectively. The lifetime of the mRNA is typically much smaller than that of the protein and throughout the paper we use the approximation $d_x/a_x \ll 1$. As the average lifetime of a mRNA is $1/a_x$ and proteins are made from it at rate L_x , $B_x = L_x/a_x$ denotes the number of proteins produced per mRNA, which is referred to as the *burst size* of the gene $GeneX$. Here and in the sequel we denote by \mathbf{m}_x and \mathbf{x} , the number of molecules of the mRNA and protein X , respectively. As a continuous deterministic model based on chemical rate equations does not provide information about the stochastic fluctuation in the protein, we consider a stochastic formulation that treats births and deaths of the mRNA and the protein as probabilistic events. Given that $\mathbf{x}(t) = x$ and $\mathbf{m}_x(t) = m_x$, the probabilities of the four reactions corresponding to births and deaths of the mRNA and the protein happening in the infinitesimal time interval $(t, t+dt]$ are given by

$$\Pr\{\mathbf{x}(t+dt) = x, \mathbf{m}_x(t+dt) = m_x + 1\} = T_x dt \quad (7a)$$

$$\Pr\{\mathbf{x}(t+dt) = x, \mathbf{m}_x(t+dt) = m_x - 1\} = a_x m_x dt \quad (7b)$$

$$\Pr\{\mathbf{x}(t+dt) = x + 1, \mathbf{m}_x(t+dt) = m_x\} = L_x m_x dt \quad (7c)$$

$$\Pr\{\mathbf{x}(t+dt) = x - 1, \mathbf{m}_x(t+dt) = m_x\} = d_x x dt. \quad (7d)$$

As shown in [7], [8], a convenient way to model the time evolution of the number of molecules \mathbf{x} and \mathbf{m}_x is through

a Stochastic Hybrid System (SHS), the state of which is $\mathbf{y} = [\mathbf{m}_x, \mathbf{x}]^T$. This SHS is characterized by trivial continuous dynamics

$$\dot{\mathbf{y}} = 0, \quad (8)$$

four reset maps $\phi_i(\mathbf{y})$

$$\mathbf{y} \mapsto \phi_1(\mathbf{y}) = \begin{bmatrix} \mathbf{m}_x + 1 \\ \mathbf{x} \end{bmatrix}, \quad \mathbf{y} \mapsto \phi_2(\mathbf{y}) = \begin{bmatrix} \mathbf{m}_x - 1 \\ \mathbf{x} \end{bmatrix} \quad (9a)$$

$$\mathbf{y} \mapsto \phi_3(\mathbf{y}) = \begin{bmatrix} \mathbf{m}_x \\ \mathbf{x} + 1 \end{bmatrix}, \quad \mathbf{y} \mapsto \phi_4(\mathbf{y}) = \begin{bmatrix} \mathbf{m}_x \\ \mathbf{x} - 1 \end{bmatrix} \quad (9b)$$

and corresponding transition intensities $\lambda_i(\mathbf{y})$

$$\lambda_1(\mathbf{y}) = T_x, \quad \lambda_2(\mathbf{y}) = a_x \mathbf{m}_x, \quad \lambda_3(\mathbf{y}) = L_x \mathbf{m}_x, \quad \lambda_4(\mathbf{y}) = d_x \mathbf{x} \quad (10)$$

corresponding to transcription, translation, mRNA and protein degradation. 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{y} , i.e. the expected values $\mathbf{E}[\mathbf{m}_x]$, $\mathbf{E}[\mathbf{x}]$, $\mathbf{E}[\mathbf{x}^2]$, $\mathbf{E}[\mathbf{m}_x^2]$ and $\mathbf{E}[\mathbf{m}_x \mathbf{x}]$. Using the Dynkin's formula for the SHS (8)-(10) we have

$$\frac{d\mathbf{E}[\mathbf{m}_x]}{dt} = T_x - a_x \mathbf{E}[\mathbf{m}_x], \quad \frac{d\mathbf{E}[\mathbf{x}]}{dt} = L_x \mathbf{E}[\mathbf{m}_x] - d_x \mathbf{E}[\mathbf{x}] \quad (11a)$$

$$\frac{d\mathbf{E}[\mathbf{m}_x^2]}{dt} = T_x + a_x \mathbf{E}[\mathbf{m}_x] + 2T_x \mathbf{E}[\mathbf{m}_x] - 2a_x \mathbf{E}[\mathbf{m}_x^2] \quad (11b)$$

$$\frac{d\mathbf{E}[\mathbf{x}^2]}{dt} = L_x \mathbf{E}[\mathbf{m}_x] + d_x \mathbf{E}[\mathbf{x}] + 2L_x \mathbf{E}[\mathbf{m}_x \mathbf{x}] - 2d_x \mathbf{E}[\mathbf{x}^2] \quad (11c)$$

$$\frac{d\mathbf{E}[\mathbf{m}_x \mathbf{x}]}{dt} = L_x \mathbf{E}[\mathbf{m}_x^2] + T_x \mathbf{E}[\mathbf{x}] - d_x \mathbf{E}[\mathbf{m}_x \mathbf{x}] - a_x \mathbf{E}[\mathbf{m}_x \mathbf{x}] \quad (11d)$$

[9], [10]. The steady-state moments are then given by

$$\mathbf{E}[\mathbf{m}_x(\infty)] = \frac{T_x}{a_x}, \quad \mathbf{E}[\mathbf{x}(\infty)] = \frac{L_x T_x}{d_x a_x} = \frac{B_x T_x}{d_x} \quad (12a)$$

$$\mathbf{E}[\mathbf{m}_x^2(\infty)] = \frac{a_x T_x + T_x^2}{a_x^2} \quad (12b)$$

$$\mathbf{E}[\mathbf{x}^2(\infty)] = \frac{L_x T_x}{d_x a_x} + \frac{L_x (d_x a_x L_x T_x + d_x L_x T_x^2 + a_x L_x T_x^2)}{d_x^2 a_x^2 (d_x + a_x)} \quad (12c)$$

$$\mathbf{E}[\mathbf{m}_x(\infty) \mathbf{x}(\infty)] = \frac{d_x a_x L_x T_x + d_x L_x T_x^2 + a_x L_x T_x^2}{d_x a_x^2 (d_x + a_x)} \quad (12d)$$

where $\mathbf{E}[\mathbf{m}_x^i(\infty) \mathbf{x}^j(\infty)]$ denotes the steady-state value of the moment $\mathbf{E}[\mathbf{m}_x^i \mathbf{x}^j]$. Replacing the above steady-states in (1) and using $d_x/a_x \ll 1$ we conclude that

$$CV_{intX}^2 = \frac{1 + B_x}{\mathbf{E}[\mathbf{x}(\infty)]} = \frac{1 + B_x}{\frac{L_x T_x}{d_x a_x}}, \quad B_x = \frac{L_x}{a_x}. \quad (13)$$

We refer to CV_{intX} as the *intrinsic stochasticity* of the gene *GeneX*, which represents the noise in a protein which arises due to the random transcription and translation events occurring in gene expression.

III. INCORPORATING FLUCTUATIONS IN ACTIVATING PROTEIN

In the previous section the protein *X* was constitutively expressed from gene *GeneX* at a constant rate. We now consider the situation where *GeneX* is not constitutively expressed but activated by a different protein *Z*. This activation is modeled by assuming that the transcription rate of gene *GeneX* is given by $\alpha \mathbf{z}(t)$ where $\mathbf{z}(t)$ represents the number of molecules of protein *Z* at time t and α is a constant. Our goal is to evaluate the noise in the protein *X* as a function of the noise in the protein *Z*.

We assume that the protein *Z* is expressed from a gene *GeneZ* at a constant rate T_z and its corresponding mRNA is translated at a constant rate L_z . Both the protein *Z* and its corresponding mRNA decay at rates d_z and a_z , respectively. Then from (13) we conclude that the noise in the protein *Z* is given by

$$CV_{intZ}^2 = \frac{1 + B_z}{\frac{T_z L_z}{d_z a_z}}, \quad B_z = \frac{L_z}{a_z}. \quad (14)$$

In order to see the effect of noise in the protein *Z* on the noise in the protein *X*, we augment the state of the SHS in the previous section with $\mathbf{m}_z(t)$ and $\mathbf{z}(t)$, where $\mathbf{m}_z(t)$ represents the number of molecules of the mRNA transcribed from gene *GeneZ*. The state of the modified SHS is now given by

$$\dot{\mathbf{y}} = 0, \quad \mathbf{y} = [\mathbf{m}_z, \mathbf{z}, \mathbf{m}_x, \mathbf{x}] \quad (15)$$

with eight reset maps

$$\mathbf{y} \mapsto \phi_1(\mathbf{y}) = \begin{bmatrix} \mathbf{m}_z + 1 \\ \mathbf{z} \\ \mathbf{m}_x \\ \mathbf{x} \end{bmatrix}, \quad \mathbf{y} \mapsto \phi_2(\mathbf{y}) = \begin{bmatrix} \mathbf{m}_z - 1 \\ \mathbf{z} \\ \mathbf{m}_x \\ \mathbf{x} \end{bmatrix}, \quad (16a)$$

$$\mathbf{y} \mapsto \phi_3(\mathbf{y}) = \begin{bmatrix} \mathbf{m}_z \\ \mathbf{z} + 1 \\ \mathbf{m}_x \\ \mathbf{x} \end{bmatrix}, \quad \mathbf{y} \mapsto \phi_4(\mathbf{y}) = \begin{bmatrix} \mathbf{m}_z \\ \mathbf{z} - 1 \\ \mathbf{m}_x \\ \mathbf{x} \end{bmatrix}, \quad (16b)$$

$$\mathbf{y} \mapsto \phi_5(\mathbf{y}) = \begin{bmatrix} \mathbf{m}_z \\ \mathbf{z} \\ \mathbf{m}_x + 1 \\ \mathbf{x} \end{bmatrix}, \quad \mathbf{y} \mapsto \phi_6(\mathbf{y}) = \begin{bmatrix} \mathbf{m}_z \\ \mathbf{z} \\ \mathbf{m}_x - 1 \\ \mathbf{x} \end{bmatrix}, \quad (16c)$$

$$\mathbf{y} \mapsto \phi_7(\mathbf{y}) = \begin{bmatrix} \mathbf{m}_z \\ \mathbf{z} \\ \mathbf{m}_x \\ \mathbf{x} + 1 \end{bmatrix}, \quad \mathbf{y} \mapsto \phi_8(\mathbf{y}) = \begin{bmatrix} \mathbf{m}_z \\ \mathbf{z} \\ \mathbf{m}_x \\ \mathbf{x} - 1 \end{bmatrix} \quad (16d)$$

and corresponding transition intensities

$$\lambda_1(\mathbf{y}) = T_z, \quad \lambda_2(\mathbf{y}) = a_z \mathbf{m}_z, \quad \lambda_3(\mathbf{y}) = L_z \mathbf{m}_z, \quad \lambda_4(\mathbf{y}) = d_z \mathbf{z} \quad (17a)$$

$$\lambda_5(\mathbf{y}) = \alpha \mathbf{z}, \quad \lambda_6(\mathbf{y}) = a_x \mathbf{m}_x, \quad \lambda_7(\mathbf{y}) = L_x \mathbf{m}_x, \quad \lambda_8(\mathbf{y}) = d_x \mathbf{x}. \quad (17b)$$

Again using the Dynkin's formula for the SHS (15)-(17) one can write the time evolution of all the first and second

order moments of \mathbf{y} . In particular, if we construct a vector $\boldsymbol{\mu}$ containing all the first and second order moments of \mathbf{y} its time derivative is given by

$$\dot{\boldsymbol{\mu}} = \mathbf{v} + A\boldsymbol{\mu} \quad (18)$$

for some vector \mathbf{v} and matrix A . Solving for the steady-state moments from (18) we obtain that the steady-state means and noise in the protein X are given by

$$\mathbf{E}[\mathbf{z}(\infty)] = \frac{T_z B_z}{d_z}, \quad \mathbf{E}[\mathbf{x}(\infty)] = \frac{\alpha B_x}{d_x} \mathbf{E}[\mathbf{z}(\infty)], \quad (19a)$$

$$CV_X^2 = \frac{1+B_x}{\mathbf{E}[\mathbf{x}(\infty)]} + \frac{d_x}{d_x+d_z} CV_{intZ}^2. \quad (19b)$$

Using (13), the above expression for the noise in the protein X can be re-written as

$$CV_X^2 = CV_{intX}^2 + \frac{d_x}{d_x+d_z} CV_{intZ}^2. \quad (20)$$

The above equation shows that the noise in the protein X is given as the sum of the intrinsic stochasticity in gene *GeneX* plus the contribution in noise from the activating protein Z which is given by

$$\frac{d_x}{d_x+d_z} CV_{intZ}^2. \quad (21)$$

IV. SCALING OF INTRINSIC NOISE OVER NETWORKS

In this section we extend the network in the previous section to a cascade of N genes *GeneX*₁, *GeneX*₂, ..., *GeneX* _{N} where gene *GeneX* _{i} for $i \in [1, \dots, N-1]$ expresses protein X_i . The protein X_i then activates gene *GeneX* _{$i+1$} to express protein X_{i+1} (see Figure 1). As in the previous section we model the activation by assuming that the transcription rate of gene *GeneX* _{$i+1$} is $\alpha \mathbf{x}_i$ where α is a constant and \mathbf{x}_i is the number of molecules of protein X_i . In the sequel we denote the burst size of *GeneX* _{i} and degradation rate of protein X_i by B_i and d_i respectively. This implies from (19) that

$$\mathbf{E}[\mathbf{x}_{i+1}(\infty)] = \frac{\alpha B_{i+1}}{d_{i+1}} \mathbf{E}[\mathbf{x}_i(\infty)] =: M_{i+1} \mathbf{E}[\mathbf{x}_i(\infty)], \quad (22)$$

for all $i \in [1, \dots, N-1]$ where the constant $M_{i+1} = \alpha B_{i+1}/d_{i+1}$. We consider a cascade where the the input signal (i.e. $\mathbf{E}[\mathbf{x}_i(\infty)]$) is magnified at each stage, and hence, assume $M_{i+1} > 1$. Using (20) we express the noise in the protein X_{i+1} as a function of noise in the protein X_i as

$$CV_{X_{i+1}}^2 = CV_{intX_{i+1}}^2 + \frac{d_{i+1}}{d_i+d_{i+1}} CV_{X_i}^2 \quad (23)$$

where $CV_{intX_{i+1}}$ represents the intrinsic stochasticity of the gene *GeneX* _{$i+1$} . Substituting (13) in (23) we have

$$CV_{X_{i+1}}^2 = \frac{1+B_{i+1}}{\mathbf{E}[\mathbf{x}_{i+1}(\infty)]} + \frac{d_{i+1}}{d_i+d_{i+1}} CV_{X_i}^2. \quad (24)$$

Assuming that for all i , B_i is bounded from above and $d_{i+1}/(d_i+d_{i+1})$ is strictly less than one, we have

$$\lim_{i \rightarrow \infty} CV_{X_i}^2 = 0. \quad (25)$$

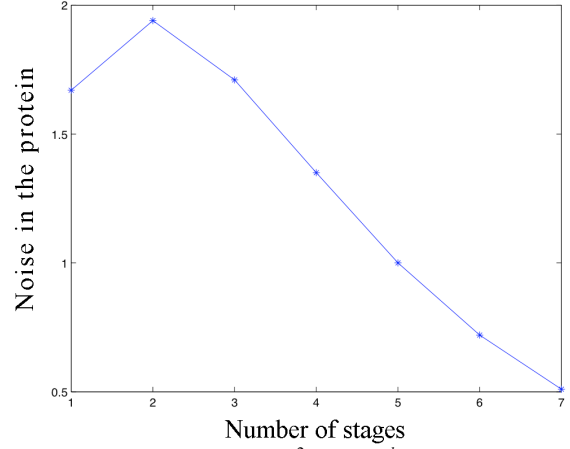


Fig. 3. Noise in the protein X_i ($CV_{X_i}^2$) at the i^{th} stage of the gene cascade when all genes have per stage magnification of $M = 1.5$, $\alpha = d = \gamma = 1$.

Thus in this case the cascade acts like a noise attenuator where as we go down the cascade, the noise in the downstream proteins becomes smaller.

For a homogeneous cascade of genes with the transcription rate of the first gene given by $\gamma\alpha$ for some constant γ and

$$B_i = B, \quad d_i = d, \quad M_i = M = \frac{\alpha B}{d} > 1, \quad \forall i \quad (26)$$

we conclude from (22) that

$$\mathbf{E}[\mathbf{x}_i(\infty)] = \gamma M^i \quad (27)$$

and from (24) that

$$\begin{aligned} CV_{X_{i+1}}^2 &= \frac{1+B}{\gamma M^{i+1}} + \frac{1}{2} CV_{X_i}^2 \\ &= \frac{(1+B)}{M\gamma} \frac{1}{2^i} \sum_{j=0}^i \left(\frac{2}{M}\right)^j = \frac{(1+B)((M/2)^{i+1} - 1)}{(M/2 - 1)M^{i+1}\gamma}. \end{aligned} \quad (28)$$

Analysis of this expression shows that if the per stage magnification (M) is less than two, then $CV_{X_i}^2$ first increases and then decreases (see Figure 3), while for larger for values of $M > 2$, $CV_{X_i}^2$ decreases monotonically with i (see Figure 4).

We now compare the final protein noise in two different gene cascades of different sizes but same final average number of molecules. We assume that both the cascades have equal parameters, except for the per stage magnification M . For both cascades to have the same average number of molecules of the final protein, the cascade with the lower number of stages will need to have a larger magnification per stage. More specifically, if the cascades are of size N and K with per stage magnification of M_N and M_K , respectively, then

$$M_N^N = M_K^K. \quad (29)$$

From (28) we conclude that the ratio of the noises in the final protein are

$$\frac{CV_{X_N}^2}{CV_{X_K}^2} = \frac{(\alpha + dM_N)(M_K/2 - 1)(M_N/2)^N - 1}{(\alpha + dM_K)(M_N/2 - 1)(M_K/2)^K - 1} \quad (30)$$

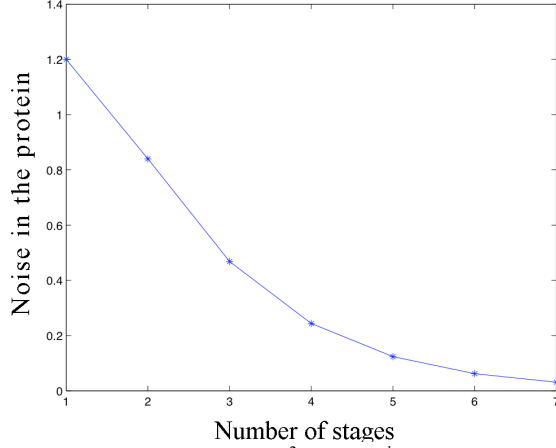


Fig. 4. Noise in the protein X_i ($CV_{X_i}^2$) at the i^{th} stage of the gene cascade when all genes have per stage magnification of $M = 5$, $\alpha = d = \gamma = 1$.

where $CV_{X_N}^2$ and $CV_{X_K}^2$ denote the noise in the final protein of a gene cascade with N and K stages, respectively. If the per stage magnification of both the cascades is less than two then for sufficiently large N and K we have

$$\frac{CV_{X_N}^2}{CV_{X_K}^2} \approx \frac{(\alpha + dM_N)(1 - M_K/2)}{(\alpha + dM_K)(1 - M_N/2)}. \quad (31)$$

If $N > K$, which implies $M_N < M_K$, then one can show that the above ratio is always less than one, i.e. the gene cascade with the larger number of stages has less noise in the final protein. In the other limit, when the magnification is much larger than two, we have from (30)

$$\frac{CV_{X_N}^2}{CV_{X_K}^2} \approx \frac{dM_N M_K/2 (M_N/2)^N}{dM_K M_N/2 (M_K/2)^K} = \left(\frac{1}{2}\right)^{N-K}. \quad (32)$$

This again shows that a gene cascade which achieves magnification with a larger number of stages and less magnification per stage will have lower stochastic fluctuation in the final protein compared to a cascade with a smaller number of stages and a higher magnification per stage.

V. SCALING OF EXTRINSIC AND INTRINSIC NOISE OVER NETWORKS

In the previous section we assumed that the transcription rate of gene $GeneX_{i+1}$ is equal to $\alpha \mathbf{x}_i$ where α was a constant. As α corresponds to the number of molecules of RNAP (the enzyme involved in transcription) present in cell, an implicit assumption in the previous section was that there were no significant fluctuations in the enzyme levels. We now incorporate these stochastic fluctuations in RNAP in the gene cascade. These fluctuations are often referred to as the *extrinsic noise*. We now modify the transcription rate of gene $GeneX_{i+1}$ to $\beta \mathbf{r}_{i+1}(t) \mathbf{x}_i(t)$ where β is a constant and $\mathbf{r}_{i+1}(t)$ is the number of molecules of RNAP involved in transcribing the gene $GeneX_{i+1}$ at time t . As in Section II we assume that this enzyme is constitutively expressed from a gene through random transcription, translation and degradation events and characterize the stochastic fluctuations in $\mathbf{r}_i(t)$ by

its coefficient of variation CV_{ext_i} . For simplicity we assume that \mathbf{r}_i and \mathbf{r}_j for $i \neq j$ are independent of each other which would correspond to different genes being transcribed by different types of RNA polymerases. The above assumption implies that \mathbf{r}_{i+1} is independent of \mathbf{x}_i , and hence we have

$$\mathbf{E}[\mathbf{x}_{i+1}(\infty)] = \frac{\beta \mathbf{E}[\mathbf{r}_{i+1}(\infty)] B_{i+1}}{d_{i+1}} \mathbf{E}[\mathbf{x}_i(\infty)] =: M_{i+1} \mathbf{E}[\mathbf{x}_i(\infty)] \quad (33)$$

for all $i \in [1, \dots, N-1]$ where $\mathbf{E}[\mathbf{r}_i(\infty)]$ denotes the steady-state count of the corresponding RNAP and as in the previous section we assume $M_{i+1} > 1$. As before, by using Dynkin's formula to write the moment dynamics and then performing a steady-state analysis we can relate the coefficients of variation of protein X_{i+1} and X_i as follows

$$CV_{X_{i+1}}^2 = \frac{1 + B_{i+1}}{\mathbf{E}[\mathbf{x}_{i+1}(\infty)]} + \xi_i CV_{X_i}^2 + \zeta_i CV_{ext_i}^2 + \delta_i CV_{ext_i}^2 CV_{X_i}^2 \quad (34)$$

where

$$\xi_i = \frac{d_{i+1}}{d_i + d_{i+1}} < 1, \quad \zeta_i = \frac{d_{i+1}}{g_{i+1} + d_{i+1}} < 1 \quad (35a)$$

$$\delta_i = \frac{d_{i+1}}{g_{i+1} + d_{i+1} + d_i} < 1 \quad (35b)$$

and g_i represents the degradation rate for the RNAP transcribing the gene $GeneX_i$. One can now conclude from (34) that unlike in Section IV where $CV_{X_i}^2$ goes to zero for sufficiently large i , $CV_{X_i}^2$ now approach some non-zero value determined by the amount of extrinsic noise in the cascade. Furthermore, if $\xi_i + \delta_i CV_{ext_i}^2 > 1$, for all i , then

$$\lim_{i \rightarrow \infty} CV_{X_i}^2 = \infty, \quad (36)$$

in which case the gene cascade is a noise magnifier where the downstream proteins have increased noise. Hence, by altering the amount of extrinsic noise in the cascade one could change a cascade from being a noise attenuator (as in Section IV where $CV_{ext_i}^2 = 0$) to a noise amplifier.

For a homogeneous cascade where

$$d_i = d, \quad \beta \mathbf{E}[\mathbf{r}_i(\infty)] = \alpha, \quad CV_{ext_i}^2 = CV_{ext}^2, \quad g_i = g, \quad B_i = B, \quad (37)$$

$$M_i = M = \frac{\alpha B}{d} > 1,$$

with the transcription rate of the first gene given by $\gamma \alpha$ for some constant γ , we have from (34) that

$$CV_{X_{i+1}}^2 = \frac{1+B}{\gamma M^{i+1}} + \frac{d}{d+g} CV_{ext}^2 + \left(\frac{1}{2} + \frac{d}{2d+g} CV_{ext}^2\right) CV_{X_i}^2. \quad (38)$$

From (38) we conclude that

$$\lim_{i \rightarrow \infty} CV_{X_i}^2 = \begin{cases} S & \text{for } CV_{ext}^2 < 1 + \frac{g}{2d} \\ \infty & \text{for } CV_{ext}^2 \geq 1 + \frac{g}{2d} \end{cases} \quad (39)$$

where

$$S = \frac{d CV_{ext}^2}{d+g} \frac{2(2d+g)}{2d(1 - CV_{ext}^2) + g}. \quad (40)$$

We can explicitly solve the difference equation (38) to obtain the noise in protein X_i to be

$$CV_{X_i}^2 = \frac{(\alpha + dM)((Mv)^i - 1)}{\alpha(Mv - 1)\gamma M^i} + \zeta CV_{ext}^2 \frac{v^i - 1}{v - 1} \quad (41)$$

where

$$v = \frac{1}{2} + \frac{d}{2d + g} CV_{ext}^2, \quad \zeta = \frac{d}{d + g}. \quad (42)$$

For sufficiently large M the above expression reduces to

$$CV_{X_i}^2 \approx \frac{d v^{i-1}}{\alpha \gamma} + \zeta CV_{ext}^2 \frac{v^i - 1}{v - 1}. \quad (43)$$

As in the previous section we now consider two cascades of lengths N, K with $N > K$ and per stage magnification M_N and M_K chosen such that both of them have the same average number of molecules for the final protein (i.e. $M_N^N = M_K^K$). We showed in Section IV that if $CV_{ext}^2 = 0$ then $CV_{X_N}^2 < CV_{X_K}^2$. However, from (43), when $v = 1$, i.e. $CV_{ext}^2 = 1 + g/2d$ we have

$$CV_{X_i}^2 \approx \frac{d}{\alpha \gamma} + i \zeta \left(1 + \frac{g}{2d}\right). \quad (44)$$

Thus when $CV_{ext}^2 = 1 + g/2d$, $CV_{X_N}^2 > CV_{X_K}^2$ as $N > K$. Since $CV_{X_i}^2$ varies continuously with CV_{ext}^2 , there must exist a $CV_{crit}^2 < 1 + g/2d$ such that if $CV_{ext}^2 > CV_{crit}^2$, then $CV_{X_N}^2 > CV_{X_K}^2$. This result illustrates that if the extrinsic noise present is larger than a critical value, then the gene cascade which achieves magnification with a smaller number of stages and higher magnification per stage will have lower stochastic fluctuation in the final protein compared to a cascade with a larger number of stages and a smaller magnification per stage.

VI. CONCLUSION AND FUTURE WORK

This paper presented results relating the stochastic noise in proteins of a gene cascade with the number of stages and the per stage magnification of the cascade. We provided explicit formulas to compute the noise in the proteins both in the absence and presence of extrinsic noise. We showed that when there is no extrinsic noise the noise in the i^{th} protein decreases with i and the cascade acts like a *noise attenuator*. Furthermore, for two different cascades with the same average final protein level, the cascade with the larger number of stages will have lower stochastic noise in the final protein. When sufficiently large extrinsic noise is present in the cell there is a role reversal for the cascade. More specifically, the cascade now acts like a *noise amplifier* where downstream proteins have increased noise. Thus depending on whether noise in the final product is deleterious or advantageous to the cell, stochastic noise can be appropriately modulated via cascades. We also showed that there exists a critical level of extrinsic noise above which, the gene cascade with the lower number of stages has lesser stochastic noise in the final protein.

We assumed in Section V that the molecular counts of the RNAP transcribing the i^{th} gene and j^{th} gene were independent of each other (i.e. \mathbf{r}_i and \mathbf{r}_j for $i \neq j$ were independent

of each other). However, if the same RNAP is involved in transcribing both the genes, \mathbf{r}_i and \mathbf{r}_j would be positively correlated. A direction of future work is to investigate scenarios where correlations exist between reaction rates and study its consequences on the scaling of stochasticity in cascades.

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] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, "Stochastic gene expression in a single cell," *Science*, vol. 297, pp. 1183–1186, 2002.
- [4] W. J. Blake, M. Krn, C. R. Cantor, and J. J. Collins, "Noise in eukaryotic gene expression," *Nature*, vol. 422, pp. 633–637, 2003.
- [5] J. L. Spudich and D. E. K. Jr, "Non-genetic individuality: chance in the single cell," *Nature (London)*, vol. 262, pp. 467–471, 1976.
- [6] 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.
- [7] 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.
- [8] A. Singh and J. P. Hespanha, "Stochastic analysis of gene regulatory networks using moment closure," in *Proc. of the 2007 Amer. Control Conference*, 2007.
- [9] M. H. A. Davis, *Markov models and Optimization*. Chapman and Hall, 1993.
- [10] 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.