Frequency modulation of stochastic gene expression bursts by strongly interacting small RNAs

Niraj Kumar, Tao Jia, Kourosh Zarringhalam, and Rahul V. Kulkarni

Phys. Rev. E **94**, 042419 — Published 24 October 2016

DOI: 10.1103/PhysRevE.94.042419

Frequency modulation of stochastic gene expression bursts by strongly interacting small RNAs

Niraj Kumar,[1,*] Tao Jia,[2,†] Kourosh Zarringhalam,[3,‡] and Rahul V. Kulkarni[1,§]

[1] *Department of Physics, University of Massachusetts Boston, Boston MA 02125, USA*
[2] *College of Computer and Information Science, Southwest University, Chongqing, 400715, P. R. China*
[3] *Department of Mathematics, University of Massachusetts Boston, Boston MA 02125, USA*

The sporadic nature of gene expression at the single cell level - long periods of inactivity punctuated by bursts of mRNA/protein production - plays a critical role in diverse cellular processes. To elucidate the cellular role of bursting in gene expression, synthetic biology approaches have been used to design simple genetic circuits with bursty mRNA/protein production. Understanding how such genetic circuits can be designed with the ability to control burst-related parameters requires the development of quantitative stochastic models of gene expression. In this work, we analyze stochastic models for the regulation of gene expression bursts by strongly interacting small RNAs (sRNAs). For the parameter range considered, results based on mean-field approaches are significantly inaccurate and alternative analytical approaches are needed. Using simplifying approximations, we obtain analytical results for the corresponding steady-state distributions that are in agreement with results from stochastic simulations. These results indicate that regulation by small RNAs, in the strong interaction limit, can be used to effectively modulate the frequency of bursting. We explore the consequences of such regulation for simple genetic circuits involving feedback effects and switching between promoter states.

PACS numbers: 87.10.Mn, 02.50.r, 87.17.Aa

I. INTRODUCTION

The stochastic nature of biochemical reactions in combination with small numbers of reacting molecules often leads to large fluctuations (noise) in the process of gene expression [1–8]. At the single-cell level, experimental observations reveal that mRNAs are often produced in bursts [9–11], characterized by long periods of inactivity punctuated by significantly shorter periods of transcriptional activity leading to creation of mRNAs [12]. Furthermore, since mRNAs are typically short-lived relative to proteins, the production of proteins is also expected to occur in bursts; as indeed observed in multiple single-cell experiments [13, 14]. Several studies have shown that bursting in gene expression plays a critical role in diverse cell-fate decisions ranging from viral latency [15] to cellular stress responses[16]. These observations have highlighted the need for understanding molecular mechanisms for the control of bursting in gene expression and for analyzing the consequences of such control in cellular genetic circuits.

Recent work has demonstrated that synthetic biology approaches can be used to develop simple genetic circuits that provide novel insights into the functioning of more complex cellular circuits found in nature [5, 17, 18]. In particular, analysis of these circuits at the single-cell level can elucidate how different sources of noise in gene expression contribute to observed variability at the population level. Furthermore, by appropriately designing components of the synthetic genetic circuit, the sources of noise such as bursting can be regulated and the impact of this regulation on the functioning of downstream circuits can be analyzed. A quantitative understanding of such molecular mechanisms to control and to exploit noise in gene expression can lead to both fundamental insights into cellular processes and to novel applications using synthetic genetic circuits.

Typically, synthetic biology research involving control of gene expression has focused on its regulation by proteins such as transcription factors; however in recent years there has been growing interest in characterizing and utilizing regulation by non-coding small RNAs (sRNAs). Regulation by sRNAs is known to play a key role in diverse cellular processes [19, 20] ranging from responses to stress [21] to virulence gene expression [22–25], whereas dysregulation of and by sRNAs is implicated in several diseases including cancer [26, 27]. Furthermore, it has been proposed that sRNAs play key roles in canalization during development by acting to buffer noise in gene expression [28, 29]. Approaches using synthetic biology have demonstrated the ability of sRNA based circuits to buffer protein synthesis rates [30] and to generate thresholds in target gene expression [31, 32]. These developments have highlighted the need for development of theoretical approaches that guide and complement experimental efforts to elucidate the impact of noise regulation by sRNAs in simple genetic circuits.

Previous work has led to the development of theoretical models for regulation by sRNAs. In particular, the canonical model [33] predicted that sRNAs generate thresholds in target gene expression as observed in recent

*Electronic address: niraj.kumar@umb.edu
†Electronic address: tjia@swu.edu.cn
‡Electronic address: kourosh.zarringhalam@umb.edu
§Electronic address: rahul.kulkarni@umb.edu

experiments [31]. It should be noted though that this analysis is based on a mean-field approach that does not take into account bursting in gene expression. While approaches that take bursting into account have been developed, they are typically based on the linear noise approximation [34], which is effectively an expansion around mean-field results. However, it has been shown that in the limit of infrequent transcription events and strong mRNA-sRNA interactions, mean-field results are not accurate [35]. In this limit, the role of sRNAs in regulating noise in gene expression is not well understood and alternative analytical approaches are needed.

In this work, we analyze the interplay between bursting in gene expression and its regulation by sRNAs in the limit of strong interactions and infrequent bursting. The key factors that are problematic for mean-field approaches (bursting and strong interactions) can be used to develop simplifying assumptions that lead to new analytical results in these limits. In the parameter regime analyzed, we find that regulation by sRNAs can be used to effectively modulate the frequency of bursting. Such frequency modulation of protein bursts (FMPB) can potentially play an important role in cellular processes[36], as seen in the activation of Crz1 transcription factor in yeast wherein the frequency of Crz1 bursts was shown to be regulated by extracellular calcium concentration [37]. In combination with synthetic biology based approaches, the results obtained can be used to gain new insights into the potential roles as well as functional capabilities of stochastic gene expression circuits involving sRNAs.

II. MODEL

We consider the canonical model [33] of mRNA-sRNA interactions generalized to allow for mRNA production in bursts. A schematic representation of the model is shown in Fig. 1. We assume that mRNA bursts arrive at a constant rate k_m and that each burst results in the production of a geometrically distributed number of mRNAs with mean burst size $\langle m_b \rangle$, consistent with experimental observations [9–11]. The production of sRNAs is taken to be a Poisson process occurring with rate k_s. The degradation of mRNAs and sRNAs can occur in two ways: individual degradation wherein mRNAs(sRNAs) degrade with rate $\mu_m(\mu_s)$, and mutual-degradation wherein a single mRNA and a single sRNA jointly degrade with rate γ. Translation of proteins from surviving mRNA occurs with rate k_p and protein degradation occurs with rate μ_p. The stochastic dynamics of the system is governed by the Master equation which describes the evolution of the joint probability distribution $P(s,m,p,t)$, wherein s, m, and p denote the number of sRNAs, mRNAs and proteins respectively, at time t.

$$\begin{aligned}\frac{\partial P(s,m,p,t)}{\partial t} &= k_m \sum_{n=1}^{m} P_{mb}(n) P(s,m-n,p,t) + \mu_m(m+1)P(s,m+1,p,t) + \mu_s(s+1)P(s+1,m,p,t) \\ &+ \gamma(s+1)(m+1)P(s+1,m+1,p,t) + k_s P(s-1,m,p,t) + k_p m P(s,m,p-1,t) + \mu_p(p+1)P(s,m,p+1,t) \\ &- (k_m + k_s + k_p m + \mu_m m + \mu_s s + \gamma m s + \mu_p p)\, P(s,m,p,t), \end{aligned} \tag{1}$$

where $P_{mb}(n)$ is the probability of producing n mRNA molecules in a geometrically distributed transcriptional burst (conditioned on the production of at least one mRNA) and is given by

$$P_{mb}(n) = (1-q_m)^{n-1} q_m, \tag{2}$$

with parameter q_m such that the mean burst size is

$$\langle m_b \rangle = \frac{1}{q_m}.$$

The different terms on the right hand side of Eq. (1) correspond to either transitions from other states to the state (s,m,p) or transition from the state (s,m,p) to other possible states. For example, the first term corresponds to transition to the state (s,m,p) from other states due to production of geometrically distributed random number of mRNAs in a burst. In order to explore the impact of sRNA regulation on the steady-state distribution of proteins, we need to solve Eq.(1) in the long-time limit. However, the nonlinear term associated with interaction parameter γ makes the solution analytically intractable. The nonlinear term results in the equations for lower moments getting coupled to equations for higher moments. For example, multiplying Eq. (1) by m, s and p and summing over all possible values results in the rate equations for the corresponding mean values, $\langle j \rangle = \sum_{m,s,p} j P(s,m,p,t)$, $j = m, s, p$:

$$\begin{aligned}\frac{\partial \langle m \rangle}{\partial t} &= k_m \langle m_b \rangle - \mu_m \langle m \rangle - \gamma \langle ms \rangle, \\ \frac{\partial \langle s \rangle}{\partial t} &= k_s - \mu_s \langle s \rangle - \gamma \langle ms \rangle, \\ \frac{\partial \langle p \rangle}{\partial t} &= k_p \langle m \rangle - \mu_p \langle p \rangle, \end{aligned} \tag{3}$$

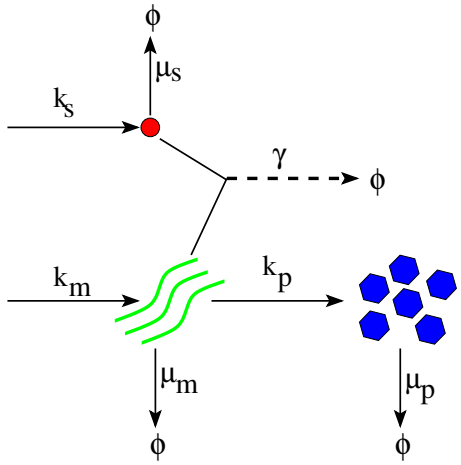

FIG. 1: Schematic representation of the model. Bursty synthesis of mRNAs is shown by Green lines, blue hexagons represent proteins, and red circles denote sRNA.

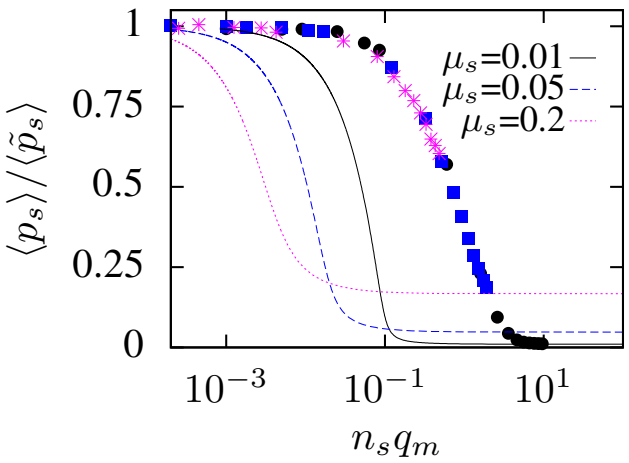

FIG. 2: Comparison of mean-field predictions with results from stochastic simulations. The figure shows variation of $\langle p_s\rangle/\langle\tilde{p}_s\rangle$ with $n_s q_m$, where $q_m = 1/\langle m_b\rangle$ and $\langle m_b\rangle$ is the mean burst size. $\langle\tilde{p}_s\rangle = (k_m k_p\langle m_b\rangle)/(\mu_m\mu_p)$ is the unregulated steady-state protein level and $n_s = k_s/\mu_s$ is the mean sRNA level in the absence of interaction with mRNAs. Lines are results from the mean field predictions for three different values of μ_s: $n_s q_m$ is varied by changing q_m, and the other parameters are fixed at $k_m = 0.001, \mu_m = 1, k_p = 50, \mu_p = 0.01, k_s = 0.1$ and $\gamma = 10$. Corresponding simulation results are shown as points.

where $\langle . \rangle$ represents the ensemble average over different realizations of the stochastic process. The above equations show that the evolution of mean mRNA or sRNA levels are coupled due to the nonlinear term $\gamma\langle ms\rangle$. This term, which appears due to sRNA-mRNA interactions, can be interpreted as the mean rate for coupled degradation of mRNAs and sRNAs. Furthermore, we can see that these equations are not closed and correspondingly obtaining exact expressions for even the mean steady-state levels of proteins and mRNAs is currently an open problem.

The correlation between mRNA and sRNA levels at any time t is given by $C = \langle ms\rangle - \langle m\rangle\langle s\rangle$. A commonly-used approximation to solve Eq. (3) is to use the mean-field approach [33] which corresponds to assuming $C = 0$ i.e. $\langle ms\rangle = \langle m\rangle\langle s\rangle$. In the steady-state, it is straightforward to obtain the solution of Eq. (3) using this approximation. The mean-field solution, which is exact in the absence of mRNA-sRNA interactions (i.e. $\gamma = 0$), can serve as a reasonable approximation for low γ values. However, as γ increases, the interactions between sRNAs and mRNAs are expected to lead to significant deviations from the approximation ($C = 0$) used to obtain mean field results. To get further insight, let us introduce the following dimensionless quantities:

$$x = \frac{\langle m\rangle}{k_m(\langle m_b\rangle/\mu_m)},\ y = \frac{\langle s\rangle}{k_s/\mu_s},\ z = \frac{\langle p\rangle}{(k_m k_p\langle m_b\rangle)/(\mu_m\mu_p)},$$

$$\epsilon_1 = \frac{\gamma k_s}{\mu_m\mu_s},\ \epsilon_2 = \frac{\gamma k_m\langle m_b\rangle}{\mu_m\mu_s}.$$

In the steady-state, using these dimensionless quantities, Eq. (3) reduces to

$$\begin{aligned} x + \epsilon_1 xy &= 1, \\ y + \epsilon_2 xy &= 1, \\ z &= x. \end{aligned} \tag{4}$$

The above set of equations can be solved to obtain the mean-field predictions for mean steady-state protein levels. Fig. 2 compares the mean-field predictions with the results from stochastic simulations for the variation in (scaled) steady-state protein levels as the mean burst size is varied. As can be seen in the figure, the mean-field predictions for the crossover behavior, i.e. the transition from suppression to expression in the mean protein levels, is highly inaccurate. In the parameter range considered, results from simulations show a sharp transition which is not reproduced by the mean-field analysis. Furthermore, the simulation data for different values of μ_s collapse onto a single curve when plotted using the scaled variables, another feature that is not reproduced by the mean-field analysis. In the next section, we develop an analytical approach based on simplifying assumptions that explains these features emerging from stochastic simulations.

III. FREQUENCY MODULATION OF PROTEIN BURSTS(FMPB) IN THE LIMIT OF STRONG INTERACTIONS AND INFREQUENT BURSTING

As noted in the previous section, in the limit of infrequent bursting and strong mRNA-sRNA interactions, results from mean-field approaches are significantly inaccurate and cannot serve as a guide for analyzing the quantitative or even the qualitative behavior of the corresponding genetic circuits. Since synthetic genetic cir-

cuits can be designed to operate in the limiting cases considered, it is of interest to explore if qualitatively new features arise when considering regulation of bursty gene expression by sRNAs in these limits. As noted, an exact solution of the master equation is intractable and therefore we need to consider simplifying assumptions.

In the limit that the mRNA degradation rate is much higher than the protein degradation rate ($\mu_m \gg \mu_p$), which is generally a valid assumption, the protein dynamics can be decoupled from mRNAs using the bursty synthesis approximation [38**?** , 39]. The approximation proceeds in two steps: 1) First obtain the protein burst distribution; this is the distribution of proteins produced from a single mRNA. 2) Assume that the protein burst is produced instantaneously, i.e. each burst arrival results in the instantaneous creation of proteins drawn from the burst distribution. In the absence of sRNA-based regulation, using this approximation leads to a master equation governing the evolution of proteins alone [38]. Notably, the derived equation can be solved to obtain an analytical expression for the steady-state distribution of proteins for arbitrary protein burst distributions [39].

We now focus on the limits of infrequent transcription and strong mRNA-sRNA interactions and make the following simplifying assumptions.

1) The limit of infrequent transcription implies that the arrival rate of mRNA bursts is small compared to the mRNA degradation rate. In this limit, mRNAs are degraded rapidly (either by natural or coupled degradation) after being created in transcriptional burst, following which there is typically a long time period wherein no mRNAs are present in the cell. Furthermore, we note that sRNAs and mRNAs can interact only during a transcriptional burst, i.e. the short time-scale over which the mRNAs are created and subsequently degraded. Since the interaction between mRNAs and sRNAs occurs only during a transcriptional burst, in the time interval *between* two consecutive mRNA transcriptional bursts, the evolution of sRNAs follows a simple birth-death process. In the limit of infrequent transcription, the mean time between consecutive mRNA bursts is large relative to the relaxation time for sRNA dynamics (i.e. $k_m \ll \mu_s$). Correspondingly, we assume that sRNA distribution *prior* to the arrival of the mRNA burst can be taken to be the steady-state distribution of sRNAs in the absence of mRNAs. For the model considered, this corresponds to the Poisson distribution, $\rho(s) = (n_s^s/s!)e^{-n_s}$, with mean sRNA level given by $n_s = k_s/\mu_s$.

2) To explore the domain of strong mRNA-sRNA interactions, we consider the limit $\gamma/\mu_m \to \infty$. This assumption implies that the effect of regulation by sRNAs is an instant modification of the mRNA burst distribution (see below).

3) Finally we assume that the sRNA transcription rate is such that we can neglect synthesis of new sRNAs *during* a mRNA burst.

Using these assumptions, we derive analytical expressions for the corresponding steady-state protein distributions which provide insights into the role of sRNA-based regulation in these limits. The results obtained are in good agreement with results from exact stochastic simulations for the model considered and for simple genetic circuits building on the basic model in Fig. 1.

Let us begin by analyzing how the mRNA burst distribution is modified due to interaction with sRNAs. Based on assumption 2 noted above, mRNAs and sRNAs are degraded instantaneously, with only the RNA molecules present in greater numbers surviving. If the number of mRNAs (m) in a transcriptional burst is less than or equal to that of sRNAs (s) at the beginning of transcription, all mRNAs are co-degraded with sRNA, and therefore no mRNA is left for translation. Proteins are expressed only when $m > s$, out of m mRNAs only the remaining $m - s$ mRNAs proceed for translation into proteins, as if there is no regulation by sRNA. We define $G_{pb}(z) = \sum_n z^n P_{pb}(n)$ as the generating function of the protein burst distribution (including contributions from all mRNAs created in the burst) and $G'_{pb}(z) = \sum_n z^n P'_{pb}(n)$ as the generating function of protein burst distribution from a single mRNA. Correspondingly, we obtain

$$G_{pb}(z) = \sum_{s,m} \Big(P(m \leq s) + P(m > s)(G'_{pb}(z))^{m-s} \Big). \quad (5)$$

Denoting the mRNA burst distribution by $P_{mb}(m)$, the above equation can be recast as

$$G_{pb}(z) = \sum_{i=j}^{\infty}\sum_{j=1}^{\infty} \rho(i) P_{mb}(j) + \sum_{i=0}^{\infty}\sum_{j=1}^{\infty} \rho(i) P_{mb}(i+j) G'^{\,j}_{pb}(z). \quad (6)$$

Using Eq. (2) in the above equation leads to

$$\begin{aligned} G_{pb}(z) &= 1 - \sum_{i=0}^{\infty}(1-q_m)^i \rho(i) \sum_{j=1}^{\infty} q_m (1-q_m)^{j-1} \\ &+ \sum_{i=0}^{\infty}(1-q_m)^i \rho(i) \sum_{j=1}^{\infty} q_m (1-q_m)^{j-1} G'^{\,j}_{pb}(z) \\ &= 1 - e^{-n_s q_m} + e^{-n_s q_m} \frac{q_m G'_{pb}(z)}{1 - G'_{pb}(z)(1-q_m)}. \quad (7) \end{aligned}$$

Note that the generating function for the protein burst distribution without regulation ($n_s = 0$) can be expressed as [40]

$$\tilde{G}_{pb}(z) = \frac{q_m G'_{pb}(z)}{1 - G'_{pb}(z)(1-q_m)}, \quad (8)$$

where

$$G'_{pb}(z) = 1/(1 + \frac{k_p}{\mu_m}(1-z)).$$

Correspondingly, the protein burst distribution (arising from all the mRNAs in the burst) can be expressed as

$$G_{pb}(z) = 1 - e^{-n_s q_m} + e^{-n_s q_m} \tilde{G}_{pb}(z). \quad (9)$$

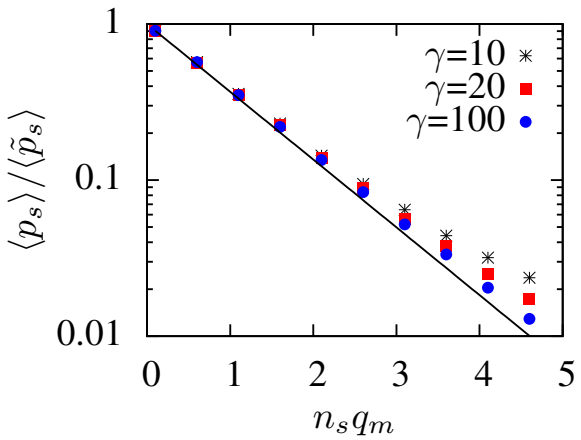

FIG. 3: Exponential suppression of protein levels due to sRNA regulation and its dependence on interaction strength γ: Scaled mean steady-state protein levels ($\langle p_s \rangle / \langle \tilde{p}_s \rangle$) as a function of $n_s q_m$ is plotted, $\langle \tilde{p}_s \rangle$ is the unregulated steady-state protein level. Points are simulation results for three different values of γ and the line represents analytic result using Eq. (10). Other parameters are: $k_m = 0.001, \mu_m = 1, k_p = 50, \mu_p = 0.01, k_s = 0.1, \mu_s = 0.01$ and $n_s q_m$ was varied by changing q_m.

The above result for $G_{pb}(z)$ has several interesting features. First, note that the probability that 1 or more mRNAs in the burst survive after the interaction with the sRNAs is given by $e^{-n_s q_m}$. This indicates that a key parameter controlling whether proteins are expressed during a transcriptional burst is $n_s q_m$, i.e. the ratio $n_s/\langle m_b \rangle$: the probability of protein expression during a burst decreases exponentially with increasing $n_s/\langle m_b \rangle$. This feature (exponential dependence on the ratio $n_s/\langle m_b \rangle$) is reflected in the results from exact stochastic simulations seen in Fig. 3 and the result derived thus identifies the underlying mechanism. Furthermore, it indicates that, conditional on 1 or more burst mRNAs surviving after the interaction with sRNAs, the regulated burst distribution is *identical* to the unregulated protein burst distribution. This implies that the effect of regulation by sRNAs (in the limit $\gamma/\mu_m \to \infty$) is to simply renormalize the rate of arrival of protein bursts, since the burst distribution (conditional on mRNA survival) is unchanged. If k_m denotes the rate of arrival of protein bursts in the unregulated case, then in the regulated case, this rate is renormalized to $k_m e^{-n_s q_m}$ and all other parameters remain effectively unchanged.

The arguments made above can be verified by explicitly deriving an expression for the protein steady-state distribution. Using previously derived results [39], we obtain that the protein steady-state distribution for the regulated case is given by (see Appendix A):

$$G_s(z) = \left(1 + \frac{k_p}{q_m \mu_m}(1-z)\right)^{-\frac{k_m}{\mu_p}\exp(-n_s q_m)}, \qquad (10)$$

which is the generating function for a negative binomial distribution. Note that the steady-state distribution for the unregulated case ($n_s = 0$) is also a negative binomial distribution; the result of sRNA regulation is a rescaling of the burst frequency by the factor $\exp(-n_s q_m)$, i.e. $k_m \to k_m \exp(-n_s q_m)$. Thus, regulation by sRNAs, in this strongly interacting limit, gives rise to frequency modulation of protein bursts (FMPB). Further, using Eq. (10) we obtain the steady-state expression for mean protein level as

$$\langle p_s \rangle \;=\; G_s'(1) = e^{-n_s q_m} \langle \tilde{p}_s \rangle, \qquad (11)$$

where prime ($\prime$) denotes differentiation of $G_s(z)$ w.r.t z at $z = 1$ and $\langle \tilde{p}_s \rangle$ is the mean value in the absence of regulation i.e. $\langle \tilde{p}_s \rangle = (k_m k_p)/(q_m \mu_m \mu_p)$. Thus, the observed frequency modulation of protein bursts leads to exponential suppression of mean protein levels with increasing mean sRNA levels. Note that, while these features were derived considering the limit $\gamma/\mu_m \to \infty$, results from simulations indicate that the limiting case is a good approximation for interaction strengths as low as $\gamma/\mu_m = 10$ (Fig. 3)

While the above result focuses on the specific case of regulation by sRNAs, the derivation clarifies that frequency modulation of bursts is a general feature that arises when we have: 1) geometric mRNA bursts and 2) a regulatory mechanism that results in preventing protein expression from some (or all) of the burst mRNAs. The main assumption is that the regulatory mechanism draws from the same distribution (for inactivation of mRNAs) for each burst and that each burst can be treated independently. Correspondingly, the key quantity is the probability P_s that 1 or more burst mRNAs survive the regulatory mechanism; for the model considered in Fig. 1 we have $P_s = \exp(-n_s q_m)$. For more general models, following the steps outlined in the derivation given above shows that the result of the regulatory mechanism is that the burst frequency is modulated as $k_m \to k_m P_s$. Thus the approach developed also provides insights into how general mechanisms for frequency modulation of bursts can be achieved.

IV. APPLICATIONS TO SIMPLE GENETIC CIRCUITS

In the following, we consider potential applications of the features noted, namely, exponential suppression and frequency modulation of protein bursts, in simple genetic circuits involving small RNAs.

(1) Suppression of bursts due to leakage: A potential application of the preceding results for sRNA-mediated regulation is filtering noise from a transcriptional signal that is composed of bursts of widely different burst sizes. A typical situation is that response to a signal (e.g. binding of a transcription factor) leads to mRNA bursts with high mean burst size; however bursts with much lower mean burst size can also occur in the

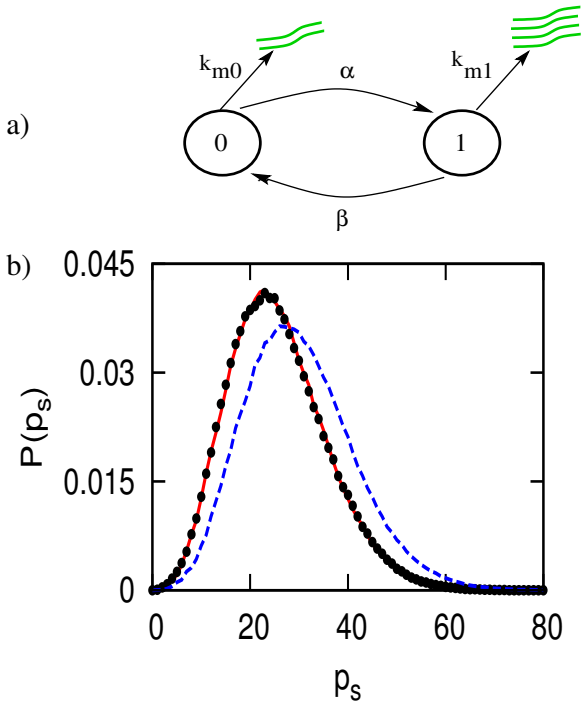

FIG. 4: Leakage suppression due to strong sRNA-mRNA interaction: In (a), kinetic scheme for the bursty production of mRNAs from the two-states of a promoter. In (b), simulation results for the steady-state probability distribution of proteins for such kinetic scheme are shown for the no-regulation cased $\gamma = 0$ (dashed line) and for $\gamma = 10$ (solid line) with other parameters as: $\alpha = 0.003$, $\beta = 0.001$, $k_{m0} = k_{m1} = 0.01$, $q_{m0} = 1$, $q_{m1} = 0.01$, $\mu_m = 0.5$, $k_p = 0.01$, $\mu_p = 0.0005$, $k_s = 1$ and $\mu_s = 0.1$. Points correspond to the steady-state protein distribution for the case when mRNAs are produced only from state 1 (i.e. $k_{m0} = 0$) with $\gamma = 10$ and keeping all the other parameters same.

absence of a signal, also known as leakage [41]. If the two mean burst sizes are widely separated, then, by appropriate tuning of the mean levels of sRNA regulator, the bursts due to leakage can be exponentially suppressed, whereas the bursts due to the signal will be minimally affected.

To exemplify the preceding scheme for filtering bursts due to leakage, we consider a kinetic scheme that takes into account promoter fluctuations (Fig. 4a). As indicated, the promoter has two states which are denoted by 1 and 0. The rate of promoter transition from state 0 to state 1 is denoted by α whereas the rate from 1 to 0 is denoted by β. Both states 0(1) generate transcriptional bursts with rates $k_{m0}(k_{m1})$, which are again geometrically distributed, but with different mean sizes characterized by $m_{b0}(m_{b1})$. Further, let us take $m_{b1} \gg m_{b0}$, i.e state 1 produces bursts of much larger mean burst size. In this case, the results derived indicate that by tuning the mean sRNA levels (n_s) such that $m_{b0} \ll n_s \ll m_{b1}$, the bursts due to leakage from state 0 will be exponentially suppressed, whereas the the bursts from state 1 will be minimally affected.

To test this prediction, we carried out stochastic simulations for the genetic circuit shown in Fig. 4. As can be seen in Fig. 4b, when γ is large, the steady-state probability distribution of proteins of the model where bursts are generated only from the state 1, i.e. $k_{m0} = 0$, is indistinguishable from the case where both states produce bursts with $m_{b0} \ll n_s \ll m_{b1}$. This implies that regulation by sRNAs has effectively suppressed expression from promoter state 0 (with burst size less than n_s) while minimally impacting expression from promoter state 1 (with burst size greater than n_s).

(2) Modulation of promoter switching in a positive feedback circuit: Switching between different promoter states can be a significant source of noise in gene expression [7]. In some cases, switching rates between different promoter states are driven by feedback due to the fluctuating levels of proteins expressed from the promoter [15, 42–44]. The switching induced by such feedback mechanisms can act to either enhance (positive feedback) or reduce (negative feedback) overall protein production. An important example is the well-studied genetic circuit involved in HIV-1 viral infections, wherein Tat proteins enhance their own production through a positive feedback mechanism [15]. Given the importance of bursting and feedback in diverse systems, it is of interest to analyze the impact of adding sRNA-based regulation to such genetic circuits.

We consider the gene expression model outlined in Fig.5a, with two promoter states denoted 0 and 1. The rate of switching from 0 to 1 has two contributions: first, a spontaneous switching rate α, and second, a feedback rate proportional to the number of proteins: $\tilde{\alpha}p$. The linear dependence of the feedback term is consistent with experimental observations for the genetic circuit in HIV-1 viral infections [15]. The rate of switching from 1 to 0 is denoted by β. The state 0(1) produces geometrically distributed mRNA bursts (with parameter $q_{m0}(q_{m1})$) at rate $k_{m0}(k_{m1})$. As in the model shown in Fig. 1, each mRNA created in the burst then degrades either due to self-degradation with rate μ_m or due to the interaction with a sRNA with rate γ. mRNAs that survive the interaction with sRNAs are translated with rate k_p into proteins that degrade with rate μ_p. In the bursty protein synthesis limit, this corresponds to the creation of geometrically distributed protein bursts with mean burst size $\langle p_{b0} \rangle = \frac{k_p}{q_{m0}\mu_m}$ ($\langle p_{b1} \rangle = \frac{k_p}{q_{m1}\mu_m}$).

The stochastic evolution of the system is encapsulated by the Master equation for $P(\sigma, s, m, p)$, where $\sigma = 0, 1$ denotes distinct promoter states. While the exact solution of the Master equation appears to be intractable due to the presence of nonlinear interaction terms, approximate analytical insights can be gained by applying the results derived in combination with recently obtained exact results for a stochastic gene expression model with bursting and feedback [42].

Our preceding results for large γ suggest that the effect of regulation by sRNAs can be captured in the model without sRNAs (i.e. $\gamma = 0$) by rescaling the

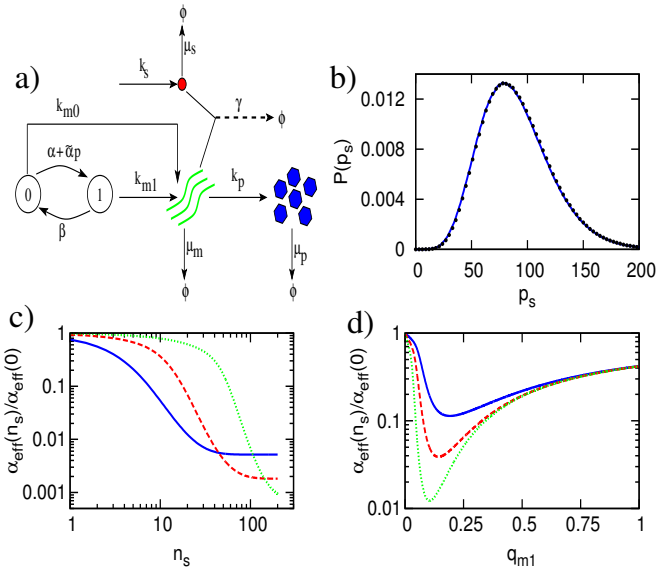

FIG. 5: Effects of strong sRNA-mRNA interaction on the effective promoter switching rate in a model with feedback shown in (a). In (b), simulation results for the probability distribution of proteins in the steady-state has been shown (line) for $\alpha = 0.003$, $\tilde{\alpha} = 0.1$, $\beta = 0.001$, $k_{m0} = 0.01$, $k_{m1} = 0.05$, $q_{m0} = 1$, $q_{m1} = 0.01$, $k_p = 0.05$, $k_s = 1$, $\mu_m = 0.5$, $\mu_p = 0.005$, $\mu_s = 0.1$ and $\gamma = 10$. Solid circles correspond to the effective/reduced model with $\gamma = 0$ but with k_{mi} scaled by the factor $\exp(-n_s q_{mi})$, $i = 0, 1$. c) Variation of the ratio of effective switching rate in the presence of sRNAs, $\alpha_{eff}(n_s)$, to that in the absence of sRNAs, $\alpha_{eff}(0)$, with n_s for different values of q_{m1}, 0.1 (solid), 0.05 (dashed), 0.02 (dotted). d) Variation of the ratio $\alpha_{eff}(n_s)/\alpha_{eff}(0)$ as a function of q_{m1} for three different values of n_s, 5 (solid), 10 (dashed), 20 (dotted). In both c) and d) other parameters are: $\alpha = 0.003$, $\tilde{\alpha} = 0.1$, $\beta = 0.001$, $k_{m0} = 0$, $k_{m1} = 0.05$, $k_p = 0.05$, $\mu_m = 0.5$, $\mu_p = 0.005$ and $\gamma = 10$.

rates k_{m0} and k_{m1} as: $k_{m0} \to k_{m0}\exp(-n_s q_{m0})$ and $k_{m1} \to k_{m1}\exp(-n_s q_{m1})$. As shown in Fig. 5b, in the limit of large γ, the steady-state protein distribution matches reasonably well with the corresponding distribution for a model with $\gamma = 0$ but k_{m0} and k_{m1} scaled by the exponential factors.

Let us consider the above circuit with $k_{m0} = 0$. The reduced model replaces the sRNAs with a rescaling of the the rate k_{m1}; correspondingly the Master equation is given by:

$$\begin{aligned}
\partial_t P_{0,p} &= \mu_p(p+1)P_{0,p+1} + \beta P_{1,p} - [\alpha + \tilde{\alpha}p + \mu_p p]P_{0,p}, \\
\partial_t P_{1,p} &= k_{m1}\exp(-n_s q_{m1}) \sum_{n=0}^{p} g(n) P_{1,p-n} \\
&+ \mu_p(p+1)P_{1,p+1} + (\alpha + \tilde{\alpha}p)P_{0,p} \\
&- [k_{m1}\exp(-n_s q_{m1}) + \beta + \mu_p p]P_{1,p},
\end{aligned} \tag{12}$$

where

$$g(n) = \langle p_b\rangle^n/(1+\langle p_b\rangle)^{n+1}$$

with $\langle p_b\rangle = \frac{k_p}{q_{m1}\mu_m}$ is the protein burst distribution. Following [42], an expression for the steady-state mean protein level, $\langle p_s\rangle$, and higher moments can be derived (see Appendix B). The results obtained can thus be used to determine the mean sRNA levels needed for adjusting the mean protein output of the genetic circuit to a desired level.

Besides the mean protein levels, we can also study how sRNAs can be used to tune the mean switching rate between promoter states. In the previous work [42], we have shown how to derive an effective model (with no feedback) characterized by a constant rate α_{eff} that is a good approximation to the original model with feedback. The effective switching rate α_{eff} is determined by the condition that the mean switching rate in both the effective and original models are identical (keeping the other parameters fixed). Correspondingly the expression for effective switching rate from the state 0 to 1 can be written as [42]

$$\alpha_{\rm eff} = \frac{\langle p_s\rangle \mu_p \beta}{k_{m1}\exp(-n_s q_{m1})\langle p_b\rangle - \langle p_s\rangle \mu_p}, \tag{13}$$

Using $\langle p_s\rangle$ from Eq. (B2) in this expression, we study the effects of adjusting sRNA levels on the switching rate from the state 0 to 1. In Fig. 5c we have shown the variation of this effective rate with n_s. As can be seen, by increasing the number of sRNAs for a given value of q_{m1}, this rate decreases monotonically. However, we notice that for a given n_s the rate shows non-monotonic variation with q_{m1} and there is a critical value of q_{m1} where the switching rate is minimum (Fig.5d). Thus controlling the mean levels of sRNAs can be used to effectively tune different properties of the feedback circuit such as mean protein levels and switching rates.

(3) Effects of FMPB on a simple two-state switch: Many important cellular systems can be represented by a simple two-state switch, with switching dynamics controlled by upstream signals. A classic example is the switching between run and tumble states in bacterial chemotaxis [45–49], where binding of an upstream regulator (CheY-P) to the flagellar motor induces a transition in the flagellar motor to a state with clockwise rotation (leading to tumbling) from a state with counter-clockwise rotation (leading to smooth runs).

Regulation of switching in a two-state system due to a fluctuating upstream regulator has been studied recently to analyze how upstream noise affects switching statistics [42, 49]. Considering an upstream noise source that is produced in bursts, it is of interest to analyze how frequency modulation of protein bursts (FMPB) by sRNAs can be used to tune noise in switching statistics.

A schematic representation of a two-state switch driven by an upstream regulatory protein that is created in geometric bursts is shown in Fig. 6a, with p_b as the mean

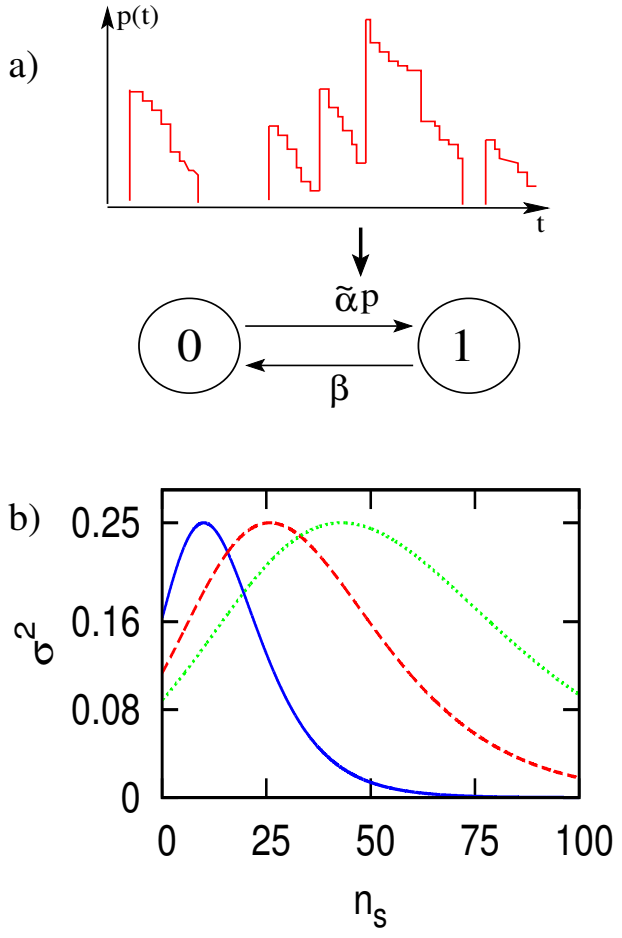

FIG. 6: Effects of burst frequency modulation due to sRNA regulation on a simple two-state switch. In (a), schematic figure showing a bursty input signal characterized by its mean burst size $\langle p_b \rangle$ and modulated burst frequency due to strong sRNA interaction $k_m \to k_m \exp(-n_s q_m)$. (b) Variance of the switch is plotted as a function of n_s for three different value of $\langle p_b \rangle$, 10 (blue), 20(red), 30 (green). Other parameters are: $\tilde{\alpha} = \beta = \mu_p = k_m = 1$.

burst size and k_m as the constant rate of arrival of bursts. The two states of the switch are represented by 0 and 1. Switching from state 0 to state 1 is driven by the upstream regulator, with rate $\tilde{\alpha}p$ (where p denotes protein number), whereas state 1 switches to state 0 with rate β.

We now note that the above kinetic scheme can be seen as a special case of the model shown in Fig. 5a [42]. Consider the special case $q_{m0} = q_{m1} = q_m$ and $k_{m0} = k_{m1} = k_m$ such that both promoter states produce geometrically distributed protein bursts of mean burst size $\langle p_b \rangle = \frac{k_p}{q_m \mu_m}$ with equal rate k_m. In this case, protein production is independent of promoter state, thus we have geometric protein bursts (with mean $\langle p_b \rangle$) arriving with constant rate k_m. Thus the kinetic scheme represents a two-state switch driven by a bursty input noise having mean size $\langle p_b \rangle$ and burst arrival rate k_m.

Following [42] and scaling k_m as $k_m \to k_m \exp(-n_s q_m)$, the steady-state probability that switch is in the state 0, P_0, is given as

$$P_0 = \mathcal{C}\, {}_2F_1[u+1, 1, u+2-w, 1-\phi], \tag{14}$$

where

$$\mathcal{C} = \frac{\phi\beta}{\beta + \frac{k_m \exp(-n_s q_m)\tilde{\alpha}\langle p_b \rangle}{\mu_p + \tilde{\alpha}(1+\langle p_b \rangle)}},$$

${}_2F_1$ is the Gaussian Hypergeometric function, and the quantities u, w and ϕ are given as

$$\begin{aligned} u &= \frac{\beta - \tilde{\alpha} k_m \exp(-n_s q_m)/\mu_p}{\mu_p + \tilde{\alpha}}, \\ w &= \frac{\mu_p + \tilde{\alpha}(1+\langle p_b \rangle)(1 - k_m \exp(-n_s q_m)/\mu_p)}{\mu_p + \tilde{\alpha}(1+\langle p_b \rangle)}, \\ \phi &= \frac{\mu_p + \tilde{\alpha}}{\mu_p + \tilde{\alpha} + \langle p_b \rangle \tilde{\alpha}}. \end{aligned} \tag{15}$$

Using Eqs. (14) and (15), we can analyze the effects of frequency modulation of protein bursts due to variation in mean sRNA number. Fig. 6b illustrates the variation of noise in switching statistics, $\sigma^2 = P_0 P_1$, with mean sRNA number n_s. We observe that σ^2 varies with n_s non-monotonically: In the limit $n_s \to \infty$, the upstream protein regulator is completely suppressed; thus in the long-time limit, the switch is localized in state 0 leading to zero variance. In the other limit $n_s \to 0$, switching frequency is independent of sRNA regulation: In this limit, switch variance changes with the burst size of the input signal and vanishes in the limit $\langle p_b \rangle \to \infty$. In between these two extreme limits ($n_s \to 0, \infty$), the variance changes non-monotonically and reaches the maximum at intermediate n_s values. Thus, adjusting the mean levels of sRNAs in the circuit can be used to effectively tune noise in a downstream two-state switch.

V. DISCUSSION

In summary, we have analyzed simple genetic circuits involving transcriptional bursting and post transcriptional regulation by sRNAs. Under the conditions of strong regulation and infrequent transcription, we derive analytical expression for the corresponding regulated protein burst and steady-state distributions. The results derived indicate that sRNA-based regulation can lead to exponential suppression of protein copy numbers by frequency modulation of protein bursts. These qualitative features can be utilized in developing strategies for suppressing leakage and modulating noise in simple genetic circuits.

While our analytical result is based on the assumption that γ/μ_m goes to infinity, we find through numerical simulations that lower γ/μ_m values are also well approximated by this limiting case (Fig. 3). For example, for

values as low as $\gamma/\mu_m = 10$ the features noted using the approximate analytical approach are also observed in the simulation results (Figs. 4, 5 and 6). It is interesting to note that the ratio $\gamma/\mu_m = 10$ is at the lower end of the range chosen in the previous work, which was taken from 10 to 100 [32, 50] . For much lower values of γ/μ_m, the results will change and the features noted will not be observed. Thus strong interactions, in the sense outlined above, are necessary to observe the features noted and thus to engineer the corresponding synthetic circuits. Recent work [51, 52] provides multiple examples of sythetic genetic circuits involving small RNA based regulation with tunable interaction strength. For example, in applications for metabolic engineering [53], it has been shown that naturally occuring small RNAs can be used to rationally design synthetic RNAs whose repression strength (effectively the parameter γ) correlates with the mRNA-sRNA binding energy which can be tuned by changes in the target-binding sequence. Thus it should be possible to engineer synthetic genetic circuits with strong mRNA-sRNA interactions as noted above.

Our approach provides insights into the impact of sRNA-based regulation in parameter regimes that have previously been unexplored. In particular, for the parameter regime considered, mean-field results do not serve as a good approximation thus highlighting the need for alternative approaches. The analytical results obtained in this work in limiting cases can serve as the base for developing approximate analytical approaches that are valid over larger regions of parameter space. Finally, as demonstrated in this work, the results obtained can serve as useful building blocks for the analysis of more complex genetic circuits.

Acknowledgments

This work was supported by the NSF through grant DMS- 1413111. T.J. is supported by the Natural Science Foundation of China (No. 6160309).

Appendix A: Derivation of steady-state protein distribution

Here we provide the details of deriving the steady-state protein distribution using the regulated protein burst distribution. In the limit that mRNA degradation rate is much higher than that of protein (the bursty synthesis approximation), bursts of proteins can be assumed to be produced instantaneously after the production of mRNAs, i.e. protein bursts can be assumed to arrive at the rate k_m. This simplification allows us to decouple protein evolution from the evolution of mRNA and sRNA. The corresponding evolution equation for proteins can then be written as

$$\begin{aligned}\frac{\partial P(p,t)}{\partial t} &= k_m \sum_{n=0}^{p} P_{pb}(n) P(p-n,t) + \mu_p (p+1) P(p+1,t) \\ &- (k_m + \mu_p p) P(p,t), \end{aligned} \tag{A1}$$

where $P_{pb}(n)$ is the probability of creating n proteins in a translational burst. We use Eq. (A1) and define the generating functions for protein burst distribution and protein steady-state distribution as

$$G_{pb}(z) = \sum_{n=0}^{\infty} z^n P_{pb}(n), \quad G(z,t) = \sum_{n=0}^{\infty} z^n P(n,t), \tag{A2}$$

Correspondingly, the evolution equation for the generating function reads

$$\partial_t G(z,t) = \mu_p (1-z) \partial_z G(z,t) + k_m [G_{pb}(z) - 1] G(z,t), \tag{A3}$$

which, in the steady-state limit, gives

$$G_s(z) = \exp\left(\frac{k_m}{\mu_p} \int_1^z \left(\frac{G_{pb}(y) - 1}{y-1} \right) dy \right). \tag{A4}$$

Thus $G_s(z)$ can be found if the expression for $G_{pb}(z)$ is known. Plugging in the expression for $G_{pb}(z)$ from Eq. (9) in Eq. (A4), we have

$$G_s(z) = \exp\left(\frac{k_m \exp(-n_s q_m)}{\mu_p} \int_1^z \left(\frac{\tilde{G}_{pb}(y) - 1}{y-1} \right) dy \right). \tag{A5}$$

Using the generating function for the protein burst distribution in the absence of regulation, given by Eq. (8), and using the generating function for protein burst from a single mRNA, $G'_{pb}(z) = \mu_m/(\mu_m + k_p(1-z))$, the steady-state expression for protein generating function can be written as

$$G_s(z) = \left(\frac{q_m \mu_m}{q_m \mu_m + k_p (1-z)} \right)^{\frac{k_m \exp(-n_s q_m)}{\mu_p}}, \tag{A6}$$

which is the generating function for a negative binomial distribution.

Appendix B: Derivation of steady-state mean for proteins(positive feedback)

The exact solution for the generating function corresponding to Eq. (12) is given by [42]:

$$\begin{aligned} G(z) &= \left[\frac{1}{1 + \langle p_b \rangle (1-z)} \right]^{k_{m1} \exp(-n_s q_{m1})/\mu_p} \\ &\times \frac{{}_2F_1[u, v | u+v+1-w | 1 - \phi\{1 + b(1-z)\}]}{{}_2F_1[u, v | u+v+1-w | 1-\phi]}, \end{aligned} \tag{B1}$$

where the quantities, u, v, w and ϕ are given by

$$u+v = \frac{\alpha+\beta-k_{m1}\exp(-n_s q_{m1})(1+\tilde{\alpha}/\mu_p)}{\mu_p+\tilde{\alpha}},$$
$$uv = -\frac{\beta k_{m1}\exp(-n_s q_{m1})}{\mu_p(\mu_p+\tilde{\alpha})},$$
$$w = \frac{\mu_p+\tilde{\alpha}(1+\langle p_b\rangle)-k_{m1}\exp(-n_s q_{m1})(1+\tilde{\alpha}(1+\langle p_b\rangle)/\mu_p)}{\mu_p+\tilde{\alpha}(1+\langle p_b\rangle)},$$
$$\phi = \frac{\mu_p+\tilde{\alpha}}{\mu_p+\tilde{\alpha}+\langle p_b\rangle\tilde{\alpha}},$$

and ${}_2F_1$ represents the Gaussian hypergeometric function. Eq. (B1) leads to the mean protein level as

$$\langle p_s\rangle/\langle p_b\rangle = \frac{k_{m1}\exp(-n_s q_{m1})}{\mu_p}+\phi\frac{uv}{u+v+1-w} \times\frac{{}_2F_1[u+1,v+1|u+v+2-w|1-\phi]}{{}_2F_1[u,v,u+v+1-w|1-\phi]}. \quad \text{(B2)}$$

[1] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Science **297**, 1836 (2002).
[2] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins, Nat. Rev. Genet. **6**, 451 (2005).
[3] J. M Paulsson, Phys. Life Rev. **2**, 157 (2005).
[4] A. Raj and A. van Oudenaarden, Cell **135**, 216 (2008).
[5] A. Eldar and M. B. Elowitz, Nature **467**, 167 (2010).
[6] A. Sanchez and I. Golding, Science **342**, 1188 (2013).
[7] A. Sanchez, S. Choubey, and J. Kondev, Annual review of biophysics **42**, 469 (2013).
[8] S. Tyagi, Mol. Syst. Biol. **11**, 805 (2015).
[9] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox, Cell **123**, 1025 (2005).
[10] A. Raj, C. S. Peskin, D. Tranchina, D.Y. Vargas, and S. Tyagi, PLoS Biol. **4**, e309 (2006).
[11] J. Chubb, T. Trcek, S. Shenoy, and R. Singer, Curr. Biol. **16**, 1018 (2006).
[12] N. Kumar, A. Singh, and R.V. Kulkarni, PLOS Comput.Biol. **11**, e1004292 (2015).
[13] L. Cai, N. Friedman, and X. S. Xie, Nature (London) **440**, 358 (2006).
[14] J. Yu, J. Xiao, X. Ren, K. Lao, and X. S. Xie, Science **311**, 1600 (2006).
[15] L. S. Weinberger, J. C. Burnett, J. E. Toettcher, A. P. Arkin, and D. V. Schafer, Cell **122**, 169 (2005).
[16] J. E. Purvis, K. W. Karhohs, C. Mock, E. Batchelor, A. Loewer, and Galit Lahav, Science **336**, 1440 (2012).
[17] T. S. Gardner, C. R. Cantor, and J. J. Collins, Nature **403 (6767)**, 339 (2000).
[18] J. A. N. Brophy and C. A. Voigt, Nat Methods **11(5)**, 508520 (2014).
[19] M. S. Ebert and P. A. Sharp, Cell **149 (3)**, 515-524 (2012).
[20] J. M. Schmiedel, S. L. Klemm, Y. Zheng, A. Sahay, N. Blthgen, D. S. Marks, and A. van Oudenaarden, Science **348**, 128132 (2015).
[21] S. Gottesman, , C. A. McCullen, , M. Guillier, C. K. Vanderpool, N. Majdalani, J. Benhammou, K. M. Thompson, P. C. FitzGerald, N. A. Sowa, and D. J. FitzGerald, Cold Spring Harb. Symp. Quant. Biol. **71**, 1(2006).
[22] T. Geissmann, Possedko, M. Possedko, E. Huntzinger, , P. Fechter, C. Ehresmann, and P. Romby, Handb Exp. Pharmacol. **173**, 9-43(2006).
[23] C. A. Schiano, L. E. Bellows and Lathem, W. W. Lathem, Infect. Immun. **78**, 2034 (2010).
[24] S. Chabelskaya, O. Gaillot and B. Felden, PLoS Pathog. **6**, e1000927 (2010).
[25] D. Podkaminski, and J. Vogel, Mol. Microbiol. **78**, 1327 (2010).
[26] I. Alvarez-Garcia, EA. Miska, Development **132(21)**, 4653 (2005).
[27] A. Esquela-Kerscher, FJ. Slack, Nat Rev Cancer **6(4)**, 259 (2006).
[28] E. Hornstein and S. Shomron, Nature Genet **38**, S20 (2006).
[29] M. Osella, C. Bosia, D. Cora and M. Caselle, PLoS Comput Biol **7(3)**, e1001101 (2011).
[30] T. J. Strovas, A. B. Rosenberg, B. E. Kuypers, R. A. Muscat, G. Seelig, ACS Synth Biol, **3(5)**, 324 (2014).
[31] S. Mukherji, M. S. Ebert, G. X. Y. Zheng, J. S. Tsang, P. A. Sharp, and A. V. Oudenaarden, Nat Genetics **43**, 854859 (2011).
[32] S. Legewie, D. Dienst, A. Wilde, H. Herzel, and I. M. Axmann, Biophys J. **95(7)**:3232 3238 (2008).
[33] E. Levine, Z. Zhang, T. Kuhlman, and T. Hwa, PLoS Biology **5(9)**, e229, (2007).
[34] J. Noorbakhsh, A. Lang, P. Mehta, PLOS One **8**, e72676, (2013).
[35] T. Platini, T. Jia, and R. V. Kulkarni, Phys. Rev. E **84**, 021928 (2011).
[36] J. H. Levine, Y. Lin and M. B. Elowitz, Science **342**, 1193 (2013).
[37] L. Cai, K. Dalal and M. B. Elowitz, Nature **455**,485 (2008).
[38] V. Shahrezaei and P. Swain, Proceedings of the National Academy of Sciences of the United States of America **105**, 17256 (2008).
[39] V. Elgart, T. Jia, A. T. Fenley, and R. V. Kulkarni, Phys Biol. **8**, 046001 (2011).
[40] T. Jia and R. V. Kulkarni, Phys. Rev. Lett. **105 (1)**, 018101 (2010).
[41] L. Huang, Z. Yuan, P. Liu, and T. Zhou, BMC systems biology **9(1)**, 1 (2015).
[42] Niraj Kumar, T. Platini, and R.V. Kulkarni, Phys. Rev. Lett. **113(26)**, 268105 (2014).
[43] Tsz-Leung To and N. Maheshri, Science **327**, 1142 (2010).
[44] L. Huang, Z. Yuan, P. Liu, and T. Zhou, Phys. Rev. E **90**, 052702 (2014).
[45] E. A. Korobkova, T. Emonet, H. Park, and P. Cluzel, Phys. Rev. Lett. **96**, 058105 (2006).
[46] J. L. Spudich and D. E. Koshland, Nature **262**, 467(1976).
[47] H. C. Berg and E. M. Purcell, Biophys. J. **20**, 193 (1977).
[48] H. Park, P. Oikonomou,C.C. Guet, and P. Cluzel, Biophys. J. **101**, 2336(2011).

[49] B. Hu, D. A. Kessler, W. J. Rappel, and H. Levine, Phys. Rev. Lett. **107**, 148101 (2011).
[50] N. Mitarai, A. M. C. Andersson, S. Krishna, S. Semsey and K. Sneppen, Physical biology **4(3)**, 164 (2007).
[51] Y. J. Lee, A. Hoynes-O'Connor, M. C. Leong, and T. S. Moon, Nucleic Acids Res. (2016).
[52] R. A. Mars, P. Nicolas, M. Ciccolini, E. Reilman, A. Reder, M. Schaffer, U. Ma der, U. Vo lker, J. M. van Dijl, and E. L. Denham, PLoS Genet. **11**, e1005046 (2015).
[53] Na D., M. S. Yoo, H. Chung, H. Park, H. J. Park, and Y. S. Lee, Nature Biotechnology, **31(2)**, 170 (2013).