


**Timing efficiency in small-RNA-regulated post-transcriptional processes**Kuheli Biswas and Anandamohan Ghosh *Indian Institute of Science Education and Research Kolkata, Mohanpur, Nadia 741246, India*

(Received 18 October 2019; revised manuscript received 6 February 2020; accepted 6 February 2020; published 26 February 2020)

Gene regulation in a cellular environment is a stochastic phenomenon leading to a large variability in mRNAs and protein numbers that are often produced in bursts. The regulation leading to varied protein dynamics can be ascribed to transcriptional or post-transcriptional mechanisms. In transcriptional regulation, the gene dynamically switches between an active and an inactive state, while in the post-transcriptional regulation small RNAs tune the activity of mRNAs. In either scenario, it is possible to calculate the time-dependent probability distribution of proteins and address the interesting question pertaining to their first passage time statistics. The coefficient of variation of first passage time can be considered to be an indicator of efficiency in controlling regulatory pathways and we show that post-transcriptional regulation performs better than simple transcriptional regulation for comparable protein yields.

DOI: [10.1103/PhysRevE.101.022418](https://doi.org/10.1103/PhysRevE.101.022418)**I. INTRODUCTION**

The process of gene expression is an intrinsically stochastic process involving random biochemical reactions governing transcription and translation. With the advancement of single-cell experimental techniques it has been observed that transcription and translation produces mRNAs and proteins, respectively, with large variation across a cellular population [1–3]. A generic feature that emerges for a wide range of operating regimes is that mRNAs and/or proteins are produced in bursts [4–6]. The experimental findings have been accurately explained by theoretical models explaining the mRNA and protein fluctuations [7–10]. In these studies, the transcriptional regulation is modeled as a stochastic process where the gene switches between an active (ON) and an inactive (OFF) state exhibiting bursts in mRNAs [11] as well as in protein dynamics described by a three-stage model [12]. It has also been observed that activity levels of mRNAs are varied resulting in different rates of translation, indicative of a possible existence of post-transcriptional regulations [13,14]. Recently, a general mathematical framework of post-transcriptional regulation mediated by small RNA (sRNA) has been proposed [15–17]. Unlike the gene switching between ON and OFF states, here transcription produces inactive mRNAs which can switch to active mRNAs leading to translational bursts as shown in the schematic in Fig. 1. In the inactive (OFF) state the ribosome binding site is shielded in the mRNA structure which is released by the sRNA turning ON the process of translation. A comprehensive analysis has been done by modeling the above scenario as stochastic differential equations (SDEs) and analytical expressions have been obtained for steady-state protein noise [17]. In this work we set up the post-transcriptional regulation as a master equation and obtain a general time-dependent solution of the probability distribution of proteins. Our analytical results, as compared with an approximate Gamma distribution, exhibit very good agreement with numerical results obtained by Gillespie simulation [18].

It will now be possible to obtain the time-dependent mean and variance of fluctuating protein numbers enabling us to determine the precision in first passage time (FPT) statistics.

The sequence of biochemical events occurring in developmental processes and cell-fate decision making necessitate temporal precision in the underlying cellular processes [19]. In the scenario of gene regulation the time taken by the expression levels of regulators to attain a certain threshold determines the onset of subsequent events, for example, microRNA-mediated control of mRNA threshold necessary for protein production [20,21]. It is thus important to understand the FPT properties of regulatory mechanisms [22]. In simple transcriptional regulation it has been possible to determine FPT statistics with a feature that it is inversely proportional to mRNA burst size [23,24]. Incorporating the protein dynamics it has been shown that timing fluctuations decreases with transcription rate and are independent of translation rate [25]. Equipped with our results for post-transcriptional regulation we estimate FPT coefficient of variation ( $\chi_T$ ) and compare with that of transcriptional regulation. We show that the timing precision, quantified by  $\chi_T$ , is determined by the competing protein burst size and the steady-state value. If both models yield comparable protein levels, we find that post-transcriptional regulation can attain superior timing precision in gene regulation.

**II. POST-TRANSCRIPTIONAL REGULATION**

Let  $P_{m,m^*,n}(t)$  be the probability of finding  $m$  number of mRNAs in the OFF state,  $m^*$  number of mRNAs in the ON state, and  $n$  number of proteins at time  $t$ . The average concentration of sRNA ( $\bar{s}$ ) switches mRNAs in the OFF state to the ON state with rate  $k_{\text{on}}\bar{s}$  while  $k_{\text{off}}$  is the rate of switching from OFF to ON. Here we do not consider the dynamics of sRNA and  $\bar{s}$  is constant, enabling us to rename  $k_{\text{on}} \equiv k_{\text{on}}\bar{s}$ . The transcription and the translation rates are  $k_m$

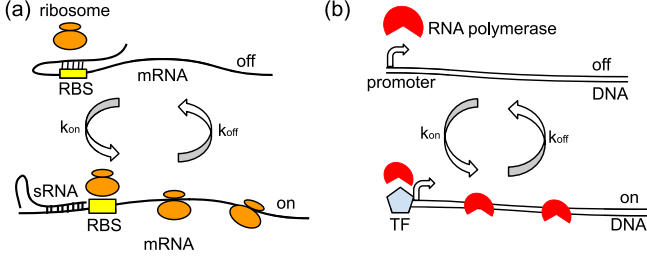


FIG. 1. Schematic representation of (a) post-transcriptional regulation (b) transcriptional regulation.

and  $k_p$ , respectively, with  $\gamma_m$  and  $\gamma_p$  being the corresponding degradation rates. The master equation for this model can be set up as follows:

$$\begin{aligned} \frac{\partial P_{m,m^*,n}}{\partial t} = & k_m [P_{m-1,m^*,n} - P_{m,m^*,n}] \\ & + k_p m^* [P_{m,m^*,n-1} - P_{m,m^*,n}] \\ & + k_{on} [(m+1)P_{m+1,m^*-1,n} - mP_{m,m^*,n}] \\ & + k_{off} [(m^*+1)P_{m-1,m^*+1,n} - m^*P_{m,m^*,n}] \\ & + \gamma_m [(m+1)P_{m+1,m^*,n} - mP_{m,m^*,n}] \\ & + \gamma_m [(m^*+1)P_{m,m^*+1,n} - m^*P_{m,m^*,n}] \\ & + \gamma_p [(n+1)P_{m,m^*,n+1} - nP_{m,m^*,n}]. \end{aligned} \quad (1)$$

A convenient approach to solve the master equation is to introduce a generating function

$$F(x, y, z, t) = \sum_{m,m^*,n} x^m y^{m^*} z^n P_{m,m^*,n}(t) \quad (2)$$

and rewrite the master equation as a partial differential equation (PDE) which can be solved by using the method of characteristics. It is possible to eliminate the variables  $(x, y)$  and impose the initial condition

$$\begin{aligned} F(z, \tau = 0) &= \sum P_n(\tau = 0) z^n \\ &= \sum \delta_{n,0} z^n = 1. \end{aligned} \quad (3)$$

The detailed calculation is shown in Appendix B and the final solution can be expressed in a simple form

$$F(z, \tau) = \left[ \frac{1 - \xi(z-1)e^{-\tau}}{1 - \xi z + \xi} \right]^\phi \quad (4)$$

by introducing the variables  $\xi$  quantifying the protein burst size,

$$\xi = b \frac{k_{on} + \gamma_m}{k_{on} + k_{off} + \gamma_m}, \quad (5)$$

where  $b = \frac{k_p}{\gamma_m}$  and  $\phi = \frac{k_m}{\gamma_p} \frac{k_{on}}{k_{on} + \gamma_m}$ . We now consider that the protein lifetime is much greater than the mRNA lifetime, which is a reasonable assumption, as typical mRNA degradation rate in *E. coli*,  $\gamma_m \sim \text{min}^{-1}$  and protein degradation rate,  $\gamma_p \sim \text{hr}^{-1}$  [26–28]. Thus, in the limit  $\gamma = \frac{\gamma_m}{\gamma_p} \gg 1$  and timescale  $\tau = t\gamma_p > \gamma^{-1}$  from the definition  $P_n =$

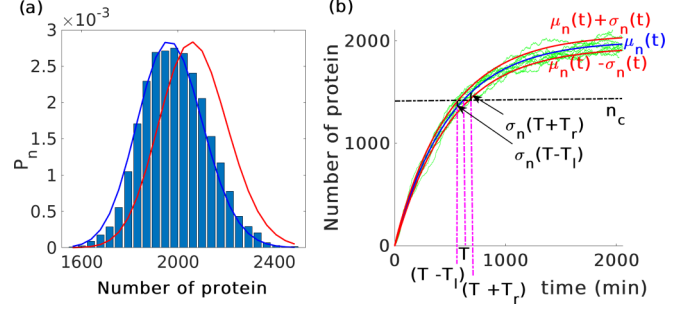


FIG. 2. (a) Post-transcriptional regulated protein steady-state distribution obtained from Gillespie simulation (histogram), analytical calculation (blue curve), mentioned in Eq. (7). The red curve is the approximate gamma distribution (8). (b) The mean protein numbers  $\mu_n(t)$  and their variance is shown as a function of time. The parameter values are  $k_{on} = 2 \text{ min}^{-1}$ ,  $k_{off} = 10 \text{ min}^{-1}$ ,  $k_m = 4.88 \text{ min}^{-1}$ ,  $k_p = 1 \text{ min}^{-1}$ ,  $\gamma_m = 0.2 \text{ min}^{-1}$ , and  $\gamma_p = 0.002 \text{ min}^{-1}$  such that the protein steady-state value  $n_{ss} = 2000$ .

$\frac{1}{n!} \frac{\partial^n}{\partial z^n} F(z, \tau) |_{z=0}$  we obtain the protein distribution function

$$\begin{aligned} P_n(\tau) &= \frac{\Gamma(\phi + n)}{\Gamma(n+1)\Gamma(\phi)} \left( \frac{\xi}{\xi + 1} \right)^n \left( \frac{\xi e^{-\tau} + 1}{\xi + 1} \right)^\phi \\ &\times {}_2F_1 \left( -n, -\phi, 1 - n - \phi; \frac{1 + \xi}{\xi + e^\tau} \right). \end{aligned} \quad (6)$$

In the steady state,  $\tau \rightarrow \infty$ , the protein distribution reduces to

$$P_n = \frac{\Gamma(\phi + n)}{\Gamma(n+1)\Gamma(\phi)} \left( \frac{\xi}{\xi + 1} \right)^n \left( \frac{1}{\xi + 1} \right)^\phi. \quad (7)$$

In Fig. 2(a) we show the steady-state protein distribution  $P_n$  obtained from the Gillespie simulation and the analytical expression in Eq. (7). Note that the agreement is good for a wide range of parameter values compared with the Gamma distribution

$$P_n = \frac{n^{(\tilde{a}-1)} e^{-n/\tilde{b}}}{\Gamma(\tilde{a}) \tilde{b}^{\tilde{a}}}, \quad (8)$$

which is an approximation in the limit  $k_{off} \gg k_{on}$ . The expressions of parameters  $\tilde{a}$ ,  $\tilde{b}$  are given in Appendix C. Comparisons of Eq. (7) with Eq. (8) are shown in Fig. 3 for different parameter values.

### III. FIRST PASSAGE TIME COEFFICIENT OF VARIATION ( $\chi_T$ )

From the generating function of protein in Eq. (4) we can derive the time-dependent moments of the protein distribution and get the mean  $\mu_n(t)$  and the variance  $\sigma_n^2(t)$  of the protein levels as

$$\mu_n(t) = n_{ss} (1 - e^{-t\gamma_p}), \quad (9)$$

$$\sigma_n^2(t) = \mu_n(t) (1 + \xi + \xi e^{-t\gamma_p}). \quad (10)$$

The steady-state value of the protein is  $n_{ss} = k_m k_p k_{on} / [\gamma_m \gamma_p (k_{off} + k_{on} + \gamma_m)]$ . In Fig. 2(b) we show the mean and the fluctuations obtained from Gillespie simulation and the theoretical curves. We are interested

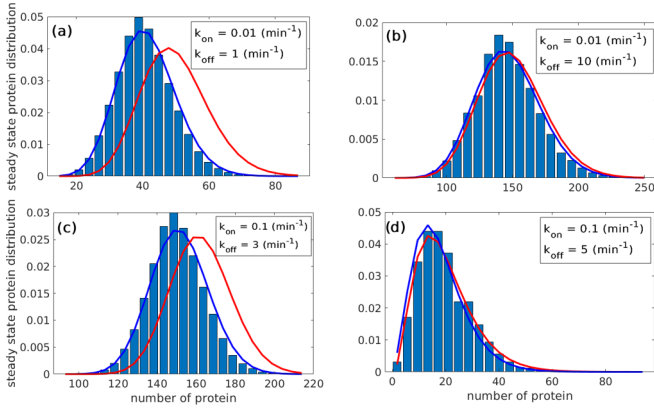


FIG. 3. For the post-transcription regulation we plot the steady-state distribution of protein obtained from a Gillespie simulation (histogram), and analytical calculations from Eq. (7) (blue curve) and Eq. (C5) (red curve) with parameter values (a)  $\xi = 0.86$ ,  $n_{ss} = 41$ ,  $k_m = 10 \text{ min}^{-1}$ ,  $k_p = 1 \text{ min}^{-1}$ ; (b)  $\xi = 3.1$ ,  $n_{ss} = 147$ ,  $k_m = 10 \text{ min}^{-1}$ ,  $k_p = 30 \text{ min}^{-1}$ ; (c)  $\xi = 0.45$ ,  $n_{ss} = 151.5$ ,  $k_m = 10 \text{ min}^{-1}$ ,  $k_p = 1 \text{ min}^{-1}$  and we take  $\gamma_m = 0.2 \text{ min}^{-1}$  and  $\gamma_p = 0.01 \text{ min}^{-1}$ ; (d)  $\xi = 4.28$ ,  $n_{ss} = 17.8$ ,  $k_m = 5 \text{ min}^{-1}$ ,  $k_p = 20 \text{ min}^{-1}$ ,  $\gamma_m = 0.5 \text{ min}^{-1}$ , and  $\gamma_p = 0.2 \text{ min}^{-1}$ .

in the first passage time i.e., the time  $T$  required for the protein numbers to reach a threshold  $n_c$ . A typical trajectory lying in  $\mu_n(t) \pm \sigma_n(t)$  cross the threshold  $n_c$  at time  $T \in [T - T_l, T + T_r]$ . Following the simple geometric argument as in Ref. [25] and outlined in detail in Appendix D, the coefficient of variation of protein FPT can be obtained as

$$\chi_T = \frac{\sigma_T}{T} = \frac{1}{T} \left[ \left( \frac{d\mu_n(t)}{dt} \right)^{-1} \sigma_n(t) \right]_T, \quad (11)$$

evaluated at mean FPT  $T \simeq -\ln(1 - \alpha_c)/\gamma_p$ . Using the form of  $\mu_n(t)$  and  $\sigma_n^2(t)$  from Eqs. (9) and (10), respectively, we get

$$\chi_T^2 = \frac{\alpha_c}{n_{ss}} \frac{1 + \xi(2 - \alpha_c)}{[(1 - \alpha_c) \ln(1 - \alpha_c)]^2}, \quad (12)$$

where  $\alpha_c = n_c/n_{ss}$ . The agreement with Gillespie simulations is good and  $\chi_T$  varies with protein burst per transcript as  $\sqrt{\xi}$  and shown in Fig. 4(a). Again a minimum  $\chi_T$  is observed as

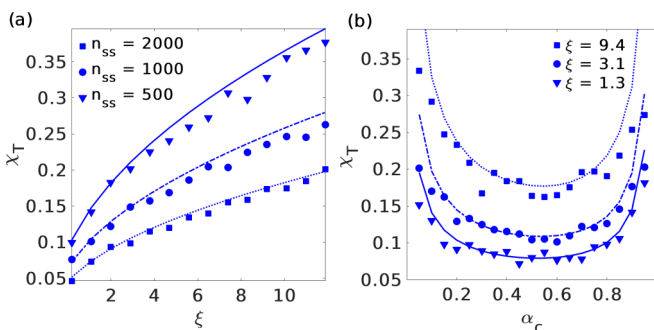


FIG. 4. (a)  $\chi_T$  vs  $\xi$  for different values of  $n_{ss}$  for a fixed  $\alpha_c = 0.5$ . (b)  $\chi_T$  vs  $\alpha_c$  for different value of  $\xi$  for constant  $n_{ss} = 2000$ . The parameter values are  $k_{on} = 2 \text{ min}^{-1}$ ,  $k_{off} = 10 \text{ min}^{-1}$ ,  $\gamma_m = 0.2 \text{ min}^{-1}$ ,  $\gamma_p = 0.002 \text{ min}^{-1}$ . The solid lines are obtained from Eq. (12) and the symbols are obtained from Gillespie simulation.

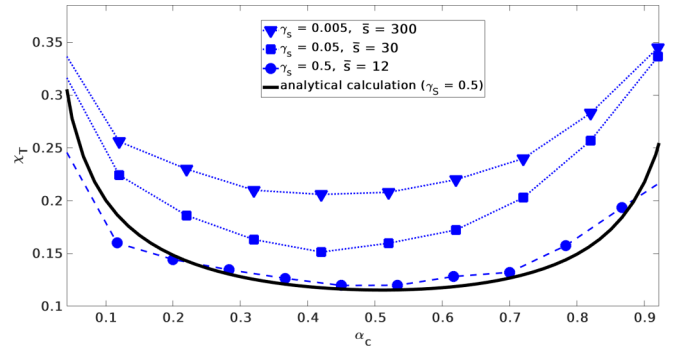


FIG. 5. Variation of  $\chi_T$  with  $\alpha_c$  for varying sRNA numbers. The dynamics of sRNAs follows the birth-death process with steady-state value  $\bar{s}$ . For small  $\bar{s}$  (large  $\gamma_s$ ), the Gillespie simulation results agree with Eq. (12) in the main text.

the threshold  $\alpha_c$  is varied, indicating an optimum operating regime Fig. 4(b).

So far we have considered the number of sRNAs to be constant at steady state  $\bar{s}$ . In general, the dynamics of sRNA can be described by a birth-death process [17] with synthesis rate  $k_s$  and degradation rate  $\gamma_s$ . In the limit  $\gamma_s > \gamma_p$ , sRNAs equilibrate faster than protein, implying that sRNA can be considered to remain constant as in the calculation above. However, for decreasing  $\gamma_s$ , numerical simulations show that  $\chi_T$  increases but the qualitative behavior is similar to the case of constant sRNA (Fig. 5).

#### IV. COMPARISON WITH TRANSCRIPTIONAL REGULATION

Coefficient of variation of FPT calculated for post-transcriptional regulation can be compared with transcriptional regulation models, namely two-stage and three-stage models [12]. In Ref. [12] the steady-state protein distribution for transcriptional regulation has already been obtained and the method can be easily extended to obtain the time-dependent protein distribution, as shown in Appendix A. In Table I we show the schematic of the models and provide expressions for steady states and protein bursts. The two-stage model is the *standard* scenario of transcription while in the three-stage model the gene switches between ON and OFF states leading to transcriptional as well as translational bursty dynamics. In the limit  $\gamma \gg 1$  the time-dependent mean and variance in protein numbers are still of the forms given in Eqs. (9) and (10) with the values of the steady state and burst parameter given in Table I. The  $\chi_T$  for the two-stage model has already been given in Ref. [25] and we compute that for the three-stage model from the expressions of moments of protein as given in Appendix A. Moreover, for all the three models considered here the expression of  $\chi_T$  is generically given as Eq. (12). The qualitative behavior of  $\chi_T$ , for instance, the variation with respect to threshold, is identical for all three models showing a minimum Fig. 6(a).

In the limits  $k_{on} \rightarrow \infty$  or  $k_{off} \rightarrow 0$ , the steady states of the transcriptional regulation and the post-transcriptional regulation reduces to  $k_m k_p / \gamma_m \gamma_p$  as in the two-stage model. Similarly, the two-stage model  $\chi_T$  is the limiting case for

TABLE I. Description of different gene regulatory networks for the protein synthesis.

Models	Chemical reactions	$n_{ss}$	$\xi$
Two-stage no regulation	$D_{on} \xrightarrow{k_m} M_{on} \xrightarrow{k_p} P$ $\downarrow \gamma_m$ $\downarrow \gamma_p$ $\emptyset$ $\emptyset$	$\frac{k_m k_p}{\gamma_m \gamma_p}$	$b = \frac{k_p}{\gamma_m}$
Three-stage transcriptional regulation	$D_{off} \xrightleftharpoons[k_{off}]{k_{on}} D_{on} \xrightarrow{k_m} M \xrightarrow{k_p} P$ $\downarrow \gamma_m$ $\downarrow \gamma_p$ $\emptyset$ $\emptyset$	$\frac{k_m k_p k_{on}}{\gamma_m \gamma_p (k_{on} + k_{off})}$	$b \left[ 1 + \frac{k_m k_{off}}{(k_{on} + k_{off})(k_{on} + k_{off} + \gamma_p)} \right]$
Post-transcriptional regulation	$D_{on} \xrightarrow{k_m} M_{off} \xrightleftharpoons[k_{off}]{k_{on}} M_{on} \xrightarrow{k_p} P$ $\downarrow \gamma_m$ $\downarrow \gamma_m$ $\downarrow \gamma_p$ $\emptyset$ $\emptyset$ $\emptyset$	$\frac{k_m k_p k_{on}}{\gamma_m \gamma_p (k_{on} + k_{off} + \gamma_m)}$	$b \left[ \frac{k_{on} + \gamma_m}{k_{on} + k_{off} + \gamma_m} \right]$

both types of regulation. It is important to note that, for any combination of rate constants,  $\xi^{\text{three stage}} \geq \xi^{\text{two stage}} \geq \xi^{\text{post-transcription}}$ . It is evident from Eq. (12) that  $\chi_T^2 \propto \xi$  and simultaneously  $\chi_T^2 \propto \frac{1}{n_{ss}}$  for constant threshold  $\alpha_c$  chosen as a fraction of the steady state  $n_{ss}$ . Thus the competition between the burst size and the steady-state value of the protein will determine the behavior of  $\chi_T$ . For example, in the regime of fast mRNA degradation ( $\gamma_m \gg 1$ ), the protein burst size is less but this will also lead to a decrease in the mean protein expression level. To compare timing precision of different mechanisms one may choose to impose a constraint of constant protein yield [22,25]. If we consider fixed protein steady state the condition

$$\chi_T^{\text{three stage}} \geq \chi_T^{\text{two stage}} \geq \chi_T^{\text{post-transcription}} \quad (13)$$

is always true for any threshold value, implying that, for the post-transcriptional regulation,  $\chi_T$  is less for a choice of rate constants satisfying the constraint.

In Fig. 6(b) we show that evaluation of Eq. (12) as well as Gillespie simulation results confirm that  $\chi_T$  for the post-transcription model is always less for any value of protein

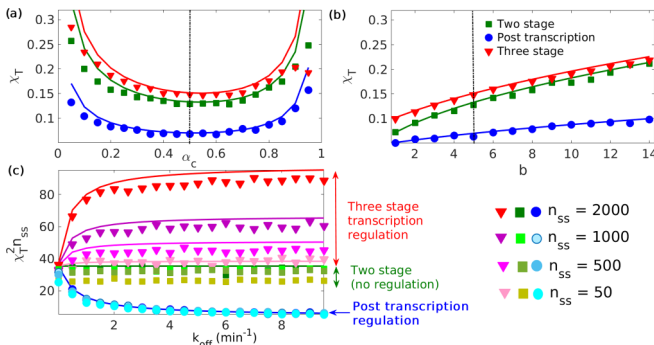


FIG. 6. Comparison of  $\chi_T$  for three models in Table I: (a)  $\chi_T$  as a function of the target expression level  $\alpha_c$  (in units of the steady-state value) for constant  $b = 5$ , (b)  $\chi_T$  vs protein burst per transcript,  $b$ , for constant threshold  $\alpha_c = 0.5$ . The protein numbers are in a steady-state value,  $n_{ss} = 2000$ , with other parameter values  $\gamma_m = 0.2 \text{ min}^{-1}$ ,  $\gamma_p = 0.002 \text{ min}^{-1}$ ,  $k_{off} = 10 \text{ min}^{-1}$  and  $k_{on} = 2 \text{ min}^{-1}$ . (c) Variation of  $\chi_T^2 n_{ss}$  with  $k_{off}$  for different values of  $n_{ss}$  for  $b = 5$  and  $\alpha_c = 0.5$ . The solid lines are obtained from Eq. (12) and symbols are obtained from the Gillespie simulation averaged over  $10^3$  realizations.

burst size and constant steady state. We can also check that, when  $k_{off} \rightarrow 0$ , the value of  $\chi_T$  for both the three-stage model and the post-transcription regulation converges to that of the two-stage model [Fig. 6(c)]. We have also varied the constant steady-state values  $n_{ss}$  over a wide range and scale  $\chi_T^2$  as  $\chi_T^2 n_{ss}$  to show that Eq. (13) is a generic feature. Moreover, as  $k_{on}$  increases for either model,  $\chi_T$  for the two-stage model is the limiting case, as shown in Fig. 7.

## V. CONCLUSIONS

In post transcriptional regulation a bacterial gene is positively regulated by sRNAs where transcription produces inactive mRNAs in which the ribosome binding site (RBS) is trapped in a strong mRNA structure and is released upon interaction between the sRNA and mRNA in the active state [16,17]. This mechanism is in contrast with the transcriptional three-stage model [12] and in this work we have studied the coefficient of variation of FPT,  $\chi_T$ , in gene-regulation models under different scenarios: no regulation (two stage), transcriptional regulation (three stage), and post-transcriptional regulation. A general expression for  $\chi_T$  is obtained in the regime of long protein lifetime  $\gamma_m > \gamma_p$  and matches with numerical simulations. The generic feature in all three models is that  $\chi_T$  (a) has an optimal level with respect to threshold

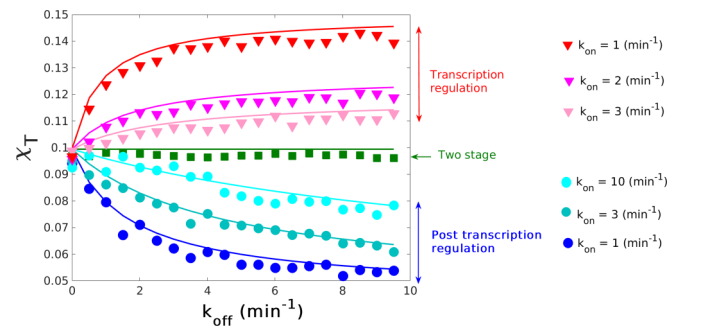


FIG. 7. Variation of  $\chi_T$  with  $k_{off}$  for different values of  $k_{on}$ . Protein steady state is held constant,  $n_{ss} = 2000$ , and other parameters are  $\alpha_c = 0.5$ ,  $\gamma_m = 0.2 \text{ min}^{-1}$ ,  $\gamma_p = 0.002 \text{ min}^{-1}$  and  $k_p = 0.5 \text{ min}^{-1}$ . The solid lines are obtained from Eq. (12) in the main text and symbols are obtained from the Gillespie simulation averaged over  $10^3$  realizations.



TABLE II. Parameter values used in Gillespie simulation to generate the figures.

Figure	Parameter	Parameter value ( $\text{min}^{-1}$ )	
Fig. 2	$\gamma_m$	0.2	
	$\gamma_p$	0.002	
	$k_{\text{on}}$	2	
	$k_{\text{off}}$	10	$n_{ss} = 2000$
	$k_p$	1	$b = 5$
	$k_m$	4.88	$\xi = 0.9016$
Fig. 4(a)	$\gamma_m$	0.2	
	$\gamma_p$	0.002	$n_{ss} = \{500, 1000, 2000\}$
	$k_{\text{on}}$	2	
	$k_{\text{off}}$	10	$\alpha_c = 0.5$
	$k_p$	{0.2 ... 10}	
	$k_m$	{24.4, ..., 0.488}	$\xi = \{0, \dots, 12\}$
Fig. 4(b)	$\gamma_m$	0.2	
	$\gamma_p$	0.002	
	$k_{\text{on}}$	2	$n_{ss} = 2000$
	$k_{\text{off}}$	10	
	$k_p$	{1.5, 3.5, 10.5}	$\xi = \{1.3, 3.1, 9.4\}$
	$k_m$	{3.2, 1.4, 0.46}	
Fig. 6(a)	$\gamma_m$	0.2	
	$\gamma_p$	0.002	
	$k_{\text{on}}$	2	
	$k_{\text{off}}$	10	$n_{ss} = 2000$
	$k_p$	1	
	$k_m$	4.88	$b = 5$
Fig. 6(b)	$\gamma_m$	0.2	
	$\gamma_p$	0.002	$n_{ss} = 2000$
	$k_{\text{on}}$	2	
	$k_{\text{off}}$	10	$\alpha_c = 0.5$
	$k_p$	{0.2, ..., 3}	
	$k_m$	{24.4, ..., 1.74}	$b = \{1, \dots, 14\}$
Fig. 6(c)	$\gamma_m$	0.2	
	$\gamma_p$	0.002	$n_{ss} = \{50, 500, 1000, 2000\}$
	$k_{\text{on}}$	0.4	$\alpha_c = 0.5$
	$k_p$	1	$b = 5$

the proteins has to attain, and (b) the burstiness of protein dynamics increases the  $\chi_T$ .

It is generically true that post-transcriptional protein burst size is less than transcriptional protein burst size, for any choice of rates. However, the same term that reduces the effective burstiness  $\xi$  also reduces the protein steady state in the post-transcriptional model. To compare the different regulatory models at equal protein yield, we therefore tuned the transcription rate  $k_m$  and the translation rate  $k_p$  because, for rapidly growing cells,  $k_m$  and  $k_p$  vary much more from gene to gene than their corresponding degradation rates [29–32]. Interestingly, with the constraint of fixed  $n_{ss}$ , our calculations, supported by numerical simulations, show that  $\chi_T$  is always less in the case of post-transcriptional regulation, and this is valid for a wide range of parameter values. The parameter values used in Gillespie simulations to generate the numerical data presented in the figures are given in Table II.

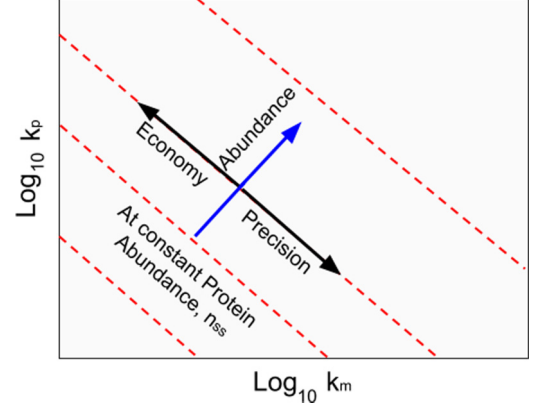


FIG. 8. Crick space [33].

Since we compare the different regulatory models also at constant burst size  $b \propto k_p$ , the only free parameter that we can then tune is  $k_m$ . For the two-stage model we have  $k_m^{\text{two stage}} = \frac{n_{ss}\gamma_p}{b}$  and we can choose  $k_m^{\text{two stage}} : k_m^{\text{three stage}} : k_m^{\text{post-transcription}} = 1 : (1 + \frac{k_{\text{off}}}{k_{\text{on}}}) : (1 + \frac{k_{\text{off}} + \gamma_m}{k_{\text{on}}})$ . Post-transcriptional regulation requires the largest value of  $k_m$ , implying increased precision but less economy by virtue of trade-off in the  $k_m$ - $k_p$  Crick space [33] (see Fig. 8). In line with this, we show that  $\chi_T$ , representative of temporal regulatory precision, requires increasing values of  $k_m$ . From a theoretical perspective, our study shows the possible functional role of small RNAs in determining the signatures of protein dynamics and is consistent with the finding of fast response to input signals [15]. It is possible to experimentally measure temporal dynamics of proteins and determine the timing efficiency in post-transcriptionally controlled gene regulatory pathways.

#### APPENDIX A: SOLUTION OF THE MASTER EQUATION FOR TRANSCRIPTIONAL REGULATION

The steady-state protein distribution of protein for three-stage model is given by Shahrezaei and Swain [12]. Here we explicitly show their calculation and also obtain the time-dependent protein distribution needed for the calculation of FPT.

Let  $P_{m,n}^{(0)}$  be the probability of having  $m$  mRNAs and  $n$  proteins when the promoter is inactive and  $P_{m,n}^{(1)}$  be the probability of having  $m$  mRNAs and  $n$  proteins when the promoter is active. We then have two coupled master equations for the three-stage model:

$$\begin{aligned} \frac{\partial P_{m,n}^{(0)}}{\partial \tau} = & \kappa_1 P_{m,n}^{(1)} - \kappa_0 P_{m,n}^{(0)} + (n+1)P_{m,n+1}^{(0)} - nP_{m,n}^{(0)} \\ & + \gamma [(m+1)P_{m+1,n}^{(0)} - mP_{m,n}^{(0)} + bm(P_{m,n-1}^{(0)} - P_{m,n}^{(0)})], \end{aligned} \quad (\text{A1})$$

$$\begin{aligned} \frac{\partial P_{m,n}^{(1)}}{\partial \tau} = & -\kappa_1 P_{m,n}^{(1)} + \kappa_0 P_{m,n}^{(0)} + (n+1)P_{m,n+1}^{(1)} - nP_{m,n}^{(1)} \\ & + a(P_{m-1,n}^{(1)} - P_{m,n}^{(1)}) + \gamma [(m+1)P_{m+1,n}^{(1)} \\ & - mP_{m,n}^{(1)} + bm(P_{m,n-1}^{(1)} - P_{m,n}^{(1)})], \end{aligned} \quad (\text{A2})$$

where  $b = k_p/\gamma_m$ ,  $a = k_m/\gamma_p$ ,  $\gamma = \gamma_m/\gamma_p$ ,  $\tau = t\gamma_p$ ,  $\kappa_1 = k_{\text{off}}/\gamma_p$ ,  $\kappa_0 = k_{\text{on}}/\gamma_p$ .

By defining two generating functions

$$f^{(0)}(y, z) = \sum_{m,n} y^m z^n P_{m,n}^{(0)}, \quad f^{(1)}(y, z) = \sum_{m,n} y^m z^n P_{m,n}^{(1)}, \quad (\text{A3})$$

the master equations become

$$\frac{1}{v} \frac{\partial f^{(0)}}{\partial \tau} = \frac{1}{v} [\kappa_1 f^{(1)} - \kappa_0 f^{(0)}] - \frac{\partial f^{(0)}}{\partial v} + \gamma \left[ b(1+u) - \frac{u}{v} \right] \frac{\partial f^{(0)}}{\partial u}, \quad (\text{A4})$$

$$\frac{1}{v} \frac{\partial f^{(1)}}{\partial \tau} = \frac{1}{v} [-\kappa_1 f^{(1)} + \kappa_0 f^{(0)}] - \frac{\partial f^{(1)}}{\partial v} + \frac{au}{v} f^{(1)} + \gamma \left[ b(1+u) - \frac{u}{v} \right] \frac{\partial f^{(1)}}{\partial u}, \quad (\text{A5})$$

with  $u = (y-1)$  and  $v = (z-1)$ . The characteristic equations will be

$$\frac{d\tau}{dr} = \frac{1}{v}, \quad (\text{A6})$$

$$\frac{dv}{dr} = 1, \quad (\text{A7})$$

$$\frac{du}{dr} = -\gamma \left[ b(1+u) - \frac{u}{v} \right], \quad (\text{A8})$$

$$\frac{df^{(0)}}{dr} = \frac{1}{v} [\kappa_1 f^{(1)} - \kappa_0 f^{(0)}], \quad (\text{A9})$$

$$\frac{df^{(1)}}{dr} = \frac{1}{v} [-\kappa_1 f^{(1)} + \kappa_0 f^{(0)}] + \frac{au}{v} f^{(1)}. \quad (\text{A10})$$

Consequently, direct integration of Eqs. (A6) and (A7) implies  $r = v = v_0 e^\tau$  and from Eq. (A8) in the limit  $\gamma \gg 1$ ,  $u(v) \approx bv/(1-bv)$ . From Eqs. (A9) and (A10) we get

$$v(bv-1) \frac{d^2 f^{(0)}}{dv^2} + [(\kappa_1 + \kappa_0)(bv-1) + bv(1+a) - 1] \frac{df^{(0)}}{dv} + ab\kappa_0 f^{(0)} = 0. \quad (\text{A11})$$

The solution of the Eq. (A11) is

$$f^{(0)}(v) = C {}_2F_1(\alpha, \beta, \kappa_0 + \kappa_1 + 1; bv), \quad (\text{A12})$$

where  ${}_2F_1(a, b, c; z)$  is a hypergeometric function with

$$\alpha = \frac{1}{2}(a + \kappa_0 + \kappa_1 + \sqrt{(a + \kappa_0 + \kappa_1)^2 - 4a\kappa_0}), \quad (\text{A13})$$

$$\beta = \frac{1}{2}(a + \kappa_0 + \kappa_1 - \sqrt{(a + \kappa_0 + \kappa_1)^2 - 4a\kappa_0}), \quad (\text{A14})$$

and  $C$  is the constant of integration.

The generating function for protein numbers,  $F(z) = f^{(0)}(z) + f^{(1)}(z)$ , can be given by using Eqs. (A12) and (A9) and the relation  $c(c+1) {}_2F_1(a, b, c; z) = c(c+$

$1) {}_2F_1(a, b, c+1; z) + abz {}_2F_1(a+1, b+1, c+2; z)$  as

$$F(z) = \frac{\kappa_0 + \kappa_1}{\kappa_1} C {}_2F_1\{\alpha, \beta, \kappa_0 + \kappa_1; b(z-1)\}. \quad (\text{A15})$$

Now the constant of integration  $C$  can be obtained from the initial condition

$$F(v_0) = F(\tau = 0) = \sum P_n(\tau = 0) z^n = \sum \delta_{n,0} z^n = 1. \quad (\text{A16})$$

By using  $v_0 = v e^{-\tau} = (z-1) e^{-\tau}$  we get the integration constant  $C$  as

$$C = \frac{\kappa_1}{\kappa_0 + \kappa_1} \frac{1}{{}_2F_1\{\alpha, \beta, \kappa_0 + \kappa_1; b(z-1) e^{-\tau}\}}. \quad (\text{A17})$$

By substituting the value of  $C$  into Eq. (A15) we get

$$F(z, \tau) = \frac{{}_2F_1\{\alpha, \beta, \kappa_0 + \kappa_1; b(z-1)\}}{{}_2F_1\{\alpha, \beta, \kappa_0 + \kappa_1; b(z-1) e^{-\tau}\}}, \quad (\text{A18})$$

$$\mu_n(t) = \frac{k_m k_p k_{\text{on}}}{\gamma_m \gamma_p (k_{\text{on}} + k_{\text{off}})} (1 - e^{-\gamma_p t}), \quad (\text{A19})$$

$$\eta_n^2(t) = \frac{1}{\mu_n(t)} (1 + \xi + \xi e^{-\gamma_p t}), \quad (\text{A20})$$

where

$$\xi = b \left[ 1 + \frac{k_m k_{\text{off}}}{(k_{\text{on}} + k_{\text{off}})(k_{\text{on}} + k_{\text{off}} + \gamma_p)} \right]. \quad (\text{A21})$$

In the steady state ( $\tau \rightarrow \infty$ ),  $\mu_n$  and  $\eta_n^2$  from Eqs. (A19) and (A20) are

$$\mu_n = \frac{k_m k_p k_{\text{on}}}{\gamma_m \gamma_p (k_{\text{on}} + k_{\text{off}})}, \quad (\text{A22})$$

$$\eta_n^2 = \frac{1}{\mu_n} + \gamma^{-1} \frac{1}{\mu_m} + \frac{\gamma_p}{k_{\text{on}} + k_{\text{off}} + \gamma_p} \eta_D^2, \quad (\text{A23})$$

where  $\eta_D^2 = \frac{k_{\text{off}}}{k_{\text{on}}}$  and  $\mu_m = \frac{k_m k_{\text{on}}}{\gamma_m (k_{\text{on}} + k_{\text{off}})}$ . The expression of protein noise and average in steady state is exactly same as that of Shahrezaei and Swain [12].

## APPENDIX B: SOLUTION OF MASTER EQUATION FOR POST-TRANSCRIPTIONAL REGULATION

The master equation in the continuum limit can be written as the partial differential equation (PDE)

$$\frac{\partial F}{\partial w} + \frac{1}{w} \frac{\partial F}{\partial \tau} + \gamma \left( \frac{u}{w} - \frac{\kappa_0}{w\gamma} (v-u) \right) \frac{\partial F}{\partial u} + \gamma \left( \frac{v}{w} - \frac{\kappa_1}{w\gamma} (u-v) - b(v+1) \right) \frac{\partial F}{\partial v} = \frac{au}{w} F, \quad (\text{B1})$$

where we have introduced the dimensionless parameters  $b = k_p/\gamma_m$ ,  $a = k_m/\gamma_p$ ,  $\gamma = \gamma_m/\gamma_p$ ,  $\tau = t\gamma_p$ ,  $\kappa_1 = k_{\text{off}}/\gamma_p$ ,  $\kappa_0 = k_{\text{on}}/\gamma_p$ , and the change of variables  $u = x-1$ ,  $v = y-1$ ,  $w = z-1$ . The above PDE can be solved by the method of characteristics, yielding the set of ODEs

$$\frac{dw}{dr} = 1, \quad (\text{B2})$$

$$\frac{d\tau}{dr} = \frac{1}{w}, \quad (\text{B3})$$

$$\frac{dF}{dr} = \frac{au}{w} F, \quad (\text{B4})$$

$$\frac{du}{dr} = \gamma \left( \frac{u}{w} - \frac{\kappa_0}{w\gamma} (v - u) \right), \quad (\text{B5})$$

$$\frac{dv}{dr} = \gamma \left( \frac{v}{w} - \frac{\kappa_1}{w\gamma} (u - v) - b(v + 1) \right). \quad (\text{B6})$$

From the characteristic equations (B2) and (B3) we get  $r = w = w_0 e^\tau$ . The remaining equations are nonlinear but, assuming that the proteins are more stable compared with mRNAs, we invoke the limit  $\gamma \gg 1$ . From Eqs. (B5) and (B6) it is now easy to obtain

$$u(w) = \frac{bw\kappa_0}{(\kappa_1 + \kappa_0 + \gamma) - wb(\gamma + \kappa_0)}, \quad (\text{B7})$$

which on substitution into Eq. (B4) gives

$$\frac{dF}{dw} = \frac{ab\kappa_0}{(\kappa_1 + \kappa_0 + \gamma) - wb(\gamma + \kappa_0)} F. \quad (\text{B8})$$

With the initial condition

$$\begin{aligned} F(w_0) &= F(\tau = 0) = \sum P_n(\tau = 0) z^n \\ &= \sum \delta_{n,0} z^n = 1, \end{aligned} \quad (\text{B9})$$

the solution of Eq. (B8) can be given a simple representation

$$F(z, \tau) = \left[ \frac{1 - \xi(z - 1)e^{-\tau}}{1 - \xi z + \xi} \right]^\phi \quad (\text{B10})$$

by introducing the variables

$$\begin{aligned} \xi &= b \frac{k_{\text{on}} + \gamma_m}{k_{\text{on}} + k_{\text{off}} + \gamma_m}, \\ \phi &= \frac{a\kappa_0}{\kappa_0 + \gamma}. \end{aligned}$$

#### APPENDIX C: PROTEIN DISTRIBUTION FOR POST-TRANSCRIPTIONAL REGULATION

The protein distribution in the steady state is a negative binomial distribution

$$P_n = \frac{\Gamma(\phi + n)}{\Gamma(n + 1)\Gamma(\phi)} \left( \frac{1}{1 + 1/\xi} \right)^n \left( \frac{1}{\xi + 1} \right)^\phi \quad (\text{C1})$$

$$= \frac{\Gamma(\phi + n)}{\Gamma(n + 1)\Gamma(\phi)} \left( \frac{1}{\xi + 1} \right)^\phi \left( \frac{1}{1 + \frac{n}{\xi} + \frac{n(n-1)}{2\xi^2} + \dots} \right) \quad (\text{C2})$$

$$= \frac{n^{(\phi+n)}}{n^{(n+1)}\Gamma(\phi)} \left( \frac{1}{\xi} \right)^\phi \left( \frac{1}{e^{n/\xi}} \right) \quad (\text{If } n \gg 1 \text{ and } \xi \gg 1) \quad (\text{C3})$$

$$= \frac{n^{(\phi-1)} e^{-n/\xi}}{\Gamma(\phi)\xi^\phi}. \quad (\text{C4})$$

In the limit of large number  $n$  of proteins and large burst size  $\xi$ , the negative binomial distribution approaches the Gamma distribution with mean value  $\mu_n = \xi\phi = k_m k_p k_{\text{on}} / \gamma_m \gamma_p (k_{\text{off}} + k_{\text{on}} + \gamma_m)$ , Fano factor  $\simeq \xi$ , and  $\eta_n^2 = 1/\phi$ . In Ref. [17] the protein distribution for the post transcription regulation is given by

$$P_n = \frac{n^{(\tilde{a}-1)} e^{-n/\tilde{b}}}{\Gamma(\tilde{a})\tilde{b}^{\tilde{a}}}, \quad (\text{C5})$$

with  $\tilde{a} = \mu_n \tau$ ,  $\mu_n = k_m / \gamma_m$ ,  $\tau = \tau_p / (\tau_{\text{on}} + \tau_m + \tau_y)$ ,  $\tau_p = 1/\gamma_p$ ,  $\tau_{\text{on}} = 1/k_{\text{on}}$ ,  $\tau_m = 1/\gamma_m$ ,  $\tau_y = (k_{\text{on}} + k_{\text{off}}) / k_p k_{\text{on}}$ , and  $\tilde{b} = 1 + k_p / k_{\text{off}} + k_p k_{\text{on}} / \gamma_m (k_{\text{on}} + k_{\text{off}})$ .

The Gamma distributions (C4) and (C5) are equivalent

(1) in the limit when  $k_{\text{off}} \gg k_{\text{on}}$ ;

(2) when  $\gamma_m$  is much smaller than both  $k_{\text{on}}$  and  $k_{\text{off}}$ .

#### APPENDIX D: CALCULATION OF FIRST PASSAGE TIME FLUCTUATIONS

With an explicit knowledge of the time-dependent moments from the protein distribution, it is possible to compute the FPT, as illustrated in Co *et al.* [25]. Denoting the average dynamics as  $\mu_n(t)$  as illustrated in Fig. 9, two geometric relations hold:

$$\begin{aligned} \sigma_n(T + T_r) &= \mu_n(T + T_r) - \mu_n(T), \\ \sigma_n(T - T_l) &= \mu_n(T) - \mu_n(T - T_l), \end{aligned} \quad (\text{D1})$$

which on Taylor expansion around the FPT,  $T$ , the time to reach a threshold  $n_c = \mu_n(T)$ , can be written as

$$\begin{aligned} \sigma_n(T) + \frac{d}{dt} \sigma_n(t) \Big|_T T_r &= \frac{d}{dt} \mu_n(t) \Big|_T T_r, \\ \sigma_n(T) - \frac{d}{dt} \sigma_n(t) \Big|_T T_l &= \frac{d}{dt} \mu_n(t) \Big|_T T_l. \end{aligned} \quad (\text{D2})$$

The spread in FPT can now be given by

$$\begin{aligned} \sigma_T &\simeq \frac{T_r + T_l}{2} \\ &\simeq \left[ \sigma_n(t) \frac{d\mu_n(t)}{dt} \left[ \left( \frac{d\mu_n(t)}{dt} \right)^2 - \left( \frac{d\sigma_n(t)}{dt} \right)^2 \right]^{-1} \right]_T. \end{aligned} \quad (\text{D3})$$

Assuming that the variability in the protein level is constant in time, we get further simplification:

$$\sigma_T \simeq \left[ \sigma_n(t) \left( \frac{d\mu_n(t)}{dt} \right)^{-1} \right]_T, \quad (\text{D4})$$

from which we can obtain FPT coefficient of variation.

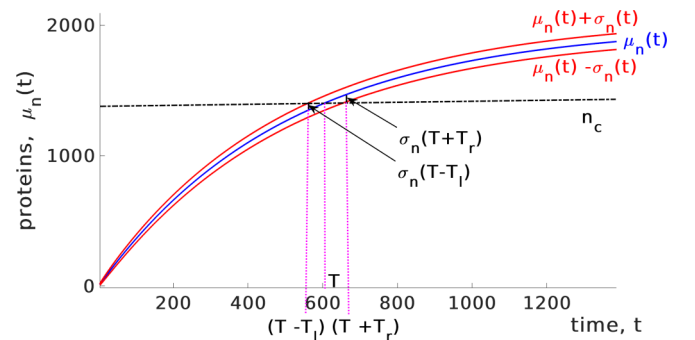


FIG. 9. The mean and variance of protein numbers as a function of time.

- [1] M. Thattai and A. Van Oudenaarden, *Proc. Natl. Acad. Sci. USA* **98**, 8614 (2001).
- [2] J. Paulsson, *Nature (London)* **427**, 415 (2004).
- [3] A. Raj and A. van Oudenaarden, *Cell (Cambridge, MA, U. S.)* **135**, 216 (2008).
- [4] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox, *Cell (Cambridge, MA, U. S.)* **123**, 1025 (2005).
- [5] L. Cai, N. Friedman, and X. S. Xie, *Nature (London)* **440**, 358 (2006).
- [6] L.-h. So, A. Ghosh, C. Zong, L. A. Sepúlveda, R. Segev, and I. Golding, *Nat. Genet.* **43**, 554 (2011).
- [7] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. Van Oudenaarden, *Nat. Genet.* **31**, 69 (2002).
- [8] P. S. Swain, M. B. Elowitz, and E. D. Siggia, *Proc. Natl. Acad. Sci. USA* **99**, 12795 (2002).
- [9] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi, *PLoS Biol.* **4**, e309 (2006).
- [10] T. Jia and R. V. Kulkarni, *Phys. Rev. Lett.* **106**, 058102 (2011).
- [11] J. Peccoud and B. Ycart, *Theor. Popul. Biol.* **48**, 222 (1995).
- [12] V. Shahrezaei and P. S. Swain, *Proc. Natl. Acad. Sci. USA* **105**, 17256 (2008).
- [13] B. Wu, C. Eliscovich, Y. J. Yoon, and R. H. Singer, *Science* **352**, 1430 (2016).
- [14] X. Yan, T. A. Hoek, R. D. Vale, and M. E. Tanenbaum, *Cell (Cambridge, MA, U. S.)* **165**, 976 (2016).
- [15] P. Mehta, S. Goyal, and N. S. Wingreen, *Mol. Syst. Biol.* **4**, 221 (2008).
- [16] G. Rodrigo, T. E. Landrain, and A. Jaramillo, *Proc. Natl. Acad. Sci. USA* **109**, 15271 (2012).
- [17] G. Rodrigo, *Phys. Rev. E* **97**, 032401 (2018).
- [18] D. T. Gillespie, *J. Phys. Chem.* **81**, 2340 (1977).
- [19] J. M. Pedraza and J. Paulsson, *Mol. Syst. Biol.* **3**, 81 (2007).
- [20] S. Bagga, J. Bracht, S. Hunter, K. Massirer, J. Holtz, R. Eachus, and A. E. Pasquinelli, *Cell (Cambridge, MA, U. S.)* **122**, 553 (2005).
- [21] S. Mukherji, M. S. Ebert, G. X. Zheng, J. S. Tsang, P. A. Sharp, and A. van Oudenaarden, *Nat. Genet.* **43**, 854 (2011).
- [22] K. R. Ghusinga, J. J. Dennehy, and A. Singh, *Proc. Natl. Acad. Sci. USA* **114**, 693 (2017).
- [23] M. Shreshtha, A. Surendran, and A. Ghosh, *Phys. Biol.* **13**, 036004 (2016).
- [24] K. Biswas, M. Shreshtha, A. Surendran, and A. Ghosh, *Eur. Phys. J. E: Soft Matter Biol. Phys.* **42**, 24 (2019).
- [25] A. D. Co, M. C. Lagomarsino, M. Caselle, and M. Osella, *Nucleic Acids Res.* **45**, 1069 (2016).
- [26] R. Milo, P. Jorgensen, U. Moran, G. Weber, and M. Springer, *Nucleic Acids Res.* **38**, D750 (2009).
- [27] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, *Science* **329**, 533 (2010).
- [28] <http://book.bionumbers.org/>.
- [29] G.-W. Li, D. Burkhardt, C. Gross, and J. S. Weissman, *Cell (Cambridge, MA, U. S.)* **157**, 624 (2014).
- [30] M. Jovanovic, M. S. Rooney, P. Mertins, D. Przybylski, N. Chevrier, R. Satija, E. H. Rodriguez, A. P. Fields, S. Schwartz, R. Raychowdhury *et al.*, *Science* **347**, 1259038 (2015).
- [31] B. Schwanhäusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach, *Nature (London)* **473**, 337 (2011).
- [32] J. J. Li, P. J. Bickel, and M. D. Biggin, *PeerJ* **2**, e270 (2014).
- [33] J. Hausser, A. Mayo, L. Keren, and U. Alon, *Nat. Commun.* **10**, 68 (2019).