Gene expression dynamics with stochastic bursts: Construction and exact results for a coarse-grained model

Yen Ting Lin and Charles R. Doering

Phys. Rev. E **93**, 022409 — Published 18 February 2016

DOI: 10.1103/PhysRevE.93.022409

Gene expression dynamics with stochastic bursts: Construction and exact results for a coarse-grained model

Yen Ting Lin*
Theoretical Physics Division, School of Physics and Astronomy, The University of Manchester, UK and Max Planck Institute for the Physics of Complex Systems, Dresden, Germany

Charles R. Doering†
Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109-1107 USA Department of Mathematics, University of Michigan, Ann Arbor, MI 48109-1043 USA and Department of Physics, University of Michigan, Ann Arbor, MI 48109-1040 USA

We present a theoretical framework to analyze the dynamics of gene expression with stochastic bursts. Begining with an individual-based model which fully accounts for the messenger RNA (mRNA) and protein populations, we propose a novel expansion of the master equation for the joint process. The resulting coarse-grained model reduces the dimensionality of the system, describing only the protein population while fully accounting for the effects of discrete and fluctuating mRNA population. Closed form expressions for the stationary distribution of the protein population and mean first-passage times of the coarse-grained model are derived and large-scale Monte Carlo simulations show that the analysis accurately describes the individual-based process accounting for mRNA population, in contrast to the failure of commonly proposed diffusion-type models.

PACS numbers: 02.50.Ey, 05.40.-a, 82.39.-k, 87.16.Yc

I. INTRODUCTION

Intrinsic noise originating from the discreteness of interacting particles plays an important role in genetic expression: it diversifies the distribution of protein population, promotes transition between different cellular phenotypes on a population level, and in turn enhances organisms' ability to adapt to changing environments without the need of genetic mutation [1]. There are several sources of intrinsic noise in the context of gene expression: *transcriptional bursting noise* from the stochastic transition between active and repressed states of DNA transcription, *translational bursting noise* from the relatively fast action of mRNA to produce the proteins [1, 2], and finally the *demographic noise* from the finite and discrete nature of the protein molecules. Bursts of protein production are experimentally observed to be the predominant form of intrinsic noise in gene expression dynamics [3, 4]. While many stochastic models have been proposed to model gene circuits [5–14], only a few studies quantitatively account for the effects of bursting noise [11, 12, 15, 16] at a mesoscopic level. To our knowledge, current theoretical investigation of the dynamical properties of such bursting processes is limited to stationary properties of the protein distribution on the population level [15, 16].

Our aim of this work is to construct a systematic scheme to expanding the master equations of the individual-based model while retaining the signature of the bursting noise. Starting from an individual-based model including both mRNA and protein populations we construct a novel coarse-grained process describing only the protein population dynamics that fully accounts for the discreteness effects and fluctuations in the mRNA population. When the mRNA degrades at a much shorter time scale, our proposed method nicely bridge the existing bursting models [15, 16] and the individual-based models [7, 11], and mathematically identifies that the *translational bursting noise* originates from the *demographic stochasticity* of the mRNA population dynamics. The resulting process from our proposed expanding scheme allows straightforward formulation a computations of the mean first-passage statistics [17] which quantifies the timescale of the gene regulatory network. We present analytic solutions along with computational verification from large-scale Monte-Carlo simulations of the individual-based model, our proposed approximating model, and the diffusion approximations [9, 10, 13, 14] of the dynamics. The key conclusion is that the conventional diffusion approximation of the master equation, while vastly adopted because of its well-studied mathematical properties, fails to accurately estimate the timescales of the individual-based model. Alternatively, our proposed approximating process faithfully captures the signature of the bursting dynamics and serves as a candidate of coarse-grained models.

II. INDIVIDUAL-BASED MODEL

A simple individual-based model of autoregulated gene expression including both the mRNA and protein populations contains four reaction steps [2, 15] as summarized

* yenting@umich.edu
† doering@umich.edu

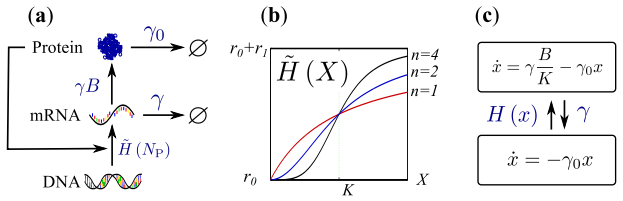

FIG. 1. (a) Schematic diagram of the individual-based model. (b) Hill functions with Hill coefficients $n = 1, 2, 4$. (c) The piecewise deterministic Markov process Eq. (4).

in Fig. 1(a):

$$\varnothing \xrightarrow{\tilde{H}(N_{\rm P})} \text{mRNA}, \quad \text{(Transcription)} \tag{1a}$$

$$\text{mRNA} \xrightarrow{\gamma B} \text{mRNA} + \text{Protein}, \quad \text{(Translation)} \tag{1b}$$

$$\text{mRNA} \xrightarrow{\gamma} \varnothing, \quad \text{(Degradation of mRNA's)} \tag{1c}$$

$$\text{Protein} \xrightarrow{\gamma_0} \varnothing. \quad \text{(Degradation of proteins)} \tag{1d}$$

We use $\varnothing$ to denote the empty set. In the first reaction, each mRNA with transcribed genetic sequence from the DNA is produced with a rate $\tilde{H}$. The mRNA is then translated to produce proteins with a rate γB. mRNA's and proteins degrade with constant rates γ and γ_0 respectively. In this autoregulated genetic circuit, proteins plays a role of transcription factors and self-regulate the transcription rate, so $\tilde{H}$ is a function of the random population of proteins $N_{\rm P}$. Specifically, we use a Hill function $\tilde{H}(X) := r_0 + r_1 X^n/(K^n + X^n)$ to model the transcription rate, where n is the Hill coefficient, which quantifies the effect of cooperative binding of the transcription factors to the DNA [2]. To illustrate the qualitative behaviour of $\tilde{H}(X)$, we plot Hill functions with $n = 1$, 2, 4 in Fig. 1(b). Finally, we remark that the implementation of the Hill function $\tilde{H}$ results from an adiabatic approximation when the gene switches between on and off states on a much shorter timescale; interested readers are referred to [2] and references therein.

We refer to the process in Fig. 1(a) as the individual-based (IB) model. Although the IB model provides a detail description of both the mRNA and protein populations, it is generally difficult to analyze theoretically except for linear cases [18, 19]. Single-species models describing only the protein populations are often adopted, especially for more complicated genetic circuits [9, 10, 13, 14]. However, fluctuations in the mRNA population are an important dynamical factor [11, 12] and our objective is to construct a coarse-grained model describing only the protein population accounting for contributions from fluctuation in the mRNA population.

Generally mRNA's degrade much faster than proteins. In the model organism *Escherichia coli* for example, the mean lifetime of the mRNA is about 2 min while protein lifetimes are $45 \sim 60$ min [15]. As a consequence, a large number of proteins is produced in a relatively short period of time—a phenomenon termed *translational bursting*. In addition, due to small system size (the volume of *E. coli* are $\sim 10^{-18}\text{m}^3$), the onset of the transcription and the lifetime of the synthesized mRNA are observed to be stochastic [20].

III. CONSTRUCTION OF A COARSE-GRAINED MODEL

Motivated by the observation of translational bursting, we propose a novel expansion to approximate the master equation of the IB process in Fig. 1(a). First, we notice that in the IB model, for any given mRNA number m, the protein population $N_P(t)$ is a birth-death processes with constant birth rate $m\gamma B$ and constant per capita death rate γ_0. Therefore, it is convenient to expand the process describing the protein dynamics *conditioning on* the mRNA population: each "state" of the system is labeled by the mRNA number m. The transition rate from state m to $m+1$ mRNA molecules is the autoregulated transcription rate $\tilde{H}(N_P)$, and the transition rate from state $m+1$ to m mRNA molecules is the mRNA degradation rate γ. Within each state of the system we perform a Kramers–Moyal expansion of the birth-death process [17, 21] with respect to the system size $K \gg 1$. In the lowest order approximation only the advection terms describing the mean-field dynamics are retained [22]. Formally letting the protein concentration be $x := N_{\rm P}/K \geq 0$ and the number of mRNA molecules be $m \in \{0, 1, 2, \ldots\}$, in each state the protein density evolves according to the deterministic equation

$$\dot{x}(t) = m\gamma \frac{B}{K} - \gamma_0 x, \tag{2}$$

with transition rates between different states

$$m \xrightarrow{H(x)} m+1 \text{ and } m \xrightarrow{\gamma m} m-1 \tag{3}$$

where $H(x) := \tilde{H}(Kx)$ is the scaled Hill-function.

Next we note that the mean lifetime of the mRNA is $\mathcal{O}(1/\gamma)$ and in the fast-degrading mRNA limit $\gamma \gg 1$, most of the time the system has either $m = 0$ or $m = 1$. We therefore neglect states $m \geq 2$ and formulate a closed forward equation for $p_m(x,t)$, the joint probability density that the system presents $m \in \{0, 1\}$ mRNA molecules and protein density x at time t:

$$\frac{\partial}{\partial t}\begin{bmatrix} p_1(x,t) \\ p_0(x,t) \end{bmatrix} = L^\dagger \begin{bmatrix} p_1(x,t) \\ p_0(x,t) \end{bmatrix}, \tag{4}$$

where the forward operator [23] is defined to be

$$L^\dagger := \begin{bmatrix} -\gamma - \partial_x(\gamma b - \gamma_0 x) & H(x) \\ \gamma & -H(x) + \partial_x f(x) \end{bmatrix} \tag{5}$$

where we have defined $b := B/K$ to be a dimensionless parameter characterizing the strength of the bursts, and $f(x) := \gamma_0 x$. A more rigorous derivation of the model (4) can be found in Appendix A. We shall refer to (4) as the piecewise deterministic Markov process (PDMP [24, 25]; schematic diagram Fig. 1(c)) and remark that the process in x alone is non-Markovian [25].

IV. STATIONARY DISTRIBUTION OF THE COARSE-GRAINED MODEL

To proceed with our analytic investigation, an infinitely fast-degrading mRNA limit $\gamma \to \infty$ is taken. Although in such a limit the mean duration when the system stays in $m = 1$ state is $1/\gamma \to 0$, the protein concentration in $m = 1$ state increases with a rate $b\gamma \to \infty$, preserving exponentially distributed random burst with an average burst strength b. In this limit the process stays in $m = 0$ state almost surely (i.e. $p_1(x,t) \to 0\ \forall t$), and the probability distribution p_0 satisfies a closed and second-order differential equation

$$(1 + b\partial_x)\,\partial_t p_0 = -\partial_x \left[-x + bH(x) - b\partial_x f(x)\right] p_0. \quad (6)$$

The stationary probability distribution is obtained by direct integration:

$$p_{\text{stat}}(x) = \frac{\mathcal{N}}{f(x)} \exp\left\{\frac{-x}{b} + \int^x \frac{H(\xi)d\xi}{f(\xi)}\right\} \quad (7)$$

where $\mathcal{N}$ is the normalization factor. Substituting the explicit form of the Hill function we find the analytic expression for the stationary distribution

$$p_{\text{stat}}(x) = \frac{\mathcal{N}}{\gamma_0} e^{-\frac{x}{b}} x^{\frac{r_0}{\gamma_0}-1} \left(x^n + 1^n\right)^{\frac{r_1}{n\gamma_0}}. \quad (8)$$

We remark that in the limit $\gamma \to \infty$ the PDMP model reduces to the bursting model described by a continuous master equation (presented in Appendices B and C), and that this result confirms [16].

V. MEAN FIRST-PASSAGE TIME

At this stage, the PDMP formulation provides a mathematical relation bridging the previously bursting models [15, 16] and the individual-based model [7, 11]: the mesoscopic bursting models are the "continuum limit" of the indidvidual-based models, assuming that (a) the system size of the the protein population is infinite $K \to \infty$ and (b) the lifetime of the mRNA is random and infinitely fast and consequently the proteins are synthesized in a stochastic and bursting fashion.

In some parameter regimes the stationary distribution (8) exhibits bi-stability [26] and can be adopted to model a biological switch [11, 16]. Our formulation (4) can be used to derive the mean switching time (MST) between two modes of gene expression in a straightforward way [27]. We begin by deriving the mean first-exit time to leave a domain (x_1, x_2) where $0 < x_1 < x_2 < \infty$.

If the initial protein concentration is $x \in (x_1, x_2)$ and the initial number of mRNA is m, then the mean time to exit the domain $T_m(x)$ satisfies the inhomogeneous equation [17]

$$-\begin{bmatrix}1\\1\end{bmatrix} = L\begin{bmatrix}T_1(x,t)\\T_0(x,t)\end{bmatrix}, \quad (9)$$

where the generator L is the adjoint of the forward operator in (5),

$$L = \begin{bmatrix}-\gamma + \left[\gamma b - f(x)\right]\partial_x & \gamma\\ H(x) & -H(x) - f(x)\,\partial_x\end{bmatrix} \quad (10)$$

with boundary conditions

$$T_1(x_2) = 0,\, T_0(x_1) = 0. \quad (11)$$

The physical meaning of the boundary conditions is clear: when the system starts with the state $m = 1$—a state with fast production of proteins—at upper boundary x_2, and when the state $m = 0$—a state with only degrading proteins—at lower boundary x_1, immediately the flow leaves the domain (x_1, x_2).

Taking the limit $\gamma \to \infty$ we deduce a closed second-order differential equation for T_0,

$$-T_0'' - \left[\frac{H}{f} - \frac{1}{b} + \frac{H}{x}\left(\frac{x}{H}\right)'\right] T_0' = \frac{1}{bf} + \frac{H'}{fH} \quad (12)$$

where prime denote derivative with respect to x. The boundary conditions for (12) follow from (11):

$$T_0(x_1) = 0, 1 = H(x_2)\,T_0(x_2) + \gamma_0 x_2 T_0'(x_2). \quad (13)$$

We remark that while formally deriving the backward equations of the bursting models [15, 16] considering only the protein concentration is possible, imposing the correct boundary conditions (13) is not trivial.

The solution (derived in the Appendix E) is

$$T_0(x) = C\int_{x_1}^x e^{-M(y)}dy + \int_{x_1}^x e^{-M(y)}V(y)dy, \quad (14)$$

where the auxiliary functions $M(x)$, $V(x)$, and the constant C using $f(x) = \gamma_0 x$ are

$$M(x) := \int^{x} \left[\frac{H(y)}{\gamma_0 y} - \frac{1}{b} + \frac{d}{dy}\left(\ln \frac{y}{H(y)} \right) \right] dy, \tag{15}$$

$$V(x) := \int^{x} \left[\frac{-1}{b\gamma_0 y} + \frac{-1}{\gamma_0 y H(y)} \frac{dH(y)}{dy} \right] e^{M(y)} dy, \tag{16}$$

$$C \equiv \left[-V(x_2) e^{-M(x_2)} \gamma_0 x_2 - H(x_2) \int_{x_1}^{x_2} V(y) e^{-M(y)} dy + 1 \right] \left[\gamma_0 x_2 e^{-M(x_2)} + H(x_2) \int_{x_1}^{x_2} e^{-M(y)} dy \right]^{-1}. \tag{17}$$

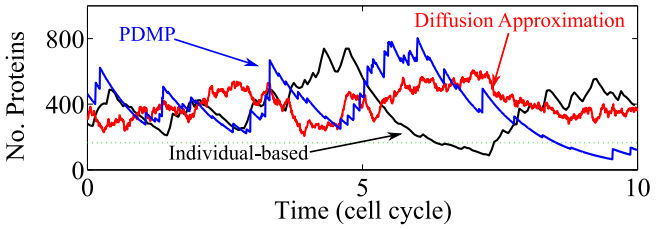

FIG. 2. Sample paths of the models. Dotted green line denotes 165 molecules which separates low and high protein abundance mode.

This solution is a generalization of results in [28, 29].

The exact solution (14) resembles the form of the mean first passage times of widely studied diffusion-type processes [17, 30]. For the diffusion approximation the integrand of $M(x)$ would be simply replaced by the dirft divided by the (possibly multiplicative) diffusion—a measure of the relative strength of the "dissipation" to "fluctuations", and the function $M(x)$ is often implicitly interpreted as a "potential" of the process.

When the system exhibits bi-modality, the mean switching times between two modes of protein expression can be obtained by taking appropriate limits of (14). First, we define a critical density x_c separating the low and high protein abundance modes, then take $x_1 \to 0$ and $x_2 \to x_c$ for the low mode, and $x_1 \to x_c$ and $x_2 \to \infty$ for the high mode. Careful analysis is needed because (12) is singular at $x = 0$ and $x \to \infty$ (and is presented in Appendices E 1 and E 2). The analytic expressions for the mean switching times (MSTs) are

$$T_{\text{low}\to\text{high}} \equiv \int_0^x e^{-M(y)} V(y) dy + C_2, \tag{18}$$

$$T_{\text{high}\to\text{low}} \equiv \int_{x_c}^x e^{-M(y)} \left[V(y) - V(\infty) \right] dy, \tag{19}$$

where the constant C_2 is

$$C_2 := \frac{1 - x_c V(x_c) e^{-M(x_c)}}{H(x_c)} - \int_0^{x_c} e^{-M(y)} V(y) dy. \tag{20}$$

VI. COMPARISON TO THE DIFFUSION APPROXIMATION

We now turn to the diffusion approximation (DA) of the IB process. To our knowledge there is no standard way to derive DA models for general bursting kernels. In Appendix D we present the straightforward Kramers–Moyal expansion [17, 21] of the master equation of the IB process in the limit $\gamma \to \infty$ and $B \gg 1$ yielding the Itō stochastic differential equation

$$dX_t = \left[bH(X_t) - \gamma_0 X_t \right] dt + \sqrt{\Gamma b^2 H(x)}\, dW_t \tag{21}$$

where X_t is the random population density of the proteins, W_t is the standard Wiener process and the scaling factor $\Gamma = 2$. An alternative and phenomenological construction the diffusion approximation is to insert the mean and variance of the bursting kernel in the individual-based process (see Appendix D) which yields (21) with the scaling factor $\Gamma = 1$. To avoid leaking probabilities to negative densities, we put a reflective boundary at the origin $x = 0$. Analytic expressions for the stationary distribution and the mean switching times of the diffusion equation are derived by standard analysis [17].

We performed numerical simulations to measure the stationary distributions and the mean switching times (MST) in all three models to verify the theoretical analysis. For the IB model, exact sample paths are generated by standard continuous time Markov chain simulations [31]. For the PDMP model, kinetic Monte Carlo simulations can be constructed by generating exact random waiting times to the next transition events [32]. In the limit $\gamma \to \infty$, we adopted a previously proposed algorithm [33]. For the diffusion approximations we construct a standard Euler–Maruyama integrator of (21). One sample path of each of the models are presented in Fig. 2.

The parameters were chosen to be in a biologically relevant regime [15, 16, 34]: $K = 200$, $n = 4$, $B = 40$, $r_0 = 2$, $r_1 = 10$, $\gamma = 30$, while $\gamma_0 := 1$ is chosen to normalize the unit of the time by a natural cell cycle.

Fig. 3 presents the stationary probability distributions of the IB, PDMP, and DA models. Note that the low-mode is noise induced and does not exist in the mean-field dynamics (top panel of Fig 3). While the PDMP model captures the stationary distribution of the IB model ex-

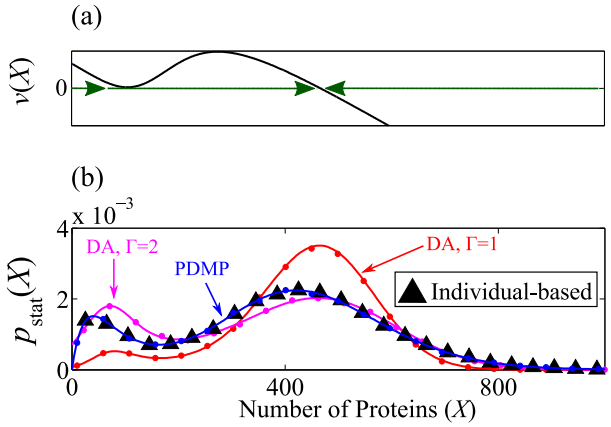

FIG. 3. (a) Drift of the mean-field dynamics $\dot{x} = bH(x) - \gamma_0 x$ showing a single fixed point. (b) Stationary probability distributions of individual-based model, piecewise deterministic Markov process (PDMP), and the diffusion approximations (DA). Solid lines are analytic solutions. Discrete markers represent numerically measured probability distributions from Monte Carlo simulations.

tremely well, directly expanding the IB stochastic bursting model by Kramers–Moyal expansion (DA with $\Gamma = 2$) qualitatively captures the stationary distribution, and the phenomenological DA model with $\Gamma = 1$ failed to capture the stability of the low mode.

Fig. 4 presents the MST between low and high protein-abundance modes in all three models. Again, the PDMP model well estimates the mean switching times of the IB model, and both the DA models fail by a large amount. When the state is initially below x_c, both the DA models under-estimate the transition time because the bursting kernels of the DA model have a thinner (Gaussian) tail compared to to the geometric bursting kernel of the IB model. When the initial state is above x_c, the DA model with $\Gamma = 1$ over-estimate the MST because the the approximation does not capture the high probabilities of low-density bursts, and the DA model with $\Gamma = 2$ underestimate the MST because the approximation fail to capture that the bursting kernel is always positive. Due to the nontrivial boundary conditions (13), as $x \nearrow x_c$, $T_{\text{low}\to\text{high}}(x)$ does not converge to 0 as the solutions of the Markovian diffusion-type models do. We highlight that the non-Markovian PDMP formulation faithfully captures this feature of the IB model.

VII. DISCUSSION AND OUTLOOK

The PDMP approximation works well even for models with a strong noise strength. In our example, the low-mode is of an order of 100 protein molecules, and the noise strength (per each burst) is of order 40 protein molecules. In addition, the PDMP approximation performs well even though an infinitely-fast degrading mRNA limit $\gamma \to \infty$ is taken and consequently almost

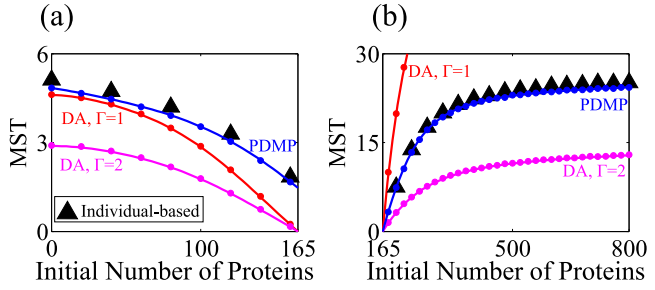

FIG. 4. Mean switching times (MST) of the individual-based model, piecewise deterministic Markov process (PDMP), and the diffusion approximations (DA). Solid lines are analytic predictions, and discrete markers are measured from Monte Carlo simulations. (a) $T_{\text{low}\to\text{high}}$; (b) $T_{\text{high}\to\text{low}}$. $x_c := 165/K$.

surely there is no mRNA presented in the system. Meanwhile for a finite $\gamma = 30$, we observe an average 0.3188 mRNA in the stationary state of the IB model. Fixing other parameters, we observe that the bursting PDMP (i.e., $\gamma \to \infty$) provides relatively good estimates when $\gamma \geq 10$, while the diffusion approximations systematically fail to capture the stationary distribution and the MST (see Appendex F).

The PDMP model can be easily generalized. For example, finite population and lifetime of mRNA can be considered by generalize (4) to include p_m with $m \in \{0, 1, 2, \ldots\}$. Models with further downstream protein reactions [8, 35] can be modeled by including more subpopulations of the proteins. These generalizations merit future investigations. We remark that an investigation of higher dimensional genetic circuits along with the sensitivity analysis of the parameters B and K have been performed and published separately [36].

We conclude that bursting originating from the discreteness of the fast-living mRNA molecules and the stochastic transcription events is the dominating noise in individual-based autoregulated gene expression model. The key finding in this study is that while the conventional diffusion-type models qualitatively approximate the stationary distributions of the individual-based model, they are no longer adequate to analyze the dynamical properties of bursting systems while the novel expansion described here faithfully captures the dynamical properties of the individual-based model in a biologically realistic parameter regime and serves as a new analytic tool to investigate more complex models with bursting noise.

ACKNOWLEDGEMENT

YTL was supported by the visitors program at MPI-PKS and EPSRC (grant reference EP/K037145/1). CRD was supported in part by US-NSF Awards PHY-1205219 and DMS-1515161, and as Simons (Foundation) Fellow in Theoretical Physics.

Appendix A: Construction of the PDMP

The master equation of process (1) is

$$\begin{aligned}\dot{P}_{m,n} = & -\left[\tilde{H}(n)+\gamma Bm+\gamma m+\gamma_0 n\right]P_{m,n} \\ & +\gamma\left(m+1\right)P_{m+1,n}+\tilde{H}(n)P_{m-1,n} \\ & +\gamma bmP_{m,n-1}+\gamma_0\left(n+1\right)P_{m,n+1}. \end{aligned} \tag{A1}$$

In the fast degrading mRNA limit ($\gamma/\gamma_0 \gg 1$), the system only presents only 1 or 0 mRNA in a majority portion of the time. Our proposed approximation is to consider the process (1) conditioning on whether or not the system presents an mRNA, and truncate the probabilities associated with mRNA number greater than 1:

$$P_{m,n} = 0,\ \forall m > 1, \tag{A2a}$$

$$\begin{aligned}\dot{P}_{1,n} = & -\left[\gamma B+\gamma+\gamma_0 n\right]P_{1,n}+\tilde{H}(n)P_{0,n} \\ & +\gamma BP_{1,n-1}+\gamma_0\left(n+1\right)P_{1,n+1}, \end{aligned} \tag{A2b}$$

$$\begin{aligned}\dot{P}_{0,n} = & -\left[\tilde{H}(n)+\gamma_0 n\right]P_{0,n}+\gamma P_{1,n} \\ & +\gamma_0\left(n+1\right)P_{1,n+1}. \end{aligned} \tag{A2c}$$

Next, for each of the master equations of the protein number n conditioning on the mRNA number m, we perform the conventional Kramers–Moyal expansion [17]. Denote a typical population scale of the protein by $N_\Omega \gg 1$. Note that in the autocorrelated circuit, it is convenient to choose $N_\Omega = K$. In the continuum limit, the population density is defined by $x := n/K$, and the mean "burst" size is defined is $b := B/N_\Omega$. The evolution of the probability distributions $p_0(x,t) := P_{0,n}(t)/K$ and $p_1(x,t) := P_{1,n}(t)/K$ is well-approximated by two coupled Fokker–Planck equations [21], which are expressed in a compact matrix form:

$$\partial_t\begin{pmatrix}p_1\\p_0\end{pmatrix} = L^\dagger\begin{pmatrix}p_1\\p_0\end{pmatrix}, \tag{A3}$$

with

$$\left(L^\dagger\right)_{11} := -\gamma+\partial_x\left(\gamma_0 x-\gamma b\right)+\frac{1}{2K}\partial_x^2\left(\gamma_0 x+\gamma b\right) \tag{A4a}$$

$$\left(L^\dagger\right)_{12} := H\left(x\right), \tag{A4b}$$

$$\left(L^\dagger\right)_{21} := \gamma, \tag{A4c}$$

$$\left(L^\dagger\right)_{22} := -H\left(x\right)+\gamma_0\partial_x x+\frac{\gamma_0}{2K}\partial_x^2 x. \tag{A4d}$$

We again remind the reader that the differential operators ∂_x and ∂_x^2 act on $p_{0,1}$ too.

It should be clear that the the *discrete population* of the *proteins* causes the demographic stochasticity, which is described by those terms with a prefactor $1/K$. We further propose to take the limit $K \to \infty$ [21] and leave only the advection terms in (A4) to consider exclusively the **bursting noise**, which is a result of *randomly production and degradation events of the mRNA*. In such a limit, the process becomes a piecewise deterministic Markov process: in each state of $m = 0$ or $m = 1$, the process is deterministic but the switching between the states is Markovian. We emphasize that, in such a limit, the demographic noise which comes from the discreteness of the protein population does not exist—condition on a m state, the concentration of the protein on its own is always evolving in a deterministic fashion.

Appendix B: Individual-Based Bursting Model

Back to process (1), when $\gamma \gg \gamma_0$ and $\tilde{H}(N_{\text{Protein}})$, there is a time-scale separation and the mRNA's degrade at a very rapid rate. As a consequence, when one mRNA is formed, almost surely the next happening events before its final degradation are the even more rapid productions of proteins.

Due to the time-scale separation, the production of other mRNA's is and the protein degradation are negligible in one mRNA's lifetime. In such a limit, the distribution of the total number of proteins an mRNA could ever synthesized before its final degradation can be computed. Define the total number of proteins an mRNA could ever synthesize to be N_Σ, a non-negative random variable. Because the mRNA has only have two choices—either to degrade or to produce a protein—at any time before the final degradation, the probability that the mRNA produce a protein is $B/(1+B)$ from reading off the ratio of the rates in the process (1). Therefore, the distribution of N_Σ is a geometric distribution

$$\mathbb{P}\left\{N_\Sigma = n\right\} \equiv \left(\frac{B}{1+B}\right)^n\left(\frac{1}{1+B}\right). \tag{B1}$$

As a consequence, process (1) in the limit $\gamma \to \infty$ can be re-formulated to neglect the mRNA population

$$\begin{aligned} \varnothing &\longrightarrow N_\Sigma \times \text{Protein} && \text{with a rate } \tilde{H}\left(N_{\text{P}}\right), \\ \text{Protein} &\longrightarrow \varnothing && \text{with a rate } \gamma_0, \\ N_\Sigma+1 &\sim \text{Geom}\left(B/\left(1+B\right)\right). \end{aligned} \tag{B2}$$

We remind the reader that the parameter B is the mean number of the proteins an mRNA can produce. We shall refer to model (B2) as the "individual-based bursting model".

Let P_n to be the probability when the system has exactly n proteins. The master equation of process (B2) can be derived

$$\begin{aligned}\dot{P}_n = & -\left[\tilde{H}(n)+\gamma_0 n\right]P_n+\gamma_0\left(n+1\right)P_{n+1} \\ & +\sum_{m=0}^{n}\tilde{H}(m)\left(\frac{B}{1+B}\right)^{n-m}\frac{1}{1+B}P_m. \end{aligned} \tag{B3}$$

Appendix C: Equivalence between the PDMP and continuous state bursting models

We now apply Kramers-Moyal system-size expansion is performed *only to* the degradation dynamics in (B3). The expansion of (B3) yields

$$\partial_t p(x,t) = \left[\gamma_0 \partial_x x + \frac{\gamma_0}{2K}\partial_x^2 x\right] p(x,t) + \int_0^x W(x-y)H(y)p(y)dy, \tag{C1}$$

where $p(x,t) := P_n(t)/N_\Omega$ is the continuum probability distribution, $x := n/N_\Omega$ is the population density of the protein, and $W(x-y)$ is a kernel of the bursting process, defined by the approximating the discrete by the trapezoid rule:

$$\int_0^x W(x-y)f(y)dy := -f(x) + \frac{1}{2}\left(\frac{f(0)}{1+bK} + \frac{f(x)}{1+bK}\right) + \int_0^x \frac{1}{1/K+b} e^{-bN_\Omega(x-y)\log\left(1+\frac{1}{bN_\Omega}\right)} f(y)dy. \tag{C2}$$

In the infinity population limit $K \to \infty$, (C1) reduces to

$$\partial_t p(x,t) = \partial_x \left[x p(x,t)\right] + \int_0^x \frac{e^{-\frac{x-y}{b}}}{b} H(y)p(y)dy - H(x)p(x), \tag{C3}$$

which is exactly the continuous master equation in Friedman *et al.* [16] It is straightforward to establish the equivalence of this model to the the piecewise deterministic Markov process: acting an operator $1 + b\partial_x$ to (C3) and sending $K \to \infty$, we arrive at the forward equation(6).

Appendix D: Derivation of the diffusion approximation

Often in higher dimensional systems, diffusion processes are adopted to analyze complex genetic circuits [9, 10, 13, 14]. This section present the derivation to the diffusion approximation of the process (B2).

In the fast-degrading mRNA limit [i.e. $\gamma \to \infty$ in (1)], diffusion approximation can be obtained by by performing Kramers–Moyal expansion to (B2). The corresponding Fokker–Planck approximation reads

$$\partial_t p_n(x) = -\partial_x\left[\left(H(x)\frac{\mathbb{E}\left[N_\Sigma\right]}{K} - \gamma_0 x\right) p_n\right] + \frac{1}{2}\partial_x^2\left[\left(H\left(x\right)\frac{\mathbb{E}\left[\left[N_\Sigma^2\right]\right]}{K^2} + \frac{\gamma_0 x}{K^2}\right) p_n\right] \tag{D1}$$

N_Σ is a geometric distribution, and the exact expression of the first two moments are

$$\mathbb{E}\left[N_\Sigma\right] = B \tag{D2a}$$

$$\mathbb{E}\left[N_\Sigma^2\right] = B\left(1+2B\right). \tag{D2b}$$

When the bursting number is large $B \gg 1$ (typical biological value $10^1 \sim 10^2$ in *E. Coli* [15, 34]), we arrive at the final diffusion approximation of (B2):

$$\partial_t p_n(x) = -\partial_x\left[\left(bH(x) - \gamma_0 x\right) p_n\right] + \frac{1}{2}\partial_x^2\left[\left(2b^2 H\left(x\right) + \frac{\gamma_0 x}{K}\right) p_n\right]. \tag{D3}$$

We finally remark that in the large population limit, b scales $\mathcal{O}\left(K^{-1}\right)$. A sensible population scaling suggests that near the mean-field fixed points, $\mathcal{O}\left(bH(x)\right) = \mathcal{O}\left(\gamma_0 x\right) = \mathcal{O}\left(K^0\right)$, which in turn indicates that $\mathcal{O}\left(H(x)\right) = \mathcal{O}\left(K\right)$. It is clear that the diffusion term can then be simplified if we *neglect demographic noise* due to the protein degradation, when $B \gg 1$.

$$\partial_t p_n(x) = -\partial_x\left[\left(bH(x) - \gamma_0 x\right) p_n\right] + \partial_x^2\left[\left(b^2 H\left(x\right)\right) p_n\right], \tag{D4}$$

or equivalently the Itō stochastic differential equation

$$dX_t = \left[bH(X_t) - \gamma_0 X_t\right] dt + \sqrt{2b^2 H\left(x\right)} dW_t, \tag{D5}$$

where dW_t is the Wiener process.

In the main text, we refer the diffusion process (D4) to be the diffusion approximation of process (1) in the fast-degrading mRNA limit ($\gamma \to \infty$), with a reflective boundary at $x = 0$.

A more phenomenological way is to assert the drift and the diffusion terms of the stochastic differential equation to be the mean (b) and variance (b^2) of the exponentially distributed burst size. It can be shown [36] that this approach corresponds to a *constant-burst* model. In this case, the stochastic differential equation is clearly

$$dX_t = \left[bH(X_t) - \gamma_0 X_t\right] dt + \sqrt{b^2 H\left(x\right)} dW_t \tag{D6}$$

Finally, we remark that (D4) is derived from expanding the master equation (B2) where effect that the bursting noise only enhance the population of the proteins, and (D5) clearly over-estimate the noise. On the other hand, the phenomenological approach (D6) clearly under-estimate the low-density bursts of the exponentially distributed kernel.

Appendix E: Solving the backward equation

On the domain $\Omega := \{x : 0 < x_1 \le x \le x < \infty\}$, the backward equation reads

$$-\begin{pmatrix} 1 \\ 1 \end{pmatrix} = L \begin{pmatrix} T_1(x,t) \\ T_0(x,t) \end{pmatrix}, \tag{E1}$$

where L is defined in (A4). In the limit with fast-degrading mRNA $\gamma \to \infty$, it converges to

$$-\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 + b\partial_x & 1 \\ H(x) & -H(x) - f(x)\partial_x \end{pmatrix} \begin{pmatrix} T_1(x,t) \\ T_0(x,t) \end{pmatrix}, \tag{E2}$$

where we denote $\gamma_0 x$ by $f(x)$. It is elementary to eliminate the variable T_1 and obtain

$$\frac{d^2T_0}{dx^2}+\left[\frac{H}{f}-\frac{1}{b}+\frac{H}{x}\frac{d}{dx}\left(\frac{x}{H}\right)\right]\frac{dT_0}{dx}=-\left(\frac{1}{bf}+\frac{1}{fH}\frac{dH}{dx}\right). \tag{E3}$$

Define an auxiliary function $M(x)$ and $V(x)$

$$M(x):=\int^x\left[\frac{H(x')}{f(x')}-\frac{1}{b}+\frac{H(x')}{x'}\frac{d}{dx'}\left(\frac{x'}{H(x')}\right)\right]dx', \tag{E4}$$

$$V(x):=-\int^x\left(\frac{1}{bf(x')}+\frac{1}{f(x')H(x')}\frac{dH}{dx}\right)e^{M(x')}dx'. \tag{E5}$$

With the expression $H(x):=r_0+r_1x^n/(x^n+k^n)$ and $f(x):=\gamma_0 x$, $M(x)$ has a closed form

$$M(x):=\log\left[e^{-\frac{x}{b}}x^{\frac{r_0}{\gamma_0}+1}\left(x^n+k^n\right)^{\frac{r_1}{n\gamma_0}}\frac{1}{r_0+\frac{r_1x^n}{x^n+k^n}}\right]. \tag{E6}$$

Now (E3) can be expressed as

$$\frac{d}{dx}\left(e^M(x)\frac{dT_0}{dx}\right)=-\left(\frac{1}{b\gamma_0x}+\frac{1}{\gamma_0xH(x)}\frac{dH}{dx}\right)e^{M(x)}, \tag{E7}$$

and the formal solution is

$$T_0\left(x\right)=C_0+C_1\int_{x_1}^x e^{-M(x')}dx'+\int_{x_1}^x e^{-M(x')}V(x')dx' \tag{E8}$$

With two constants of integration C_0 and C_1.

Since $T_0(x_1)=0$, clearly $C_0=0$. The second constant C_1 can be determined by the second boundary condition

$$T_1(x_2)=0\iff -1=-H(x_2)T_1(x_2)-f(x_2)\frac{dT_1}{dx}(x_2). \tag{E9}$$

After some algebra, we arrive at

$$C_1=\frac{-V(x_2)e^{-M(x_2)}-\frac{1}{f(x_2)}\left[H(x_2)\int_{x_1}^{x_2}V(x')e^{-M(x')}dx'-1\right]}{e^{-M(x_2)}+\frac{H(x_2)}{f(x_2)}\int_{x_1}^{x_2}e^{-M(x')}dx'.} \tag{E10}$$

For the mean switching times between the high- and the low-concentration mode, we have to impose either $x_1\to 0$ (for initial state in the low mode) or $x_2\to\infty$ (for the initial state in the high mode).

1. Limiting case: $x_1\to 0$

Note that $\exp\left[-M(x)\right]$ has a singularity at $x=0$. However, $V(x)$ is a well-behave function near $x=0$, and we claim

$$\int_0^x e^{-M(y)}V(y)dx'<\infty. \tag{E11}$$

To show this, first let $0<\epsilon\ll 1$, and

$$\int_0^x e^{-M(y)}V(y)dy=\int_0^\epsilon e^{-M(y)}V(y)dy+\int_\epsilon^x e^{-M(y)}V(y)dy. \tag{E12}$$

The second term is bounded. As for the first term, since $y<\epsilon\ll 1$, we have

$$e^{-M(y)}=\frac{e^{\frac{y}{b}}}{y^{\frac{r_0}{\gamma_0}+1}\left(y^n+k^n\right)^{\frac{r_1}{n\gamma_0}}}\left(r_0+\frac{r_1y^n}{y^n+k^n}\right)\le\frac{e^{\frac{\epsilon}{b}}\left(r_0+r_1\right)}{k^{\frac{r_1}{\gamma_0}}}\frac{1}{y^{\frac{r_0}{\gamma_0}+1}}=:\frac{B_1}{y^{\frac{r_0}{\gamma_0}+1}} \tag{E13}$$

with a constant $B_1<\infty$. Similarly,

$$e^{M(y)}\le y^{\frac{r_0}{\gamma_0}+1}\frac{\left(\epsilon^n+k^n\right)^{\frac{r_1}{n\gamma_0}}}{r_0}=:B_2y^{\frac{r_0}{\gamma_0}+1} \tag{E14}$$

with a constant $B_2<\infty$. Similarly, $V(y)$ can be bounded:

$$V(y)\equiv\int^y\left(\frac{1}{b\gamma_0z}+\frac{1}{\gamma_0zH(z)}\frac{dH(z)}{dz}\right)e^{M(z)}dz\le B_3\int^y z^{\frac{r_0}{\gamma_0}}dz=\frac{B_3}{\frac{r_0}{\gamma_0}+1}y^{\frac{r_0}{\gamma_0}}, \tag{E15}$$

with some $B_3 < \infty$. Finally, we have

$$\int_0^\epsilon e^{-M(y)}V(y)dy \le \frac{B_1 B_3}{\frac{r_0}{\gamma_0}+1}\int_0^\epsilon dy < \infty, \tag{E16}$$

which establishes our claim (E11).

Next, we proceed to show the following statements:

$$\lim_{x_1\to 0}\left[\int_{x_1}^{x_2} e^{-M(x)}dx\right]^{-1} = 0, \text{ and } \lim_{x_1\to 0}\frac{\int_{x_1}^{x_2} e^{-M(x)}dx}{\int_{x_1}^{x_3} e^{-M(x)}dx} = 1 \tag{E17}$$

for any $x_3 < x_2$. To show this, again we separate the integral

$$\int_{x_1}^{x} e^{-M(y)}dy = \int_{x_1}^{\epsilon} e^{-M(y)}dy + \int_{\epsilon}^{x} e^{-M(y)}dy. \tag{E18}$$

The second term is again bounded, and for simplicity define

$$B_5 := \int_{\epsilon}^{x} e^{-M(y)}dy < \infty. \tag{E19}$$

As for the first term, we begin with the lower bound of the integrand:

$$e^{-M(x)} \ge \frac{r_0}{y^{\frac{r_0}{\gamma_0}+1}\left(\epsilon^n + k^n\right)^{\frac{r_1}{n\gamma_0}}} =: B_6\frac{1}{y^{\frac{r_0}{\gamma_0}+1}}. \tag{E20}$$

As a consequence,

$$\int_{x_1}^{\epsilon} e^{-M(y)}dy \ge \frac{B_6\gamma_0}{r_0}\left(\frac{1}{x_1^{\frac{r_0}{\gamma_0}}} - \frac{1}{\epsilon^{\frac{r_0}{\gamma_0}}}\right), \tag{E21}$$

and finally

$$\lim_{x_1\to 0}\left[\int_{x_1}^{x_2} e^{-M(y)}dy\right]^{-1} \le \lim_{x_1\to 0}\frac{r_0}{B_6\gamma_0}\frac{x^{\frac{r_0}{\gamma_0}}}{x^{\frac{r_0}{\gamma_0}} + \epsilon^{\frac{r_0}{\gamma_0}}} = 0. \tag{E22}$$

Similarly, it is straightforward to apply the L'Hôpital's law to show

$$\lim_{x_1\downarrow 0}\frac{\int_{x_1}^{x_2} e^{-M(x)}dx}{\int_{x_1}^{x_3} e^{-M(x)}dx} = 1. \tag{E23}$$

To sum up, upon taking the limit $x_1 \to 0$, the solution (E8)—the mean switching time if the system starts with the low-protein-abundance mode—can be expressed as

$$T_0(x) \equiv \lim_{x_1\to 0} T_0(x) = \int_0^x e^{-M(x')}V(x')dx' + T_0(0) \tag{E24}$$

with

$$\begin{aligned} T_0(0) &= \lim_{x_1\to 0} C_1\int_{x_1}^{x} e^{-M(x')}dx' = \lim_{x_1\to 0}\frac{-V(x_2)e^{-M(x_2)} - \frac{1}{f(x_2)}\left[H(x_2)\int_{x_1}^{x_2}V(x')e^{-M(x')}dx' - 1\right]}{\frac{e^{-M(x_2)} + \frac{H(x_2)}{f(x_2)}\int_{x_1}^{x_2}e^{-M(x')}dx'}{\int_{x_1}^{x}e^{-M(x')}dx'}} \\ &= \frac{-V(x_2)e^{-M(x_2)} - \frac{1}{f(x_2)}\left[H(x_2)\int_{x_1}^{x_2}V(x')e^{-M(x')}dx' - 1\right]}{\frac{H(x_2)}{f(x_2)}} \\ &= \frac{1}{H(x_2)}\left[1 - x_2V(x_2)e^{-M(x_2)}\right] - \int_0^{x_2} e^{-M(x')}V(x')dx' \end{aligned} \tag{E25}$$

2. Limiting case:$x_2 \to \infty$

Note that $\exp[-M(x)]$ has a singularity at $x \to \infty$. On the other hand, $V(x)$ is again well-behaving. Rewrite

$$C_1 = \frac{-V(x_2) - \frac{1}{f(x_2)} e^{M(x_2)} \left[H(x_2) \int_{x_1}^{x_2} V(x') e^{-M(x')} dx' - 1 \right]}{1 + e^{M(x_2)} \frac{H(x_2)}{f(x_2)} \int_{x_1}^{x_2} e^{-M(x')} dx',} \tag{E26}$$

and we aim to show that $\lim_{x_2 \to \infty} C_1 = -\lim_{x \to \infty} V(x) < \infty$.
First, we show that for a given and strictly positive x_1,

$$\lim_{x_2 \to \infty} e^{M(x_2)} \int_{x_1}^{x_2} e^{-M(y)} dy = 0. \tag{E27}$$

Clearly from the exact expression of $M(x_2)$, we have

$$e^{M(x_2)} \int_{x_1}^{x_2} e^{-M(y)} dy = \int_{x_1}^{x_2} e^{-(x_2 - y)} \frac{x_2^{\frac{r_0}{\gamma_0}+1} \left(x_2^n + k^n\right)^{\frac{r_1}{n\gamma_0}} \frac{1}{r_0 + \frac{r_1 x_2^n}{x_2^n + k^n}}}{y^{\frac{r_0}{\gamma_0}+1} \left(y^n + k^n\right)^{\frac{r_1}{n\gamma_0}} \frac{1}{r_0 + \frac{r_1 y^n}{y^n + k^n}}} dy \tag{E28}$$

$$\leq \left(1 + \frac{r_1}{r_0}\right) \left(1 + \frac{k^n}{x_2^n}\right)^{\frac{r_1}{n\gamma_0}} \int_{x_1}^{x_2} e^{y - x_2} \left(\frac{x}{y}\right)^{\frac{r_0}{\gamma_0} + \frac{r_1}{\gamma_0} + 1} dy. \tag{E29}$$

For $x_2 \gg x_1$, it is elementary to show that the integrand has a single maxima at $y_* := r_0/\gamma_0 + r_1/\gamma_0 + 1 < \infty$. As a consequence,

$$e^{M(x_2)} \int_{x_1}^{x_2} e^{-M(y)} dy \leq \left(1 + \frac{r_1}{r_0}\right) \left(1 + \frac{k^n}{x_2^n}\right)^{\frac{r_1}{n\gamma_0}} e^{y_* - x_2} \left(\frac{x_2}{y_*}\right)^{\frac{r_0}{\gamma_0} + \frac{r_1}{\gamma_0} + 1} \int_{x_1}^{x_2} dy \tag{E30a}$$

$$= \left(1 + \frac{r_1}{r_0}\right) \left(1 + \frac{k^n}{x_2^n}\right)^{\frac{r_1}{n\gamma_0}} e^{y_* - x_2} \left(\frac{x_2}{y_*}\right)^{\frac{r_0}{\gamma_0} + \frac{r_1}{\gamma_0} + 1} (x_2 - x_1), \tag{E30b}$$

and apparently,

$$\lim_{x_2 \to \infty} e^{M(x_2)} \int_{x_1}^{x_2} e^{-M(y)} dy = 0. \tag{E31}$$

Note that $V(x)$ is a strictly decreasing and well-behaved function, so $0 > V(x) > V_{\min} := \lim_{x \to \infty} V(x)$ for $x_1 < x < \infty$. As a consequence,

$$\lim_{x_2 \to \infty} \left| e^{M(x_2)} \int_{x_1}^{x_2} V(y) e^{-M(y)} dy \right| \leq |V_{\min}| \lim_{x_2 \to \infty} e^{M(x_2)} \int_{x_1}^{x_2} e^{-M(y)} dy = 0. \tag{E32}$$

The above Eqs. (E31) and (E32) suggest that

$$\lim_{x_2 \to \infty} C_1 \to |V_{\min}|, \tag{E33}$$

and the final solution of the mean switching time if the system begin with a high-protein-abundance state is

$$T_0(x) = |V_{\min}| \int_{x_1}^{x} e^{-M(x')} dx' + \int_{x_1}^{x} e^{-M(x')} V(x') dx'. \tag{E34}$$

Appendix F: Cases with finite γ

We present preliminary sensitivity analysis of the parameter γ which controls the inverse lifetime of the mRNA molecule. When γ is finite, proteins are no longer produced in bursts. By observing Figs. 5 and 6, we conclude that when fixing other parameters and varying the γ value, the bursting PDMP model provides reliable estimation (error $< 30\%$) of the stationary distribution and MST when $\gamma \gtrsim 10$. For systems with smaller $\gamma < 10$ the bursting

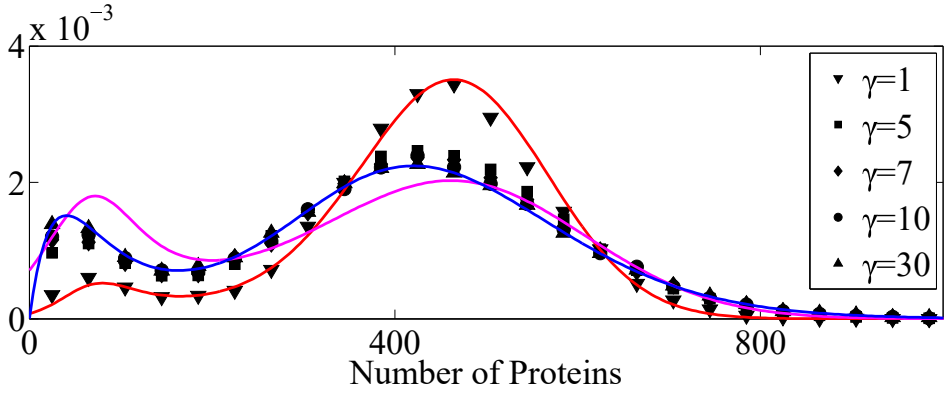

FIG. 5. Stationary distributions of the individual-based model (discrete markers), analytic predictions of piecewise deterministic Markov process (PDMP, solid blue) and the diffusion approximations (DA, solid red line for $\Gamma = 1$ and solid purple line $\Gamma = 2$).

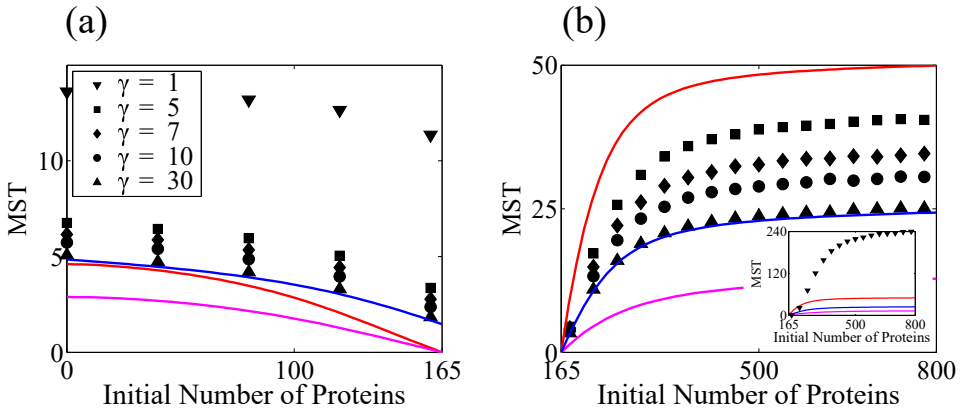

FIG. 6. Mean switching times (MST) of the individual-based model (discrete markers), analytic predictions of piecewise deterministic Markov process (PDMP, solid blue) and the diffusion approximations (DA, solid red line for $\Gamma = 1$ and solid purple line $\Gamma = 2$). (a) $T_{\text{low}\to\text{high}}$ for finite $\gamma = 1, 2, 7, 10, 30$; (b) $T_{\text{high}\to\text{low}}$ for $\gamma = 2, 7, 10, 30$ and the inset is for $\gamma = 1$. $x_c := 165/K$.

assumption no longer holds and the model needs to include mRNA with finite life time. The results suggests that the critical timescale is roughly the timescale of the transcription events, i.e., $\approx \tilde{H}$. Further investigation is needed to verify the relations between these two timescales. We remark that in any case the diffusion approximations capture neither the stationary distributions ($\gamma > 5$) or the MST ($\forall\gamma$).

[1] M. Kaern, T. Elston, W. Blake, and J. Collins, Nature Rev. Genet. **6**, 451 (2005).
[2] A. Walczak, A. Mugler, and C. Wiggins, Methods in molecular biology **880**, 273 (2011).
[3] E. Ozbudak, M. Thattai, I. Kurtser, A. Grossman, and A. van Oudenaarden, Nature Genet. **31**, 69 (2002).
[4] W. Blake, M. Kaern, C. Cantor, and J. Collins, Nature **422**, 633 (2003).
[5] T. Kepler and T. Elston, Biophys. J **81**, 3116 (2001).
[6] J. Hornos, D. Schultz, G. Innocentini, J. Wang, A. Walczak, J. Onuchic, and P. Wolynes, Phys. Rev. E **72**, 051907 (2005).
[7] A. Walczak, M. Sasai, and P. Wolynes, Biophys. J. **88**, 828 (2005).
[8] P. B. Warren and P. R. ten Wolde, J. Phys. Chem. B **109**, 6812 (2005).

[9] J. Wang, L. Xu., E. Wang, and S. Huang, Biophys. J. **99**, 29 (2010).

[10] J. Wang, K. Zhang, L. Xu, and E. Wang, Proc. Natl. Acad. Sci. U.S.A. **108**, 8257 (2010).

[11] M. Assaf, E. Roberts, and Z. Luthey-Schulten, Phys. Rev. Lett **106**, 248102 (2011).

[12] M. Strasser, F. Theis, and C. Marr, Biophys. J **102**, 19 (2012).

[13] J. Zhou, M. Aliyu, E. Aurell, and S. Huang, J. R. Soc. Interface **9**, 3539 (2012).

[14] M. Lu, J. Onuchic, and E. Ben-Jacob, Phys. Rev. Lett. **113**, 078102 (2014).

[15] M. Thattai and A. van Oudenaarden, Proc. Natl. Acad. Sci. U.S.A. **98**, 8614 (2001).

[16] N. Friedman, L. Cai, and X. Xie, Phys. Rev. Lett. **97**, 168302 (2006).

[17] N. van Kampen, "Stochastic processes in physics and chemistry," (North-Holland, Amsterdam, 2007).

[18] V. Shahrezaei and P. S. Swain, Proc. Natl. Acad. Sci. U.S.A. **105**, 17256 (2008).

[19] N. Kumar, T. Platini, and R. V. Kulkarni, Phys. Rev. Lett **113**, 5 (2014).

[20] L. Cai, N. Friedman, and X. Xie, Nature **440**, 358 (2006).

[21] T. Kurtz, J. Appl. Probab. **8**, 344 (1971).

[22] T. Kurtz, J. Appl. Probab. **7**, 49 (1970).

[23] Notation: the differential operator ∂_x acts as a total differential including the probability distributions $p_m(x,t)$ outside the matrix.

[24] A. Faggionato, D. Gabrielli, and M. R. Crivellari, J. Stat. Phys. **137**, 259 (2009).

[25] I. Bena, Int. J. Mod. Phys. B **20**, 2825 (2006).

[26] W. Horsthemke and R. Lefever, "Noise induced transitions: Theory and applications in physics, chemistry, and biology," (Springer, Berlin-Heidelberg, 2006).

[27] C. Doering, Phys. Rev. A **34**, 2564 (1986).

[28] J. Masoliver, K. Lindenberg, and B. West, Phys. Rev. A **34**, 2351 (1986).

[29] C. Doering, Phys. Rev. A **35**, 3166 (1987).

[30] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer, New York, 2009).

[31] R. Schwartz, *Biological modeling and simulation : a survey of practical models, algorithms, and numerical methods* (MIT Press, Cambridge, Mass., 2008).

[32] The waiting times can be generated by deriving the survival functions of each state [28] and utilizing the inverse transform sampling.

[33] P. Bokes, J. King, A. Wood, and M. Loose, Bull. Math. Biol. **75**, 351 (2013).

[34] Y. Taniguchi, P. Choi, G. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. Xie, Science **329**, 533 (2011).

[35] A. Arkin, J. Ross, and H. H. McAdams, Genetics **149**, 1633 (1998).

[36] Y. T. Lin and T. Galla, J. R. Soc. Interface **13**, 20150772 (2016).