bioRxiv preprint doi: https://doi.org/10.1101/2022.06.09.494908; this version posted June 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Threshold-crossing time statistics for gene expression in growing cells

César Nieto[1], Khem Raj Ghusinga[2], César Vargas-García [3], and Abhyudai Singh[4]

***Abstract*— Many intracellular events are triggered by attaining critical concentrations of their corresponding regulatory proteins. How cells ensure precision in the timing of the protein accumulation is a fundamental problem, and contrasting predictions of different models can help us understand the mechanisms involved in such processes. Here, we formulate the timing of protein threshold-crossing as a first passage time (FPT) problem focusing on how the mean FPT and its fluctuations depend on the threshold protein concentration. First, we model the protein-crossing dynamics from the perspective of three classical models of gene expression that do not explicitly accounts for cell growth but consider the dilution as equivalent to degradation: (*birth-death process*, *discrete birth with continuous deterministic degradation*, and *Fokker-Planck approximation*). We compare the resulting FPT statistics with a fourth model proposed by us (*growing cell*) that comprises size-dependent expression in an exponentially growing cell. When proteins accumulate in growing cells, their concentration reaches a steady value. We observe that if dilution by cell growth is modeled as degradation, cells can reach concentrations higher than this steady-state level at a finite time. In the growing cell model, on the other hand, the FPT moments diverge if the threshold is higher than the steady-state level. This effect can be interpreted as a transition between noisy dynamics when cells are small to an almost deterministic behavior when cells grow enough. We finally study the mean FPT that optimizes the timing precision. The growing cell model predicts a higher optimal FPT and less variability than the classical models.**

I. Introduction

At the level of an individual cell, gene products exhibit stochastic fluctuations (noise) [1], [2]. This noise stems from the inherent randomness of the synthesis and degradation of specific gene products [3], [4] as well as from variability in cell-cycle dependent variables such as gene replication [5], cell size dynamics [6], molecule segregation at division [7], [8], and growth rate [9], [10]. Although stochastic models of gene-expression have been studied for quite some time, effects of global processes such as cellular growth, cell size regulation, etc. have been incorporated only recently to these models [11]–[14].

Both number and concentration of a gene-product are the main quantities of interest in gene-expression models and they are characterized by computing their probability distributions or a few lower-order statistical moments. Gene-product fluctuations are studied also by the first passage times (FPT). Formally, FPT is the time spent to reach a threshold being relevant in many cases [15]–[17]. For example, accumulation of a threshold number of molecules is thought to be a mechanism for cell division [18], [19] that can be modeled as an FPT problem [20]. Other kind of FPT problems such as bacterial sporulation [21], cell differentiation [22] and apoptosis [23] involve reaching a threshold concentration, not a molecule number.

The difference between reaching either an amount or a concentration is often not clear. Some studies [16], [24], [25], for instance, consider reaching concentration as equivalent to molecule accumulation since they assume that cell size does not change appreciably during the molecule accumulation [4], [26]. When cell size grows exponentially, FPT problems comprising concentration threshold are expected to have different properties from those comprising molecule number threshold.

To illustrate this effect, consider a cell whose size grows exponentially over time and production rate of a protein of interest scales with cell size [27]–[29]. In case of numbers threshold, the mean FPT is smaller for a cell with a larger size at birth [20]. On the contrary, for a concentration threshold, we expect a rather weak dependency between mean FPT and cell size at birth since both the production rate as well as the number of molecules required to achieve a set concentration scale with cell size. In other words, production is balanced by dilution of gene-product concentration in a growing cell. In most models, effects of deterministic dilution by growth are taken to be equivalent to a random process such as either degradation or segregation at division [30]–[33]. It is however unclear whether explicitly accounting for cell-size and growth would result in the same FPT statistics as the models that do not do it.

Here, we study FPT statistics (mean FPT and noise in FPT, quantified by the squared coefficient of variation) for the protein concentration threshold-crossing in exponentially growing cells with protein synthesis rate proportional to the size. We compare these FPT statistics with the statistics of three other models that do not consider cell size dynamics: the discrete birth-death Poisson process [34], discrete stochastic production and continuous dilution [30], and the continuous process with Brownian noise by the Fokker-Planck equation [32]. An important finding is that the model that explicitly accounts for cell size dependency on gene expression has lower noise in FPT than the other three models.

[1]Department of Electrical and Computer Engineering, University of Delaware, Newark, DE. USA. 19716. cnieto@udel.edu

[3] AGROSAVIA. Corporación Colombiana de Investigación Agropecuaria, Bogotá, Colombia. cavargas@agrosavia.co

[2] Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA 27599. khem@email.unc.edu

[4] Department of Electrical and Computer Engineering, Biomedical Engineering, Mathematical Sciences, Center for Bioinformatics and Computational Biology, University of Delaware, Newark, DE USA 19716. e-mail: absingh@udel.edu

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.09.494908; this version posted June 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

II. Modelling the FPT distribution for thresholding

Let $c(t)$ denote the concentration of a gene-product at time t. Assuming $c(0) = 0$, the first passage time T to cross a threshold C is given by

$$T := \inf \{t : c(t) \geq C | c(0) = 0\}. \tag{1}$$

The probability distribution and moments of T depend upon the description of $c(t)$. Below, we describe each of the four models considered herein. Table 1 defines this c and other variables used in the article.

TABLE I

Variables employed throughout the article.

Variable	Interpretation
v	Cell size.
v_0	Cell size at the beginning of the process.
α	Growth rate.
k	Molecule synthesis rate.
t	Time.
τ	Doubling time.
T	First passage time (FPT).
c	Molecule concentration.
C	Threshold concentration.
x	Number of molecules.
X	Threshold number.
$\lceil X \rceil$	The least integer greater than or equal to X
$\rho_T(t)$	Probability density function (PDF) for the FPT $T = t$.
$p(x,t)$	Probability of having x molecules at a time t (x discrete).
$\rho(x,t)$	PDF for x concentration at time t (x continuous).

A. Cell growth and gene expression

Let $v(t)$ be the cell size at time t. We assume exponential growth [35], [36] described by:

$$\frac{dv}{dt} = \alpha v; \quad \alpha = \frac{\ln(2)}{\tau}, \tag{2}$$

where α is the growth rate related to the doubling time τ. By solving (2), we know that size follows an exponential function of time $v(t) = v_0 e^{\alpha t}$ with $v_0 := v|_{t=0}$.

Now, consider that cell has x molecules. The concentration c, in this case is related to x through the size v by:

$$c = \frac{x}{v} = \frac{x}{v_0 e^{\alpha t}}. \tag{3}$$

As suggested in some previous studies [20], [27], [28], we can consider the molecule number to increase at a size-dependent rate $r(t) = kv(t)$, where k is a constant. In this case, the mean molecule number $\langle x \rangle$ follows [27]:

$$\frac{d\langle x \rangle}{dt} = kv(t) = kv_0 e^{\alpha t}. \tag{4}$$

Since cells are growing exponentially, the mean concentration $\langle c \rangle = \frac{\langle x \rangle}{v}$, using (2) and (4), obeys:

$$\begin{aligned} \frac{d\langle c \rangle}{dt} &= \frac{d}{dt}\left(\frac{\langle x \rangle}{v}\right) = \frac{1}{v}\frac{d\langle x \rangle}{dt} - \frac{\langle x \rangle}{v^2}\frac{dv}{dt} \\ &= k - \alpha \langle c \rangle, \end{aligned} \tag{5}$$

which, with the initial condition $\langle c \rangle|_{t=0} = 0$, it has solution

$$\langle c \rangle(t) = \frac{k}{\alpha}\left(1 - e^{-\alpha t}\right). \tag{6}$$

In this article, we will present some approaches to describe the random fluctuations of the process associated with (5), and how the FPT statistics change depending on each description. First, we start with three common models that do not consider explicitly the growth and a fourth model where cells grow along the time.

B. Birth-Death Poisson Discrete Process

In a Birth-Death model, protein synthesis is described as a Poisson process with size-independent production rate k while degradation, dilution, and effects of size division are taken as a death process with occurrence rate αc (Fig. 1A). Since this model considers the size constant, the concentration c is equivalent to the molecule number x. In the same way, threshold amount X is identical to the threshold concentration C.

To solve the FPT distribution, we set the threshold concentration C, let $\lceil C \rceil$ being the least integer greater than or equal to C. The concentration c is in the set $c \in \{0, 1, \cdots \lceil C \rceil\}$ and, considering the state $\lceil C \rceil$ as an absorbent one:

$$\begin{aligned} \frac{dp(0,t)}{dt} &= -kp(0,t) + \alpha p(1,t) \\ &\vdots \\ \frac{dp(c,t)}{dt} &= k[p(c-1,t) - p(c,t)] \\ &\quad -\alpha[p(c,t) - p(c+1,t)] \\ &\vdots \\ \frac{dp(\lceil C \rceil - 1,t)}{dt} &= k[p(\lceil C \rceil - 2,t) - p(\lceil C \rceil - 1,t)] \\ &\quad -\alpha[p(\lceil C \rceil - 1,t)] \\ \frac{dp(\lceil C \rceil,t)}{dt} &= kp(\lceil C \rceil - 1,t) \\ p(c, t=0) &= \delta(c). \end{aligned} \tag{7}$$

The FPT distribution $\rho_T(t)$ can be estimated after integration of the system (7) through:

$$\rho_T(t) = \frac{dp(\lceil C \rceil, t)}{dt} = kp(\lceil C \rceil - 1, t), \tag{8}$$

which can be used to find the n-th moment $\langle T^n \rangle$ of $\rho_T(t)$:

$$\langle T^n \rangle = \int_0^\infty t^n \rho_T(t) dt. \tag{9}$$

Some studies [16] obtained analytic expressions for $\langle T^n \rangle$ solving (7). In this article we are interested on study the average FPT $\langle T \rangle$ and the variability of these FPTs measured by the squared coefficient of variation $CV^2(T)$ as follows:

$$CV^2(T) = \frac{\langle T^2 \rangle - \langle T \rangle^2}{\langle T \rangle^2} = \frac{\langle T^2 \rangle}{\langle T \rangle^2} - 1 \tag{10}$$

The mean FPT $\langle T \rangle$, for example, can be compared to the FPT of a deterministic process T_C derived from (6):

$$T_C = \frac{1}{\alpha} \ln\left(1 - \frac{C}{k/\alpha}\right). \tag{11}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.09.494908; this version posted June 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

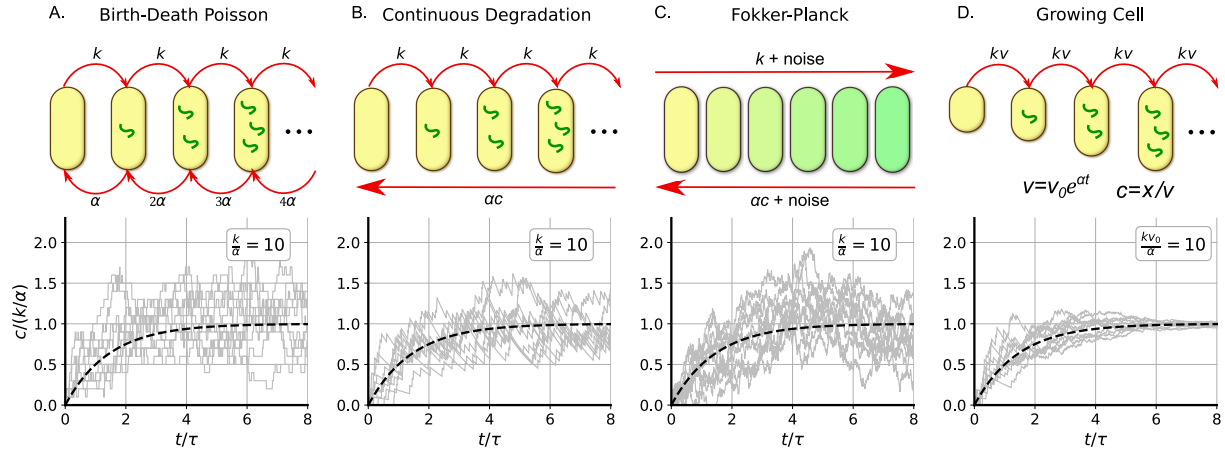

Fig. 1. **Diagram of the considered models for describing the protein accumulation.** A. *Top:* Discrete synthesis of molecules inside the cells can be approximated as a Birth-Death Poisson process with the birth rate k and death rate α times the molecule number. *Bottom:* An example of some stochastic trajectories in gray. B. *Top*: Discrete synthesis at a rate k plus deterministic dilution at a rate α times the concentration c. *Bottom:* An example of some stochastic trajectories in gray. . C. *Top*: Concentration dynamics as a continuous stochastic process following the Fokker-Planck equation with production rate k and degradation-dilution rate α. *Bottom:* An example of some stochastic trajectories in gray. D. *Top*: Discrete protein synthesis in an exponentially growing cell. The synthesis rate is proportional to the cell size. The concentration c is the molecule number x over the size v. *Bottom:* An example of some stochastic trajectories in gray. A Black dashed line is presented in the bottom plots corresponding to the deterministic trajectory $x = \frac{k}{\alpha}\left(1 - e^{-\alpha t}\right)$, derived in (6).

As this equation suggest, we can compare different relative dynamics by normalizing the concentration threshold C by k/α. Fig. 2A shows the dependence of $\langle T \rangle$ for different values of $\frac{C}{k/\alpha}$. This T_C also defines an upper boundary for $\langle T \rangle$ being its limit as $k/\alpha \to \infty$ (Fig. 2A dashed line).

This variable k/α is the mean number of molecules in steady state ($t >> \alpha^{-1}$) as (6) suggest. This variable is also quantifies the protein fluctuations. In steady conditions $CV^2(c) = \frac{\langle c^2 \rangle - \langle c \rangle^2}{\langle c \rangle^2}$ obeys [4]:

$$\lim_{t\to\infty} CV^2(c) = \frac{1}{\langle c \rangle} = \frac{\alpha}{k}. \tag{12}$$

We expect that $CV^2(T)$ has a similar order of magnitude as $CV^2(c)$. Thus, in this article, we will present the noise in T in terms of $CV^2(T)/(\alpha/k)$.

C. Discrete production and Continuous degradation.

Although molecule synthesis exhibits discreteness in production, dilution by growth is a continuous deterministic process since the cell is elongating exponentially. We can model dilution by considering c as a continuous variable and taking the dilution as a drift term, at a growth rate α in the associated forward differential Chapman-Kolmogorov equation (Fig. 1B) [37]:

$$\begin{aligned} \frac{d\rho(c,t)}{dt} &= -\frac{d}{dc}\left(-\alpha c \rho(c,t)\right) + k(\rho(c-1,t) - \rho(c,t)) \\ \rho(c,0) &= \delta(c), \end{aligned} \tag{13}$$

where the jumps from c to $c+1$ reflect the discrete protein synthesis. Fig. 2B and Fig. 2E present results of $\langle T \rangle$ and $CV^2(T)$, respectively. To estimate the solution of (13), we used the stochastic simulation algorithm as in Algorithm 1.

The result of algorithm 1 is an array of the FPT T_m with $m \in \{1, 2, \cdots, M\}$ with M being the total number of simulated cells (we took $M = 20000$). From this array, we can estimate the non-central moment $\langle T^n \rangle$ as:

$$\langle T^n \rangle \approx \frac{1}{M}\sum_{m=1}^{M} (T_m)^n. \tag{14}$$

Data: X, M, k, α
Result: $[T_1, \cdots, T_m, \cdots, T_M]$
for $m \leftarrow 1$ **to** M ***by*** 1 **do**
  $T_m = 0$, $x_m = 0$;
  while $x_m < X$ **do**
    $r = random(0, 1)$;
    $\Delta t = -\frac{1}{k}\ln(r)$;
    $x_m = x_n e^{-\alpha \Delta t}$;
    $x_m = x_m + 1$;
    $T_m = T_m + \Delta t$;
  end
end

Algorithm 1: Stochastic Simulation Algorithm for estimating the FPT T_m for each of the M cells.

D. Approximation of continuous stochastic dynamics: The Fokker-Planck equation

A Wiener process can approximate discrete jumps, under the assumption that the burst sizes are small relative to the molecule number (Fig. 1C) [4]. Under this approximation, c takes continuous values with stochastic dynamics that can be described by the SDE

$$dc = (k - \alpha c)dt + \sqrt{(k + \alpha c)}dW, \tag{15}$$

with dW being a random Wiener increment satisfying $\langle dW(t)dW(t') \rangle = \delta(t - t')$ [32]. From this SDE, the

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.09.494908; this version posted June 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

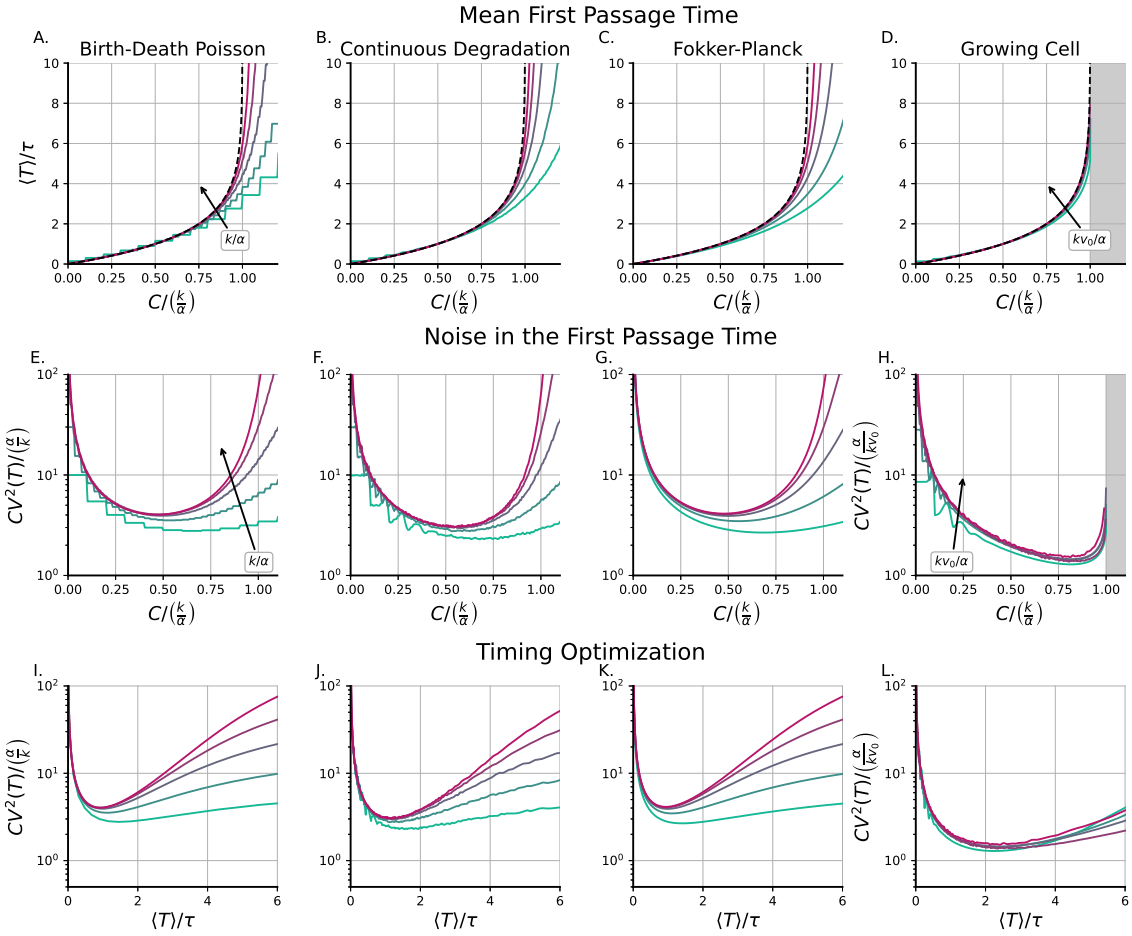

Fig. 2. **FPT statistics for different threshold concentrations considering the four models for protein accumulation.** *Top*: Mean first passage time (FPT) as a function of different concentration thresholds (C) for the four models explored here and different noise regimes ($k/\alpha \in \{10, 30, 100, 300, 1000\}$ cyan line being $k/\alpha = 10$ or $kv_0/\alpha \in \{10, 30, 100, 300, 1000\}$ for growing cell, cyan line being $kv_0/\alpha = 10$). *Middle*: Stochastic fluctuations of the FPT measured as the squared coefficient of variation relative to the protein concentration in steady level α/k as a function of different concentration thresholds for the four explored models and different noise regimes. *Bottom*: Stochastic fluctuation of the FPT's $CV^2(T)$ as a function of the mean FPT with C as the parameter for the four models explored here and different regimes of noise.

distribution $\rho(c,t)$ follows the Fokker-Planck equation

$$\begin{aligned} \frac{d\rho(c,t)}{dt} &= -\frac{d}{dc}\left[(k-\alpha c)\rho(c,t)\right] + \frac{d^2[(k+\alpha c)\rho(c,t)]}{dc^2} \\ \rho(c,0) &= \delta(c). \end{aligned} \tag{16}$$

Analytic approaches for obtaining the FPT moments of $\rho_T(t)$ are elaborated in [32] for models similar to (16). To solve numerically these moments we follow the method elaborated in Ref. [38], obtaining the moments $\langle T^n\rangle$:

$$\frac{1}{2}(k+\alpha c_0)\frac{\partial^2\langle T^n\rangle}{\partial^2 c_0} + (k-\alpha c_0)\frac{\partial\langle T\rangle}{\partial c_0} = -n\langle T^n\rangle, \tag{17}$$

where $c_0 := c|_{t=0}$, is the concentration at the beginning of the process. We solve the problem with boundary conditions $\langle T^n\rangle|_{c_0=C} = 0$ and $\lim_{c_0\to-\infty}\langle T^n\rangle|_{c_0} = 0$ and evaluate the solution $\langle T^n\rangle$ at $c_0 = 0$. Fig. 2C and Fig. 2F show $\langle T\rangle$ and $CV^2(T)$ respectively as function of $C/(k/\alpha)$.

E. Size dependent propensity and continuous growth: FPT for a threshold concentration in a growing cell.

This model considers that concentration c and molecule number x are related through the size v as shown in (3). The protein accumulation can be described using a master equation similar to (7) If $p_i(t) := p(x = i, t)$ is the probability for the cell having i molecules at a time t, protein accumulation is described by:

$$\begin{aligned} \frac{dp_0}{dt} &= -kvp_0 \\ \frac{dp_1}{dt} &= kvp_0 - kvp_1 \\ &\vdots \\ \frac{dp_i}{dt} &= kvp_{i-1} - kvp_i, \\ &\vdots \\ \frac{dp_{\lceil X\rceil}}{dt} &= kvp_{\lceil X\rceil-1}, \end{aligned} \tag{18}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.09.494908; this version posted June 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

with $v = v_0 e^{\alpha t}$ and $\lceil X \rceil$ being the lowest integer greater than or equal to the threshold amount $X = Cv$. Here, the number of molecules x is incremented by 1 at a rate proportional to cell size $kv(t)$ as explained in (4).

Since C is fixed and cell size v is growing, X is a function of time:

$$X = Cv = Cv_0 e^{\alpha t} = X_0 e^{\alpha t}, \quad X_0 : X|_{t=0}. \tag{19}$$

To explain the effects of having a growing X as in (19), consider the case where the initial volume v_0 is small enough, such as $X_0 = Cv_0 < 1$. We can define the time $t_1 > 0$ such as the instant when $X := 1$:

$$1 = Cv(t_1) = Cv_0 e^{\alpha t_1}. \tag{20}$$

This means:

$$t_1 = \frac{1}{\alpha} \ln\left(\frac{1}{Cv_0}\right). \tag{21}$$

During $0 < t < t_1$ the system follows the master equation:

$$\begin{aligned} \frac{dp_0}{dt} &= -kvp_0 = -kv_0 e^{\alpha t} p_0 \\ \frac{dp_1}{dt} &= kvp_0. \end{aligned} \tag{22}$$

During $0 < t < t_1$, cells in the state $x = 1$ have already crossed the threshold. The analytic solution of (22) for $p_1(t)$ is given by:

$$p_1(t) = 1 - \exp\left(-\frac{kv_0}{\alpha}\left(e^{\alpha t} - 1\right)\right). \tag{23}$$

This helps us to estimate the FPT distribution:

$$\begin{aligned} \rho_T(t) &= \frac{dp_1}{dt} = kvp_0 \\ &= kv_0 \exp\left(-\frac{kv_0}{\alpha}\left(e^{\alpha t} - 1\right) + \alpha t\right), t \in (0, t_1). \end{aligned} \tag{24}$$

When $t_1 < t < t_2$, the remaining cells, those ones still in $x = 0$ have to jump to state $x = \lceil X \rceil = 2$ to cross the threshold concentration. The master equation is now

$$\begin{aligned} \frac{dp_0}{dt} &= -kvp_0 \\ \frac{dp_1}{dt} &= kvp_0 - kvp_1 \\ \frac{dp_2}{dt} &= kvp_1. \end{aligned} \tag{25}$$

The initial conditions of the distribution p_i at time t_1 are:

$$\begin{aligned} p_0(t_1) &= 1 - p_1(t_1) = \exp\left(-\frac{kv_0}{\alpha}\left(e^{\alpha t_1} - 1\right)\right) \\ p_1(t_1) &= 0 \\ p_2(t_1) &= 0. \end{aligned} \tag{26}$$

Hence, the distribution of FPTs during $t \in (t_1, t_2)$ is $\rho_T(t) = \frac{dp_2}{dt}$ which is solved using (25).

In general, we define the time t_X where the cell grows so large that the threshold number is X molecules:

$$t_X := \frac{1}{\alpha} \ln\left(\frac{X}{Cv_0}\right). \tag{27}$$

During $t_{\lceil X \rceil - 1} < t < t_{\lceil X \rceil}$, the master equation is (18) and the initial conditions, similar to (26), correspond to the distribution vector $p_i(t_{\lceil X \rceil - 1}) \forall i$.

The vector $p_i(t)$, with $t \in (t_{\lceil X \rceil - 1}, t_{\lceil X \rceil}))$ and $X = X_0 e^{\alpha t}$, can be calculated knowing the values of $p_i(t_{\lceil X \rceil - 1})$, $v_{\lceil X \rceil - 1} := v|_{t = t_{\lceil X \rceil - 1}} = v_0 e^{\alpha t_{\lceil X \rceil - 1}}$, and using the recursive formula:

$$\begin{aligned} p_0(t) &= \exp\left(-\frac{kv_0}{\alpha}(e^{\alpha t} - 1)\right) \\ p_i(t) &= \exp\left(-\frac{kv_{\lceil X \rceil - 1}}{\alpha} e^{\alpha(t - t_{\lceil X \rceil - 1})}\right) \\ &\quad \times [p_i(t_{\lceil X \rceil - 1}) e^{\frac{k}{\alpha} v_{\lceil X \rceil - 1}} \\ &\quad + kv_{\lceil X \rceil - 1} \int_0^{t - t_{\lceil X \rceil - 1}} \exp\left(\frac{kv_{\lceil X \rceil - 1}}{\alpha} e^{\alpha t'}\right) \\ &\quad \times e^{\alpha t'} p_{i-1}(t_{\lceil X \rceil - 1} + t') dt']. \end{aligned} \tag{28}$$

with the additional initial condition $p_{\lceil X \rceil - 1}(t_{\lceil X \rceil - 1}) = 0$.

After solving $p_{\lceil X \rceil}$ using (28), the FPT distribution is:

$$\rho_T(t) = \frac{dp_{\lceil X \rceil}}{dt}; \quad t_{\lceil X \rceil - 1} < t < t_{\lceil X \rceil}, \tag{29}$$

In Fig. 2D and Fig. 2H, we present numerical computations for $\langle T \rangle$ as a function of the threshold concentration $C/(k/\alpha)$ for different values of $\frac{kv_0}{\alpha}$. This parameter can be associated to the parameter k/α in the classical models. The interpretation of this ratio will be discussed in the next section.

III. Discussion

In this work, we revisited three models of molecule accumulation by cells up to a threshold concentration. These models do not account for cell size dynamics and they consider that concentration is equivalent to molecule number. The parameter k/α defines the mean concentration at steady state. As explained in (12) and articles such as [4], [5], this number also defines the relative noise of the process. In the growing cell model, the regime of noise is defined by the parameter $\frac{kv_0}{\alpha}$. This quantity is interpreted as the number of molecules needed to reach steady concentration k/α when the cell has size v_0. Observe that for a fixed concentration k/α, by changing v_0, the noise of the process also changes.

In the growing cell model (Fig. 2D and 2H), we observe that the FPT moments diverge as the threshold concentration exceeds the steady value k/α. To understand this effect, we derived the Theorem 1 in the appendix and present some stochastic trajectories as examples (Fig. 1D). As the theorem 1 explains, as time passes, the concentration fluctuations decrease exponentially. As a result, some cells can cross thresholds higher than the steady concentration at the beginning of the process. However, cells that did not cross this threshold have less probability of doing it in the future.

In previous studies [16], [32], protein accumulation FPT problem focused on *E. coli* lysis by lambda phage. Once

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.09.494908; this version posted June 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the process starts, the protein holin is accumulated on the bacterial membrane and punctures it after its concentration crosses a threshold concentration [39]. Some evolutionary studies observed that the parameters of the chemical process evolved such as the lysis occurs at a concentrations with an associated FPT having minimal fluctuations [40], [41].

Considering this experimental context, we can show (Fig. 2I, 2J, 2K and 2L) the profile $CV^2(T)$ vs $\langle T\rangle$ predicted by each model. In classical models, $CV^2(T)$ shows a a global minimum close to $\langle T\rangle \approx \tau$ while in the model of growing cell, this minimum corresponds to a larger FPT $\langle T\rangle \approx 2\tau$. As discussed above, the fact that in growing cells the fluctuations are progressively smaller can also explain why the position of the optimum noise time is higher in the growing cell model than in the classical ones.

IV. Conclusions

This article explores the statistics of the first passage times (FPT's) for the process of protein accumulation up to a threshold concentration. We revisited three classical models as a context: the Birth-Death Poisson Process, the discrete production with continuous degradation, and the Fokker-Planck approximation. We compare their results on FPT statistics with a fourth model proposed by us, which considers a growing cell with protein produced at a rate proportional to the cell size. We see how the first three models are equivalent in these timing statistics and how they differ from the last model.

The first three models predict that, on average, cells can reach concentrations higher than the mean concentration at steady-state. On the other hand, the growing cell model predicts that all the FPT moments diverge if the threshold concentration is higher than this steady concentration.

Regarding the timing variability, we observe how the growing cell model predicts fewer fluctuations in timing than the classical models. After finding the average FPT that minimizes these timing fluctuations, this first model predicts an optimum FPT almost twice the one found by the other models.

Acknowledgment

AS is supported by NIH 1R01GM124446-01.

Appendix

Theorem 1: If the dynamics of the vector $\vec{\mathbf{P}} = (p_0, p_1, \cdots, p_x, \cdots, P_{\lceil X\rceil})^\dagger$ is described by the system:

$$\frac{d}{dt}\vec{\mathbf{P}} = kv_0 e^{\alpha t}\mathbf{M}\vec{\mathbf{P}}, \tag{30}$$

with k, v_0 and α constants, and $\mathbf{M}$ a matrix of constant elements $M_{i,j}$ with properties:

$$\begin{aligned} M_{i,j} &> 0 \quad i \neq j, \\ \sum_j M_{i,j} &= 0, \end{aligned} \tag{31}$$

and defining the concentration $c := x/(v_0 e^{\alpha t})$, then, the fluctuations in c decay asymptotically as an exponential function of time, this is:

$$CV^2(c) < Ke^{-\alpha t}, \quad t \ll \alpha^{-1}, \tag{32}$$

for a constant K.

Proof: Consider the mapping:

$$\tau \longleftrightarrow \frac{kv_0}{\alpha}\left(e^{\alpha t} - 1\right), \quad d\tau = kv_0 e^{\alpha t} dt, \tag{33}$$

which after been applied to (30), reduces the system into a master equation of a Poisson process:

$$\frac{d}{d\tau}\vec{\mathbf{P}} = \mathbf{M}\vec{\mathbf{P}}. \tag{34}$$

which has the property:

$$\langle x\rangle = \mathrm{var}(x), \tag{35}$$

for any τ, with $\langle x\rangle = \sum_x x p_x$ and $\mathrm{var}(x) = \sum_x x^2 p_x - \left(\sum_x x p_x\right)^2$. Since the mapping (33) is monotonically increasing function of the time and state independent, this property also holds for the system (30).

Given that the protein x is produced at a rate proportional to the size, the mean protein follows:

$$\langle x\rangle = v_0 e^{\alpha t} g(t) < v_0 e^{\alpha t} G_{max}, \tag{36}$$

with G_{max} a constant since $g(t)$ is bounded up to time t. Now, using (35), we obtain:

$$\mathrm{var}(x) = \langle x\rangle < v_0 e^{\alpha t} G_{max}. \tag{37}$$

While the mean concentration follows

$$\langle c\rangle = \frac{\langle x\rangle}{v_0 e^{\alpha t}} = G(t) < G_{max}, \tag{38}$$

the fluctuations of $c = k/(v_0 e^{\alpha t})$ satify:

$$\begin{aligned} \mathrm{var}(c) &= \langle\left(\frac{x}{v_0 e^{\alpha t}}\right)^2\rangle - \langle\frac{x}{v_0 e^{\alpha t}}\rangle^2 \\ &= v_0^2 e^{-2\alpha t}\mathrm{var}(x) < v_0^3 e^{-2\alpha t} e^{\alpha t} G_{max}. \end{aligned} \tag{39}$$

Now, the squared coefficient of variation follows:

$$CV^2(c) = \frac{\mathrm{var}(c)}{\langle c\rangle^2} < Ke^{-\alpha t}. \tag{40}$$

■

References

[1] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, "Stochastic gene expression in a single cell," *Science*, vol. 297, no. 5584, pp. 1183–1186, 2002.

[2] W. J. Blake, M. Kærn, C. R. Cantor, and J. J. Collins, "Noise in eukaryotic gene expression," *Nature*, vol. 422, no. 6932, pp. 633–637, 2003.

[3] B. Schwanhäusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach, "Global quantification of mammalian gene expression control," *Nature*, vol. 473, no. 7347, pp. 337–342, 2011.

[4] J. Paulsson, "Models of stochastic gene expression," *Physics of life reviews*, vol. 2, no. 2, pp. 157–175, 2005.

[5] R. Dessalles, V. Fromion, and P. Robert, "Models of protein production along the cell cycle: An investigation of possible sources of noise," *Plos one*, vol. 15, no. 1, p. e0226016, 2020.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.09.494908; this version posted June 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[6] M. P. Swaffer, D. Chandler-Brown, M. Langhinrichs, G. Marinov, W. Greenleaf, A. Kundaje, K. M. Schmoller, and J. M. Skotheim, "Size-independent mrna synthesis and chromatin-based partitioning mechanisms generate and maintain constant amounts of protein per cell," *bioRxiv*, 2020.

[7] C. Jia, A. Singh, and R. Grima, "Concentration fluctuations due to size-dependent gene expression and cell-size control mechanisms," *bioRxiv*, 2021.

[8] M. Soltani, C. A. Vargas-Garcia, D. Antunes, and A. Singh, "Intercellular variability in protein levels from stochastic expression and noisy cell cycle processes," *PLoS computational biology*, vol. 12, no. 8, p. e1004972, 2016.

[9] P. Thomas, G. Terradot, V. Danos, and A. Y. Weiße, "Sources, propagation and consequences of stochasticity in cellular growth," *Nature communications*, vol. 9, no. 1, pp. 1–11, 2018.

[10] S. Klumpp, Z. Zhang, and T. Hwa, "Growth rate-dependent global effects on gene expression in bacteria," *Cell*, vol. 139, no. 7, pp. 1366–1375, 2009.

[11] J. Lin and A. Amir, "Disentangling intrinsic and extrinsic gene expression noise in growing cells," *Physical Review Letters*, vol. 126, no. 7, p. 078101, 2021.

[12] P. Thomas and V. Shahrezaei, "Coordination of gene expression noise with cell size: analytical results for agent-based models of growing cell populations," *Journal of the Royal Society Interface*, vol. 18, no. 178, p. 20210274, 2021.

[13] C. Jia and R. Grima, "Frequency domain analysis of fluctuations of mrna and protein copy numbers within a cell lineage: theory and experimental validation," *Physical Review X*, vol. 11, no. 2, p. 021032, 2021.

[14] C. H. Beentjes, R. Perez-Carrasco, and R. Grima, "Exact solution of stochastic gene expression models with bursting, cell cycle and replication dynamics," *Physical Review E*, vol. 101, no. 3, p. 032403, 2020.

[15] D. Gomez, R. Marathe, V. Bierbaum, and S. Klumpp, "Modeling stochastic gene expression in growing cells," *Journal of theoretical biology*, vol. 348, pp. 1–11, 2014.

[16] K. R. Ghusinga, J. J. Dennehy, and A. Singh, "First-passage time approach to controlling noise in the timing of intracellular events," *Proceedings of the National Academy of Sciences*, vol. 114, no. 4, pp. 693–698, 2017.

[17] P. J. Piggot and D. W. Hilbert, "Sporulation of bacillus subtilis," *Current opinion in microbiology*, vol. 7, no. 6, pp. 579–586, 2004.

[18] K. Sekar, R. Rusconi, J. T. Sauls, T. Fuhrer, E. Noor, J. Nguyen, V. I. Fernandez, M. F. Buffing, M. Berney, S. Jun, *et al.*, "Synthesis and degradation of ftsz quantitatively predict the first cell division in starved bacteria," *Molecular systems biology*, vol. 14, no. 11, p. e8623, 2018.

[19] F. Si, G. Le Treut, J. T. Sauls, S. Vadia, P. A. Levin, and S. Jun, "Mechanistic origin of cell-size control and homeostasis in bacteria," *Current Biology*, vol. 29, no. 11, pp. 1760–1770, 2019.

[20] K. R. Ghusinga, C. A. Vargas-Garcia, and A. Singh, "A mechanistic stochastic framework for regulating bacterial cell division," *Scientific reports*, vol. 6, no. 1, pp. 1–9, 2016.

[21] K. Carniol, P. Eichenberger, and R. Losick, "A threshold mechanism governing activation of the developmental regulatory protein σf in bacillus subtilis," *Journal of Biological Chemistry*, vol. 279, no. 15, pp. 14860–14870, 2004.

[22] T. Koyama, M. Iwami, and S. Sakurai, "Ecdysteroid control of cell cycle and cellular commitment in insect wing imaginal discs," *Molecular and cellular endocrinology*, vol. 213, no. 2, pp. 155–166, 2004.

[23] J. Roux, M. Hafner, S. Bandara, J. J. Sims, H. Hudson, D. Chai, and P. K. Sorger, "Fractional killing arises from cell-to-cell variability in overcoming a caspase activity threshold," *Molecular systems biology*, vol. 11, no. 5, p. 803, 2015.

[24] Z. Vahdat, K. R. Ghusinga, and A. Singh, "Comparing feedback strategies for minimizing noise in gene expression event timing," in *2021 29th Mediterranean Conference on Control and Automation (MED)*, pp. 450–455, IEEE, 2021.

[25] A. D. Co, M. C. Lagomarsino, M. Caselle, and M. Osella, "Stochastic timing in gene expression for simple regulatory strategies," *Nucleic acids research*, vol. 45, no. 3, pp. 1069–1078, 2017.

[26] D. R. Rigney, "Stochastic model of constitutive protein levels in growing and dividing bacterial cells," *Journal of Theoretical Biology*, vol. 76, no. 4, pp. 453–480, 1979.

[27] C. Nieto-Acuña, J. C. Arias-Castro, C. Vargas-García, C. Sánchez, and J. M. Pedraza, "Correlation between protein concentration and bacterial cell size can reveal mechanisms of gene expression," *Physical Biology*, vol. 17, no. 4, p. 045002, 2020.

[28] X.-M. Sun, A. Bowman, M. Priestman, F. Bertaux, A. Martinez-Segura, W. Tang, C. Whilding, D. Dormann, V. Shahrezaei, and S. Marguerat, "Size-dependent increase in rna polymerase ii initiation rates mediates gene expression scaling with cell size," *Current Biology*, 2020.

[29] K.-L. Claude, D. Bureik, P. Adarska, A. Singh, and K. M. Schmoller, "Transcription coordinates histone amounts and genome content," *bioRxiv*, 2020.

[30] Z. Vahdat, Z. Xu, and A. Singh, "Modeling protein concentrations in cycling cells using stochastic hybrid systems," *IFAC-PapersOnLine*, vol. 54, no. 9, pp. 521–526, 2021.

[31] A. Singh and J. J. Dennehy, "Stochastic holin expression can account for lysis time variation in the bacteriophage λ," *Journal of The Royal Society Interface*, vol. 11, no. 95, p. 20140140, 2014.

[32] K. Rijal, A. Prasad, A. Singh, and D. Das, "Exact distribution of threshold-crossing times for protein concentrations: Implication for biological timekeeping," *bioRxiv*, 2021.

[33] S. Dey, S. Kannoly, P. Bokes, J. J. Dennehy, and A. Singh, "The role of incoherent feedforward circuits in regulating precision of event timing," *bioRxiv*, 2020.

[34] T. Williams, "The basic birth-death model for microbial infections," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 27, no. 2, pp. 338–360, 1965.

[35] W. K. Sinclair and D. W. Ross, "Modes of growth in mammalian cells," *Biophysical journal*, vol. 9, no. 8, pp. 1056–1070, 1969.

[36] S. Cooper, "Distinguishing between linear and exponential cell growth during the division cycle: single-cell studies, cell-culture studies, and the object of cell-cycle research," *Theoretical Biology and Medical Modelling*, vol. 3, no. 1, pp. 1–15, 2006.

[37] N. Friedman, L. Cai, and X. S. Xie, "Linking stochastic dynamics to population distribution: an analytical framework of gene expression," *Physical review letters*, vol. 97, no. 16, p. 168302, 2006.

[38] H. C. Tuckwell and F. Y. Wan, "First-passage time of markov processes to moving barriers," *Journal of applied probability*, vol. 21, no. 4, pp. 695–709, 1984.

[39] R. Young, "Phage lysis: three steps, three choices, one outcome," *Journal of microbiology*, vol. 52, no. 3, pp. 243–258, 2014.

[40] J. J. Dennehy and N. Wang, "Factors influencing lysis time stochasticity in bacteriophage λ," *BMC microbiology*, vol. 11, no. 1, p. 174, 2011.

[41] S. Kannoly, T. Gao, S. Dey, N. Wang, A. Singh, and J. J. Dennehy, "Optimum threshold minimizes noise in timing of intracellular events," *Iscience*, vol. 23, no. 6, p. 101186, 2020.