Cell-division time statistics from stochastic exponential threshold-crossing

Kuheli Biswas¹ and Naama Brenner¹

¹Technion - Israel Institute of Technology, Haifa, 3200003, Israel

For cell division to take place, proteins that carry it out need to accumulate to a functional threshold. Most of these *divisome* proteins are highly abundant in the cell, and accumulate smoothly and approximately exponentially throughout the cell cycle. In this threshold-crossing process, stochastic components arise from variation from one cycle to the next of accumulation rate and division fraction, and from fluctuations of the threshold itself. How these combine to determine the statistical properties of division times is still not well understood. Here we formulate this stochastic process and calculate the statistical properties of cell division times by using *first passage* time (FPT) techniques. We find that the distribution shape is determined by a ratio between two coefficient of variations (CVs), interpolating between Gaussian-like and long-tailed. Mean, variance and skewness of division times are predicted to follow well-defined relationships with model parameters. Publicly available single-cell data span a broad range of values in parameter space; the measured distribution shape and moment scaling agree well with the theory over the entire range. Because of balanced biosynthesis, the accumulation dynamics of any abundant protein - as well as cell size – predicts division time statistics equally well using our model. These results suggest that cell division is a multi-variable emergent process, which is nevertheless predictable by a single variable thanks to coupling and correlations inside the system.

Introduction: What makes a cell divide? this question has been at the center of scientific investigation for decades. Over the years many attributes have been considered as triggering cell division: elapsed time, cell size, DNA replication, accumulation of proteins to threshold, as well as combinations of these events (1-9). In particular, divisome proteins - which carry out the biophysical steps leading to division - need to accumulate to a functioning threshold in order for division to be carried out (10-12). This, as well as other cellular events that depend on the presence of sufficient protein content, marks the problem of protein accumulation to threshold as a central one in several different contexts.

Most previous work has considered the accumulation of low-copy-number proteins, which is noisy and strongly affected by irregularities of bursting to a fixed threshold (13–17). However, many functionally important proteins are expressed at high copy numbers in the

2

cell; their accumulation generally proceeds smoothly through the cell cycle (18–20), and noise is negligible over this timescale. As an important quantitative result of single-cell tracking, it was shown that the accumulation of many proteins, as well as cell size, between consecutive divisions is exponential to a good approximation. Such behavior was found for yeast and different bacterial cells (19, 21–24).

These dynamics define a different threshold crossing problem, in which smooth exponential accumulation needs to reach a threshold. Previous work considered such a process with white noise on top of a fixed exponential accumulation rate, and a constant threshold (20, 24–26), or a constant accumulation rate and a stochastic threshold (27). Thus in practically all previous models and analyses, the basic rate of exponential accumulation was considered fixed. This was motivated by single-cell measurements in *E. coli* that showed a narrow distribution of exponential rates (24, 28, 29). However, other data-sets in similar experiments on the same bacteria shows a more significant variability (18–20). It is therefore crucial to develop a theoretical framework that covers this range of behaviors and considers both the stochastic threshold and cycle-to-cycle variability in accumulation rate.

In what follows we develop such a framework and compare the results to multiple published data-sets of microbial single-cell tracking. First, we define and analyze a stochastic process in which a single protein accumulates exponentially to cross a stochastic threshold. We analytically calculate the distribution of division times using the *first passage time* (FPT) approach, identify the important variables controlling this distribution and develop limiting approximations to the general expression. Next we investigate the agreement of these quantitative results with published single-cell data. These data span a significant range of parameters of the threshold-crossing problem and exhibit excellent agreement with the theory in the entire range. Finally, we show that, thanks to balanced biosynthesis, a similar level of agreement holds when applying the threshold-crossing model to different highly expressed proteins as well as to cell size. These results suggest that cell division is an emergent property arising from multiple coupled (and therefore correlated) variables.

Results

Model Development: Consider a protein whose copy number $n_k(t)$ accumulates across the k^{th} cell cycle, finally reaching a threshold c that triggers division. Continuous measurements of fluorescently tagged, highly expressed proteins over time reveal smooth, exponential-like accumulation throughout each cell cycle interrupted by abrupt drops at division. Fig. 1a shows a small portion of such a measurement, which provides the inspiration to our model. Over the k^{th} cycle of growth and division, protein content can thus be described as

$$n_k(t) = n_k(0)e^{\alpha_k t} + \xi(t), \quad 0 < t < T_k$$
(1)

$$n_{k+1}(0) = n_k(T_k)f_k, (2)$$





FIG. 1. Stochastic exponential accumulation to a fluctuating threshold. (a) The accumulation of highly expressed protein in bacteria proceeds in a smooth, exponential-like manner. Coarse-grained parameters α_k and $n(T_k)$ are estimated directly from the data by exponential fitting. (b) Distribution of effective exponential accumulation rates α_k (best fit to data within each cycle), collected across cycles. Coefficient of variation $\chi_{\alpha} = \sigma_{\alpha}/\mu_{\alpha} \approx 0.3$. (c) Distribution of final protein content $n(T_k)$ across cycles, provides an indirect measure of fluctuations in threshold *c* under the assumption of division triggered by threshold crossing. Coefficient of variation ≈ 0.2 . Data in (a),(b),(c) from (18). (d) Simulated protein trajectories (lighter blue) follow smooth exponential accumulation with varying rates across cycles. Cell division occurs when crossing a fluctuating threshold (darker blue; distribution on the right). Top: Resulting distribution of division times, model simulation (histogram) and analytic solution (line).

where α_k is the exponential accumulation rate during cell cycle k, T_k is its duration and f_k is the division fraction at its end. The cell divides when the protein reaches a threshold $n(T_k) = c(T_k)$, which itself is a stochastic process, described by an Ornstein-Uhlenbeck (OU) process over the long timescale of many cycles

$$\frac{dc}{dt} = \gamma(\mu_c - c) + \eta(t), \tag{3}$$

Here the threshold has mean μ_c , and δ -correlated Gaussian noise $\langle \eta(t) \rangle = 0$ and $\langle \eta(t)\eta(t') \rangle = 2D\delta(t-t')$. The noise controls the variance of c by $\sigma_c^2 = D/\gamma$, while the restoring force γ determines the typical correlation time as $\tau_c = 1/\gamma$.

There are generally four sources of noise in this model: accumulation rate α_k and division fraction f_k can vary across cycles; $\xi(t)$ is added white noise to the smooth accumulation; and the threshold c fluctuates with variance σ_c^2 . An example of the distribution of α_k is presented in Fig. 1b, showing an approximately Gaussian shape with CV of 0.3. Division noise, in contrast, shows a CV of only a few percent in all available data-sets (30, 31). We thus neglect this noise and take $f_k = 1/2$. We shall also neglect the white noise $\xi(t)$, since

4

it contributes negligibly to the statistical properties when variation in α_k are taken into account (18, 19). An estimate of the stochastic effect of threshold can be inferred from the distribution of final protein values just before division, $n_k(T_k)$. Fig. 1c shows that this can be a significant noise source, with $CV \approx 0.2$, and is not neglected. Our model thus consists of smooth exponential accumulations, with rate varying randomly from one cycle to the next, crossing a threshold that follows OU dynamics and dividing in half.

Model Solution: The stochastic threshold crossing model described above can be analytically solved as a first passage time (FPT) problem, with the Ornstein-Uhlenbeck Green's function as input (32–35). Recent work has derived the distribution of threshold crossing events in this model for a fixed accumulation rate α (27). To extend to variable accumulation rates, we assume that α_k is drawn at each cycle k from a Gaussian distribution with mean μ_{α} and standard deviation σ_{α} (see Fig. 1b).

We begin by considering the case of rapid threshold fluctuations, $\tau_c \ll T$; the threshold is then a white Gaussian variable with coefficient of variation χ_c . Using a Taylor expansion and ignoring higher order terms of α (detailed description in Appendix A), we find the cell division time distribution

$$P(T) = \frac{(\mu_{\alpha}\chi_c^2 + \sigma_{\alpha}^2 T(\chi_c^2 + \ln 2))}{\sqrt{2\pi}(\chi_c^2 + (\sigma_{\alpha}T)^2)^{3/2}} \exp\left[-\frac{(\mu_{\alpha}T - \chi_c^2 - \ln 2)^2}{2(\chi_c^2 + (T\sigma_{\alpha})^2)}\right].$$
(4)

In this expression, the mean accumulation rate μ_{α} sets the scale for the division time, whereas the two CVs – χ_{α} of the accumulation rates and χ_c of the threshold – determine the distribution shape. Two simple limits can be defined, when one noise source is dominant over the other. In the first, $\chi_c \ll \chi_{\alpha}$, random exponents have to cross an almost constant threshold. Then, the crossing times are given by the random variable $\ln 2/\alpha$; for Gaussian α we have

$$P(T) = \frac{\ln 2}{\sqrt{2\pi\sigma_{\alpha}^2}T^2} \exp\left[-\frac{\left(\ln 2/T - \mu_{\alpha}\right)^2}{2\sigma_{\alpha}^2}\right].$$
(5)

This is the reciprocal of the Gaussian accumulation rates, scaled by $\ln 2$. This distribution has a heavy tail: in the limit of large T, i.e., $T \gg \ln 2/\mu_{\alpha}$, it decreases as $P(T) \sim T^{-2}$. While it does not admit finite moments, we shall see that it provides a good approximation to the measured distributions in the relevant regime.

At the other extreme, where threshold fluctuations are dominant $\chi_c \gg \chi_{\alpha}$, we find

$$P(T) = \frac{\mu_{\alpha}}{\sqrt{2\pi\chi_{c}^{2}}} \exp\left[-\frac{(\mu_{\alpha}T - \ln 2 - \chi_{c}^{2})^{2}}{2\chi_{c}^{2}}\right],$$
(6)

which is a Gaussian distribution with mean $(\ln 2 + \chi_c^2)/\mu_{\alpha}$, and variance χ_c^2/μ_{α}^2 . We see that threshold noise shifts the peak of the distribution and at the same time defines its variance.

5

In summary, under the condition of a rapidly fluctuating threshold, $\tau_c \ll T$, we find a non-universal distribution whose shape depends on the relative size of the two variances, accumulation rate and threshold. It interpolates between a Gaussian shape when rates are rather uniform, to a skewed heavy-tailed shape when they exhibit significant variation.

Solving the threshold crossing problem analytically in the limit of slowly fluctuating threshold, $\tau_c \gg T$, is more difficult. However, numerical simulations reveal that once rate variability is present in the model, the division time distribution is insensitive to the threshold correlation time (Fig. 8 in Appendix D). In contrast, for the exponential threshold-crossing problem with a fixed accumulation rate, the distribution shape varies continuously with the threshold characteristic time. Since all experimental data have variable exponential rates, it is sufficient to use our results on the white threshold limit (Eq. 4) when comparing our results to experimental data. This comparison is presented next.

Comparison to data: In the comparison of our results to experiments, the parameters that define the process $-\mu_{\alpha}$, σ_{α} , χ_c – will be estimated from the data to provide input to the model. These will be used to predict statistical properties of the division time, its distribution and relation between moments, as predicted by the results presented above. These predictions, in turn, will be tested against the empirical statistical properties of division times in the data.

Many groups have utilized mother machine microfluidic devices to monitor single cells as they grow and divide over multiple cycles (11, 18–20, 23, 28, 29, 36, 37). In these experiments typically cell size, and sometimes fluorescently labeled proteins, are measured. Thus, accumulation rates are directly measurable from the data, and their mean and variance are estimated. Additionally, we estimate the stochastic properties of the threshold – mean and variance – from the final value of these variables just before division.

Fig. 2 displays, in its middle panel, a parameter plane consisting of these two empirically estimated CVs: the plane of (χ_c, χ_α) . Each point represents one experiment where single-cell traces were pooled to estimate the parameters. It can be seen that χ_α (y axis, quantifying variability in accumulation rates) spans a broad range of approximately ten-fold between 0.05 to 0.5. In comparison, χ_c (x-axis, quantifying variability in threshold) is more restricted between 0.1 and 0.27. The two parameters are weakly correlated across experiments, so that extreme cases where one is dominant over the other are not found in the data. Nevertheless, they span both regions where either one is large than the other – both sides of the diagonal. Two of the data-sets contain both cell size and protein measurements, shown in the same color with different symbols. The location of these two types of measurements in the plane is very close to one another relative to the entire spread; the significance of this observation will become clear shortly.



FIG. 2. Single-cell data agrees with theoretical predictions. Plot of χ_{α} vs χ_c obtained from several experimental data sets (11, 18–20, 28, 29, 36, 37), which span two different regimes, $\chi_{\alpha} > \chi_c$ and $\chi_{\alpha} < \chi_c$. The distribution of cell division times are more skewed and deviate more from the Gaussian (dashed lines) when $\chi_{\alpha} > \chi_c$ (histograms on the left); this deviation is smaller when $\chi_{\alpha} < \chi_c$ (histograms on the right). In the plot of cell division time distribution, the bars are obtained from different experiments, solid lines are obtained from Eq. (4), and dashed lines are Gaussian fits.

We used these two parameters and the average μ_{α} to compute the distribution of threshold-crossing events from our model, and plotted them (solid lines) together with the measured division time distribution (colored histograms); four plots are displayed in the figure, spanning different regimes of the parameter plane. Other data sets are shown in Fig. (4) of Appendix A. Note that there are no fitting parameters in this procedure. The agreement is excellent in all cases, and even in the Gaussian-like limit our model provides a better fit than a Gaussian distribution (dashed black lines). This agreement indicates that the stochastic process we defined indeed captures the main determinants of cell division at a statistical level, over a range of qualitatively different distribution.

For cell division to occur, several different proteins must accumulate to their functional threshold (28). The question arises, which protein should we measure in order to compare to predictions of the theory? In particular, the data of Fig. 2 correspond to cell size and arbitrary proteins - not necessarily related to cell division. It is not *a-priori* obvious why these measurements would predict division time distributions as well as they do. To understand the relevance of these data, we note that different proteins' accumulation rates are tightly correlated to one another and to the accumulation rate of cell size, while most variability is expressed between one cell cycle and the next (18–20). This is demonstrated in Fig. 5(b) of (19) and Fig. S5 of (38). It suggests that division time statistics may not

be sensitive to the phenotype chosen – specific protein, or cell size – as long as it is an abundant component, accumulating smoothly and approximately exponentially. In support of this hypothesis, in Fig. 5 of Appendix A we plot the generation time distributions (Eq. 4) derived from parameter values { $\mu_{\alpha}, \sigma_{\alpha}, \chi_c$ } of both the cell size and protein data, for those experiments where the two sets of measurements are available. Both predictions (solid and dash-dot lines) are shown together with the empirical histograms, highlighting the similarity between them and their consistency with the data.



FIG. 3. Moments of cell division time depend on model parameters as predicted. If exponential accumulation to threshold is a strong determinant of cell division time, then scaling relations between moments of two random variables – exponential accumulation rate and cell division time – should obey specific predictions. (a) The means are inversely proportional and (b) the coefficient of variations are linearly proportional. (c) Coefficient of variation and (d) skewness of cell division time distribution with $\chi = \chi_{\alpha}/\chi_c$. The symbols are obtained from several experimental data and the green dots are obtained from the theory (Eq. 4). The value of the parameters μ_{α} , χ_{α} , and χ_c are chosen randomly from the uniform distribution, respectively, $\mathcal{U}_{[0.003,0.2]}$, $\mathcal{U}_{[0.03,0.45]}$, and $\mathcal{U}_{[0.08,0.25]}$, which span the broad range of experimental data. The green line is the green dots' mean plot, and the green shaded region is the error across the mean line.

Another way to test our theory is to examine the relations between moments of division times and parameters of the model. Fig. 3 shows these relations as estimated from the data (colored symbols), together with the theoretical prediction of our model (solid lines). First, we note that empirically the mean division time is governed by the mean inverse accumulation rate (Fig. 3a). The solid line is very close to $\mu_T = \ln 2/\mu_{\alpha}$ (dashed line), which is the hallmark of the exponential threshold-crossing model. A more refined calculation shows that in the presence of threshold fluctuations our model predicts $\mu_T \approx \ln 2(1+\chi_{\alpha})/\mu_{\alpha}$,

providing a small deviation from the deterministic limit (Appendix B). This theoretical finding also agrees well with experimental data, as plotted in Fig. 6 of Appendix B.

Moving to the second moment, our model predicts that the CV of division times increases with CV of accumulation rates (Fig. 3b) and with CV of the threshold fluctuations (Fig. 3c). In each of these panels, the parameter not presented – χ_c in b and χ_{α} in c – take on a range of values. The solid line shows the prediction of the model using the average of this parameter, whereas the light shaded region depicts the margins obtained for the range of different hidden parameters. Dark green points show the result of the theory for randomly chosen values of the hidden parameters within the experimental range. Finally, the skewness depends on the ratio between two CVs (Fig. 3d); as mentioned above, the distribution interpolates Gaussian when $\chi_{\alpha} < \chi_c$ to heavy-tailed when $\chi_{\alpha} > \chi_c$, a property reflected by the increase in skewness. These results, again without fitting parameters, indicate that the exponential threshold-crossing model captures correctly the dependence of division time moments on the parameters of the accumulation rate and threshold fluctuations.

Discussion

Cell division time, i.e., the time between consecutive divisions, is an important phenotype that has been of interest in studies since the early 20th century. In the last decade, as large samples of quantitative measurements became available, interest was renewed in bacterial growth and division at the single cell level; the statistical properties of division times provide important information for the understanding of these processes. In this work, we theoretically analyzed the cell-division time statistics of an exponentially accumulating cellular component by considering division as a threshold crossing problem, i.e., a cell will divide when this component crosses a threshold copy number. This model, with the component being a divisome protein that needs to accumulate in order to carry out division, was shown to be consistent with single-cell data in several recent publications, describing both steady-state (11) and transient (12) growth. A new ingredient in our model, not considered in previous work, is the variability in exponential accumulation rates between consecutive cycles in the same cell. Such variability is observed, to different degrees, in multiple datasets (see Fig. 2) and should therefore be accounted for. Indeed, our results show that the statistical properties of accumulation rates are major predictors of the division time statistics. We provide an analytical calculation of division time distribution for symmetric division and stochastic accumulation rate and threshold, which agrees with a large number of experimental data sets.

We find that cell-division time statistics are determined by three model parameters. The mean accumulation rate (μ_{α}) sets the scale of division times; its CV (χ_{α}) , and the CV of the stochastic threshold (χ_c) , determine the distribution shape. In agreement with the

9

exponential threshold crossing model, the mean division time measured across a large set of single-cell measurements is determined by the mean accumulation rate: To very high precision (Fig. 3a), $\mu_T = \ln 2/\mu_{\alpha}$, reflecting the average fold-growth of 2 that balances symmetric division. The CV of division time in these experiments is strongly correlated with the CVs of the two random component, that of growth rate: $\chi_T \sim \chi_{\alpha}$ (Fig. 3b), and that of threshold $\chi_T \sim \chi_c$ (Fig. 3c), in excellent agreement with theory.

Our model predicts also the full distribution of division times: this is found to be of non-universal shape, that depends primarily on the ratio between the two CVs, χ_{α} and χ_c . If $\chi_{\alpha} \ll \chi_c$ then the distribution is Gaussian, while in the other limit $\chi_{\alpha} \gg \chi_c$ it is the reciprocal of the Gaussian distribution, which has a heavy power-law tail $\sim T^{-2}$. In between our general solution interpolates between these behaviors. Although experimental data do not lie in these extremes, they nevertheless span a broad enough range of $\chi = \chi_{\alpha}/\chi_c$ to exhibit both types of distribution in good agreement with the theory. This prediction is reflected also in the increase of division time skewness with the ratio χ , also in good agreement with the data (Fig. 3d).

The range of division time distributions is revealed when considering different levels of variability in accumulation rates. Previous work has argued for Gaussian-like division time distributions that collapse by scaling (24, 28, 29), and can be reconstructed by models with deterministic exponential accumulation rates (25, 39). While this framework holds for some bacterial data-sets, others exhibit a much broader variability in accumulation rates (11, 18–20, 37). Our analysis suggests that different experiments, possibly due to slightly different conditions or culture details, span a range of behaviors from highly uniform to highly varying accumulation rates. The success of our model to describe this range of behaviors indicates that exponential threshold-crossing captures a fundamental principle of cell division common to all these conditions. In future work it will be interesting to test this model further in other microorganisms. More generally, it remains an open question whether the same principles found here for bacteria apply to cell division in higher organisms.

A common heuristic approach for distinguishing candidate control strategies is to examine correlation plots, and compare them to the prediction of several candidate models. Until now, these models have assumed negligible variability in exponential accumulation rates and fixed thresholds. The empirical plot of a variable added over the cell cycles, as a function of its initial value at cycle start, has served as a common such critirion (6, 28, 39). Specifically it is reasoned that in threshold control, smaller initial values would result in larger added values to reach the same threshold, thus inducing a negative correlation in the empirical plot. However, this categorization breaks down when threshold dynamics are taken into account. It was recently shown that the same control mechanism – an exponential threshold-crossing process with a fixed exponent – can produce different empirical slopes,

10

depending on the threshold correlation time (27). This result is reproduced both analytically and numerically in Appendix C, showing that the correlation between added protein or cell size Δn and its initial value n(0), changes from a positive to zero to a negative correlation, as the ratio between cell division time and threshold correlation time (Fig. 7a). We extend this calculation to a model with variable accumulation rate in Fig. 7c, and find that the same phenomenon occurs. This highlights the caution that needs to be exercised when reverse-engineering correlation plots by a limited set of alternative simple models.

Another important correlation is that of division time with initial values. This is expected to be negative for a threshold-crossing process regardless of the source of noise (see Appendix C). Indeed, Fig. 7c,d show that with or without variability of accumulation rate, the empirical slope is negative. Nevertheless, the presence of rate variability has an effect on the correlation plot: it renders the negative slope insensitive to the correlation time. We have seen that also the distribution shape looses its sensitivity to threshold correlation time in the presence of accumulation rate variability (see Appendix D, Fig. 8d,e). These results provide some intuition as to why the statistics of division times are well predicted by a relatively simple model with the main input being statistics of accumulation rates.

An intriguing finding that emerges when comparing our theory to single-cell data, is that division time statistics can be equally well predicted from protein accumulation measurements of various proteins (not necessarily divisome proteins), or even from cell-length measurements. Fig. 5 compares the data with the two predicting curves. This may seem surprising at first, but can be understood once we recall the strong correlations between accumulation rates of different highly expressed proteins and cell size across cycles (18–20). A related phenomenon arises in (12), where dynamics of an arbitrary measured protein was used as a proxy for divisiome proteins and was useful in explaining transition between growth media. This behavior, where the dynamics of a multi-dimensional system can be reduced to one or a few variables, is a hallmark of balanced growth, which in turn characterizes strong interactions and dynamic coupling among system components. Previous models have pointed to ribosomal proteins as providing the auto-catalytic component that indirectly induces exponential-like accumulation of other proteins as well as cell size (19, 20, 40-42). More generally, different interactions among cellular components can indirectly induce autocatalytic cycles that cause correlated exponential-like accumulation in multiple components (24, 40). Further theoretical work is required to understand minimal model requirements that account for variable but correlated dynamics across components (19).

Acknowledgements: This research is supported in part by the Israeli Science Foundation (grant number 155/18, joint N.B. and K.B.), and Binational Science Foundation (grant number2016376, N.B.).

11

Appendix A: Calculation of division time with rate noise and stochastic threshold

In this work, we have applied the first passage problem (FPP) framework to analytically solve the cell division time statistics of exponentially growing bacteria. Cells divide when the copy number of a specific divisome proteins n(t) reaches the fluctuating threshold, c(t)for the first time. Thus cell division is equivalent to the FPP with the stochastic c(t) to a shifting absorbing boundary at n(t). If $G(c(t), t|c(t_0), t_0)$ is the probability density of c(t)conditioned by known $c(t_0)$, then the survival probability that the cell has not divided till time t can be expressed as

$$S(t|c(t_0), t_0) = \int_{n(t)}^{\infty} G(c(t), t|c(t_0), t_0) dc,$$
(A1)

which relates to the probability density of first passage time as below

$$F(t|c(t_0), t_0) = -\frac{\partial S(t|c(t_0), t_0)}{\partial t}.$$
(A2)

Thus, the distribution of cell division time follows $P(T) = F(t = T + t_0)$. If the threshold c(t) follows the OU process (Eq. 3) then the probability, $G(\tilde{c}, t | \tilde{c}', t')$, with normalised variable $\tilde{c} = c/\mu_c - 1$, can be given as a Gaussian distribution (t > t') (35)

$$G(\tilde{c},t|\tilde{c}',t') = \sqrt{\frac{1}{2\pi\chi_c^2(1-e^{2\gamma(t-t')})}} \exp\left[-\frac{(\tilde{c}-\tilde{c}'e^{2\gamma(t-t')})^2}{2\chi_c^2(1-e^{2\gamma(t-t')})}\right],$$
(A3)

where χ_c is the CV of threshold with $\chi_c^2 = \sigma_c^2/\mu_c^2 = D/\gamma\mu_c^2$. For symmetric division and fixed accumulation rate, Eq.(A1)-Eq.(A3) lead to the cell division time distribution for exponentially growing cell, $n(t) = \frac{\mu_c}{2}e^{\alpha(t-t')}$ as (27)

$$P(T|\alpha) = \frac{\gamma(1 - e^{-2\gamma T})^{-3/2}}{\sqrt{2\pi\chi_c^2}} \left(\frac{\alpha}{2\gamma}e^{\alpha T}(1 - e^{-2\gamma T}) + e^{-2\gamma T}\left(1 - \frac{1}{2}e^{\alpha T}\right)\right) exp\left[-\frac{\left(1 - \frac{1}{2}e^{\alpha T}\right)^2}{2\chi_c^2(1 - e^{-2\gamma T})}\right]$$
(A4)

In the limit $\tau_c \ll T$, i.e., $\gamma T \gg 1$ then the cell division time distribution in Eq. (A4) will be

$$P(T|\alpha) = \frac{\alpha e^{\alpha T}}{2\sqrt{2\pi\chi_c^2}} exp\left[-\frac{1}{2\chi_c^2}\left(1-\frac{1}{2}e^{\alpha T}\right)^2\right].$$
 (A5)

Now, to incorporate the fluctuations in the accumulation rate we have to multiply the probability distribution function of α , $P(\alpha)$ with Eq. (A5) and integrate over α :

$$P(T) = \int_0^\infty P(T|\alpha) P(\alpha) d\alpha.$$
 (A6)

12

The distribution of exponential accumulation rate α follows Gaussian as shown in Fig. 1c, which is obtained from the single cell experimental date of *E. coli* (18).

$$P(\alpha) = \frac{1}{\sqrt{2\pi\sigma_{\alpha}^2}} exp\left(-\frac{(\alpha - \mu_{\alpha})^2}{2\sigma_{\alpha}^2}\right).$$
 (A7)

But, one of the α dependent terms inside the integrand is, $\exp\left[-\frac{1}{2\chi_c^2}\left(1-\frac{1}{2}e^{\alpha T}\right)^2\right]$, which is not exactly integrable. So, we do the Taylor expansion of the exponent around the maximum value of α (= ln 2/T) and neglect higher order terms $\mathcal{O}((\alpha T)^3/\chi_c^2)$. Using this approximation, Eq. (A5) turns into the following form:

$$P(T|\alpha) = \frac{\alpha e^{\alpha T}}{2\sqrt{2\pi\chi_c^2}} \exp\left[-\frac{1}{2\chi_c^2} \left(\alpha T - \ln 2\right)^2\right] = \frac{\alpha e^{\chi_c^2/2}}{\sqrt{2\pi\chi_c^2}} \exp\left[-\frac{\left(\alpha T - \ln 2 - \chi_c^2\right)^2}{2\chi_c^2}\right].$$
 (A8)

This approximation renders the probability density in Eq. (A8) unnormalization, with an additional term $e^{\chi_c^2/2}$, which is very close to 1 as $\chi_c^2 \ll 1$. Approximating $e^{\chi_c^2/2} \approx 1$ restores normalization and we have a Gaussian

$$P(T|\alpha) = \frac{\alpha}{\sqrt{2\pi\chi_c^2}} \exp\left[-\frac{\left(\alpha T - \ln 2 - \chi_c^2\right)^2}{2\chi_c^2}\right].$$
 (A9)

Inserting Eqs. (A7) and (A9) into Eq. (A6) we can obtain the cell division time distribution

$$P(T) = \frac{(\mu_{\alpha}\chi_c^2 + \sigma_{\alpha}^2 T(\chi_c^2 + \ln 2))}{\sqrt{2\pi}(\chi_c^2 + (\sigma_{\alpha}T)^2)^{3/2}} \exp\left[-\frac{(\mu_{\alpha}T - \chi_c^2 - \ln 2)^2}{2(\chi_c^2 + (T\sigma_{\alpha})^2)}\right],$$
(A10)

as described in Eq. (4). Furthermore, scaling the division time by $Z = \mu_{\alpha} T$ gives the distribution of the new scaled variable

$$P(Z) = \frac{\chi_c^2 + Z\chi_\alpha^2(\chi_c^2 + \ln 2)}{\sqrt{2\pi}(\chi_c^2 + (Z\chi_\alpha)^2)^{3/2}} \exp\left[-\frac{(Z - \chi_c^2 - \ln 2)^2}{2(\chi_c^2 + (Z\chi_\alpha)^2)}\right],$$
(A11)

which depends only on CVs, χ_{α} and χ_c . This result implies that μ_{α} is the scale variable, and two CVs play a role as the shape parameter of division time distribution.





FIG. 4. Comparison of theoretical division time distribution to single-cell data. Cell division time distribution from different published experimental data (bar histograms): (a-d)(19), (e) (18), (f) (37), (g-i) (11), (j-l) (29), (m-r) (28). The solid curves are obtained from our theoretical model (Eq. 4), where the parameters, μ_{α} , σ_{α} , and χ_c are estimated directly from the respective experimental data of cell length, which are also displayed in Fig. (3).

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.21.504683; this version posted August 21, 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-ND 4.0 International license.



FIG. 5. Division time is insensitive to cell length measurements or protein. Cell division time distribution in Eq. 4 depends on three quantities, μ_{α} , σ_{α} , and χ_c , which can be estimated either from protein data or from cell size data. We plot experimental division time distributions (bar histogram) along with P(T) (Eq. 4) for two input types; $\{\mu_{\alpha}, \sigma_{\alpha}, \chi_c\}_{cell length}$ (solid line) and $\{\mu_{\alpha}, \sigma_{\alpha}, \chi_c\}_{protein number}$ (dash-dot line), estimated from the single cell experimental data of *E. coli* (19, 20). The measured parameter values of cell length and protein data for both experiments are exhibited in Fig. (3).

15

Appendix B: Approximated calculation of mean division time with rate noise

In the presence of variability in the accumulation rate from one cycle to another and stochastic threshold, we find an exact distribution of cell division time as presented in Eq. (4). However, it is almost impossible to calculate an exact mathematical form of cell division moments from this distribution. In the limit when $\chi_c > \chi_{\alpha}$ we obtain Gaussian distribution (Eq. 6) with well defined mean, $\mu_T = (\ln 2 + \chi_c^2)/\mu_{\alpha} \approx \ln 2/\mu_{\alpha}$, as $\chi_c^2 \ll 1$. In the other limit, $\chi_{\alpha} > \chi_c$, there are no well-defined moments of division time distribution as it is reciprocal of Gaussian (Eq. 5). However, we can do an approximated calculation of mean cell division time using the reciprocal Gaussian distribution (Eq. 5) as below:

$$\mu_{T} = \frac{\ln 2}{\sqrt{2\pi}\sigma_{\alpha}} \int \frac{1}{T} \exp\left[-\frac{\left(\frac{\ln 2}{T} - \mu_{\alpha}\right)^{2}}{2\sigma_{\alpha}^{2}}\right] dT$$

$$= \frac{\ln 2}{\sqrt{2\pi}\sigma_{\alpha}} \int \frac{1}{Z} \exp\left[-\frac{\left(Z - \mu_{\alpha}\right)^{2}}{2\sigma_{\alpha}^{2}}\right] dZ \quad [\text{change of variable } Z = \ln 2/T \approx (\mu_{\alpha} \pm \sigma_{\alpha})]$$

$$\approx \frac{\ln 2}{\mu_{\alpha}} (1 + \chi_{\alpha}) \quad \left[\frac{1}{Z} = \frac{1}{\mu_{\alpha}(1 \pm \chi_{\alpha})} \approx \frac{1}{\mu_{\alpha}} (1 + \chi_{\alpha}) \because \chi_{\alpha} < 1\right], \quad (B1)$$

which provides that the mean cell division time is proportional with the noise in accumulation rate, χ_{α} , as observed from several single-cell experimental data (Fig. 6).



FIG. 6. Mean cell division time is proportional with CV of accumulation rate. Variation of $\mu_T \mu_{\alpha}$ with χ_{α} , which exhibits positive correlation. The symbols are obtained from several experimental data (11, 18–20, 28, 29, 36, 37), and green dots are obtained from the theory (Eq. 4), where μ_{α} , σ_{α} , and χ_c are chosen randomly from the uniform distribution within the range of experimental values. The green line is the mean plot of the green dots and green shaded region is the error across the mean line.

16

Appendix C: Covariation of initial and added values

There have been extensive experimental studies and analyses on the dependence of the added cell size Δn and the division time T on the initial value n(0) of exponentially growing bacteria under different growth conditions. Here, we analyze joint variability among these quantities in our model, namely in the presence of accumulation noise and threshold dynamics.

The joint variability of two random variables can be measured by covariance, defined as cov(X, Y) = E(XY)-E(X)E(Y), where E is the expectation (average) value. Following this definition one can exactly calculate the $cov(\Delta n, n(0))$ in the case of symmetric division (n(0) = c(0)/2) and n(T) = c(T), as bacterial cell division is a threshold crossing process. Thus, $cov(\Delta n, n(0))$ will be given as following

$$C_n = \frac{\sigma_c^2}{4} \left(2e^{-\gamma T} - 1 \right), \tag{C1}$$

by using the $\operatorname{cov}(c(T), c(0)) = \sigma_c^2 e^{-\gamma T}$ (32). From Eq. (C1), one can clearly understand that the covariance between Δn and n(0) does not depend on the accumulation rate α ; however, it can span from positive to negative depending on the value of threshold correlation time $\tau_c = 1/\gamma$ (Fig. 7a, c). If $\tau_c \gg T$, then $e^{-\gamma T} \approx 1$, which gives $C_n > 0$ (positive correlation); on the other hand, if $\tau_c \ll T$, then $e^{-\gamma T} \approx 0$ and $C_n < 0$ (negative correlation). Similarly, $\tau_c \approx T$ gives $e^{-\gamma T} \approx 0.5$, which implies $C_n \approx 0$ i.e., weak correlation.

Similarly, we can measure the joint variability among T and n(0) by calculating cov(T, n(0)), which is not exactly analytically solvable. However, using a crude 1st order approximation of $\ln c(T) \approx c(T) - 1$, we can roughly understand the relation between T and n(0). Using this approximation the cov(T, n(0)) leads to

$$C_T \approx -\frac{\sigma_c^2}{2\alpha} (1 - e^{-\gamma T}).$$
(C2)

From the expression of C_T (Eq. C2), it is clear that the cov(T, n(0)) is always negative independent of the value of threshold correlation time τ_c ; though, the value of $|C_T|$ will decrease if we increase τ_c (Fig. 7b). However, unlike C_n , the slope C_T does depend on the accumulation rate α . Due to the limitations of approximated calculation, the analytical form of C_T is unable to explain why the noise in accumulation rate χ_{α} makes this correlation almost insensitive to threshold correlation time (Fig. 7d).

For the numerical analysis we do the time evolution of cell size/protein number for k^{th} cell cycle, $n_k(t) = n_k(0)e^{\alpha_k t}$ until $n_k(t)$ hits the threshold c(t). For the threshold dynamics as presented in Eq. (3), we simultaneously do the Euler Maruyama simulation with time interval 0.1 min. We record the time and cell size at the point when $n_k(t)$ crosses the

17

threshold, which give, respectively, the cell division time and cell size at division. We continue these steps for 10^4 realizations to get cell division time, and elongated cell size.



FIG. 7. The presence of accumulation rate variability renders the negative $\operatorname{cov}(T, n(0))$ insensitive to the threshold correlation time. Variation of (a, c) Δn and (b, d) cell division time T with cell size at birth, n(0) for three different time scales of threshold. Here the plots are obtained from the simulation with $\mu_{\alpha} = 0.023 \ min^{-1}$, $\chi_c = 0.1$, $\chi_{\alpha} = 0$ (a, b), $\chi_{\alpha} = 0.3$ (c, d), and $\tau_c = 5$ (< T), $40(\approx T)$, 200 (> T) and the unit of time is min.

18



Appendix D: Simulated and calculated P(T) for different τ_c and χ_{α} with $\chi_c = 0.1$

FIG. 8. Noise in accumulation rate makes the distribution insensitive to the threshold correlation time. In the left panel, for constant accumulation rate ($\chi_{\alpha} = 0$), we plot (a) the dynamics of protein number, threshold, and (b) cell division time distribution for $\tau_c = 200$ min (> T) and $\tau_c = 2 \min (< T)$. In the right panel we do the same plots; (c) trajectories of threshold and protein number for $\chi_{\alpha} \neq 0$ and division time distribution for (d) $\chi_{\alpha} = 0.06 (< \chi_c)$ and (e) χ_{α} = 0.4 (> χ_c). For (b) the solid lines are obtained from Eq. (A8), and Eq. (4) gives the solid lines of (d), (e). Histograms are obtained from the simulation with $\chi_c = 0.1$, and $\mu_{\alpha} = 0.02 \text{ min}^{-1}$.

- S. Cooper and C. E. Helmstetter, "Chromosome replication and the division cycle of Escherichia coli Br," *Journal of molecular biology*, vol. 31, no. 3, pp. 519–540, 1968.
- M. Slater and M. Schaechter, "Control of cell division in bacteria," *Bacteriological reviews*, vol. 38, no. 2, pp. 199–221, 1974.
- [3] W. D. Donachie, "The cell cycle of escherichia coli," Annual review of microbiology, vol. 47, no. 1, pp. 199–230, 1993.
- [4] M. Osella, E. Nugent, and M. Cosentino Lagomarsino, "Concerted control of Escherichia coli cell division," *Proceedings of the National Academy of Sciences*, vol. 111, no. 9, pp. 3431–3435, 2014.
- [5] A. Zaritsky and C. L. Woldringh, "Chromosome replication, cell growth, division and shape: a personal perspective," *Frontiers in microbiology*, vol. 6, p. 756, 2015.
- [6] L. Willis and K. C. Huang, "Sizing up the bacterial cell cycle," *Nature Reviews Microbiology*, vol. 15, no. 10, pp. 606–620, 2017.
- [7] M. Osella, S. J. Tans, and M. C. Lagomarsino, "Step by step, cell by cell: quantification of the bacterial cell cycle," *Trends in microbiology*, vol. 25, no. 4, pp. 250–256, 2017.
- [8] S. Jun, F. Si, R. Pugatch, and M. Scott, "Fundamental principles in bacterial physiology—history, recent progress, and the future with focus on cell size control: a review," *Reports* on Progress in Physics, vol. 81, no. 5, p. 056601, 2018.
- [9] P. Y. Ho, J. Lin, and A. Amir, "Modeling cell size regulation: From single-cell-level statistics to molecular mechanisms and population-level effects," *Annual review of biophysics*, vol. 47, pp. 251–271, 2018.
- [10] T. den Blaauwen, L. W. Hamoen, and P. A. Levin, "The divisome at 25: the road ahead," *Current opinion in microbiology*, vol. 36, pp. 85–94, 2017.
- [11] 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.
- [12] M. Panlilio, J. Grilli, G. Tallarico, I. Iuliani, B. Sclavi, P. Cicuta, and M. C. Lagomarsino, "Threshold accumulation of a constitutive protein explains E. coli cell-division behavior in nutrient upshifts," *Proceedings of the National Academy of Sciences*, vol. 118, no. 18, 2021.
- [13] 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.
- [14] 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*,

20

vol. 114, no. 4, pp. 693–698, 2017.

- [15] K. Biswas, M. Shreshtha, A. Surendran, and A. Ghosh, "First-passage time statistics of stochastic transcription process for time-dependent reaction rates," *The European Physical Journal E*, vol. 42, no. 2, pp. 1–11, 2019.
- [16] K. Biswas and A. Ghosh, "Timing effciency in small-RNA-regulated post-transcriptional processes," *Physical Review E*, vol. 101, no. 2, p. 022418, 2020.
- [17] K. Biswas and A. Ghosh, "First passage time in post-transcriptional regulation by multiple small RNAs," *The European Physical Journal E*, vol. 44, no. 2, pp. 1–10, 2021.
- [18] N. Brenner, E. Braun, A. Yoney, L. Susman, J. Rotella, and H. Salman, "Single-cell protein dynamics reproduce universal fluctuations in cell populations," *The European Physical Journal E*, vol. 38, no. 9, pp. 1–9, 2015.
- [19] L. Susman, M. Kohram, H. Vashistha, J. T. Nechleba, H. Salman, and N. Brenner, "Individuality and slow dynamics in bacterial growth homeostasis," *Proceedings of the National Academy of Sciences*, vol. 115, no. 25, pp. E5679–E5687, 2018.
- [20] A. S. Sassi, M. Garcia-Alcala, M. Aldana, and Y. Tu, "Protein Concentration Fluctuations in the High Expression Regime: Taylor's Law and Its Mechanistic Origin," *Physical review X*, vol. 12, no. 1, p. 011051, 2022.
- [21] S. D. Talia, J. M. Skotheim, J. M. Bean, E. D. Siggia, and F. R. Cross, "The effects of molecular noise and size control on variability in the budding yeast cell cycle," *Nature*, vol. 448, no. 7156, pp. 947–951, 2007.
- [22] M. Godin, F. F. Delgado, S. Son, W. H. Grover, A. K. Bryan, A. Tzur, P. Jorgensen, K. Payer, A. D. Grossman, M. W. Kirschner, *et al.*, "Using buoyant mass to measure the growth of single cells," *Nature methods*, vol. 7, no. 5, pp. 387–390, 2010.
- [23] P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun, "Robust growth of Escherichia coli," *Current biology*, vol. 20, no. 12, pp. 1099–1103, 2010.
- [24] S. Iyer-Biswas, C. S. Wright, J. T. Henry, K. Lo, S. Burov, Y. Lin, G. E. Crooks, S. Crosson, A. R. Dinner, and N. F. Scherer, "Scaling laws governing stochastic growth and division of single bacterial cells," *Proceedings of the National Academy of Sciences*, vol. 111, no. 45, pp. 15912–15917, 2014.
- [25] S. Iyer-Biswas, G. E. Crooks, N. F. Scherer, and A. R. Dinner, "Universality in stochastic exponential growth," *Physical review letters*, vol. 113, no. 2, p. 028101, 2014.
- [26] 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.
- [27] L. Luo, Y. Bai, and X. Fu, "Stochastic threshold in cell size control," arXiv preprint arXiv:2203.02987, 2022.

21

- [28] S. Taheri-Araghi, S. Bradde, J. T. Sauls, N. S. Hill, P. A. Levin, J. Paulsson, M. Vergassola, and S. Jun, "Cell-size control and homeostasis in bacteria," *Current biology*, vol. 25, no. 3, pp. 385–391, 2015.
- [29] Y. Tanouchi, A. Pai, H. Park, S. Huang, R. Stamatov, N. E. Buchler, and L. You, "A noisy linear map underlies oscillations in cell size and gene expression in bacteria," *Nature*, vol. 523, no. 7560, pp. 357–360, 2015.
- [30] A. G. Marr, R. Harvey, and W. Trentini, "Growth and division of Escherichia coli," *Journal of Bacteriology*, vol. 91, no. 6, pp. 2388–2389, 1966.
- [31] J. Männik, F. Wu, F. J. Hol, P. Bisicchia, D. J. Sherratt, J. E. Keymer, and C. Dekker, "Robustness and accuracy of cell division in Escherichia coli in diverse cell shapes," *Proceedings* of the National Academy of Sciences, vol. 109, no. 18, pp. 6957–6962, 2012.
- [32] C. W. Gardiner, Handbook of stochastic methods, vol. 3. springer Berlin, 1985.
- [33] L. M. Ricciardi and S. Sato, "First-passage-time density and moments of the ornsteinuhlenbeck process," *Journal of Applied Probability*, vol. 25, no. 1, pp. 43–57, 1988.
- [34] N. G. Van Kampen, Stochastic processes in physics and chemistry, vol. 1. Elsevier, 1992.
- [35] H. Risken, "Fokker-planck equation," in *The Fokker-Planck Equation*, pp. 63–95, Springer, 1996.
- [36] N. Nordholt, J. H. van Heerden, and F. J. Bruggeman, "Biphasic cell-size and growth-rate homeostasis by single bacillus subtilis cells," *Current Biology*, vol. 30, no. 12, pp. 2238–2247, 2020.
- [37] A. Stawsky, H. Vashistha, H. Salman, and N. Brenner, "Multiple timescales in bacterial growth homeostasis," *Iscience*, vol. 25, no. 2, p. 103678, 2022.
- [38] N. Brenner, E. Braun, J. S. Rotella, and H. Salman, "Single-cell protein dynamics reproduce universal fluctuations in cell populations,"
- [39] A. Amir, "Cell size regulation in bacteria," *Physical review letters*, vol. 112, no. 20, p. 208102, 2014.
- [40] P. P. Pandey, H. Singh, and S. Jain, "Exponential trajectories, cell size fluctuations, and the adder property in bacteria follow from simple chemical dynamics and division control," *Physical Review E*, vol. 101, no. 6, p. 062406, 2020.
- [41] L. Li, "Balanced biosynthesis and trigger threshold resulting in a double adder mechanism of cell size control," *Communications in Theoretical Physics*, vol. 73, no. 8, p. 085601, 2021.
- [42] K. C. Cylke and S. Banerjee, "Super-exponential growth and stochastic shape dynamics in rod-like bacteria," *bioRxiv*, 2022.















Experimental data



Simulation & Calculation





