

# Protein Concentration Fluctuations in the High Expression Regime: Taylor's Law and Its Mechanistic Origin

Alberto Stefano Sassi<sup>1,†</sup>, Mayra Garcia-Alcala<sup>2,3,†</sup>, Maximino Aldana<sup>3,4</sup> and Yuhai Tu<sup>1,\*</sup>

<sup>1</sup>IBM T. J. Watson Research Center, Yorktown Heights, New York 10598, USA

<sup>2</sup>Department of Molecular and Cellular Biology, John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA

<sup>3</sup>Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos 62210, México

<sup>4</sup>Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Ciudad de México 04510, México

 (Received 9 April 2021; revised 17 August 2021; accepted 25 January 2022; published 17 March 2022)

Protein concentration in a living cell fluctuates over time due to noise in growth and division processes. In the high expression regime, variance of the protein concentration in a cell is found to scale with the square of the mean, which belongs to a general phenomenon called Taylor's law (TL). To understand the origin for these fluctuations, we measure protein concentration dynamics in single *Escherichia coli* cells from a set of strains with a variable expression of fluorescent proteins. The protein expression is controlled by a set of constitutive promoters with different strength, which allows one to change the expression level over 2 orders of magnitude without introducing noise from fluctuations in transcription regulators. Our data confirm the square TL, but the prefactor has a cell-to-cell variation independent of the promoter strength. Furthermore, distributions of the normalized protein concentration for different promoters collapse onto the same curve. To explain these observations, we use a minimal mechanistic model to describe the stochastic growth and division processes in a single cell with a feedback mechanism for regulating cell division. In the high expression regime where extrinsic noise dominates, the model reproduces our experimental results quantitatively. By using the mean-field approximation in the minimal model, we show that the stochastic dynamics of protein concentration is described by a Langevin equation with multiplicative noise. The Langevin equation has a scale invariance which is responsible for the square TL. By solving the Langevin equation, we obtain an analytical solution for the protein concentration distribution function that agrees with experiments. The solution shows explicitly how the prefactor depends on strength of different noise sources, which explains its cell-to-cell variability. By using this approach to analyze our single-cell data, we find that the noise in production rate dominates the noise from cell division. The deviation from the square TL in the low expression regime can also be captured in our model by including intrinsic noise in the production rate.

DOI: [10.1103/PhysRevX.12.011051](https://doi.org/10.1103/PhysRevX.12.011051)

Subject Areas: Biological Physics, Statistical Physics

## I. INTRODUCTION

Protein concentrations inside a single cell determine functions and behaviors of the cell [1–5]. Given the small size of a cell, the dynamics of protein concentration is highly stochastic, and there is a significant cell-to-cell variability [6–10]. Protein concentration dynamics in a cell

is determined by gene expression and protein synthesis processes as well as the growth and division processes of the cell, all of which can introduce significant noise [11,12].

In a genome-wide study of protein concentration fluctuations in individual cells by Taniguchi *et al.* [13], it is found that almost all protein concentration distributions can be described by the Gamma distribution characterized by its mean  $\langle c \rangle$  and variance  $\sigma_c^2$ . For genes with low expression levels (fewer than ten proteins per cell cycle),  $\sigma_c^2$  is found to scale linearly with  $\langle c \rangle$ , which can be explained by a simple model of the transcription process and its intrinsic noise [13]. For genes with higher expression levels,  $\sigma_c^2$  is found to scale with  $\langle c \rangle^2$ , which is attributed phenomenologically to contributions from noise in processes that are “extrinsic” to the transcriptional process. However, what the main

\*Corresponding author.

yuhai@us.ibm.com

†These authors contributed equally to this work.

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

sources of this extrinsic noise are and how they lead to the observed scaling behavior in the high expression regime remain unclear.

A power law scaling relationship between variance and mean is generally called Taylor’s law, named after Lionel Taylor, who first observed it in ecology [14] and later found to exist in diverse fields from number theory to epidemiology [15–18]. For protein concentration fluctuations in the high expression regime, the observed scaling relation  $\sigma_c^2 = A\langle c \rangle^2$  corresponds to a square Taylor’s law (TL) with power  $p = 2$ . However, for protein concentration fluctuations, the prefactor  $A$  is not a constant, and it varies within a roughly twofold range [13]. Here, by combining experimental and theoretical studies, we aim to understand the origin of single-cell protein concentration fluctuations in the high expression regime. Our goal is twofold. We want to understand the fundamental reason for the universal scaling behavior observed in protein concentration fluctuations. Equally importantly, we aim to understand how different (dominant) sources of noise contribute to the prefactor  $A$  to explain its observed variation in individual cells and/or under different conditions.

We construct a set of *Escherichia coli* strains that express the fluorescent protein Venus at levels that varies over 2 orders of magnitude. In each strain, we introduce a plasmid that contains a constitutive promoter controlling Venus. This promoter belongs to a set of artificial promoters with variable strength [19]. We study dynamics of the protein concentration by measuring the expression level (fluorescence intensity) and the cell size dynamics in growing and dividing cells by using the mother machine microfluidic device. This allows us to measure single cells for many generations under stable environmental conditions [8,20–24]. Furthermore, the use of the constitutive promoters in plasmids allows us to vary the gene expression level by orders of magnitude without introducing significant additional sources of noise (e.g., those caused by fluctuations in regulators for nonconstitutive promoters or due to the specific chromosomal location of the gene). This setup helps us isolate some of the common factors that contribute to the protein concentration fluctuations, which are difficult to distill from the genome-wide study [13].

The protein copy number fluctuations are studied in *E. coli* by Salman *et al.* [25], where a square TL for the protein copy number fluctuations in growing and dividing cells is reported. By using a phenomenological model [4,26], Brenner *et al.* show that the protein number  $N$  in a growing and dividing cell follows approximately a log-normal distribution under the assumption that  $N$  increases exponentially with respect to time. The only protein number scale comes from the mean, which is fixed in the phenomenological model as a mathematical constraint. However, while previous models [4,26] are able to explain the Taylor’s law in protein number fluctuations phenomenologically, they cannot be used to explain fluctuations of

protein concentration, which depend on fluctuations in both protein number and cell size that are highly correlated in a growing and dividing cell.

In this work, we first present a minimal mechanistic model to describe dynamics of both the protein number and cell size during growth and division in a coherent framework. In our model, the growth of both the protein number and cell size depends on the number of a common production machinery (complex) with production (growth) rates different for the protein number and cell size. The production machinery, which can be characterized by the number of ribosomes ( $R$ ) in the cell, has its own dynamics, which is controlled by the same production and division processes as for a specific protein. Instead of enforcing a mean protein number as an *ad hoc* mathematical constraint as in previous models for protein number fluctuations [26] and cell size regulation [27], a division control variable (molecule  $Z$ ) is introduced in our model.  $Z$  also follows the same production and division dynamics as that of a protein (or  $R$ ). The probability of cell division increases sharply when the number of molecule  $Z$  crosses a certain threshold  $Z_0$  [28–32]. We then use this minimal mechanistic model to study protein concentration fluctuations and compare the theoretical results quantitatively with single-cell data from the mother machine experiments. Furthermore, by using a mean-field approximation in the minimal model, we find that the stochastic dynamics of the concentration  $c$  is governed by a single Langevin equation with multiplicative noise. We solve the Langevin equation analytically, which allows us not only to prove the square TL, but also to show explicitly how the prefactor  $A$  depends on different microscopic noise sources. The deviations from the square TL in the low expression regime are also discussed.

## II. RESULTS

### A. Protein concentration fluctuations follow Taylor’s law with a variable prefactor

We constructed a set of *E. coli* strains that produce the fluorescent protein Venus [Fig. 1(a)]. Each strain contains a plasmid that expresses the Venus gene controlled by a promoter from a set of constitutive promoters with different strengths (see the Appendix A for details about the experiments) [19]. The constitutive promoters are derived from the same DNA strand with small modifications that affect the affinity for the sigma factor. Using the mother machine technique [8,10,22,23] and fluorescent microscopy, we monitor individual cells of such strains under steady-state conditions, with a constant flow of either rich or poor media. For each combination of nutrient conditions and promoter strength, we measure the cell size and the fluorescent intensity protein for approximately 20 different mother cells and approximately 50 cell generations per mother cell. We assume that the intensity of the fluorescence of a cell at time  $t$ ,  $N(t)$ , is proportional to the number

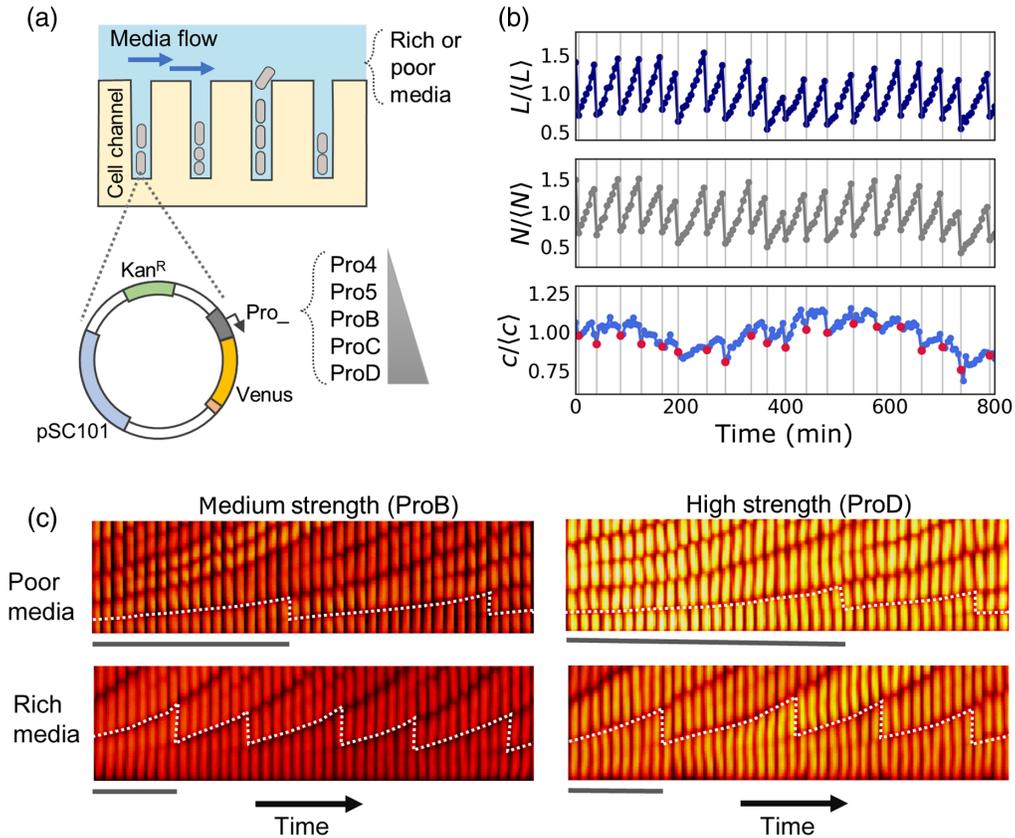


FIG. 1. (a) Schematics of the experimental setup. Using the microfluidic device *mother machine* and fluorescent microscopy, we monitor individual cells of various *E. coli* strains growing under a constant flow of either rich or poor media. Each strain contains a plasmid that codifies for the fluorescent protein Venus under the control of an artificial promoter from a set of promoters with increasing strengths,  $\text{Pro}\{4, 5, B, C, D\}$  [19]. (b) Typical time traces of a single cell containing the ProB promoter and growing in rich medium: normalized cell length  $L$  (top), total fluorescence  $N$  (middle), and fluorescence density  $c (= N/L)$  (bottom). The vertical lines and the red dots at the bottom plot indicate the times of cell division. (c) Kymographs illustrating the fluorescence level of strains with the promoters ProB (left) and ProD (right) controlling Venus expression, growing in poor (top) and rich (bottom) media. Each frame is separated from the following one by a time interval of 5 min. The bars below the kymographs show the span of a cell trace from birth until division.

of Venus protein, and, thus, the fluorescence density (fluorescence divided by cell size, indicated as  $c$ ) can be considered a proxy of the protein concentration.

In Fig. 1(b), we show a typical time trace of the normalized cell length (size)  $L$  (top), total fluorescence  $N$  (middle), and fluorescence density  $c (\equiv N/L)$  (bottom) for many cell generations. Both the protein number (fluorescence intensity) and cell length follow continuous exponential growth that is interrupted by periodic “discontinuous” reductions due to cell division. The fluorescence density  $c$  is much more continuous than  $N$  and  $L$ , and the influence of cell division is not strong. This is easy to understand, because the gross effect of cell division for both  $N$  and  $L$  is canceled out in  $c$ . However, even though cell division does not affect the protein concentration the same way as the protein number or cell size, we note from Fig. 1(b) (bottom) that there is a small but observable change in  $c$  at cell division times (red dots), which suggests that cell division is likely a source of noise for protein concentration.

To study how fluctuations in cell size and protein number depend on the mean growth and transcription rates, we do experiments using several constitutive promoters with different strengths [22] in two nutrient conditions [8]. Qualitatively, as shown in Fig. 1(c), a stronger promoter leads to a higher fluorescence intensity, whereas a rich nutrient condition leads to a faster growth and a shorter average division time  $\Delta T$ .

From the measured time traces  $c_i(t)$  [such as the one shown in Fig. 1(b)] for each mother cell lineage  $i$ , we compute the mean  $\mu_i \equiv \langle c_i \rangle_t$  and variance  $\sigma_{c,i}^2 \equiv \langle c_i^2 \rangle_t - \mu_i^2$ , where the average  $\langle \cdot \rangle_t$  is taken over measurements at different time points. In Fig. 2(a), we show the log-log plot of variance versus the mean for all individual lineages for all ten experimental conditions studied in this paper—different symbols represent the five different promoters, and different colors represent the two different growth conditions. Our data clearly show that on average the variance-mean relation follows the square Taylor’s law with power  $p = 2$

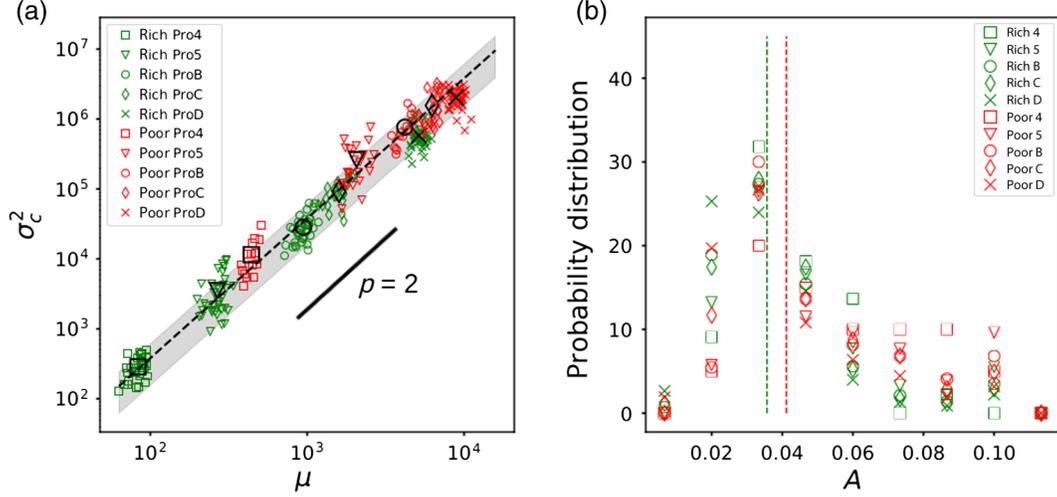


FIG. 2. TL in protein concentration fluctuations. (a) Variance ( $\sigma_c^2$ ) of the protein concentration versus the mean value  $\mu (= \langle c \rangle)$  for individual cells with all experimental conditions. The poor and rich nutrient conditions are labeled by red and green, respectively. Different symbols correspond to cells with promoters with different affinity to the sigma factor. All data from different cells with different promoters and growth condition are overlaid with the square TL:  $\sigma_c^2 = A\mu^2$  with  $A = 0.038 \pm 0.015$ . The black dashed line and the gray thick line around it reflect the mean and standard deviation of the distribution of  $A$ , respectively. The bold black symbols represent the mean and variance averaged over cells with the same experimental conditions. (b) Distributions of  $A$  for all ten experimental conditions. The dashed lines indicate the means of  $A$  for all the cells in rich (green) and poor (red) conditions.

[the dashed line in Fig. 2(a) has slope 2]. However, it is clear from our data that there is significant variability among individual cells.

To characterize the cell-to-cell variability, we define the strength of fluctuation as  $A_i \equiv \sigma_{c,i}^2/\mu_i^2$  for each cell  $i$ . The variance-mean relation can, thus, be expressed as

$$\sigma_{c,i}^2 = A_i \mu_i^2. \quad (1)$$

From Fig. 2(a), it is clear that the prefactor  $A_i$  varies from cell to cell. Importantly, we find that the variation in  $A_i$  is independent of the expression level  $\mu_i$  across cells in all strains we study. We compute the correlation coefficient:

$$C_{A\mu} = \frac{\langle A\mu \rangle - \langle A \rangle \langle \mu \rangle}{\sigma_A \sigma_\mu}, \quad (2)$$

where  $\langle \cdot \rangle$  means averaging over all cells in our experiments and  $\sigma_A^2 = \langle A^2 \rangle - \langle A \rangle^2$  and  $\sigma_\mu^2 = \langle \mu^2 \rangle - \langle \mu \rangle^2$  are the variance for  $A$  and  $\mu$ , respectively. We find that the correlation ( $C_{A\mu} = -0.06$ ) is statistically insignificant, which suggests that the dominant noise sources responsible for the variations of  $A_i$  are independent of the factors that control the gene expression level in our experiments.

In Fig. 2(b), we show the distributions of  $A$  in cells for each of the ten experimental conditions. The distributions do not vary significantly or systematically with experimental conditions. The mean is independent of the promoter strength and has only a weak dependence on the growth condition. Quantitatively, the average  $\langle A \rangle = 0.038$  and the standard deviation  $\sigma_A \approx 0.015$  are found for all the

cells studied here. It is evident from Fig. 2(b) that variations in  $A$  are dominated by cell-to-cell variability among cells within the same experimental condition (promoter strength and growth condition). Consequently, when we average over the population of cells with a specific experimental condition  $\mathcal{E}$ , the resulting averaged variance [ $\sigma_c^2(\mathcal{E})$ ] and mean [ $\mu(\mathcal{E})$ ] satisfy the square TL [33]:

$$\sigma_c^2(\mathcal{E}) = \langle A \rangle \mu^2(\mathcal{E}), \quad (3)$$

with a constant prefactor  $\langle A \rangle$ . The population-averaged variance [ $\sigma_c^2(\mathcal{E})$ ] and mean [ $\mu(\mathcal{E})$ ] are represented by coordinates of the bold black symbols in Fig. 2(a), where the averaged square TL [Eq. (3)] is shown as the dashed line (there is a small error bar much smaller than  $\sigma_A$  due to the finite number of cells for each experimental condition  $\mathcal{E}$ ), and the gray thick line represents the range of  $A$  given by  $\sigma_A$ .

The prefactor  $A$  and its distribution depend on microscopic details of the system such as noise level in protein production and cell division as we show later in the paper. To test the generality of our results, we analyze data from our previous experiments [22], where the same family of promoters used in this study are integrated in the chromosome. As shown in Fig. 10 in Appendix G, the square TL with  $p = 2$  is preserved and the variation of the prefactor  $A$  is independent of the mean expression level with a correlation coefficient  $C_{A\mu} = -0.07$ . Quantitatively, the distribution of the prefactor  $A$  has a mean and standard deviation  $A = 0.083 \pm 0.046$  that are different from those found in our current experiments using plasmids.

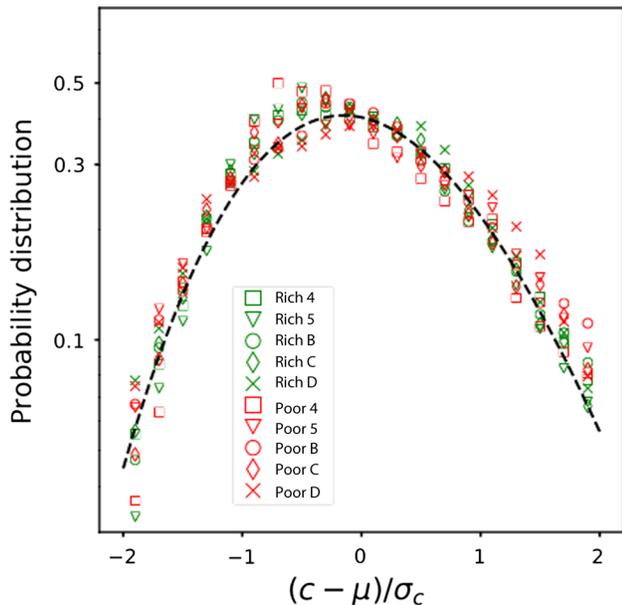


FIG. 3. Distributions of the normalized protein concentration fluctuations under different experimental conditions. The dashed line is obtained from the analytical expression Eq. (13) from our model.

Analogous to scaling laws in physics such as those in critical phenomena, while the power  $p = 2$  remains universal, the prefactor  $A$  in the square Taylor’s law for protein concentration fluctuations is not a universal constant, and it can vary from gene to gene and from condition to condition. However, since the variation in  $A$  is not correlated with  $\mu$  in the high expression regime, the square TL is preserved statistically despite the variation in  $A$  for individual cases.

From our single-cell measurements, we compute the probability distribution of the normalized concentration  $(c_i - \mu_i)/\sigma_{c,i}$  for cells under each experimental condition. As shown in Fig. 3, all ten distributions for the normalized concentration collapse onto the same master curve, which agrees with an analytical solution (dashed line in Fig. 3) obtained from our model study.

In the following, we develop a minimal mechanistic model to understand the origin of the square TL; to identify and quantify some of the dominant noise sources that contribute to the variations in the prefactor  $A$ ; and to explain the collapse of the normalized protein concentration distributions.

### B. A minimal mechanistic model to describe protein concentration fluctuations in growing and dividing cells

Cell growth and division are complex processes involving many regulatory proteins and pathways, which are beyond the scope of this paper. Here, we aim to develop a simple model that captures the most salient features in the underlying molecular mechanism governing protein

concentration fluctuations in growing and dividing cells in a mother machine setup.

In the minimal model, the growth and division processes in a cell are controlled by a *production* variable  $R$  and a *division* variable  $Z$ , respectively. All dynamic variables in a cell such as the cell length  $L$  and protein number  $N$ , as well as  $R$  and  $Z$  themselves, are controlled by  $R$  and  $Z$ .

The use of a common production variable  $R$  that is shared by other variables of the system is supported by the fact that growth rates of  $L$  and  $N$  during cell growth are not independent. As shown in Fig. 4(a), during each growth period between two consecutive divisions, we can fit the cell length and protein number dynamics to exponential functions and determine the growth rates  $\lambda_l$  and  $\lambda_n$  for  $L$  and  $N$ , respectively. In Fig. 4(b), we plot  $\lambda_n$  versus  $\lambda_l$  for all growth periods and for both nutrient conditions, which clearly shows that the two growth rates are highly correlated. To account for this strong correlation, all growth rates in our model are assumed to be proportional to the common production variable  $R$ , which can be interpreted as the number of active ribosomes in the cell. In previous experiments [34], it is shown that the growth rate depends linearly on the RNA/protein ratio. Since the total RNA content in a cell is a good measure of the ribosome number, these experiments support the assumption that growth rates are proportional to  $R$ .

In order to maintain cell size homeostasis, a feedback mechanism is needed to control cell division. Indeed, as pointed out in Refs. [4,26], if cell divisions were to occur at independent random time intervals, even with a fixed mean, the accumulation of variation in division times away from their mean would lead to divergence of cell length fluctuation. In bacterial cells, cell division is regulated by the division protein FtsZ [35–37], which assemble into a ring (the  $Z$  ring) localized at the future division site of the cell. The completion of the  $Z$  ring, together with other proteins, is critical for cell division. Motivated by these experimental facts and by following previous theoretical work [28–32], we assume that the probability for cell division increases sharply with  $Z$  when it crosses a certain threshold  $Z_0$ . Since the  $Z$  ring formation depends on oligomerization of FtsZ at a localized site, the threshold is set over the total number of proteins as opposed to its concentration.

Taken together, our model is represented schematically in Fig. 5(a) (top). There are four extensive variables: The cell size (length)  $L$  and the protein number  $N$  are “observable” variables that can be directly measured in the mother machine experiments; the number of growth complexes (ribosomes)  $R$  and the number of division proteins (FtsZ)  $Z$  are “hidden” variables that control the growth and division of all variables including themselves. Since we do not have detailed information on all the relevant biochemical processes that govern protein concentration fluctuations, instead of modeling specific biochemical reactions explicitly as in previous works [12,13,38], we use the following stochastic ordinary differential equations to study

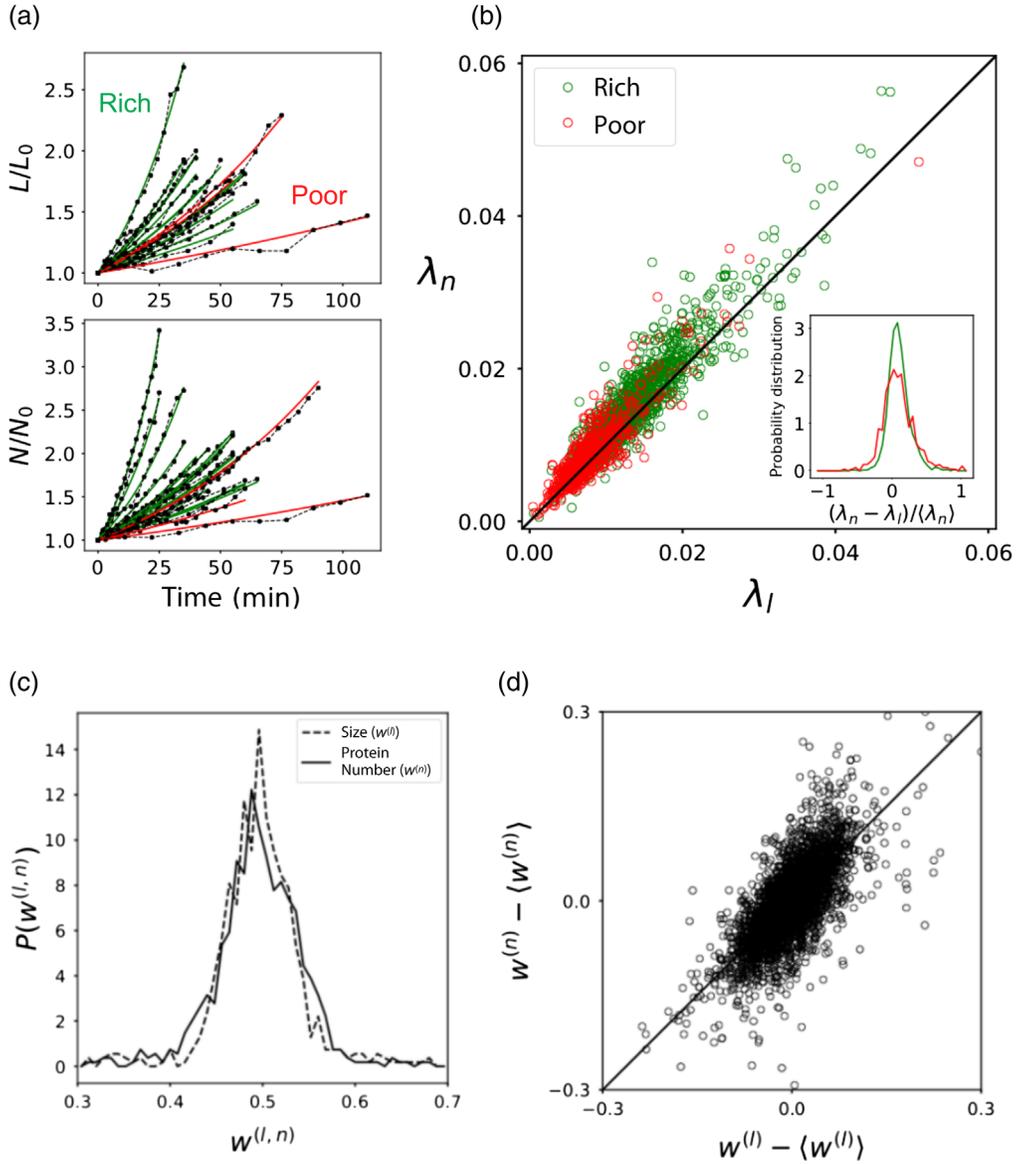


FIG. 4. Noise sources due to growth and division. (a) The effective growth rates  $\lambda_l$  and  $\lambda_n$  are obtained by fitting the cell size and protein number dynamics during each growth period (between two consecutive division events) with exponential functions. An example of the fitting is shown for  $L$  (top) and  $N$  (bottom); different colors correspond to different nutrient conditions. (b)  $\lambda_l$  versus  $\lambda_n$  shows high but not perfect correlation. Deviations from the diagonal correspond to a noise source for  $c$  due to growth. The inset shows the distribution of the relative difference between  $\lambda_l$  and  $\lambda_n$ . (c) Distribution of the partition factors  $w^{(l,n)}$  for cell size and protein number. (d) The correlation between  $w^{(l)}$  and  $w^{(n)}$  is high but not perfect. Deviations from the diagonal correspond to another noise source for  $c$  due to division.

the coarse-grained dynamics of the four key variables ( $L$ ,  $N$ ,  $R$ , and  $Z$ ):

$$\frac{dL}{dt} = RK_l(1 + \eta_l) - L \sum_{i=1}^{n_d(t)} w_i^{(l)} \delta[t - t_i(Z)], \quad (4)$$

$$\frac{dN}{dt} = RK_n(1 + \eta_n) - N \sum_{i=1}^{n_d(t)} w_i^{(n)} \delta[t - t_i(Z)], \quad (5)$$

$$\frac{dR}{dt} = RK_r(1 + \eta_r) - R \sum_{i=1}^{n_d(t)} w_i^{(r)} \delta[t - t_i(Z)], \quad (6)$$

$$\frac{dZ}{dt} = RK_z(1 + \eta_z) - Z \sum_{i=1}^{n_d(t)} w_i^{(z)} \delta[t - t_i(Z)], \quad (7)$$

which all have the same general form. The first and second terms on the right side of each equation represent effects of

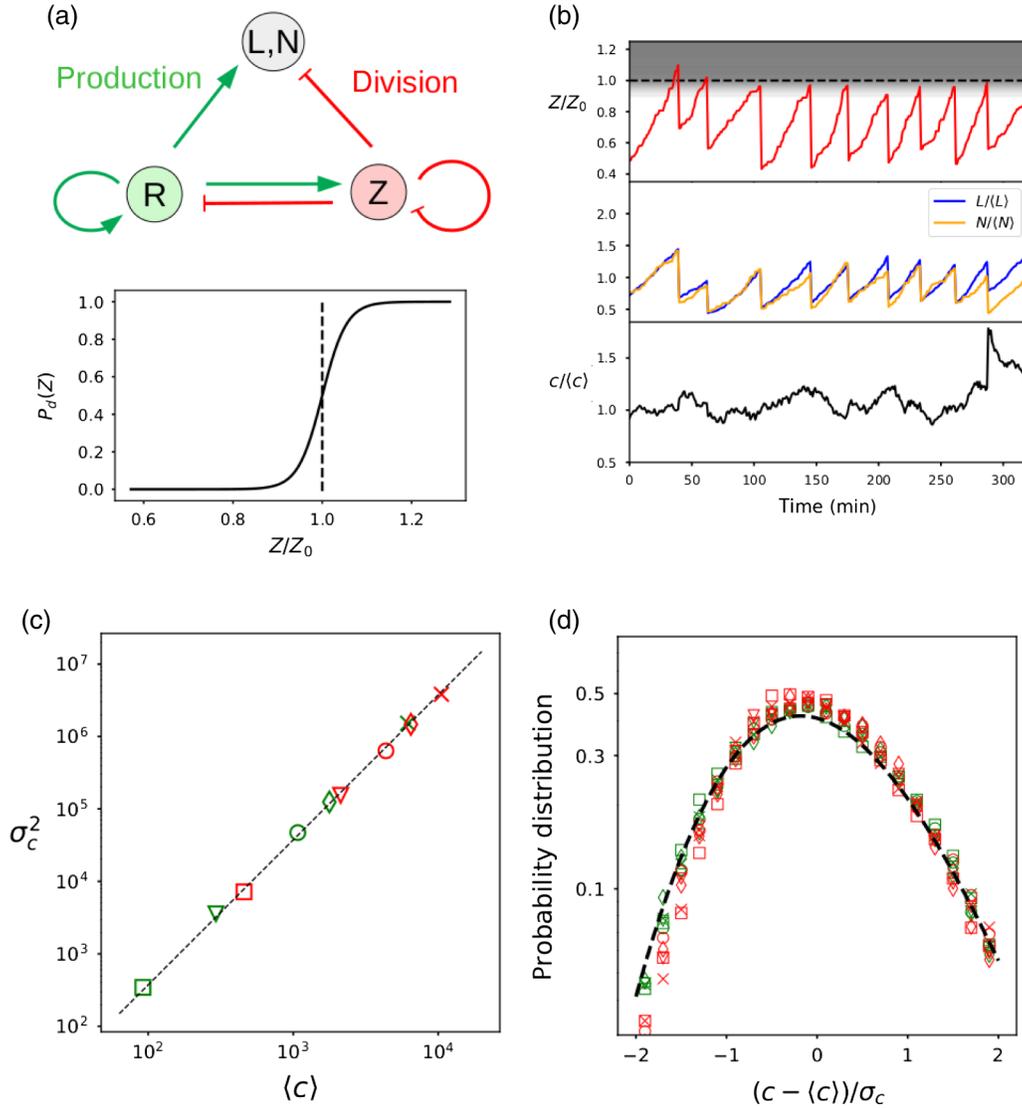


FIG. 5. The minimal mechanistic model and its behaviors. (a) Top: illustration of the four-node model. All production (growth) processes (green arrows) depend on  $R$ . The division process, which reduces all variables (red lines), is controlled by  $Z$ . Bottom: the probability rate of division  $P_d(Z)$  as a function of the division protein number  $Z$ . (b) Time traces of normalized division protein (red), protein number  $N$  (orange), cell size  $L$  (blue), and protein concentration  $c = N/L$  (black). In the top, we show in gray the region where the division probability is large (the dotted line is for  $Z = Z_0$ ). (c) Variance of the concentration versus its mean from our model for different promoters (different symbols) and under different nutrient conditions (green, rich medium; red, poor medium). The same symbols and colors are used as experimental data shown in Fig. 2(a) for easy comparison. (d) Distributions of the normalized fluctuations in the concentration obtained with the same parameters as in (c) collapse onto the same curve. The dashed line is from the analytical expression Eq. (13) obtained from our model. The parameters for  $P_d(Z)$  are  $\Delta Z/Z_0 = 2.7 \times 10^{-2}$  and  $\Delta t = 2$  min. See Appendix B for details of the simulations and parameters used.

growth and division, [39] respectively. The variables  $w_i^{(j)}$  ( $j = n, l, r, z$ ) are the partition factors, which is the fraction of variable  $j$  (e.g., protein number) that goes to the tracked daughter cell upon the  $i$ th division event, and they follow a Gaussian distribution around the mean  $\langle w_i^{(j)} \rangle = 1/2$ ;  $t_i(Z)$  is the time of the  $i$ th division that depends on  $Z$  (see below);  $K_j$  ( $j = n, l, r, z$ ) are the mean growth (production) rates;  $\eta_j$  ( $j = n, l, r, z$ ) are the relative noise in the rates; and  $n_d(t)$  is the number of divisions up to a time  $t$ :

$$n_d(t) = \int_0^t dt' \sum_i \delta[t' - t_i(Z)].$$
 Since we are interested in protein concentration dynamics in timescales comparable to or longer than the cell division time, the noise  $\eta_j$  is approximated as a white noise:  $\langle \eta_j(t) \eta_j(t') \rangle = \Delta_j \delta(t - t')$  with the noise strength given by  $\Delta_j$ . We use a truncated Gaussian distribution for  $\eta_j$  (truncated to maintain  $1 + \eta_j > 0$ ) in our study, but other forms of the noise distribution function such as a uniform distribution function can be used without affecting the general results.

We mainly focus on the high expression regime where the intrinsic noise, i.e., noise in the process that is used to control the mean expression level in the experiments, can be neglected. The extrinsic noise [40], which is defined as noise in the “external” processes that are not involved in controlling the mean expression level in the experiments, is captured by the multiplicative factor  $(1 + \eta_n)$ , where  $\eta_n$  characterizes fluctuations of the external factor(s) relative to its mean value (set to be 1 for simplicity). For example, when a gene is expressed by using plasmids as in our experiments shown in Fig. 1, fluctuations in the plasmid copy number can be a source of extrinsic noise, and they are uncorrelated with the mean expression level. More specifically, we can write the number of plasmids in a cell as  $n_p = \langle n_p \rangle + \delta n_p$  with mean  $\langle n_p \rangle$  and fluctuation  $\delta n_p$ ; our model remains the same with the mean production rate  $K_n = \tilde{K}_n \langle n_p \rangle$ , where  $\tilde{K}_n$  is the mean production rate per plasmid; and the noise term  $\eta_n = \eta_n^{(0)} + \eta_p$ , which includes the noise  $\eta_p = \delta n_p / \langle n_p \rangle$  from the plasmid number fluctuations and the noise  $\eta_n^{(0)}$  from all other extrinsic noise sources. In our study, the overall noise strength  $\Delta_n$  is obtained by fitting our stochastic model to experimental data without separately considering contributions from different noise sources. Since  $\eta_n$  describes fluctuations in the external processes, its strength  $\Delta_n$  as well as all other parameters in our model is assumed to be independent of  $K_n$  (see Sec. III for possible implications when this assumption is not valid).

The intrinsic noise can also be considered in our model by including an additive noise term  $\delta K_n$  in  $K_n$ . In the low expression regime, the intrinsic noise becomes dominant, and it changes the scaling behavior [13], which is captured by the modified model as shown later in this section.

Following previous work [32,41,42], we introduce a feedback control for cell division based on the division protein number  $Z$ . In particular, we implement a soft-thresholding process with a simple logistic function [Fig. 5(a) (bottom)]:

$$P_d(Z) = \frac{\Delta t^{-1}}{1 + \exp[(Z_0 - Z)/\Delta Z]}, \quad (8)$$

which is the probability rate of cell division for a given  $Z$ .  $P_d(Z)$  is characterized by three parameters, each with clear biological meaning. The threshold value  $Z_0$  is the value of  $Z$  at which the division probability increases sharply, with the sharpness determined by  $\Delta Z \ll Z_0$  ( $P_d$  is simply a step function when  $\Delta Z = 0$ ). Once cell division starts, it can take a finite time to complete. This small but finite timescale is given by  $\Delta t$  in Eq. (8). Other forms of  $P_d(Z)$  with the same general properties considered here are used without affecting the general results. The choice of parameters used in our simulations is discussed in Appendix B.

There are two sources of noise due to fluctuations in growth rates and in partition factors, respectively. The noise

in growth rates ( $\eta_{n,l}$ ) for  $N$  and  $L$  can be estimated from measurements and analysis shown in Figs. 4(a) and 4(b). Similarly, the noise in partition factors ( $w^{(n,l)}$ ) can be determined from the measured distributions for  $w^{(n,l)}$  as shown in Fig. 4(c). Though there is a strong correlation between these two partition factors [Fig. 4(d)], their difference remains significant, and it gives rise to another source of noise for  $c$ . Dynamics of the two hidden variables  $R$  and  $Z$  are not directly measured. However, their dynamics and effects of their noise can be inferred by fitting our stochastic model to the measured dynamics of  $N$  and  $L$ .

We study our model by solving the stochastic equations [Eqs. (4)–(7)] numerically with physiologically reasonable parameters, some of which are estimated from the mother machine experiments. In Fig. 5(b), we show the time traces of different variables as well as the dynamics of the protein concentration. Cell division controlled by the variable  $Z$  can be seen from the top in Fig. 5(b), which shows that  $Z$  follows the same growth and division dynamics as other variables and division occurs with a high probability as  $Z$  (red) crosses the threshold  $Z_0$  (dashed line). As shown in the middle in Fig. 5(b), both  $L$  (blue) and  $N$  (orange) follow the same general growth and division pattern as that of  $Z$ , but their dynamics are not identical. As a result, the protein concentration  $c$  (black) shows smaller but finite and continuous fluctuations that are different from those in  $N$  or  $L$ . The behaviors of  $N$ ,  $L$ , and  $c$  from our model closely resemble those from experiments shown in Fig. 1.

We also study cell size homeostasis during growth [8,43], i.e., the dependence of elongation  $\Delta L$  and division time  $\Delta T$  on initial cell size  $L_0$  during each cell cycle under different growth conditions. The model results are in quantitative agreement with experiments, which validates the model for describing cell growth and division (see Appendix C for details).

To quantitatively compare the results from the model with the experimental results, we tune the rates ( $K_{n,l,r,z}$ ) and the noise strength ( $\Delta_{n,l,r,z}$ ), in accordance with experimental data for different promoters and in different nutrient conditions. Given that the promoter strength primarily influences the expression rate of the corresponding proteins, we assume that the change in promoter is captured in the model by a change in the value of  $K_n$ , so that the rate of increase in  $N$  would be directly affected. However, all the other parameters ( $K_{l,r,z}$  and  $\Delta_{n,l,r,z}$ ) are kept the same for all experimental conditions, which reduces the number of parameters in our model.

The mechanism by which the nutrient condition influences the kinetic rates of growth and expression is beyond the scope of this work. Here, we treat the nutrient dependence within our model phenomenologically based on experiments. In particular, from the experiments, the average division time  $\langle \Delta T \rangle$  is longer in the poor nutrient condition, whereas the average length size  $\langle L \rangle$  is smaller. We find that the simplest way to account for this observed

difference in our model is by assuming that the production rate for the division variable  $Z$ ,  $K_z$ , is roughly independent of nutrient conditions, while the other two rates for growth-related variables ( $R$  and  $L$ ) are scaled by a common nutrient-dependent constant  $\xi$ , i.e.,  $K_i \rightarrow \xi K_i$  with  $i = r, l$ ,  $\xi = 1$  for rich medium and  $\xi = 0.6$  for poor medium (see Appendix C for details). Operationally, we first tune  $K_l$  and  $K_r$  to match the observed cell growth statistics in different nutrient conditions before we tune  $K_n$  to properly fit the value of  $\langle c \rangle$ . See Appendix B for details of the simulations and parameter choice.

As shown in Fig. 5(c), where the variance of  $c$  is plotted as a function of its mean, the simulation results from our model obey the same square Taylor's law as in the experiments with the coefficient of proportionality  $A = 0.034 \pm 0.006$  that is in quantitative agreement with the experiments (the error bar here is small, because we do not consider cell-cell variability in our study). Furthermore, the normalized distribution function for  $c$  shows the same collapse for all values of  $K_n$  as shown in Fig. 5(d).

### C. The Langevin equation for stochastic protein concentration dynamics

In the previous section, we study protein concentration fluctuations by simulating the full model [Eqs. (5)–(7)] numerically. Here, by using a mean-field approximation, we derive the Langevin equation for protein concentration fluctuations, which leads to a closed form relation between the variance and the mean of  $c$  and an analytical expression for the probability distribution of  $c$ .

We start by considering the equation for  $c$ , obtained from Eqs. (4) and (5) (see Appendix D for details):

$$\frac{dc}{dt} = \hat{K}_n - \hat{K}_l c + c \sum_i^{n_d(t)} \{(w_i^{(l)} - w_i^{(n)})\delta[t - t_i(Z)]\}, \quad (9)$$

where we define the effective rates  $\hat{K}_l = rK_l(1 + \eta_l)$  and  $\hat{K}_n = rK_n(1 + \eta_n)$ , with  $r = R/L$ . The difference  $\delta w_i = w_i^{(l)} - w_i^{(n)}$  has zero mean  $\langle \delta w_i \rangle = 0$  and a delta-function correlation  $\langle \delta w_i \delta w_j \rangle = 2\Delta_w \delta_{ij}$ , as  $\delta w_i$  for different division events can be considered independent. For a timescale longer than the division time, we can define a mean-field cell division noise  $\eta_d$  whose correlation has the form  $\langle \eta_d(t_1)\eta_d(t_2) \rangle = 2\Delta_d \delta(t_1 - t_2)$  with  $\Delta_d = \Delta_w / \langle \Delta T \rangle$  the average (mean-field) cell division noise strength (see Appendix H for a derivation of the division time distribution and the mean-field noise).

The variables  $\hat{K}_l$  and  $\hat{K}_n$  in Eq. (9) can be written as  $\hat{K}_{l,n} = \langle \hat{K}_{l,n} \rangle (1 + \chi_{l,n})$ , where  $\chi_{l,n} = [(\delta r) / \langle r \rangle] + \eta_{n,l}$  is the fractional noise with  $\delta r (= r - \langle r \rangle)$  and  $\langle r \rangle$  the fluctuation and the mean of  $r$ , respectively. Thus, by taking the mean-field approximation in the long time limit, the Langevin equation [Eq. (9)] can be rewritten as

$$\frac{dc}{dt} = [\langle \hat{K}_l \rangle (\mu - c) + \mu \eta_a + c \eta_m], \quad (10)$$

where  $\mu = \langle \hat{K}_n \rangle / \langle \hat{K}_l \rangle$  is the average of  $c$ :  $\langle c \rangle = \mu$ , which is varied experimentally by changing the promoter strength or the nutrient condition, whereas  $\eta_a = \langle \hat{K}_l \rangle \chi_n$  and  $\eta_m = \eta_d - \langle \hat{K}_l \rangle \chi_l$  are the two noise terms for  $c$ .

It is interesting to note that both noise terms ( $\eta_a$  and  $\eta_m$ ) in Eq. (10) are multiplied by either the mean concentration ( $\mu$ ) or the instantaneous concentration ( $c$ ) itself. As a result of the multiplicative nature of the noise terms, Eq. (10) is invariant if  $c$  and  $\mu$  are scaled by an arbitrary constant factor. [44] It is easy to see that the scale invariance is absent if the noise is additive with a constant strength. This scale invariance in the multiplicative noise Langevin equation [Eq. (10)] immediately suggests that the distribution of  $c/\mu$  is independent of  $\mu$ . As a consequence, the variance of  $c$  is proportional to the square of its mean, which is just the square TL. Indeed, by solving the Fokker-Planck equation corresponding to Eq. (10), we derive an exact relationship between the variance and the mean (see Appendix E for details):

$$\sigma_c^2 = A \langle c \rangle^2, \quad (11)$$

and the prefactor  $A$  can be determined analytically:  $A = (\Delta_g + \Delta_d) / (\langle \hat{K}_l \rangle - \Delta_m)$ , where  $\Delta_g = \langle \hat{K}_l \rangle^2 (\Delta_n + \Delta_l)$  is the strength of the growth-dependent noise due to fluctuations in growth and production rates for cell size ( $\Delta_l$ ) and protein number ( $\Delta_n$ );  $\Delta_m = \Delta_d + \langle \hat{K}_l \rangle^2 (\Delta_r + \Delta_l)$  is the strength of the noise  $\eta_m$  with  $\Delta_r$  the noise strength for  $\delta r / \langle r \rangle$ . The typical timescale for the  $c$  dynamics is given by  $\tau_c \equiv \langle \hat{K}_l \rangle^{-1}$ , which is comparable to the average cell division time  $\tau_c \sim \langle \Delta T \rangle$ . The strength of the noise averaged over the typical timescale  $\tau_c$  is given by  $\Delta_{n(l)}/\tau_c$ , which is found to be small:  $\Delta_{n(l)}/\tau_c \ll 1$ . Taken together with the fact that  $\Delta_w \ll 1$ , we have  $\Delta_m \ll \langle \hat{K}_l \rangle$  and the prefactor  $A$  can be expressed as

$$A \approx \langle \hat{K}_l \rangle^{-1} (\Delta_g + \Delta_d), \quad (12)$$

where the two terms on the right-hand side of the equation above represent contributions to  $A$  from the noise sources in the growth and division processes, respectively.

Finally, by solving the steady-state Fokker-Planck equation exactly, we obtain an analytical expression for the probability distribution of the protein concentration:

$$P(c) = \frac{1}{\mathcal{Z}} \left( \frac{\mu^2 \Delta_a + 2\mu \Delta_{am} c + \Delta_m c^2}{\langle \hat{K}_l \rangle} \right)^{-1 - (\langle \hat{K}_l \rangle / 2\Delta_m)} \times \exp \left[ \frac{2\langle \hat{K}_l \rangle (\Delta_{am} + \Delta_m) \tan^{-1} \left[ \frac{\mu \Delta_{am} + \Delta_m c}{\mu \rho_\Delta} \right]}{2\Delta_m \rho_\Delta} \right], \quad (13)$$

and  $P(c < 0) = 0$ , where  $\mathcal{Z}$  is the normalization constant,  $\Delta_a = \langle \hat{K}_l \rangle^2 (\Delta_{(r)} + \Delta_n)$  is the noise strength for  $\eta_a$ ,  $\Delta_{am} = -\langle \hat{K}_l \rangle^2 \Delta_{(r)}$  is the correlation between  $\eta_a$  and  $\eta_m$ , and  $\rho_\Delta = \sqrt{\Delta_a \Delta_m - \Delta_{am}^2}$ . There is a negligibly small value for  $P(c)$  at  $c = 0$ , which is caused by assuming  $\eta_a$  to be an unbounded Gaussian noise. For  $c \gg \mu$ ,  $P(c)$  decays as a power law  $\sim c^{-2[1+(\langle \hat{K}_l \rangle/2\Delta_m)]}$ , which is different from a log-normal distribution but similar to an inverse Gamma distribution, which is the solution for  $P(c)$  in the limit of negligible additive noise ( $\Delta_a = \Delta_{am} = \rho_\Delta = 0$ ). In Figs. 2(b) and 5(d), we plot the analytical distribution function given in Eq. (13) (black dashed line), which quantitatively agrees with experimental results and simulation results from the full stochastic model.

#### D. Microscopic origins of protein concentration fluctuations

Our model not only demonstrates the square TL in protein concentration fluctuations in single cells, it also reveals the microscopic origins of the concentration fluctuations. In particular, Eq. (12) shows that the prefactor  $A$  depends on parameters such as the extrinsic noise strengths  $\Delta_n$  and  $\Delta_l$ , the noise strength of division  $\Delta_d$ , and the effective average cell growth rate  $\langle \hat{K}_l \rangle$ , all of which can be different in different cells and/or under different experimental conditions. For example, the use of plasmids as opposed to chromosomal integration of the promoters leads to a different noise strength in the protein expression rate (see also Refs. [45,46]) and, consequently, a different distribution of the prefactor  $A$ , which is consistent with experimental data shown in Fig. 10 in Appendix G. Moreover, external regulators that are

involved in the transcription process can also affect the noise in the production rate, which explains the broader distribution of  $A$  with a larger mean  $\langle A \rangle$  observed in the genome-wide study [13]. Note that this additional source of noise is reduced in our study by the use of constitutive promoters, which are not directly regulated by specific transcriptional regulators. However, from Eq. (12), we see that  $A$  does not depend on the mean protein expression rate  $K_n$ , which is consistent with the absence of correlation between  $A$  and the mean expression level  $\mu$  observed in all existing experiments. This independence of  $A$  on  $\mu$  is critical for the existence of the square TL in the high expression regime.

It is shown explicitly in our model [Eq. (12)] that there are two contributions to  $A$ :  $A_g = \langle \hat{K}_l \rangle^{-1} \Delta_g = \langle \hat{K}_l \rangle (\Delta_n + \Delta_l)$  and  $A_d = \langle \hat{K}_l \rangle^{-1} \Delta_d = \Delta_w / (\langle \hat{K}_l \rangle \langle \Delta T \rangle)$  represent contributions from the growth and production noise and the cell-division partition noise, respectively. Quantitatively,  $A_g$  and  $A_d$  can be determined by the different noise strengths ( $\Delta_n$ ,  $\Delta_l$ , and  $\Delta_w$ ) and timescales ( $\langle \Delta T \rangle$  and  $\langle \hat{K}_l \rangle^{-1}$ ), which can be estimated by fitting our model to the experimental data (see Appendixes D and F for details). In Fig. 6(a), the values of  $A$  obtained from our model  $A_{\text{model}} = A_g + A_d$  are shown to be highly consistent with those determined directly from concentration fluctuations  $A_{\text{exp}} = \sigma_c^2 / \langle c \rangle^2$  for individual cells for different promoters and different growth conditions. The distribution of the log-ratio  $[\ln(A_g/A_d)]$  for all individual cells is shown in Fig. 6(b), which reveals that the contribution from growth-related noise is significantly stronger than that from cell-division partition noise.

The expression for the coefficient  $A$  shown in Eq. (12) also highlights an important difference between statistics of

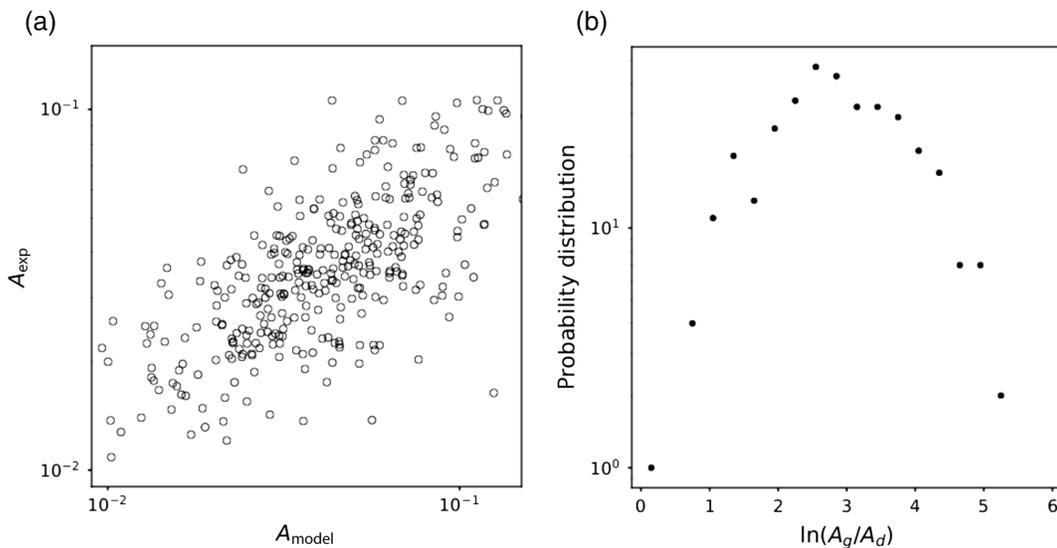


FIG. 6. Fluctuation strength (prefactor in TL)  $A$  from experiments and model. (a) Comparison of  $A$  obtained from direct measurements ( $A_{\text{exp}}$ ) and from model ( $A_{\text{model}}$ ) for individual cells (circles) across different promoters and different growth conditions. (b) Distribution of the log-ratio  $[\ln(A_g/A_d)]$  for all individual cells.  $A_g$  and  $A_d$ , which represent contributions to  $A$  from growth-related processes and cell division, are obtained from our model. It is clear that  $A_g$  dominates over  $A_d$  with  $\ln(A_g/A_d) = 2.8 \pm 1.0$ .

the protein concentration and the protein number  $N$ . Even though they both satisfy the square Taylor's law, the coefficient  $A$  for  $c$  depends on noise in the cell size growth rate, whereas the coefficient  $A_n$  for  $N$  does not. To verify this difference, we run simulations for different noise strengths  $\Delta_l$  and compare  $A$  with  $A_n$ , obtained by computing the variance and the mean value of  $N$ . As we show in Fig. 12 in Appendix I, while  $A$  increases with  $\Delta_l$ ,  $A_n$  is independent of  $\Delta_l$ . Another significant difference is that  $A_n \gg A$ , as the variation of  $N$  is dominated by the large change (approximately twofold) of  $N$  during each growth and division cycle even in the absence of any noise effect.

### E. The low expression regime

For low copy number genes (below ten proteins per cycle), it is shown experimentally [13] that the prefactor  $A$  is proportional to the inverse of the mean value:  $A \sim 1/\langle c \rangle$ , which is understood by considering a model of key biochemical reactions in the transcriptional process and the underlying intrinsic noise [13]. Here, we ask whether the protein concentration fluctuations in the low expression regime can also be studied in our coarse-grained model.

To include the intrinsic noise in the transcriptional process, we change the parameter  $K_n$  in our model to a time-dependent random variable  $K_n = \bar{K}_n + \delta K_n(t)$  with mean  $\bar{K}_n$  and a noise term  $\delta K_n$  with variance  $\langle \delta K_n^2 \rangle = \sigma_k^2$ . As the chemical reactions underlying  $K_n$  are Poisson processes, we assume the variance is proportional to the

mean:  $\sigma_k^2 = \Delta_k \bar{K}_n$  with a constant  $\Delta_k$ , which characterizes the strength of the intrinsic noise. We run simulations of the modified model for different values of  $\bar{K}_n$  and study how the protein concentration variance  $\sigma_c^2$  scales with its mean  $\langle c \rangle$  or, equivalently, how the prefactor  $A \equiv \sigma_c^2 / \langle c \rangle^2$  varies with  $\langle c \rangle$ . As shown in Fig. 7, for smaller values of  $\langle c \rangle$  when  $\bar{K}_n < \Delta_k / \Delta_n$ ,  $A$  is a decreasing function of  $\langle c \rangle$ :  $A \sim \langle c \rangle^{-1}$ . For larger values of  $\langle c \rangle$ ,  $A$  becomes a constant, which indicates the existence of the square Taylor's law consistent with the results from the previous sections when the intrinsic noise is neglected. The transition from the low expression regime behavior ( $\sigma_c^2 \propto \langle c \rangle$ ) to the square Taylor's law ( $\sigma_c^2 \propto \langle c \rangle^2$ ) occurs at a crossover concentration  $\langle c \rangle^* \propto \Delta_k^{-1}$ , which depends inversely on the intrinsic noise strength  $\Delta_k$ .

Our results suggest that the modified model that includes the intrinsic noise in  $K_n$  can be used to study single-cell protein concentration dynamics and the variance-mean relation in the whole range of gene expression.

## III. DISCUSSION

In this paper, we measure protein concentration dynamics in single cells for many generations using the mother machine microfluidic device. By using a set of constitutive promoters with strength varying over 2 orders of magnitude, we confirm the square TL. In addition, we find that the prefactor in the square TL varies from cell to cell independent of the mean expression level. Guided by these observations, we develop a minimal mechanistic model with four biologically meaningful variables ( $R$ ,  $Z$ ,  $N$ , and  $L$ ) to study stochastic dynamics of protein concentration in growing and dividing cells. A Langevin equation to describe the stochastic dynamics of the concentration  $c$  is derived from the minimal model. The balance between growth and division processes is key to maintain a dynamic equilibrium for cells. By considering these two stochastic processes consistently in the minimal model and the Langevin equation, we not only prove the square TL, but also obtain an analytical expression for the distribution function of the normalized concentration, which is in quantitative agreement with our mother machine experiments with different promoters and under different nutrient conditions. In addition, our study shows how the prefactor  $A$  in the square TL depends on the strength of the two main noise sources (growth and division), which explains the observed cell-to-cell variability in  $A$ . We also use our model to analyze previous experiments [25] under different experimental conditions and obtain consistent results, although the data there cover a much smaller range (approximately 1/2 decade) of mean protein concentrations (see Fig. 11 in Appendix G for details).

There are two central regulatory variables in our model,  $R$  and  $Z$ , which control growth and division, respectively. Both  $R$  and  $Z$  can be considered as large complexes with

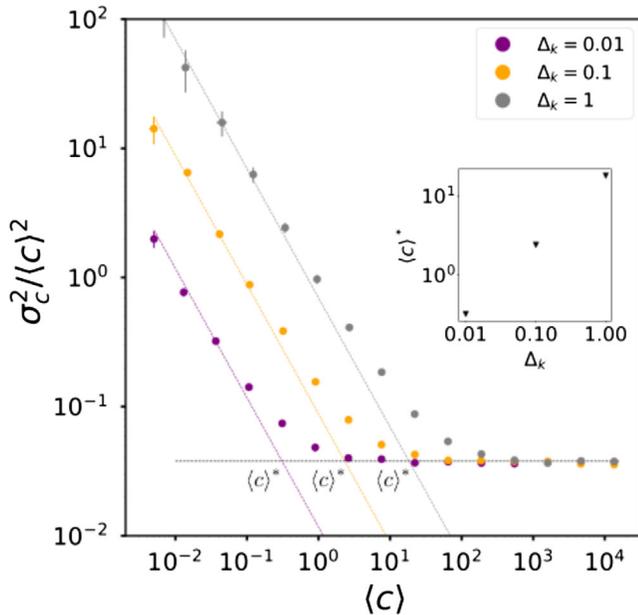


FIG. 7.  $A$  as a function of  $\langle c \rangle$  when  $K_n$  is generated from a Gamma distribution with mean value  $\bar{K}_n$  and variance  $\sigma_k^2 = \Delta_k \bar{K}_n$ , for three values of  $\Delta_k$ . The mean value  $\langle c \rangle$  is changed by changing  $\bar{K}_n$ . The error bars are obtained by running the simulations for five realizations and calculating the resulting standard deviation.

multiple components; however, each of them has identifiable key components:  $R$  is associated with ribosome, whereas  $Z$  is associated with FtsZ. In our minimal model, we implement only the simplest possible interactions between  $R$  and  $Z$ —the production rates of all proteins (including  $R$  and  $Z$  themselves) are proportional to  $R$ ; and the probability of cell division increases sharply when  $Z$  crosses a certain threshold  $Z_0$ . Despite its simplicity, the minimal model is able to reproduce the observed statistics of protein concentration fluctuations and cell size homeostasis. More importantly, the minimal model provides a mathematical framework to ask further questions regarding molecular origins of the growth and division processes. For the division control, one key question is about the molecular origin for the cell division regulation, which is modeled here by the assumption (hypothesis) that the probability of division increases sharply when  $Z$  crosses a threshold  $Z_0$ . For the growth process, we do not take into account the way resources are allocated to the production of different protein classes in our model where all production rates are linearly proportional to  $R$  [47]. The effect of proteome allocation can be studied in our model by allowing different substrates (mRNAs) to compete for the same finite pool of  $R$  with different affinities, which results in a nonlinear dependence of the production rates on  $R$ . It would be interesting to study proteome allocation [41] in dividing cells by introducing nonlinear growth rates in our model.

Though we focus on explaining TL with exponent  $p = 2$  for protein concentration fluctuations in the high expression regime, our model also indicates how the exponent  $p$  may deviate from 2. In particular, from the analytical expression for the variance-average dependence [Eq. (11)], it is easy to see that if the prefactor  $A$  is correlated with the mean  $\mu (\equiv \langle c \rangle)$ , then the variance-mean relation would deviate from the square TL; e.g., if  $A$  depends on  $\mu$  in a power law form,  $A \sim \mu^\alpha$ , then the variance-mean relation follows the TL with an exponent  $p = 2 + \alpha$ . In the low expression regime, the intrinsic noise can be considered as a special case with  $\alpha = -1$ . In the high expression regime, from the expression for the prefactor  $A$  [Eq. (12)], such a correlation may exist if there is a correlation between the noise strength ( $\Delta_{g,d}$ ) and  $\mu$ . In our experiments, the mean concentration is varied by changing the promoter strength characterized by  $K_n$  in our model. Since  $K_n$  does not appear in the expression for  $A$ , there is no correlation between  $A$  and the mean  $\mu$ , which leads to the observed square TL. However,  $\mu$  also depends on the elongation rate  $K_l$ , which affects  $A$ . Therefore, if the mean concentration  $\mu$  can be varied by changing the elongation rate  $K_l$ , our model would predict a deviation from the square TL, which may be tested experimentally. More generally, changes in the coefficient of variation in the protein content are discussed in several previous works [40,45,46,48]. It would be interesting to do an extensive analysis of the biological conditions that determine the noise parameters that appear in Eq. (12)

and infer in this way the classes of systems in which the Taylor’s law is satisfied with the same coefficient.

Taylor’s law is a ubiquitous scaling law observed in a plethora of different systems, from the occurrence of measles cases [17] to the share price fluctuations in stock market [49]. However, these observations of Taylor’s law remain largely empirical with little understanding of their mechanistic origins. Here, we show the existence of Taylor’s law (with exponent  $p = 2$ ) for a set of constitutive promoters with different strengths and operating in two different conditions. Given the generic nature of these promoters, there is no *a priori* reason for the exponent to be dependent on specific properties of these promoters in the high expression regime where the intrinsic noise can be neglected. Indeed, as we show in this paper, the square Taylor’s law exists generally in systems that can be described by a Langevin equation with a relaxational deterministic linear term and a multiplicative noise; i.e., the noise term is multiplied by the variable of interest. Even though we focus on studying protein concentration fluctuations in bacterial cells such as *E. coli*, the general theoretical framework used in this paper may be applicable to other systems where the dominant noise source is multiplicative. Different TL exponents observed in different systems may be caused by possible correlations between key parameters in the system (e.g., the relaxation timescale and the noise strength) and the mean.

## ACKNOWLEDGMENTS

We thank Dr. Philippe Cluzel for many insightful discussions and members of the Cluzel lab where the original dataset reported in this manuscript was collected. We thank Dr. Hanna Salman for providing data from Ref. [25] and Dr. Joel Cohen for discussions regarding Taylor’s law and comments on our manuscript. The work by A. S. S. and Y. T. are supported by a NIH grant (R35GM131734 to Y. T.). The work by M. G. in the Cluzel lab are supported by a NIH grant (R01GM134275 to Philippe Cluzel) and a NSF grant (1615487 to Philippe Cluzel). The work by M. A. is partially supported by PAPIIT-UNAM Grant IN-111322.

## APPENDIX A: EXPERIMENTAL METHODS

A vector with the artificial Pro5 promoter [19] upstream of the gene of the fluorescent protein Venus followed by the *rrnB* T1 terminator is inserted into a pSC101 plasmid with the kanamycin antibiotic resistance gene. It is assembled using the isothermal assembly protocol to construct pPro5Venus plasmid. It is transformed into DH5 $\alpha$  strain (New England Biolabs) for selection and amplification. The plasmid sequence is verified by sequencing and named pPro5Venus.

The rest of the plasmids are constructed out of pPro5Venus. The complete sequence of the plasmid is amplified with primers that help swap the 10 box with the respective sequence from other promoters from the set

published in Ref. [19]. Each plasmid is circularized using the NEB KLD enzyme mix (New England Biolabs).

The background strain for this work is MGR-E98K [22], which is *E. coli* MG1655 (CGSC No. 6300) with a point mutation in MotA that disables rotation and prevents cells from swimming out of the cell channels in the microfluidic device. Each plasmid is transformed into the strain by electroporation, and the final set of strains is named ProVenus set. Selected colonies have kanamycin resistance and are fluorescent under the microscope. The content is verified by colony polymerase chain reaction. Using flow cytometry, it is confirmed that the sets of strains have different fluorescent levels.

Two growth media are used to grow the cells in the microfluidic device. They are selected considering that wild-type *E. coli*'s growth rate is significantly higher in one medium compared to the other on a batch experiment. *Rich media*.—MOPS EZ rich defined medium from Teknova, supplemented with 0.4% glycerol. *Poor media*.—M91X, 2 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>, 0.5% casamino acids, 0.4% glycerol, and 1 ug/ml thiamine. In both media, Pluronic F-108 (Sigma-Aldrich) is added to final concentration of 0.85 g/l, to act as a surfactant in the microfluidic device.

Each ProVenus strain is grown overnight in 1 ml of the same growth media that is used in the microfluidic experiment. In each microfluidic experiment, there are four different stains in the same device to ensure they grow in the same conditions, avoiding day-to-day variations in the setup. In all experiments, the wild-type strain is included, and the remaining strains are chosen so that they have distinguishable fluorescence levels. The overnight cultures are centrifuged at 6000 g for 1 min, and from each of them the same volume of the pellet is taken and mixed by gently pipetting.

The microfluidic device used in this work is an adaptation of the mother machine described in Ref. [23]. The design of the device and the protocol for the mold construction can be found in Ref. [22]. For every experiment, a microfluidic device is cast from the same mold. The cells are loaded into the device by pipetting into it the high-density mixture of strains. The loaded microfluidic device is connected to a pump that delivers growth media with a constant flow of 5 ul/min. It is placed under the inverted microscope in an incubated box at 30 °C during all the experiment.

Cells are tracked using imaging microscopy by taking phase contrast and fluorescent images with the YFP channel every 5 min for 24 h. The microscope setup is controlled using custom software on MATLAB 2013a interfacing with  $\mu$  Manager 1.4. Multiple positions of the device are captured.

We use the software Bacman for cell segmentation and tracking [24].

## APPENDIX B: DETAILS OF THE NUMERICAL SIMULATIONS

We solve our stochastic model [Eqs. (4)–(7) in the main text] numerically. Between two consecutive divisions, the

integration is performed with the Euler-Maruyama method, which is a standard method to find the numerical solution of stochastic differential equations. At each time step  $\delta t$ , a division could take place with a probability  $P_d[Z(t)]\delta t$ , where  $Z(t)$  is the value of the division protein number at the time  $t$ . In case of division, every variable is multiplied by a factor  $f$ , where  $f = 1/2(1 + \epsilon_d)$ , where  $\epsilon_d$  is a Gaussian variable with variance  $\sigma_d^2$ . We denote  $\tau$  ( $= 45$  min) as the average division time in rich nutrient conditions, and a small time step is chosen to be  $\delta t = 10^{-3}\tau$ . The parameters chosen for  $P_d(Z)$  are  $\Delta Z/Z_0 = 2.7 \times 10^{-2}$  and  $\Delta t = 2$  min. The rate of expression of the division protein is  $K_z = 1 \text{ min}^{-1}$ . The noise strengths for  $\eta_n$ ,  $\eta_z$ ,  $\eta_l$ , and  $\eta_r$  are  $\Delta_n = 1.2$  min,  $\Delta_z = 1$  min,  $\Delta_l = 0.2$  min, and  $\Delta_r = 1$ , respectively.

The standard deviation for the partition error is  $\sigma_d = 0.1$ , inferred from experiments. We tune the parameter  $K_l$  in such a way that the average size at division is the same as in the experiments and the parameter  $K_n$  for the concentration to match the value of the fluorescence traces for every nutrient condition (see Table I for the numerical value of  $K_l$  and  $K_n$ ). In most cases, the simulation is run for a total time  $T_t = 5 \times 10^3$  min, but in the case to obtain the probability distribution we use  $T_t = 5 \times 10^4$  min to sample a larger statistics.

We do not have direct measurements of the ribosome number. However, there are some choices of the parameters that are constrained by the experimental results. Indeed, if we consider the equation for  $R$ , for small noise, the solution between two cell divisions is given by

$$R(t) = R_0 \exp[K_r t]. \quad (\text{B1})$$

Since we want to have  $\langle R(t_f) \rangle = 2 \langle R(t_{in}) \rangle$ , where  $t_f$  and  $t_{in}$  are the final and initial time of a given cell cycle, respectively, in order to have a stationary state, the average division time must be

$$\tau_d = \langle t_f - t_{in} \rangle = \ln 2 / K_r, \quad (\text{B2})$$

independently of any other parameter. Therefore, if we set  $K_r = \ln 2 / \tau_d$ , where  $\tau_d = \langle \Delta T \rangle$  is taken from the experiments, the simulation leads to an average division time that is coherent with the results of the experiments performed

TABLE I. Rates  $K_n$  and  $K_l$  used in the simulations for the two nutrient conditions. The units of measurement are such that  $RK_n$  is in units of fluorescence per minute, while  $RK_l$  is in microns per minute, where  $R$  is a variable indicating the number of ribosomes.

	Promoters	Rich	Poor
$K_n$	Pro4	0.014	0.04
	Pro5	0.04	0.2
	ProB	0.16	0.4
	ProC	0.26	0.61
	ProD	0.92	0.97
$K_l$		$1.5 \times 10^{-4}$	$9 \times 10^{-5}$

with the mother machine technique. To study the deviation from the TL for low expression, we generate the value of  $K_n$  from a Gamma distribution with shape parameter  $k = \hat{K}_n/\Delta_k$  and scale parameter  $\theta = \Delta_k$ , where  $\hat{K}_n$  and  $\Delta_k$  are two constants, so that the mean value would be  $k\theta = \hat{K}_n$  and the variance  $k\theta^2 = \hat{K}_n\Delta_k$ .

### APPENDIX C: THE MINIMAL GROWTH-DIVISION MODEL EXPLAINS OBSERVED CELL DIVISION STATISTICS

There have been extensive experimental studies and analysis on the dependence of the elongation  $\Delta L$  and the division time  $\Delta T$  on the initial cell size  $L_0$  during each cell cycle under different growth conditions. Here, we first briefly describe correlations among these quantities from our mother machine experiments, and then we check if our

model can reproduce the observed results quantitatively. In Fig. 8(a), the division time  $\Delta T$  is plotted as a function of the size at birth  $L_0$  for every cell cycle (lighter dots), together with the mean corresponding to a specific division time (darker dots), for two nutrient conditions from our mother machine experiments. Consistent with previous experiments [8], the size at birth has a negative correlation with respect to the division time; i.e., longer cells tend to divide faster. The mean value of the division time decreases for richer nutrient condition, suggesting that the rate of cell growth is influenced by the nutrient condition.

In Fig. 8(b), values of  $\Delta L$  for each generation are shown as a function of  $L_0$  (lighter dots). It is clear from the mean values (darker dots) that the elongation  $\Delta L$  does not depend on the initial size  $L_0$ , but it slightly increases when the nutrient conditions are richer. The independence of  $\Delta L$  on  $L_0$  is also consistent with previous experiments [8]. In fact, this

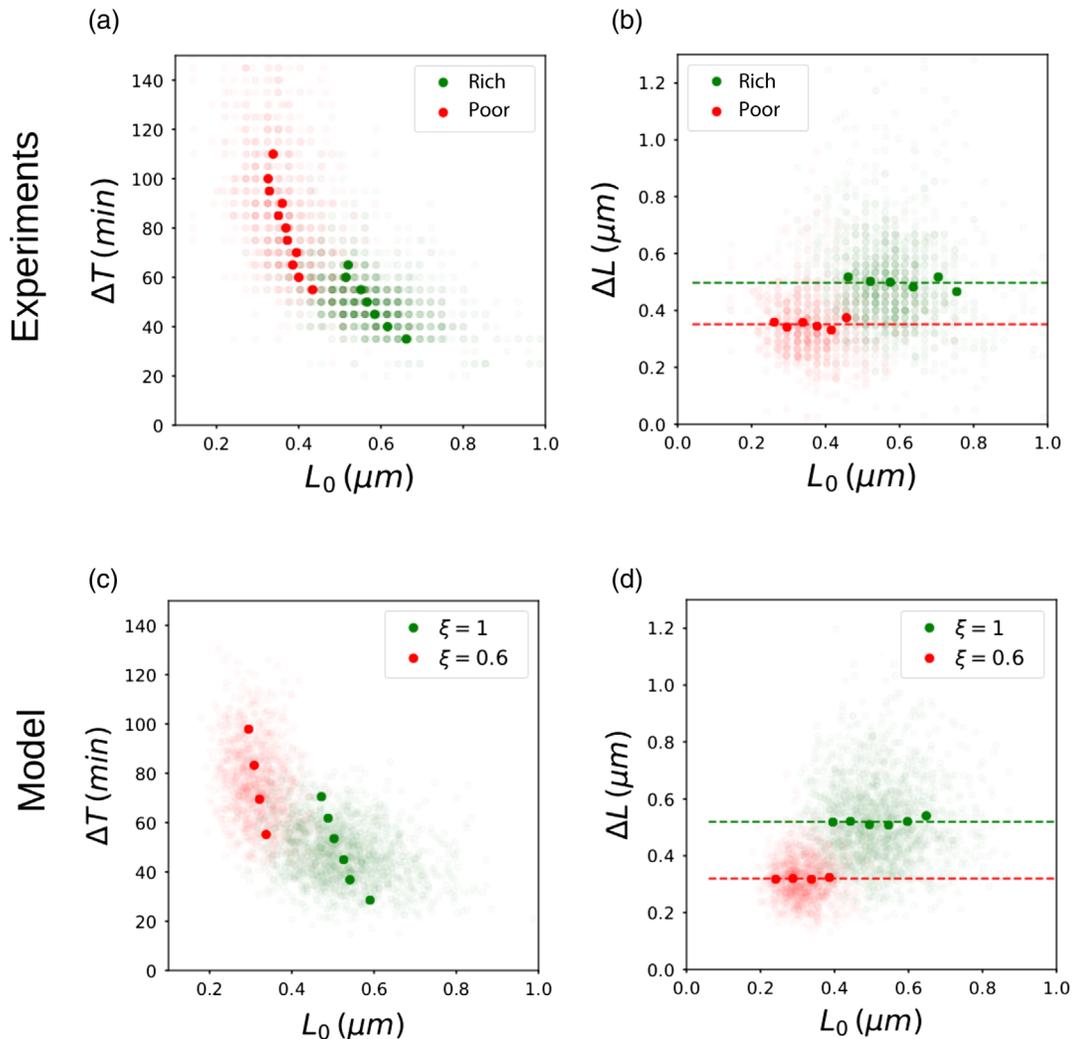


FIG. 8. Comparison between experiments and model results. (a) Division time  $\Delta T$  versus the cell size at birth  $L_0$  from the mother machine experiment for rich and poor nutrient conditions. Each lighter dot represents a single cell cycle data. Darker dots are bin averages. (b) Elongation  $\Delta L$  versus  $L_0$  for the same experiments as in (a). (c),(d) The same plots as (a) and (b), respectively, from numerical simulations of our model. The same  $P_d(Z)$  as in Fig. 5 is used. See Appendix B for details on the simulations.

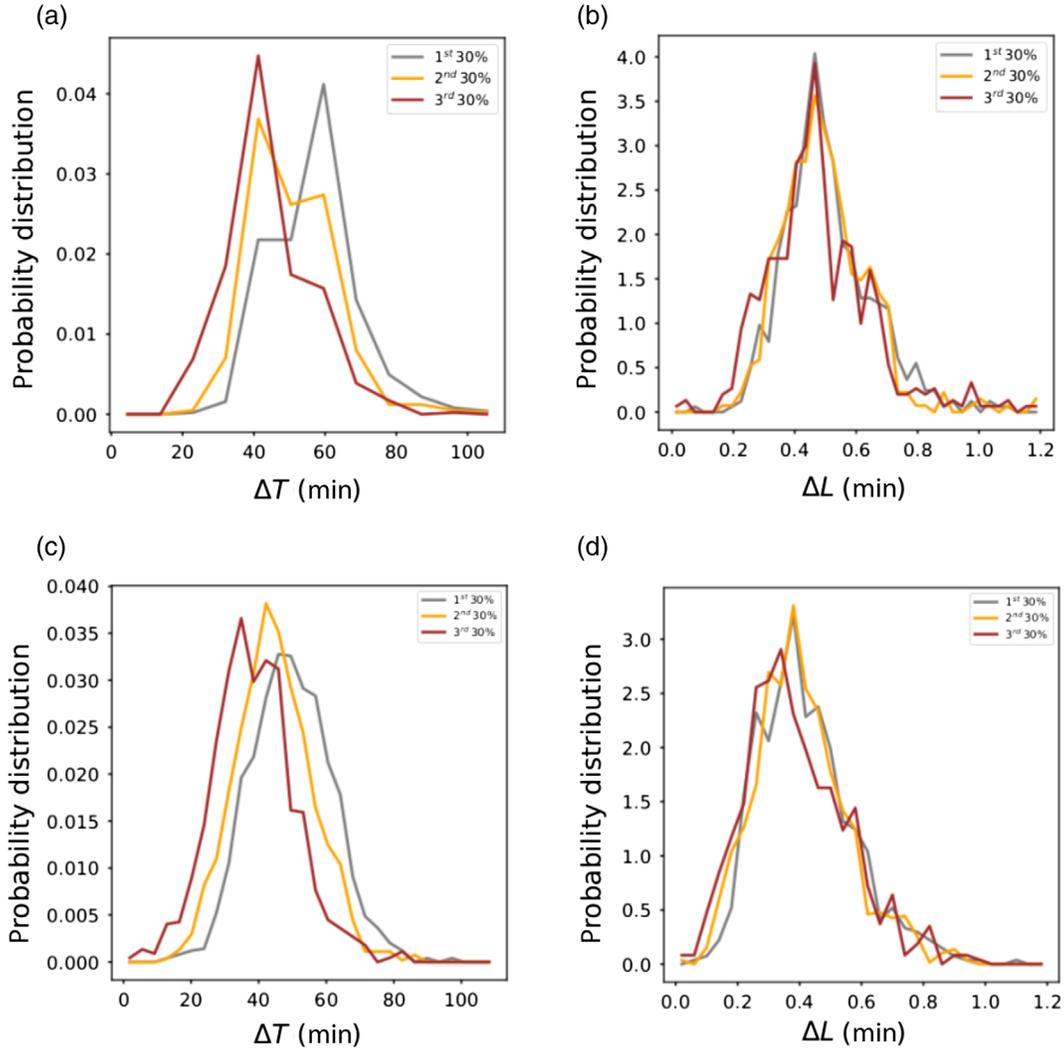


FIG. 9. Conditional distributions in experiments and model. (a),(b) The distributions of the division time ( $\Delta T$ ) and the length elongation ( $\Delta L$ ) when the size at birth  $L_0$  is in three different percentiles from experimental measurements. The corresponding distributions from model results are shown in (c),(d).

empirical observation is incorporated into the phenomenological *adder model* [27,43], which starts from the assumption that a cell divides when its elongation reaches a fixed value independent of its size at birth  $L_0$ . The negative correlation between  $\Delta T$  and  $L_0$  and the independence of  $\Delta L$  on  $L_0$  are the two key features in cell size control we wish to reproduce from our growth-division model.

In Fig. 8(c), we show the division time  $\Delta T$  versus the size at birth  $L_0$  for each individual cell cycle (lighter dots) from our model. We divide the values of  $L_0$  into equally spaced bins and calculate the average of  $\Delta T$  and  $L_0$  for each bin (darker dots). The same analysis is used to plot the increment  $\Delta L$  as a function of the size at birth  $L_0$  for different cell cycles in Fig. 8(d). Our model reproduces the negative correlation between  $L_0$  and  $\Delta T$  as well as the independence of  $\Delta L$  on  $L_0$ . Furthermore, the dependence of  $\Delta T$ ,  $L_0$ , and  $\Delta L$  on the nutrient conditions is also in quantitative agreement with our mother machine experiments. Moreover, for a given nutrient

condition, it is possible to appreciate the dependency of  $\Delta T$  and  $\Delta L$  on  $L_0$  by considering the conditional probability distributions. In Fig. 9, we plot the distributions of  $\Delta T$  and  $\Delta L$  for cell cycles in which the initial length  $L_0$  is in specific ranges. We can, thus, notice that, while  $\Delta T$  tends to decrease for larger  $L_0$ , the distributions of  $\Delta L$  collapse, and the results from the simulations are in line with the ones from the experiments.

#### APPENDIX D: DERIVATION OF THE LANGEVIN EQUATION FOR THE CONCENTRATION USING A MEAN-FIELD APPROXIMATION

In this section and the next one, we derive the differential equation of the probability distribution  $P(c)$  and the relation between the mean and the variance of the protein concentration in our mean-field model. To do so, we first write the time derivative of  $c$  in terms of derivatives of  $N$  and  $L$ :

$$\frac{dc}{dt} = \frac{1}{L} \frac{dN}{dt} - \frac{c}{L} \frac{dL}{dt}. \quad (\text{D1})$$

Using Eqs. (4) and (5) in the main text, the Langevin equation becomes

$$\frac{dc}{dt} = \hat{K}_n - \hat{K}_l c + c \sum_i^{n_d(t)} [(w_i^{(l)} - w_i^{(n)}) \delta(t - t_i)]. \quad (\text{D2})$$

The two variables for the division term are

$$w_i^{(l)} = \frac{L_a}{L_b}, \quad (\text{D3})$$

$$w_i^{(n)} = \frac{N_a}{N_b}, \quad (\text{D4})$$

where  $L_b$ ,  $N_b$ ,  $L_a$ , and  $N_a$  are the values of the size and the protein number before and after the division, respectively.

The difference  $\delta w_i = w_i^{(l)} - w_i^{(n)}$  has the following properties:

$$\begin{aligned} \langle \delta w_i \rangle &= 0, \\ \langle \delta w_i \delta w_j \rangle &= 2\Delta_w \delta_{ij}, \end{aligned} \quad (\text{D5})$$

where  $\Delta_w$  is the variance of the partition difference. We can, thus, define the noise due to cell division:

$$\eta_d(t) \equiv \sum_i^{n_d(t)} \delta w_i \delta[t - t_i(Z)]. \quad (\text{D6})$$

This new stochastic variable has still mean equal to zero, and it is delta-function correlated:

$$\langle \eta_d(t) \rangle = 0 \quad (\text{D7})$$

and

$$\langle \eta_d(t_1) \eta_d(t_2) \rangle = 2\Delta_d \delta(t_2 - t_1), \quad (\text{D8})$$

where

$$\Delta_d = \frac{1}{2\langle \Delta T \rangle} \left\langle \left( \frac{L_a}{L_b} - \frac{N_a}{N_b} \right)^2 \right\rangle. \quad (\text{D9})$$

The other sources of noise are discussed in Appendix F.

### APPENDIX E: DERIVATION OF THE SQUARE TAYLOR'S LAW

From the Langevin equation [Eq. (9) in the main text], the steady-state protein concentration distribution function  $P(c)$  satisfies the stationary Fokker-Planck equation [50]:

$$\langle \hat{K}_l \rangle \frac{d}{dc} [(\mu - c)P] = \frac{d^2}{dc^2} [D(c)P], \quad (\text{E1})$$

where  $D(c) = \mu^2 \Delta_a + 2\mu c \Delta_{am} + c^2 \Delta_m$  with  $\Delta_{(a,m,am)}$  representing the noise strength for  $\eta_a$ ,  $\eta_m$ , and their correlation.

The solution of Eq. (E1) is reported in the main text [Eq. (13)]. Here, we derive the relation between the mean and the variance.

After a first integration over  $c$ , we obtain

$$\langle \hat{K}_l \rangle [(\mu - c)P(c)] = \frac{d}{dc} [D(c)P(c)]. \quad (\text{E2})$$

If we integrate on both sides and we use the fact that  $P(c)D(c) \rightarrow 0$  when  $c \rightarrow \infty$ , we have

$$\langle \hat{K}_l \rangle (\mu - \langle c \rangle) = 0, \quad (\text{E3})$$

and, thus,  $\langle c \rangle = \mu$ . If instead we multiply Eq. (E2) on both sides for  $(\mu - c)/\langle \hat{K}_l \rangle$  and we integrate over  $c$ , on the left-hand side we simply have  $\sigma_c^2$ . Therefore, after an integration by parts, the equation reduces to

$$\begin{aligned} \sigma_c^2 &= \frac{1}{\langle \hat{K}_l \rangle} \int D(c)P(c)dc \\ &= \frac{1}{\langle \hat{K}_l \rangle} [\mu^2 \Delta_a + 2\mu^2 \Delta_{am} + (\mu^2 + \sigma_c^2) \Delta_m]. \end{aligned} \quad (\text{E4})$$

By solving Eq. (E4), we obtain the variance as a function of the noise strengths and  $\mu$ :

$$\sigma_c^2 = \frac{\Delta_a + 2\Delta_{am} + \Delta_m}{\langle \hat{K}_l \rangle - \Delta_m} \mu^2 = \frac{\Delta_d + \Delta_g}{\langle \hat{K}_l \rangle - \Delta_m} \mu^2, \quad (\text{E5})$$

where  $\Delta_a = \langle \hat{K}_l \rangle^2 (\Delta_{(r)} + \Delta_n)$  is the noise strength for  $\eta_a$ ,  $\Delta_{am} = -\langle \hat{K}_l \rangle^2 \Delta_{(r)}$  is the correlation between  $\eta_a$  and  $\eta_m$ , and  $\Delta_m = \Delta_d + \langle \hat{K}_l \rangle^2 (\Delta_{(r)} + \Delta_l)$  is the strength of the noise  $\eta_m$  with  $\Delta_{(r)} = \langle \delta r^2 \rangle / \langle r \rangle^2$  the noise strength for  $\delta r / \langle r \rangle$ . The above equation [Eq. (E5)] is the same as Eq. (11) in the main text.

### APPENDIX F: DETAILED NOISE ANALYSIS IN DATA BASED ON THE MODEL

The averages of the effective growth and expression rates  $\langle \hat{K}_{l,n} \rangle$  can be calculated from experiments in the following way:

$$\begin{aligned} \langle \hat{K}_l \rangle &= \left\langle \frac{1}{L} \frac{\delta L}{\delta t} \right\rangle, \\ \langle \hat{K}_n \rangle &= \left\langle \frac{1}{L} \frac{\delta N}{\delta t} \right\rangle, \end{aligned} \quad (\text{F1})$$

where  $\delta L$ ,  $\delta N$ , and  $\delta t$  are the smallest increment allowed by the experimental setup. In our case, the time step is  $\delta t = 5$  min. From the experiments, we do not have direct information about  $R$ , and the noise from the observable variables includes contributions from  $r$  and contributions from  $N$  and  $L$ . In other words, we can directly measure  $\chi_l$  and  $\chi_n$  but not  $\eta_l$  and  $\eta_n$ . The noise strengths of  $\chi_l$  and  $\chi_n$  are given by the following formula:

TABLE II. Relevant parameters inferred from the experiments for different promoter strengths and nutrient conditions. These values are calculated by averaging over different mother cells with the same promoter and under the same growth conditions.

Conditions	$\Delta_{\chi_l}$ (min)	$\Delta_{\chi_n}$ (min)	$\langle \hat{K}_l \rangle (\text{min}^{-1})$	$\Delta_{(r)}$ (min)	$\Delta_m (\text{min}^{-1})$	$\Delta_d (\text{min}^{-1})$	$\Delta_g (\text{min}^{-1})$
Rich 4	0.619	2.4	0.017	0.66	0.000 23	$5.5 \times 10^{-5}$	$4.8 \times 10^{-4}$
Rich 5	0.412	2.8	0.017	0.12	0.000 15	$3.7 \times 10^{-5}$	$8.1 \times 10^{-4}$
Rich B	0.429	1.7	0.017	0.56	0.000 17	$4.5 \times 10^{-5}$	$3.1 \times 10^{-4}$
Rich C	0.654	2.3	0.016	0.63	0.0002	$3.7 \times 10^{-5}$	$4.4 \times 10^{-4}$
Rich D	0.904	1.8	0.015	0.85	0.000 22	$3 \times 10^{-5}$	$2.1 \times 10^{-4}$
Poor 4	9.42	8.5	0.0075	5.3	0.000 57	$3.9 \times 10^{-5}$	$4.3 \times 10^{-4}$
Poor 5	8.2	6.25	0.0092	4.3	0.000 74	$3.9 \times 10^{-5}$	$4.7 \times 10^{-4}$
Poor B	8.2	5.15	0.0087	3.5	0.000 63	$2.5 \times 10^{-5}$	$5.5 \times 10^{-4}$
Poor C	4.7	4.1	0.0092	2.2	0.000 42	$2.1 \times 10^{-5}$	$3.8 \times 10^{-4}$
Poor D	5.49	6.3	0.0087	3.4	0.000 44	$1.9 \times 10^{-5}$	$3.6 \times 10^{-4}$

$$\Delta_{(\chi_l, \chi_n)} = \frac{1}{2} \int ds \langle \chi_{(l,n)}(t) \chi_{(l,n)}(t+s) \rangle. \quad (\text{F2})$$

Here, the average is over an ensemble of time traces, and the result does not depend on the specific time  $t$ .

Since the noises  $\eta_n$  and  $\eta_l$  are independent, when we calculate the correlation between  $\chi_l$  and  $\chi_n$ , we obtain the strength of the noise  $\delta r / \langle r \rangle$ —the only nonzero term in the correlation—that we call  $\Delta_{(r)}$  (not to be confused with the strength of  $\eta_r$ ):

$$\Delta_{(r)} = \frac{1}{2} \int ds \langle \chi_l(t) \chi_n(t+s) \rangle. \quad (\text{F3})$$

Once we have  $\Delta_{(r)}$ , by subtracting it from  $\Delta_{\chi_l}$  and  $\Delta_{\chi_n}$ , we obtain  $\Delta_l$  and  $\Delta_n$ , respectively.

We can now write  $\Delta_a$ ,  $\Delta_{am}$ , and  $\Delta_m$  in terms of these noise strengths:

$$\begin{aligned} \Delta_a &= \langle \hat{K}_l \rangle^2 \Delta_{\chi_n}, \\ \Delta_m &= \Delta_d + \langle \hat{K}_l \rangle^2 \Delta_{\chi_l}, \\ \Delta_{am} &= -\langle \hat{K}_l \rangle^2 \Delta_{(r)}. \end{aligned}$$

The *growth* noise strength is defined as

$$\Delta_g = \langle \hat{K}_l \rangle^2 (\Delta_{\chi_n} + \Delta_{\chi_l} - 2\Delta_{(r)}). \quad (\text{F4})$$

The noise strengths and other relevant parameters averaged over all cells with each of the combinations of nutrient conditions and promoter strengths are reported in Table II.

## APPENDIX G: TAYLOR'S LAW IS CONFIRMED IN ADDITIONAL EXPERIMENTS

We investigate whether Taylor's law would still be valid in case the constitutive promoters are integrated in the chromosome. We plot the variance of the fluorescence

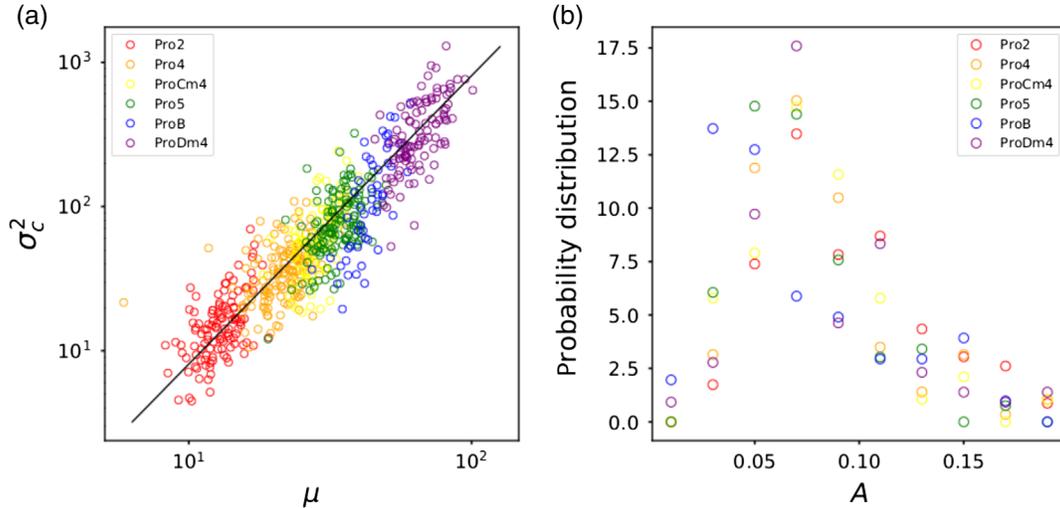


FIG. 10. Taylor's law in strains with chromosomal integrated promoters. (a) Variance as a function of the mean for fluorescence data in which constitutive promoters from the same family of promoters used in the current study are integrated in the chromosome. The prefactor determined from the data for individual mother cells is found to lie in the range  $A = 0.083 \pm 0.046$ . The slope of the black line is 2. (b) Distributions of the prefactor  $A \equiv \sigma_c^2 / \mu^2$  for different promoters. No statistically significant correlation between  $A$  and  $\mu$  is found with  $C_{A\mu} = -0.07$ , which is consistent with experiments using plasmids. Data taken from Ref. [22].

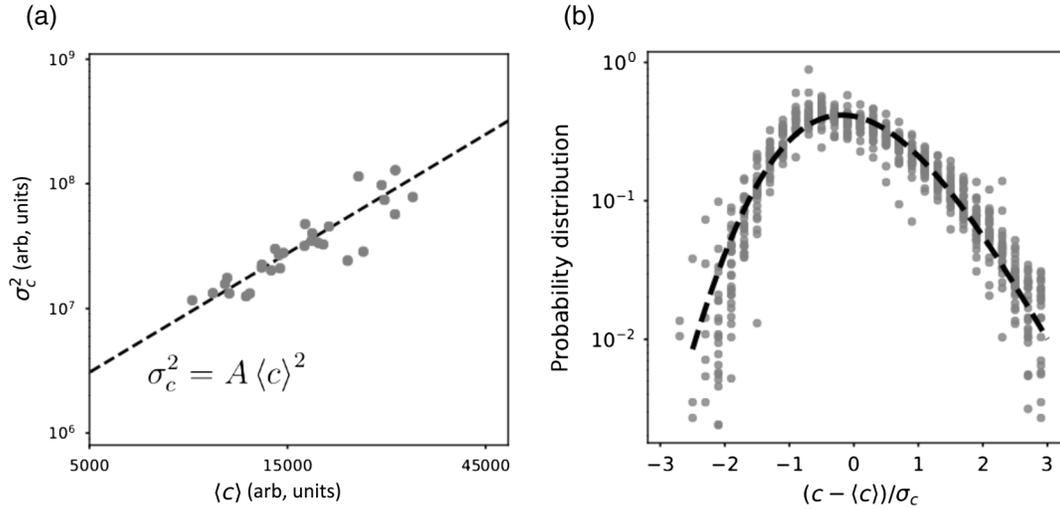


FIG. 11. (a) The variance versus mean of the fluorescence density, both in arbitrary units (arb. units), from the experiments reported in Ref. [25]. The dashed line is obtained using an average value for the coefficient  $A = \sigma_c^2/\langle c \rangle^2$ . The value obtained in this set of data is  $\langle A \rangle = 0.12$ . (b) The scaled protein concentration distributions from the experimental data reported in Ref. [25]. The black curve is the distribution that we show in the main text.

intensity as a function of the mean value in the case of experiments using the same constitutive promoters, with the difference that they are integrated in the chromosome rather than using plasmids. As shown in Fig. 10, the relation  $\sigma_c^2 = A \langle c \rangle^2$  is confirmed for almost one order of magnitude of the mean value, with  $A = 0.083 \pm 0.046$ .

Moreover, the same relation between the variance and the mean with a different prefactor  $A$  is confirmed by data collected from fluorescence experiments described in Ref. [25], kindly shared by Hanna Salman (see Fig. 11).

#### APPENDIX H: DERIVATION OF THE DIVISION TIME DISTRIBUTION

We see that the probability rate of division is a function  $P_d(Z)$ , depending on the stochastic variable  $Z$ , that is the division protein number, following the formula

$$P_d(Z) = \frac{\Delta t^{-1}}{1 + \exp[(Z_0 - Z)/\Delta Z]}. \quad (\text{H1})$$

Then, the survival probability  $S(t)$  for a division not to take place before a time  $t$  is the solution of the differential equation:

$$\frac{dS}{dt} = -S(t)P_d[Z(t)], \quad (\text{H2})$$

and, thus, we have

$$S(t) = \exp \left[ - \int_0^t P_d[Z(t')] dt' \right]. \quad (\text{H3})$$

The survival probability can be used to determine the distribution of the division time  $\Delta T$ , that we call  $p(\Delta T)$ :

$$p(\Delta T) = - \frac{dS}{dt} \Big|_{t=\Delta T} = P_d[Z(\Delta T)] \exp \left[ - \int_0^{\Delta T} P_d[Z(t')] dt' \right]. \quad (\text{H4})$$

Note that the variance of  $\Delta T$  is finite, given that the distribution decreases exponentially for large times, and as a consequence  $n_d$  has finite variance as well.

#### APPENDIX I: ADDITIONAL FIGURES 11 AND 12

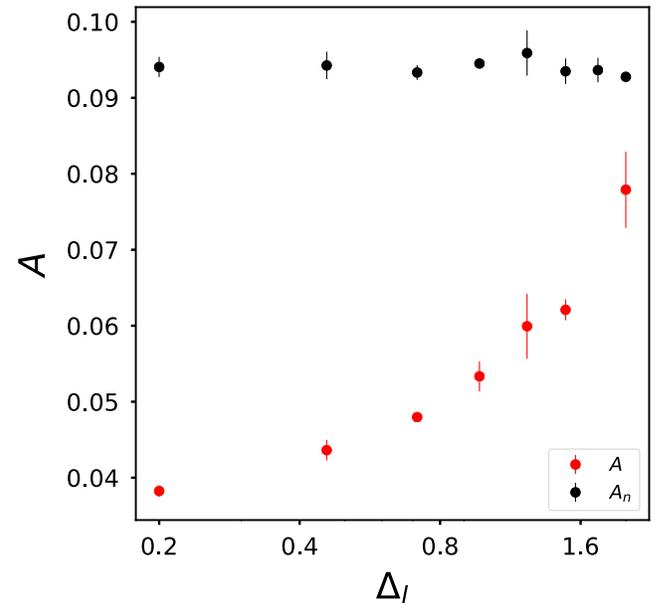


FIG. 12. Prefactors for the concentration ( $A$ ) and the protein number ( $A_n$ ) obtained from the numerical solution of the stochastic differential equations for different values of the parameter  $\Delta_l$ . The error bars are obtained by using the standard deviation of the fluctuations of the prefactor for different promoter strengths.

- [1] S. Di Talia, J. M. Skotheim, Ja. 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 (London)* **448**, 947 (2007).
- [2] A. Eldar and M. B. Elowitz, *Functional Roles for Noise in Genetic Circuits*, *Nature (London)* **467**, 167 (2010).
- [3] J. Lin and A. Amir, *Homeostasis of Protein and mRNA Concentrations in Growing Cells*, *Nat. Commun.* **9**, 4496 (2018).
- [4] N. Brenner, E. Braun, A. Yoney, L. Susman, J. Rotella, and H. Salman, *Single-Cell Protein Dynamics Reproduce Universal Fluctuations in Cell Populations*, *Eur. Phys. J. E* **38**, 102 (2015).
- [5] Z. Cao and R. Grima, *Analytical Distributions for Detailed Models of Stochastic Gene Expression in Eukaryotic Cells*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 4682 (2020).
- [6] L. Susman, M. Kohram, H. Vashistha, J. T. Nechleba, H. Salman, and N. Brenner, *Individuality and Slow Dynamics in Bacterial Growth Homeostasis*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, E5679 (2018).
- [7] M. Wallden, D. Fange, E. G. Lundius, O. Baltekin, and J. Elf, *The Synchronization of Replication and Division Cycles in Individual E. Coli Cells*, *Cell* **166**, 729 (2016).
- [8] 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*, *Curr. Biol.* **25**, 385 (2015).
- [9] 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 (London)* **523**, 357 (2015).
- [10] A. S. Sassi, M. Garcia-Alcala, M. J. Kim, P. Cluzel, and Y. Tu, *Filtering Input Fluctuations in Intensity and in Time Underlies Stochastic Transcriptional Pulses without Feedback*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 26608 (2020).
- [11] D. Huh and J. Paulsson, *Non-genetic Heterogeneity from Stochastic Partitioning at Cell Division*, *Nat. Genet.* **43**, 95 (2011).
- [12] C. Jia and R. Grima, *Frequency Domain Analysis of Fluctuations of mRNA and Protein Copy Numbers within a Cell Lineage: Theory and Experimental Validation*, *Phys. Rev. X* **11**, 021032 (2021).
- [13] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, *Quantifying E. Coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells*, *Science* **329**, 533 (2010).
- [14] L. R. Taylor, *Aggregation, Variance and the Mean*, *Nature (London)* **189**, 732 (1961).
- [15] J. E. Cohen, *Species-Abundance Distributions and Taylor's Power Law of Fluctuation Scaling*, *Theor. Ecol.* **13**, 607 (2020).
- [16] M. Xu and J. E. Cohen, *Spatial and Temporal Autocorrelations Affect Taylor's Law for US County Populations: Descriptive and Predictive Models*, *PLoS One* **16**, e0245062 (2021).
- [17] M. Keeling and B. Grenfell, *Stochastic Dynamics and a Power Law for Measles*, *Phil. Trans. R. Soc. B* **354**, 769 (1999).
- [18] J. E. Cohen, *Taylor's Power Law of Fluctuation Scaling and the Growth-Rate Theorem*, *Theor. Popul. Biol.* **88**, 94 (2013).
- [19] J. H. Davis, A. J. Rubin, and R. T. Sauer, *Design, Construction and Characterization of a Set of Insulated Bacterial Promoters*, *Nucl. Acids Res.* **39**, 1131 (2011).
- [20] J. R. S. Newman, S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. DeRisi, and J. S. Weissman, *Single-Cell Proteomic Analysis of S. Cerevisiae Reveals the Architecture of Biological Noise*, *Nature (London)* **441**, 840 (2006).
- [21] I. Soifer, L. Robert, and A. Amir, *Single-Cell Analysis of Growth in Budding Yeast and Bacteria Reveals a Common Size Regulation Strategy*, *Curr. Biol.* **26**, 356 (2016).
- [22] J. M. Kim, M. Garcia-Alcala, E. Balleza, and P. Cluzel, *Stochastic Transcriptional Pulses Orchestrate Flagellar Biosynthesis in E. Coli*, *Sci. Adv.* **6** (2020).
- [23] P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun, *Robust Growth of Escherichia Coli*, *Curr. Biol.* **20**, 1099 (2010).
- [24] J. Ollion, M. Elez, and L. Robert, *High-Throughput Detection and Tracking of Cells and Intracellular Spots in Mother Machine Experiments*, *Nat. Protoc.* **14**, 3144 (2019).
- [25] H. Salman, N. Brenner, C.-K. Tung, N. Elyahu, E. Stolovicki, L. Moore, A. Libchaber, and E. Braun, *Universal Protein Fluctuations in Populations of Microorganisms*, *Phys. Rev. Lett.* **108**, 238105 (2012).
- [26] N. Brenner, C. M. Newman, D. Osmanović, Y. Rabin, H. Salman, and D. L. Stein, *Universal Protein Distributions in a Model of Cell Growth and Division*, *Phys. Rev. E* **92**, 042713 (2015).
- [27] A. Amir, *Cell Size Regulation in Bacteria*, *Phys. Rev. Lett.* **112**, 208102 (2014).
- [28] P. A. Fantes, W. D. Grant, R. H. Pritchard, P. E. Sudbery, and A. E. Wheals, *The Regulation of Cell Size and the Control of Mitosis*, *J. Theor. Biol.* **50**, 213 (1975).
- [29] S. Wold, K. Skarstad, H. B. Steen, T. Stokke, and E. Boye, *The Initiation Mass for DNA Replication in Escherichia Coli k-12 Is Dependent on Growth Rate*, *EMBO J.* **13**, 2097 (1994).
- [30] E. Boye and K. Nordstrom, *Coupling the Cell Cycle to Cell Growth*, *EMBO Rep.* **4**, 757 (2003).
- [31] W. D. Donachie and G. W. Blakely, *Coupling the Initiation of Chromosome Replication to Cell Size in Escherichia Coli*, *Curr. Opin. Microbiol.* **6**, 146 (2003).
- [32] M. Basan, M. Zhu, X. Dai, M. Warren, D. Svin, Y.-P. Wang, and T. Hwa, *Inflating Bacterial Cells by Increased Protein Synthesis*, *Mol. Syst. Biol.* **11**, 836 (2015).
- [33] In Eq. (3), we neglect the variation of  $\mu_i$  among cells within each experimental condition  $\mathcal{E}$ , which is much smaller than their average  $\mu(\mathcal{E})$ .
- [34] M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa, *Interdependence of Cell Growth and Gene Expression: Origins and Consequences*, *Science* **330**, 1099 (2010).
- [35] 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*, *Curr. Biol.* **29**, 1760 (2019).
- [36] K. Dai and J. Lutkenhaus, *ftsZ Is an Essential Cell Division Gene in Escherichia Coli*, *J. Bacteriol.* **173**, 3500 (1991).

- [37] A. Mukherjee, K. Dai, and J. Lutkenhaus, *Escherichia Coli Cell Division Protein FtsZ Is a Guanine Nucleotide Binding Protein*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1053 (1993).
- [38] 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 Comput. Biol.* **12**, e1004972 (2016).
- [39] We consider only cell division as the cause of protein number reduction here. However, including protein degradation does not change the main results in this paper due to the multiplicative nature of the degradation process.
- [40] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, *Stochastic Gene Expression in a Single Cell*, *Science* **297**, 1183 (2002).
- [41] F. Bertaux, J. von Kgelgen, S. Marguerat, and V. Shahrezaei, *A Bacterial Size Law Revealed by a Coarse-Grained Model of Cell Physiology*, *PLoS Comput. Biol.* **16**, e1008245 (2020).
- [42] 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*, *Phys. Rev. E* **101**, 062406 (2020).
- [43] L. Willis and K. C. Huang, *Sizing up the Bacterial Cell Cycle*, *Nat. Rev. Microbiol.* **15**, 606 (2017).
- [44] This scale invariance is rooted in the invariance of the full model [Eqs. (4)–(7)] under the transformation:  $N \rightarrow \lambda N$ ,  $K_n \rightarrow \lambda K_n$  for arbitrary positive  $\lambda$ .
- [45] M. Jahn, S. Gunther, and S. Muller, *Non-random Distribution of Macromolecules as Driving Forces for Phenotypic Variation*, *Curr. Opin. Microbiol.* **25**, 49 (2015).
- [46] J. W. Ng, D. Chatenay, J. Robert, and M. G. Poirier, *Plasmid Copy Number Noise in Monoclonal Populations of Bacteria*, *Phys. Rev. E* **81**, 011909 (2010).
- [47] J. Kim, A. Darlington, M. Salvador, J. Utrilla, and J. I. Jiménez, *Trade-offs between Gene Expression, Growth and Phenotypic Diversity in Microbial Populations*, *Curr. Opin. Biotechnol.* **62**, 29 (2020).
- [48] J. R. S. Newman, S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. DeRisi, and J. S. Weissman, *Single-Cell Proteomic Analysis of *S. Cerevisiae* Reveals the Architecture of Biological Noise*, *Nature (London)* **441**, 840 (2006).
- [49] Z. Eisler, I. Bartos, and J. Kertsz, *Fluctuation Scaling in Complex Systems: Taylor's Law and beyond*, *Adv. Phys.* **57**, 89 (2008).
- [50] We use the Ito interpretation for the multiplicative noise here. Using the Stratonovich interpretation does not change the qualitative results.