

**Relevant parameters in models of cell division control**Jacopo Grilli,<sup>1,\*</sup> Matteo Osella,<sup>2</sup> Andrew S. Kennard,<sup>3,4</sup> and Marco Cosentino Lagomarsino<sup>5,6,7,†</sup><sup>1</sup>*Department of Ecology and Evolution, University of Chicago, 1101 E 57th Street, Chicago, Illinois 60637, USA*<sup>2</sup>*Dipartimento di Fisica and INFN, University of Torino, V. Pietro Giuria 1, Torino, I-10125, Italy*<sup>3</sup>*Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, United Kingdom*<sup>4</sup>*Biophysics Program, Stanford University, Stanford, California 94305, USA*<sup>5</sup>*Sorbonne Universités, UPMC Univ Paris 06, UMR 7238, Computational and Quantitative Biology, 15 rue de l'École de Médecine Paris, France*<sup>6</sup>*CNRS, UMR 7238, Paris, France*<sup>7</sup>*FIRC Institute of Molecular Oncology (IFOM), 20139 Milan, Italy*

(Received 5 July 2016; revised manuscript received 4 February 2017; published 17 March 2017)

A recent burst of dynamic single-cell data makes it possible to characterize the stochastic dynamics of cell division control in bacteria. Different models were used to propose specific mechanisms, but the links between them are poorly explored. The lack of comparative studies makes it difficult to appreciate how well any particular mechanism is supported by the data. Here, we describe a simple and generic framework in which two common formalisms can be used interchangeably: (i) a continuous-time division process described by a hazard function and (ii) a discrete-time equation describing cell size across generations (where the unit of time is a cell cycle). In our framework, this second process is a discrete-time Langevin equation with simple physical analogues. By perturbative expansion around the mean initial size (or interdivision time), we show how this framework describes a wide range of division control mechanisms, including combinations of time and size control, as well as the constant added size mechanism recently found to capture several aspects of the cell division behavior of different bacteria. As we show by analytical estimates and numerical simulations, the available data are described precisely by the first-order approximation of this expansion, i.e., by a “linear response” regime for the correction of size fluctuations. Hence, a single dimensionless parameter defines the strength and action of the division control against cell-to-cell variability (quantified by a single “noise” parameter). However, the same strength of linear response may emerge from several mechanisms, which are distinguished only by higher-order terms in the perturbative expansion. Our analytical estimate of the sample size needed to distinguish between second-order effects shows that this value is close to but larger than the values of the current datasets. These results provide a unified framework for future studies and clarify the relevant parameters at play in the control of cell division.

DOI: [10.1103/PhysRevE.95.032411](https://doi.org/10.1103/PhysRevE.95.032411)**I. INTRODUCTION**

Today, quantitative data of single dividing cells across generations and lineages can be produced with high throughput and spatiotemporal resolution. Such state-of-the-art data have renewed the investigation of bacterial growth and division. Due to the intrinsic stochasticity of these systems, approaches based on statistical physics play a primary role. One example is the decision mechanism by which a cell divides, which has a key role in its size determination.

Several recent findings have progressed this field by joint use of theoretical models and single-cell experiments. Namely, (i) interesting scaling behavior emerges for the distributions of key variables such as doubling times and cell sizes across conditions and species [1–4], suggesting the existence of universal parameters setting these variables; (ii) fluctuations of different quantities bear intriguing relationships—for example, cell size and doubling time fluctuations are interlinked relative to the average growth rate [1,5]; (iii) the mechanisms of division control can be explored using theoretical models, formulated as stochastic processes (of different kinds) whose dynamic variables are cell size, time, and division events [1,5–8].

These mechanisms of division control are the most studied due to their direct biological relevance. The data typically rule out control mechanisms based on measurement of either cell size or cell cycle time alone [5,7,9,10]. Concerted controls where multiple variables (e.g., time and size) may enter jointly have been proposed [6,7]. Several studies in *Escherichia coli* [5,8] and other microbes [5,10–12] have argued for a mechanism in which the size extension in a single cell cycle is nearly constant and independent of the initial size of the cell (sometimes called “adder” mechanism of division control). However, it is clear that the constant added size is not the only trend found in the data [5,8,13–15], and that it is not a necessary and sufficient condition for the observed scaling behavior and fluctuation patterns [1]. More broadly, the question of how much a mechanism can be isolated and specified with available data is still open.

Additionally, existing studies so far have relied on different modeling approaches and raise the need for a unified framework. Specifically, two dominant formalisms emerge. The first describes the continuous-time division process by a hazard function. The hazard function defines the probability per unit time that a cell divides, as a function of the values of measurable variables such as initial and/or current size, incremental or multiplicative growth, and elapsed time from cell division [5,7,16]. The second formalism describes cell size across generations as a discrete-time autoregressive process (where a unit of time is a cell cycle) [6,10,12].

\*jgrilli@uchicago.edu

†marco.cosentino-lagomarsino@upmc.fr

Here, we propose a unified framework linking explicitly these two formalisms and we pose the question of the general possibility to distinguish mechanisms from data. By expanding around the mean initial size or interdivision time (generalizing the approach of Ref. [6]), we show explicitly how this framework describes a wide range of division control mechanisms, including combinations of time and size control, as well as control by constant added size. As we show by analytical estimates and numerical simulation, the available data are characterized with great precision by the first-order approximation of this expansion. Hence, a single dimensionless parameter defines the strength and the action of the division control against cell-to-cell variability. However, this parameter may emerge from several mechanisms, which are distinguished only by higher-order terms in our perturbative expansion. Finally, we estimate the sample size needed to distinguish between second-order effects, and show that it is larger than the size of currently available datasets.

## II. BACKGROUND

### A. Theoretical description of division control

This section describes the basic definitions and ingredients of the modeling framework. We will assume exponential growth of the cell size  $x(t) = x_0 e^{\alpha t}$ , which is well supported in the literature [2,5,7,8] and, as in previous studies, we neglect fluctuations of the growth rate  $\alpha$  [2,5,7]. A cell divides at a size  $x_f$  into two cells of equal size  $x_f/2$  (we thus do not consider the small fluctuations around binary fission, the process of filamentation and recovery, or species with nonbinary division [7,10,14,17]).

A control mechanism defines the division size  $x_f$ . In absence of this control, fluctuations of cell size may grow indefinitely in time. The full information on division control is encoded by the function  $p(x_f|x_0, \alpha)$ , the conditional probability that a cell, born at size  $x_0$  and growing with a growth rate  $\alpha$ , divides at size  $x_f$ . The growth-division process is therefore defined by four variables  $x_0, x, t, \alpha$ , with the constraint of exponential growth (which reduces to three the number of independent variables) and the model assumption of negligible fluctuations in  $\alpha$ . This allows different equivalent parametrization of the process. A quantity of interest is the size at birth of a cell, followed across generations. Given the probability  $p_b^i(x_0|\alpha)$  of observing a cell at generation  $i$  with initial size  $x_0$ , the following Chapman-Kolmogorov equation gives the same probability at the subsequent generation:

$$p_b^{i+1}(x_0|\alpha) := 2 \int_0^\infty dy p(2x_0|y, \alpha) p_b^i(y|\alpha), \quad (1)$$

where  $p(x_f|x_0, \alpha)$  plays the role of a transition probability. These probabilities quantify the variability of single cells. Strictly speaking, Eq. (1) applies only along isolated, or at least nonoverlapping, lineages. Population averages (when averages are computed over a population of cells that share common ancestors) are in fact different from single-cell averages, since the latter neglect the correlations among cells in a population [4]. On the other hand, the approach of Eq. (1) has been previously applied to cell lineages within a population in microcolony data with fairly satisfactory results [1].

The assumption of exponential single-cell growth implies that in this process the noise on doubling times has a multiplica-

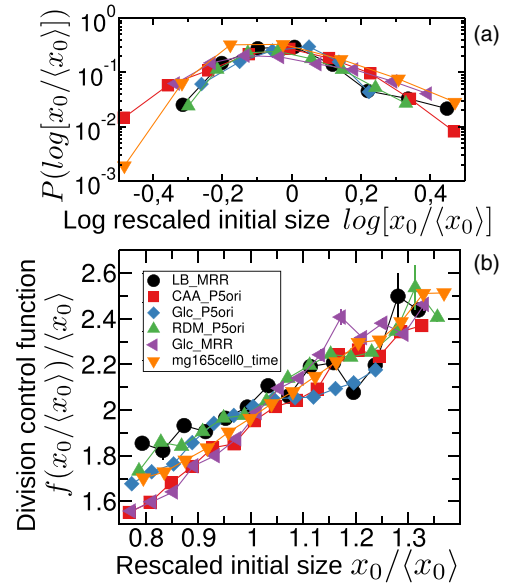


FIG. 1. The division-control function follows finite-size scaling. (A) Collapse of the probability distribution of rescaled logarithmic cell size  $\log(x_0/\langle x_0 \rangle)$  across different conditions and strains (colors) [equivalent to the condition in Eq. (3)]. (B) Collapse of the functions  $f(\cdot)$ , defining the mechanism of division control in the discrete-time Langevin framework [Eq. (7)], obtained from experimental data in different conditions. The collapse of  $f(\cdot)$  is obtained, as predicted by Eq. (9). The function is evaluated as the conditional average of a cell's birth size (y axis) given the mother's birth size [x axis, see Eq. (7)]. Data from Ref. [1] refer to two strains, P5-ori (a BW25113 derivative strain) and MRR, grown on agarose pads in four nutrient conditions (Glc, CAA, RDM, and LB). Data from Ref. [19] (orange triangle) refer to MG1655 strain in a microfluidic device with LB as growth medium.

tive effect. Consequently, it is useful to introduce the quantity  $q = \log(x/x^*)$ , which measures logarithmic deviations in size. At this stage,  $x^*$  is an arbitrary scale, necessary to make the argument of the logarithm dimensionless. With this choice of parametrization the exponential growth maps into the linear relation  $q(t) = q_0 + \alpha t$ . The mechanism of division control can be equivalently specified in terms of  $q$ , by introducing the transition probability

$$\rho(q_f|q_0, \alpha) := x^* e^{\alpha q_f} p(x^* e^{\alpha q_f} | x^* e^{\alpha q_0}, \alpha). \quad (2)$$

The transition probability  $p(x_f|x_0, \alpha)$ , and therefore the mechanism of division control, determines the stationary distribution (if it exists) of initial sizes observed in a steadily dividing population or genealogy, denoted by  $p_b^*(x_0, \alpha)$ . This quantity can be obtained as the fixed point of the iteration defined by Eq. (1) (see Appendix A). Equivalently, the stationary distribution for interdivision times  $t_d$  derives from the mechanism of division control. A change of condition, e.g., nutrients or temperature, corresponds to a change of the growth rate  $\alpha$ , which in turn affects division control. Experimental data (Fig. 1) show that the stationary distributions of both initial size and interdivision time, measured in different conditions, collapse when rescaled by their means [1,2,5,18]. In the following, we will assume this scaling property, which implies some constraints on the control defined by  $p(x_f|x_0, \alpha)$  [1].

### B. Scaling laws for size and doubling-time distributions as a result of division control

This section derives explicitly the constraints on division control emerging from finite-size scaling [Fig. 1(a)], following Ref. [1]. In this case, the scaling condition reads

$$p_b^*(x_0|\alpha) = \frac{1}{\langle x_0 \rangle_\alpha} F\left(\frac{x_0}{\langle x_0 \rangle_\alpha}\right), \quad (3)$$

where we defined

$$\langle x_0 \rangle_\alpha := \int dx p_b^*(x|\alpha)x, \quad (4)$$

and the scaling function  $F(\cdot)$  describes the shape of the plot in Fig. 1(a), assumed to be invariant in all conditions [1,3]. A similar equation applies to the interdivision time distribution using  $\langle t_d \rangle_\alpha$ , i.e., the average interdivision time conditional on a growth rate. While the scaling properties for cell size have been observed in numerous studies [1–3,14], the scaling of interdivision time distribution is less widely studied and appears to vary across strains and conditions [1,4].

When the fluctuations of  $\alpha$  are neglected, the collapse can be explained as a result of division control but does not by itself isolate a specific mechanism [1]. Specifically, the observed collapse of the doubling-time and initial-size distributions implies that the conditional distribution  $p(x_f|x_0,\alpha)$  (for a growth condition with a given mean growth rate) has to collapse when both variables are rescaled by  $\langle x_0 \rangle_\alpha$ ,

$$p(x_f|x_0,\alpha) = \frac{1}{\langle x_0 \rangle_\alpha} G\left(\frac{x_f}{\langle x_0 \rangle_\alpha}, \frac{x_0}{\langle x_0 \rangle_\alpha}\right). \quad (5)$$

The function  $G(\cdot)$  represents the invariant functional form of the conditional probability distribution, analogous to  $F(\cdot)$  in Eq. (3). This calculation is discussed in Appendix A. Another important consequence of the scaling of the doubling time distributions is that the product  $\alpha \langle \tau \rangle_\alpha$  does not depend on the mean growth rate in a given condition  $\alpha$  (see Appendix A).

Finally, Eq. (5) implies that the division control depends on a single “internal” size scale, which, in turn, sets the value of  $\langle x_0 \rangle_\alpha$ . In conclusion, the joint universality in doubling time and size distributions can be explained by division control mechanisms based on a single size (volume) scale and  $1/\alpha$  as the unique time scale. While this condition does not imply any mechanism, it can be applied to different modeling frameworks, allowing model-independent predictions.

## III. RESULTS

### A. A unified modeling framework connects different descriptions of the growth-division process

This section shows the equivalence of the two theoretical formalisms discussed in the Introduction and explicitly provides the map connecting them. The result is a generic framework describing the cellular growth and division process, which can be compared with data to investigate the possible mechanism of division control. Specifically, the map leads us to a discrete-time equation, where the function describing the control is linked explicitly to a hazard rate. Finally, we show how the defining equation of this model is constrained by finite-size scaling of size and doubling-time distributions.

The continuous-time approach [5,7,16] supposes an underlying “decisional process” for cell division, which is entirely specified by the dependency of the division rate  $h_d$  on the measured dynamic parameters, such as instantaneous and initial cell size, added size, elapsed time from the previous cell division, and growth rate. The function  $h_d$  is analogous to a hazard rate in survival models. In particular, since division control is fully specified by  $p(x_f|x_0,\alpha)$  (as discussed in Ref. [1]), one has the relation

$$p(x|x_0,\alpha) = -\frac{d}{dx} \exp\left(\int_0^x ds h_d(s,x_0,\alpha)\right). \quad (6)$$

This hazard rate  $h_d$  can be inferred directly from data, as done in survival analysis in statistics [7]. Alternatively, a specific functional form can be assumed to test specific mechanisms [5,7]. Previous work [5] has shown that data are well reproduced by models where the division hazard rate depends on added size  $x - x_0$ , or by more complex “concerted control” models where the rate is allowed to depend on two variables, instantaneous size  $x$  and initial size  $x_0$  or elapsed time  $t$  (the latter two variables are essentially interchangeable since the distribution of elongation rates is generally quite peaked) [7]. This approach works very well in reproducing essentially all available observations. However, it leads to the problem of finding an interpretation of  $h_d$ , which is not simple. In fact, the link between  $h_d$  and “molecular” variables such as concentrations or absolute amounts of cell-cycle related proteins is still an open question. The other problem with the continuous-time approach is that  $h_d$  is a function, and, while it can be inferred directly from data, its parametrization may be far from obvious.

In order to comply with the empirical scaling properties of initial, final, and added size, and of interdivision time, the hazard rate function must collapse when the variables are rescaled by  $\langle x_0 \rangle_\alpha$  [see Eq. (5)],

$$h_d(x,x_0,\alpha) = \tilde{h}\left(\frac{x}{\langle x_0 \rangle_\alpha}, \frac{x_0}{\langle x_0 \rangle_\alpha}\right).$$

We now set out to discuss the connection of the hazard rate with the discrete-time formalism used in Refs. [5,6,10,12], which gives up the ambition of capturing doubling-time fluctuations, in order to obtain a clearer view of the dynamics of cell size. Importantly, this discrete-time approach makes an assumption for doubling-time fluctuations, defining the doubling time conditional to a certain initial size  $x_0$  as a random variable with a pre-defined mean  $\tau_0$  and “noise”  $\xi$ , accounting for intrinsic cell-to-cell variability. Hence,  $\tau = \tau_0 + \xi$ , where the distribution of the zero-mean variable  $\xi$  must be specified. One can verify *a posteriori* whether these assumptions are reasonable in the data. This choice leads to discrete-time Langevin equations for the initial size  $x_0(i)$  where  $i$  is the cell-cycle index,

$$x_0(i+1) = f(x_0(i),\alpha) + \eta(x_0(i),\alpha), \quad (7)$$

where the function  $f(\cdot)$  specifies cell division control, while  $\eta$  is a random noise with mean zero and arbitrary distribution, that, in principle, can depend on the size at generation  $i$ . In other words,  $x_0(i)$  is implied to be a random variable whose distribution depends on the value of  $\alpha$ . Note that there is no time discretization involved in obtaining Eq. (7), because

the generations are by definition discrete. The dependency of these two terms on the growth rate  $\alpha$  reflects the variation of the transition probability  $p(x|x_0, \alpha)$  with this parameter. In particular,  $f(\cdot)$  is given by

$$f(x_0, \alpha) = \frac{1}{2} \int_0^\infty dx p(x|x_0, \alpha)x. \quad (8)$$

Different forms of this function correspond to different kinds of controls on cell division. For instance a perfect sizer (division triggered by an absolute cell size  $x^*$ ) corresponds to  $f(x_0, \alpha) = x^*$  while an adder (division triggered by a noisy constant added size) is defined by  $f(x_0, \alpha) = (x_0 + \Delta)/2$ .

This function can be estimated from empirical data as  $f(x_0, \alpha) = \langle x_0(i+1) \rangle_{x_0(i)=x_0}$ , i.e., as the average size at birth of the daughter conditional on the size at birth of the mother. Figure 1(b) reports  $f(x_0/\langle x_0 \rangle_\alpha)/\langle x_0 \rangle_\alpha$  in empirical data for different growth conditions experiments and strains, showing the expected collapse.

The scaling relation in Eq. (5) imposes that  $f(x_0, \alpha)/\langle x_0 \rangle_\alpha$  is solely a function of the ratio  $x_0/\langle x_0 \rangle_\alpha$ . In particular, by combining Eq. (8) with Eq. (6), one can derive a simple relation between  $f$  and the hazard rate function, obtaining

$$\begin{aligned} & \frac{1}{\langle x_0 \rangle_\alpha} f(x_0, \alpha) \\ &= \frac{1}{2} \left\{ -\frac{x_0}{\langle x_0 \rangle} + \int_{\frac{x_0}{\langle x_0 \rangle}}^\infty dy \exp \left[ \int_y^\infty dz \tilde{h} \left( z, \frac{x_0}{\langle x_0 \rangle} \right) \right] \right\}. \quad (9) \end{aligned}$$

Considering the discrete framework, we can write an equivalent process for the initial logarithmic size  $q$ , which, after having imposed the constraints given by the scaling of the stationary distribution, reads

$$q_0(i+1) = \langle q_0 \rangle_\alpha + g(q_0(i) - \langle q_0 \rangle_\alpha) + \xi(q_0(i) - \langle q_0 \rangle_\alpha), \quad (10)$$

where  $\langle q_0 \rangle_\alpha$  is the average value of  $q_0$  given a growth rate  $\alpha$  and  $g(\cdot)$  specifies cell division control in log-space, analogous to  $f(\cdot)$  in Eq. (7) (see Appendix B). By imposing stationarity, we have that  $g(0) = 0$ . The noise term is again drawn from a zero-mean distribution. The function  $g(\cdot)$  can be estimated from empirical data by evaluating  $g(\Delta q) = \langle q_0(i+1) - \langle q_0 \rangle_\alpha \rangle_{q_0(i) - \langle q_0 \rangle_\alpha = \Delta q(i)}$ .

The two functions  $f(\cdot)$  and  $g(\cdot)$ , appearing in Eqs. (7) and (10) are interchangeable. Both expressions, once defined, correspond unequivocally to a specific division control mechanism. To obtain the hazard rate function, one must specify the distribution of the noise terms. Since in empirical data the initial and final size are approximately lognormally distributed [1], the steady-state distribution of  $q$  can be well approximated by a Gaussian. It is therefore reasonable to assume that the distribution of the noise is Gaussian itself,

$$\Delta q_0(i+1) = g[\Delta q_0(i)] + \sigma(\Delta q_0(i))\xi, \quad (11)$$

where  $\xi$  in this expression is a Gaussian random variable of zero mean and unit variance and  $\sigma(\cdot)$  a proper function of  $\Delta q_0(i) = q_0(i) - \langle q_0 \rangle_\alpha$ . Under this assumption, we obtain (see Appendix B)

$$h_d(x, x_0, \alpha) = \frac{1}{x} g_\sigma \left\{ \frac{\log(x/\langle x_0 \rangle_\alpha) - g[\log(x_0/\langle x_0 \rangle_\alpha)]}{\sqrt{2}\sigma[\log(x_0/\langle x_0 \rangle_\alpha)]} \right\},$$

where

$$g_\sigma(y) = \frac{1}{\sqrt{2\pi}\sigma} \frac{\exp(-y^2)}{1 - \text{Erf}(y)},$$

where  $\text{Erf}(\cdot)$  is the error function. However, note that, for unspecified  $g(\cdot)$  and  $\sigma(\cdot)$ , the stationary distribution of this process is not a Gaussian in general. Our direct calculation of the hazard rate  $h_d$  from the control function  $g$  links the discrete-time to the continuous-time formalism through a quantitative map. We will now focus on the parametrization defined in Eq. (11), showing how it can be reduced to a single relevant parameter, using a perturbative approach.

### B. A perturbative approach identifies the conditions for a steady-state size distribution (homeostasis)

This section extends the approach of Ref. [6] to nonlinear terms in cell-size control. A general perturbative expansion around the mean initial logarithmic cell size allows us to investigate the conditions for stationarity of cell sizes beyond linear order. As we will see, it is possible to assign a simple interpretation to the coefficients of the expansion and use them to estimate the division control mechanism. This kind of expansion is justified by empirical observations as follows. The collapse of the initial-size distributions implies that the standard deviation  $\sigma_{x_0}$  (which depends on the condition through the mean growth rate  $\alpha$ ), scales as  $\text{cv}_x \langle x_0 \rangle_\alpha$ , where  $\text{cv}_x$  is the coefficient of variation, and it is a constant independent of  $\alpha$ . In particular, the measured relative size fluctuations  $\text{cv}_x$  have an empirical value around 0.15, which is, as expected, approximately constant in different growth conditions [1,5]. Such value implies that the fluctuations of sizes around their mean are small, suggesting therefore to expand the control function over size fluctuations around their mean.

Instead of expanding the feedback control in powers of the ratio  $\sigma_{x_0}/\langle x_0 \rangle_\alpha$ , we will focus on logarithmic size, i.e., the previously introduced variable  $q$ . Starting from Eq. (11), one can expand  $g(\cdot)$  and  $\sigma(\cdot)$  around  $q_0(i) = \langle q_0 \rangle_\alpha$ . In this case, taking the first-order term of the expansion, we obtain

$$\Delta q_0(i+1) = (1 - \lambda)\Delta q_0(i) + \sigma\xi, \quad (12)$$

where  $\lambda = 1 - g'(0)$  and  $\sigma = \sigma(0)$ . The process defined by Eq. (11) is a discrete-time Langevin equation in a quadratic potential with stiffness  $\lambda$ , and thus has multiple continuous-time physical analogs (e.g., Brownian motion of colloids), as well as being widely used for time series modeling, e.g., in econophysics. Its exact solution is a Gaussian distribution of  $q_0$  with mean  $\langle q_0 \rangle_\alpha$  and variance  $\sigma_q^2 = \sigma^2/(\lambda(2 - \lambda))$  (see Appendix E and Ref. [6]), which corresponds to a lognormal distribution of  $x_0$ . This relation can be considered as a discrete-time version of a fluctuation-dissipation theorem, as it connects the fluctuations of cell size  $\sigma_q$  with the strength of the response  $\lambda$  to deviation of the size from the mean.

Figure 2 shows the scaling of  $g(\cdot)$ , comparing data and models. Equation (12) predicts a linear scaling with slope  $1 - \lambda$ , which is well in agreement with the data. A value of  $\lambda \sim 1/2$  captures the data, even if there are some variations among different conditions.

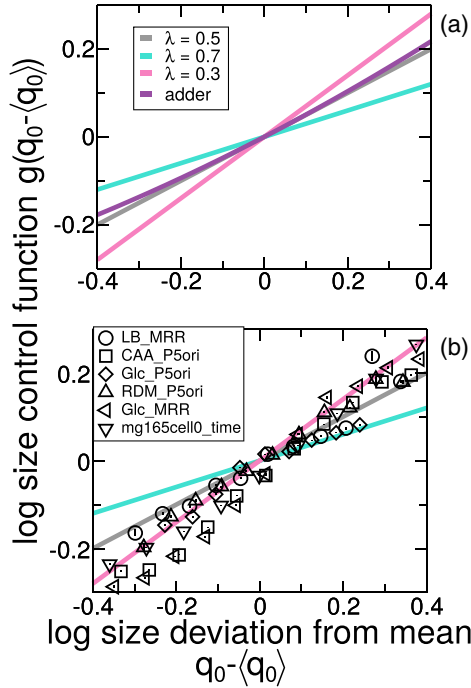


FIG. 2. Unified framework of division control and comparison with data. (a) Division control function  $g(\cdot)$ , for an adder model (purple solid line) and linearized models with different values of the control parameter  $\lambda$  (cyan, gray, and magenta solid lines). This function defines the control mechanism in the model [Eq. (10)]. The adder mechanism is near linear and closest to the linearization with  $\lambda \sim 0.5$  [6] (see Appendix F). (b) Comparison between data (symbols) and the linearized discrete-time Langevin framework, for different values of the single control parameter  $\lambda$ . The function  $g(\cdot)$  can be estimated as the conditional average of the log-size fluctuation of a cell, given the log-size fluctuation of the mother [see Eq. (10)]. The roughly linear scaling of the symbols suggests that the data are close to a simple “linear response” scenario, and the collapse across different strains and conditions confirms the results of Fig. 1. Values of  $\lambda$  around 1/2 well reproduce the data, but deviations are visible. Data (from Refs. [1] and [19]) refer to different strains of dividing *E. coli* cells grown in different conditions (see Fig. 1).

Equation (12) can be solved exactly. Starting from an arbitrary initial condition, we derive the distribution of sizes after any number of generations (Appendix E). In particular, it is possible to calculate how fluctuations of size are dampened in time. Starting at generation 0 with an initial size corresponding to  $q_0(0)$ , the expected size at birth after  $n$  generations is

$$\langle \Delta q_0(n) \rangle = \Delta q_0(0)(1 - \lambda)^n. \quad (13)$$

It is simple to see from this expression that, as expected, a steady-state size distribution (corresponding to “size homeostasis” in biological terms) is possible only if  $|1 - \lambda| < 1$  and that  $1 < \lambda < 2$  would lead to oscillatory sizes around the mean [12]. The role played by  $\lambda$  is therefore to set the correlation time scale, measured in generations.

Equations (12) and (13) show that size homeostasis is possible if  $|1 - \lambda| = |g'(0)| < 1$  [see Fig. 3(a)]. Note that this condition is necessary but not sufficient, as it only implies local stability of the deterministic solution (Appendix D). While values of  $\lambda$  between 1 and 2 guarantee homeostasis,

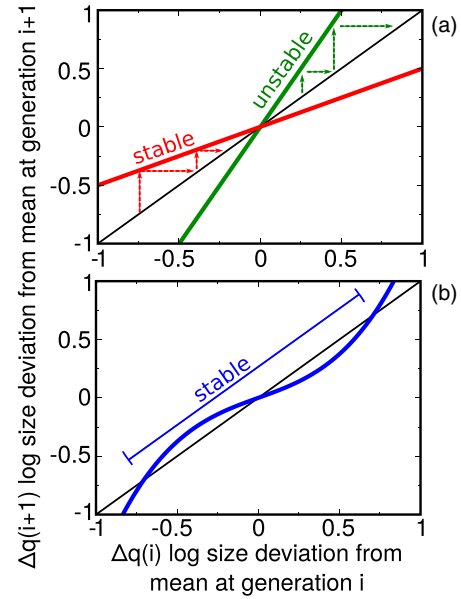


FIG. 3. Conditions for stability of the deterministic part of cell-size control. (a) Under linear control with negligible noise,  $\Delta q(i+1) = (1 - \lambda)\Delta q$ . If  $(1 - \lambda) > 1$  (green line) the system is unstable, while if  $(1 - \lambda) < 1$  (red line) the system is stable. The black line represents the marginally stable case  $\Delta q(i+1) = \Delta q$ . By using a similar argument (see Appendix D), it is possible to show that if  $|g(\Delta q)| < |\Delta q|$  for any  $|\Delta q|$ , then the system is globally stable. (b) In the more general case of a locally stable point the basin of attraction can be obtained as the set of values of  $\Delta q$  such that  $|g(\Delta q)| < |\Delta q|$ .

current data suggests that they are not biologically relevant. In particular, a value larger than one would correspond to overcorrection of size leading to oscillatory damping of size fluctuations (specifically, the genealogy of a cell born with a larger-than-average size would typically at some point generate progeny with smaller-than-average birth size, but closer to the mean, etc.) This behavior is not observed in experiments [5]. Hence, we restrict our analysis to the case  $0 < \lambda < 1$ .

Considering the next orders in the expansion, one can obtain precise criteria on the conditions leading to homeostasis. When only the deterministic part of Eq. (10) is considered (i.e.,  $\sigma = 0$ ), it is possible to show that the equilibrium is unique and globally stable if  $|g(\Delta q)| < |\Delta q|$ . If the noise is additive (i.e.,  $\sigma(\Delta q)$  is independent of  $\Delta q$ ), then global stability implies that the process is stable and always reaches a stationary distribution. On the other hand, what is relevant for homeostasis is that the basin of attraction determined by  $g(\cdot)$  is large enough compared to the typical fluctuations. The basin of attraction of the deterministic equation corresponds to the values of  $\Delta q$  such that  $|g(\Delta q)| < |\Delta q|$  see Fig 3(b)].

When the noise in Eq. (12) is not additive (i.e.,  $\sigma(\Delta q)$  depends on  $\Delta q$ ), a general condition is unknown. A perturbative approach gives conditions on the parameters of the expansion that determine homeostasis. For instance, considering the first orders in the expansion of  $\sigma[\Delta q_0(i)]$  around 0, we obtain that

the variance of initial logarithmic size distribution is finite only if  $\sigma'(0) < \sqrt{\lambda(2-\lambda)}$  (see Appendix E).

### C. Inequalities defining the relevant parameters given a set of experimental observations

This section derives general constraints on the estimates of relevant parameters given the number of observations through simple quantitative estimates. The calculations of the previous section indicate that  $\lambda$  is the most important parameter at play, together with another parameter defining the width of the noise (the intrinsic cell-to-cell variability). A further question is whether  $\lambda$  is effectively the only relevant control parameter. For instance, a sizer correspond to  $g(\Delta q) = 0$ . Both a sizer and a cubic control  $g(\Delta q) = \Delta q^3$  would correspond to  $\lambda = 1$ . Therefore, mechanism can be associated to a value of  $\lambda$ , while the converse is not true. In order to address this question, one has to consider higher-order terms in the expansion, and ask when those terms play a role, and whether they can be identified from data, given the number of available observations.

The expansion around  $\langle x_0 \rangle$  [Eq. (11)] is effective as long as the fluctuations of size are sufficiently small. In order to estimate precisely the regime where the approximation is valid, we include the second order in the expansion,

$$q_0(i+1) = \langle q_0 \rangle_\alpha + (1-\lambda)(q_0(i) - \langle q_0 \rangle_\alpha) + \gamma \frac{(q_0(i) - \langle q_0 \rangle_\alpha)^2}{2} + \sigma \xi, \quad (14)$$

where  $\gamma = g''(0)$  is the second derivative of the control function. The parameter  $\gamma$  characterizes in further detail the mechanism of division control and can be used to distinguish between different control functions that have the same value of  $\lambda$ . Importantly, two different functions  $g(\Delta q)$  describing completely different mechanisms of size control could, in fact, have the same value of  $g'(0) = 1 - \lambda$ , and therefore produce the same predictions to linear order. We provide one specific example in the next section. At a given fixed  $\lambda$ , estimating  $\gamma$  may help in distinguishing different mechanisms.

The quadratic term proportional to  $\gamma$  destabilizes the deterministic dynamics, and, using the condition  $|g(\Delta q)| < |\Delta q|$  to determine the basin of attraction of the deterministic dynamics, one obtains that fluctuations are dampened if  $|\Delta q| < 2\lambda/\gamma$ , imposing a limit on the possible values of  $\gamma$  based on the extent of the experimentally observed fluctuations.

We set out to evaluate the difference between this process and the one defined by Eq. (12). The quadratic term is measurable from stochastic trajectories if it is sufficiently large compared to stochastic fluctuations. Thus, we evaluate the distribution of  $q_0(i+1) - \langle q_0 \rangle_\alpha - (1-\lambda)(q_0(i) - \langle q_0 \rangle_\alpha)$  and ask whether, for given sample size and value of  $q_0(i)$ , its mean is significantly different from zero or not.

The error on the mean is given by the standard deviation divided by the square root of the sample size. Hence, the quadratic term is detectable if

$$\frac{\sigma}{\sqrt{T(q)}} < \gamma \frac{(q_0(i) - \langle q_0 \rangle_\alpha)^2}{2}, \quad (15)$$

where  $T(q)$  is the number of cells with initial size  $q$ . Since the distribution of  $q$  is approximately Gaussian (in the limit of

$\gamma \approx 0$ ), the number of cells with initial size in a bin of width  $\delta q$  around  $q$  is estimated by

$$T(q) = N \frac{\exp\left(-\frac{(q - \langle q_0 \rangle_\alpha)^2}{2\sigma_q^2}\right)}{\sqrt{2\pi\sigma_q^2}} \delta q, \quad (16)$$

where  $N$  is the total number of cells. The bin size  $\delta q$  must be smaller than the standard deviation of the distribution, and we can parameterize it by defining  $\delta q = \epsilon \sigma_q$ . The quantity  $1/\epsilon$  represents therefore the number of bins per standard deviation. Choosing  $\epsilon$  too large increases the sampling at the cost of introducing some error, by grouping together cells with different sizes. The constraint on the total number of cells measured in order to recognize higher-order terms then reads

$$N > \frac{\sqrt{2\pi}\sigma_q \sigma^2}{\epsilon \sigma_q} \frac{4}{\gamma^2 (q - \langle q_0 \rangle_\alpha)^4} \exp\left(\frac{(q - \langle q_0 \rangle_\alpha)^2}{2\sigma_q^2}\right). \quad (17)$$

The above expression reveals an important tradeoff [illustrated in Fig. 4(a)]. Choosing values of  $q$  close to the mean will correspond to larger sample sizes but also causes the effect to be measured to be very small, and increasingly close to the experimental resolution. Conversely, choosing a value of  $q$  very far from the mean corresponds to larger effects, but their detection would be limited by smaller sample sizes. The optimal choice of  $q$  to evaluate deviations from a division control based on linear response in the data minimizes the left side of Eq. (17). We have, therefore,

$$N > \frac{1}{\epsilon} \frac{1}{\sigma_q^2 \gamma^2} \lambda(2-\lambda) \min_t \frac{4\sqrt{2\pi}}{t^4} \exp\left(\frac{t^2}{2}\right) \approx \frac{4.6}{\epsilon} \lambda(2-\lambda) \frac{1}{\gamma^2 \sigma_q^2} = \frac{4.6}{\epsilon} \lambda(2-\lambda) \frac{1}{\gamma^2 \log(1 + cv_x^2)}, \quad (18)$$

where 4.6 is approximately the minimum value of the function of  $t$ , and  $cv_x$  is the coefficient of variation of the distribution of  $x$ . We shall assume  $\epsilon = 0.1$ ; we know that  $\epsilon$  must be a small number—if it were not, one would have to consider other sources of errors. For instance, in Ref. [5] the initial size was binned using a bin size equal to about  $0.06 \mu\text{m}$ . Since the standard deviation varied between  $0.4 \mu\text{m}$  and  $1.3 \mu\text{m}$  across conditions, the bins used in Ref. [5] correspond to values of  $\epsilon$  between 0.03 and 0.15. Assuming that  $\lambda(2-\lambda)$  is a number of order 1 (which should be the case if  $\lambda \approx 1/2$ ), we obtain  $N \approx 1.5 \times 10^3 / \gamma^2$ . Note that the factor  $\gamma^2$ , which is set by the second derivative of  $g(\cdot)$  and is therefore mechanism-dependent, plays a very important role here, as, at fixed  $\lambda$ , its value sets the scale at which specific mechanisms can be distinguished.

## IV. INTERPRETABILITY OF MECHANISMS OF DIVISION CONTROL

This section relates the perturbative expansion of division control to specific examples of mechanisms discussed in the literature. We treat the case of the concerted control mechanism, and we also specifically consider the case of the adder, proposed as a mechanism of division control across

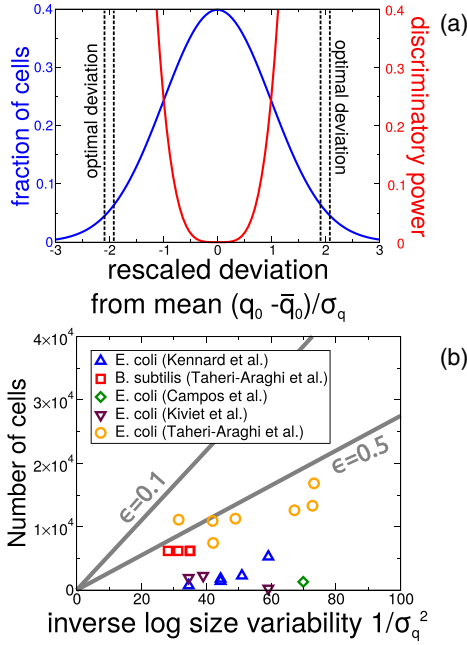


FIG. 4. Estimated threshold for detection of non-linear contributions to cell-division control. (a) The tradeoff between two competitive terms determines an optimal cell fluctuation to identify cell-size control. On one hand, large deviations of size from the average correspond to stronger corrections and make differences between mechanisms more detectable. On the other hand, fewer cells have large fluctuations, reducing the statistical power and increasing the sampling noise. The optimal fluctuation value is the one that minimizes the error on the inferred cell-size control mechanism. Our calculations [Eq. (17)] show that when fluctuations are rescaled by the variance of the distribution, the optimal value is independent of the cell-size control mechanism, as shown in the plots of the contributions described by Eqs. (15) (discriminatory power, green line) and (17) (sampling level, blue line). (b) Number of cells measured vs. the inverse of  $\sigma_q$  [which is related to the coefficient of variation via equation (E15)] for the available datasets (points). The two gray lines represent the threshold of detectability obtained using Eq. (21), with  $\epsilon = 0.1$  or  $\epsilon = 0.5$ . All the available data sets lay below the threshold of detectability (with one exception in the case of  $\epsilon = 0.5$ ), suggesting that a linear model of division control is sufficient to describe these data.

different conditions, obtaining the parameters of its expansion. More specifically, we compare the adder model (characterized by  $\lambda = 1/2$  and  $\gamma = 1/4$ ) with a model with the same linear control strength, which can be interpreted as a combination of a timing and a sizing mechanism (characterized by  $\lambda = 1/2$  and  $\gamma = 0$ ).

#### A. The “concerted control” mechanism

Equation (12) provides a generic description of division control for small fluctuations. When it is interpreted in terms of mechanisms of control, it corresponds to the simplest “concerted control” model, i.e., to a mix of sizer and timer behavior (as in the framework of Ref. [6]) that has been shown

to be in agreement with data [7]. Specifically, since the time between divisions is  $(q_f - q_0)/\alpha$ , setting  $\tau = \tau_0 + \xi$ , where  $\xi$  are Gaussian, independent, zero-mean random variables, one obtains

$$\tau_0 = (1 - \lambda) \frac{\log 2}{\alpha} + \frac{\lambda}{\alpha} \log \frac{x_\alpha^*}{x_0}. \quad (19)$$

The above equation can be interpreted as implementing the control on cell division as a mixture of timer and sizer behavior [6]. Indeed, the doubling time is set by a convex combination with mixing parameter  $\lambda$  of a fixed time (set by the inverse mean growth rate  $1/\alpha$ ) and a perfect sizer (set by a limit threshold size  $x^*$  for cell division). A pure sizer model, recovered for  $\lambda = 1$ , would set this conditional interdivision time as  $\tau_0 = \frac{\lambda}{\alpha} \log \frac{x_\alpha^*}{x_0}$ , while a pure timer model ( $\lambda = 0$ ) defines  $\tau_0$  as  $\log 2/\alpha$  [6]. It is straightforward to show that, in the small noise limit,  $x_\alpha^* = 2\langle x_0 \rangle = \langle x_f \rangle$ . The concerted control is a consequence of the combination of these two decision processes, set by the parameter  $\lambda$ . As shown in Sec. III B this process leads to stationary distributions of sizes if  $0 < \lambda < 2$ . Our previous calculations show that such an effective cell-cycle model (equivalent to the approach introduced in Ref. [6]) can be characterized as the autoregressive model giving a discrete-time Langevin equation with harmonic potential. As discussed above, this has a number of consequences, including a strict relation between the noise in  $\tau$  and in  $\log x_0$ , and the fact that the characteristic times (in generations) for damping of fluctuations and perturbations (fluctuation-dissipation theorem) is  $1/\lambda$ .

As shown in Sec. III B and previously suggested in Ref. [6], the linear dependency of  $\tau_0$  on  $\log x_0$  can be seen as a first-order approximation of a generic function relating the doubling time to the initial size. Thus, nearly all models where one of the two terms in Eq. (19) is not strictly null are expected to behave similarly to this concerted control model as long as the probed initial cell sizes  $x_0$  are close to their mean  $\langle x_0 \rangle_\alpha$ , or equivalently as long as the noise in  $\alpha\tau$  is small.

#### B. The constant added size mechanism

We now consider the constant added size mechanism, with the goal of comparing it with its linearization ( $\lambda = 1/2$  and  $\gamma = 0$ ). The control part is defined by the function

$$g(q - \langle q_0 \rangle_\alpha) = q - \langle q_0 \rangle_\alpha + \int dz F(z) \log \left( \frac{c + ze^{(q_0)_\alpha - q_0}}{2} \right), \quad (20)$$

where  $c$  is determined by imposing  $g(0) = 0$  and  $F(\cdot)$  is the probability distribution of the relative fluctuations of the added size around its mean (Appendix F). Expansion of this function gives  $\lambda = 1/2$ , consistently with previous results [6], which guarantees stationarity of the process. The second-order term gives  $\gamma = 1/4$ . Having determined all the parameters, one can evaluate Eq. (17), obtaining

$$N > \frac{55}{\epsilon} \frac{1}{\log(1 + cv_x^2)}. \quad (21)$$

This estimate approximates the full adder model with a second-order model with  $\lambda = 1/2$  and  $\gamma = 1/4$ . It sets a threshold on the number of cells that one needs to measure

in order to achieve enough statistical power to observe nonlinearity in the size control function  $g(\Delta q)$ . Figure 4(b) compares the current available datasets with the estimated threshold, for different choices of  $\epsilon$ , showing the while some data set may be close to the requirements, all of them are below. This suggests that, for most available experimental datasets, a linearization of  $g(\Delta q)$  should be sufficient to describe all the main observations. The choice of  $\epsilon$  is somewhat arbitrary and is justified as long as it is much smaller than 1. Our results are robust up to to the (extreme) choice of  $\epsilon \simeq 0.5$ .

To make the result on the estimated threshold for detectability more concrete, we consider explicitly the case of the adder mechanism and its distinguishability from the linearized model [Eq. (12)]. By definition, the adder mechanism predicts that the conditional mean (and distribution) of the added size, given initial size, is independent on initial size. While the first-order expansion of the framework defined here with  $\lambda = 1/2$  (and analogously for the model in Ref. [6]) does not follow this functional trend (i.e., the next orders in the expansion are different), it shows a very small difference with the adder model in the empirical range of sizes, which might not be discerned with the sampling of available empirical data.

In order to further support this point, we employed direct numerical simulations at different sample sizes (mimicking experimental sampling levels). As explained above, the most complete information on the process is the transition probability  $p(x_f|x_0, \alpha)$ . For an adder, this probability depends only on the difference  $x_f - x_0$ , i.e.,  $p(x_f - x_0|x_0)$  is independent of  $x_0$ , or, in other words, the conditional distribution of added size given initial size does not depend on the initial size. The fact that  $p(x_f - x_0|x_0)$ , obtained for different  $x_0$ , collapses has been interpreted in Ref. [5] as evidence in favor of the adder mechanism of division control. To gain more insight into this conclusion, we simulated the first-order process [Eq. (14)] with  $\lambda = 1/2$ . Figure 5 reports the binned histograms of rescaled added sizes for cells with different initial sizes, and using similar bin sizes as in Ref. [5], the very good collapse shows that the difference between the probability distributions is barely detectable at the available level of sampling.

We quantified the error on the collapse measuring the average  $L_2$  distance between all the pairs of curves plotted in Fig. 5(a). This error, measured for different values of  $\lambda$  and different sample sizes, can be compared with the expected error due to fluctuations in the adder (also estimated as the average  $L_2$  distance between the conditional distributions of added size given the birth size). As expected, Fig. 5(b) shows that the error is minimal when  $\lambda = 0.5$ . Interestingly, the measured error does not depend on the sample size, while the expected error from the adder model decreases as the number of measured cells increases. Figure 5(b) shows that the two error measures become comparable when  $N$  is between 10 000 and 20 000, which is around the same order of magnitude of the number of cells measured in Ref. [5]. The plot shows that with  $N = 10\,000$  cells, the collapse discriminates between different values of  $\lambda$ , but cannot easily discriminate if higher-order terms are relevant (and in particular whether the data support the full adder model or just a linear model). However, the test presented in Fig. 5(b) may work if existing data are pooled together.

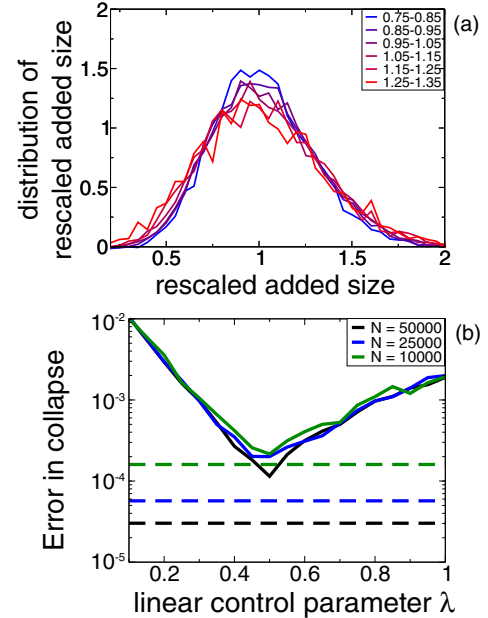


FIG. 5. Detectability of adder mechanism from simulated models and direct test. Plotted data refer to simulations of models with realistic parameters and sampling of cells. (a) Conditional distribution of added size given initial size for different initial sizes (different colors) obtained with the linearized model [Eq. (14)] with  $\lambda = 0.5$ . The linearized model reproduces visually the collapse expected for the adder model. The simulations consider conservatively a coefficient of variation of the added size equal to 0.3 (which is larger than the observed values [5]) and a total number of cells  $N = 50000$ . (b) Error test on the collapse of the distribution of added size (estimated as the average  $L_2$  distance between all pairs of curves) for different values of  $N$  and  $\lambda$ . The horizontal dashed lines represent the expected error in collapse in the adder model due to fluctuations. For these parameter values, the error in the collapse for the model with  $\lambda = 0.5$  starts to be relevant when  $N \sim 10000$ . This test may be applied to empirical data: in order for an adder to be detectable, the solid line should stay above the dashed line.

## V. DISCUSSION

Our approach provides a map between an autoregressive discrete-time formalism of cell division control and a continuous-time description based on hazard-rate functions, showing the impact on both formalisms of the observed scaling behavior of cell sizes and doubling times. This map connects the approaches used in Refs. [1,7,14] with those of Refs. [6,8,10], and leads us to propose a unified framework (with discrete-time Langevin equations) embracing both formalisms to develop and explore effective models of cell division. The framework has the additional advantage of showing how parameter-poor models describing different kinds of cell division control are possible for multiple mechanisms.

The use of a discrete-time Langevin formalism for the logarithmic size leads to a simple physical analogy with a fluctuating system and guides the interpretation of the model parameters. The same formalism also enables the applicability of familiar concepts in the statistical physics of fluctuating systems, such as correlation, response, and fluctuation-dissipation



relations. We anticipate that such concepts will become useful for future studies of dividing cells in fluctuating environments. An important extension of the present framework should incorporate growth fluctuations. Recent studies [2,15] show clear indications that the assumption of constant growth rate  $\alpha$  is an oversimplification, and that size homeostasis needs to be understood by addressing contributions from both growth and cell division. Possibly the most relevant result from recent experimental work is the existence of a mechanism governed by a single size scale. The “microscopic” origin of this length scale is a relevant question that is not solved by any of the mechanisms proposed in the literature.

The most important results of our study are connected to the perturbative expansion of the model, which shows how the unified framework defined here can lead to similar equations to the ones introduced in Ref. [6], with the advantage of elucidating the direct link with the hazard rate function. The main difference is found in the dependence of the noise term on the growth rate  $\alpha$ . In the setting defined here, there is no dependency, while Ref. [6] assumes a dependency (see Sec. IV). The importance of this difference is that in the model used here the distributions of division times collapse as observed experimentally. The possibility to carry out this expansion makes it possible to analyze which terms are important to describe the experimental data. In other words, the framework presented here can be used to compute the next orders of the expansion, and to study hierarchically the mechanisms leading to homeostasis. While these higher-order terms makes the discrete Langevin equation untractable, it has been recently shown [20], also in the context of cell-size control [21], that discrete Langevin equations can be approximated by analytically tractable continuous Langeving equations.

This perturbative approach also leads to relevant insight on the ability to distinguish different control mechanisms from data. Overall, our results indicate that, for most if not all the currently available experimental data sets, a linearization of the control function  $g(\Delta q)$  should be sufficient to describe the main observations. Thus, for practical purposes, the linear response regime appears to be an excellent description.

The evaluation of higher-order “anharmonic” terms is necessary to pinpoint precise mechanisms, and depends on an important tradeoff. On the one hand, testing sizes that deviate a great deal from the average will show increasingly detectable differences between mechanisms. On the other hand, fewer cells have large fluctuations, reducing the statistical power and increasing the sampling noise of such measurements. We define the optimal fluctuation value as the one that minimizes the error on the inferred cell-size control mechanism. Importantly, the calculations show that when fluctuations are rescaled by the variance of the distribution, this optimal value is independent of the cell-size control mechanism. Thus, we expect that different division control functions should be distinguishable without requiring an ad hoc number of observations. Comparing the detection threshold with the number of observations made in available studies, we find that at the current sampling levels, it is not simple to distinguish the underlying mechanism of division control from linear response, even with sophisticated tests such as the collapse of the conditional distribution of added

size. In fact, our results support the view that, given the current experimental resolution, the adder model should be in most cases indistinguishable from its linearization. This poses important caveats on the interpretation of measured trends as “microscopic” mechanisms of size control.

We propose the method developed in Fig. 5(b) as an effective way, applicable to empirical data, to test for deviations from the behavior of the linearized model, which should work with sampling levels that can be attained experimentally with existing approaches. We are currently working on extending this approach using Bayesian statistics and producing reliable statistical estimators of the relevant parameters ( $\lambda$  and  $\gamma$ ) of the division control function.

## ACKNOWLEDGMENTS

This work was supported by the International Human Frontier Science Program Organization (HFSP), Grant No. RGY0070/2014.

## APPENDIX A: COLLAPSE OF THE INITIAL SIZE AND DOUBLING TIME DISTRIBUTIONS

This section discusses the implications of the observed collapse of doubling time and initial size distributions on the division rate function  $h_d$ .

The initial size distribution  $p_b^*(x_0|\alpha)$  in a given condition characterized by mean growth rate  $\alpha$  is given by

$$p_b^*(x_0|\alpha) = 2 \int_0^{+\infty} dx'_0 \theta(2x_0 - x'_0) p_b^*(x'_0|\alpha) p(2x_0|x'_0, \alpha), \quad (\text{A1})$$

where  $\theta(\cdot)$  is the Heaviside function.

The collapse of initial sizes implies that  $p_b^*(y|\alpha) = p_b^*(y)$  is independent of  $\alpha$ , with  $y = x_0/\langle x_0 \rangle_\alpha$ . Imposing this condition in Eq. (A1) implies that

$$p_b^*(y) = 2 \int_0^\infty dy' \theta(2y - y') p_b^*(y') p(2y|y', \alpha). \quad (\text{A2})$$

This equation immediately shows that a necessary and sufficient condition for the collapse is that the conditioned distribution does not depend on  $\alpha$ , i.e.,

$$p(y_f|y_0, \alpha) = \tilde{p}(y_f|y_0). \quad (\text{A3})$$

The division rate function  $h_d(x, x_0)$  is related to the above conditioned distribution by the following equation:

$$\begin{aligned} h_d(x, x_0, \alpha) &= -\log \int_{x_0}^x dz p(z|x_0, \alpha) \\ &= -\log \int_{x_0/\langle x_0 \rangle_\alpha}^{x/\langle x_0 \rangle_\alpha} dy \tilde{p}(y|x_0/\langle x_0 \rangle_\alpha). \end{aligned} \quad (\text{A4})$$

The collapse of initial size distributions is therefore equivalent to collapse of the division hazard rate when rescaled by mean initial sizes, i.e.,

$$h_d(x, x_0, \alpha) = \tilde{h} \left( \frac{x}{\langle x_0 \rangle_\alpha}, \frac{x_0}{\langle x_0 \rangle_\alpha} \right). \quad (\text{A5})$$

We now consider the collapse of doubling-time distributions. The conditioned distribution for final sizes can be written as

$$p(x_f|x_0, \alpha) = \tilde{p}\left(\frac{x_f}{\langle x_0 \rangle_\alpha}, \frac{x_0}{\langle x_0 \rangle_\alpha}\right) = \hat{p}\left(\frac{x_f}{x_0}, \frac{x_0}{\langle x_0 \rangle_\alpha}\right). \quad (\text{A6})$$

Since  $\log(x_f/x_0) = \alpha\tau$ , the above expression, combined with Eq. (A3), implies the following condition for the collapse of the distribution of doubling times:

$$\alpha p_i^*(\tau|x_0, \alpha) = \hat{p}\left(\alpha\tau, \frac{x_0}{\langle x_0 \rangle_\alpha}\right). \quad (\text{A7})$$

The joint collapse of the distribution of doubling times and initial cell sizes impose conditions on size control. In other words, a control of cell division obeying to the condition described in Eqs. (A3) and (A7) will generate universal size and doubling-time distribution.

In particular, a necessary condition for this to hold is that the product of the mean doubling time and the mean growth rate  $\alpha\langle\tau\rangle_\alpha$  does not depend on the mean growth rate in a given condition  $\alpha$ .

## APPENDIX B: FULL DERIVATION OF THE MAPPING BETWEEN DISCRETE-TIME LANGEVIN EQUATION AND DIVISION HAZARD RATE

This section shows in full generality the mapping between a discrete-time Langevin formalism and the corresponding division hazard rate.

The discrete equation for the logarithm of the initial size is

$$q(i+1) = \langle q_0 \rangle_\alpha + g(q(i) - \langle q_0 \rangle_\alpha) + \eta(q(i) - \langle q_0 \rangle_\alpha), \quad (\text{B1})$$

where

$$g(q_0 - \langle q_0 \rangle_\alpha) = \int dq \rho(q|q_0, \alpha) q - \langle q_0 \rangle_\alpha - \log(2), \quad (\text{B2})$$

while  $\eta$  is a random variable with distribution  $\rho(q - \langle q_0 \rangle_\alpha - \log(2) - g(q_0 - \langle q_0 \rangle_\alpha)|q_0, \alpha)$ .

Using Eq. (2) we can write

$$g(q_0 - \langle q_0 \rangle_\alpha) = \int_{x^*e^{q_0}}^{\infty} dx p(x|x^*e^{q_0}, \alpha) \log\left(\frac{x}{x^*}\right) - \langle q_0 \rangle_\alpha - \log(2), \quad (\text{B3})$$

and introducing Eq. (6) we obtain

$$\begin{aligned} & g(q_0 - \langle q_0 \rangle_\alpha) \\ &= \int_{x^*e^{q_0}}^{\infty} dx \left\{ -\frac{d}{dx} \exp\left[\int_{x_0}^x ds h_d(s, x_0, \alpha)\right] \right\} \\ & \quad \times \log\left(\frac{x}{x^*}\right) - \langle q_0 \rangle_\alpha - \log(2) \\ &= -\exp\left[\int_{x^*e^{q_0}}^x ds h_d(s, x_0, \alpha) \log\left(\frac{x}{x^*}\right)\right] \Big|_{x=x^*e^{q_0}}^{\infty} \\ & \quad + \int_{x^*e^{q_0}}^{\infty} dx \exp\left[\int_{x_0}^x ds h_d(s, x_0, \alpha)\right] \frac{1}{x} \\ & \quad - \langle q_0 \rangle_\alpha - \log(2), \end{aligned} \quad (\text{B4})$$

and the final expression reads

$$\begin{aligned} & g(q_0 - \langle q_0 \rangle_\alpha) \\ &= q_0 - \langle q_0 \rangle_\alpha + \int_{x^*e^{q_0}}^{\infty} \frac{dx}{x} \\ & \quad \times \exp\left[\int_{x^*e^{q_0}}^x ds h_d(s, x^*e^{q_0}, \alpha)\right] - \log(2). \end{aligned} \quad (\text{B5})$$

The hazard rate function cannot be derived from Eq. (B1) without specifying the form of the noise  $\eta(q(i) - \langle q_0 \rangle_\alpha)$ . Assuming that the distribution of the noise is Gaussian, we obtain

$$\Delta q_0(i+1) = g[\Delta q_0(i)] + \sigma[\Delta q_0(i)]\xi, \quad (\text{B6})$$

where  $\xi$  in this expression is a Gaussian random variable of zero mean and unit variance and  $\sigma(\cdot)$  a proper function of  $\Delta q_0(i) = q_0(i) - \langle q_0 \rangle_\alpha$ . The division probability at log-size  $q$  given and initial log-size  $q_0$  is therefore

$$\begin{aligned} \rho(q|q_0, \alpha) &= \frac{1}{\sqrt{2\pi}\sigma(q_0 - \langle q_0 \rangle_\alpha)} \\ & \quad \times \exp\left[-\frac{(q - \langle q_0 \rangle_\alpha - g(q_0 - \langle q_0 \rangle_\alpha))^2}{2\sigma(q_0 - \langle q_0 \rangle_\alpha)^2}\right], \end{aligned} \quad (\text{B7})$$

using the fact that

$$h_d(x, x_0, \alpha) = -\frac{d}{dx} \log P_0(x, x_0, \alpha),$$

where

$$P_0(x, x_0, \alpha) = \int_{x_0}^x dz p(z|x_0, \alpha),$$

we obtain that

$$\begin{aligned} h_d(x, x_0, \alpha) &= -\frac{d}{dx} \log \int_{x_0}^x dy p(y|x_0, \alpha) = -\frac{dq}{dx} \frac{d}{dq} \log \int_{q_0}^q dp \rho(p|q_0, \alpha) \\ &= -\frac{1}{x} \frac{d}{dq} \log \frac{1}{2} \left\{ 1 - \text{Erf}\left[\frac{q - \langle q_0 \rangle_\alpha - g(q_0 - \langle q_0 \rangle_\alpha)}{\sqrt{2}\sigma(q_0 - \langle q_0 \rangle_\alpha)}\right] \right\} \Big|_{q_0=\log(x_0/x^*)}^{q=\log(x/x^*)}, \end{aligned} \quad (\text{B8})$$

where the error function Erf is defined as

$$\text{Erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x dt e^{-t^2}.$$

We finally obtain

$$h_d(x, x_0, \alpha) = \frac{1}{x} \left\{ g_\sigma \left[ \frac{q - \langle q_0 \rangle_\alpha - g(q_0 - \langle q_0 \rangle_\alpha)}{\sqrt{2}\sigma(q_0 - \langle q_0 \rangle_\alpha)} \right] \right\}_{q=\log(x/x^*)}^{q_0=\log(x_0/x^*)},$$

where

$$g_\sigma(y) = \frac{1}{\sqrt{2\pi}\sigma} \frac{\exp(-y^2)}{1 - \text{Erf}(y)}.$$

In the next session we show the explicit calculation in the case of linear  $g(\cdot)$  and constant  $\sigma(\cdot)$ .

### APPENDIX C: DIVISION RATE FOR LINEARIZED MODEL

As explained in the main text, one can linearize Eq. (B1) around its equilibrium, obtaining

$$q_0(i+1) = \langle q_0 \rangle_\alpha + (1-\lambda)[q_0(i) - \langle q_0 \rangle_\alpha] + \sigma\xi. \quad (\text{C1})$$

In this case, the Eq. (B8) in Appendix B reads

$$h_d(x, x_0, \alpha) = \frac{1}{\alpha x} g_\sigma \left( \frac{1}{\sqrt{2}\sigma\alpha} \log \frac{x}{x_0^{(1-\lambda)} x^{*\lambda}} \right).$$

### APPENDIX D: CONDITIONS FOR STATIONARITY

This section discusses under which conditions Eq. (10) admits a well-defined stationary size distribution. The scaling of stationary distribution is the only assumption that we used to derive Eq. (10). Any division control must, by definition, regulate sizes and stabilize size fluctuations. A necessary condition is therefore that the deterministic equation corresponding to Eq. (10) has a fixed point and that fixed point is (at least) locally asymptotically stable.

The fixed point of the deterministic part of Eq. (10) is a solution of the equation  $q^* = \langle q_0 \rangle_\alpha + g(q^* - \langle q_0 \rangle_\alpha)$ . This fixed point is asymptotically locally stable if and only if

$$\left| \left( \frac{dg}{dq} \right)_{q=q^*} \right| = |1 - \lambda| < 1. \quad (\text{D1})$$

This condition is necessary, but not sufficient to guarantee stationarity of the process.

More generally, the deterministic part of Eq. (10) implies that the equilibrium is unique and globally stable if and only if  $|g(\Delta q)| < |\Delta q|$  for any  $\Delta q$ , where  $\Delta q_0(i) = q_0(i) - \langle q_0 \rangle_\alpha$ . In particular, if the function  $g(\cdot)$  is monotonic and has only one fixed point that is locally stable, then that fixed point is globally stable. This property sets a minimal condition that division control has to fulfill to guaranty stationarity of cell-size distribution. If the fixed point was not globally stable, then a large enough fluctuation would not be corrected by feedback control. When the stochasticity is taken into account, since the noise in Eq. (10) can be multiplicative, global stability does not guarantee stationarity of the size distribution in general. On the other hand, the requirement of having a stationary distribution is not necessarily biologically relevant and is not needed to have homeostasis. The basin of attraction of the fixed point  $\Delta q = 0$  is determined by the values of  $\Delta q$  such that  $|g(\Delta q)| < |\Delta q|$ . What is relevant for homeostasis is that the

basin of attraction determined by  $g(\cdot)$  is large enough compared to the typical fluctuations. This would guarantee that most of the cells are able to control fluctuation of their size, and loss of control is a rare event.

To characterize the effect of multiplicative noise on the existence of a stationary size distribution, we study a general expansion of  $\sigma[\Delta q_0(i)]$  in Eq. (11),

$$\Delta q_0(i+1) = (1-\lambda)\Delta q_0(i) + (\sigma + \beta\Delta q_0(i))\xi, \quad (\text{D2})$$

where  $\beta = \sigma'(0)$ . If  $\beta = 0$ , then this process guarantees homeostasis for any  $|1-\lambda| < 1$ . In the case  $\beta \neq 0$ , one can write recursive equations for the moments of the distribution of sizes, given an arbitrary initial condition. The equation for the mean corresponds, obviously, to the deterministic equation. The recursive equation for the variance reads

$$\begin{aligned} \langle \Delta q_0(i+1)^2 \rangle &= (1-\lambda)^2 \langle \Delta q_0(i)^2 \rangle + \sigma^2 + \beta^2 \langle \Delta q_0(i)^2 \rangle \\ &\quad + 2\sigma\beta \langle \Delta q_0(i) \rangle. \end{aligned} \quad (\text{D3})$$

Starting from a deterministic initial condition  $\Delta q_0(0)$ , using the result of Eq. (13) and solving the recursive equation one can obtain the time evolution of the variance. In the case of  $\Delta q_0(0) = 0$  it reads

$$\langle (q_0(n) - \langle q_0 \rangle_\alpha)^2 \rangle = \sigma \frac{1 - ((1-\lambda)^2 + \beta^2)^n}{\lambda(2-\lambda) - \beta^2}. \quad (\text{D4})$$

It is simple to see that the variance converges to a constant if and only if  $(1-\lambda)^2 + \beta^2 < 1$ , i.e.,  $\beta^2 < \lambda(2-\lambda)$ . In case of multiplicative noise, the stationarity of the size distribution depends, in a nontrivial concerted way, from both the strength of control and the magnitude of the noise.

### APPENDIX E: SOLUTION OF THE LINEARIZED MODEL

In this section we discuss the solution of the linearized model defined by the discrete Langevin equation

$$\Delta q_0(i+1) = (1-\lambda)\Delta q_0(i) + \sigma\xi. \quad (\text{E1})$$

This equation defines the distribution of initial size at generation  $i+1$  given the one of generation  $i$  as

$$\rho^{i+1}(\Delta q) = \int_{-\infty}^{\infty} d\Delta q' \rho^i(\Delta q') \varrho \left( \frac{\Delta q - (1-\lambda)\Delta q'}{\sigma} \right), \quad (\text{E2})$$

where  $\varrho(\cdot)$  is a Gaussian distribution with zero mean and unit variance. One can iterate this equation, and, exploiting the fact that the Gaussian is stable under convolution, one obtains

$$\begin{aligned} \rho^{i+1}(\Delta q) &= \int_{-\infty}^{\infty} d\Delta q' \rho^0(\Delta q') \varrho \\ &\quad \times \left( \frac{\Delta q - \langle \Delta q(i) \rangle_{\Delta q'}}{\sqrt{\langle \Delta q(i)^2 \rangle_{\Delta q'} - \langle \Delta q(i) \rangle_{\Delta q'}^2}} \right), \end{aligned} \quad (\text{E3})$$

where  $\langle \Delta q_0(i) \rangle_{\Delta q'}$  is the average of  $\Delta q$  at generation  $i$  given that the initial log-size displacement at the first generation  $i=0$  was  $\Delta q'$ . In order to have an explicit equation, we need just to calculate the  $\langle \Delta q_0(i) \rangle_{\Delta q(0)}$  and  $\langle \Delta q_0(i)^2 \rangle_{\Delta q(0)}$ .

The mean displacement can be calculated by solving

$$\langle \Delta q_0(i+1) \rangle_{\Delta q(0)} = (1-\lambda) \langle \Delta q_0(i+1) \rangle_{\Delta q(0)} + \sigma \xi, \quad (\text{E4})$$

with initial condition  $\Delta q_0(0)$ . The solution reads

$$\langle \Delta q_0(n) \rangle_{\Delta q(0)} = \Delta q(0)(1-\lambda)^n. \quad (\text{E5})$$

A similar equation can be written for the second moment,

$$\langle \Delta q_0(i+1)^2 \rangle_{\Delta q(0)} = (1-\lambda)^2 \langle \Delta q_0(i)^2 \rangle_{\Delta q(0)} + \sigma^2, \quad (\text{E6})$$

whose solution is

$$\langle \Delta q_0(n)^2 \rangle_{\Delta q(0)} = \sigma^2 \frac{1 - (1-\lambda)^{2n}}{\lambda(2-\lambda)} + (1-\lambda)^{2(n-1)} \Delta q(0)^2. \quad (\text{E7})$$

Therefore, we finally obtain

$$\begin{aligned} & \langle \Delta q_0(n)^2 \rangle_{\Delta q(0)} - \langle \Delta q_0(n) \rangle_{\Delta q(0)}^2 \\ &= \sigma^2 \frac{1 - (1-\lambda)^{2n}}{\lambda(2-\lambda)} + (1-\lambda)^{2n} \frac{\lambda(2-\lambda)}{(1-\lambda)^2} \Delta q(0)^2. \end{aligned} \quad (\text{E8})$$

By taking the limit  $n \rightarrow \infty$  of Eq. (E7) we obtain the stationary variance, which reads

$$\sigma_q^2 = \frac{\sigma^2}{\lambda(2-\lambda)}. \quad (\text{E9})$$

The stationary distribution is therefore

$$\rho_b^*(q) = \frac{1}{\sqrt{2\pi\sigma_q^2}} \exp\left(-\frac{(q - \langle q_0 \rangle_\alpha)^2}{2\sigma_q^2}\right), \quad (\text{E10})$$

and therefore the one of the sizes at birth is

$$p_b^*(x_0) = \frac{1}{\sqrt{2\pi\sigma_q^2}x_0} \exp\left(-\frac{(\log(x_0/x^*) - \langle q_0 \rangle_\alpha)^2}{2\sigma_q^2}\right), \quad (\text{E11})$$

which has mean

$$\langle x_0 \rangle_\alpha = x^* e^{\langle q_0 \rangle_\alpha + \sigma_q^2/2}, \quad (\text{E12})$$

and variance

$$\sigma_{x_0}^2 = (x^*)^2 e^{2\langle q_0 \rangle_\alpha} (e^{\sigma_q^2} - 1) e^{\sigma_q^2}. \quad (\text{E13})$$

The coefficient of variation of the size at birth is defined as

$$\text{cv}_x^2 = \frac{\sigma_{x_0}^2}{\langle x_0 \rangle_\alpha^2} = (e^{\sigma_q^2} - 1), \quad (\text{E14})$$

and therefore we have

$$\sigma_q^2 = \log\left[1 + \left(\frac{\sigma_{x_0}}{\langle x_0 \rangle_\alpha}\right)^2\right]. \quad (\text{E15})$$

#### APPENDIX F: PERTURBATIVE EXPANSION AND IDENTIFICATION OF PARAMETERS FOR THE ADDER MODEL

An adder mechanism of division control corresponds to a division probability of the form

$$p(x_f|x_0, \alpha) = F_\alpha(x_f - x_0). \quad (\text{F1})$$

Using the scaling of the stationary distributions of Eq. (A6), we obtain

$$F_\alpha(x_f - x_0) = p(x_f|x_0, \alpha) = \frac{1}{\langle x_0 \rangle_\alpha} F\left(\frac{x_f - x_0}{\langle x_0 \rangle_\alpha}\right). \quad (\text{F2})$$

This equation is consistent with the collapse of the probabilities of added size as observed in Ref. [5].

The hazard rate  $h(z)$  for the adder model can be easily obtained using

$$F(z) = h(z) \exp\left[\int_0^z h(y) dy\right]. \quad (\text{F3})$$

By inverting this equation, one obtains that

$$h(z) = \frac{d}{dz} \log\left[\int_0^z F(y) dy\right]. \quad (\text{F4})$$

By using Eq. (B2) and introducing  $q = \log(x/x^*)$ , one obtains the functional form of  $g(\cdot)$ ,

$$\begin{aligned} g(q_0 - \langle q_0 \rangle_\alpha) &= \int dx p(x|x^* e_0^q, \alpha) \log(x/x^*) \\ &\quad - \langle q_0 \rangle_\alpha - \log(2), \end{aligned} \quad (\text{F5})$$

and, by introducing the scaling of Eq. (F4), this expression reads

$$\begin{aligned} g(q_0 - \langle q_0 \rangle_\alpha) &= \frac{1}{\langle x_0 \rangle_\alpha} \int dx F\left(\frac{x - x^* e^{q_0}}{\langle x_0 \rangle_\alpha}\right) \log(x/x^*) \\ &\quad - \langle q_0 \rangle_\alpha - \log(2) \\ &= c e^{-\langle q_0 \rangle_\alpha} \int dx F\left(\frac{x}{\langle x_0 \rangle_\alpha} - c e^{q_0 - \langle q_0 \rangle_\alpha}\right) \\ &\quad \times \log(x/x^*) - \langle q_0 \rangle_\alpha - \log(2) \\ &= \int ds F(s - c e^{q_0 - \langle q_0 \rangle_\alpha}) \log(s) - \log(2) \\ &= \int dz F(z) \log(z + c e^{q_0 - \langle q_0 \rangle_\alpha}) - \log(2), \end{aligned} \quad (\text{F6})$$

leading to the final expression,

$$\begin{aligned} g(q_0 - \langle q_0 \rangle_\alpha) &= q_0 - \langle q_0 \rangle_\alpha + \int dz F(z) \log\left(\frac{c + z e^{\langle q_0 \rangle_\alpha - q_0}}{2}\right), \end{aligned} \quad (\text{F7})$$

where the variable  $c = \langle q_0 \rangle_\alpha - \log(\langle x_0 \rangle_\alpha)$  can be estimated by imposing  $g(0) = 0$ . Under this condition,  $c$  is defined as the solution of

$$\int dz F(z) \log\left(\frac{c+z}{2}\right) = 0. \quad (\text{F8})$$

In a similar way, the value of the control parameter  $\lambda$  can be obtained from the following equation:

$$\lambda = 1 - g'(0) = \int dz F(z) \frac{z}{c+z}. \quad (\text{F9})$$

Since the value of  $\lambda$  appears as first order in an expansion around the mean initial logarithmic cell size, we can neglect size fluctuations in calculating its value, since they correspond to subleading terms. Note, however, that these subleading

terms have to be considered when other terms than the first order are included in the expansion. Since in the adder model  $\langle x_f - x_0 \rangle = \langle x_0 \rangle$ , we have (up to subleading terms) from Eq. (F4)

$$\langle z \rangle = \int dz F(z)z = 1. \quad (\text{F10})$$

Therefore, by neglecting fluctuations in Eq. (F8), i.e., by imposing  $F(z) = \delta(z - 1)$ , we obtain  $c = 1$  and therefore  $\lambda = 1/2$ . In case we are also considering quadratic term in the expansion of  $g(\cdot)$ , we should include also a correction on this value, by a factor that depends on the variance of the added size.

In the following we consider an explicit case of the adder model. We assume that  $p(x_f|x_0, \alpha)$  is a log-normal distribution, which corresponds to

$$F(z) = \frac{1}{\sqrt{2\pi z\sigma_a}} \exp\left[-\frac{(\log z)^2}{2\sigma_a^2}\right]. \quad (\text{F11})$$

We have, therefore,  $\langle z \rangle = \exp(\sigma_a^2/2)$  and

$$\langle z^2 \rangle - \langle z \rangle^2 = e^{2\sigma_a^2} - e^{\sigma_a^2}.$$

By introducing this expression in Eq. (F7) we obtain

$$g(q) = q + \int dz \frac{1}{\sqrt{2\pi z\sigma_a}} \exp\left(-\frac{(\log z)^2}{2\sigma_a^2}\right) \times \log\left(\frac{c + ze^{-q}}{2}\right). \quad (\text{F12})$$

By expanding this expression up to second order in  $q$ , we obtain

$$g(q) \approx q \left\langle \frac{c}{c+z} \right\rangle + \frac{q^2}{2} \left\langle \frac{cz}{(c+z)^2} \right\rangle. \quad (\text{F13})$$

Assuming that the fluctuations are small, it is natural to expand the terms in  $z$  around  $z = \langle z \rangle$ . Equation (F8) then reduces to

$$0 = \left\langle \log \frac{c+z}{2} \right\rangle \approx \log \frac{c + \langle z \rangle}{2} + \frac{\langle (z - \langle z \rangle)^2 \rangle}{2(c + \langle z \rangle)^2} \\ \approx \log \frac{c+1}{2} + \frac{\sigma_z^2}{2(1+c)}, \quad (\text{F14})$$

and therefore

$$c \approx 1 - \frac{\sigma_z^2}{4}. \quad (\text{F15})$$

Expanding Eq. (F13) around  $\langle z \rangle$ , and introducing the explicit dependence on  $\sigma_z$ , we obtain

$$g(q) \approx q \left( \frac{1}{2} - \frac{\sigma_z^2}{16} + o(\sigma_z^4) \right) + \frac{q^2}{2} \left( \frac{1}{4} + o(\sigma_z^2) \right), \quad (\text{F16})$$

which corresponds to  $\lambda = 1/2 + \sigma_z^2/16$ .

In a similar way, it is possible to estimate the variance of the noise term in the discrete Langevin formalism

$$\sigma(q_0 - \langle q_0 \rangle_\alpha)^2 = \int dx p(x|x^* e^q, \alpha) [\log(x/x^*) - \langle q_0 \rangle_\alpha \\ - \log(2)]^2 - g(q_0 - \langle q_0 \rangle_\alpha)^2, \quad (\text{F17})$$

and, by substituting for  $F(z)$  in this expression,

$$\sigma(q)^2 = \int dx F(z) \left[ q + \log\left(\frac{c + ze^{-q}}{2}\right) \right]^2 \\ - g(q_0 - \langle q_0 \rangle_\alpha)^2. \quad (\text{F18})$$

By expanding it up to the first order, one obtains

$$\sigma(q)^2 \approx \left\langle \left[ \log\left(\frac{c+z}{2}\right) \right]^2 \right\rangle + q \left\langle \frac{2c \log\left(\frac{c+z}{2}\right)}{c+z} \right\rangle. \quad (\text{F19})$$

We can calculate explicitly the two terms in the case of a log-normal  $F(z)$  in the limit of small  $\sigma_z$ , and we obtain

$$\sigma(q)^2 \approx \frac{\sigma_z^2}{4} - \frac{\sigma_z^2}{4} q, \quad (\text{F20})$$

and finally

$$\sigma(q) \approx \frac{\sigma_z}{2} - \frac{\sigma_z}{4} q. \quad (\text{F21})$$

In general, a log-normal distribution does not result from a discrete-time Langevin process with a normal noise as in Eq. (11). On the other hand, since we are expanding for small fluctuations, the errors made approximating it with a normal noise are subleading. Using the notation

$$\Delta q(i+1) = (1-\lambda)\Delta q(i) + \gamma \frac{[\Delta q(i)]^2}{2} + [\sigma + \beta \Delta q(i)]\xi, \quad (\text{F22})$$

we have that, for the adder model,

$$\lambda = \frac{1}{2} + \frac{\sigma_z^2}{16} + o(\sigma_z^3), \\ \gamma = \frac{1}{4} - \frac{\sigma_z^2}{16} + o(\sigma_z^3), \\ \sigma = \frac{\sigma_z}{2} + o(\sigma_z^2), \\ \beta = -\frac{\sigma_z}{2} + o(\sigma_z^2). \quad (\text{F23})$$

[1] A. S. Kennard, M. Osella, A. Javer, J. Grilli, P. Nghe, S. J. Tans, P. Cicuta, and M. C. Lagomarsino, Individuality and universality in the growth-division laws of single *E. coli* cells, *Phys. Rev. E* **93**, 012408 (2016).

[2] 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, *Proc. Natl. Acad. Sci. USA* **111**, 15912 (2014).

- [3] A. Giometto, F. Altermatt, F. Carrara, A. Maritan, and A. Rinaldo, Scaling body size fluctuations, *Proc. Natl. Acad. Sci. USA* **110**, 4646 (2013).
- [4] M. Hashimoto, T. Nozoe, H. Nakaoka, R. Okura, S. Akiyoshi, K. Kaneko, E. Kussell, and Y. Wakamoto, Noise-driven growth rate gain in clonal cellular populations, *Proc. Natl. Acad. Sci. USA* **113**, 3251 (2016).
- [5] 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).
- [6] A. Amir, Cell Size Regulation in Bacteria, *Phys. Rev. Lett.* **112**, 208102 (2014).
- [7] M. OSELLA, E. Nugent, and M. C. Lagomarsino, Concerted control of *Escherichia coli* cell division, *Proc. Natl. Acad. Sci. USA* **111**, 3431 (2014).
- [8] M. Campos, I. V. Surovtsev, S. Kato, A. Paintdakhi, B. Beltran, S. E. Ebmeier, and C. Jacobs-Wagner, A constant size extension drives bacterial cell size homeostasis, *Cell* **159**, 1433 (2014).
- [9] L. Robert, M. Hoffmann, N. Krell, S. Aymerich, J. Robert, and M. Doumic, Division in *Escherichia coli* is triggered by a size-sensing rather than a timing mechanism, *BMC Biology* **12**, 17 (2014).
- [10] 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).
- [11] M. Deforet, D. van Ditmarsch, and J. B. Xavier, Cell-size homeostasis and the incremental rule in a bacterial pathogen, *Biophys. J.* **109**, 521 (2015).
- [12] 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* **523**, 357 (2015).
- [13] S. Iyer-Biswas, G. E. Crooks, N. F. Scherer, and A. R. Dinner, Universality in Stochastic Exponential Growth, *Phys. Rev. Lett.* **113**, 028101 (2014).
- [14] S. Jun and S. Taheri-Araghi, Cell-size maintenance: Universal strategy revealed, *Trends Microbiol.* **23**, 4 (2015).
- [15] L. K. Harris and J. A. Theriot, Relative rates of surface and volume synthesis set bacterial cell size, *Cell* **165**, 1479 (2016).
- [16] P. R. Painter and A. G. Marr, Mathematics of microbial populations, *Annu. Rev. Microbiol.* **22**, 519 (1968).
- [17] K. M. Schmoller, J. J. Turner, M. Kõivomägi, and J. M. Skotheim, Dilution of the cell cycle inhibitor Whi5 controls budding-yeast cell size, *Nature* **526**, 268 (2015).
- [18] F. J. Trueba, O. M. Neijssel, and C. L. Woldringh, Generality of the growth kinetics of the average individual cell in different bacterial populations, *J. Bacteriol.* **150**, 1048 (1982).
- [19] 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).
- [20] D. A. Kessler and S. Burov, Stochastic maps, continuous approximation and stable distribution, [arXiv:1612.08703](https://arxiv.org/abs/1612.08703).
- [21] D. A. Kessler and S. Burov, Effective potential for cellular size control, [arXiv:1701.01725](https://arxiv.org/abs/1701.01725).