Intergenerational scaling law determines the precision kinematics of stochastic individual-cell-size homeostasis

Kunaal Joshi[*]

Department of Physics and Astronomy, Purdue University, West Lafayette, IN 47907.

Rudro R. Biswas[†]

Department of Physics and Astronomy, Purdue University, West Lafayette, IN 47907.

Srividya Iyer-Biswas

Department of Physics and Astronomy, Purdue University, West Lafayette, IN 47907. and Santa Fe Institute, Santa Fe, NM 87501.[‡]

Abstract

Individual bacterial cells grow and divide stochastically. Yet they maintain their characteristic sizes across generations within a tightly controlled range. What rules ensure intergenerational stochastic homeostasis of individual cell sizes? Valuable clues have emerged from high-precision longterm tracking of individual statistically-identical *Caulobacter crescentus* cells as reported in [1]: Intergenerational cell size homeostasis is an inherently stochastic phenomenon, follows Markovian or memory-free dynamics, and cells obey an intergenerational scaling law, which governs the stochastic map characterizing generational sequences of characteristic cell sizes. These observed emergent simplicities serve as essential building blocks of the first-principles-based theoretical framework we develop here. Our exact analytic fitting-parameter-free results for the predicted intergenerational stochastic map governing the precision kinematics of cell size homeostasis are remarkably well borne out by experimental data, including extant published data on other microorganisms, *Escherichia coli* and *Bacillus subtilis*. Furthermore, our framework naturally yields the general exact and analytic condition, necessary and sufficient, which ensures that stochastic homeostasis can be achieved and maintained. Significantly, this condition is more stringent than the known heuristic result for the popular quasi-deterministic adder-sizer-timer frameworks. In turn the fully stochastic treatment we present here extends and updates extant frameworks, and challenges the notion that the mythical "average cell" can serve as a reasonable proxy for the inherently stochastic behaviors of actual individual cells.

[*] joshi84@purdue.edu
[†] rrbiswas@purdue.edu
[‡] iyerbiswas@purdue.edu

I. INTRODUCTION

The processes of cellular growth and division are fundamental to the survival and propagation of life. An outstanding open question is how the characteristic generational size of an individual cell remains "constant" across generations, even as that cell repeatedly grows and divides, given the significant stochasticity in both growth and division processes. Homeostasis is the process of maintaining "constancy" to within a specified tolerance of a macroscopic characteristic or state variable, despite the complex internal processes that prevail in even the simplest of organisms [2, 3]. The question of how cells maintain size homeostasis is, in turn, connected to the broader goal of understanding the quantitative principles and mechanisms by which complex processes are controlled and regulated to ensure proper organismal functioning in the face of inescapable stochastic fluctuations in both internal and external environments [4, 5].

Bacterial cells serve as uniquely convenient systems to characterize the dynamics of cell size homeostasis, since the entire organism consists of a single cell. Recently developed approaches for multigenerational single-cell imaging of microorganisms provide the means to observe growth and division tracks of individual bacterial cell sizes, division upon successive division [1, 6-12]. Popular single-cell technologies typically use the concept of the Mother Machine, which takes advantage of designed confinement and one-dimensional crowding of cells in narrow channels and alleviates the problem of exponential crowding of the imaging region of interest [6, 8, 10]. The "mother" cell confined to the bottom of the narrow channel can be tracked for extended durations since it cannot easily escape and flow away. Thus, these approaches yield long-term intergenerational trajectories of individual cell growth and division; the size of the relevant dataset is set by the number of such mother cells that can be imaged in a microfluidic device at reasonably frequent intervals of time with sufficient spatial resolution to record the size at birth and at division.

Significant effort has been expended in characterizing how the size at division ("final size") is dependent—on average—on the size at birth ("initial size") for a given cell in a given generation [10, 13-17]. Typically, data from different cells and generations are pooled together on a scatter plot of size at division versus size at birth, and the relationship between these quantities used to infer the underlying phenomenology. In this quasi-deterministic picture, homeostasis is characterized by a target "set point" of cell size alongside the sensitivity of final size to initial size in a given generation. This sensitivity is typically categorized into one of three schemes, referred to as sizer (wherein the size at division is independent of the size at birth), timer (wherein the size

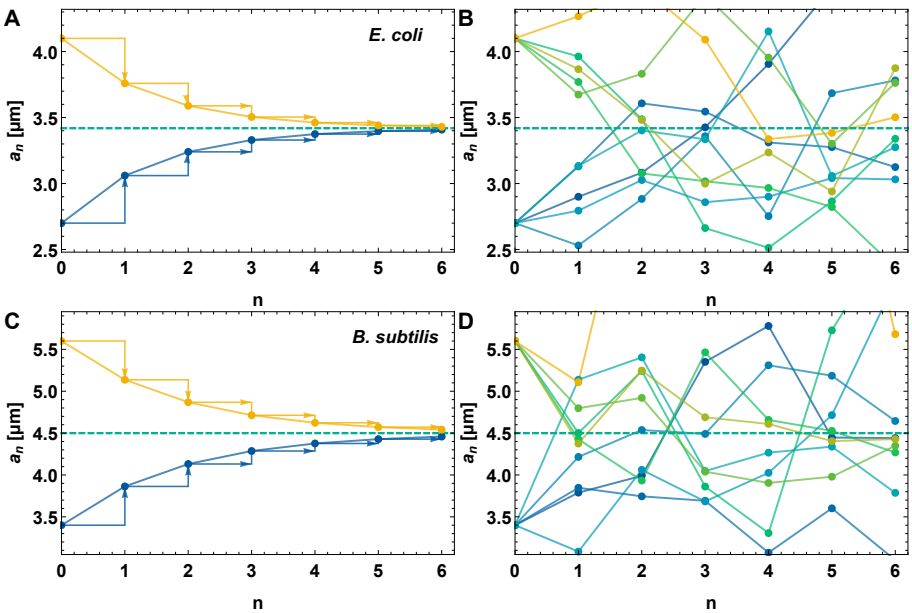

FIG. 1. **Heuristic understanding of homeostasis motivated by quasi-deterministic frameworks vis-a-vis experimentally obtained individual-cell data** The left panels show idealized deterministic initial cell size (a_n) trajectories corresponding to the sizer-timer-adder framework. Within this perspective, through evolution over successive generations (n), cells starting with larger (yellow) and smaller (blue) than average sizes in the $n = 0$ generation converge exponentially to the population mean (dashed green line) as shown for the two different species featured in [10]. In this deterministic picture, the deviation from the mean decreases in each subsequent generation in proportion to the slope of the linear dependence of next generation's initial size to current generation's (see Fig. 5 B). Each of the panels on the right show the actual trajectories of five random cells each, starting from the initial values used in the panels on the left, and for each of the experimental conditions. Evidently, the actual trajectories look dramatically different from the corresponding deterministic picture provided by the sizer-timer-adder heuristic, thus motivating the need for a first-principles-based stochastic framework to realistically capture the observed phenomenology. **(A-B)** *E. coli* in MOPS glucose rich media, **(C-D)** *B. subtilis* in LB media. Data source: [10].

at division is a multiple of the size at birth), and adder (wherein the size at division is a constant added to the size at birth) [10, 13-17]. For cellular homeostasis, this scheme is used to motivate a heuristic argument for a deterministic exponential approach to the target cell size over successive generations (see Fig. 1) [18-20]. The sensitivity value is reproducible in experimental replicates; however, it varies across species, as well as across growth conditions for the same species. That said, the preponderance of current interest rests on the "adder" scheme [19-22]. In contrast, as we have shown in [1], the emergence of cell size homeostasis is the result of an inherently stochastic and intergenerational process *for each individual cell*. Thus, a deterministic exponential relax-

ation picture, motivated by population averages characterizing the mythical "average cell", does not faithfully capture observed intergenerational phenomenology. This divergence points to important lacunae in our quantitative understanding of stochastic intergenerational homeostasis.

In this work we take advantage of recently reported "emergent simplicities" in the stochastic intergenerational size homeostasis of individual cells of *Caulobacter crescentus* [1]. These datasets were obtained using the SChemostat technology, in which the original "mother" cell is retained after each division, and the newborn "daughter" cell leaves the experimental arena; this facilitates high-precision characterization of all cells being imaged [9, 11]. These datasets focus on the same ensemble of statistically identical non-interacting cells over the course of tens of generations, imaged in precisely controlled time invariant growth conditions. In [1], we reported direct evidence of stochastic cell size homeostasis using these data, demonstrating that the intergenerational dynamics are Markovian and that stochastic homeostasis follows an intergenerational scaling law for cell size (see Fig. 2). In the present work, we build on these experimental observations in *C. crescentus* to develop a first-principles-based theoretical framework, and derive the exact analytic intergenerational stochastic map governing cell size dynamics. We then reanalyze published multigenerational datasets for *Escherichia coli* and *Bacilus subtilis* obtained using the Mother Machine technology [10]. Our exact, analytic results for the predicted intergenerational stochastic map governing the precision kinematics of cell size homeostasis, developed for *C. crescentus* data without fitting parameters [1], are remarkably well borne out by these experimental data on other microorganisms [10]. (See Figs. 3 and 4.) Furthermore, our framework naturally yields the general exact, analytic condition, which is both necessary and sufficient to ensure that stochastic homeostasis can be achieved and maintained (Eq. (17) and Fig. 5). Significantly, this condition is more stringent than the well-known heuristic result for the popular quasi-deterministic sizer–timer–adder frameworks. The fully stochastic treatment we present here extends and updates previous frameworks, and challenges the notion that the mythical "average cell" can serve as a reasonable proxy for the inherently stochastic behaviors of actual individual cells.

II. EMERGENT SIMPLICITIES IN THE STOCHASTIC INTERGENERATIONAL HOMEOSTASIS OF INDIVIDUAL CELLS

We now develop a theoretical framework for the generation-to-generation evolution ("kinematics") of an individual cell's size (see Fig. 2). We use the random variable, a_n, to denote the cell size in the nth generation. Since cell size grows from division to division, we need to choose a rep-

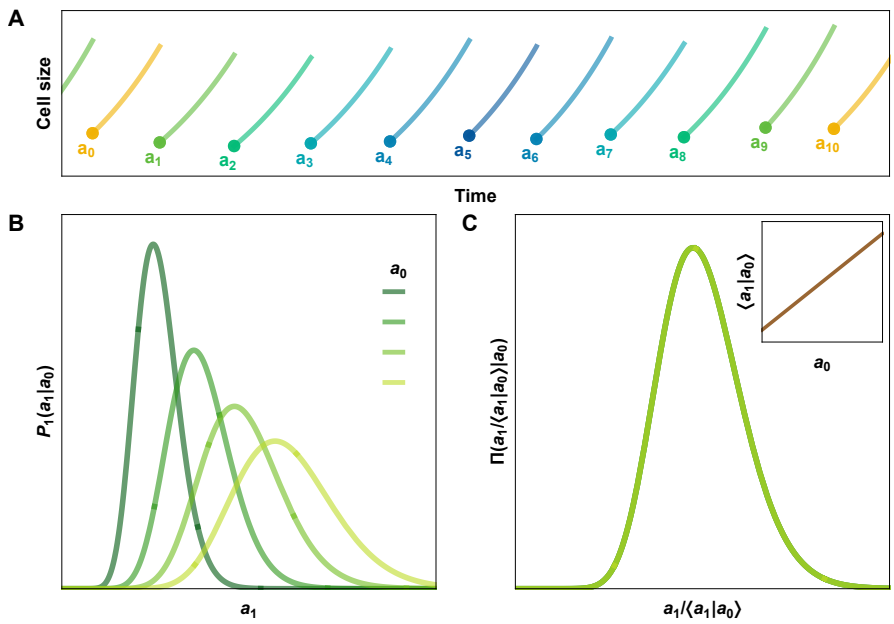

FIG. 2. **Schematic representation of typical intergenerational single-cell growth tracks and the intergenerational scaling law observed. (A)** A typical intergenerational size (growth) trajectory of an individual cell. The sizes at birth, or "initial sizes", are labeled a_n, where n is the generation number since the starting, or zeroth, generation. **(B)** Schematic conditional distributions, P_1, of the next generation's initial size, conditioned on the current generation's initial size. This characterization uses the experimental observation that the intergenerational cell size dynamics in balanced growth conditions are Markovian (see [1]). **(C)** The invariant distribution, Π, obtained by rescaling the distributions in (B) by their respective mean values. As reported in [1], the distributions in (B) have the property that when rescaled by their mean values, they result in universal distribution (a_0 independent but growth condition specific). Inset: The mean values of next generation's initial size given current generation's ($\langle a_1|a_0\rangle = \mu(a_0)$), are found to vary linearly with the current generation's initial size (a_0) in experimental observations (see [1] and Fig. 5 B). Using this calibration curve (inset), one can recover the non-universal distributions in (B) for any a_0 starting with the universal distribution in (C).

resentative size from within the cell cycle: we choose a_n to be the cell size-at-birth (see Fig. 2). We treat $n = 0$ as the "initial" generation. In addition to the usual convention that $P(X|Y)$ denotes the conditional probability distribution of the random variable X given a specific realization of the random variable Y, we reserve the specific notation $P_n(a|a_0)$ to denote the conditional probability distribution of a_n at $a_n = a$, given the initial size of the initial generation, a_0. This general setup permits various possibilities for attaining or violating cell size homeostasis, depending on general properties of the multivariate functions $P_n(a|a_0)$, which in principle can depend on history.

Significant reduction in the complexity of the general problem results from the experimental observations, and corresponding emergent simplicities, for *C. crescentus* cells reported in [1]. The first emergent simplicity is that in each growth condition, the conditional distribution of next generation's initial size, conditioned on the current and previous generations' initial sizes, in fact only depends on the current generation's initial size. In other words, it is independent of previous generations. Equivalently, the intergenerational dynamics of cell sizes are Markovian under constant growth conditions; thus, the initial size of a given cell in the current generation is the sole determinant of the statistics of future sizes-at-birth of the same cell. It follows that stochastic intergenerational cell size dynamics can be characterized completely by properties of the single-generational stochastic map P_1. In other words, the function P_n can be written down only in terms of P_1:

$$P_n(a_n|a_0) = \left[\int_0^\infty\right]^{n-1}\left[P_1(a_n|a_{n-1})\prod_{j=1}^{n-1}\{P_1(a_j|a_{j-1})da_j\}\right]. \tag{1}$$

Within this perspective, attainment of cell size homeostasis corresponds to:

$$\lim_{n\to\infty} P_n(a|a_0) = P_\infty(a), \qquad \forall a_0, \tag{2}$$

where P_∞ is the asymptotic homeostatic distribution, independent of the initial size. Furthermore, we require that the homeostatic distribution have well-defined moments. Thus, for a cell that starts with initial size a_0, considering all possible futures after a large number of generations, the size probability distribution converges to a well-behaved homeostatic distribution which is independent of the precise value of a_0. In this work, we seek to define under what conditions the kinematics of intergenerational cell sizes specified by the experimentally observed emergent simplicities lead to the homeostatic distribution, P_∞.

In [1] we report a second emergent simplicity for *C. crescentus* cells: the mean-rescaled distribution of cell sizes-at-birth (initial sizes) in a given generation, conditioned on the initial size of the previous generation, is independent of the previous generation's initial size. We refer to this result as the "intergenerational scaling law" (see Fig. 2). Mathematically,

$$P_1(a_n|a_{n-1}) = \frac{1}{\mu(a_{n-1})}\Pi\left(\frac{a_n}{\mu(a_{n-1})}\right), \tag{3}$$

where Π is the universal mean-rescaled distribution that is independent of a cell's history of size dynamics. Π has mean $m_1 = 1$, and $\mu(a_{n-1})$ is the mean of a_n when conditioned on the previous generation's initial size, a_{n-1}. We note that the shape of Π may vary depending on growth condition.

Experimentally, $\mu(a)$ is found to be well-described by a linear function over the physiologically relevant range of cell sizes observed (see Fig. 5 B and [1]):

$$\mu(a) = \alpha a + \beta. \tag{4}$$

The different values of the slope, α (observed in different growth conditions for different organisms), have been scrutinized in great detail, leading to vigorous debates regarding their mechanistic implications for growth control [10, 19–22]. The specific values to which meanings are attached are $\alpha = 0$ (sizer), $\alpha = 1$ (timer), and $\alpha = r$ (adder, with $r < 1$ being the division ratio). As indicated previously, these approaches view cell size homeostasis as a quasi-deterministic process in which the size in a given generation, if different from the homeostatic setpoint, exponentially relaxes to the setpoint size over successive generations with rate $\ln(1/|\alpha|)$ (see Fig. 1 A and [18]). Since this rate must be positive for homeostasis to be achieved, this heuristic treatment requires that the condition $|\alpha| < 1$ be obeyed to permit the possibility of homeostasis. Thus, the value of α needs to be less than the slope of the timer ($\alpha = 1$) for homeostasis to be attainable in this quasi-deterministic picture.

In stark contrast to the sizer–timer–adder quasi-deterministic perspective, our main focus here is on the intergenerational scaling law, Eq. (3)—we view the value of the slope α, in Eq. (4) as simply a calibration parameter, and do not imbue it with any special mechanistic meaning. We now proceed to develop the precise kinematics governing stochastic intergenerational cell size dynamics. We use the intergenerational scaling law Eq. (3) and the calibration curve Eq. (4) as the essential building blocks of our theory.

III. RESULTS AND DISCUSSIONS

A. The intergenerational scaling law universally determines the precision kinematics of stochastic intergenerational homeostasis.

We now proceed to develop the following universal framework for intergenerational evolution of an individual cell's size. Eq. (3) and the Markov property imply that the sizes-at-birth of a given cell follow the stochastic map:

$$a_{n+1} = s_{n+1}\mu(a_n). \tag{5}$$

In the preceding equation, the rescaled random variables, $\{s_n\}$, are uncorrelated. They are drawn independently from the mean-rescaled probability distribution, Π, introduced in Eq. (3), which

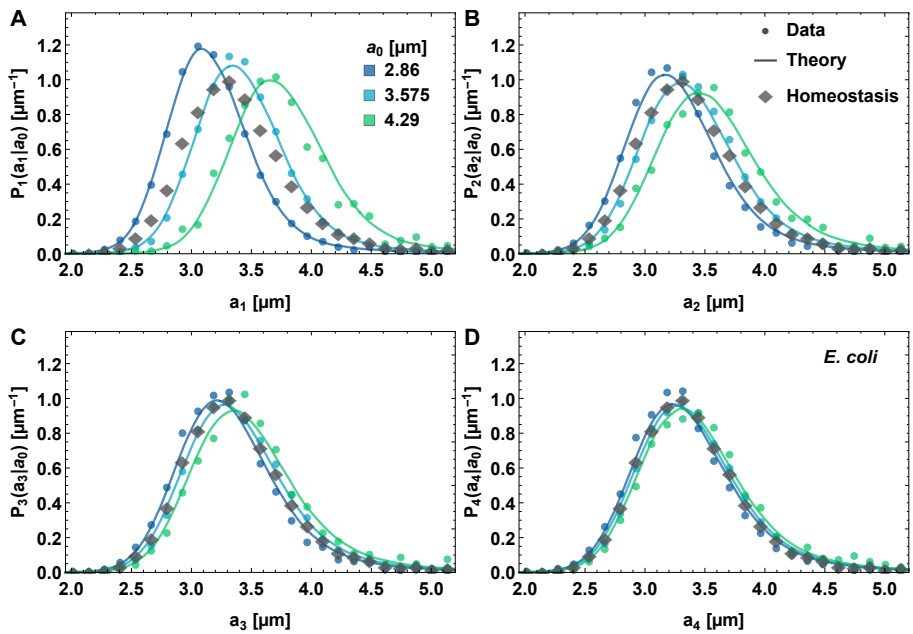

FIG. 3. **Our theoretical framework accurately predicts, with no fitting parameters, the stochastic intergenerational cell size dynamics leading to homeostasis, for *E. coli* cells.** Sequence of panels shows the intergenerational evolution of distributions of initial cell sizes, a_n (n is the generation number), starting from three different initial sizes in the zeroth generation (a_0, marked by different colors). Panels (A-D) correspond to generations $1-4$ respectively. The filled circles correspond to experimental data. Curves correspond to fitting-free predictions from our theoretical intergenerational framework. The homeostatic distribution, obtained by pooling together the data for initial sizes, is overlaid in all panels (diamond markers) for ready comparison. Our theoretical framework accurately predicts the stochastic dynamics governing the precise convergence of conditional initial size distributions to the homeostatic distribution over successive generations. Data are taken from [10]; these correspond to *E. coli* cells grown in glucose rich media. Cell length was used as a measure of cell size in these datasets. See Fig. 4 for similar application to *B. subtilis* cells.

has mean $m_1 = 1$. (We denote the kth moment of the probability distribution Π by m_k.) The random variable a_n denotes the initial cell size in the nth generation (see Fig. 2). Using Eq. (5) recursively, the size-at-birth in the nth generation can be related to the size-at-birth in the 0th generation:

$$\begin{aligned} a_n &= s_n \mu(a_{n-1}) = s_n \mu(s_{n-1}\mu(a_{n-2})) = \dots \\ &= s_n \mu(s_{n-1}\mu(s_{n-2}\mu(\dots\mu(s_1\mu(a_0))\dots))). \end{aligned} \tag{6}$$

Homeostasis is attained when the probability distribution for a_n, and hence all moments of a_n, reach finite asymptotic limits that are independent of a_0 as $n \to \infty$.

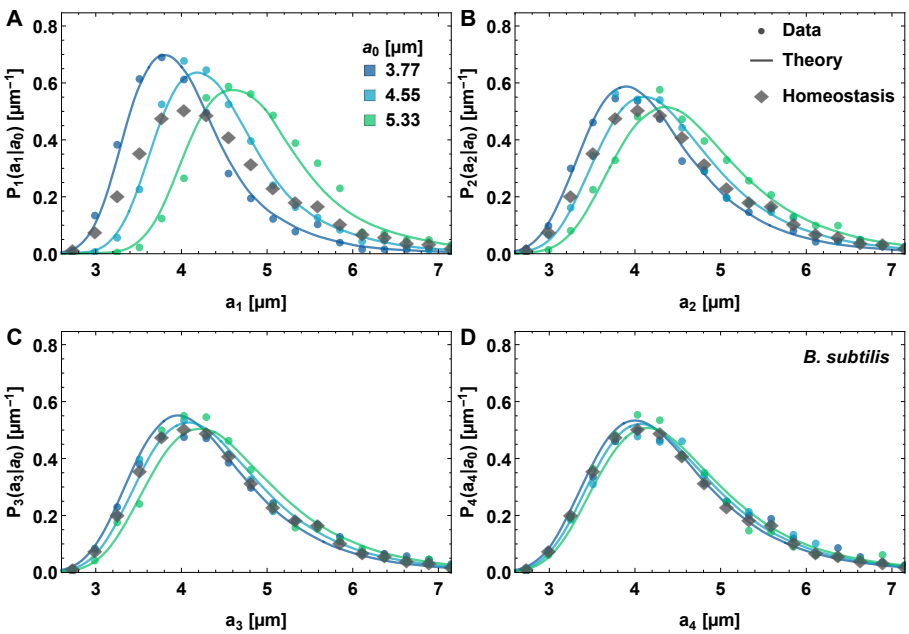

FIG. 4. **Our theoretical framework accurately predicts, with no fitting parameters, the stochastic intergenerational cell size dynamics leading to homeostasis, for *B. subtilis* cells.** Sequence of panels shows the intergenerational evolution of distributions of initial cell sizes, a_n (n is the generation number), starting from three different initial sizes in the zeroth generation (a_0, marked by different colors). Panels (A-D) correspond to generations $1-4$ respectively. The filled circles correspond to experimental data. Curves correspond to fitting-free predictions from our theoretical intergenerational framework. The homeostatic distribution, obtained by pooling together the data for initial sizes, is overlaid in all panels (diamond markers) for ready comparison. Our theoretical framework accurately predicts the stochastic dynamics governing the precise convergence of conditional initial size distributions to the homeostatic distribution over successive generations. Data are taken from [10]; these correspond to *B. subtilis* cells grown in LB media. Cell length was used as a measure of cell size in these datasets. See Fig. 3 for similar application to *E. coli* cells.

For experimentally relevant scenarios $\mu(a)$ is typically a linear function, $\mu(a) = \alpha a + \beta$, as written down in Eq. (4). Thus Eq. (6) becomes:

$$a_n = s_n(\beta + \alpha a_{n-1}) = \beta s_n + \alpha s_n s_{n-1}(\beta + \alpha a_{n-2}) = \ldots = a_0 \alpha^n \prod_{i=1}^{n} s_i + \beta \sum_{j=1}^{n} \alpha^{n-j} \prod_{k=j}^{n} s_k. \tag{7}$$

This equation shows the explicit connection between: the starting size, a_0; the stochastic variable denoting the size in the nth generation, a_n; and the sequence of intermediate independent stochastic scaling factors $\{s_1, s_2, \ldots, s_n\}$.

Using this map, we use our theoretical framework to predict the distributions of initial sizes

over successive generations, for cells starting from any given initial size, using only the mean-rescaled distribution Π and the calibration curve μ. Our fitting parameter-free predictions (Eqs. (1), (3) and (4)) match excellently when applied to published *E. coli* (Fig. 3) and *B. subtilis* (Fig. 4) data, taken from [10]. Moreover, compelling data-theory matches were also obtained for the *C. crescentus* data as detailed in [1]. Thus we have comprehensively validated the framework we propose here.

B. The deterministic limit of stochastic intergenerational homeostasis.

In what follows, we take a_0 to have a fixed value. All averages $\langle\cdots\rangle$ are taken with respect to the random variables $\{s_m\}$. When a_0 is explicitly taken to be random, i.e., the initial generation size is drawn from an arbitrary distribution, we denote further averaging over initial sizes by double angular brackets, $\langle\langle\cdots\rangle\rangle$. To evaluate the moments characterizing the probability distribution of a_n in terms of α, β and the raw moments, $\{m_k\}$, of the generation-independent scaling factor distribution Π, we raise Eq. (7) to different powers and take averages. Directly averaging Eq. (7), we find the mean initial size in the nth generation:

$$\langle a_n\rangle = \left\langle a_0\alpha^n \prod_{i=1}^{n} s_i + \beta \sum_{j=1}^{n} \alpha^{n-j} \prod_{k=j}^{n} s_k \right\rangle = \frac{\beta}{1-\alpha} + \alpha^n\left(a_0 - \frac{\beta}{1-\alpha}\right). \tag{8}$$

After a large number of generations have elapsed (when $n \to \infty$), the second term diverges unless $|\alpha| < 1$. *Thus* $|\alpha| < 1$ *is a necessary condition for homeostasis*. When $|\alpha| < 1$:

$$\langle a\rangle_\infty \stackrel{\text{def}}{=} \lim_{n\to\infty} \langle a_n\rangle = \frac{\beta}{1-\alpha}. \tag{9}$$

Thus, the asymptotic mean, $\langle a\rangle_\infty$, also becomes independent of a_0 as $n \to \infty$ when $|\alpha| < 1$, showing that *the condition* $|\alpha| < 1$ *is also sufficient for homeostasis at the level of the mean*.

The deviation of the population mean from its asymptotic (homeostatic setpoint) value decreases exponentially over successive generations, with the exponential rate given by $\ln(1/\alpha)$. To see this, we use Eq. (9) in Eq. (8) and average over the probability distribution of a_0:

$$\frac{\langle\langle a_n\rangle\rangle - \langle a\rangle_\infty}{\langle\langle a_0\rangle\rangle - \langle a\rangle_\infty} = \alpha^n = e^{-n\ln(1/\alpha)}. \tag{10}$$

The preceding result from the fully stochastic picture presented here is reminiscent of the deterministic heuristic picture sketched to motivate how homeostasis is attained in the sizer–timer–adder framework (for an example, see Fig. 1 in [18] and the panels in the left column of Fig. 1).

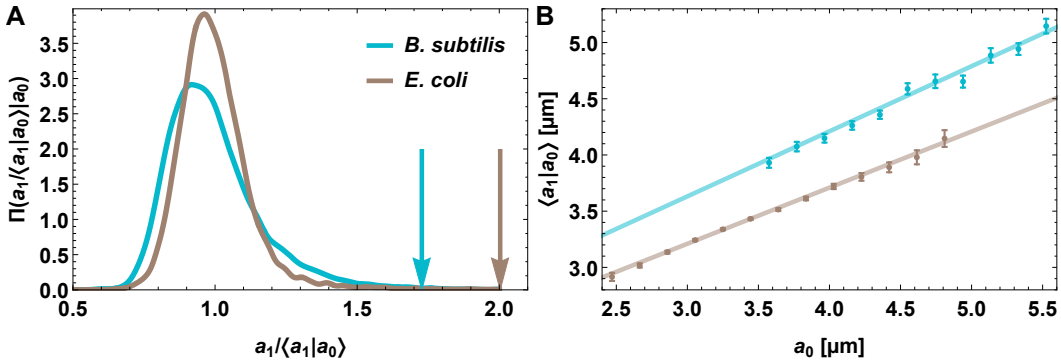

FIG. 5. **Input functions for predicting initial size distributions over successive generations, and depiction of the necessary and sufficient condition for homeostasis.** **(A)** Experimentally obtained mean rescaled conditional distribution of next generation's initial size, conditioned on current generation's initial size, plotted for two different species (distinguished by color). For each distribution, the value of $1/\alpha$ for the corresponding species is marked by an arrow. Both arrow placements agree with our theoretically derived necessary and sufficient condition for homeostasis: the upper limit of the support of the mean rescaled distribution should be less than or equal to the corresponding $1/\alpha$. Here, α is the slope of the line in (B). **(B)** Binned means of next generation's initial sizes (a_1), conditioned on the current generation's initial sizes (a_0), are plotted as functions of a_0 along with the corresponding linear fits for the two species in (A). These calibration curves, along with the distributions in (A), are used as inputs by our theoretical framework that can then predict the distributions of successive generation's initial sizes for cells starting from any given initial size (see Fig. 3 and 4). The data used are *E. coli* in glucose rich media (brown), and *B. subtilis* in LB media (blue), obtained from [10].

C. Beyond the mean: fluctuations impose additional requirements for attainment of stochastic homeostasis.

To include the effects of fluctuations, we find the variance of size in the nth generation by subtracting Eq. (8) from Eq. (7), and squaring and averaging the result:

$$\begin{aligned}\mathrm{Var}(a_n) = \beta^2 &\left(-\left(\frac{1-\alpha^n}{1-\alpha}\right)^2 + \frac{(\alpha+1)m_2\left(1-\left(\alpha^2 m_2\right)^n\right)}{(1-\alpha)(1-\alpha^2 m_2)} - \frac{(2m_2)\left(\alpha^n - \left(\alpha^2 m_2\right)^n\right)}{(1-\alpha)(1-\alpha m_2)}\right) \\ &+ 2a_0\beta\left(\frac{m_2\left(\alpha^n - \left(\alpha^2 m_2\right)^n\right)}{1-\alpha m_2} - \frac{\alpha^n(1-\alpha^n)}{1-\alpha}\right) + (a_0)^2\alpha^{2n}(m_2^n - 1).\end{aligned} \tag{11}$$

This result implies an additional condition that must be met for homeostasis of the variance, beyond the condition $|\alpha| < 1$ required for homeostasis of the mean. That additional condition is $|\alpha^2 m_2| < 1$, i.e., $|\alpha| < 1/\sqrt{m_2}$, needed for achieving both a_0-independence and a finite limit for

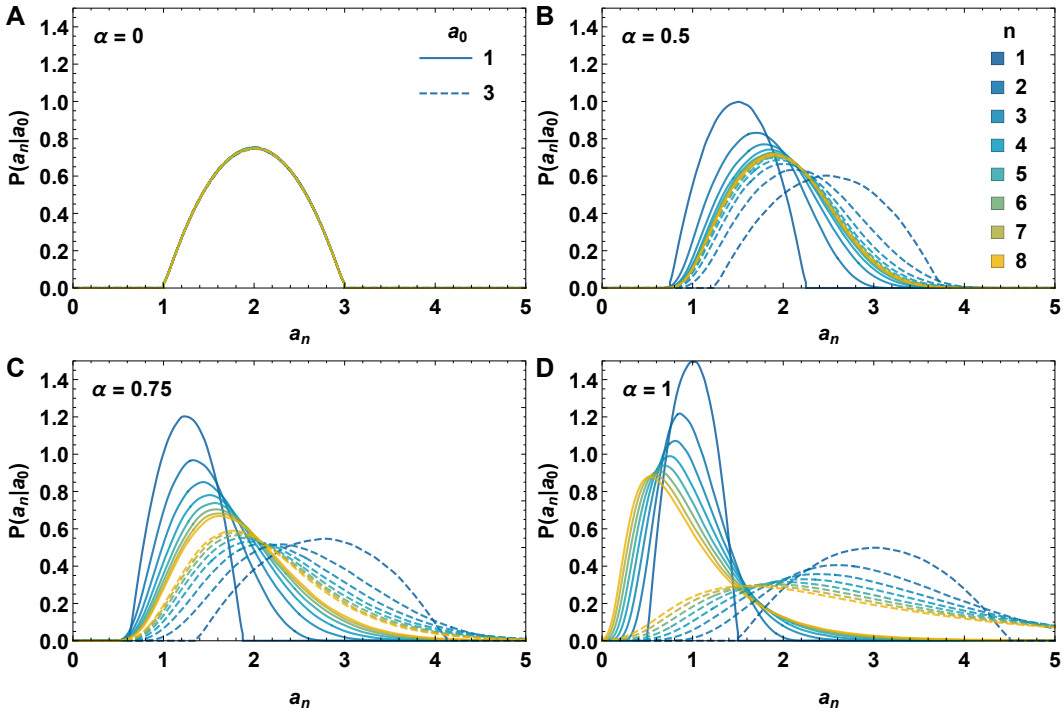

FIG. 6. **In-silico evolution of conditional initial size distributions over successive generations, for different values of α, for cells starting out with a given initial size.** These plots are made using the inter-generational kinematic theory presented in Eqs. (1), (3) and (4). Each plot shows the conditional initial size distributions over successive generations (different colors) for cells starting from two different initial sizes, $a_0 = 1$ (solid line) and 3 (dashed). $1/s_{max} \simeq 0.67$ for our hypothetical mean-rescaled distribution, $\Pi(s) = 3(1+\Delta-s)(s-1+\Delta)/(4\Delta^3)$ with $\Delta = 0.5$. The mean of next generation's initial size given the current generation's is $\langle a_1|a_0\rangle = \mu(a_0) = \alpha a_0 + 2(1-\alpha)$, ensuring that the homeostatic mean setpoint is 2 and allowing easy comparison between different α's. **(A)** $\alpha = 0$. Cells reach the homeostatic distribution within a generation. This corresponds to "sizer model" in the deterministic picture. **(B)** $\alpha = 0.5$. Cells reach the homeostatic distribution less rapidly than in (A). Since this value of α also satisfies the condition for homeostasis ($\alpha \leq 1/s_{max}$), all moments are expected to converge to finite values. This corresponds to "adder model" in the deterministic picture. **(C)** $\alpha = 0.75$. The condition for homeostasis is no longer satisfied. The lower moments, including mean, still converge, thus the overall location and shape of the distribution appears to converge. But the higher moments diverge, manifesting in the form of a long tail of the distribution. In the purely deterministic picture, this value of α would successfully result in homeostasis, but not in the exact stochastic framework we present here. **(D)** $\alpha = 1$. Here, even the mean does not converge to an a_0-independent value. This corresponds to the "timer model". Homeostasis is visibly absent.

$\mathrm{Var}(a_n)$ at large n:

$$\lim_{n\to\infty} \mathrm{Var}(a_n) = \frac{\beta^2 (m_2 - 1)}{(1-\alpha)^2 (1-\alpha^2 m_2)}, \text{ iff } |\alpha|, \alpha^2 m_2 < 1. \tag{12}$$

Comparing Eqs. (11) and (12), we note that the variance's approach to its asymptotic value is through a superposition of exponentials with rates given by $\ln(\alpha)$, $2\ln(\alpha)$ and $\ln(\alpha^2 m_2)$. Thus, it is not a simple exponential as in the case of the mean (Eq. (10)).

Since $m_2 \geq (m_1)^2 = 1$, the new condition $|\alpha| < 1/\sqrt{m_2} \leq 1$ is more restrictive than $|\alpha| < 1$. Therefore we have obtained a new condition on cell size homeostasis, which directly results from including the precise intergenerational stochastic behavior of cell growth and division.

D. General necessary and sufficient conditions for stochastic intergenerational homeostasis:

Our analysis of the asymptotic behavior of all raw moments yields the following general constraints for achieving size homeostasis, when $\alpha \geq 0$:

> For the kth moment of the initial cell size to display homeostasis (i.e., it has a finite a_0-independent limit after a large number of generations), all quantities $\{|\alpha^r m_r|\}$, for $1 \leq r \leq k$, need to be less than 1. In particular, since $m_1 = 1$, the first of these conditions is simply $|\alpha| < 1$, as can be derived in the deterministic picture. For size homeostasis to hold, however, all moments of the cell size need to display finite a_0-independent asymptotic behavior as the number of generations grows, thus requiring $|\alpha^k m_k| < 1$ for all $k \geq 1$.

We derive these constraints, and also an exact expression for the general raw moment of a_n, in the Appendix. There we also extend our analysis to $\alpha < 0$, showing an additional constraint needs to be met for homeostasis:

> When $\alpha < 0$, for size homeostasis to hold, i.e., all moments of the cell size display finite a_0-independent asymptotic behaviors as the number of generations grows and a_n never becomes negative, we require $|\alpha^k m_k| < 1$ for all $k \geq 1$, and $0 < a_0 < \beta/|\alpha|$.

Further, when above conditions are met, the asymptotic values of the raw moments of a_n are:

$$\lim_{n\to\infty} \left\langle (a_n)^k \right\rangle = \frac{\beta^k m_k}{1-\alpha^k m_k} \sum_{\tilde{v}_1,\ldots,\tilde{v}_{k-1}=0}^{\infty} \frac{k!}{q_k(\{\tilde{v}_1,\ldots,\tilde{v}_{k-1}\})} \prod_{j=1}^{k-1} \left(\alpha^j m_j\right)^{\tilde{v}_j}, \tag{13}$$

where $q_k(\{\tilde{v}_1,\ldots,\tilde{v}_{k-1}\})$ equals one times $(l+1)!$ for each set of consecutive zeros of length l in the sequence $\tilde{v}_i$'s. To see it in action, we now demonstrate equivalence of this formula to Eqs. (9) and (12) (cases $k=1$ and 2).

Setting $k=1$ in Eq. (13), the summation is absent and we get back Eq. (9) (recall: $m_1=1$). For comparing with the $k=2$ case, first we calculate the second raw moment from Eqs. (9) and (12) :

$$\lim_{n\to\infty}\left\langle (a_n)^2\right\rangle = \frac{(1+\alpha)\beta^2 m_2}{(1-\alpha)(1-\alpha^2 m_2)}. \tag{14}$$

Then, setting $k=2$ in Eq. (13) and using $m_1=1$:

$$\lim_{n\to\infty}\left\langle (a_n)^2\right\rangle = \frac{\beta^2 m_2}{1-\alpha^2 m_2}\sum_{\tilde{v}_1=0}^{\infty}\frac{2!}{q_2(\{\tilde{v}_1\})}(\alpha m_1)^{\tilde{v}_1}. \tag{15}$$

Now, $q_2(\{\tilde{v}_1\}) = (1+1)! = 2!$ when $\tilde{v}_1 = 0$ (a 'string' of zeros with length 1), and 1 otherwise. Using this,

$$\lim_{n\to\infty}\left\langle (a_n)^2\right\rangle = \frac{\beta^2 m_2}{1-\alpha^2 m_2}\left(1+2\sum_{\tilde{v}_1=1}^{\infty}\alpha^{\tilde{v}_1}\right) = \frac{\beta^2 m_2}{1-\alpha^2 m_2}\left(1+\frac{2\alpha}{1-\alpha}\right) = \frac{\beta^2 m_2}{1-\alpha^2 m_2}\times\frac{1+\alpha}{1-\alpha}. \tag{16}$$

This is the same as Eq. (14), demonstrating their equivalence.

In Fig. 7 we show the generational evolution of the first two moments of the cell size distribution, as the value of α is varied through successive homeostasis thresholds, showing successive levels of breakdown in homeostasis.

E. Restrictions on the mean rescaled distribution $\Pi(s)$ and the slope of the calibration curve α.

We recapitulate the general necessary and sufficient homeostasis condition previously derived:

$$|\alpha^k m_k| < 1 \text{ for all } k \geq 1. \tag{17}$$

(Recall that for $\alpha < 0$ we also require $0 < a_0 < \beta/|\alpha|$.) This relation places strong constraints on the mean-rescaled distribution $\Pi(s)$ and its relationship with α. We first show that $\Pi(s)$ must have finite support. In other words, the mean-rescaled intial size of the current generation given the initial size of the previous generation must have a maximum allowed value. To show this, we proceed as follows. Consider what happens when $\Pi(s)$ has infinite support. For *any* nonzero value $\alpha = \alpha_0$,

$$m_k = \int_0^{\infty} s^k \Pi(s)ds > \left(\frac{2}{|\alpha_0|}\right)^k \int_{2/|\alpha_0|}^{\infty}\Pi(s)ds. \tag{18}$$

Denoting the integral on the right hand side of the inequality sign by $\mathcal{I}$,

$$|\alpha_0|^k m_k > 2^k \mathcal{I} > 1 \qquad \text{for} \quad k > \log_2\left(\frac{1}{\mathcal{I}}\right). \tag{19}$$

Since the preceding condition on k can always be satisfied because the infinite support for Π implies $\mathcal{I} > 0$, we conclude that all homeostasis conditions, Eq. (17), cannot be satisfied for any arbitrary non-zero value of α, however small. *Hence we conclude that* $\Pi(s)$ *needs to have finite support if cell size homeostasis is to be achieved and sustained.* In turn, the implication is that the support of $\Pi(s)$ extends to a maximum value, $s_{\max}$.

We now proceed to determine the allowed values of α that satisfy the homeostasis conditions Eq. (17), when $\Pi(s)$ has finite support. Since

$$|\alpha|^k m_k = |\alpha|^k \int_0^{s_{max}} s^k \Pi(s) ds < (|\alpha| s_{max})^k, \tag{20}$$

the homeostatis conditions $|\alpha|^k m_k < 1$ will always be satisfied for $|\alpha| \leq 1/s_{\max}$ (provided Π is not a Dirac Delta function, in which case the necessary and sufficient condition for homeostasis simply reduces to $\alpha < 1$). In the complementary scenario when $|\alpha| > 1/s_{\max}$, consider any value b satisfying $1 < b < |\alpha| s_{\max}$. Then,

$$|\alpha|^k m_k > |\alpha|^k \left(\frac{b}{|\alpha|}\right)^k \int_{b/|\alpha|}^{s_{\max}} \Pi(s) ds. \tag{21}$$

Denoting by $\mathcal{I}$ the integral on the right hand side of the inequality,

$$|\alpha|^k m_k > b^k \mathcal{I} > 1 \quad \text{for } k > \log_b\left(\frac{1}{\mathcal{I}}\right). \tag{22}$$

Since the condition on k can always be satisfied by large enough values of k, all homeostasis conditions cannot hold for $|\alpha| > 1/s_{\max}$. Combining the above proofs that the homeostasis conditions Eq. (17) can always be satisfied when $|\alpha| \leq 1/s_{\max}$, and a subset of those conditions will always be violated when $|\alpha| > 1/s_{\max}$, we have derived a remarkable condition:

> Cell size homeostasis is obtained for all α satisfying $|\alpha| \leq 1/s_{\max}$, where $s_{\max}$ is the upper limit of the support of the universal mean-rescaled distribution, $\Pi(s)$, independent of other details of the shape of $\Pi(s)$. When $\alpha < 0$ we also require $0 < a_0 < \beta/|\alpha|$.

To gain physical insight into this simple limit on α, we note that it can be rationalized by requiring that the coefficient of a_0 in the expression for a_n, Eq. (7), decrease as $n \to \infty$. Since increasing n by one introduces one extra factor of the form αs, which can have a maximum value of $\alpha s_{\max}$, the coefficient would be non-increasing as n increases, as long as $\alpha s_{\max} \leq 1$. But provided the

distribution of s is not a Dirac Delta, as n grows large, the probability of choosing the maximum value of s each draw decreases to zero, ensuring the coefficient of a_0 decreases. The preceding heuristic argument is also rigorously borne out by our exact probabilistic analysis. If Π is a Dirac Delta function, the only possible value of s is 1, and the condition for homeostasis is $\alpha < 1$, which again ensures that the coefficient of a_0 decreases over successive generations.

Fig. 5 A shows a comparison between the experimentally measured values of $1/\alpha$ and $s_{\max}$ in *E. coli* and *B. subtilis*. From this comparison it is clear that our predicted general condition for homeostasis, $s_{\max} \leq 1/\alpha$, is satisfied by experimentally observed *E. coli* and *B. subtilis* growth and division dynamics, under the growth conditions shown.

To explicate and visualize the breakdown of homeostasis as the theoretical condition $|\alpha| \leq 1/s_{\max}$ is violated, we take advantage of numerical simulations based on the theoretical framework proposed here. In Fig. 6 we show the simulated distributions of initial sizes over successive generations for a range of α values for hypothetical cells. These include cases for which the condition for homeostasis is not satisfied. Fig. 7 shows the generational evolution of the first two moments of cell size, as α passes through different homeostasis thresholds. When $|\alpha| < 1/s_{\max}$, the cell size distribution converges to a well-defined distribution with a short tail. When $1 > |\alpha| > 1/s_{\max}$, the mean and some lower moments converge to values independent of the starting generation size a_0, but higher moments do not, serving as indicators of the expected breakdown of homeostasis. This breakdown is reflected in the distributions developing long tails after sufficiently many generations have elapsed. Finally, when $\alpha > 1$, we predict that even the mean cannot attain an a_0-independent value (Fig. 7 G); in confirmation, the distributions become qualitatively different for distinct a_0 values, showing a spectacular breakdown of homeostasis (see Fig. 6).

IV. CONCLUDING REMARKS

In summary, using the "emergent simplicities" reported in [1] as essential building blocks, we have written down a first-principles-based theoretical framework to characterize the fully stochastic dynamics of cell size homeostasis of bacterial cells in balanced growth conditions. Prior works typically describe homeostasis building on a quasi-deterministic scheme (the sizer–timer–adder framework), with noise "added on top" in adhoc ways motivated by analytical or computational tractability [19, 23-26]. The deterministic condition for homeostasis obtained from the sizer-timer-adder framework, $\alpha < 1$, applies literally only to the mythical "average" cell. This heuristic falls short when the fully stochastic picture is considered. In contrast, in this work we

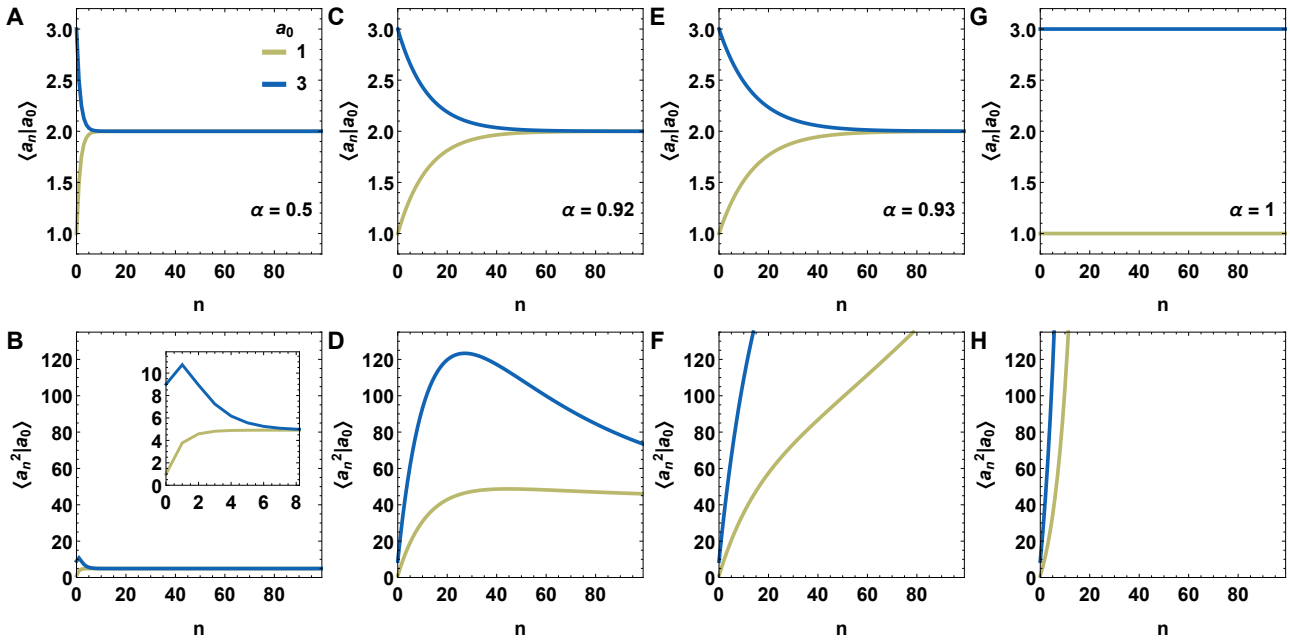

FIG. 7. **Demonstrating the conditions for convergence of moments.** These plots are made using the expressions in Eqs. (8) and (11). Each plot shows the evolution of the mean (left panels) or the second moment (right panels) of the conditional initial size distributions with generation n, for cells starting from two different initial sizes, $a_0 = 1$ (orange) and 3 (brown). $1/s_{max} \simeq 0.526$ and $1/\sqrt{m_2} \simeq 0.928$ for our hypothetical mean rescaled distribution, $\Pi(s) = 3(1+\Delta-s)(s-1+\Delta)/(4\Delta^3)$ with $\Delta = 0.9$. The mean of next generation's initial size given the current generation's is $\langle a_1|a_0\rangle = \mu(a_0) = \alpha a_0 + 2(1-\alpha)$, ensuring that the homeostatic mean setpoint is 2 and allowing easy comparison between different α's. **(A,B)** $\alpha = 0.5$. α satisfies the necessary and sufficient condition for homeostasis ($\alpha \leq 1/s_{max}$), thus all moments converge to their homeostatic limits. (B) inset is a zoomed in version of (B), showing rapid convergence. **(C,D)** $\alpha = 0.92$. The general condition for homeostasis is not satisfied and higher moments do not converge. However, since $\alpha < 1/\sqrt{m_2} < 1$, the second moment and mean converge, showing partial attainment of homeostasis in their case. **(E,F)** $\alpha = 0.93$. The value of α is greater than $1/\sqrt{m_2}$, thus variance does not converge. However, since $\alpha < 1$, the mean still converges. **(G,H)** $\alpha = 1$. Even the condition for a_0-independent convergence of mean is no longer satisfied, resulting in complete breakdown of homeostasis. In summary, these plots demonstrate that as α increases, the rates of convergence of moments slow down, and as it crosses $1/s_{\max}$ followed by the limits $(m_k)^{-1/k}$, determined by the moments m_k of Π as derived in the main text, the corresponding moments no longer converge and homeostasis is broken.

have shown that the necessary and sufficient condition for homeostasis, when considering the data-informed [1] inherently stochastic formulation of the problem, is $|\alpha| \leq 1/s_{\max}$, where $s_{\max} > 1$ is the upper limit of the support of the mean-rescaled distribution Π, introduced in Eq. (3). By reanalyzing published experimental data in *E. coli* and *B. subtilis* (Figs. 3 and 4 respectively), we have

not only shown compelling data-theory matches, but also that this new homeostatis condition, $|\alpha| \leq 1/s_{\max}$, is indeed satisfied by these data (see downward pointing arrows in Fig. 5). While the theoretical framework of intergenerational size dynamics presented here is data-informed and accurately describes the kinematics of cell size homeostasis, it is mechanistically agnostic. In complementary work [27] we will address possible architectural underpinnings of the observed intergenerational scaling law leading to these precision kinematics.

Our theoretical framework is built on two experimentally-observed emergent simplicities [1]: first, the inter-generational initial size dynamics are Markovian, and second, the conditional distribution of the next generation's initial size, conditioned on the current generation's initial size, when rescaled by its mean value, results in a distribution (Π) that is invariant of this generation's initial size. Since the mean-rescaled distribution changes from growth condition to growth condition, even for the same organism, as shown in [1], how the results generalize to time-varying growth conditions remains to be seen.

AUTHOR CONTRIBUTIONS

K.J., R.R.B., and S.I.-B. conceived of and designed the research; K.J. developed the theoretical framework under the guidance of S.I.-B.; K.J., R.R.B. and S.I.-B. performed analytic calculations; K.J. performed data analyses and simulations under the guidance of R.R.B. and S.I.-B.; K.J., R.R.B., and S.I.-B. wrote the paper; S.I.-B. supervised the research.

ACKNOWLEDGEMENTS

We thank Purdue University Startup funds, Purdue Research Foundation, the Purdue College of Science Dean's Special Fund, and the Showalter Trust for financial support. K.J. and S.I.-B. acknowledge support from the Ross-Lynn Fellowship award. We are grateful to Charles Wright for insightful discussions and detailed feedback on the manuscript. We thank Suckjoon Jun and Fangwei Si for generously sharing the previously published single-cell datasets that are utilized in this study, along with the details of their methodology and analysis.

[1] K. Joshi, C. S. Wright, K. F. Ziegler, E. M. Spiers, J. T. Crosser, S. Eschker, R. R. Biswas, and S. Iyer-Biswas, bioRxiv (2023), 10.1101/2023.01.18.524627.

[2] W. B. Cannon, in *À Charles Richet: ses amis, ses collègues, ses élèves (in French)*, edited by A. Pettit (Les Éditions Médicales, Paris, 1926) p. 91.

[3] W. B. Cannon, *The Wisdom of the Body* (W.W. Norton & Company, Inc., New York, 1932).

[4] J. D. Hardy, Harvey Lectures **49**, 242 (1953–1954).

[5] G. E. Billman, Frontiers in Physiology **11** (2020), 10.3389/fphys.2020.00200.

[6] P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun, Current Biology **20**, 1099 (2010).

[7] J. R. Moffitt, J. B. Lee, and P. Cluzel, Lab on a Chip **12**, 1487 (2012).

[8] T. M. Norman, N. D. Lord, J. Paulsson, and R. Losick, Nature **503**, 481 (2013).

[9] 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, Proceedings of the National Academy of Sciences of the United States of America **111**, 15912 (2014).

[10] S. Taheri-Araghi, S. Bradde, J. T. Sauls, N. S. Hill, P. A. Levin, J. Paulsson, M. Vergassola, and S. Jun, Current biology **25**, 385 (2015).

[11] C. S. Wright, S. Banerjee, S. Iyer-Biswas, S. Crosson, A. R. Dinner, and N. F. Scherer, Scientific reports **5**, 1 (2015).

[12] F. Jafarpour, C. S. Wright, H. Gudjonson, J. Riebling, E. Dawson, K. Lo, A. Fiebig, S. Crosson, A. R. Dinner, and S. Iyer-Biswas, Physical Review X **8**, 021007 (2018).

[13] T. W. Spiesser, C. Müller, G. Schreiber, M. Krantz, and E. Klipp, The FEBS journal **279**, 4213 (2012).

[14] M. Deforet, D. van Ditmarsch, and J. B. Xavier, Biophysical Journal **109**, 521 (2015).

[15] J. T. Sauls, D. Li, and S. Jun, Current Opinion in Cell Biology **38**, 38 (2016).

[16] M. M. Logsdon, P.-Y. Ho, K. Papavinasasundaram, K. Richardson, M. Cokol, C. M. Sassetti, A. Amir, and B. B. Aldridge, Current Biology **27**, 3367 (2017).

[17] L. Willis and K. C. Huang, Nature Reviews Microbiology **15**, 606 (2017).

[18] S. Jun and S. Taheri-Araghi, Trends in microbiology **23**, 4 (2015).

[19] A. Amir, Physical Review Letters **112**, 208102 (2014).

[20] J. Lin and A. Amir, Cell systems **5**, 358 (2017).

[21] M. Campos, I. V. Surovtsev, S. Kato, A. Paintdakhi, B. Beltran, S. E. Ebmeier, and C. Jacobs-Wagner, Cell **159**, 1433 (2014).

[22] F. Si, G. L. Treut, J. T. Sauls, S. Vadia, P. A. Levin, and S. Jun, Current Biology **29**, 1760 (2019).

[23] L. Susman, M. Kohram, H. Vashistha, J. T. Nechleba, H. Salman, and N. Brenner, Proceedings of the National Academy of Sciences **115**, E5679 (2018).

[24] W. F. Marshall, Annual review of biophysics **45**, 49 (2016).

[25] M. B. Ginzberg, R. Kafri, and M. Kirschner, Science **348**, 1245075 (2015).

[26] S. Sanders, K. Joshi, P. A. Levin, and S. Iyer-Biswas, PLoS genetics **19**, e1010505 (2023).

[27] K. Joshi, R. R. Biswas, and S. Iyer-Biswas, biorXiv (2023).