CONTENTS

I. SINGLE-CELL STOCHASTIC HELMSTETTER-COOPER MODEL

A. Deterministic Helmstetter-Cooper model

Here we model the growth of individual *Escherichia coli* cells. Based on experimental measurements, we posit that a single-cell of size S elongates exponentially [S12–S14]:

$$\frac{\mathrm{d}S}{\mathrm{d}t}(t) = \lambda S(t), \tag{1}$$

where λ is the growth rate. For rod-shaped bacteria such as *E. coli* the width is nearly constant, so the size means either the length or volume of an individual cell (to a proportionality constant). In order to determine the division timing, we adopt the Helmstetter-Cooper model, which couples the replication of the chromosome to the cell division (see **Helmstetter-Cooper model**). This model proposes that cell division occurs after a prescribed time has elapsed since replication initiation. This time is the duration of the cell cycle, denoted τ_{cyc} [S14,S15]. It comprises the time required to fully replicate the chromosome, known as the C-period, and the time following replication termination until division, known as the D-period. Hence $\tau_{\mathrm{cyc}} = C + D$. In other words, an initiation event commits the cell to division after the duration of one cell cycle. In bacteria, multiple cell cycles can overlap. This occurs when the cell cycle duration is larger than the generation time: $\tau_{\mathrm{cyc}} > \tau$. In order to maintain its DNA content, a cell still needs to initiate chromosome replication once per generation (see **Helmstetter-Cooper model**). At this stage, the problem of division timing is thus transfered to the problem of initiation timing. This is solved by considering Donachie's conjecture [S16], which is that a new round of replication initiates at a fixed size per number of origins of replication (*oriC*), denoted s_{i}.

In summary, a model for the growth of a single cell is completely parametrized by the following "physiological variables":

- the growth rate λ;
- the cell cycle duration τ_{cyc};
- the initiation size per *oriC* s_{i}.

This model predicts that the cell at division is given by:

$$S_{\mathrm{d}} = s_{\mathrm{i}} \exp(\lambda \tau_{\mathrm{cyc}}), \tag{2}$$

which was indeed verified experimentally at the population level [S1]. Single-cell measurement of the initiation timing revealed that the initiation size is indeed tightly regulated. In fact, s_{i} is the physiological variable with the narrowest distribution (**Comparison of experimental measurements with simulations of the Helmstetter-Cooper model (1)**, **Comparison of experimental measurements with simulations of the Helmstetter-Cooper model (2) Comparison of experimental measurements with simulations of the Helmstetter-Cooper model (3)** and **Comparison of experimental measurements with simulations of the Helmstetter-Cooper model (4)**).

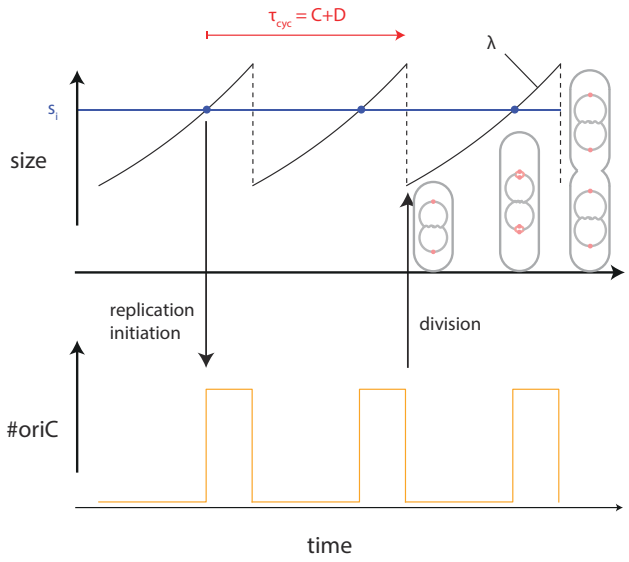
B. Stochastic Helmstetter-Cooper model

In order to account for the experimental fluctuations in the physiological variables between individual cells, we treat them as stochastic variables. That is to say, at cell birth, the growth rate, the cell cycle duration and the unit cell are drawn from independent Gaussian distributions (**Stochastic Helmstetter-Cooper model**). Once the physiological variables are determined, a single-cell follows the deterministic growth as described in the previous section.

In reality, we should expect that values taken by the physiological variables are not independent from each other. For example, the growth rate and the cell cycle duration are anti-correlated [S14]. Another example comes from the observation that cells growing faster than the average growth rate also tend to produce fast growing daughter cells. We call the former type of correlation "cross-correlation" and the latter ones "autocorrelation". In order to take into account those correlations, we modify the way physiological variables are determined at cell birth. Let us denote the physiological state of a single cell by the three-dimensional vector $\mathbf{x}^n = (\lambda^n, \tau^n_{\mathrm{cyc}}, S^n_0)$. At cell birth, the physiological state of cell $n+1$ is determined according to the rule:

$$\mathbf{y}^{n+1} = D \cdot \mathbf{y}^n + A \cdot \mathbf{z}, \qquad \text{where} \quad \mathbf{y}^n = \mathbf{x}^n - \langle \mathbf{x}^n \rangle. \tag{3}$$

Here, $D = \mathrm{Diag}(\alpha_1, \alpha_2, \alpha_3)$ is a diagonal matrix enforcing mother/daughter correlations, A is a real symmetric matrix enforcing the cross-correlations (to be determined below) and $\mathbf{z}$ is a vector of three independent unit Gaussian variables.



Helmstetter-Cooper model. In bacteria, several rounds of chromosome replication can overlap when the duration of the cell cycle τ_{cyc}, including both C- and D-periods, is larger than the generation time.

Equation (3) defines a discrete stochastic process. Being a sum of Gaussian random vectors, $\mathbf{y}^n$ has a (multivariate) Gaussian distribution. Furthermore, it can be shown that it converges toward the Gaussian distribution:

$$\begin{aligned} &\mathrm{P}(\mathbf{y}^n) \xrightarrow[n\to\infty]{} \mathrm{P}(\mathbf{y}), \\ &\mathrm{P}(\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^3 \det \Sigma}} \exp\left(-\frac{1}{2}\mathbf{y}^T \cdot \Sigma^{-1} \cdot \mathbf{y}\right), \end{aligned} \tag{4}$$

whose covariance matrix $\Sigma = [\sigma_{ij}]_{i,j}$ has the elements:

$$\sigma_{ij} = \frac{1}{1-\alpha_i\alpha_j} \sum_{k=1}^{3} a_{ik}a_{kj}. \tag{5}$$

The previous relation can be inverted to express the matrix A as a function of the covariance matrix of the limiting distribution. We obtain:

$$A = \sqrt{B}, \qquad \text{with} \qquad B = [\sigma_{ij}(1-\alpha_i\alpha_j)]_{i,j}. \tag{6}$$

In short, with an appropriate choice of the matrix A, the stochastic process in Equation (3) will sample random physiological vectors distributed according to a Gaussian distribution with the desired covariance matrix Σ, in agreement with experimental measurements. For the stochastic process in Equation (3), the variances of the physiological variables are given by:

$$\langle \delta x_i^2 \rangle = \sigma_{ii}, \tag{7}$$

and the cross-correlations between the physiological variables are expressed as:

$$\langle \delta x_i \cdot \delta x_j \rangle = \sigma_{ij}. \tag{8}$$

Note that the Pearson correlation coefficients between the physiological variables are expressed as:

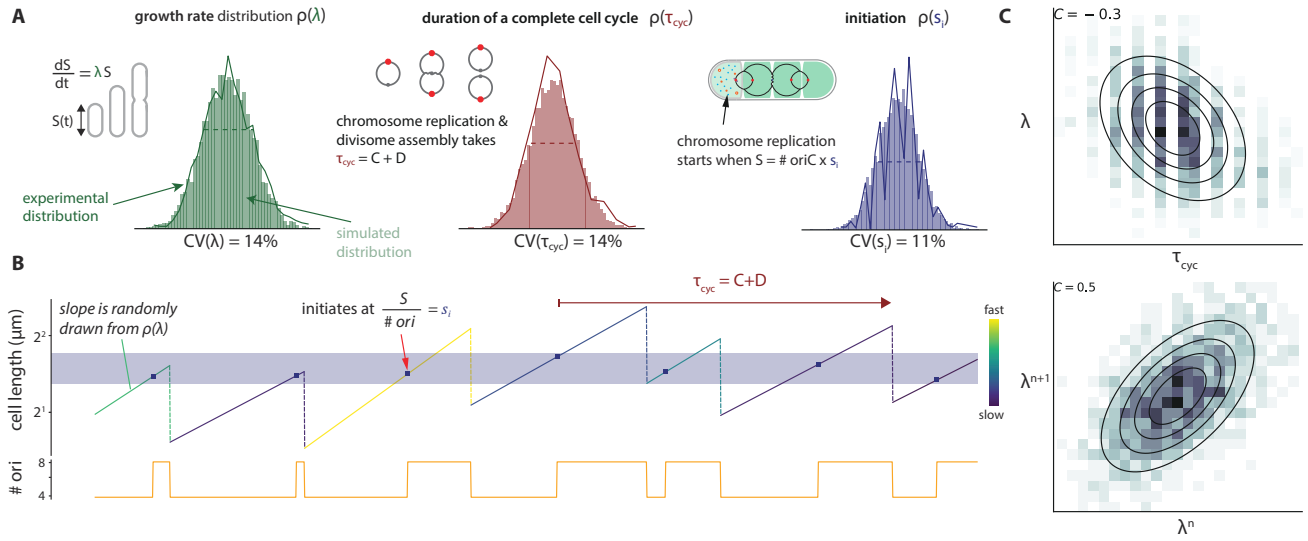
$$r(x_i, x_j) = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}}. \tag{9}$$

Concerning mother/daughter correlations, it can be shown that:

$$\langle \delta x_i^n \cdot \delta x_j^{n+1} \rangle = \alpha_i \sigma_{ij}^n \xrightarrow[n\to\infty]{} \alpha_i \sigma_{ij}, \tag{10}$$

with the corresponding Pearson correlations:

$$r(x_i^n, x_j^{n+1}) = \alpha_i r(x_i, x_j). \tag{11}$$



Stochastic Helmstetter-Cooper model. **(A)** At each new generation, the physiological variables are drawn from a multivariate Gaussian distribution with means and variances matching the experimental values. **(B)** In this example, cell division is coupled to an initiation event happening in the grandmother cell: there is three overlapping cell cycles. **(C)** Examples of negative cross-correlation between λ and τ_{cyc} and of positive mother/daughter correlation for λ.

In particular, the autocorrelation Pearson coefficients are:

$$r(x_i^n, x_i^{n+1}) = \alpha_i. \tag{12}$$

In summary, we draw the physiological variables at each new generation according to Equation (3). This stochastic process is parametrized by the experimental values measured for the means, variances, cross-correlations and autocorrelations of the three physiological variables.

C. Implementation

The implementation used to generate a lineage of cells according to the stochastic Helmstetter-Cooper model is described in Algorithms 1 to 3. The simulations rely on the following input: (i) the means $\langle\lambda\rangle$, $\langle\tau_{\mathrm{cyc}}\rangle$ and $\langle s_{\mathrm{i}}\rangle$; (ii) the variances $\langle\delta\lambda^2\rangle$, $\langle\delta\tau_{\mathrm{cyc}}^2\rangle$ and $\langle\delta s_{\mathrm{i}}^2\rangle$; (iii) the Pearson cross-correlation coefficients $r_{ij} := r(x_i, x_j)$ defined above; and (iv) the Pearson autocorrelation coefficients α_i defined above. These inputs can be measured experimentally and were used to set the joint-probability distribution of the physiological variables. Unless specified otherwise, we generated a single lineage of 10 000 cells in one simulation.

D. Comparison with experiments

We performed simulations according to the stochastic Helmstetter-Cooper for several experimental conditions (**Comparison of experimental measurements with simulations of the Helmstetter-Cooper model (1)**, **Comparison of experimental measurements with simulations of the Helmstetter-Cooper model (2)**, **Comparison of experimental measurements with simulations of the Helmstetter-Cooper model (3)**, **Comparison of experimental measurements with simulations of the Helmstetter-Cooper model (4)**, **Comparison of experimental measurements with simulations of the Helmstetter-Cooper model (4)** and **Comparison of experimental measurements with simulations of the Helmstetter-Cooper model (5)**). Overall, the agreement between experiments and simulations was good. For non-overlapping cell cycles conditions, most of the distributions for the cell size at birth S_{b}, the cell size at division S_{d}, the generation time τ and the added size between divisions $\Delta_{\mathrm{d}} = S_{\mathrm{d}} - S_{\mathrm{b}}$ were well reproduced. This observation was less true for overlapping cell cycles. For example in the latter case, the distribution of division size predicted by stochastic Helmstetter-Cooper model was systematically broader than the experimental one. In general, we found that the predicted correlations between variables were in good agreement with the experimental measurements.

The simulations of the stochastic Helmstetter-Cooper model reproduced well the experimental behavior for cell size homeostasis, namely the adder behavior. In **Comparison of the experimental adder behavior with simulation of the**

Algorithm 1: Stochastic Helmstetter-Cooper simulation.

Input:

- Means: $\mu_i = \langle x_i \rangle$, $i = 1, 2, 3$.
- Variances: $\sigma_i^2 = \langle \delta x_i^2 \rangle$, $i = 1, 2, 3$.
- Pearson cross-correlations: r_{ij}, for $i, j = 1, 2, 3$.
- Pearson autocorrelations: α_i, $i = 1, 2, 3$.
- Number of generations to simulate: N.

Output: Lineage of N cells.

▷ *Initialize model random vector generator*
Define the covariance matrix Σ
for $i = 1$ **to** 3 **do**
  $\sigma_{ii} = \sigma_i^2$
  for $j = 1$ **to** $i - 1$ **do**
    $\sigma_{ij} = \sigma_{ji} = r_{ij} \cdot \sqrt{\sigma_{ii} \cdot \sigma_{jj}}$
  end
end
Define the matrix S according to Equation (6)
Define the matrix $D = \mathrm{Diag}(\alpha_1, \alpha_2, \alpha_3)$
Rn = `ModelRandomVecGen(`μ, S, D`)` ▷ *implementation of Equation* (3)
▷ *Initialize time, size and physiological variables*
$t = 0$, $v = 1$
$x = Rn(\mu)$
▷ *Initialize cell cycles*
$P = 0$ ▷ $\log_2(N_{\mathrm{ori}})$
Allocate memory for $\{A_i\}_{i=1}^{P_{\max}}$ ▷ *array of initiation times*
for $i = 1$ **to** $P_{\max}$ **do** $A_i = NULL$
▷ *Simulate lineage of cells*
`GenerateLineage(`x,t,v,A,P,N,Rn`)`

stochastic Helmstetter-Cooper model, we illustrate how the adder behavior converges toward the experimental value when cross-correlations and autocorrelations are added to the model. Clearly, in the absence of cross-correlations and/or autocorrelations, the behavior deviates from the experimental measurements. This suggests that such correlations are essential for the Helmstetter-Cooper model to reproduce the experimental cell size homeostasis behavior.

E. Co-regulation hypothesis of chromosome replication and division

The Helmstetter-Cooper is often interpreted as to impose a fixed cell cycle duration, τ_{cyc}. If in addition the growth rate is fixed, Equation (2) implies that the cell size at division S_{d} is proportional to the cell size per origin of replication s_{i} at the initiation event that led to the division. The response of division sizes is then linked to the response of initiation size to perturbations. In particular, their Pearson autocorrelation coefficients are equal:

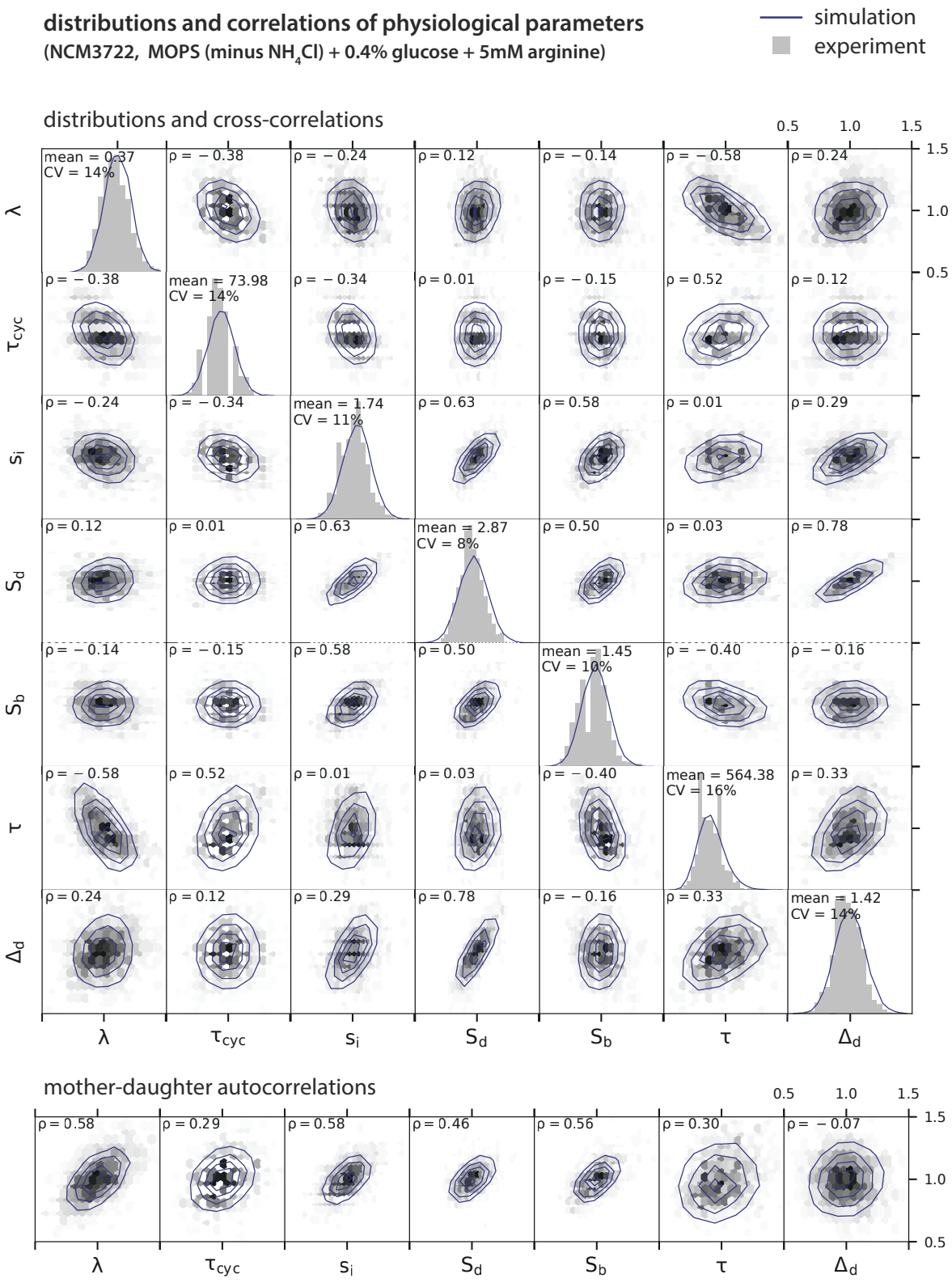
$$\rho(S_{\mathrm{d}}^{n+1}, S_{\mathrm{d}}^{n}) = \rho(s_{\mathrm{i}}^{n+1}, s_{\mathrm{i}}^{n}). \tag{13}$$

In fact, the Pearson correlation coefficient for the adder at division (resp. at initiation) is uniquely related to the Pearson autocorrelation coefficient of the cell size at division (resp. at initiation). Therefore, the previous equality implies:

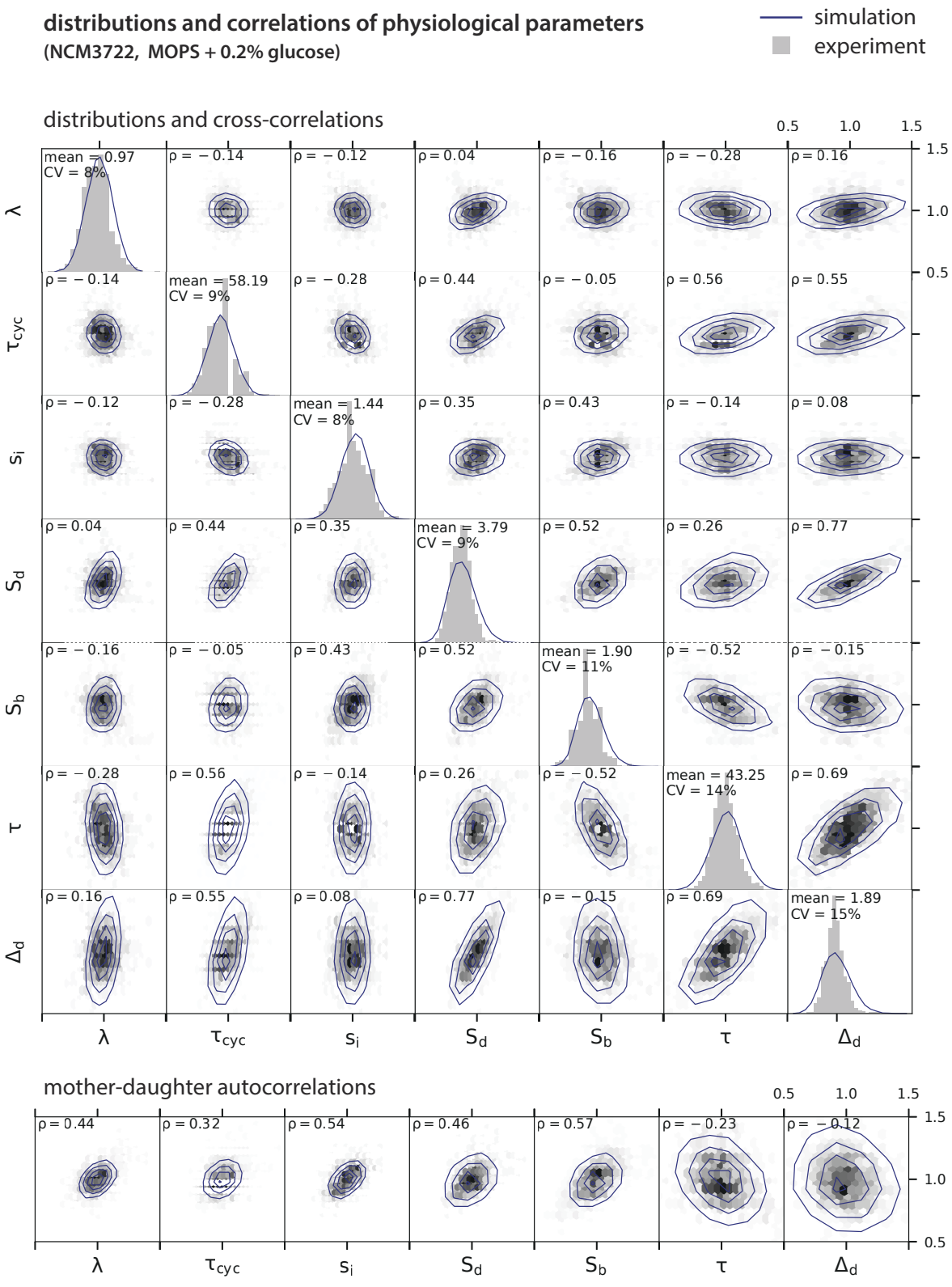
$$\rho(\Delta_{\mathrm{d}}, S_{\mathrm{b}}) = \rho(\delta_{\mathrm{i}}, s_{\mathrm{i}}), \tag{14}$$

where $\Delta_{\mathrm{d}} = 2S_{\mathrm{b}}^{n+1} - S_{\mathrm{b}}^{n}$ and $\delta_{\mathrm{i}} = 2s_{\mathrm{i}}^{n+1} - s_{\mathrm{i}}^{n}$.

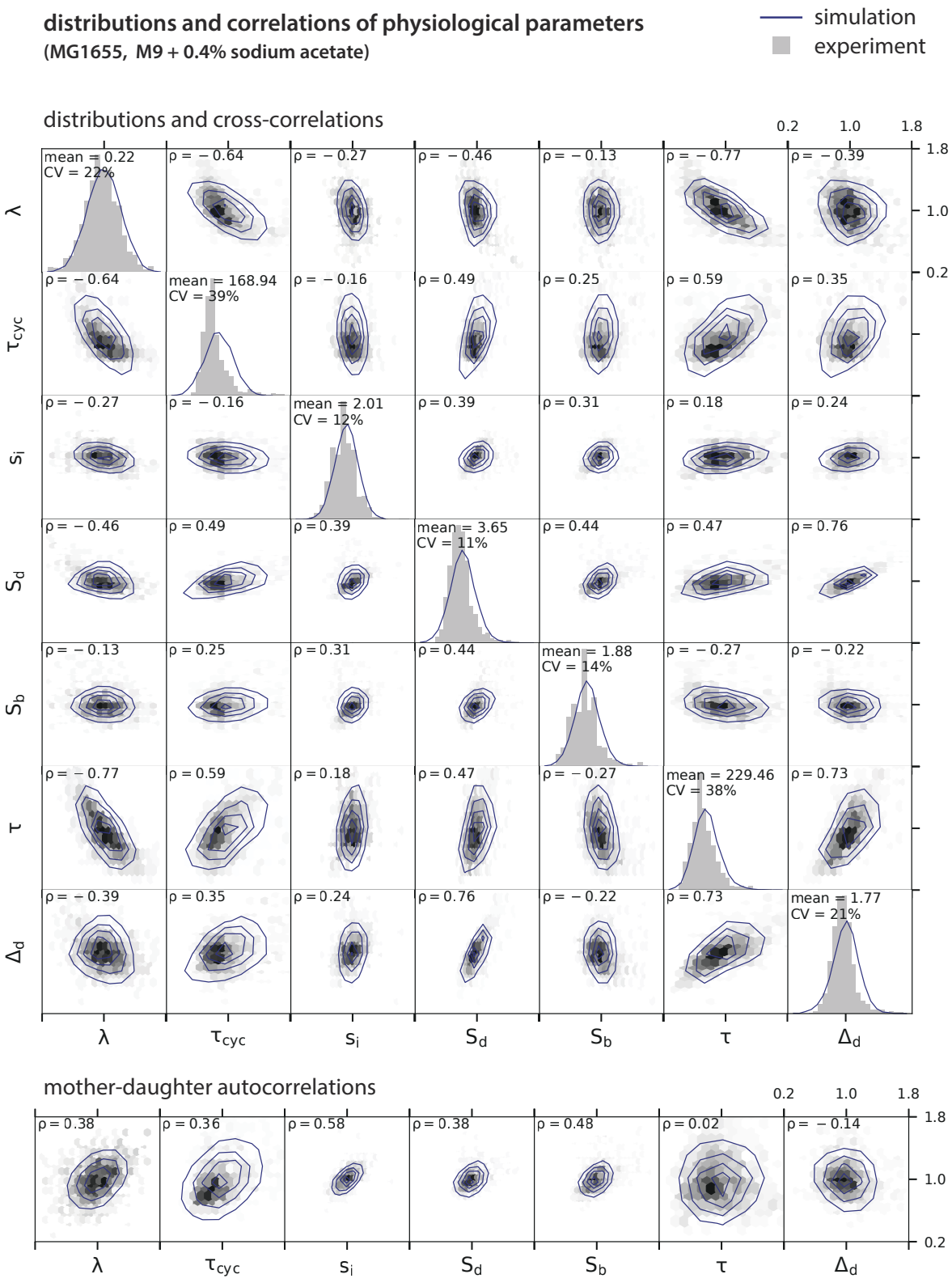
When relaxing the condition that τ_{cyc} is fixed, Equation (14) no longer holds. In **Co-regulation between division adder and initiation adder in the stochastic Helmstetter-Cooper model**, we investigated whether the two types of adder are still equivalent when τ_{cyc} is allowed to fluctuate. We started from reference values for the parameters and then varied each of the coefficient-of-variations (CVs), cross-correlation and autocorrelation away from their reference value. More accurately, the CVs were varied between 1%-30% and the Pearson correlations between -0.9 and $+0.9$. However, we did not perturb $\rho(\lambda, s_{\mathrm{i}})$ and $\rho(\tau_{\mathrm{cyc}}, s_{\mathrm{i}})$ because experimental data indicate that s_{i} should remain very robust to perturbations [S1].



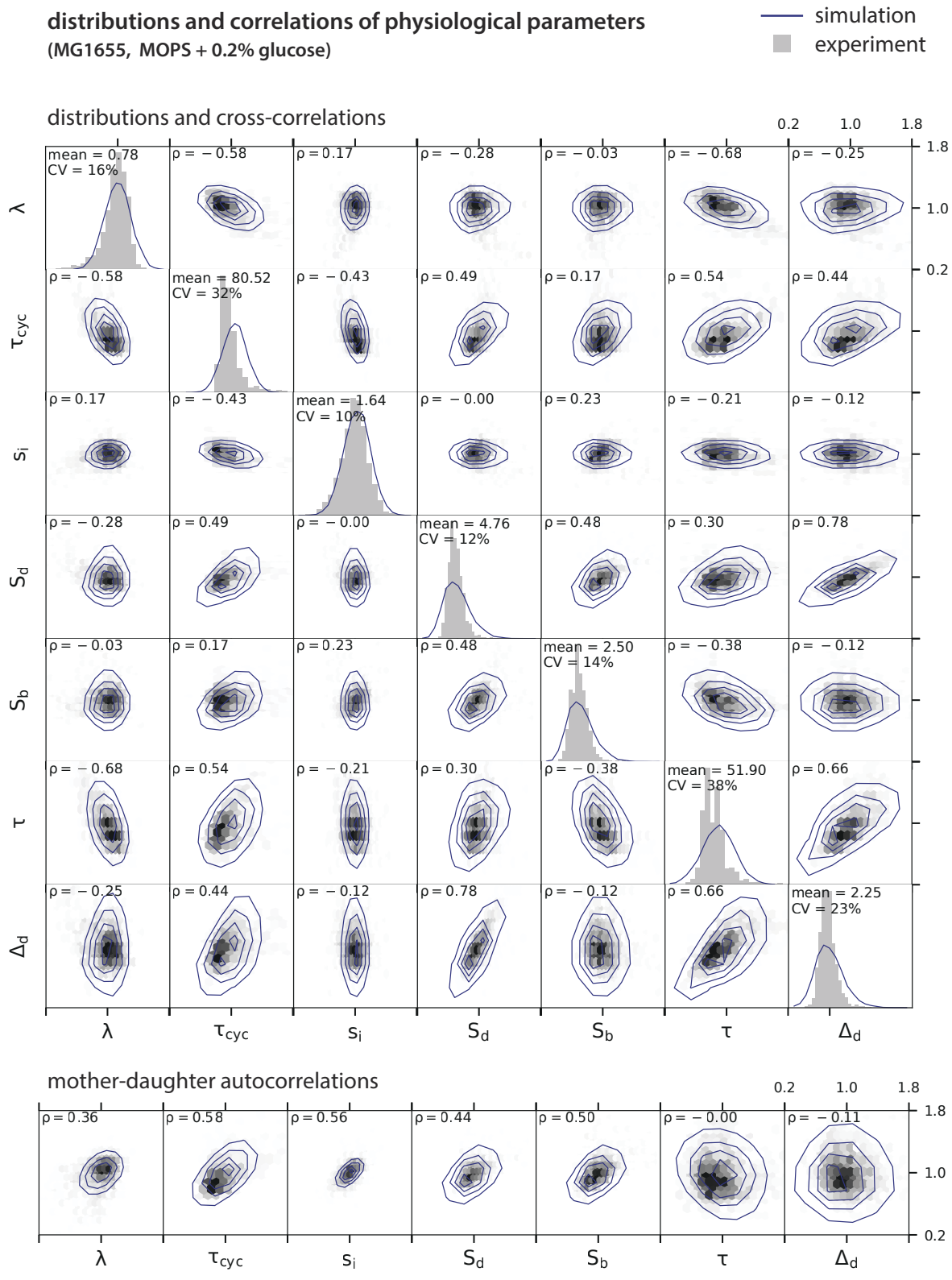
Comparison of experimental measurements with simulations of the Helmstetter-Cooper model (1). NCM3722 strain with no overlapping cell cycles as in Figure 1D.



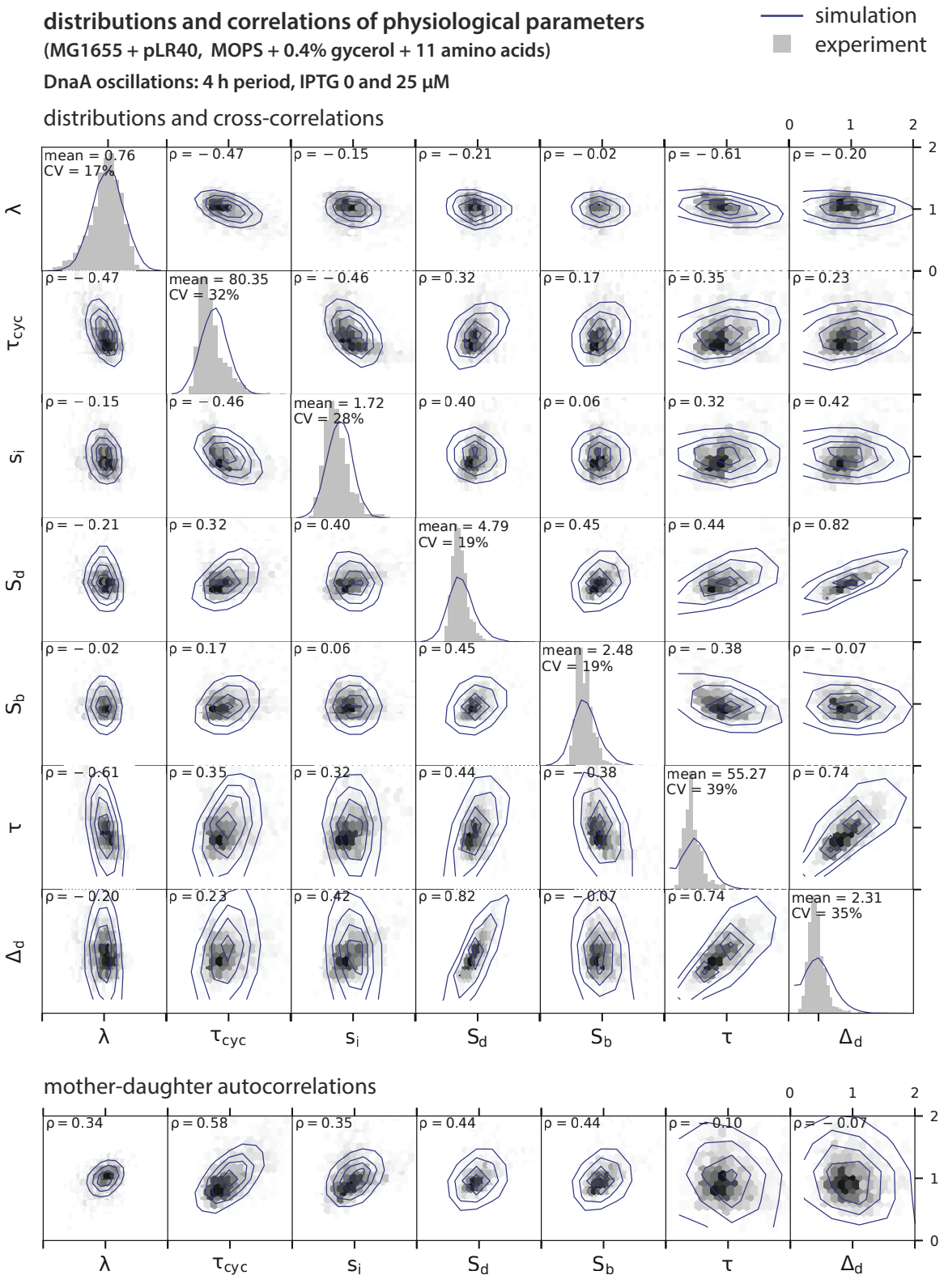
Comparison of experimental measurements with simulations of the Helmstetter-Cooper model (2). NCM3722 strain with two overlapping cell cycles as in Figure 1D.



Comparison of experimental measurements with simulations of the Helmstetter-Cooper model (3). MG1655 strain with no overlapping cell cycles as in Figure 1D.



Comparison of experimental measurements with simulations of the Helmstetter-Cooper model (4). MG1655 strain with two overlapping cell cycles as in Figure 1D.



Comparison of experimental measurements with simulations of the Helmstetter-Cooper model (5). MG1655 strain transformed with pLR40 plasmid for DnaA oscillatory induction as in Figure 3A.

Algorithm 2: Function `GenerateLineage`.

Function `GenerateLineage`$(x,t,v,A,P,N_{\max},Rn)$

Input:

- x: physiological variables λ, τ_{cyc} and s_{i}.
- t: time.
- v: cell size.
- $\{A_i\}_{i=1}^{P_{\max}}$: array holding initiation times of active replication cycles.
- P: number of active replication cycles.
- $N_{\max}$: number of generations to simulate.
- Rn: random number generator (defined according to our model)

▷ *Initializations*
$t_{\text{b}} = t$, $v_{\text{b}} = v$, $N = 1$
▷ *Start loop for cell generations*
while $N <= N_{\max}$ **do**
  ▷ *Name physiological variables*
  $\lambda = x_1$
  ▷ *Initiate replications until division occurs*
  `InitiateReplications`(x,t,v,A,P,t_d)
  ▷ *Cell division*
  $v_{\text{d}} = v_{\text{b}} \cdot \exp(\lambda(t_{\text{d}} - t_{\text{b}}))$
  dump N, t_{b}, t_{d}, v_{b}, v_{d}, A_0, λ, τ_{cyc}, s_{i}
  ▷ *Update variables for next generation*
  $N = N + 1$, $t = t_{\text{b}} = t_{\text{d}}$, $v = v_{\text{b}} = v_{\text{d}}/2$, $P = P - 1$
  for $i = 1$ **to** P **do** $A_i = A_{i+1}$
  ▷ *Draw next physiological variables from random vector generator*
  $x = Rn(x)$
end
return

We chose $\langle\lambda\rangle = \ln 2$, making the generation time $\langle\tau\rangle = \ln 2/\langle\lambda\rangle$ the unit of time. We chose $\langle s_{\text{i}}\rangle = 1$ as unit of cell size. We investigated values of $\langle\tau_{\text{cyc}}\rangle = 0.5, 1.5, 2.5$, corresponding to non-overlapping, 2 overlapping and 3 overlapping cell cycles, respectively. The reference values for the CVs were:

$$\text{CV}(\lambda) = 10\,\%, \quad \text{CV}(s_{\text{i}}) = 10\,\%. \tag{15}$$

Note that for τ_{cyc}, we took different CVs for each scenario so as to keep the same amplitude of fluctuations in the cell cycle duration. Specifically, we chose a standard deviation $\sigma_{\tau_{\text{cyc}}} = 0.05$, defining a CV of 10 % for the non-overlapping cell cycle scenario but resulting in smaller CVs for the other. The reference matrix of cross-correlations was set to:

$$\begin{array}{c} \\ \lambda \\ \tau_{\text{cyc}} \\ s_{\text{i}} \end{array} \begin{array}{c} \begin{array}{ccc} \lambda & \tau_{\text{cyc}} & s_{\text{i}} \end{array} \\ \begin{pmatrix} 0 & -0.5 & 0 \\ -0.5 & 0 & 0 \\ 0 & 0.0 & 0 \end{pmatrix} \end{array}, \tag{16}$$

and the reference autocorrelations were taken to be:

$$\rho(\lambda^{n+1}, \lambda^n) = 0.5, \quad \rho(\tau_{\text{cyc}}^{n+1}, \tau_{\text{cyc}}^n) = 0, \quad \rho(s_{\text{i}}^{n+1}, s_{\text{i}}^n) = 0.5. \tag{17}$$

In **Co-regulation between division adder and initiation adder in the stochastic Helmstetter-Cooper model** panel A, the co-regulation hypothesis from Equation (14) holds for non-overlapping cell cycles. For two and three overlapping cell cycles, deviations from the co-regulation hypothesis are seen. This is due to the sources of noise still present in the system, which tend to uncouple distant generations. For example, reducing the CV of the growth rate to $\text{CV}(\lambda) = 1\,\%$ dramatically reduces these deviations (not shown). Similarly, when increasing the noise in τ_{cyc} to $\sigma_{\tau_{\text{cyc}}} = 0.1$, deviations from the co-regulation hypothesis are more pronounced (**Co-regulation between division adder and initiation adder in the stochastic Helmstetter-Cooper model** panel B). However, despite the fact that other parameters can affect the division adder correlation

Algorithm 3: Function `InitiateReplication`.

Function `InitiateReplications`(x,t,v,A,P,t_d)

Input:

- x: physiological variables λ, τ_cyc and s_i.
- t: current time.
- v: current cell size.
- $\{A_i\}_{i=1}^{P_\mathrm{max}}$: array holding initiation times of active replication cycles.
- P: number of active replication cycles.

Output:

- t_d: next division time.

▷ *Name physiological variables*
$\lambda = x_1$, $\tau_\mathrm{cyc} = x_2$, $s_\mathrm{i} = x_3$
▷ *Initializations*
$t_\mathrm{d} = \infty$, initiate = true
▷ *Start loop for replication cycles*
while *initiate* **do**
  ▷ *Determine next division time*
  if $P > 0$ **then** $t_\mathrm{d} = A_1 + \tau_\mathrm{cyc}$
  ▷ *Determine next initiation time*
  $N_\mathrm{ori} = 2^P$
  $v_\mathrm{i} = s_\mathrm{i} \cdot N_\mathrm{ori}$
  $\delta t = \max[\frac{1}{\lambda}\ln(\frac{v_\mathrm{i}}{v}),0]$
  $t_\mathrm{i} = t + \delta t$
  ▷ *Stop initiating if next initiation is after next division*
  if $t_i > t_d$ **then**
    initiate = false
  else
    dump t_i, v_i, P
    $t = t_\mathrm{i}$, $v = v_\mathrm{i}$, $P = P + 1$, $A_P = t_\mathrm{i}$
  end
end
return

(especially for overlapping cell cycles), the effect of the unit cell autocorrelation $\rho(s_\mathrm{i}^{n+1}, s_\mathrm{i}^n)$ on the value of $\rho(\Delta_\mathrm{d}, S_\mathrm{b})$, was more systematic than cross-correlations or CVs. Therefore we concluded that even when fluctuations are introduced into the Helmstetter-Cooper model, altering the homeostasis of s_i should affect the cell size homeostasis.

F. Adder properties

The size autocorrelation can be used to characterize the cell size behavior. We focus now on the division size properties, yet the following development can also be applied to the initiation size.

We denote $p(S_\mathrm{d}^n, S_\mathrm{d}^{n+1})$ the joint probability distribution of cell size at division for a pair of mother/daughter cells. In the first approximation, we assume it is a Gaussian bivariate distribution with means $\langle S_\mathrm{d}^n\rangle = \langle S_\mathrm{d}^{n+1}\rangle = \mu(S_\mathrm{d})$ and covariance matrix:

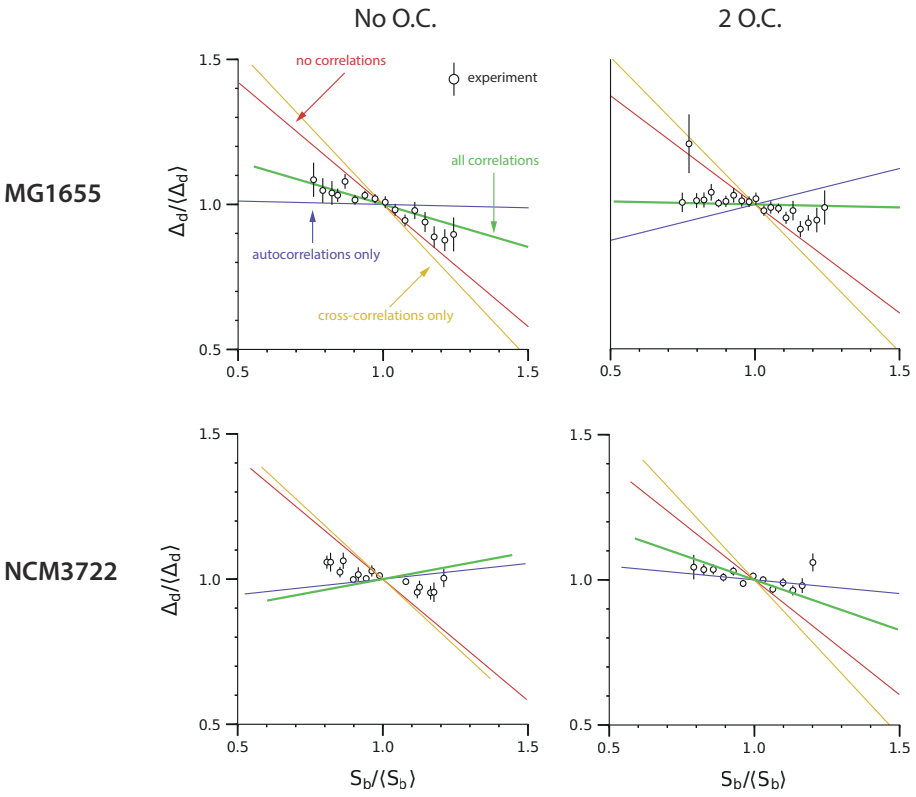
$$\sigma^2 \begin{pmatrix} 1 & \gamma \\ \gamma & 1 \end{pmatrix}, \tag{18}$$

with $\sigma = \mu(S_\mathrm{d}) \cdot \mathrm{CV}(S_\mathrm{d})$. Consequently, the conditional value of the daughter division size is obtained as (see appendix A):

$$\langle S_\mathrm{d}^{n+1} \mid S_\mathrm{d}^n\rangle = \gamma S_\mathrm{d}^n + (1-\gamma)\mu(S_\mathrm{d}). \tag{19}$$

The added size between division for generation n is expressed as

$$\begin{aligned} \Delta_\mathrm{d}^n &= S_\mathrm{d}^n - S_\mathrm{b}^n, \\ &= 2S_\mathrm{b}^{n+1} - S_\mathrm{b}^n, \end{aligned} \tag{20}$$



Comparison of the experimental adder behavior with simulation of the stochastic Helmstetter-Cooper model. The simulations are performed without correlations (red), with cross-correlations only (yellow), with autocorrelations only (blue) and with all correlations (green) between physiological variables. The experimental data for each condition were used to set the mean values, coefficient-of-variations and correlations of the joint distribution of the physiological variables.

assuming symmetric division. Using the previous equation, one can show that:

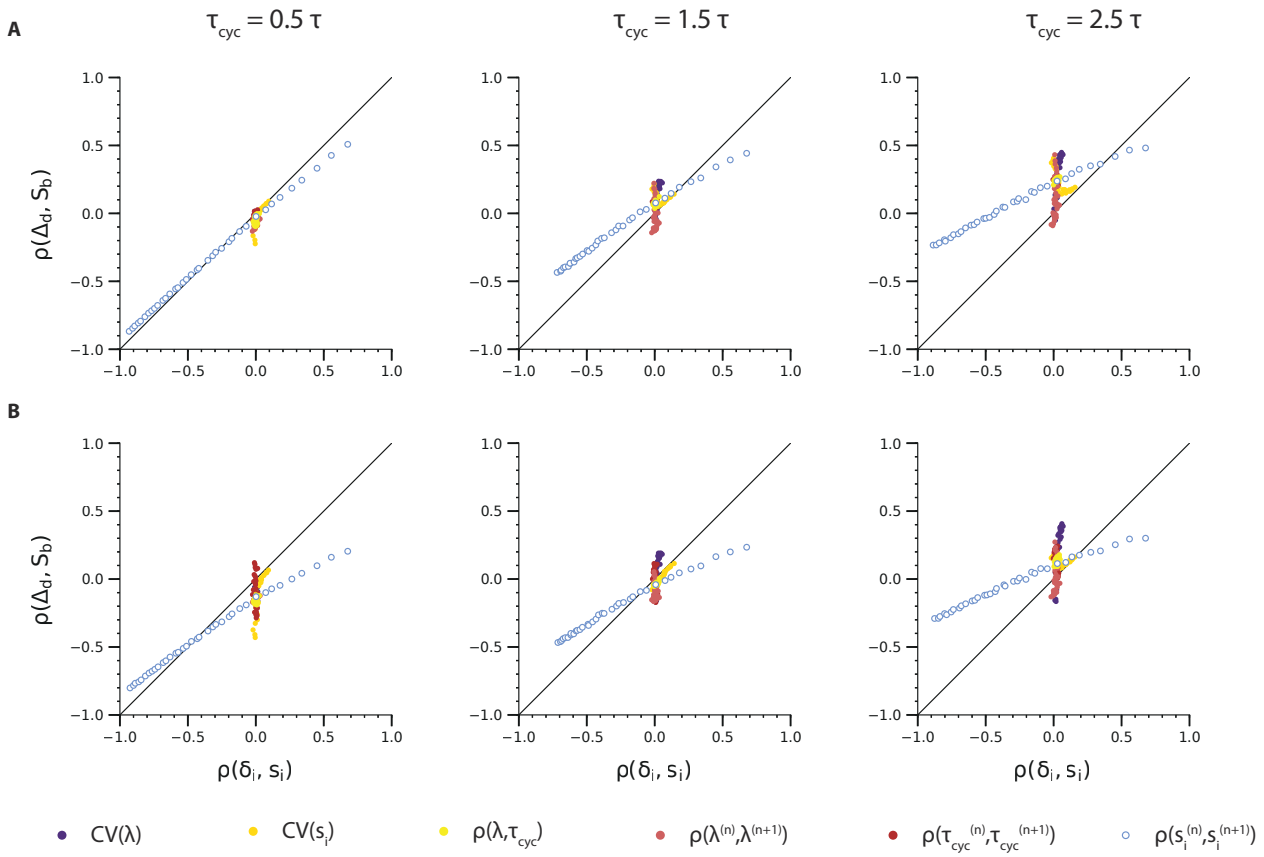
$$\begin{aligned} \mathrm{CV}(\Delta_\mathrm{d}) &= \sqrt{5-4\gamma}\,\mathrm{CV}(S_\mathrm{d}), \\ \langle \Delta_\mathrm{d} | S_\mathrm{b} \rangle &= (2\gamma-1)S_\mathrm{b} + 2(1-\gamma)\mu(S_\mathrm{b}). \\ \langle \delta\Delta_\mathrm{d}\delta S_\mathrm{b}\rangle &= (2\gamma-1)\langle \delta S_\mathrm{b}^2\rangle, \\ \rho(\Delta_\mathrm{d}, S_\mathrm{b}) &= \frac{2\gamma-1}{\sqrt{5-4\gamma}}. \end{aligned} \tag{21}$$

The last expression is shown in **Functional dependence of $\rho(\Delta_\mathbf{d}, S_\mathbf{b})$ on $\rho(S_\mathbf{d}^n, S_\mathbf{d}^{n+1})$**. Therefore, when $\gamma = 1/2$, the adder principle holds, meaning that the added size is independent of birth size: $\langle \delta\Delta_\mathrm{d}\delta S_\mathrm{b}\rangle = 0$ (or $\rho(\Delta_\mathrm{d}, S_\mathrm{b}) = 0$).

II. THRESHOLD MODELS IN BALANCED GROWTH

A. Control of replication initiation

Replication initiation is influenced by several factors, the most important being probably the DnaA protein [S17–S19]. The DnaA protein is active when bound to ATP (DnaA-ATP) and inactive when bound to ADP (DnaA-ADP). While both active and inactive forms can bind the *oriC*, evidence indicates that only the active form can trigger initiation [S20]. Approximately 10-20 DnaA-ATPs are required at the *oriC* to form a functional complex that can lead to replication initiation [S21]. Here we neglect the role of DnaA-ADP, namely as a competitor to *oriC* binding. We therefore consider that replication initiation is under the exclusive control of DnaA-ATP. DnaA binds primarily ATP after being synthesized in the cytoplasm [S18, S19], therefore the DnaA-ATP production coincides with the DnaA production. For these reasons we will abusively denote DnaA-ATP as DnaA, which we consider as the replication initiator. We also adopt the simple autorepressor model for the *dnaA* operon [S22], *i.e.* the DnaA protein is maintained at a nearly fixed concentration by repressing its own expression.



Co-regulation between division adder and initiation adder in the stochastic Helmstetter-Cooper model [Equation (14)]. **(A)** $\sigma_{\tau_{\rm cyc}} = 0.05$. **(B)** $\sigma_{\tau_{\rm cyc}} = 0.1$, with $\sigma_{\tau_{\rm cyc}} = \langle \tau_{\rm cyc} \rangle \cdot {\rm CV}(\tau_{\rm cyc})$.

B. Control of division

Z-ring formation is the predominant division process prior to constriction [S23]. Once the Z-ring is functional, division and cell wall machinery proteins bind to this scaffold in order to complete cytokinesis. The Z-ring is made of protofilaments of the essential protein FtsZ. In our experimental assays, we have adopted a nearly functional FtsZ-mVenus fusion protein [S24] in order to monitor the assembly of the Z-ring in single-cells. Our observations suggest that FtsZ accumulates to a threshold at the Z-ring. Indeed, the maximum intensity (in a cell life time) at the Z-ring was found to be independent of the cell size at division. This means that on average a fixed, critical amount of FtsZ in the Z-ring is required to trigger the assembly of the division machinery and cell constriction. This threshold mechanism parallels the control of replication initiation by DnaA.

In addition, our experimental assays also suggested that the concentration of FtsZ remains relatively constant during the division cycle, and across many generations. As far as we know, the FtsZ protein does not repress its own expression, like DnaA does. However, we explain this fixed concentration at steady state by postulating that FtsZ production is in balanced growth (see section III).

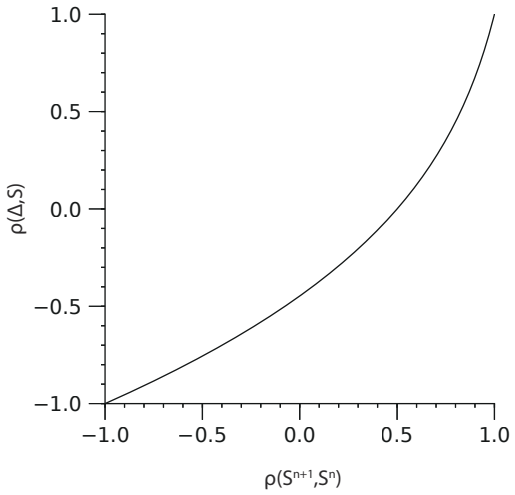
C. Threshold model

Let us now consider a generic protein responsible for the initiation of cell division (note that the same reasoning applies for the control of replication initiation). We assume that this protein accumulates and triggers cell division when its copy number N reaches a fixed threshold:

$$N(t_{\rm d}) = N_0. \tag{22}$$

Following cell division, each daughter cell receives $N_0/2$ copies of the protein. Under balanced growth (see section III), the protein copy number increases in proportion to the cell volume:

$$\frac{{\rm d}N}{{\rm d}t} = c^* \frac{{\rm d}S}{{\rm d}t}, \tag{23}$$



Functional dependence of $\rho(\Delta_{\mathbf{d}}, S_{\mathbf{b}})$ on $\rho(S^n_{\mathbf{d}}, S^{n+1}_{\mathbf{d}})$ (see Equation (21)).

where c^* is the steady-state protein concentration. Therefore, one obtains that the added volume from birth to division is given by:

$$S_{\mathrm{d}} - S_{\mathrm{b}} = \frac{N_0}{2c^*}, \tag{24}$$

which is the adder principle.

The fixed threshold N_0 does not necessarily imply that the cell physically senses the copy number N of proteins in the cytoplasm. Instead, N can be a proxy for the number of initiators bound to a cell compartment, namely the Z-ring. For example, let us assume that the binding/unbinding dynamics of the proteins in the cytoplasm (N) with a cell compartment (with copy number n) can be modeled with the linear system:

$$\begin{aligned} \frac{\mathrm{d}n}{\mathrm{d}t} &= \alpha N - \beta n, \\ \frac{\mathrm{d}N}{\mathrm{d}t} &= -\alpha N + \beta n + \lambda c^* S. \end{aligned} \tag{25}$$

In the previous system, α and β are the rates of binding and unbinding to the cellular compartment, respectively. $\lambda c^* S$ is the production rate of proteins in the cytoplasm in balanced growth (see section III). Denoting $X = n + N$, and assuming exponential growth of the cell volume at the rate λ, we immediately have:

$$X(t) = X(0) + c^* \underbrace{(S(t) - S(0))}_{\Delta S(t)}. \tag{26}$$

Using the previous equation, we obtain:

$$\begin{aligned} n(t) &= n(0)e^{-(\alpha+\beta)t} + \frac{\alpha}{\alpha+\beta}(X(0) - c^*S(0))(1 - e^{-(\alpha+\beta)t}) \\ &+ \frac{\alpha}{\alpha+\beta+\lambda}c^*(S(t) - S(0)e^{-(\alpha+\beta)t}), \\ &\approx \frac{\alpha}{\alpha+\beta}\left(X(t) - c^*S(t)\left(1 - \frac{\alpha+\beta}{\alpha+\beta+\lambda}\right)\right), \\ &\approx \frac{\alpha}{\alpha+\beta}\left(X(t) - c^*S(t)\frac{\lambda}{\alpha+\beta}\right), \end{aligned} \tag{27}$$

where in the first approximation we assumed that $(\alpha+\beta)^{-1} \ll t$, and in the second approximation that $\lambda \ll \alpha+\beta$. Therefore, as long as the elongation rate is much smaller than the binding/unbinding rates, the copy number of proteins bound to the cell compartment can be seen as a fixed fraction of the total copy number: $n = \alpha/(\alpha+\beta)X$, and similarly $N = \beta/(\alpha+\beta)X$. In other words, the protein dynamics is fast compared to growth of the cell, therefore the cytoplasm reservoir and the cell compartment

are always at equilibrium. When λ becomes comparable to $\alpha + \beta$, deviations from this equilibrium appear. Namely, the number of proteins bound to the cell compartment is below its equilibrium value: $n < \alpha/(\alpha+\beta)X$, meaning that some delay is observed for the cell compartment to reach its threshold. For simplicity, we have considered here that the cytoplasm reservoir and cell compartment are at equilibrium. Eventually, a threshold n_0 to be reached in the cell compartment translates into a threshold N_0 to be reached in the cell cytoplasm, and in a global threshold X_0 to be reached for the total protein copy number.

D. Relation to cell size homeostasis

We now consider the general case where c^* is subject to fluctuations, and focus on the division size homeostasis. The same reasoning applies to the initiation size homeostasis. Using the definition $c(t_\mathrm{d}) = N_0/S(t_\mathrm{d})$, the Pearson correlation coefficient for division size between consecutive generations is:

$$\rho\left(S_\mathrm{d}^n, S_\mathrm{d}^{n+1}\right) = \rho\left(\frac{1}{c(t_\mathrm{d}^n)}, \frac{1}{c(t_\mathrm{d}^{n+1})}\right), \tag{28}$$

where t_d^n and t_d^{n+1} are the times at division for generations n and $n+1$. Provided that the fluctuations in concentration are not too large, the previous expression can be approximated to (see appendix A):

$$\rho\left(S_\mathrm{d}^n, S_\mathrm{d}^{n+1}\right) \approx \rho\left(c(t_\mathrm{d}^n), c(t_\mathrm{d}^{n+1})\right), \tag{29}$$

Therefore, cell size homeostasis appears to be linked to the initiator concentration homeostasis. In particular, if fluctuations in the protein concentration occur on time scales much shorter than the generation time, the division size correlation between consecutive generations should vanish, resulting in a "sizer" behavior.

We now relate the mother/daughter division concentrations correlation from Equation (29) to the time autocorrelation of the protein concentration. Let us consider L lineages of cells in a time interval $[0, W]$. Let us denote by $c_{a,i}$ the concentration at division for the cell corresponding to generation i of lineage a. The mother/daughter concentration correlation is computed as:

$$\langle c(t_\mathrm{d}^n) c(t_\mathrm{d}^{n+1})\rangle = \frac{1}{W}\sum_{i=0}^{W-1}\frac{1}{L}\sum_{a=1}^{L} c_{a,i}c_{a,i+1}. \tag{30}$$

The previous average should converge to a fixed value for large W and large L. We now assume ergodicity. Specifically, the average in the previous equation with $W \to \infty$ and $L = 1$ is equal the average with $W = 1$ and $L \to \infty$:

$$\begin{aligned}\langle c(t_\mathrm{d}^n) c(t_\mathrm{d}^{n+1})\rangle &= \lim_{W\to\infty}\frac{1}{W}\sum_{i=0}^{W-1} c_i c_{i+1},\\ &= \lim_{L\to\infty}\frac{1}{L}\sum_{a=1}^{L} c_{a,0}c_{a,1}.\end{aligned} \tag{31}$$

Let us introduce the conditional probability distribution $p(t'|t)$ that a daughter cell divides at times t' given that its mother cell divided at time t. We rewrite the last expression as:

$$\begin{aligned}\langle c(t_\mathrm{d}^n) c(t_\mathrm{d}^{n+1})\rangle &= \lim_{L\to\infty}\frac{1}{L}\sum_{a=1}^{L}\int_0^\infty \mathrm{d}t\; p(t_0+t|t_0)c_a(t_0)c_a(t_0+t),\\ &\approx \lim_{L\to\infty}\frac{1}{L}\sum_{a=1}^{L} c_a(t_0)c_a(t_0+\tau),\\ &= \langle c(t_0)c(t_0+\tau)\rangle,\end{aligned} \tag{32}$$

where in the approximation, we assumed $p(t'|t) = \delta(t' - t - \tau)$. This approximation is valid as long as the fluctuations of the concentration are small in an interval $[\tau - \sigma_\tau, \tau + \sigma_\tau]$ centered around the average generation time τ, with σ_τ being the width of its distribution (the typical fluctuations of the generation time). In the last expression, the concentration $c(t)$ should be understood as a continuous stochastic process. Therefore, to a particular lineage of cells a corresponds one stochastic

continuous process $c_a(t)$. The brackets means that an average is taken over all the realizations of $\{c_a(t)\}_{a=1...L}$. The previous equation is also valid for the centered concentration $\delta c(t) = c(t) - \langle c \rangle$. As a result, we obtain:

$$\begin{aligned} \rho\left(S_{\mathrm{d}}^{n}, S_{\mathrm{d}}^{n+1}\right) &= \rho\left(\frac{1}{c_{\mathrm{d}}^{n}}, \frac{1}{c_{\mathrm{d}}^{n+1}}\right), \\ &\approx \rho\left(c_{\mathrm{d}}^{n}, c_{\mathrm{d}}^{n+1}\right), \\ &\approx \rho\left[c\right](\tau), \end{aligned} \tag{33}$$

where:

$$\rho\left[c\right](t) = \frac{\langle \delta c(t_0) \delta c(t_0 + t) \rangle}{\delta c^2}. \tag{34}$$

E. Adder property

In the next section, we will show that under balanced growth, $\rho[c](\tau) = 1/2$. From Equation (33), and using the relations from Equation (21), we conclude that:

$$\rho(\Delta_{\mathrm{d}}, S_{\mathrm{b}}) = 0, \tag{35}$$

which is an other formulation of the adder principle.

III. REPROGRAMMING CELL SIZE HOMEOSTASIS BY BREAKING BALANCED GROWTH

A. Balanced growth

Consider a type of protein whose mass fraction in the cell is ϕ^* at steady state. If we denote by m the mass of these proteins, and M the total dry mass of the cell, we have in balanced growth [S25, S26]:

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \phi^* \frac{\mathrm{d}M}{\mathrm{d}t}. \tag{36}$$

To reformulate the previous equation in terms of the protein copy number N in the cell, we introduce the mass of one protein m_P, the cell volumic mass ρ_c and the cell size or volume S. A simple rewriting leads to:

$$\frac{\mathrm{d}N}{\mathrm{d}t} = c^* \frac{\mathrm{d}S}{\mathrm{d}t}, \tag{37}$$

where $c^* = \phi^* \rho_c / m_P$ is the protein concentration at steady state. The protein concentration is $c = N/S$, thus satisfies the first order differential equation:

$$\frac{\mathrm{d}c}{\mathrm{d}t} + \lambda c = \lambda c^*. \tag{38}$$

In Equation (38), steady state is achieved when the protein synthesis rate per unit of volume, λc^*, balances the decrease in protein concentration due to dilution, λc.

B. Time-dependent production rate

Let us suppose now that the production rate in Equation (38) is not constant. Instead, the protein synthesis allocation is a time-dependent function $p(t)$. The protein concentration obeys the differential equation:

$$\frac{\mathrm{d}c}{\mathrm{d}t} + \lambda c = \lambda p(t), \tag{39}$$

The solution of this ODE is:

$$c(t) = c(0) e^{-\lambda t} + \lambda \int_0^t \mathrm{d}s \, e^{-\lambda(t-s)} p(s). \tag{40}$$

When the production rate is a periodic function of time, the steady state solution for the protein concentration will also be a periodic function with same period. For example, we give $c(t)$ for the cases of cosine and periodic square production rates in **Protein concentration for a cosine production rate** and **Protein concentration for a periodic production rate with pulses**. In practice, a time-dependent production rate can be achieved by imposing a time-dependent induction of a promoter. In particular, a periodic square production is obtained by switching between a medium without the inducer and a medium with the inducer every half-period.

$p(t)$	$c^*\left(1+\mu\sin^2(\omega t/2)\right)$
$c(t)$	$(c(0)-c_\infty(0))\,e^{-\lambda t}+c_\infty(t)$
$c_\infty(t)/c^*$	$1+\mu(1-\cos\varphi\cos(\omega t-\varphi))/2$
Notations	$\omega=2\pi/T$ $\tan\varphi=\omega/\lambda$

Protein concentration for a cosine production rate.

$p(t)$	$c^*(1+\mu S(t))$
$c(t)$	$(c(0)-c_\infty(0))\,e^{-\lambda t}+c_\infty(t)$
$c_\infty(t)/c^*$	$1+\mu\left(\psi(a)e^{-\lambda rT}+\left(1-e^{-\lambda(r-a)T}\right)S(t)\right)$
Notations	$t=(n+r)T$, where n is the integer part of t/T and $0\le r<1$. $S(t)=\begin{cases}0 & \text{if} \quad r<a,\\ 1 & \text{otherwise.}\end{cases}$ $\psi(a)=(1-e^{-\lambda(1-a)T})/(1-e^{-\lambda T})$

Protein concentration for a periodic production rate with pulses. The periodic square function is obtained when $a=1/2$.

C. Stochastic production rate

Here we consider that the protein synthesis allocation undergoes stochastic fluctuations. The protein concentration obeys the differential equation:

$$\frac{\mathrm{d}c}{\mathrm{d}t}+\lambda c=\lambda(p(t)+\eta(t)), \tag{41}$$

where $\eta(t)$ is a Gaussian white noise such that $\langle\eta(t)\rangle=0$ and $\langle\eta(t)\eta(t')\rangle=2\Gamma\delta(t-t')$. The brackets denote an average over different realizations of the noise, for example over many different cells subject to the same production rate (*e.g.* through the same induction). We may decompose the deterministic and stochastic contributions by writing $c(t)=\langle c(t)\rangle+y(t)$. The average concentration $\langle c(t)\rangle$ follows the deterministic ODE in Equation (39) while the fluctuations around the average are expressed as:

$$y(t)=\lambda\int_0^t \mathrm{d}s\, e^{-\lambda(t-s)}\eta(s). \tag{42}$$

Being a sum of Gaussian random variables $y(t)$ is also a Gaussian random variable, with mean $\langle y(t)\rangle=0$, and variance:

$$\begin{aligned}\langle y(t)^2\rangle &= \lambda^2\int_0^t \mathrm{d}s\int_0^t \mathrm{d}s'\, e^{-\lambda(2t-s-s')}\langle\eta(s)\eta(s')\rangle,\\ &= \Gamma\lambda\left(1-e^{-2\lambda t}\right),\\ &\xrightarrow[t\to\infty]{} \Gamma\lambda.\end{aligned} \tag{43}$$

Similarly, the two-point correlation is:

$$\langle y(t_0)y(t_0+t)\rangle=\Gamma\lambda e^{-\lambda t}\left(1-e^{-2\lambda t_0}\right)\xrightarrow[t_0\to\infty]{}\Gamma\lambda e^{-\lambda t}. \tag{44}$$

D. Concentration autocorrelation

In section II we have presented a model in which cell size homeostasis is driven by the autocorrelation function of division proteins concentration. Here we first give this time autocorrelation function in balanced growth, when the production rate of these protein is fixed. We then show how the autocorrelation function is modified when the production rate oscillates.

D1) Fixed production rate

In balanced growth, the production rate of proteins is fixed, namely $p(t)=\phi^*$. Thus $\langle c\rangle=c^*$. The Pearson time autocorrelation coefficient for protein concentration is defined as:

$$\rho[c](t)=\frac{\langle\delta c(t_0+t)\delta c(t_0)\rangle}{\langle\delta c^2\rangle}, \tag{45}$$

where $\delta c(t) = c(t) - \langle c \rangle$ and the brackets denote an average over different realizations of the stochastic process in Equation (41) (*i.e.* different lineages). Using Equation (44), we obtain

$$\rho[c](t) = e^{-\lambda t}. \tag{46}$$

In particular, for $t = \tau = \ln 2/\langle \lambda \rangle$, we have $\rho[c](\tau) = 1/2$, which together with Equation (33) ensures the adder behavior for cell size homeostasis in balanced growth.

D2) Time-dependent production rate

For a time-dependent production rate, the expression in Equation (45) must be revised because time translational invariance is broken, and it is necessary to take into account variations in time for the production rate. In particular, the average concentration $\langle c(t) \rangle$ is a function of time. For a periodic production rate with period T, ergodicity can still be assumed, but Equation (31) is modified to:

$$\begin{aligned} \overline{c(t_{\rm d}^n) c(t_{\rm d}^{n+1})} &= \lim_{W\to\infty} \frac{1}{W} \sum_{i=0}^{W-1} c_i c_{i+1}, \\ &= \lim_{L\to\infty} \frac{1}{L} \sum_{a=1}^{L} \int_0^T \mathrm{d}t_0 \, c_{a,0}(t_0) c_{a,1}(t_0), \end{aligned} \tag{47}$$

where $c_{a,0}(t_0)$ and $c_{a,1}(t_0)$ denote the concentrations of division proteins at division for a pair of mother/daughter cells such that the mother divides at t_0. We used a bar symbol for this average to make a distinction from the previous average with brackets. Thus, Equation (32) becomes:

$$\overline{c(t_{\rm d}^n) c(t_{\rm d}^{n+1})} = \frac{1}{T} \int_0^T \mathrm{d}t_0 \, \langle c(t_0) c(t_0 + \tau) \rangle, \tag{48}$$

where the brackets as before denote an average over different lineages. The average concentration now reads:

$$\overline{c} = \frac{1}{T} \int_0^T \mathrm{d}t \, \langle c(t) \rangle. \tag{49}$$

We rewrite the connected correlation for mother/daughter concentrations at division:

$$\begin{aligned} \overline{\delta c(t_{\rm d}^{n+1}) \delta c(t_{\rm d}^n)} &= \overline{\left(c(t_{\rm d}^{n+1}) - \overline{c} \right) \left(c(t_{\rm d}^n) - \overline{c} \right)}, \\ &= \frac{1}{T} \int_0^T \mathrm{d}t_0 \, \langle \delta c(t_0 + \tau) \delta c(t_0) \rangle, \end{aligned} \tag{50}$$

where $\delta c(t) = c(t) - \overline{c}$ and τ is the mean generation time.

The last expression in Equation (50) is the two-point correlation evaluated at $t = \tau$. It can be decomposed as a sum of a deterministic contribution due to the time variations of the production rate, and a stochastic contribution due to the stochasticity in Equation (41):

$$\begin{aligned} S(t) &= \frac{1}{T} \int_0^T \mathrm{d}t_0 \, \langle \delta c(t_0 + t) \delta c(t_0) \rangle, \\ &= \underbrace{\frac{1}{T} \int_0^T \mathrm{d}t_0 \, \left(\langle c(t_0) \rangle - \overline{c} \right) \left(\langle c(t_0 + t) \rangle - \overline{c} \right)}_{\text{deterministic}} + \underbrace{\frac{1}{T} \int_0^T \mathrm{d}t_0 \, \langle y(t_0) y(t_0 + t) \rangle}_{\text{stochastic}}, \end{aligned} \tag{51}$$

where as before $y(t) = c(t) - \langle c(t) \rangle$. Finally, the Pearson time autocorrelation coefficient is expressed as:

$$\rho[c](t) = \frac{S(t)}{S(0)} \tag{52}$$

For example, for a cosine production rate (**Protein concentration for a cosine production rate**), we find:

$$S(t) = \underbrace{\frac{1}{2}\left(\frac{\mu}{2}c^* \cos\varphi\right)^2}_{A} \cos(\omega t) + \langle \delta c^2 \rangle e^{-\lambda t}. \tag{53}$$

There are two contributions in the two-point correlation from the previous expression. The second is due to the inherent stochasticity in the protein production, while the first is imposed by the specific shape of the production rate function. We see that a careful choice of the amplitude of the oscillations μ, and of the period of the oscillations T (which determines the value of $\cos\varphi$), can lead to $S(\tau) < 0$. For example, taking $T = 2\tau$ leads to:

$$\rho[c](\tau) = \frac{1}{2}\frac{\langle \delta c^2 \rangle - 2A}{\langle \delta c^2 \rangle + A}, \tag{54}$$

which can be made negative by increasing μ. Similarly, taking $T = 4\tau$ leads to:

$$\rho[c](\tau) = \frac{1}{2}\frac{\langle \delta c^2 \rangle}{\langle \delta c^2 \rangle + A}, \tag{55}$$

which is smaller than the value of $1/2$ from balanced growth and converges to zero when increasing μ. As can be seen, the adjustable parameter to tune the autocorrelation coefficient is the ratio $\langle \delta c^2 \rangle / A$ which essentially quantifies the stochastic fluctuations of protein concentration versus the amplitude of the induced oscillations of mean protein concentration. In particular when $\langle \delta c^2 \rangle \gg A$, one retrieves the static result: $\rho(\tau) \to 1/2$. On the contrary in the limit of vanishing noise $\langle \delta c^2 \rangle \ll A$, one obtains as expected $\rho[c](\tau) = -1$ when $T = 2\tau$ and $\rho[c](\tau) = 0$ when $T = 4\tau$.

An undesired property of the time-autocorrelation function obtained from Equation (53) is that for a period $T \gg \tau$, it does not converge to the exponential function from balanced growth. This comes from the fact that deviations in the concentration are taken with respect to the total average defined in Equation (49). However, when $T \gg \tau$, the concentration is approximately constant in time intervals of length τ. Namely, when analyzing fluctuations, deviations should be considered around the average concentration in this interval, say $[t, t + \tau]$, rather than the average in Equation (53). To circumvent this problem, we define the concentration average in a window of size T_w:

$$E_w(c)(t) = \frac{1}{T_w}\int_{-T_w}^{T_w} \mathrm{d}s \, \langle c(t+s) \rangle. \tag{56}$$

Substituting $\bar{c} \leftarrow E_w(c)(t)$ in Equation (51), we obtain for the cosine induction:

$$S(t) = \underbrace{\frac{1}{2}\left(\frac{\mu}{2}c^* \cos\varphi \left(1 - \mathrm{sinc}\left(\frac{\omega T_w}{2}\right)\right)\right)^2}_{A} \cos(\omega t) + \langle \delta c^2 \rangle e^{-\lambda t}, \tag{57}$$

where $\mathrm{sinc}(x) = \sin(x)/x$. Effectively, this corresponds to rescaling the amplitude by a factor $(1 - \mathrm{sinc}(\omega T_w/2))$. Considering that the generation time is the relevant time-scale for studying protein concentration fluctuations, we may take for simplicity $T_w = 2\tau$. Therefore $\omega T_w/2 = 2\pi\tau/T$, and for $T \gg \tau$, we retrieve the exponential function in Equation (46) corresponding to a fixed induction, namely balanced growth. With such a definition, we show in **Concentration autocorrelation function for a cosine induction or for a square induction** how the concentration time-autocorrelation function varies when the induction amplitude μ and the period of oscillations T are changed for both cosine and square inductions.

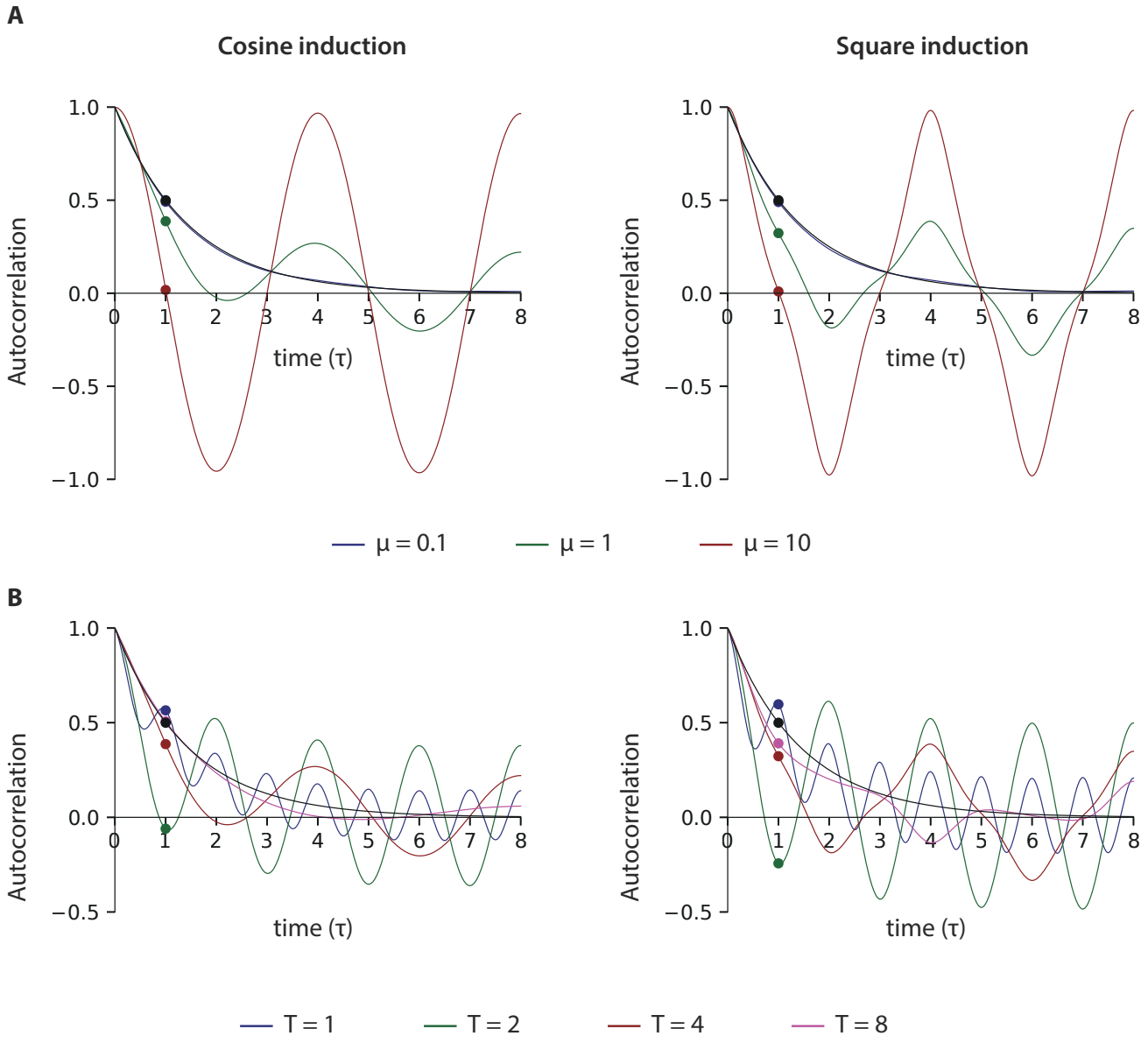
E. Simulations of the combined thresholds model

We used simulations to generate lineages of cells according to the combined thresholds model. Here we start considering the growth rate is time-dependent, as discussed in appendix B. Following Equation (94), we assume that each cell grows exponentially its size according to:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \lambda(t) S(t). \tag{58}$$

To describe the time-evolution of the instantaneous growth rate $\lambda(t)$, we linearized Equation (97) around the steady state growth rate λ^*, and introduced stochastic fluctuations:

$$\frac{\mathrm{d}\lambda}{\mathrm{d}t} = \lambda^*(\lambda^* - \lambda(t)) + \sigma_\lambda \sqrt{2\lambda^*}\eta(t), \tag{59}$$



Concentration autocorrelation function for a cosine induction or for a square induction. **(A)** Dependence of the autocorrelation function on the amplitude of induction μ (we used a period $T = 4$). **(B)** Dependence of the autocorrelation function on the period of induction T (we used an amplitude $\mu = 1$). The autocorrelation functions were computed by simulating the stochastic dynamics as defined in Equation (41) for 10 000 trajectories and using Equation (51). We defined the generation time $\tau = \ln 2/\lambda$ as unit of time and we defined the unit of concentration so that the lower steady state value is $c^* = 1$. The level of noise was set by taking $\sqrt{\Gamma\lambda} = 0.1$, so that the CV of the concentration for a constant induction is 10 %. To compute the autocorrelation we used the running window average from Equation (56) with $T_w = 2\tau$ instead of the total average from Equation (49). The autocorrelation function for a constant production rate, *i.e.* $2^{-t/\tau}$, is denoted by a black line.

where $\eta(t)$ is a Gaussian white noise with correlator $\langle \eta(t) \cdot \eta(t') \rangle = \delta(t' - t)$. The normalization ensures that the linearized process (Ornstein-Uhlenbeck type) is such that the growth rate fluctuations at steady-state are $\langle \delta\lambda^2 \rangle = \sigma_\lambda^2$. Similarly to describe the time-evolution of the concentration of initiation proteins, c_I, and of the concentration of division proteins, c_D, we linearized Equation (100) around the steady states c_I^* and c_D^*, and introduced stochastic fluctuations:

$$\begin{aligned} \frac{\mathrm{d}c_\mathrm{I}}{\mathrm{d}t} &= \lambda^*(c_\mathrm{I}^* - c_\mathrm{I}(t)) + \sigma_\mathrm{I}\sqrt{2\lambda^*}\eta(t), \\ \frac{\mathrm{d}c_\mathrm{D}}{\mathrm{d}t} &= \lambda^*(c_\mathrm{D}^* - c_\mathrm{D}(t)) + \sigma_\mathrm{D}\sqrt{2\lambda^*}\xi(t), \end{aligned} \tag{60}$$

where $\langle \eta(t) \cdot \eta(t') \rangle = \delta(t' - t)$ and $\langle \xi(t) \cdot \xi(t') \rangle = \delta(t' - t)$. Again, the normalizations ensure that the fluctuations of protein concentrations at steady state are such that $\langle \delta c_{\mathrm{I}}^2 \rangle = \sigma_{\mathrm{I}}^2$ and $\langle \delta c_{\mathrm{D}}^2 \rangle = \sigma_{\mathrm{D}}^2$. Note that in our implementation, the steady state concentrations c_{I}^* and c_{D}^* are time-dependent too since they vary with the induction level of the protein of interest. However, due to the periodic square induction, they remain constant between switches.

Initiation occurred when the total copy number of of initiation proteins per origin reached a fixed threshold: $c_{\mathrm{I}} \cdot S = n_0^{\mathrm{I}} \times N_{oriC}$. Similarly, division occurred when the total copy number reached a fixed threshold: $c_{\mathrm{D}} \cdot S = n_0^{\mathrm{D}}$. The practical implementation is described in Algorithms 4 to 6.

We have applied this method to simulate oscillation experiments performed in the laboratory. In **Overlay of experimental results and simulation of the combined threshold model**, we used $\lambda^* = 0.57\,\mathrm{h}^{-1}$, $\sigma_\lambda/\lambda^* = 25\,\%$, $n_0^{\mathrm{I}} = 200$, $n_0^{\mathrm{D}} = 1000$, $n_0^{\mathrm{I}}/c_{\mathrm{I}} = 2\,\mu$m, $\sigma_{\mathrm{I}}/c_{\mathrm{I}}^* = 10\,\%$, $n_0^{\mathrm{D}}/c_{\mathrm{D}} = 4.5\,\mu$m, $\sigma_{\mathrm{D}}/c_{\mathrm{D}}^* = 20\,\%$, a C period of 40 min. Note that for simplicity, we expressed concentrations per unit of length since the cell width is constant. We generated 100 lineages of 100 cells. As can be seen in **Overlay of experimental results and simulation of the combined threshold model** panel A, the experimental distributions are well reproduced in our simulations. Similarly the adder plot is consistent with experimental data (**Overlay of experimental results and simulation of the combined threshold model** panel B). The autocorrelation function for the concentration of division proteins is well reproduced for short times, however some discrepancies arise from intermediate to long times (**Overlay of experimental results and simulation of the combined threshold model** panel C). We suspect this is due to some uncontrolled noise in our experimental readout for division protein concentration, because we use a fluorescent signal as a proxy. Similarly the autocorrelation function for division proteins is in good agreement with experimental data for short lags and more discrepancy arise for long lags (**Overlay of experimental results and simulation of the combined threshold model** panel D). Finally note the agreement between our model and the experimental measurements for protein concentration dynamics is very good (**Overlay of experimental results and simulation of the combined threshold model** panel E). The second plot emphasizes that our model for protein concentration dynamics based on balanced growth (Equations (97) and (100)) and instantaneous change in the fixed protein allocation c^* is accurate. The third plot in **Overlay of experimental results and simulation of the combined threshold model** panel E uses the threshold model assumption $c_{\mathrm{D}}(t_{\mathrm{d}}) \cdot S(t_{\mathrm{d}}) = n_0^{\mathrm{D}}$. Despite some discrepancy, probably due to the simplicity of this model, we note a good agreement with the observed oscillations of cell size.

Algorithm 4: Combined threshold models simulation.

Input:

- Mean growth rate $\langle \lambda \rangle$.
- Steady state concentration of replication initiators $c_{\mathrm{I}}^*(t)$ (can be time-dependent).
- Steady state concentration of division initiators $c_{\mathrm{D}}^*(t)$ (can be time-dependent).
- CV for growth rate: $\mathrm{CV}(\lambda)$.
- CV for concentration of replication initiators: $\mathrm{CV}(c_{\mathrm{I}})$.
- CV for concentration of division initiators: $\mathrm{CV}(c_{\mathrm{D}})$.
- Threshold for replication initiators, n_0^{I}.
- Threshold for division initiators, n_0^{D}.
- Number of generations to simulate per lineage: N.
- Number of lineages to simulate: L.
- Time resolution: Δt.
- C-period: C.

Output: L lineages of N cells.

▷ *Initial time and size*

$t = 0$, $v = 1$

▷ *Initialize array of steady state functions*

Allocate memory for $\{x_i^*\}_{i=1}^4$ ▷ *array of steady state functions*

$x_1 = \mathrm{NULL}$, $x_2 = \langle \lambda \rangle$, $x_3 = c_{\mathrm{I}}^*$, $x_4 = c_{\mathrm{D}}^*$

▷ *Initialize array of CVs*

Allocate memory for $\{\eta_i\}_{i=1}^4$ ▷ *array of CVs*

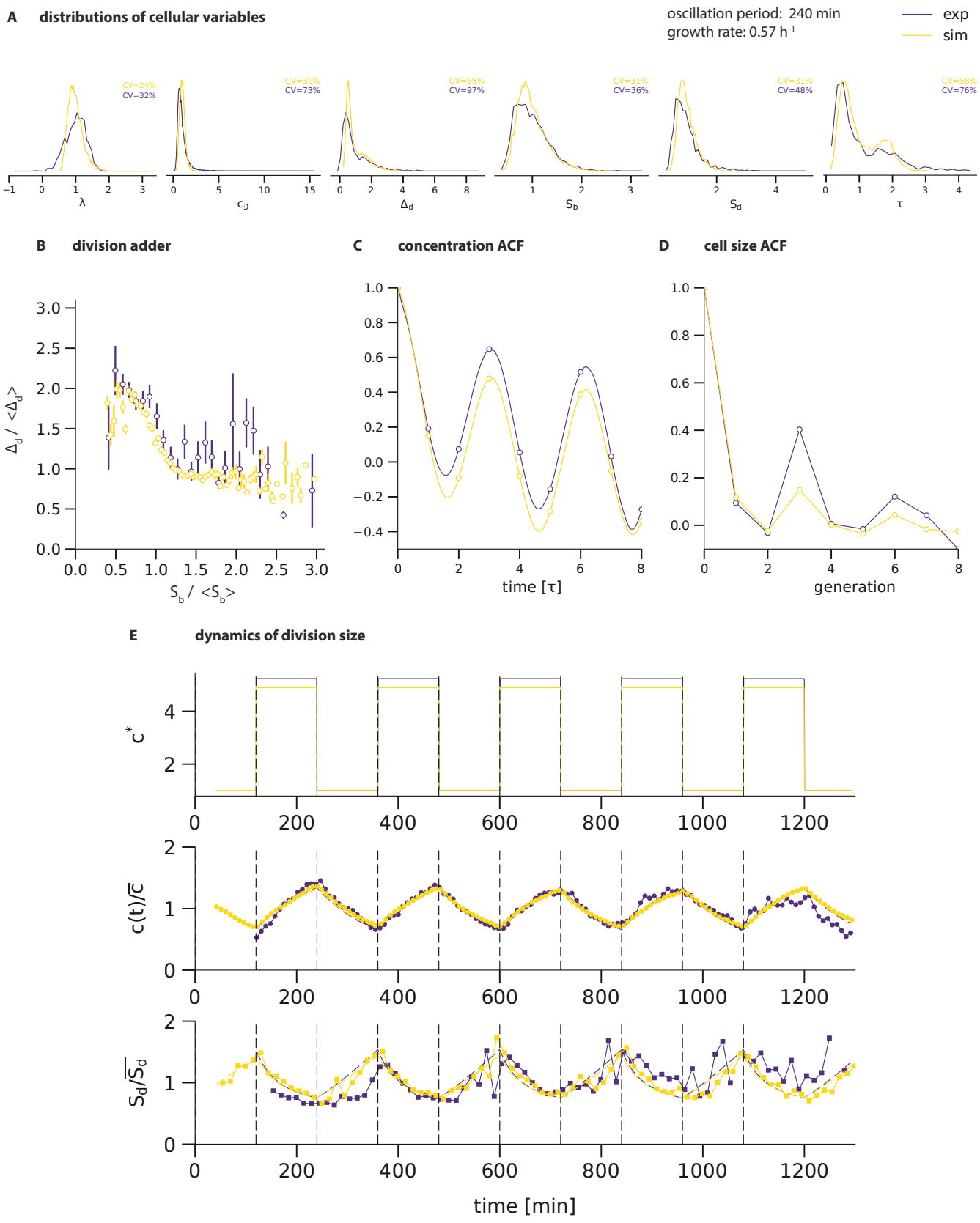
$\eta_1 = \mathrm{NULL}$, $\eta_2 = \mathrm{CV}(\lambda)$, $\eta_3 = \mathrm{CV}(c_{\mathrm{I}})$, $\eta_4 = \mathrm{CV}(c_{\mathrm{D}})$

▷ *Simulate lineages of cells*

for $l = 1$ **to** L **do**

  `GenerateLineage(`$t, v, x^*, \eta, n_0^{\mathrm{I}}, n_0^{\mathrm{D}}, N, \Delta t, C$`)`

end



Overlay of experimental results and simulation of the combined threshold model

Algorithm 5: `GenerateLineage` function for threshold models simulation.

Function `GenerateLineage`(t,v,x^*,η,n_0^{I}, n_0^{D}, $N_{\max}$, Δt,C)

Input:

- Initial time t.
- Initial volume v.
- Array of steady state functions x^*.
- Array of CVs η.
- Threshold for replication initiators, n_0^{I}.
- Threshold for division initiators, n_0^{D}.
- Number of generations to simulate: $N_{\max}$.
- Time resolution: Δt.
- C-period: C.

▷ *Initialize state vector*

Allocate memory for $\{x_i\}_{i=1}^{4}$ ▷ *array of variables*

$x_1 = v$, $x_2 = x_2^*(t)$, $x_3 = x_3^*(t)$, $x_4 = x_4^*(t)$

▷ *Initialize replication cycles*

$P = 0$, $N_{\mathrm{ori}} = 1$

Allocate memory for $\{A_i\}_{i=1}^{P_{\max}}$ ▷ *array of initiation times*

for $i = 1$ **to** $P_{\max}$ **do** $A_i = NULL$

▷ *Start loop on cell generations*

$N = 1$

while $N <= N_{\max}$ **do**

  ▷ *Write birth state*

  dump t, x, N_{ori},x^*,N

  while *true* **do**

    ▷ *Deterministic integration*

    $t' = t + \Delta t$

    dx = `dxDeterministic`(t, x, t', x^*)

    ▷ *Stochastic integration*

    $dx = dx$ + `dxStochastic`(t, x, t', x^*, η, Δt)

    ▷ *Update values*

    $t = t'$, $x = x + dx$

    ▷ *Track cellular events*

    `Initiate`(t,x, N_{ori}, P, A) ▷ *initiation*

    `Terminate`(t,x, N_{ori}, P, A,C) ▷ *termination*

    mustDivide = `Divide`(t,x,n_0^{D}) ▷ *division*

    if *mustDivide* **then break**

    ▷ *Write*

    dump t, x, N_{ori},x^*,N

  end

  ▷ *Symmetrical division*

  $x_1 = x_1/2$, $N_{\mathrm{ori}} = \max(N_{\mathrm{ori}}/2, 1)$, $N = N + 1$

end

Algorithm 6: Functions for threshold models simulation.

Function `dxDeterministic`$(t,\ x,\ t',\ x^*,\ \Delta t)$

Allocate memory for $\{dx_i\}_{i=1}^4$ ▷ *array of displacements*

$dx_1 = x_2 \cdot x_1 \cdot \Delta t$ ▷ *exponential volume growth*

$dx_2 = x_2^*(t) \cdot (x_2^*(t) - x_2) \cdot \Delta t$

$dx_3 = x_2^*(t) \cdot (x_3^*(t) - x_3) \cdot \Delta t$

$dx_4 = x_2^*(t) \cdot (x_4^*(t) - x_4) \cdot \Delta t$

return

Function `dxStochastic`$(t,\ x,\ t',\ x^*,\ \eta)$

Allocate memory for $\{dx_i\}_{i=1}^4$ ▷ *array of displacements*

Draw 4 random unit Gaussian variables $\{\xi_i\}_{i=1}^4$

$dx_1 = 0$

$dx_2 = \eta_2 \cdot x_2^*(t) \cdot \sqrt{x_2^*(t)} \cdot \sqrt{2\Delta t} \cdot \xi_2$

$dx_3 = \eta_3 \cdot x_3^*(t) \cdot \sqrt{x_2^*(t)} \cdot \sqrt{2\Delta t} \cdot \xi_3$

$dx_4 = \eta_4 \cdot x_4^*(t) \cdot \sqrt{x_2^*(t)} \cdot \sqrt{2\Delta t} \cdot \xi_4$

return

Function `Initiate`$(t,\ x,\ N_{\mathrm{ori}},\ P,\ A,\ n_0^{\mathrm{I}})$

if $x_1 \cdot x_3 / N_{\mathrm{ori}} \geq n_0^{\mathrm{I}}$ **then**

$P = P + 1,\ N_{\mathrm{ori}} = 2 \cdot N_{\mathrm{ori}},\ A_P = t$

end

return

Function `Terminate`$(t,\ x,\ N_{\mathrm{ori}},\ P,\ A,\ C)$

if $P > 0$ **then**

if $t - A_1 > C$ **then** **for** $i = 1$ **to** P **do** $A_i = A_{i+1}$

$P = P - 1$

end

return

Function `Divide`$(t,\ x,\ n_0^{\mathrm{D}})$

if $x_1 \cdot x_4 \geq n_0^{\mathrm{D}}$ **then**

return true

else

return false

end

Parameter	Campos *et al.*	Notation
Growth rate	α	λ
Size at birth	L_b	S_b
Size at division	L_d	S_d
Added size from birth to division	ΔL	Δ_d
Added size between X-events	ΔL^*	Δ_d
Elapsed time between birth and division	T	τ
Elapsed time between X-event and division	δt	τ_X

Correspondence with notations in Campos *et al.* [S12].

IV. DISCUSSION ON THE PHASE-SHIFTED MODEL PREVIOUSLY REPORTED

A. Overview

In their paper [S12], Campos and colleagues presented experimental evidences of a "constant elongation model", stating that each individual cell grows in average of a constant mass between birth and division. This result is also known today as the adder principle [S13]. Comparison of the distributions of the added size and of the birth size between experimental data and simulations served to validate this model.

They also used their results to discredit the conjecture that replication initiation and division are coupled. Specifically, they considered the alternative hypothesis that instead cells would add a constant mass between specific events ("X-events") of the cell cycle, such as chromosome replication initiation. This defined a "phase-shifted model". By comparison with their experimental results, they rejected such as model and concluded that the "constant elongation model" must hold and that division is therefore not coupled to a replication initiation event. In this comparative analysis, two main points were put forward. (i) The distribution of the added size between cell birth and cell division, ΔL, and the distribution of the cell size at birth, L_b, were aberrantly broad in simulations of the phase-shifted model. These wide fluctuations were attributed by the authors to the fact that the number of X-events per generation could fluctuate a lot. (ii) The phase-shifted model resulted in correlations between mother/daughter cells for ΔL, in contradiction with the absence of correlations seen in experimental data. The authors argued that this is because the added size between X-events can overlap several generations in the phase-shifted model, resulting in correlation in ΔL.

In this section, we will show instead that the wide fluctuations obtained result from the choice of parameters for the phase-shifted model. Actually, it will appear that the cell size convergence in the phase-shifted model critically depends on the value of the phase shift. Specifically, it can deviate significantly from the adder convergence, and even become an unstable model. We conclude this discussion by suggesting an alternative model for the cell cycle. This model relies on an adder principle holding between replication initiation events, and assume that division and replication initiation are coupled. Yet for this model, cell size convergence is consistent with adder and the simulated data would be consistent with the experimental data from Campos and colleagues [S12]. Altogether, this suggests that discarding the co-regulation hypothesis between replication initiation and division based on the simulation results of the phase-shifted model is not reasonable.

B. Cell size convergence with the phase-shifted model

To be consistent with our manuscript we adopt notations different from Campos and colleagues. The correspondence between their and our notations are summarized in **Correspondence with notations in Campos *et al.***.

B1) Model

We now describe the phase-shifted model proposed by Campos and colleagues [S12]. First, they assumed that cells elongate their size exponentially according to Equation (1). Second, they introduced a cellular event, denoted by the lower script X, which determines division timing. Specifically, provided that a cellular event occured at time t_X, cell division is bound to happen at time $t_d = t_X + \tau_X$. Such an event does not necessarily coicinde with cell birth. Instead, it will typically represent chromosome replication initiation. Also, the cellular event triggering cell division may occur in the mother cell or other ancestors. In this model, cell division timing is therefore related to the timing of these specific cellular events, or "X-events". Third, they proposed that an X-event occurs when a fixed size Δ_d has been added since the last X-event. More accurately, we introduce the quantity $\Delta_X S$ which is reset to $\Delta_X S = 0$ when an X-event occurs, and otherwise increases according to:

$$\frac{\mathrm{d}\Delta_X S}{\mathrm{d}t} = \frac{\mathrm{d}S}{\mathrm{d}t}. \tag{61}$$

In particular, if the last X-event happened in the current generation, then:

$$\Delta_X S = S - S_X \tag{62}$$

Yet if the last X-event happened in the previous generation (say with index $n-1$), then:

$$\Delta_X S = (S - S_b) + (S_d^{(n-1)} - S_X^{(n-1)}). \tag{63}$$

Whenever the added size since the last X-event reaches a fixed quantity, $\Delta_X S = \Delta_X$, an X-event occurs and $\Delta_X S$ is reset to zero.

B2) Cell size convergence for small perturbations

We now investigate the convergence of cell size from a perturbed initial value in the phase shifted model. We distinguish two cases, depending on the value of τ_X compared to the generation time $\tau = \ln 2/\lambda$. Note that the results below are derived assuming that there is exactly one X-event per generation. In order for this assumption to hold, we restrain ourselves to small perturbations around the steady state cell size value. For larger perturbations, multiple X-events may occur in one generation, which will be dealt with numererically in the next section.

B3) Case 1: $0 \leq \tau_X < \tau$

In this scenario, the X-event leading to cell division occurs in the same generation. As such the division size is expressed as:

$$S_d^{(n)} = S_X^{(n)} \underbrace{e^{\lambda \tau_X}}_{\alpha}. \tag{64}$$

Therefore the convergence in cell size at division, is determined by the convergence of $S_X^{(n)}$. We now express the cell size at the X-event:

$$\begin{aligned} S_X^{(n)} &= S_X^{(n)} - \frac{1}{2} S_d^{(n-1)} + \frac{1}{2} S_d^{(n-1)}, \\ &= \Delta_X + \left(1 - \frac{\alpha}{2}\right) S_X^{(n-1)}, \end{aligned} \tag{65}$$

where we used that $S_X^{(n)} - S_d^{(n-1)}/2 = \Delta_X - (S_d^{(n-1)} - S_X^{(n-1)})$. This is a first order recurrent series. We obtain:

$$S_d^{(n)} = \left(S_d^{(0)} - S_d^{(\infty)}\right) r^n + S_d^{(\infty)}, \quad \text{with} \quad \begin{cases} S_d^{(\infty)} & = 2\Delta_X, \\ r & = 1 - \alpha/2. \end{cases} \tag{66}$$

Note that Equation (65) holds only when there is one X-event per generation. Namely, if $S_d^{(n-1)} - S_X^{(n-1)} > \Delta_X$, then $S_X^{(n)} = S_X^{(n-1)} + \Delta_X$ instead. In Equation (66), we see that cell size converges exponentially to the steady state value $S_X^{(\infty)}$. When $\tau_X = 0$, we obtain $r = 1/2$ which is the adder convergence. Indeed, in the latter case, the phase-shifted model reduces to the adder principle [S13]. However, when $\tau_X > 0$, then $r < 1/2$ and the convergence is faster than adder.

B4) Case 2: $\tau \leq \tau_X < 2\tau$

In this scenario, the X-event leading to cell division occurs in the previous cell generation because $\tau_X > \tau$. As such, the division size is expressed as:

$$S_d^{(n)} = \frac{1}{2} \underbrace{e^{\lambda \tau_X}}_{\alpha} S_X^{(n-1)}. \tag{67}$$

Similarly as before, we express the cell size at the X-event as:

$$S_X^{(n)} = S_X^{(n)} - \frac{1}{2} S_d^{(n-1)} + \frac{1}{2} S_d^{(n-1)}, \tag{68}$$

$$= \Delta_X + S_X^{(n-1)} - \frac{\alpha}{4} S_X^{(n-2)}. \tag{69}$$

Therefore, the series $S_{\mathrm{d}}^{(n)}$ satisfies the second order recurrence relation:

$$S_{\mathrm{d}}^{(n)} - S_{\mathrm{d}}^{(n-1)} + \frac{\alpha}{4} S_{\mathrm{d}}^{(n-2)} = \frac{\alpha}{2} \Delta_{\mathrm{X}}. \tag{70}$$

Equation (70) is solved using standard results on series. The homogeneous solution is obtained by considering the characteristics equation:

$$u^2 - u + \frac{\alpha}{4} = 0, \tag{71}$$

with imaginary solutions (because $\alpha > 1$):

$$u_{1/2} = \frac{1}{2} \pm \mathrm{i} \frac{\sqrt{\alpha - 1}}{2}. \tag{72}$$

The general solution must be a linear combination of the series $[u_1^n]$ and $[u_2^n]$. Using the particular solution $S_{\mathrm{d}}^{(n)} = 2\Delta_{\mathrm{X}}$ to Equation (70) and the fact that the solution must be real, we finally find the solution:

$$S_{\mathrm{d}}^{(n)} = A r^n \cos(n\theta + \varphi) + S_{\mathrm{d}}^{(\infty)}, \quad \text{with} \quad \begin{cases} S_{\mathrm{d}}^{(\infty)} & = 2\Delta_{\mathrm{X}}, \\ r & = \sqrt{\alpha}/2, \\ \tan\varphi & = \sqrt{\alpha - 1}, \end{cases} \tag{73}$$

where A and φ are two constants determined by the initial condition $(S_{\mathrm{d}}^{(0)}, S_{\mathrm{d}}^{(1)})$. Equation (73) defines a regime in which the convergence is slower than for adder since $r > 1/2$. In addition, the presence of oscillations in the response to perturbations to cell size suggests that in the presence of stochastic fluctuations the distribution of cell size would be quite large.

B5) Cell size convergence for general perturbations

As emphasized earlier, the analytical expressions Equations (66) and (73) are only valid for small perturbations. For larger perturbations, more than one X-event may occur per cell cycle. The actual generation time of individual cells during convergence may then vary significantly, resulting in the cell size convergence to be a combination of the scenari discussed previously. We investigated numerically cell size relaxation from a perturbed initial condition (**Deterministic cell size relaxation in the phase-shifted model**). We defined $\tau = \ln 2/\lambda = 1$ as unit of time and $\Delta_{\mathrm{d}} = 1$ as unit of size. We took the initial condition $S_{\mathrm{X}}^{(0)} = 4$. We observed cell size convergence in agreement with the analytical cases discussed above. In particular, for $\tau_{\mathrm{X}} = 0$ the cell size convergence is like adder. For $\tau_{\mathrm{X}} = 0.5\tau$ we find that the cell size converges faster than adder. Gradually as τ_{X} increases, the cell size convergence becomes slower than adder, and even oscillations appear.

B6) Comments

In summary, both analytical expressions and numerical simulations indicate that the cell size convergence in the phase-shifted model can deviate significantly from adder. This stems from the very definition of the phase-shifted model. In particular, for the values tested by Campos and colleagues [S12], $\tau_{\mathrm{X}} = 1.3\tau$ and $\tau_{\mathrm{X}} = 2.2\tau$, cell size convergence not only is slower than adder but also exhibits an oscillatory response to perturbations. As such, it is expected that the distribution of cell size in a stochastic implementation of this model will be very broad, which is one of the reasons invoked to reject the phase-shifted model and consequently refute the idea that division is controlled by chromosome replication initiation. However, this feels somehow excessive since there are other models implementing a control of division by initiation that would not lead to such an aberrant convergence property for cell size.

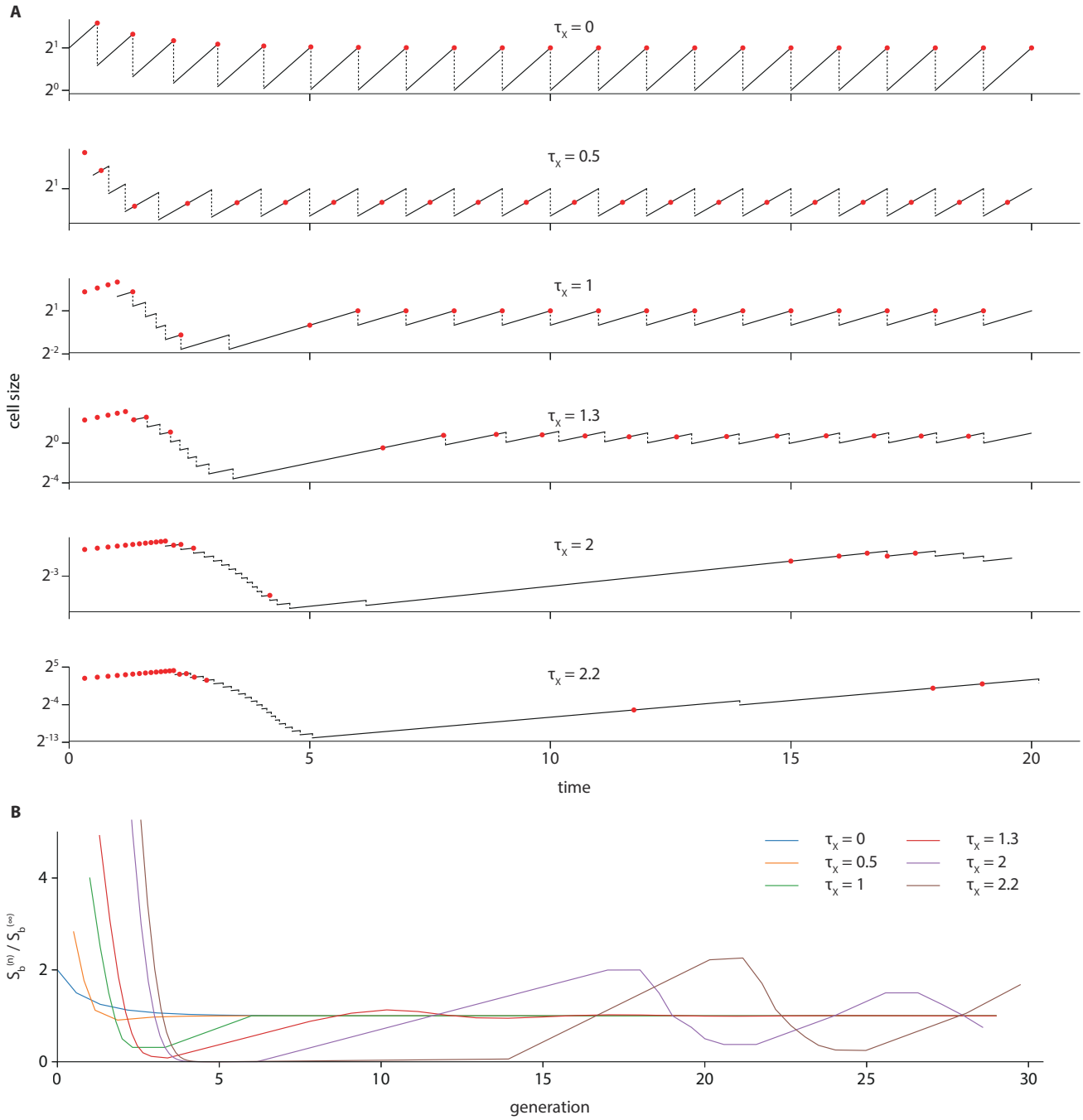
C. Alternative adder model for cell cycle based on replication initiation control

In this section, we present an alternative model for the cell cycle controlled by initiation events, yet satisfying the adder convergence for cell size and the absence of correlations for the added size. Namely, we consider that chromosome replication initiates after a fixed volume per origin of replication has been added since the last replication initiation. Since at division, the number of origins of replication is divided by two, Equations (65) and (69) become:

$$s_{\mathrm{i}}^{(n)} = s_{\mathrm{i}}^{(n-1)} + \delta_{\mathrm{i}}, \tag{74}$$

where $s_{\mathrm{i}} = S_{\mathrm{i}}/N_{\mathrm{ori}}$ is the volume per origin of replication at initiation and $\delta_{\mathrm{i}} = \Delta_{\mathrm{d}}/N_{\mathrm{ori}}$. This ensures that:

$$s_{\mathrm{i}}^{(n)} = (s_{\mathrm{i}}^{(0)} - \delta_{\mathrm{i}}) \frac{1}{2^n} + \delta_{\mathrm{i}}, \tag{75}$$



Deterministic cell size relaxation in the phase-shifted model [S12]. **(A)** Simulated lineages with different values of the phase shift τ_X. **(B)** Overlay of cell size convergence for the simulated lineages.

which is the adder convergence.

We now consider that each initiation event leads to a cell division event after a time equal to $\tau_{\rm cyc}$ has elapsed, hence assuming that division is regulated by chromosome replication initiation. In general, $\tau_{\rm cyc}$ may be larger than the generation time $\tau = \lambda/\ln 2$. Therefore, the size at cell division is given by:

$$S_{\rm d}^{(n)} = 2^p s_{\rm i}^{(n-p)} e^{\lambda \tau_{\rm cyc}}, \tag{76}$$

where p is the integer part of τ_{cyc}/τ. Note that the number of origin of replication is $N_{\rm ori} = 2^p$. For simplicity, here we assume that p is fixed, meaning that the replication initiation event leading to the cell division of the current generation always occurs

in the same relative ancestor (*e.g.* mother, grand-mother, *etc*.). The added size between division events is then expressed as:

$$\begin{aligned}\Delta_{\mathrm{d}}^{(n)} &= S_{\mathrm{d}}^{(n)} - S_{\mathrm{d}}^{(n-1)},\\ &= \underbrace{2^p e^{\lambda \tau_{\mathrm{cyc}}}}_{A} \delta_{\mathrm{i}}^{(n)}.\end{aligned} \tag{77}$$

We thus obtain that the mother/daughter correlation for the division adder is related to the mother/daughter correlation for the initiation adder:

$$\langle \Delta_{\mathrm{d}}^{(n+1)} \Delta_{\mathrm{d}}^{(n)} \rangle - \langle \Delta_{\mathrm{d}} \rangle^2 = A^2 \left(\langle \delta_{\mathrm{i}}^{(n+1)} \delta_{\mathrm{i}}^{(n)} \rangle - \langle \delta_{\mathrm{i}} \rangle^2 \right). \tag{78}$$

Provided that the added size per origin of replication in Equation (74) is uncorrelated to the next, we retrieve that the added size between divisions is uncorrelated from mother to daughter cells.

APPENDICES

Appendix A: Properties of Gaussian bivariate distributions

A1) Conditional probability

Let us consider two stochastic variables X and Y distributed according to a Gaussian bivariate distribution. We shall consider for simplicity that both X and Y are centered:

$$\langle X \rangle = 0, \qquad \langle Y \rangle = 0. \tag{79}$$

The distribution of a random Gaussian vector $\boldsymbol{R} = (X, Y)$ is characterized by the covariance matrix:

$$K_0 = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{pmatrix}. \tag{80}$$

The variance of X (resp. Y) is σ_{xx} (resp. σ_{yy}) and the covariance between variables X and Y is given by $\mathrm{cov}(X, Y) = \sigma_{xy}$. The Pearson correlation coefficient between variables X and Y is expressed as:

$$\rho = \frac{\sigma_{xy}}{\sqrt{\sigma_{xx}}\sqrt{\sigma_{yy}}}. \tag{81}$$

The probability distribution of the random vector $\boldsymbol{R} = (X, Y)$ is given by:

$$p(\mathbf{r}) = \frac{1}{(2\pi)\sqrt{\det(K_0)}} \exp\left(-\frac{1}{2}\mathbf{r}^T \cdot K_0^{-1} \cdot \mathbf{r}\right). \tag{82}$$

Denoting $p(x, y) = p(\mathbf{r})$, and using the definition for conditional probabilities: $p(x|y) = p(x, y)/p(y)$, we obtain:

$$p(x|y) = \frac{1}{Z} \exp\left(-\frac{1}{2}\frac{1}{\sigma_{xx}\sigma_{yy} - \sigma_{xy}^2}\left[\sigma_{yy}x^2 + \sigma_{xx}y^2 - 2\sigma_{xy}xy\right]\right), \tag{83}$$

where Z is a normalization constant depending on y. This normalization is obtained by ensuring that $\int \mathrm{d}x\, p(x|y) = 1$. We finally obtain:

$$p(x|y) = \frac{1}{\sqrt{2\pi\sigma(x|y)^2}} \exp\left(-\frac{1}{2}\frac{(x - E(x|y)^2)}{\sigma(x|y)^2}\right), \tag{84}$$

which is a Gaussian distribution with mean:

$$E(x|y) = \rho\sqrt{\frac{\sigma_{xx}}{\sigma_{yy}}}y, \tag{85}$$

and variance:

$$\sigma(x|y)^2 = \sigma_{xx}(1 - \rho^2). \tag{86}$$

When $\rho = 0$, we find that this Gaussian distribution does not depends on y. When $\rho = \pm 1$, the variance $\sigma(x|y)^2 \to 0$, *i.e.* x becomes a deterministic variable, with value y (or $-y$) if $\sigma_{xx} = \sigma_{yy}$.

A2) Correlation of the inverse

We first consider the random vector $f(\boldsymbol{R}) = (f(X), f(Y))$, where f is a quadratic function:

$$f(x) = \frac{1}{2}a_2x^2 + a_1x + a_0. \tag{87}$$

We now ask what is the Pearson correlation between variables $f(X)$ and $f(Y)$ given the correlation between variables X and Y. Introducing the vector $\Lambda = (\lambda, \eta)$, one can for instance consider the characteristic function:

$$\begin{aligned} \varphi(\Lambda) &= \left\langle e^{\lambda f(X) + \eta f(Y)} \right\rangle, \\ &= \int \mathrm{d}x\, \mathrm{d}y\, p(x, y) \exp\left(\Lambda^T f(R)\right), \\ &= e^{(\lambda+\eta)a_0} \int \mathrm{d}^2R\, \frac{1}{2\pi\sqrt{\det(K_0)}} \exp\left(-\frac{1}{2}R^T K_\Lambda^{-1} R + a_1\Lambda^T R\right), \\ &= e^{(\lambda+\eta)a_0} \sqrt{\frac{\det(K_\Lambda)}{\det(K_0)}} \exp\left(\frac{1}{2}a_1^2\Lambda^T K_\Lambda \Lambda\right), \end{aligned} \tag{88}$$

where we introduced the matrix:

$$K_\Lambda = \frac{1-\rho^2}{g(\lambda,\sigma_{xx})g(\eta,\sigma_{yy})-\rho^2}\begin{pmatrix}\sigma_{xx}g(\lambda,\sigma_{xx}) & \sigma_{xy}\\ \sigma_{xy} & \sigma_{xy}g(\eta,\sigma_{yy})\end{pmatrix}, \tag{89}$$

and the function $g(\lambda,\sigma^2) = 1-\lambda\sigma^2 a_2(1-\rho^2)$. We then obtain the desired correlations by taking the derivative of the logarithm of the characteristic function:

$$\begin{aligned}
\langle \delta f(X)^2\rangle &= \left.\frac{\partial^2 \ln\varphi}{\partial\lambda^2}\right|_{\lambda=0,\eta=0},\\
&= a_1^2\sigma_{xx} + \frac{1}{2}\sigma_{xx}^2 a_2^2.\\
\langle \delta f(Y)^2\rangle &= \left.\frac{\partial^2 \ln\varphi}{\partial\eta^2}\right|_{\lambda=0,\eta=0},\\
&= a_1^2\sigma_{yy} + \frac{1}{2}\sigma_{yy}^2 a_2^2.\\
\langle \delta f(X)\delta f(Y)\rangle &= \left.\frac{\partial^2 \ln\varphi}{\partial\lambda^2}\right|_{\lambda=0,\eta=0},\\
&= a_1^2\sigma_{xy} + \frac{1}{2}\sigma_{xx}\sigma_{yy} a_2^2\rho^2.
\end{aligned} \tag{90}$$

We therefore obtain for the Pearson correlation coefficient:

$$\rho(f(X),f(Y)) = \frac{\langle \delta f(X)\delta f(Y)\rangle}{\sqrt{\langle \delta f(X)^2\rangle}\sqrt{\langle \delta f(Y)^2\rangle}}, \tag{91}$$

$$= \frac{\rho\left(1+\frac{1}{2}\sqrt{\sigma_{xx}\sigma_{yy}}\left(\frac{a_2}{a_1}\right)^2\rho\right)}{\sqrt{1+\frac{1}{2}\sigma_{xx}\left(\frac{a_2}{a_1}\right)^2}\sqrt{1+\frac{1}{2}\sigma_{yy}\left(\frac{a_2}{a_1}\right)^2}}. \tag{92}$$

Equation (92) is an exact result. We see that when $\sqrt{\sigma_{xx}} \ll |a_1/a_2|$ and $\sqrt{\sigma_{yy}} \ll |a_1/a_2|$, the correlation between the transformed variables is equal to the correlation between the two variables: $\rho(f(X),f(Y)) \approx \rho(X,Y)$.

Let us consider now the case where the function of interest is the inverse function:

$$f(x) = \frac{1}{1+x}. \tag{93}$$

The result in Equation (92) does not strictly hold because Equation (93) is not a quadratic form. However, if the fluctuations in X are not too large, one might expect that the fluctuations of $f(X)$ around 1 are not too large either. In this case, one might approximate $f(x)$ to a Taylor expansion. We thus obtain a quadratic form as in Equation (87), with $a_2 = 2$, $a_1 = -1$ and $a_0 = 1$. Considering that the two variables have the same variance $\sigma_{xx} = \sigma_{yy} = \sigma^2$, we show in **Pearson correlation between $f(X)$ and $f(Y)$** that when $\sigma \ll |a_1/a_2| = 0.5$, the Pearson correlation of the transformed variables is approximately equal to the Pearson correlation between the two variables: $\rho(f(X),f(Y)) \approx \rho(X,Y)$ as long as σ is not too large.

Appendix B: Time-dependent growth rate in balanced growth

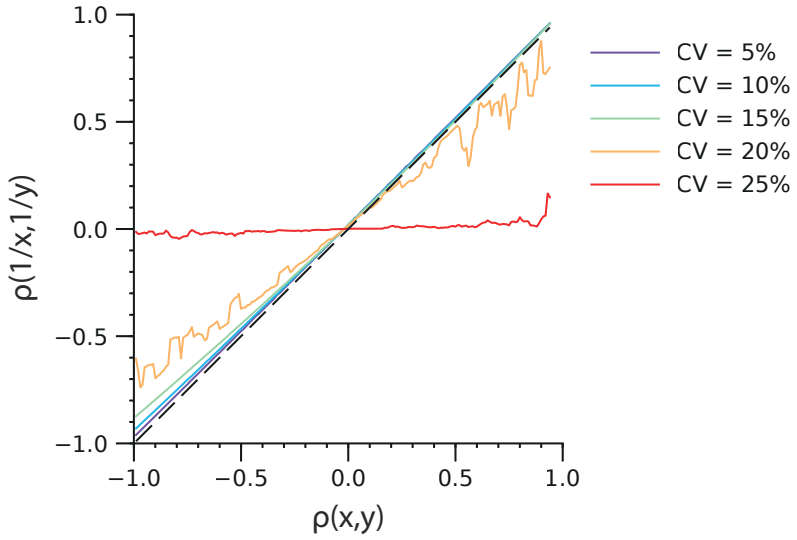
In this section, we derive the equation describing the time evolution of the instantaneous growth rate in single cells (*i.e.* the elongation rate for rod-shaped bacteria). We then generalize the equation describing concentration dynamics, namely Equation (39).

Let us denote M the dry mass of an individual cell. We assume that the total mass increase is directly proportional to the number of ribosomes in the cytoplasm. Thus we have:

$$\frac{\mathrm{d}M}{\mathrm{d}t}(t) = \sigma M_R(t), \tag{94}$$

where σ corresponds to the amount of new biomass produced per ribosome and M_R is the mass of ribosomes in the cell. Note that we assume that σ is constant, that is to say that the translation load of ribosomes is invariant through time. Furthermore, in balanced growth, a fixed fraction of the mass increase is allocated to ribosome synthesis:

$$\frac{\mathrm{d}M_R}{\mathrm{d}t}(t) = \phi_R^*\frac{\mathrm{d}M}{\mathrm{d}t}(t), \tag{95}$$



Pearson correlation between $f(X)$ and $f(Y)$ when $f(x)$ is the function in Equation (93). (X, Y) is a random vector distributed according to a Gaussian bivariate distribution, with means $\langle X \rangle = \langle Y \rangle = 1$, variances $\sigma_{xx} = \sigma_{yy} = \sigma^2$ and covariance $\sigma_{xy} = \rho(x, y)\sigma$. In this case, the coefficient-of-variation is $CV = \sigma$.

where ϕ_R^* is the fixed fraction of the mass flux allocated to ribosome synthesis. Introducing the instantaneous mass fraction of ribosomes $\phi_R(t) = M_R(t)/M(t)$, we obtain from Equation (94) the equation for exponential growth:

$$\frac{\mathrm{d}M}{\mathrm{d}t}(t) = \lambda(t)M(t), \quad \text{with} \quad \lambda(t) = \sigma\phi_R(t). \tag{96}$$

Using Equations (94) and (95), we obtain the equation describing the time-evolution of the instantaneous growth rate:

$$\frac{\mathrm{d}\lambda}{\mathrm{d}t} = \lambda(t)\left(\lambda^* - \lambda(t)\right), \tag{97}$$

where $\lambda^* = \sigma\phi_R^*$ is the steady-state growth rate.

Let us now consider a generic protein with instantaneous mass M_P in the cell. Again we assume that this protein is produced under balanced growth:

$$\frac{\mathrm{d}M_P}{\mathrm{d}t}(t) = \phi_P^* \frac{\mathrm{d}M}{\mathrm{d}t}(t), \tag{98}$$

where ϕ_P^* is the fixed fraction of the mass flux allocated to the biosynthesis of protein P. Using Equation (96), we find that the instantaneous mass fraction $\phi_P(t) = M_P(t)/M(t)$ satisfies:

$$\frac{\mathrm{d}\phi_P}{\mathrm{d}t}(t) = \lambda(t)(\phi_P^* - \phi_P(t)), \tag{99}$$

or in terms of concentrations:

$$\frac{\mathrm{d}c_P}{\mathrm{d}t}(t) = \lambda(t)(c_P^* - c_P(t)). \tag{100}$$

SUPPLEMENTAL REFERENCES

[S1] Si, F., Li, D., Cox, S.E., Sauls, J.T., Azizi, O., Sou, C., Schwartz, A.B., Erickstad, M.J., Jun, Y., Li, X., and Jun, S. (2017). Invariance of initiation mass and predictability of cell size in *Escherichia coli*. Curr. Biol. *27*, 1278–1287.

[S12] Campos, M., Surovtsev, I., Kato, S., Paintdakhi, A., Beltran, B., Ebmeier, S., and Jacobs-Wagner, C. (2014). A constant size extension drives bacterial cell size homeostasis. Cell *159*, 1433–1446.

[S13] Taheri-Araghi, S., Bradde, S., Sauls, J., Hill, N., Levin, P., Paulsson, J., Vergassola, M., and Jun, S. (2015). Cell-size control and homeostasis in bacteria. Curr. Biol. *25*, 385–391.

[S14] Wallden, M., Fange, D., Lundius, E., Baltekin, Ö., and Elf, J. (2016). The synchronization of replication and division cycles in individual *E. coli* cells. Cell *166*, 729–739.

[S15] Cooper, S., and Helmstetter, C.E. (1968). Chromosome replication and the division cycle of *Escherichia coli* B/r. J. Mol. Biol. *31*, 519–540.

[S16] Donachie, W.D. (1968). Relationship between cell size and time of initiation of DNA replication. Nature *219*, 1077–1079.

[S17] Mott, M.L., and Berger, J.M. (2007). DNA replication initiation: mechanisms and regulation in bacteria. Nat. Rev. Microbiol. *5*, 343–354.

[S18] Katayama, T., Ozaki, S., Keyamura, K., and Fujimitsu, K. (2010). Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and oriC. Nat. Rev. Microbiol. *8*, 163–170.

[S19] Skarstad, K., and Katayama, T. (2013). Regulating DNA replication in bacteria. Cold Spring Harb. Perspect. Biol. *5*, a012922.

[S20] Sekimizu, K., Yung, B.Y., and Kornberg, A. (1988). The DnaA protein of *Escherichia coli*. abundance, improved purification, and membrane binding. J. Biol. Chem. *263*, 7136–40.

[S21] Messer, W. (2002). The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. FEMS Microbiol. Rev. *26*, 355–374.

[S22] Sompayrac, L., and Maaløe, O. (1973). Autorepressor model for control of DNA replication. Nature *241*, 133–135.

[S23] Chien, A.-C., Hill, N.S., and Levin, P.A. (2012). Cell size control in bacteria. Curr. Biol. *22*, R340–R349.

[S24] Moore, D.A., Whatley, Z.N., Joshi, C.P., Osawa, M., and Erickson, H.P. (2017). Probing for binding regions of the FtsZ protein surface through site-directed insertions: Discovery of fully functional FtsZ-fluorescent proteins. J. Bacteriol. *199*, e00553–16.

[S25] Scott, M., Gunderson, C.W., Mateescu, E.M., Zhang, Z., and Hwa, T. (2010). Interdependence of cell growth and gene expression: Origins and consequences. Science *330*, 1099–1102.

[S26] Scott, M., and Hwa, T. (2011). Bacterial growth laws and their applications. Curr. Opin. Biotechnol. *22*, 559–565.