

# UC San Diego

## UC San Diego Previously Published Works

### Title

Bacterial Replication Initiation as Precision Control by Protein Counting.

### Permalink

<https://escholarship.org/uc/item/8fh4k2wv>

### Authors

Fu, Haochen

Xiao, Fangzhou

Jun, Suckjoon

### Publication Date

2023

### DOI

10.1103/prxlife.1.013011

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



# HHS Public Access

Author manuscript

*PRX Life*. Author manuscript; available in PMC 2024 March 28.

Published in final edited form as:

*PRX Life*. 2023 ; 1(1): . doi:10.1103/prxlife.1.013011.

## Bacterial Replication Initiation as Precision Control by Protein Counting

Haochen Fu<sup>\*</sup>,

Fangzhou Xiao<sup>\*</sup>

Department of Physics, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA

Suckjoon Jun<sup>†</sup>

Department of Physics and Department of Molecular Biology, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA

### Abstract

Balanced biosynthesis is the hallmark of bacterial cell physiology, where the concentrations of stable proteins remain steady. However, this poses a conceptual challenge to modeling the cell-cycle and cell-size controls in bacteria, as prevailing concentration-based eukaryote models are not directly applicable. In this study, we revisit and significantly extend the initiator-titration model, proposed 30 years ago, and we explain how bacteria precisely and robustly control replication initiation based on the mechanism of protein copy-number sensing. Using a mean-field approach, we first derive an analytical expression of the cell size at initiation based on three biological mechanistic control parameters for an extended initiator-titration model. We also study the stability of our model analytically and show that initiation can become unstable in multifork replication conditions. Using simulations, we further show that the presence of the conversion between active and inactive initiator protein forms significantly represses initiation instability. Importantly, the two-step Poisson process set by the initiator titration step results in significantly improved initiation synchrony with  $CV \sim 1/N$  scaling rather than the standard  $1/\sqrt{N}$  scaling in the Poisson process, where  $N$  is the total number of initiators required for initiation. Our results answer two long-standing questions in replication initiation: (i) Why do bacteria produce almost two orders of magnitude more DnaA, the master initiator proteins, than required for initiation? (ii) Why does DnaA exist in active (DnaA-ATP) and inactive (DnaA-ADP) forms if only the active form is competent for initiation? The mechanism presented in this work provides a satisfying general solution to how the cell can achieve precision control without sensing protein concentrations, with broad implications from evolution to the design of synthetic cells.

---

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

<sup>†</sup>Corresponding author: suckjoon.jun@gmail.com.

<sup>\*</sup>These authors contributed equally to this work.

## I. INTRODUCTION

Most biology textbooks explain biological decision-making by emphasizing the control and sensing of key protein concentrations through programmed gene expression and protein degradation in eukaryotes. Protein concentration gradients can encode spatial or temporal information across different scales, such as morphogen gradients in the French flag model in developmental biology [1], or cyclin oscillations in eukaryotic cell-cycle controls [2] [Fig. 1(a)]. However, in bacterial cell physiology, balanced biosynthesis has been the hallmark since the 1950s at the population and single-cell levels [3–5]. Balanced biosynthesis means that the synthesis rate of all cellular components is the same as the cell’s growth rate in steady-state growth, wherein the concentrations of stable proteins are steady by the balance of their production and dilution [Fig. 1(b)].

However, balanced biosynthesis poses a fundamental conceptual challenge to modeling the cell-cycle and cell-size controls, as the prevailing concentration-based models are not directly applicable if the concentration of cell-cycle proteins remains constant (within stochasticity). Indeed, for the billion-year divergent model bacterial organisms *Escherichia coli* and *Bacillus subtilis*, their size control is based on (i) balanced biosynthesis of division initiator protein FtsZ and (ii) its accumulation to a threshold number (not concentration) [6]. These two conditions lead to the adder phenotype [6]. Unfortunately, a mechanistic investigation of threshold FtsZ number sensing is a formidable challenge because division initiation involves multiple interacting proteins with unknown properties [7].

Replication initiation in bacteria, which is exclusively controlled by the widely conserved master regulator protein, DnaA, is an attractive problem for mechanistic investigation because it exhibits the adder phenotype [8–11]. That is, the added cell size between two consecutive initiation events is independent of the cell size at initiation, as originally suggested by Sompayrac and Maaloe [12]. The adder phenotype implies that cells likely accumulate the DnaA molecules to a threshold number [6], and the synthesis of DnaA is balanced [13]. Furthermore, DnaA has been extensively studied, and most properties required for modeling are known or can be estimated [14–18]. Therefore, we view *E. coli* replication initiation as a tractable problem to understand the mechanism of protein copy-number sensing to control the cell cycle and cell size, and gain mechanistic insight into the general class of precision control in biology.

In this work, we revisit and significantly extend the initiator-titration model proposed by Hansen, Christensen, and Atlung 30 years ago [19], the model closest to the protein-number-sensing idea (see Sec. II A). In Sec. II A, we summarize the original initiator-titration model and introduce our initiator-titration model v2. In Sec. II B, we first introduce the “procell” model, a minimal version of the initiator-titration model, and derive the first expression of the procell size at initiation (known as the “initiation mass”). In Sec. II C, we perform a dynamical stability analysis of the procell model and show the existence of initiation instability. In Sec. IID, we extend the procell to our “initiator-titration model v2” and derive an analytical expression for the initiation mass in a special case (the  $\Delta 4$  mutant [13]) based on three mechanistic biological control parameters: the expression level of DnaA, the ratio of the active versus passive forms of DnaA, namely  $[\text{DnaA-ATP}]/[\text{DnaA-ADP}]$ ,

and the number of DnaA titration boxes on the chromosome. In the same section, we show that adding the replication-dependent, biologically observed DnaA-ATP  $\rightarrow$  DnaA-ADP conversion element (RIDA) restores initiation stability [20,21]. In Sec. IIE, we discuss initiation asynchrony and cell-to-cell variability using the concept of intrinsic and extrinsic noise in the framework of initiator-titration model v2.

Our model provides a quantitative and mechanistic explanation for several long-standing questions in bacterial replication initiation with the following findings: DnaA titration boxes are the protein-counting device that measures the threshold number of initiator proteins, and the two forms of DnaA (DnaA-ATP and DnaA-ADP), and especially the replication-dependent DnaA-ATP  $\rightarrow$  DnaA-ADP, are needed to suppress initiation instability. Given the fundamental nature of replication initiation and its profound differences from eukaryotic cell-cycle control, we anticipate broad applications of our results, from the design of synthetic cells to the evolution of biological mechanisms in precision control.

## II. RESULTS AND DISCUSSION

### A. The “initiator-titration model v2 “ and intuition

Consider engineering a synthetic cell capable of self-replication. For such a cell to be viable, it must meet a fundamental requirement for cell-cycle control: initiating replication only once during cell division. A possible “simple” strategy to implement this requirement could be as follows [Fig. 2(a)]: (i) The chromosome has one origin of replication. (ii) The cell produces one initiator protein during the division cycle. (iii) The initiator protein binds to *ori* (the replication origin) and immediately triggers initiation. (iv) Upon initiation, the cell destroys the initiator protein. While this seemingly straightforward strategy could limit the replication origin to a single site and produce a single initiator protein during cell division, the underlying mechanisms required to achieve this are likely more complex. For instance, how would the cell “know” when to produce the initiator protein and when to degrade it?

While *E. coli* exhibits characteristics similar to the hypothetical strategy described above, there are notable differences. *E. coli* has one replication origin (*ori*), but replication initiation requires 10–20 master regulator DnaA molecules binding to the 11 DnaA boxes at *ori* [15–17,22]. Furthermore, DnaA is stable and not degraded upon initiation [15,16]. Strikingly, *E. coli* produces approximately 300 copies of DnaA per *ori*, or 30 times more than required at *ori*, with almost all being titrated by DnaA boxes encoded on the chromosome [15,16].

In 1991, Hansen and colleagues proposed the initiator-titration model to explain these observations [Fig. 2(b)] [19]. Their model posits that DnaA is first titrated by high-affinity DnaA boxes on the chromosome, which allows it to bind *ori* with weak affinity and initiate replication only after the chromosomal DnaA boxes are nearly saturated. This highlights the importance of DnaA boxes on the chromosome as the timing device for replication initiation.

Our model builds upon the initiator-titration model and incorporates the knowledge in DnaA accumulated in the past 30 years [15–17,22]. Specifically, we have learned that DnaA exists in two forms, DnaA-ATP and DnaA-ADP, with different binding affinities to DNA [23].

DnaA-ATP is the active form that can trigger initiation, while DnaA-ADP is inactive as it cannot bind *ori* specifically [24,25]. Further genetic, biochemical, and bioinformatic studies have revealed that approximately 300 high-affinity DnaA boxes are distributed across the circular chromosome [15,26,27]. By contrast, *ori* contains a cluster of 11 DnaA binding sites, wherein only three have high affinities [25,28]. Therefore, most DnaA, whether DnaA-ATP or DnaA-ADP, will first bind the high-affinity chromosomal DnaA boxes. Only after the titration step do DnaA-ATP molecules bind the weak binding sites within *ori* and trigger initiation. We refer to this updated model as the initiator-titration model v2, in recognition of the pioneering work of Hansen *et al.* [19,29].

Figure 2(c) illustrates how our initiator-titration model v2 works in more detail. To provide intuition without losing the generality of our ideas, let us consider a naked circular chromosome without bound DnaA.

(1) As DnaA binds to ATP or ADP tightly [23] and the cellular concentration of ATP is almost 10× higher than ADP [30,31], newly synthesized DnaA molecules become DnaA-ATP. During steady-state growth, both DnaA-ATP and DnaA-ADP exist in the cell due to multiple interconversion mechanisms [16]. (See Sec. IID and Appendix D for a detailed discussion.)

(2) DnaA-ATP and DnaA-ADP will first bind to around 300 high binding-affinity chromosomal DnaA boxes ( $K_D \approx 1$  nM) [26], whereas only DnaA-ATP can bind to around 10 low-affinity boxes within *ori* ( $K_D \approx 100$  nM) [26,32].

(3) When most chromosomal DnaA boxes are saturated, the probabilities for DnaA-ATP binding to *ori* versus the remaining chromosomal DnaA boxes become comparable. Initiation is triggered once the low-affinity *ori* binding sites are saturated by DnaA-ATP.

As we elaborate below, the initiator-titration model v2 answers two long-standing fundamental questions:

- (1) Why does *E. coli* produce so many more DnaA proteins than required for initiation, only to be titrated?
- (2) Why does *E. coli* maintain two forms of DnaA in the first place if they only need DnaA-ATP for initiation?

## B. The “protocell”: A minimal initiator-titration model

To gain analytical insight, we first construct a minimal initiator-titration model, named “protocell” [Fig. 3(a)]. The protocell has the complexity between the two versions of the initiator-titration model [Figs. 2(b) and 2(c)]. The protocell has one *ori*, the active initiator protein (e.g., DnaA-ATP in *E. coli*), and the initiator binding sites on the chromosome. We assume the following based on the experimental data:

- (1) The cell grows exponentially  $V(t) = V_0 e^{\lambda t}$  in steady-state [3], where  $V(t)$  is the total cell size at time  $t$ , and  $\lambda$  is the growth rate. The mass-doubling time  $\tau$  is given by  $\tau = \frac{\ln 2}{\lambda}$ .

(2) Synthesis of the initiator protein is balanced, i.e., its concentration is constant during growth [3]. We denote the initiator protein copy number at time  $t$  as  $I(t)$  and its concentration as  $c_i$ .

(3) The rate of DNA synthesis is constant [35,36], with the duration of chromosome replication  $C$ , independent of the mass-doubling time  $\tau$  [37].

(4) The chromosome encodes specific DNA sequences for binding of the initiator proteins.  $N_B$  high-affinity sites are evenly distributed on the chromosome [15], and  $n_B$  low-affinity sites are localized at *ori* [16]. For the *E. coli* chromosome, we set  $N_B = 300$  and  $n_B = 10$ , as explained in Sec. II A. During replication, the total number of initiator binding sites increases as  $B(t)$ .

(5) Initiators tightly bind to the binding sites rather than staying in the cytoplasm, and initiators preferentially bind the chromosomal binding sites before binding to the ones at *ori*. Therefore, replication initiates at  $t = t_{\text{ini}}$  when  $I(t = t_{\text{ini}}) = B(t = t_{\text{ini}})$ , i.e., all binding sites are saturated by the initiator proteins.

For illustration purposes, we consider an intermediate growth condition, where two cell cycles slightly overlap without exhibiting multifork replication [37] [Fig. 3(b)]. In the Helmstetter-Cooper model [38], this corresponds to  $C < \tau < C + D$ , where  $D$  is the duration between replication termination and cell division. As such, the cell can have two intact chromosomes between termination and the next initiation [Fig. 3(b)].

The steady-state curves of  $I(t)$  and  $B(t)$  are shown in Fig. 3(b) (in our model, a steady state means all derived quantities are periodic with a period of  $\tau$ ). In general,  $I(t)$  increases exponentially because of exponential growth and balanced biosynthesis (Assumptions 1 and 2 above), whereas  $B(t)$  increases piecewise linearly because of replication initiation and termination (Assumptions 3 and 4). Therefore, the number of initiators catches up with the total number of binding sites between replication termination and the new round of initiation at  $I(t = t_{\text{ini}}) = B(t = t_{\text{ini}}) = 2(N_B + n_B)$  (Assumption 5). Here, the factor “2” refers to the fact that there are two entire chromosomes and two *ori*'s right before the initiation event in the specific growth condition depicted in Fig. 3(b). Upon initiation, the number of binding sites  $B(t)$  increases discontinuously by  $2n_B$  due to the duplication of both *ori*'s and the binding sites therein. After that,  $B(t)$  increases at the rate  $2N_B/C$ , steeper than the slope of  $I(t)$ . Once the cell divides,  $I(t)$  and  $B(t)$  drop by half, and the cell repeats its cycle.

From this picture, the initiation mass  $v_i$ , defined by cell volume per *ori* at initiation [37], can be easily calculated by the number of initiators at initiation,

$$v_i = \frac{I(t_{\text{ini}})}{2c_i} = \frac{1}{c_i}(N_B + n_B), \quad (1)$$

where  $c_i$  is the initiator protein concentration, and “2” reflects the copy number of *ori* before initiation.

The above result can be extended to different growth conditions. For example, in slow growth ( $\tau > C + D$ ), the replication cycles do not overlap, and all the factors “2” will vanish in the above analysis due to the single chromosome at initiation. This results in the same initiation mass  $v_i$  as in the intermediate growth condition. In fast-growth conditions ( $\tau < C$ ), replication cycles overlap, exhibiting multifork replication. Since a new round of replication starts before the previous round of replication is completed, the initiation mass is given by

$$v_i = \frac{1}{c_i}(\alpha N_B + n_B), \quad (2)$$

with the cell-cycle-dependent parameter  $\alpha \leq 1$  given as

$$\alpha = \frac{1}{2^n} + \left(2 - \frac{n+2}{2^n}\right) \frac{\tau}{C}, \quad n = \left\lfloor \frac{C}{\tau} \right\rfloor, \quad (3)$$

which applies to any growth conditions (see Appendix A for a derivation).  $\alpha = 1$  when  $\tau \geq C$  (non-multifork replication), and  $0 < \alpha < 1$  when  $\tau < C$  (multifork replication) [Fig. 3(c)]. Thus,  $\alpha$  refers to the degree of overlapping replication. Some of the most salient predictions of these results include the following: (i) The initiation mass is inversely proportional to the initiator concentration  $c_i$ , and (ii) the initiation mass linearly depends on the number of chromosomal binding sites  $N_B$ .

The basis of the protocell’s behavior is that the initiator increases exponentially, whereas the number of binding sites increases piecewise linearly only during DNA replication. This allows the cell to reach the initiation point  $I(t) = B(t)$  from any initial conditions. Therefore, the protocell can always trigger initiation by protein number counting through titration.

### C. The protocell exhibits initiation instability

In the preceding section, we addressed whether a solution exists in the minimal protocell model with a period of  $\tau$ . We showed that this periodic solution always exists [Eq. (2)]. We defined it as the “steady-state” solution in the biological sense that the cell can grow in a steady state with the periodic cell cycle. However, since the model is dynamic, convergence to a steady state from a given initial condition,  $I(0)$  and  $B(0)$ , is not guaranteed. Hence, in this section, we study how the replication cycle propagates in the lineage from an arbitrary initial condition at  $t = 0$ , and under what conditions the cycle converges to the steady-state solution.

Intuitively, if the two consecutive initiations are separated by  $\tau$ , thus periodic, the system is in a steady state. Suppose an initiation event at  $t = 0$ , and its initiation mass deviates from the steady-state solution Eq. (2). Typically, the next initiation occurs at  $t = t^+ \neq \tau$ . However, if this time interval between two consecutive initiations eventually converges to  $\tau$  after

generations, the steady-state solution is stable under perturbations on the initial conditions. Otherwise, the steady-state solution is unstable.

In the rest of this section, we analyze a dynamical system based on Assumptions 1–5 in Sec. II B on the protocell.

**1. Setup**—We consider a protocell containing one chromosome with ongoing multifork replication [Fig. 4(a)]. We block the cell division so the protocell grows indefinitely as the chromosome replicates and multiplies starting from the initial condition. As the cell size approaches infinity, does the initiation mass have a fixed value (stable) or multiple values (unstable)? The analysis is nontrivial, as we need to accommodate arbitrary initial conditions.

To this end, we start with the dynamics of  $I(t)$  and  $B(t)$ . First, we have

$$I(t) = I(0)e^{\lambda t}, \quad (4)$$

as a consequence of exponential cell growth and balanced biosynthesis of the initiator proteins. Next, the dynamics of the number of binding sites  $B(t)$  is more subtle because it increases piecewise linearly depending on the replication state of the chromosome and the number of replication forks. To accommodate the possibility of arbitrary initial conditions, we define the “multifork tracker” vector variable,  $\rho(t)$ , as follows:

$$\rho(t) \equiv \begin{cases} [\rho_1(t), \rho_2(t), \dots, \rho_d(t)] & \text{if } d \geq 1, \\ 0 & \text{if } d = 0. \end{cases} \quad (5)$$

Here, the index  $d$  is the total number of generations (namely, the total rounds of replication cycles) since the initial chromosome, so  $d$  can grow indefinitely with time. That is, at every new round of the replication cycle, the size of the vector increases by 1 from  $d$  to  $d + 1$ .  $d = 0$  is for the initial cell that is supposed to have an intact single chromosome without ongoing replications.

We use the variable  $\rho$  to indicate the relative position of a replication fork of interest between *ori* and *ter* (the replication terminus), and therefore  $0 \leq \rho(t) \leq 1$  [Fig. 4(b)]. For example,  $\rho$  would be 0.5 if a pair of forks is exactly halfway between *ori* and *ter* [Figs. 4(a) and 4(b)]. To track multifork replication, we use  $\rho_i(t)$  to represent the group of replication forks that are the  $i$ th closest to the *ori* [Fig. 4(a)]. For example,  $i = 1$  always refers to the newest group of replication forks. To record the replication history, we set  $\rho_i(t) = 1$  for those replication forks that have already reached *ter* [Fig. 4(b)]. By these definitions,  $\rho(t)$  applies to both multifork replication and non-multifork replication.

Based on the multifork tracker vector, the number of binding sites  $B(t)$  is completely determined by  $\rho$  as



$$B(\rho(t)) = N_B \left[ 1 + \sum_{i=1}^d \rho_i(t) 2^{d-i} \right] + 2^d n_B. \quad (6)$$

The dynamics of  $\rho(t)$  consists of two parts: First, between two initiation events,  $\rho_i(t)$  increases linearly with a slope of  $1/C$  until it reaches 1, as replication forks travel from *ori* to *ter* [Fig. 4(b)]. Second, at initiation, the dimension of  $\rho$  increases by 1, shifting its components to the right as  $S: \mathbb{R}^d \rightarrow \mathbb{R}^{d+1}, (\rho_1, \rho_2, \dots, \rho_d) \mapsto (0, \rho_1, \rho_2, \dots, \rho_d)$  to accommodate the new pair of replication forks at each *ori* [see also Fig. 4(a)].

**2. Properties of the steady state**—The steady-state solution assumes periodicity of dynamics so that  $I(t)$  and  $B(t)$  double in each replication cycle. We consider the mapping between two consecutive initiation events to solve for the steady-state condition. We denote the first initiation event as  $\rho(t=0) = \rho$  at  $t=0$ , and the second initiation event as  $\rho(t=t^+) = \rho^+$  at  $t=t^+$ . The mapping  $\mathcal{F}: \mathbb{R}^{d-1} \rightarrow \mathbb{R}^d, \rho \mapsto \rho^+$  requires a time-translation and a shift:

$$\rho_i^+ = \begin{cases} \rho_{i-1} + \frac{t^+}{C} & \text{if } \rho_{i-1} + \frac{t^+}{C} < 1, \\ 1 & \text{otherwise,} \end{cases} \quad (7)$$

where the initiation time  $t^+$  is determined by the initiation criteria that  $I(t=0) = B(t=0)$  and  $I(t=t^+) = B(t=t^+)$ , Eqs. (4) and (6),

$$\frac{e^{\lambda t^+}}{2} \left[ N_B \left( 2^{-(d-1)} + \sum_{i=1}^{d-1} \rho_i 2^{-i} \right) + n_B \right] = N_B \left( 2^{-d} + \sum_{i=1}^d \rho_i^+ 2^{-i} \right) + n_B. \quad (8)$$

Equations (7) and (8) describe the dynamics of the system at initiation. We can now obtain the fixed point of the mapping  $\mathcal{F}$  by setting  $d \rightarrow \infty$  and  $\rho^+ = \rho$  (Appendix B):

$$t^+ = \tau, \rho_i^{\text{ss}} = \begin{cases} i \frac{\tau}{C} & \text{if } i \leq \left\lfloor \frac{C}{\tau} \right\rfloor, \\ 1 & \text{otherwise.} \end{cases} \quad (9)$$

The resulting expression for steady-state initiation mass is the same as Eq. (2), i.e., the fixed point of  $\mathcal{F}$  is the steady-state solution (see Appendix B for more details).

Next, we study the stability of the fixed point of  $\mathcal{F}$  by calculating the Jacobian matrix of  $\mathcal{F}$  at the fixed point:

$$J = \left. \frac{\partial \rho_i^+}{\partial \rho_j} \right|_{\text{ss}}. \quad (10)$$

This matrix can be reduced to an  $n \times n$  matrix ( $n = \lfloor \frac{C}{\tau} \rfloor$ ), since all other matrix elements are zero. In *E. coli*,  $0 \leq n \leq 2$  in most growth conditions; here, we consider the range of  $0 \leq n \leq 3$  to accommodate cells theoretically doubling as frequently as at every  $\tau = 10$  min, with a C period of 40 min. Therefore, we can calculate the eigenvalues of  $J$  for each  $n$ . Stability requires the largest eigenvalue of  $J$  to be smaller than 1. Eventually, we can obtain the stable and unstable regimes in the  $n_b/N_b$  versus  $C/\tau$  phase diagram, as shown in Fig. 4(c) [see Appendix B, and also Fig. 8(a)]. Importantly, the phase diagram reveals both stable ( $n < 1$ ) and unstable (small  $n_b/N_b$  when  $n > 1$ ) steady states [Fig. 4(c)].

What happens when the system becomes unstable? As discussed earlier, in fast growth conditions,  $\alpha < 1$  in the steady-state initiation mass expression [Eq. (2)]. Indeed, using numerical simulations, we found that the initiation mass oscillates between two values [Fig. 4(c)]. This indicates that the cell cycle can oscillate between multifork and non-multifork replication. Mathematically, this oscillatory behavior means that the fixed points of  $\mathcal{F}^{o2} = \mathcal{F} \circ \mathcal{F}$  are stable, although the fixed point of  $\mathcal{F}$  is unstable. By fixing one of the fixed points of  $\mathcal{F}^{o2}$  as  $\rho_1 = 1$ , we can compute the other fixed point with  $\rho_1 < 1$  (see Appendix C). In extreme cases,  $\rho_1$  can be as small as 0.1. That is, the second round of replication starts only after 10% of the chromosome has been replicated by the replication forks from the previous initiation. When the replication forks from two consecutive rounds of initiation are too close to each other, they cannot be separated into two division cycles. This should result in two initiation events in one division cycle, and no initiation in the next division cycle.

Therefore, although initiation triggering is guaranteed, the performance of the protocell is imperfect in terms of initiation instability in certain growth conditions. We show how the initiator-titration model v2 resolves the instability issue in Sec. II D.

#### D. The initiator-titration model v2: Replication-dependent DnaA-ATP $\rightarrow$ DnaA-ADP conversion stabilizes the cell cycle

In the previous section, we showed that the protocell can show initiation instability. In understanding why wild-type *E. coli* initiation is stable, we have to consider unique features of DnaA in *E. coli*, namely its two distinct forms: the active DnaA-ATP and the inactive DnaA-ADP [23]. Several extrinsic elements, categorized into two main groups, interconvert between these DnaA forms [16,20,21,39–42] [Fig. 5(a)].

The first group catalyzes the conversion of DnaA-ATP  $\rightarrow$  DnaA-ADP. This includes the Regulatory Inactivation of DnaA (RIDA) [20,21] and datA-dependent DnaA-ATP Hydrolysis (DDAH) [39,40]. RIDA's functionality requires active replication forks [43], thus rendering it replication-dependent, while DDAH's *datA*, a DnaA binding chromosomal locus, fosters DnaA-ATP hydrolysis.

By contrast, the second group, comprised of DARS1 and DARS2 (types of DnaA Reactivating Sequences), facilitates the ADP → ATP exchange for DnaA-ADP [41,42].

Importantly, DnaA harbors intrinsic ATPase activity that facilitates its own conversion from DnaA-ATP → DnaA-ADP [23,28], a feature not depicted in Fig. 5(a). Intriguingly,  $\Delta 4$  cells—cells with a full deletion of all extrinsic DnaA-ATP ↔ DnaA-ADP interconversion pathways—exhibit a nearly identical initiation phenotype to that of wild-type cells [13], solely relying on DnaA’s intrinsic ATPase activity.

Figure 5(b) presents our numerical simulation results illuminating the alterations in the [DnaA-ATP]/[DnaA-ADP] ratio during the cell cycle in a DNA replication-dependent manner (see Appendix E for simulation details). Notably, this ratio should remain steady in  $\Delta 4$  cells during cell elongation in a steady state [13]. Furthermore, reinitiation is prohibited within a certain time frame postinitiation (approximately 10 min; the “eclipse period”), attributable to the sequestration of newly synthesized DNA by SeqA [44].

In this section, we integrate each of these features into our protocell model to formulate our initiator-titration model v2 and compute the initiation stability phase diagram. Our findings reveal that the replication-dependent DnaA-ATP → DnaA-ADP conversion by RIDA largely alleviates initiation instability, thus reinstating the stability characteristic of wild-type cells.

**1. Analytical expression of the initiation mass in the initiator-titration model 2 with a constant DnaA-ATP/DnaA-ADP ratio ( $\Delta 4$  cells)**—First, we incorporate the two forms of DnaA with the intrinsic DnaA-ATP → DnaA-ADP activity by DnaA into the protocell model to construct the  $\Delta 4$  cells, a minimal version of the initiator-titration model v2 [Fig. 2(c)]. As noted earlier, both DnaA-ATP and DnaA-ADP can bind the chromosomal DnaA boxes because of their strong binding affinity ( $K_D \sim 1$  nM [15,26,45]), whereas only DnaA-ATP can bind the weak DnaA boxes at *ori* with  $K_D \sim 10^2$  nM [15,26,46]. With the same Assumptions 1–4 in Sec. II B and this additional assumption, we can derive an analytical expression for steady-state initiation mass for  $\Delta 4$  *E. coli* (see Appendix D for the derivation):

$$v_i = \frac{\alpha N_B + \left(1 + \frac{[\text{DnaA-ADP}]}{[\text{DnaA-ATP}]}\right)^{n_B}}{[\text{DnaA}] - \left(1 + \frac{[\text{DnaA-ADP}]}{[\text{DnaA-ATP}]}\right)^{n_B} K_{\text{eff}}^{n_B}}, \quad (11)$$

where  $K_{\text{eff}}$  is the effective dissociation constant of DnaA at *ori*, and  $\alpha$  is in Eq. (3). Therefore, this equation brings together the expression level of DnaA via [DnaA], the ratio [DnaA-ATP]/[DnaA-ADP], and the degree of overlapping replication ( $\alpha$ ).

Note that, under physiological conditions,  $K_{\text{eff}}^{n_B} \ll [\text{DnaA}]$  (Appendix D). If [DnaA-ATP]  $\gg$  [DnaA-ADP], all DnaA molecules are in their active form DnaA-ATP, and the  $\Delta 4$  *E. coli* converges to the protocell [i.e., Eq. (11) converging to Eq. (2)].

**2. The  $\Delta 4$  cells show initiation instability**—We also investigated the initiation stability of the  $\Delta 4$  cells using numerical simulations [Fig. 5(c)] (see Appendix E for simulation details). The initiation stability phase diagram is analogous to that of the protocells in Fig. 4(c), showing an island of instability regime. This occurs during the transition into multifork replication, wherein initiation mass alternates between two values. Importantly, changing the [DnaA-ATP]/[DnaA-ADP] ratio does not significantly impact the stability [Fig. 8(b)].

**3. Replication-dependent DnaA-ATP  $\rightarrow$  DnaA-ADP by RIDA alone can restore initiation stability**—Next, we implemented the extrinsic DnaA-ATP  $\leftrightarrow$  DnaA-ADP conversion elements in the  $\Delta 4$  cells. In contrast to the constant [DnaA-ATP]/[DnaA-ADP] in  $\Delta 4$ , the extrinsic conversion elements induce temporal modulations in [DnaA-ATP]/[DnaA-ADP] during cell elongation [47]. This ratio reaches its maximum at initiation and its minimum at termination due to the activation/deactivation of the RIDA mechanism [Fig. 5(b)] [48].

We also investigated the initiation stability across growth conditions [Fig. 5(c) and Fig. 9 in Appendix E]. Among all the known extrinsic conversion elements we tested, the replication-dependent DnaA-ATP  $\rightarrow$  DnaA-ADP by RIDA alone was sufficient to restore initiation stability [Fig. 5(c) and Fig. 9 in Appendix E]. Other elements only had mild effects on the stability. RIDA is replication-dependent, thus it immediately decreases the level of DnaA-ATP upon initiation. This reduction in the initiation-competent DnaA-ATP level is likely the reason for suppressing premature reinitiation.

Although we found RIDA to be the initiation stabilizer, it still significantly delays initiation due to the reduced level of DnaA-ATP. Our simulations show that the delayed initiation can be alleviated by the other DnaA-ADP  $\rightarrow$  DnaA-ATP conversion elements without causing instability [Fig. 5(c) and Fig. 9 in Appendix D]. Interestingly, the initiation mass becomes nearly invariant across a wide range of growth conditions in the presence of all four extrinsic conversion elements [Fig. 5(c)], as long as the concentration [DnaA] is growth-condition-independent. We previously used this growth-condition-independent [DnaA] hypothesis to explain the invariance of initiation mass [37], and the data so far support the hypothesis [33,34].

Based on these results, we conclude that the replication-dependent DnaA-ATP  $\rightarrow$  DnaA-ADP by RIDA can significantly enhance the initiation stability, and the other DnaA-ADP  $\rightarrow$  DnaA-ATP conversion elements keep the initiation mass nearly constant against physiological perturbations.

**4. The eclipse period or origin sequestration does not improve stability**—We also tested the effect of the eclipse period [44] in our simulations (see Fig. 10 in Appendix E). During the predefined eclipse period, we did not allow the binding of the initiator to *ori*. Surprisingly, the eclipse period did not improve stability significantly in the multifork replication regime. However, the amplitude of the initiation mass oscillation decreased slightly (Fig. 10 in Appendix E). Therefore, we predict the effect of SeqA on steady-state stabilization to be modest.

**5. Comparison with previous modeling by Berger and ten Wolde and recent experimental work**—In their recent study, Berger and ten Wolde [11] conducted a thorough investigation into *E. coli* DNA replication. They utilized extensive numerical simulations that factored in the known dynamics between DnaA-ATP and DnaA-ADP conversion, as well as the aspects of DnaA titration. To our knowledge, Berger and ten Wolde were the first to suggest possible instability during multifork replication.

Under relatively fast growth conditions (with the doubling time 35 min and the  $C$  period 40 min), their observations noted oscillations in the initiation mass between two distinct values, which occurred in the absence of DnaA-ATP  $\leftrightarrow$  DnaA-ADP conversion. Our instability phase diagram [Fig. 4(c)] explains this observation. For example, in the case of  $\Delta 4$  mutant cells, the initiation mass should oscillate between two values when  $1 < C/\tau < 1.8$  [Fig. 5(c)]. However, the complexity of these instability regimes needs to be noted. Our phase diagrams show that multifork replication does not always lead to instability [Fig. 5(c) and Fig. 8 in Appendix B]

Berger and ten Wolde propose the DnaA-ATP  $\leftrightarrow$  DnaA-ADP conversion as the key mechanisms in initiation control, as DnaA-ATP  $\leftrightarrow$  DnaA-ADP conversion could avoid initiation instability in the absence of titration boxes in their simulations. By contrast, we favor the idea that titration plays a more fundamental role in initiation control, because it is the protein counting device in the protocell and also the  $\Delta 4$  cells, where DnaA-ATP  $\leftrightarrow$  DnaA-ADP conversion is absent. Furthermore, titration boxes, which are prevalent in bacteria, ensure synchronous initiation (as explained in Sec. IIE), and they explain why bacteria produce significantly more DnaA molecules than necessary for *ori*. Although titration is fundamental in our model, its performance is not perfect in terms of initiation instability, and we demonstrated that RIDA is the key conversion element required for initiation stability when titration is in place.

While the details of molecular effects on initiation are beyond the scope of this theory work, we suggest recent work by Elf and colleagues [49] and by us on various deletion mutants, including  $\Delta 4$  [13], for single-cell level experimental investigation as confirmations of some of our predictions.

## E. Asynchrony and cell-to-cell variability of initiation in the initiator-titration framework

Initiation stability raises a related issue of stochasticity in initiation. In the systems biology literature, “noise” is mainly discussed in the context of stochasticity in gene expression, decomposed into “intrinsic” versus “extrinsic” components [50–53]. In our view, there are parallel observations in replication initiation: the initiation asynchrony among *ori*'s within the same cell [13,54,55], and the cell-to-cell variability of the initiation mass [6,13,56], as illustrated in Fig. 6(a). In this section, we discuss their origins and statistical properties within our initiator-titration model v2 framework.

**1. Definition of the intrinsic and extrinsic noise**—During overlapping cell cycles, the cell contains multiple replication origins at initiation. These origins share the same biochemical environment within one cell, so their initiation events are correlated; the initiation timing in different cells can vary because of stochasticity in biological processes,

such as gene expression [50,53]. On the other hand, since these origins in the same cell do not interact with each other, they can initiate asynchronously due to the innate stochasticity of initiator accumulation at origins [29,54].

To quantify initiation asynchrony and cell-to-cell variability, we consider two overlapping replication cycles. Suppose the two *ori*'s initiate at initiation mass  $v_i^{(1)}$  and  $v_i^{(2)}$ , respectively. Similar to stochastic gene expression [50], we can define the intrinsic noise and the extrinsic noise of the initiation mass by the coefficient of variation as

$$CV_{\text{int}}^2 = \frac{\langle (v_i^{(1)} - v_i^{(2)})^2 \rangle}{2\langle v_i^{(1)} \rangle \langle v_i^{(2)} \rangle}, \quad CV_{\text{ext}}^2 = \frac{\langle v_i^{(1)} v_i^{(2)} \rangle - \langle v_i^{(1)} \rangle \langle v_i^{(2)} \rangle}{\langle v_i^{(1)} \rangle \langle v_i^{(2)} \rangle}. \quad (12)$$

Note that this definition fulfills the relation  $CV_{\text{tot}}^2 = CV_{\text{int}}^2 + CV_{\text{ext}}^2$ , where  $CV_{\text{tot}}$  is the coefficient of variation of the single variable  $v_i^{(1)}$  or  $v_i^{(2)}$  (Appendix F).

We use  $CV_{\text{int}}$  as a measure of asynchrony. Visually,  $CV_{\text{int}}$  describes the width of the off-diagonal axis of the ellipsoid, while  $CV_{\text{ext}}$  describes the elongation extent of the diagonal axis compared to the short axis [Fig. 6(a)]. For example, if  $CV_{\text{ext}} = 0$ ,  $v_i^{(1)}$  and  $v_i^{(2)}$  are fully uncorrelated, and the ellipsoid becomes a circle. In this case, the intrinsic noise is the sole source of cell-to-cell variability. Generally, while asynchrony is fully determined by the intrinsic noise, the cell-to-cell variability is a result of both the intrinsic noise and the extrinsic noise (see Appendix F for details).

**2. A first-passage-time model based on a one-step Poisson process**—To study the behavior of the extrinsic noise and the intrinsic noise, we convert the initiation mass variables,  $v_i^{(1)}$  and  $v_i^{(2)}$ , into first-passage-time (FPT) variables [57],  $T^{(1)}$  and  $T^{(2)}$ , respectively. That is, the initiator proteins bind to binding sites at *ori*, increasing its occupancy  $O(t)$ , and they initiate replication as soon as *ori* is fully saturated [ $O(t) = n_b$ ]. Although the relation between  $v_i$  and FPT is nonlinear, to the zeroth-order approximation, we have

$$CV_{\text{int}}^2 \approx \frac{\langle (T^{(1)} - T^{(2)})^2 \rangle}{2\langle T^{(1)} \rangle \langle T^{(2)} \rangle}, \quad CV_{\text{ext}}^2 \approx \frac{\langle T^{(1)} T^{(2)} \rangle - \langle T^{(1)} \rangle \langle T^{(2)} \rangle}{\langle T^{(1)} \rangle \langle T^{(2)} \rangle}. \quad (13)$$

To obtain the scaling law of the noise of FPT, we assume the production of initiator proteins as a Poisson process with a constant production rate  $\beta$  [53]. We further assume that all cells are characterized by the same set of physiological parameters without noise. (By this assumption, we are considering the lower bound of the extrinsic noise, and we discuss the contribution of parameter noises in Sec. IIE 4.)

Let us first consider a simple scenario of initiation without initiator-titration. In this scenario, there is no chromosomal binding site; all  $n_b$  binding sites are localized at each *ori*, and the initiator protein has an equal probability of binding to either *ori*. That is, the two *ori*'s accumulate the initiator proteins independently. This results in uncorrelated  $T^{(1)}$  and  $T^{(2)}$  and hence  $CV_{\text{ext}} = 0$  based on Eq. (13). The intrinsic noise then becomes

$$CV_{\text{int}} = \frac{\sigma_{T^{(1)}}}{\langle T^{(1)} \rangle}, \quad (14)$$

where  $\langle T^{(1)} \rangle$  is the mean FPT at *ori*1 and  $\sigma_{T^{(1)}}$  is the standard deviation.

In this simplest scenario, the accumulation at *ori*1 is a Poisson process with a rate of  $\beta$  followed by a binomial trial with equal probability, leading to a Gamma distribution of  $T^{(1)}$ , with the mean  $\langle T^{(1)} \rangle = 2n_b/\beta$  and the standard deviation  $\sigma_{T^{(1)}} = 2\sqrt{n_b}/\beta$  (see Appendix G). Thus, the  $CV_{\text{int}}$  is independent of  $\beta$  [57,58],

$$CV_{\text{int}} = \frac{1}{\sqrt{n_b}} = \sqrt{\frac{2}{N}}, \quad (15)$$

where  $N = 2n_b$  is the mean total number of initiator proteins needed for triggering initiation at both *ori*'s (Appendix G).

Therefore, in this one-step Poisson process, the intrinsic noise of FPT scales with the square root of the required total number of initiators  $N$ . If the number of binding sites at *ori* is  $n_b \approx 10$ , we have  $CV_{\text{int}} \approx 30\%$  [Fig. 6(b)]. If the cell localizes all  $N_b \approx 300$  DnaA boxes at each *ori* to increase the threshold, the noise will decrease to  $CV_{\text{int}} = 1/\sqrt{300} \approx 6\%$  [Fig. 6(c)].

The reason for the  $1/\sqrt{N}$  intrinsic noise scaling is that the stochasticity in gene expression fully propagates to the initiation timing, and  $T^{(1)}$  and  $T^{(2)}$  are uncorrelated. As we explain below, *E. coli* suppresses the intrinsic noise using an ingenious two-step Poisson process by compressing  $T^{(1)}$  and  $T^{(2)}$  into a narrow range during the cell cycle using titration. In other words, titration of the initiator proteins redirects most gene expression noise to the extrinsic noise, effectively synchronizing  $T^{(1)}$  and  $T^{(2)}$ .

### 3. A two-step Poisson process in the initiator-titration framework predicts the $1/N$ scaling of the intrinsic noise, leading to initiation synchrony—

Due to the significant differences in the binding affinity between the chromosomal binding sites ( $K_D \approx 1$  nM) and *ori* ( $K_D \approx 100$  nM), *E. coli* titrates DnaA sequentially in two steps: (i) saturation of the  $\sim N_b$  chromosomal DnaA boxes by DnaA-ATP and DnaA-ADP, followed by (ii) accumulation of DnaA-ATP at *ori* with  $n_b \ll N_b$  binding sites. Thus, we modify the one-step Poisson process by adding the titration step, namely a two-step Poisson process

[Fig. 6(c)]. The first step delays the accumulation processes at *ori1* and *ori2* and they synchronize their initiations, and the intrinsic noise (asynchrony) is a result of stochasticity in the second step.

To analyze the two-step Poisson process, we rewrite the two FPT variables  $T^{(1)}$  and  $T^{(2)}$  as  $T^{(1)} = T^{(0)} + \Delta T^{(1)}$ ,  $T^{(2)} = T^{(0)} + \Delta T^{(2)}$ . Here,  $T^{(0)}$  is the time required to saturate the chromosomal binding sites, whereas  $\Delta T^{(1)}$  and  $\Delta T^{(2)}$  denote the additional respective times for the two *ori*'s to accumulate the initiator proteins to trigger initiation. We assume that  $T^{(0)}$ ,  $\Delta T^{(1)}$ , and  $\Delta T^{(2)}$  are three independent stochastic variables. Specifically,  $T^{(0)}$  follows the original Poisson process with an accumulation rate of  $\beta$ , while  $\Delta T^{(1)}$  and  $\Delta T^{(2)}$  each independently follows the same Poisson process with an accumulation rate of  $\beta/2$  (initiator proteins produced at the rate  $\beta$  bind the two *ori*'s), as derived in Appendix G. By this decomposition, Eq. (13) can be rewritten as

$$CV_{\text{int}} = \frac{\sigma_{\Delta T^{(1)}}}{\langle T^{(1)} \rangle}, \quad CV_{\text{ext}} = \frac{\sigma_{T^{(0)}}}{\langle T^{(1)} \rangle}. \quad (16)$$

According to the corresponding Gamma distributions, the mean FPT reads  $\langle T^{(1)} \rangle = N/\beta$ , where  $N$  is the mean total number of initiator proteins needed for triggering initiation at both *ori*'s; the standard deviation of the first-step FPT reads  $\sigma_{T^{(0)}} = \sqrt{N - 2n_B}/\beta$ , and the standard deviation of the second-step FPT reads  $\sigma_{\Delta T^{(1)}} = 2\sqrt{n_B}/\beta$  (see Appendix G). Therefore, based on Eq. (16), we obtain the CV's scaling law as

$$CV_{\text{int}} = \frac{2\sqrt{n_B}}{N}, \quad CV_{\text{ext}} = \frac{\sqrt{N - 2n_B}}{N} \approx \frac{1}{\sqrt{N}}. \quad (17)$$

This result indicates that  $CV_{\text{int}}$  decays in  $\sim 1/N$ , much faster than the total noise  $CV_{\text{tot}} \sim 1/\sqrt{N}$ , and  $CV_{\text{ext}}$  becomes the dominant noise component when  $N$  is large. For example, if  $n_B = 10$ ,  $N_B \approx 300$ , and  $N \approx 2(N_B + n_B) \approx 600$  (two overlapping cell cycles), the noise of the two-step processes decreases dramatically from  $\sim 30\%$  to only 1%, while the extrinsic noise is around 4%.

To test the predictions of the two-step Poisson process, we conducted a simulation by considering a Poissonian protein production followed by a partitioning among three destinations: chromosomal binding sites, *ori1*, and *ori2* (see Appendix G for model settings). As shown in Fig. 6(d), the scaling behavior of the intrinsic noise and the extrinsic noise is consistent with Eq. (17).

In summary, the chromosomal titration boxes effectively synchronize the accumulation of DnaA-ATP at multiple *ori*'s by titration, compressing their initiation timing into a narrow temporal window during the cell cycle [54,59]. This is consistent with long-standing experimental observations of synchronous initiation of minichromosomes [60,61], and more



recent observations of ectopic chromosomal origins [62]. This improvement in precision by two sequential binding processes is reminiscent of the ratchetlike kinetic proofreading model, and our results are generalizable.

**4. Other noise sources not quantified in this work**—In the previous section, we have mainly discussed asynchrony and cell-to-cell variability in initiation resulting from stochastic protein production, which predicts  $CV_{\text{int}} \approx 1\%$  and  $CV_{\text{ext}} \approx 4\%$  in *E. coli*. However, experimentally measured  $CV_{\text{int}}$  is about 3–4% [13] and  $CV_{\text{tot}}$  is about 10% [6,13,56], both larger than the prediction. For mutants lacking DnaA-ATP  $\leftrightarrow$  DnaA-ADP conversion elements, the cell-to-cell variability can increase up to 20% [13]. The likely sources of additional asynchrony and cell-to-cell variability are as follows.

For the intrinsic noise, the initiator accumulation at the two *ori*'s can be negatively correlated because of the new round of replication. Once the first initiation event is triggered at one *ori*, the newly produced DnaA boxes will titrate DnaA, and the newly activated RIDA decreases the DnaA-ATP pool [Fig. 5(b)], further delaying the initiation of the second *ori*. This anticorrelation between two asynchronous initiations should increase the intrinsic noise  $CV_{\text{int}}$ .

For the extrinsic noise, we suggest two extra main sources other than the  $1/\sqrt{N}$  for titration by Poisson process [Eq. (17)]: (i) cell-to-cell variability in the initiator concentration [53,63,64], DnaA-ATP/ADP ratio, doubling time [9,65], and *C* period [37,66], and (ii) the growth-condition (*C*/ $\tau$ )-dependent initiation instability discussed in Sec. II D [Figs. 4(c) and 5(c)]. In principle, even without noises in *C* period and doubling time, instability can cause a bimodal distribution of the initiation mass that significantly increases the extrinsic noise [11]. The noise caused by initiation instability should be significant in mutants without the RIDA mechanism, such as the  $\Delta 4$  cells [13].

For the quantification of these noise contributions, we leave a more detailed analysis to future work.

### III. CONCLUSION AND PERSPECTIVE

In this work, we have provided a comprehensive quantitative explanation of how bacteria control the cell cycle under balanced growth, particularly focusing on replication initiation as a tractable problem. Our analysis builds upon the original initiator-titration model proposed by Hansen and colleagues [19], which offered valuable insights into the two-step initiation process.

Over the past three decades, significant progress has been made in understanding the conserved master replication initiator protein, DnaA. One perplexing aspect has been the coexistence of two forms of DnaA (DnaA-ATP and DnaA-ADP), with only DnaA-ATP being initiation-competent. Expanding upon the original model by Hansen and colleagues, we developed the initiator-titration model v2, which incorporates the two-state DnaA model and accounts for DnaA box distribution. We have derived an analytical expression for the initiation mass in terms of three mechanistic parameters for DnaA: its concentration, the

average ratio  $[DnaA-ATP]/[DnaA-ADP]$ , and the number of DnaA titration boxes [Eq. (11)]. However, through our dynamical stability analysis, we have also revealed a previously unexplored instability in initiation within this model [Fig. 4(c)], thereby elucidating recent observations from numerical simulations by Berger and ten Wolde [11]. We have demonstrated that the replication-dependent  $DnaA-ATP \rightarrow DnaA-ADP$  conversion (by RIDA) alone restores initiation stability [67]. Additionally, when considering all extrinsic  $DnaA-ATP \leftrightarrow DnaA-ADP$  elements, the initiation mass remains remarkably invariant across a wide range of growth conditions, in agreement with experimental observations [34,37,68].

Moreover, we have discovered that the titration process of the chromosomal DnaA boxes suppresses the intrinsic noise or asynchrony in initiation by  $CV \sim 1/N$  scaling. This finding represents a significant improvement over the naively expected standard coefficient of variation scaling  $CV \sim 1/\sqrt{N}$  for a Poisson process. It underscores the extraordinary consequences of the two-step initiation processes in the initiator-titration models, highlighting the remarkable precision achieved by bacteria.

In conclusion, we propose that titration may have been a pivotal evolutionary milestone, acting as a protein-counting mechanism that co-evolved with balanced biosynthesis. This system would not only enable bacteria to homeostatically control their size via the adder principle, but also lead to synchronous initiation by effectively separating titration and initiation in two steps. Our results thus illuminate how bacteria employ a seemingly straightforward yet efficient titration-based strategy to address fundamental biological challenges. This differentiates them from eukaryotes, which use programed gene expression and protein degradation to sense and control protein concentrations. While our findings focus on a specific case of initiation control, they also trigger intriguing questions about the potential pervasiveness of titration-based precision control in diverse biological systems. Uncovering additional examples of such mechanisms will significantly advance our overall understanding of precision control and pave the way for practical applications, including the design of synthetic cells.

## ACKNOWLEDGMENTS

We thank Flemming Hansen, Tove Atlung, Tsutomu Katayama, Anders Lobner-Olesen, Godefroid Charbon, Thias Boesen, Johan Elf, Dongyang Li, Fangwei Si, Guillaume Le Treut, Cara Jensen, Mareike Berger, Pieter-Rein ten Wolde, Alan Leonard, Julia Grimwade, Conrad Woldringh, Charles Helmstetter, and Willie Donachie for many invaluable discussions and exchange of ideas over the years that inspired and helped shape the ideas presented in this work. This work was supported by NSF MCB-2016090 and NIH MIRA (R35GM139622) to S.J.

## APPENDIX A: DERIVATION OF THE STEADY-STATE INITIATION MASS FORMULA IN THE PROTOCELL MODEL

We consider one generation from cell birth to cell division in the steady state. First, due to exponential growth and balanced biosynthesis of DnaA, we have

$$V(t) = V(0)e^{\lambda t}, \tag{A1}$$

$$I(t) = I(0)e^{\lambda t}, \quad (\text{A2})$$

where  $V(t)$  is the cell volume,  $I(t)$  is the number of initiators,  $\lambda = \frac{\ln 2}{\tau}$  is the growth rate, and  $0 \leq t \leq \tau$ .

We denote the number of chromosomal binding sites as  $\tilde{B}(t)$ . The behavior of  $\tilde{B}(t)$  is complicated and depends on the ratio of  $C/\tau$ . Suppose  $n\tau \leq C \leq (n+1)\tau$ , where  $n \equiv \left\lfloor \frac{C}{\tau} \right\rfloor$ . The shape of  $\tilde{B}(t)$  also depends on the relative timing of initiation and termination. For example, Fig. 7(a) shows when  $n = 0$  and  $t_{\text{ini}} > t_{\text{ter}}$ , and Fig. 7(b) shows when  $n = 1$  and  $t_{\text{ini}} < t_{\text{ter}}$ . We will discuss the two cases  $t_{\text{ini}} \leq t_{\text{ter}}$  and  $t_{\text{ini}} > t_{\text{ter}}$  separately.

### 1. $t_{\text{ini}} \leq t_{\text{ter}}$

As illustrated in Fig. 7(c), in this case the termination time is determined by  $\tau$ ,  $C$ , and  $t_{\text{ini}}$  as

$$t_{\text{ter}} = t_{\text{ini}} + C - n\tau. \quad (\text{A3})$$

The condition for this situation is  $t_{\text{ter}} \leq \tau$ , which gives

$$t_{\text{ini}} \leq (n+1)\tau - C. \quad (\text{A4})$$

As shown in Fig. 7(c), the curve  $\tilde{B}(t)$  consists of three segments with different slopes. By mathematical induction, we can obtain the expression of each slope,

$$k_1 = (2^n - 1)k, \quad 0 \leq t < t_{\text{ini}}, \quad (\text{A5a})$$

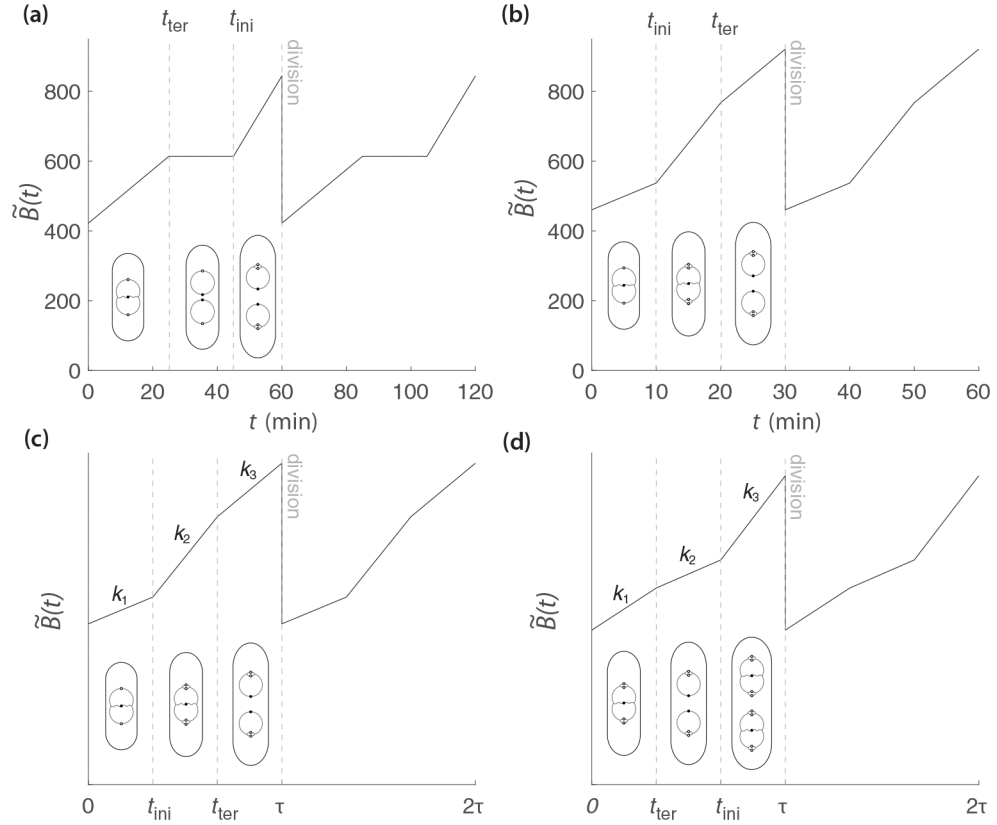
$$k_2 = (2^{n+1} - 1)k, \quad t_{\text{ini}} \leq t < t_{\text{ter}}, \quad (\text{A5b})$$

$$k_3 = (2^{n+1} - 2)k, \quad t_{\text{ter}} \leq t \leq \tau, \quad (\text{A5c})$$

where  $k$  is the replication rate of the DnaA boxes by a pair of replication forks, i.e., following Assumptions 3 and 4,

$$k \equiv \frac{N_B}{C}.$$

(A6)



**FIG. 7.** The time trajectory of the number of chromosomal initiator binding sites (titration boxes)  $\tilde{B}(t)$  in different scenarios in the steady state. (a)  $n = 0$  and  $t_{ini} > t_{ter}$ . (b)  $n = 1$  and  $t_{ini} < t_{ter}$ . (c) The general case when  $t_{ini} \leq t_{ter}$ . (d) The general case when  $t_{ini} > t_{ter}$ .

The initial value  $\tilde{B}(0)$  is  $\tilde{B}(\tau)/2$ , hence we have

$$\tilde{B}(0) = k_1 t_{ini} + k_2 (t_{ter} - t_{ini}) + k_3 (\tau - t_{ter}).$$

Using Eqs. (A3) and (A5), we obtain

$$\tilde{B}(0) = N_B \left[ 1 + (2^{n+1} - n - 2) \frac{\tau}{C} - (2^n - 1) \frac{t_{ini}}{C} \right].$$

(A7)

Once we obtain the expressions of  $\tilde{B}(0)$  and the slopes  $k_1$ ,  $k_2$ , and  $k_3$ ,  $\tilde{B}(t)$  is fully determined.

Of our particular interest, we have

$$\tilde{B}(t_{\text{ini}}) = N_{\text{B}} \left[ 1 + (2^{n+1} - n - 2) \frac{\tau}{C} \right], \quad (\text{A8})$$

which is independent of  $t_{\text{ini}}$ .

## 2. $t_{\text{ini}} > t_{\text{ter}}$

As illustrated in Fig. 7(d), in this case the termination time can be written as

$$t_{\text{ter}} = t_{\text{ini}} + C - (n + 1)\tau. \quad (\text{A9})$$

The condition for this situation is  $t_{\text{ter}} \geq 0$ , which gives

$$t_{\text{ini}} > (n + 1)\tau - C. \quad (\text{A10})$$

Similarly, we obtain

$$k_1 = (2^{n+1} - 1)k, \quad 0 \leq t < t_{\text{ter}}, \quad (\text{A11a})$$

$$k_2 = (2^{n+1} - 2)k, \quad t_{\text{ter}} \leq t < t_{\text{ini}}, \quad (\text{A11b})$$

$$k_3 = (2^{n+2} - 2)k, \quad t_{\text{ini}} \leq t \leq \tau, \quad (\text{A11c})$$

and the initial value  $\tilde{B}(0)$  reads

$$\begin{aligned} \tilde{B}(0) &= k_1 t_{\text{ter}} + k_2 (t_{\text{ini}} - t_{\text{ter}}) + k_3 (\tau - t_{\text{ini}}) \\ &= N_{\text{B}} \left[ 1 + (2^{n+2} - n - 3) \frac{\tau}{C} - (2^{n+1} - 1) \frac{t_{\text{ini}}}{C} \right]. \end{aligned} \quad (\text{A12})$$

Likewise, Eqs. (A11) and (A12) allow us to calculate  $\tilde{B}(t)$  at any time. In particular, at initiation,

$$\tilde{B}(t_{\text{ini}}) = 2N_B \left[ 1 + \left( 2^n + 1 - n - 2 \right) \frac{\tau}{C} \right], \quad (\text{A13})$$

which is again independent of  $t_{\text{ini}}$ .

### 3. Initiation mass formula

Now  $\tilde{B}(t)$  can be fully determined by the given parameters. Let us go back to Eqs. (A1) and (A2). To calculate the initiation mass, we only need to calculate  $I(t_{\text{ini}})$ . Based on our initiation criteria,  $I(t_{\text{ini}}) = B(t_{\text{ini}}) = \tilde{B}(t_{\text{ini}}) + n_B \cdot \# \text{ ori}$ , we have

$$I(t_{\text{ini}}) = \begin{cases} \tilde{B}(t_{\text{ini}}) + 2^n n_B, & t_{\text{ini}} \leq (n+1)\tau - C, \\ \tilde{B}(t_{\text{ini}}) + 2^{n+1} n_B, & t_{\text{ini}} > (n+1)\tau - C. \end{cases} \quad (\text{A14})$$

The cell volume at initiation reads

$$V(t_{\text{ini}}) = \frac{I(t_{\text{ini}})}{c_1}, \quad (\text{A15})$$

and the initiation mass, by definition, reads

$$v_i \equiv \frac{V(t_{\text{ini}})}{\# \text{ ori}} = \begin{cases} \frac{V(t_{\text{ini}})}{2^n}, & t_{\text{ini}} \leq (n+1)\tau - C, \\ \frac{V(t_{\text{ini}})}{2^{n+1}}, & t_{\text{ini}} > (n+1)\tau - C. \end{cases} \quad (\text{A16})$$

By plugging Eqs. (A15), (A14), (A8), and (A13) into this formula, we obtain the full expression of the initiation mass,

$$v_i = \frac{1}{c_1} \left\{ \left[ \frac{1}{2^n} + \left( 2 - \frac{n+2}{2^n} \right) \frac{\tau}{C} \right] N_B + n_B \right\}, \quad n = \left\lfloor \frac{C}{\tau} \right\rfloor. \quad (\text{A17})$$

This is exactly Eq. (2) in Sec. II B in the main text. As we can see, the initiation mass is independent of the initiation time  $t_{\text{ini}}$ .

## APPENDIX B: DERIVATION OF THE STABILITY REGIMES FOR THE INITIATION MAPPING $\mathcal{F}$

We start from the initiation mapping  $\mathcal{F}: \mathbb{R}^{d-1} \rightarrow \mathbb{R}^d$ ,  $\rho \mapsto \rho^+$  derived in Sec. II C in the main text,

$$\rho_i^+ = \begin{cases} \rho_{i-1} + \frac{t^+}{C} & \text{if } \rho_{i-1} + \frac{t^+}{C} < 1, \\ 1 & \text{otherwise,} \end{cases} \quad (\text{B1})$$

where the initiation time  $t^+$  is determined by

$$N_B \left( 2^{-d} + \sum_{i=1}^d \rho_i^+ 2^{-i} \right) + n_B = \frac{e^{\lambda t^+}}{2} \left[ N_B \left( 2^{-(d-1)} + \sum_{i=1}^{d-1} \rho_i 2^{-i} \right) + n_B \right]. \quad (\text{B2})$$

### 1. Steady state

First, note that for any finite  $d$ , the image space of the map  $\mathcal{F}$  is different from its preimage space, so there is no fixed point of  $\mathcal{F}$ . However, we can consider the  $d \rightarrow \infty$  limit. The reason is as follows.

Consider the left-hand side of Eq. (B2). Due to the definition of  $\rho_i$ , if  $d \rightarrow \infty$ , there exists  $n \geq 0$ , s.t.,  $\rho_{n+1} = 1$ , and we have

$$\begin{aligned} N_B \left( 2^{-d} + \sum_{i=1}^d \rho_i 2^{-i} \right) &= N_B \left( 2^{-d} + \sum_{i=1}^n \rho_i 2^{-i} + \sum_{i=n+1}^d 2^{-i} \right) \\ &= N_B \left( 2^{-d} + \sum_{i=1}^n \rho_i 2^{-i} + 2^{-n} - 2^{-d} \right) \\ &= N_B \left( 2^{-n} + \sum_{i=1}^n \rho_i 2^{-i} \right). \end{aligned} \quad (\text{B3})$$

This indicates that if  $d$  is large enough, we can seek the ‘‘fixed-point’’-like solution that

$$\rho_i^+ = \rho_i \text{ if } \rho_i < 1. \quad (\text{B4})$$

Plugging this solution into Eq. (B2) and taking Eq. (B3) into consideration, we obtain

$$e^{\lambda t_{ss}^+} = 2 \text{ or } t_{ss}^+ = \tau.$$

(B5)

This is exactly the steady-state periodic assumption in Appendix A. Thus, the steady-state solution with a period of  $\tau$  is mathematically equivalent to the “fixed-point” solution when  $d \rightarrow \infty$ . According to Eqs. (B1) and (B4), we obtain

$$\rho_i^{\text{ss}} = \begin{cases} i \frac{\tau}{C} & \text{if } i \leq n, \\ 1 & \text{otherwise,} \end{cases} \quad (\text{B6})$$

where

$$n \equiv \left\lfloor \frac{C}{\tau} \right\rfloor \quad (\text{B7})$$

is the number of overlapping replication cycles in the steady state. In the steady state, we can also calculate the number of initiators at initiation  $I(t_{\text{ini}})$ ,

$$\begin{aligned} I(t_{\text{ini}}) &= B(t_{\text{ini}}) \\ &= N_{\text{B}} \left( 2^d + \sum_{i=1}^d \rho_i^{\text{ss}} 2^{d-i} \right) + 2^d n_{\text{B}} \\ &= 2^d \left[ N_{\text{B}} \left( 2^{-d} + \sum_{i=1}^d \rho_i^{\text{ss}} 2^{-i} \right) + n_{\text{B}} \right] \\ &= 2^d \left[ N_{\text{B}} \left( 2^{-n} + \sum_{i=1}^n \rho_i^{\text{ss}} 2^{-i} \right) + n_{\text{B}} \right] \\ &= 2^d \left[ N_{\text{B}} \left( 2^{-n} + \frac{\tau}{C} \sum_{i=1}^n i 2^{-i} \right) + n_{\text{B}} \right] \\ &= 2^d \left[ N_{\text{B}} \left[ 2^{-n} + \left( 2 - \frac{n+2}{2^n} \right) \frac{\tau}{C} \right] + n_{\text{B}} \right]. \end{aligned} \quad (\text{B8})$$

The steady-state initiation mass  $v_i^{\text{ss}}$  is then given by

$$v_i^{\text{ss}} = \frac{I(t_{\text{ini}})}{2^d c_1},$$

which is the same as Eq. (A17).

## 2. The Jacobian matrix at the steady state

To analyze the stability of the steady state above, we need to compute the Jacobian matrix at the steady state,



$$J = \left. \frac{\partial \rho_i^+}{\partial \rho_j} \right|_{\text{ss}}. \quad (\text{B9})$$

Stability requires that the largest magnitude of the eigenvalues of the Jacobian matrix is smaller than 1.

Based on Eqs. (B1) and (B6), at the steady state,  $J$  is reduced to an  $n$ -dimensional matrix since other derivatives are zeros. Thus, we have

$$\left. \frac{\partial \rho_i^+}{\partial \rho_j} \right|_{\text{ss}} = \delta_{i-1,j} + \left. \frac{1}{C} \frac{\partial r^+}{\partial \rho_j} \right|_{\text{ss}}, \quad i, j = 1, 2, \dots, n, \quad (\text{B10})$$

where  $\delta_{i,j}$  is the Kronecker delta.

The partial derivatives of  $r^+$  can be computed by taking the partial derivatives at both sides of Eq. (B2) based on the implicit function theorem, and substituting Eq. (B1) and the steady-state values. The result reads

$$\left. \frac{1}{C} \frac{\partial r^+}{\partial \rho_n} \right|_{\text{ss}} = \left[ 2^n(1 - 2 \ln 2) + (n + 2) \ln 2 - 1 - \ln 2 \left( 1 + 2^n \frac{n_B}{N_B} \right) \frac{C}{\tau} \right]^{-1} \equiv a, \quad (\text{B11a})$$

$$\left. \frac{1}{C} \frac{\partial r^+}{\partial \rho_i} \right|_{\text{ss}} = 2^{n-i-1} a, \quad i = 1, 2, \dots, n-1. \quad (\text{B11b})$$

Hence, the Jacobian matrix has the following form:

$$J = \begin{pmatrix} 2^{n-2}a & 2^{n-3}a & \dots & 2a & a & a \\ 1 + 2^{n-2}a & 2^{n-3}a & \dots & 2a & a & a \\ 2^{n-2}a & 1 + 2^{n-3}a & \dots & 2a & a & a \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 2^{n-2}a & 2^{n-3}a & \dots & 1 + 2a & a & a \\ 2^{n-2}a & 2^{n-3}a & \dots & 2a & 1 + a & a \end{pmatrix}. \quad (\text{B12})$$

It may look difficult to write down a general characteristic equation for this matrix. In the following sections, we will discuss the situations when  $n \leq 3$ , which already covers almost all the experimental growth conditions in *E. coli*.

### 3. Stability regimes

#### a. $n = 0 (C < \tau)$ , slow and intermediate growth conditions

In this simplest case,  $\rho_i^+ = 1$  near the steady state, which means the map  $\mathcal{F}$  is a constant map, i.e.,  $J = 0$ . Thus, the steady state is always stable.

#### b. $n = 1 (\tau \leq C < 2\tau)$ , fast growth conditions

In this case,

$$J = a = \left[ 1 - \ln 2 - \ln 2 \left( 1 + 2 \frac{n_B}{N_B} \right) \frac{C}{\tau} \right]^{-1}.$$

The stability condition is  $|a| < 1$ , which gives out

$$\frac{n_B}{N_B} > \left( \frac{1}{\ln 2} - \frac{1}{2} \right) \frac{\tau}{C} - \frac{1}{2}.$$

(B13)

Note that at  $\tau \rightarrow C$ , there is a critical point for  $n_B$ ,

$$n_B = n_{B,1} = \left( \frac{1}{\ln 2} - 1 \right) N_B \approx 0.44 N_B.$$

If  $n_B > n_{B,1}$ , the steady state is always stable in this regime.

#### c. $n = 2 (2\tau \leq C < 3\tau)$ , very fast growth conditions

In this case,

$$J = \begin{pmatrix} a & a \\ 1 + a & a \end{pmatrix},$$

where

$$a = \left[ 3 - 4 \ln 2 - \ln 2 \left( 1 + 4 \frac{n_B}{N_B} \right) \frac{C}{\tau} \right]^{-1}.$$

Note that since  $n_B/N_B \geq 0$  and  $2 \leq C/\tau < 3$ , we have  $a \in (-1, 0)$ . The characteristic equation for  $J$  reads

$$\lambda^2 - 2a\lambda - a = 0,$$

which has two imaginary roots

$$\lambda_{\pm} = a \pm i\sqrt{-a^2 - a}.$$

Stability requires

$$|\lambda_{\pm}| = -a < 1.$$

Thus, we obtain the condition

$$\frac{n_B}{N_B} > \left(\frac{1}{\ln 2} - 1\right) \frac{\tau}{C} - \frac{1}{4}. \quad (\text{B14})$$

However, since  $\frac{\tau}{C} \in \left(\frac{1}{3}, \frac{1}{2}\right]$ , the right-hand side of the inequality is always negative, which means the steady state is always stable for any positive  $n_B$  in this regime.

**d.  $n = 3$  ( $3\tau \leq C < 4\tau$ ), extremely fast growth conditions**

In this case,

$$J = \begin{pmatrix} 2a & a & a \\ 1 + 2a & a & a \\ 2a & 1 + a & a \end{pmatrix},$$

where

$$a = \left[7 - 11 \ln 2 - \ln 2 \left(1 + 8 \frac{n_B}{N_B} \frac{C}{\tau}\right)\right]^{-1}.$$

According to the range of  $C$  and  $n_B$ , we have  $a \in (-3, 0)$ . The characteristic equation for  $J$  reads

$$\lambda^3 - 4a\lambda^2 - 2a\lambda - a = 0.$$

By computing the discriminant of this cubic equation, we know that it has one real root  $\lambda = \lambda_1$  and two imaginary roots  $\lambda = \xi \pm i\theta$ . The critical situation has three possibilities:  $\lambda_1 = 1$ ,  $\lambda_1 = -1$ , and  $\xi^2 + \theta^2 = 1$ . We find that only  $\lambda_1 = -1$  satisfies the range of  $a$ . Thus, the stability boundary requires  $a = -1/3$ . Further perturbation analysis gives out the inequality of stability condition,

$$a > -\frac{1}{3},$$

leading to

$$\frac{n_B}{N_B} > \left(\frac{5}{4 \ln 2} - \frac{11}{8}\right) \frac{\tau}{C} - \frac{1}{8}. \quad (\text{B15})$$

Note that at  $\tau \rightarrow C/3$ , there is another critical point for  $n_B$ ,

$$n_B = n_{B,2} = \left(\frac{5}{\ln 2} - 7\right) \frac{N_B}{12} \approx 0.018 N_B.$$

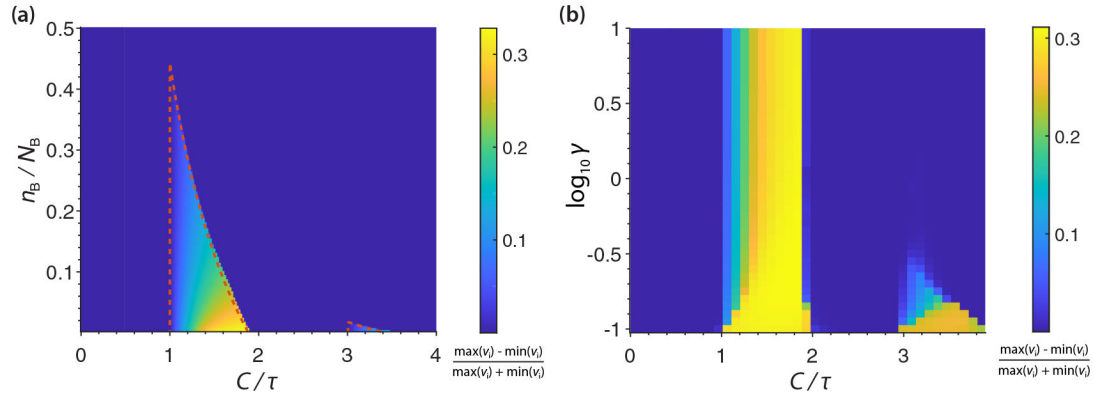
If  $n_B > n_{B,2}$ , the steady state is always stable in this regime [see Fig. 8(a)].

## APPENDIX C: A SIMPLE DERIVATION FOR THE FIXED POINTS OF MAPPING $\mathcal{F}^{02}$ IN $n = 1$

In principle, we can analyze the fixed points of  $\mathcal{F}^{02} \equiv \mathcal{F} \circ \mathcal{F}$  based on the expression of  $\mathcal{F}$  [Eqs. (B1) and (B2)]. However, since we know from simulation [Fig. 4(c)] that one of the fixed points of  $\mathcal{F}^{02}$  corresponds to

$$v_i^{(1)} = \frac{1}{c_1}(\alpha^{(1)} N_B + n_B), \quad \alpha^{(1)} = 1, \quad (\text{C1})$$

we can derive the second fixed point  $\alpha^{(2)}$  based on it.



**FIG. 8.**

Simulation of the instability on phase diagrams. Instability is quantified by  $(\max v_i - \min v_i) / (\max v_i + \min v_i)$  after 100 doubling times. (a) In the protocell, the unstable regimes match the theoretical predicted unstable regimes (boundaries shown by red dashed curves). (b) In the initiator-titration model v2 with a constant  $\gamma$ , we fixed  $n_B / N_B = 1/30$ , and we simulate the unstable regimes on the phase diagram of  $\gamma - \frac{C}{\tau}$ .

Consider the initiation event with  $\alpha^{(1)} = 1$  at  $t = 0$ . The initiator number at initiation reads

$$I(0) = 2(N_B + n_B). \quad (\text{C2})$$

After initiation,  $I(t)$  increases exponentially, whereas  $B(t)$  first jumps by  $2n_B$  and then increases linearly with a slope of  $2N_B/C$  for a duration time of  $C$ , as discussed in Sec.

II B in the main text. Suppose the next initiation time is at  $t = t^+$ . According to the initiation criterion  $I(t^+) = B(t^+)$ , we have

$$2(N_B + n_B)2\frac{t^+}{\tau} = 2(N_B + n_B) + 2n_B + \frac{2N_B}{C}t^+, \quad (C3)$$

which provides a transcendental equation to determine  $t^+$ ,

$$2\frac{t^+}{\tau} = 2 + \frac{N_B}{N_B + n_B} \left( \frac{t^+}{C} - 1 \right). \quad (C4)$$

By drawing the curves for the left-hand side and the right-hand side, it is easy to see that  $t^+ < \tau$ , and  $t^+$  exists only when  $C > \tau$ , which is consistent with the instability condition. Thus, based on Eq. (C3), the initiation mass at the second initiation event reads

$$\begin{aligned} v_i^{(2)} &= \frac{I(t^+)}{4c_I} = \left[ \frac{N_B}{2} \left( 1 + \frac{t^+}{C} \right) + n_B \right] / c_I \\ &= \frac{1}{c_I} (\alpha^{(2)} N_B + n_B). \end{aligned} \quad (C5)$$

Hence, the second fixed point reads

$$\alpha^{(2)} = \frac{1}{2} \left( 1 + \frac{t^+}{C} \right), \quad (C6)$$

where  $t^+$  is given by Eq. (C4). Because of  $t^+ < \tau$ , this value is smaller than the steady-state  $\alpha = (1 + \tau/C)/2$ , given by Eq. (3) with  $n = 1$ . This is consistent with our simulation [Fig. 4(c)]. Thus, the eventual picture is that in the unstable regime, the initiation mass oscillates around the steady-state value Eq. (3) in two values  $v_i^{(1)}$  and  $v_i^{(2)}$  given by Eqs. (C1) and (C5).

## APPENDIX D: DERIVATION OF THE STEADY-STATE INITIATION MASS FORMULA IN THE INITIATOR-TITRATION MODEL V2 WITH A STATIC DnaA-ATP/DnaA-ADP RATIO

In our initiator-titration model v2, we first considered a constant DnaA-ATP/DnaA-ADP ratio, named  $\gamma$  hereafter, which is likely the case of the  $\Delta 4$  mutant [13]. In  $\Delta 4$  mutant, there is only DnaA *de novo* synthesis and ATP hydrolysis by intrinsic ATPase activity of

DnaA [23]. We denote the intrinsic DnaA-ATP hydrolysis rate as  $v$ . We further assume that this hydrolysis rate is the same for free DnaA-ATP or bound DnaA-ATP on chromosomal binding sites. Additionally, since the free ATP/ADP ratio is high in the cytoplasm, we assume that newly expressed DnaA will immediately form DnaA-ATP. Denoting the total number of DnaA-ATP and DnaA-ADP as  $I^T$  and  $I^D$ , respectively, we have

$$\frac{dI^T}{dt} = \lambda(I^T + I^D) - vI^T, \quad (\text{D1a})$$

$$\frac{dI^D}{dt} = vI^T. \quad (\text{D1b})$$

For DnaA-ATP and DnaA-ADP concentrations,  $[I^T] = I^T/V$  and  $[I^D] = I^D/V$ , we have

$$\frac{d[I^T]}{dt} = \lambda[I^D] - v[I^T], \quad (\text{D2a})$$

$$\frac{d[I^D]}{dt} = -\lambda[I^D] + v[I^T], \quad (\text{D2b})$$

and the steady-state ratio is given by

$$\frac{[I^T]}{[I^D]} \equiv \gamma = \frac{\lambda}{v}. \quad (\text{D3})$$

The timescale for the intrinsic DnaA-ATP hydrolysis is about 15 min in wild-type *E. coli* [13]. If the doubling time ranges from 15 min to 2.5 h, then  $\gamma$  ranges from 0.1 to 1.

Now we consider that there are bound DnaA on the chromosome and free DnaA in the cytoplasm. We assume that the two forms of DnaA have the same strong binding affinity with chromosomal binding sites, but only DnaA-ATP can bind to DnaA boxes at *ori* binding sites with a relatively weak binding affinity. We denote the number of DnaA-ATP and DnaA-ADP bound to the chromosomal binding sites as  $I_b^T$  and  $I_b^D$ , respectively, the number of free DnaA-ATP and DnaA-ADP in the cytoplasm as  $I_f^T$  and  $I_f^D$ , respectively, and the number of DnaA-ATP at *ori* as  $I_o^T$ . Based on Eq. (D3), we have

$$\frac{I_b^T + I_f^T + I_o^T}{I_b^D + I_f^D} = \gamma. \quad (\text{D4})$$

The DnaA binding and unbinding processes are fast compared to the doubling time, so we can assume binding reactions are in rapid equilibrium, which gives

$$I_f^T(B_{\text{tot}} - I_b^T - I_b^D) = I_b^T K_b V, \quad (\text{D5})$$

$$I_f^D(B_{\text{tot}} - I_b^T - I_b^D) = I_b^D K_b V, \quad (\text{D6})$$

$$I_f^T(O_{\text{tot}} - I_o^T) = I_o^T K_o V, \quad (\text{D7})$$

where  $K_b$  and  $K_o$  are the dissociation constant for chromosomal binding sites and for *ori* binding sites, respectively;  $B_{\text{tot}}$  is the total number of chromosomal binding sites, and  $O_{\text{tot}}$  is the total *ori* binding sites that must be larger than  $n_b$  at each *ori*. Moreover, we have the equation of balanced biosynthesis for total DnaA,

$$c_1 V = I_b^T + I_f^T + I_o^T + I_b^D + I_f^D. \quad (\text{D8})$$

To determine the initiation mass  $\nu_i$ , suppose there are  $2^d$  *ori*'s at initiation, then  $V = 2^d \nu_i$ ,  $I_o^T = 2^d n_b$ , and  $O_{\text{tot}} - I_o^T = 2^d \Delta_b$ , where  $\Delta_b$  is the number of remaining binding sites at each *ori* at initiation. As discussed in Appendixes A and B,  $B_{\text{tot}}$  at initiation is fully determined by the replication fork progress that is a function of growth rates:  $B_{\text{tot}} = 2^d \alpha N_B$ , where  $\alpha$  was defined in Eq. (3) in the main text.

In principle, by substituting these parameters into Eqs. (D4)–(D8), we can solve the unknown variables  $I_f^T$ ,  $I_f^D$ ,  $I_b^T$ ,  $I_b^D$ , and  $\nu_i$ . However, it is hard to obtain a simple expression by solving Eqs. (D4)–(D8) directly because we need to solve a polynomial equation. To simplify the solution, notice that  $B_{\text{tot}}$  should be almost saturated at initiation because of strong binding affinity, so approximately

$$I_b^T + I_b^D \approx B_{\text{tot}}. \quad (\text{D9})$$

On the other hand, since  $0.1 \leq \gamma \leq 1$ ,  $I_b^T - \gamma(I_b^T + I_o^T)B_{\text{tot}} > I_o^T$ , so we can fairly assume  $I_b^T + I_f^T \gg I_o^T$ . By Eqs. (D5) and (D6), Eq. (D4) becomes

$$\frac{I_b^T}{I_b^D} \approx \frac{I_f^T + I_o^T}{I_f^D} = \gamma. \quad (\text{D10})$$

These approximations make it easy to solve the original Eqs. (D4)–(D8), and finally we obtain

$$v_i = \frac{\alpha N_B + \left(1 + \frac{1}{\gamma}\right)n_B}{c_i - \left(1 + \frac{1}{\gamma}\right)K_{\text{eff}}n_B}, \quad (\text{D11})$$

where  $K_{\text{eff}} = K_o/A_B$ . This is exactly Eq. (9) in the main text.

The dissociation constant  $K_o$  at *ori* should be much larger than  $K_b$  most of the time. In our simulation, we set  $c_i = 400 \mu\text{m}^{-3}$ ,  $n_B = 1$ ,  $K_b = 1 \mu\text{m}^{-3}$ , and  $K_o > 10 \mu\text{m}^{-3}$  [26]. These values result in a very large  $v_i$ . This is because we do not assume cooperativity of DnaA binding to *ori*, so that the binding process at *ori* will be slowed down significantly when the occupancy is close to the threshold  $n_B$ . To resolve this effect, we consider that  $K_o$  can decrease with  $I_o^T$  because of the stabilized filamentous structure of DnaA proteins at *ori*, which has been found in *E. coli* [16]. In this case, the rapid-equilibrium assumption may not hold, but we assume that it is not too far away. Hence, we treat  $K_{\text{eff}}$  as a fitting parameter. In our simulation, we assumed a linear decrease of  $K_o$  from 50 to  $1 \mu\text{m}^{-3}$  when the occupancy rises from 0 to the threshold. We fitted the numerical results with the initiation expression Eq. (D11), and we found that  $K_{\text{eff}} = 2 \mu\text{m}^{-3}$  is almost a perfect fit for the stable regimes [Fig. 5(c) in the main text].

## APPENDIX E: THE NUMERICAL MODEL SETUP AND SUPPLEMENTARY RESULTS

In this Appendix, we include the effects of RIDA, DDAH, DARS1, and DARS2. First, it is reported that RIDA accelerates the conversion of DnaA-ATP to DnaA-ADP by the Hda protein that binds to the DNA-loaded  $\beta$ -clamp at the replisome [20,21,43]. We assume that Hda proteins work at the saturation level, so the RIDA activity is proportional to the number of replication forks,  $N_{\text{fork}}$ . Second, DDAH promotes the conversion of DnaA-ATP to DnaA-ADP by the locus *datA* on the chromosome [39,40]. Thus, we assume DDAH activity to be proportional to the copy number of *datA* loci,  $N_{\text{datA}}$ . Third, DARS1 and DARS2 are the two loci on the chromosome that promote the reactivation of DnaA-ADP [41,42], hence we assume that their activities are proportional to the copy number of DARS1 and DARS2 loci,  $N_{\text{DARS1}}$  and  $N_{\text{DARS2}}$ , respectively. We further assume that these reactions follow the Michaelis-Menten form. Based on these assumptions, we have



$$\frac{dI^T}{dt} = \lambda(I^T + I^D) - \nu I^T - (k_{\text{RIDA}}N_{\text{fork}} + k_{\text{DDAH}}N_{\text{datA}}) \times \frac{[I^T]}{[I^T] + K_T} + (k_{\text{DARS1}}N_{\text{DARS1}} + k_{\text{DARS2}}N_{\text{DARS2}}) \times \frac{[I^D]}{[I^D] + K_D}, \quad (\text{E1a})$$

$$\frac{dI^D}{dt} = \nu I^T + (k_{\text{RIDA}}N_{\text{fork}} + k_{\text{DDAH}}N_{\text{datA}}) \frac{[I^T]}{[I^T] + K_T} - (k_{\text{DARS1}}N_{\text{DARS1}} + k_{\text{DARS2}}N_{\text{DARS2}}) \frac{[I^D]}{[I^D] + K_D}, \quad (\text{E1b})$$

where  $k_{\text{RIDA}}$ ,  $k_{\text{DDAH}}$ ,  $k_{\text{DARS1}}$ , and  $k_{\text{DARS2}}$  are the reaction rate constants corresponding to RIDA, DDAH, DARS1, and DARS2, respectively;  $K_T$  and  $K_D$  are the Michaelis-Menten constants for DnaA-ATP to DnaA-ADP conversion and the other way around, respectively. We have assumed that the Michaelis-Menten constant is the same for RIDA and DDAH, and for DARS1 and DARS2, for simplicity. When we study the absence of some of the mechanisms, we set the corresponding reaction rate constants to be zero.

We are particularly interested in the dynamics of  $\gamma = I^T/I^D$ . Based on Eqs. (E1), we can obtain the ODE for  $\gamma$ :

$$\frac{d\gamma}{dt} = \left( 1 + \gamma \right) \left[ \lambda - \nu\gamma - \frac{k_{\text{RIDA}}N_{\text{fork}} + k_{\text{DDAH}}N_{\text{datA}}}{c_1V} \times \frac{\gamma}{\frac{\gamma}{1+\gamma} + K_T^*} + \frac{k_{\text{DARS1}}N_{\text{DARS1}} + k_{\text{DARS2}}N_{\text{DARS2}}}{c_1V} \times \frac{1}{\frac{1}{1+\gamma} + K_D^*} \right], \quad (\text{E2})$$

where  $K_T^* = K_T/c_1$  and  $K_D^* = K_D/c_1$ .

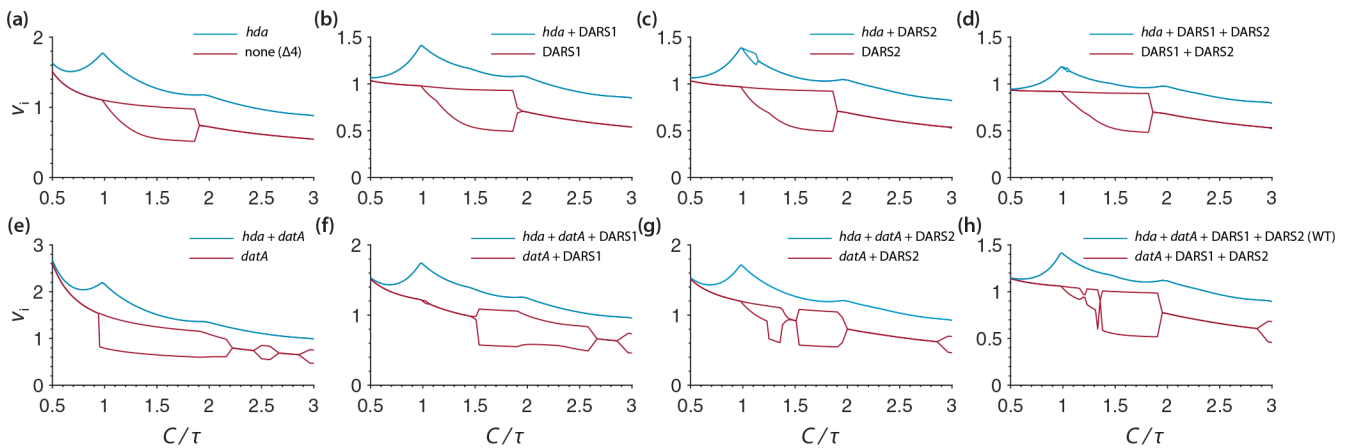


FIG. 9.

The initiation mass behavior in different combinations of the DnaA-ATP/DnaA-ADP conversion elements, i.e., *hda* (RIDA), *datA* (DDAH), DARS1, and DARS2. The situations with *hda* and without *hda* are plotted together in each plot to show the contrast.

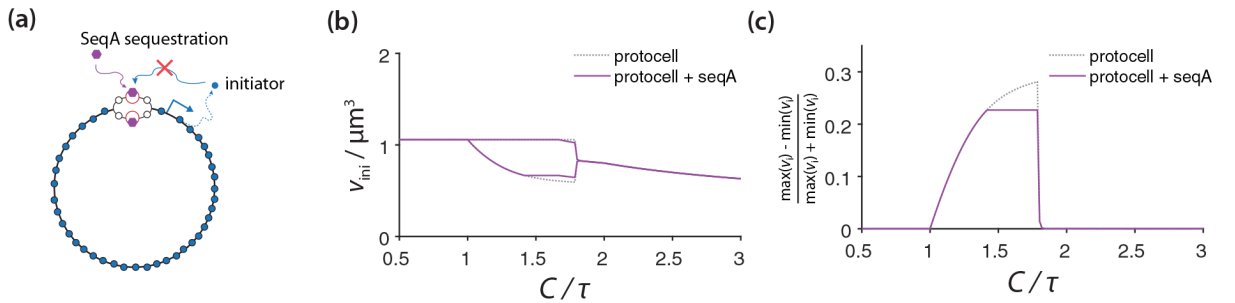
Now we consider the bound and unbound forms of DnaA: DnaA bound to chromosomal binding sites  $I_b = I_b^T + I_b^D$ , free DnaA in cytoplasm  $I_f = I_f^T + I_f^D$ , and DnaA bound to *ori* binding sites  $I_o$ . Equation (D10) is still assumed here. We denote the binding and unbinding rate constant on the chromosomal binding sites as  $k_{on}^b$  and  $k_{off}^b$ , and those on the *ori* binding sites as  $k_{on}^o$  and  $k_{off}^o$ . Thus, the ODEs for  $I_b$ ,  $I_f$ , and  $I_o$  read

$$\frac{dI_b}{dt} = k_{on}^b \frac{I_f}{V} (B_{tot} - I_b) - k_{off}^b I_b, \quad (E3a)$$

$$\frac{dI_f}{dt} = \lambda(I_f + I_b + I_o) - k_{on}^b \frac{I_f}{V} (B_{tot} - I_b) + k_{off}^b I_b - \frac{\gamma}{1+\gamma} k_{on}^o \frac{I_f}{V} (O_{tot} - I_o) + k_{off}^o I_o, \quad (E3b)$$

$$\frac{dI_o}{dt} = \frac{\gamma}{1+\gamma} k_{on}^o \frac{I_f}{V} (O_{tot} - I_o) - k_{off}^o I_o, \quad (E3c)$$

where  $B_{tot}$  follows the same dynamics determined by  $\rho(t)$  as derived in Appendix B. In this way, we conducted a deterministic simulation based on Eqs. (E2) and (E3). The results with all 16 combinations of the four mechanisms are shown in Fig. 9. Additionally, we simulated the effect of SeqA in our deterministic model [Fig. 10(a)]. We found that SeqA does not significantly change the results in any of the above 16 cases. As an example for its mild effect, Figs. 10(b) and 10(c) show how SeqA affects the initiation stability in the protocell.



**FIG. 10.** The effect of SeqA in the protocell model. (a) We consider the SeqA sequestration effect at *ori* binding sites, which results in no initiation events in a duration time of  $t_{seq} = 10$  min. (b) The unstable range does not shrink with the existence of seqA. (c) The amplitude of the initiation mass oscillation, defined by  $(\max v_i - \min v_i)/(\max v_i + \min v_i)$ , is slightly reduced when  $C/\tau$  is relatively large in the unstable regime. Here, we set  $n_B/N_B = 1/30$ .

## APPENDIX F: RELATION BETWEEN THE INTRINSIC/EXTRINSIC NOISE AND INITIATION ASYNCHRONY/CELL-TO-CELL VARIABILITY

We defined the intrinsic noise and the extrinsic noise for the initiation mass based on one of their conventional definitions [50] in Sec. IIE. However, we will show that they are not equivalent to the intuitive concepts of asynchrony and cell-to-cell variability, though these two groups of concepts are closely related [see Fig. 6(a)].

For asynchrony versus cell-to-cell variability, by intuition, we can quantify them as the coefficients of variation along the diagonal axis and off-diagonal axis on the *ori1-ori2* initiation mass plot [Fig. 6(a) in the main text], respectively. That is, if we transform the initiation mass variables  $v_i^{(1)}$  and  $v_i^{(2)}$  by a rotation of the coordinates,  $u_i^{(1)} = (v_i^{(1)} + v_i^{(2)})/\sqrt{2}$ ,  $u_i^{(2)} = (v_i^{(1)} - v_i^{(2)})/\sqrt{2}$ , initiation cell-to-cell variability ( $CV_{\text{ccv}}$ ) and initiation asynchrony ( $CV_{\text{asyn}}$ ) can be properly defined as

$$CV_{\text{ccv}}^2 = \frac{\langle u_i^{(1)2} \rangle - \langle u_i^{(1)} \rangle^2}{\langle u_i^{(1)} \rangle^2}, \quad CV_{\text{asyn}}^2 = \frac{\langle u_i^{(2)2} \rangle - \langle u_i^{(2)} \rangle^2}{\langle u_i^{(1)} \rangle^2}. \quad (\text{F1})$$

We want to use the statistics of  $v_i^{(1)}$  and  $v_i^{(2)}$  to represent these two new CVs. We assume that  $\langle v_i^{(1)} \rangle = \langle v_i^{(2)} \rangle$ , and  $\langle v_i^{(1)2} \rangle = \langle v_i^{(2)2} \rangle$ , because two *ori*'s are identical such that  $v_i^{(1)}$  and  $v_i^{(2)}$  should follow the same distribution. Therefore, their coefficients of variation should be the same and equal to the total CV:

$$CV^{(1)2} = CV^{(2)2} = CV_{\text{tot}}^2 = \frac{\langle v_i^2 \rangle - \langle v_i \rangle^2}{\langle v_i \rangle^2}. \quad (\text{F2})$$

Substituting the definition of  $u_i^{(1)}$  and  $u_i^{(2)}$  into Eq. (F1), we have

$$CV_{\text{ccv}}^2 = \frac{1+r}{2} CV_{\text{tot}}^2, \quad CV_{\text{asyn}}^2 = \frac{1-r}{2} CV_{\text{tot}}^2, \quad (\text{F3})$$

where  $r$  is the Pearson correlation coefficient between  $v_i^{(1)}$  and  $v_i^{(2)}$ ,

$$r = \frac{\langle v_i^{(1)} v_i^{(2)} \rangle - \langle v_i^{(1)} \rangle \langle v_i^{(2)} \rangle}{\sqrt{\langle v_i^{(1)2} \rangle - \langle v_i^{(1)} \rangle^2} \cdot \sqrt{\langle v_i^{(2)2} \rangle - \langle v_i^{(2)} \rangle^2}}. \quad (\text{F4})$$

On the other hand, based on the conventional definition [50], the intrinsic noise and the extrinsic noise for the initiation mass are defined in Eq. (12) in Sec. II E 1. Following the above derivations, we have

$$CV_{\text{int}}^2 = (1 - r)CV_{\text{tot}}^2, \quad CV_{\text{ext}}^2 = rCV_{\text{tot}}^2. \quad (\text{F5})$$

However, this definition is only valid when  $0 \leq r \leq 1$  given that  $CV_{\text{ext}}$  has to be a real number, which means it only applies when two variables are non-negatively correlated. Based on Eqs. (F3) and (F5), we obtain the relation between intrinsic/extrinsic noise and asynchrony/cell-to-cell variability,

$$CV_{\text{asyn}}^2 = \frac{1}{2}CV_{\text{int}}^2, \quad CV_{\text{ccv}}^2 = CV_{\text{ext}}^2 + \frac{1}{2}CV_{\text{int}}^2. \quad (\text{F6})$$

We have the summation relation of the two groups of concepts,

$$CV_{\text{ccv}}^2 + CV_{\text{asyn}}^2 = CV_{\text{ext}}^2 + CV_{\text{int}}^2 = CV_{\text{tot}}^2. \quad (\text{F7})$$

Hence, the intrinsic/extrinsic noise and asynchrony/cell-to-cell variability are just two decompositions of the total noise. The first works only at the regime  $0 \leq r \leq 1$ , while the second works for any correlation situation, i.e.,  $-1 \leq r \leq 1$ .

## APPENDIX G: DERIVATION OF THE CVS' SCALING LAWS IN THE FIRST-PASSAGE-TIME MODELS

Our first model in Sec. IIE 2 assumes that (i) the initiator proteins are produced in a Poisson process with a constant production rate  $\beta$ , and (ii) after production, the initiator protein has an equal chance to bind to either *ori1* or *ori2*, namely a Bernoulli trial. Thus, the probability of *ori1* to accumulate one protein during the time interval  $(t, t + dt]$  reads

$$\mathbb{P}[O_i(t + dt) = m + 1 \mid O_i(t) = m] = \beta dt \cdot \frac{1}{2}, \quad (\text{G1})$$

which is equivalently a Poisson process with a production rate of  $\beta' = \frac{\beta}{2}$ .

For a Poisson process with a production rate of  $\beta'$ , the waiting time for each jump event follows an exponential distribution with a decay rate  $\beta'$ , and the first-passage-time (FPT)  $T^{(1)}$  is the sum of  $n_B$  identical waiting-time variables. This gives out a Gamma distribution of  $T^{(1)}$ ,

$$P(T; n_B, \beta') = \frac{\beta'^{n_B} T^{n_B-1} e^{-\beta' T}}{(n_B - 1)!}.$$

(G2)

Thus,  $\langle T^{(1)} \rangle = n_B / \beta' = 2n_B / \beta$  and  $\sigma_{T^{(1)}} = \sqrt{n_B} / \beta' = 2\sqrt{n_B} / \beta$ . Based on Eq. (14), we obtain

$$CV_{\text{int}} = \frac{1}{\sqrt{n_B}}.$$

(G3)

On the other hand, the total number of proteins produced before time  $T$  follows the original Poisson distribution:

$$P(N; T) = \frac{(\beta T)^N}{N!} e^{-\beta T},$$

(G4)

which gives a mean value  $\langle N \rangle = \beta T$ . Thus, the mean total protein at the FPT ( $\langle T^{(1)} \rangle = 2n_B / \beta$ ) is  $\langle N \rangle = 2n_B$ . Therefore,

$$CV_{\text{int}} = \sqrt{\frac{2}{\langle N \rangle}},$$

(G5)

which is essentially Eq. (15).

Next, for the two-step Poisson process, we have decomposed  $T^{(1)}$  and  $T^{(2)}$  into three independent variables,  $T^{(1)} = T^{(0)} + \Delta T^{(1)}$ ,  $T^{(2)} = T^{(0)} + \Delta T^{(2)}$  in the main text. To calculate Eq. (16), we need to obtain  $\langle T^{(1)} \rangle$ ,  $\sigma_{\Delta T^{(1)}}$ , and  $\sigma_{T^{(0)}}$ . To avoid the instability issue, we consider the stable cases of two-overlapping cell cycles (non-multifork replication). In this situation, The threshold for the titration step is roughly  $2N_B$ , and for the *oriI* accumulation step it is  $n_B$ . The first titration step has an accumulation rate of  $\beta$ , while the second step has an accumulation rate of  $\beta' = \beta/2$ , similar to the simple Poisson process case above. Thus,

$$\langle T^{(0)} \rangle = \frac{2N_B}{\beta}, \quad \langle \Delta T^{(1)} \rangle = \frac{n_B}{\beta'} = \frac{2n_B}{\beta};$$

(G6)

$$\sigma_{T^{(0)}} = \frac{\sqrt{2N_B}}{\beta}, \quad \sigma_{\Delta T^{(1)}} = \frac{\sqrt{n_B}}{\beta'} = \frac{2\sqrt{n_B}}{\beta}.$$

(G7)

Hence,  $\langle N \rangle = \beta \langle T^{(1)} \rangle = 2(N_B + n_B) / \beta$ . We rewrite Eq. (G7) in terms of  $\langle N \rangle$  and obtain

$$CV_{\text{int}} = \frac{2\sqrt{n_B}}{\langle N \rangle}, \quad CV_{\text{ext}} = \frac{\sqrt{\langle N \rangle - 2n_B}}{\langle N \rangle} \approx \frac{1}{\sqrt{\langle N \rangle}},$$

(G8)

which is the same as Eq. (17) in the main text.

To verify that the decomposition is valid, we performed Monte Carlo simulation. In the simulation, the initiator protein is generated based on the Poisson process with a constant rate  $\beta$ . Once a protein is generated, it will undergo a trial among titration sites, *ori1*, and *ori2*. The probability of the protein to bind to different destinations is proportional to their binding affinities, and we set the binding affinity difference between titration sites and *ori*'s to be 100-fold, based on the *E. coli* case we mentioned in the main text. Further, we assumed that when the occupancy of one *ori* reaches the threshold, it will keep accumulation rather than waiting for the other *ori*. This is to assure that the stochastic process is not changed after one *ori* reaches the threshold. We notice that although these detailed settings are necessary to get numbers very close to what are predicted in Eq. (17) [Fig. 6(d), dashed gray lines], the  $1/N$  versus  $1/\sqrt{N}$  scaling laws are quite conserved in settings with different details.

## References

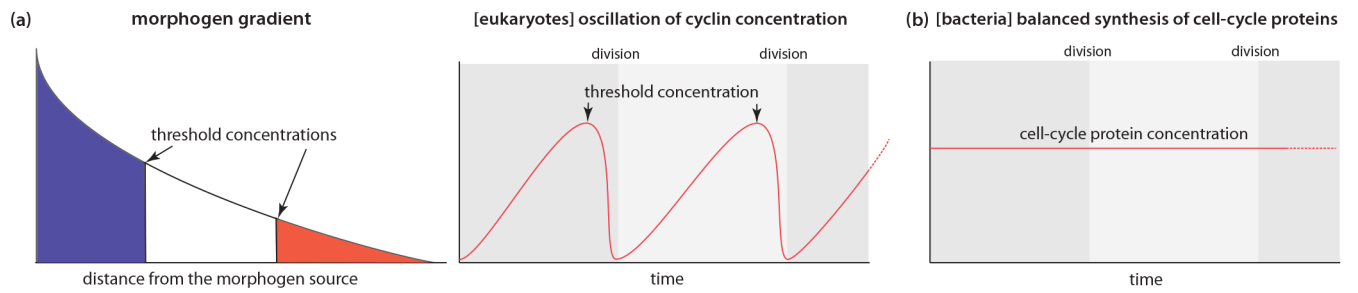
- [1]. Wolpert L, Positional information and the spatial pattern of cellular differentiation, *J. Theor. Biol* 25, 1 (1969). [PubMed: 4390734]
- [2]. Evans T, Rosenthal ET, Youngblom J, Distel D, and Hunt T, Cyclin: a protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division, *Cell* 33, 389 (1983). [PubMed: 6134587]
- [3]. Jun S, Si F, Pugatch R, and Scott M, Fundamental principles in bacterial physiology—history, recent progress, and the future with focus on cell size control: A review, *Rep. Prog. Phys* 81, 056601 (2018). [PubMed: 29313526]
- [4]. Schaechter M, Maaløe O, and Kjeldgaard NO, Dependency on medium and temperature of cell size and chemical composition during balanced growth of salmonella typhimurium, *Microbiology* 19, 592 (1958).
- [5]. Kjeldgaard NO, Maaloe O, and Schaechter M, The transition between different physiological states during balanced growth of salmonella typhimurium, *J. Gen. Microbiol* 19, 607 (1958). [PubMed: 13611203]
- [6]. Si F, Le Treut G, Sauls JT, Vadia S, Levin PA, and Jun S, Mechanistic origin of cell-size control and homeostasis in bacteria, *Curr. Biol* 29, 1760 (2019). [PubMed: 31104932]
- [7]. den Blaauwen T, Hamoen LW, and Levin PA, The divisome at 25: The road ahead, *Curr. Opin. Microbiol* 36, 85 (2017). [PubMed: 28254403]
- [8]. Campos M, Surovtsev IV, Kato S, Paintdakhi A, Beltran B, Ebmeier SE, and Jacobs-Wagner C, A constant size extension drives bacterial cell size homeostasis, *Cell* 159, 1433 (2014). [PubMed: 25480302]
- [9]. Taheri-Araghi S, Bradde S, Sauls JT, Hill NS, Levin PA, Paulsson J, Vergassola M, and Jun S, Cell-size control and homeostasis in bacteria, *Curr. Biol* 25, 385 (2015). [PubMed: 25544609]
- [10]. Jun S and Taheri-Araghi S, Cell-size maintenance: Universal strategy revealed, *Trends Microbiol* 23, 4 (2015). [PubMed: 25497321]
- [11]. Berger M and Wolde PRT, Robust replication initiation from coupled homeostatic mechanisms, *Nat. Commun* 13, 6556 (2022). [PubMed: 36344507]
- [12]. Sompayrac L and Maaloe O, Autorepressor model for control of DNA replication, *Nat. New Biol* 241, 133 (1973). [PubMed: 4573387]

- [13]. Boesen T, Charbon G, Fu H, Jensen C, Li D, Jun S, and others, Robust control of replication initiation in the absence of DnaA-ATP DnaA-ADP regulatory elements in *Escherichia coli*, *bioRxiv* (2022).
- [14]. Leonard AC, Rao P, Kadam RP, and Grimwade JE, Changing perspectives on the role of DnaA-ATP in orisome function and timing regulation, *Front. Microbiol* 10, 2009 (2019). [PubMed: 31555240]
- [15]. Hansen FG and Atlung T, The DnaA tale, *Front. Microbiol* 9, 319 (2018). [PubMed: 29541066]
- [16]. Katayama T, Kasho K, and Kawakami H, The DnaA cycle in *Escherichia coli*: Activation, function and inactivation of the initiator protein, *Front. Microbiol* 8, 2496 (2017). [PubMed: 29312202]
- [17]. Riber L, Frimodt-Møller J, Charbon G, and Løbner-Olesen A, Multiple DNA binding proteins contribute to timing of chromosome replication in *E. coli*, *Front. Mol. Biosci* 3, 29 (2016). [PubMed: 27446932]
- [18]. Katayama T, Ozaki S, Keyamura K, and Fujimitsu K, Regulation of the replication cycle: Conserved and diverse regulatory systems for DnaA and *oric*, *Nat. Rev. Microbiol* 8, 163 (2010). [PubMed: 20157337]
- [19]. Hansen FG, Christensen BB, and Atlung T, The initiator titration model: Computer simulation of chromosome and minichromosome control, *Res. Microbiol* 142, 161 (1991). [PubMed: 1925015]
- [20]. Katayama T, Kubota T, Kurokawa K, Crooke E, and Sekimizu K, The initiator function of DnaA protein is negatively regulated by the sliding clamp of the *E. coli* chromosomal replicase, *Cell* 94, 61 (1998). [PubMed: 9674428]
- [21]. Kato J and Katayama T, Hda, a novel DnaA-related protein, regulates the replication cycle in *Escherichia coli*, *EMBO J* 20, 4253 (2001). [PubMed: 11483528]
- [22]. Leonard AC and Grimwade JE, The orisome: Structure and function, *Front. Microbiol* 6, 545 (2015). [PubMed: 26082765]
- [23]. Sekimizu K, Bramhill D, and Kornberg A, ATP activates *dnaa* protein in initiating replication of plasmids bearing the origin of the *E. coli* chromosome, *Cell* 50, 259 (1987). [PubMed: 3036372]
- [24]. Nishida S, Fujimitsu K, Sekimizu K, Ohmura T, Ueda T, and Katayama T, A nucleotide switch in the *Escherichia coli* DnaA protein initiates chromosomal replication: Evidence from a mutant DnaA protein defective in regulatory ATP hydrolysis *in vitro* and *in vivo*, *J. Biol. Chem* 277, 14986 (2002). [PubMed: 11839737]
- [25]. McGarry KC, Ryan VT, Grimwade JE, and Leonard AC, Two discriminatory binding sites in the *Escherichia coli* replication origin are required for DNA strand opening by initiator DnaA-ATP, *Proc. Natl. Acad. Sci. USA* 101, 2811 (2004). [PubMed: 14978287]
- [26]. Schaper S and Messer W, Interaction of the initiator protein DnaA of *Escherichia coli* with its DNA target, *J. Biol. Chem* 270, 17622 (1995). [PubMed: 7615570]
- [27]. Fuller RS, Funnell BE, and Kornberg A, The *dnaA* protein complex with the *E. coli* chromosomal replication origin (*oric*) and other DNA sites, *Cell* 38, 889 (1984). [PubMed: 6091903]
- [28]. Kawakami H, Keyamura K, and Katayama T, Formation of an ATP-DnaA-specific initiation complex requires DnaA arginine 285, a conserved motif in the AAA+ protein family, *J. Biol. Chem* 280, 27420 (2005). [PubMed: 15901724]
- [29]. Herrick J, Kohiyama M, Atlung T, and Hansen FG, The initiation mess? *Mol. Microbiol* 19, 659 (1996). [PubMed: 8820637]
- [30]. Buckstein MH, He J, and Rubin H, Characterization of nucleotide pools as a function of physiological state in *Escherichia coli*, *J. Bacteriol* 190, 718 (2008). [PubMed: 17965154]
- [31]. Hsieh LS, Rouviere-Yaniv J, and Drlica K, Bacterial DNA supercoiling and [ATP]/[ADP] ratio: Changes associated with salt shock, *J. Bacteriol* 173, 3914 (1991). [PubMed: 1646791]
- [32]. More precisely, there are 11 DnaA boxes clustered within *ori*, with three high-affinity sites and eight low-affinity sites. Only DnaA-ATP can bind to the eight low-affinity sites. For simplicity, we used “10” to denote the number of low-affinity binding sites at *ori* throughout this paper.
- [33]. Hansen FG, Atlung T, Braun RE, Wright A, Hughes P, and Kohiyama M, Initiator (DnaA) protein concentration as a function of growth rate in *Escherichia coli* and *salmonella typhimurium*, *J. Bacteriol* 173, 5194 (1991). [PubMed: 1860829]

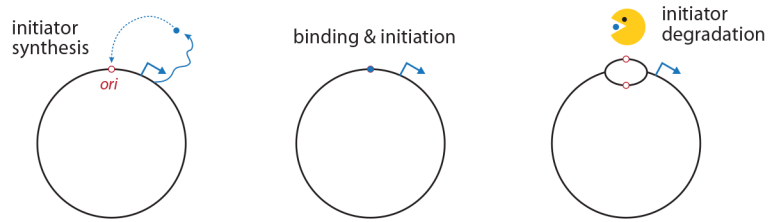
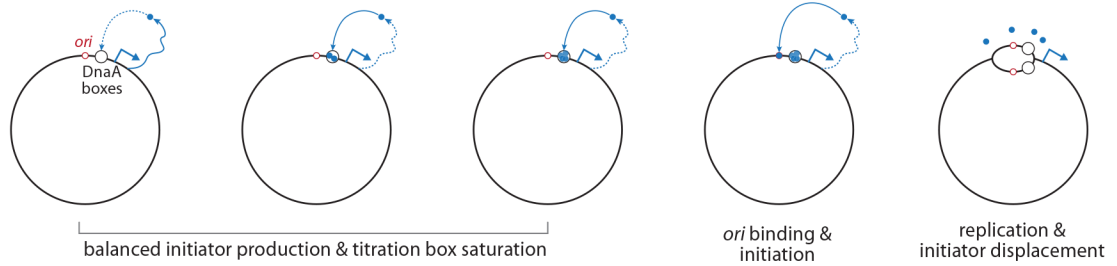
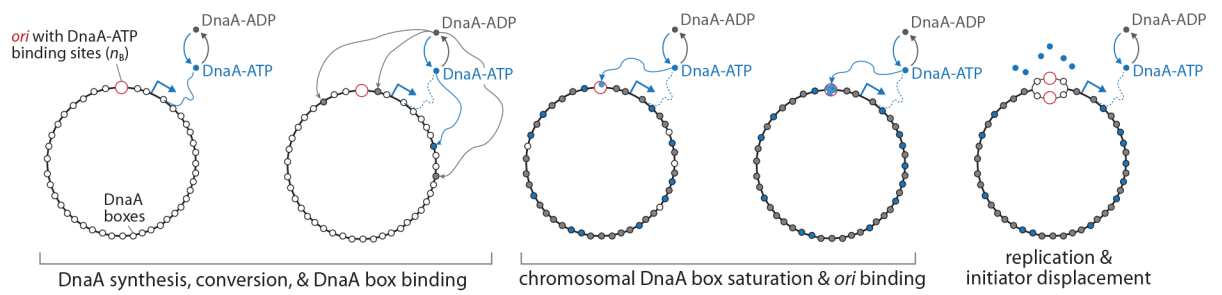
- [34]. Zheng H, Bai Y, Jiang M, Tokuyasu TA, Huang X, Zhong F, Wu Y, Fu X, Kleckner N, Hwa T, and Liu C, General quantitative relations linking cell growth and the cell cycle in *Escherichia coli*, *Nat Microbiol* 5, 995 (2020). [PubMed: 32424336]
- [35]. Pham TM, Tan KW, Sakumura Y, Okumura K, Maki H, and Akiyama MT, A single-molecule approach to DNA replication in *Escherichia coli* cells demonstrated that DNA polymerase III is a major determinant of fork speed, *Mol. Microbiol* 90, 584 (2013). [PubMed: 23998701]
- [36]. Bhat D, Hauf S, Plessy C, Yokobayashi Y, and Pigolotti S, Speed variations of bacterial replisomes, *Elife* 11, e75884 (2022). [PubMed: 35877175]
- [37]. Si F, Li D, Cox SE, Sauls JT, Azizi O, Sou C, Schwartz AB, Erickstad MJ, Jun Y, Li X, and Jun S, Invariance of initiation mass and predictability of cell size in *Escherichia coli*, *Curr. Biol* 27, 1278 (2017). [PubMed: 28416114]
- [38]. Cooper S and Helmstetter CE, Chromosome replication and the division cycle of *Escherichia coli* Br. *J. Mol. Biol* 31, 519 (1968). [PubMed: 4866337]
- [39]. Kitagawa R, Mitsuki H, Okazaki T, and Ogawa T, A novel DnaA protein-binding site at 94.7 min on the *Escherichia coli* chromosome, *Mol. Microbiol* 19, 1137 (1996). [PubMed: 8830270]
- [40]. Kasho K and Katayama T, DnaA binding locus *datA* promotes DnaA-ATP hydrolysis to enable cell cycle-coordinated replication initiation, *Proc. Natl. Acad. Sci. USA* 110, 936 (2013). [PubMed: 23277577]
- [41]. Fujimitsu K, Senriuchi T, and Katayama T, Specific genomic sequences of *E. coli* promote replicational initiation by directly reactivating ADP-DnaA, *Genes Dev* 23, 1221 (2009). [PubMed: 19401329]
- [42]. Kasho K, Fujimitsu K, Matoba T, Oshima T, and Katayama T, Timely binding of IHF and Fis to DARS2 regulates ATP-DnaA production and replication initiation, *Nucl. Acids Res* 42, 13134 (2014). [PubMed: 25378325]
- [43]. Moolman MC, Krishnan ST, Kerssemakers JWW, van den Berg A, Tulinski P, Depken M, Reyes-Lamothe R, Sherratt DJ, and Dekker NH, Slow unloading leads to DNA-bound  $\beta_2$ -sliding clamp accumulation in live *Escherichia coli* cells, *Nat. Commun* 5, 5820 (2014). [PubMed: 25520215]
- [44]. Lu M, Campbell JL, Boye E, and Kleckner N, SeqA: A negative modulator of replication initiation in *E. coli*, *Cell* 77, 413 (1994). [PubMed: 8011018]
- [45]. Hansen FG, Christensen BB, Nielsen CB, and Atlung T, Insights into the quality of DnaA boxes and their cooperativity, *J. Mol. Biol* 355, 85 (2006). [PubMed: 16298387]
- [46]. Rozgaja TA, Grimwade JE, Iqbal M, Czerwonka C, Vora M, and Leonard AC, Two oppositely oriented arrays of low-affinity recognition sites in *oric* guide progressive binding of DnaA during *Escherichia coli* pre-RC assembly, *Mol. Microbiol* 82, 475 (2011). [PubMed: 21895796]
- [47]. Donachie WD and Blakely GW, Coupling the initiation of chromosome replication to cell size in *Escherichia coli*, *Curr. Opin. Microbiol* 6, 146 (2003). [PubMed: 12732304]
- [48]. Kurokawa K, Nishida S, Emoto A, Sekimizu K, and Katayama T, Replication cycle-coordinated change of the adenine nucleotide-bound forms of DnaA protein in *Escherichia coli*, *EMBO J* 18, 6642 (1999). [PubMed: 10581238]
- [49]. Knöppel A, Broström O, Gras K, Elf J, and Fange D, Regulatory elements coordinating initiation of chromosome replication to the *Escherichia coli* cell cycle, *Proc. Natl. Acad. Sci. USA* 120, e2213795120 (2023). [PubMed: 37220276]
- [50]. Elowitz MB, Levine AJ, Siggia ED, and Swain PS, Stochastic gene expression in a single cell, *Science* 297, 1183 (2002). [PubMed: 12183631]
- [51]. Thattai M and van Oudenaarden A, Intrinsic noise in gene regulatory networks, *Proc. Natl. Acad. Sci. USA* 98, 8614 (2001). [PubMed: 11438714]
- [52]. Paulsson J, Summing up the noise in gene networks, *Nature (London)* 427, 415 (2004). [PubMed: 14749823]
- [53]. Taniguchi Y, Choi PJ, Li G-W, Chen H, Babu M, Hearn J, Emili A, and Sunney Xie X, Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells, *Science* 329, 533 (2010). [PubMed: 20671182]
- [54]. Skarstad K, Boye E, and Steen HB, Timing of initiation of chromosome replication in individual *Escherichia coli* cells, *EMBO J* 5, 1711 (1986). [PubMed: 3527695]



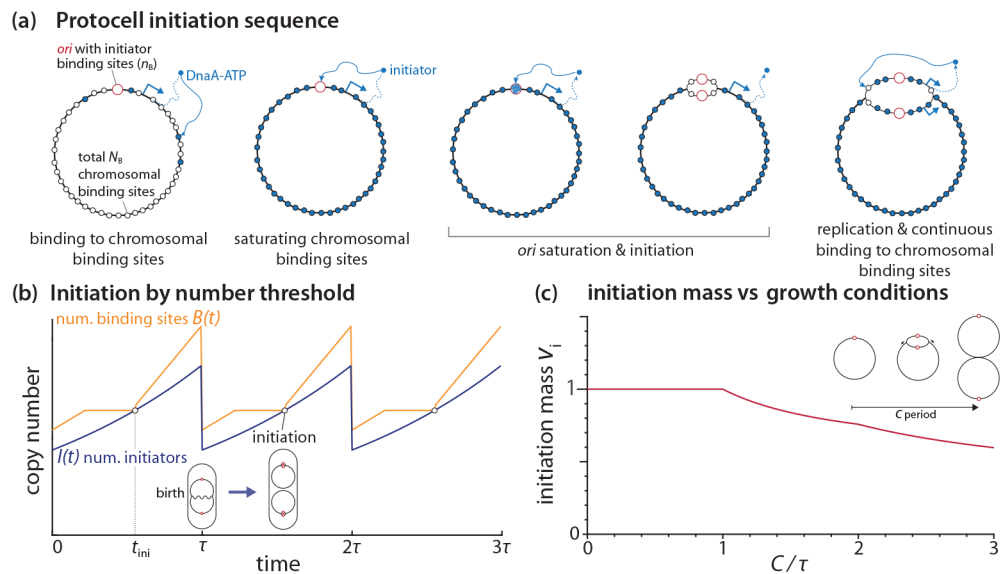
- [55]. Løbner-Olesen A, Skarstad K, Hansen FG, von Meyenburg K, and Boye E, The DnaA protein determines the initiation mass of *Escherichia coli* K-12, *Cell* 57, 881 (1989). [PubMed: 2541928]
- [56]. Sauls JT, Cox SE, Do Q, Castillo V, Ghulam-Jelani Z, and Jun S, Control of *Bacillus subtilis* replication initiation during physiological transitions and perturbations, *mBio* 10, e02205–19 (2019). [PubMed: 31848269]
- [57]. Pandey PP, Singh H, and Jain S, Exponential trajectories, cell size fluctuations, and the adder property in bacteria follow from simple chemical dynamics and division control, *Phys. Rev. E* 101, 062406 (2020). [PubMed: 32688579]
- [58]. Ghusinga KR, Dennehy JJ, and Singh A, First-passage time approach to controlling noise in the timing of intracellular events, *Proc. Natl. Acad. Sci. USA* 114, 693 (2017). [PubMed: 28069947]
- [59]. Wallden M, Fange D, Lundius EG, Baltekin Ö, and Elf J, The synchronization of replication and division cycles in individual *E. coli* cells, *Cell* 166, 729 (2016). [PubMed: 27471967]
- [60]. Leonard AC and Helmstetter CE, Cell cycle-specific replication of *Escherichia coli* minichromosomes, *Proc. Natl. Acad. Sci. USA* 83, 5101 (1986). [PubMed: 3523483]
- [61]. Helmstetter CE and Leonard AC, Coordinate initiation of chromosome and minichromosome replication in *Escherichia coli*, *J. Bacteriol* 169, 3489 (1987). [PubMed: 3301802]
- [62]. Wang X, Lesterlin C, Reyes-Lamothe R, Ball G, and Sherratt DJ, Replication and segregation of an *Escherichia coli* chromosome with two replication origins, *Proc. Natl. Acad. Sci. USA* 108, E243 (2011). [PubMed: 21670292]
- [63]. Salman H, Brenner N, Tung C-K, Elyahu N, Stolovicki E, Moore L, Libchaber A, and Braun E, Universal Protein Fluctuations in Populations of Microorganisms, *Phys. Rev. Lett* 108, 238105 (2012). [PubMed: 23003996]
- [64]. Wang P, Robert L, Pelletier J, Dang WL, Taddei F, Wright A, and Jun S, Robust growth of *Escherichia coli*, *Curr. Biol* 20, 1099 (2010) [PubMed: 20537537]
- [65]. Schaechter M, Williamson JP, Hood JR Jr., and Koch AL, Growth, cell and nuclear divisions in some bacteria, *J. Gen. Microbiol* 29, 421 (1962). [PubMed: 13976593]
- [66]. Le Treut G, Si F, Li D, and Jun S, Quantitative examination of five stochastic cell-cycle and cell-size control models for *Escherichia coli* and *Bacillus subtilis*, *Front. Microbiol* 12, 721899 (2021). [PubMed: 34795646]
- [67]. Riber L, Olsson JA, Jensen RB, Skovgaard O, Dasgupta S, Marinus MG, and Løbner-Olesen A, Hda-mediated inactivation of the DnaA protein and dnaA gene autoregulation act in concert to ensure homeostatic maintenance of the *Escherichia coli* chromosome, *Genes Dev* 20, 2121 (2006). [PubMed: 16882985]
- [68]. Donachie WD, Relationship between cell size and time of initiation of DNA replication, *Nature (London)* 219, 1077 (1968). [PubMed: 4876941]



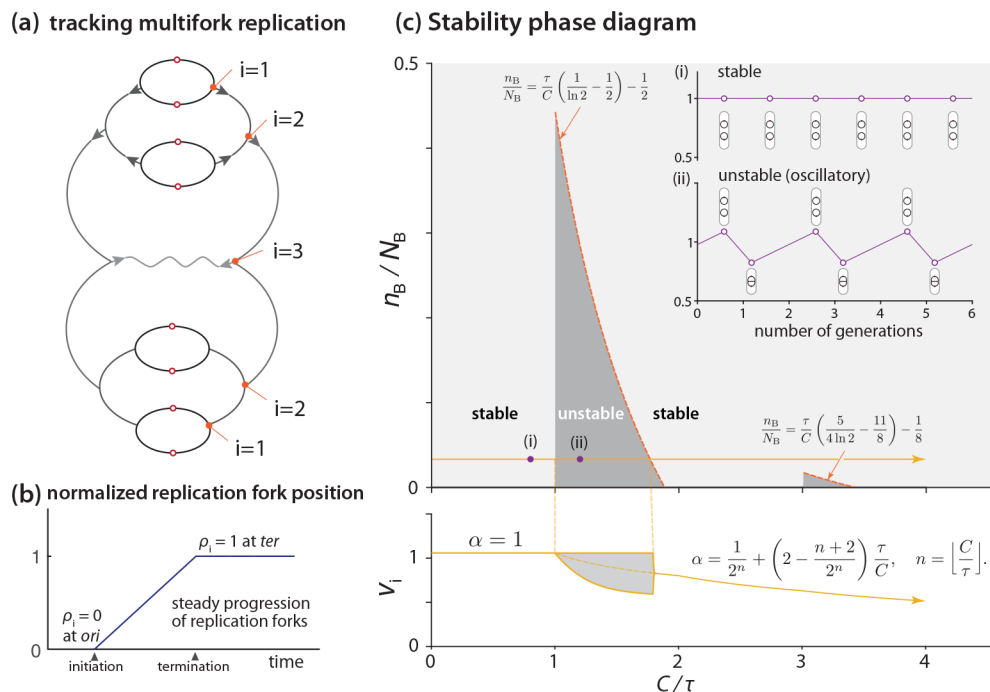
**FIG. 1.** Protein concentration in eukaryotes vs bacteria. (a) Left: morphogen gradient in the French flag model in developmental biology. Right: Oscillation of cyclin concentration for eukaryotic cell-cycle control. (b) Balanced biosynthesis in bacteria.

**(a) One-initiator minimal cell****(b) Initiator-titration model v1****(c) Initiator-titration model v2****FIG. 2.**

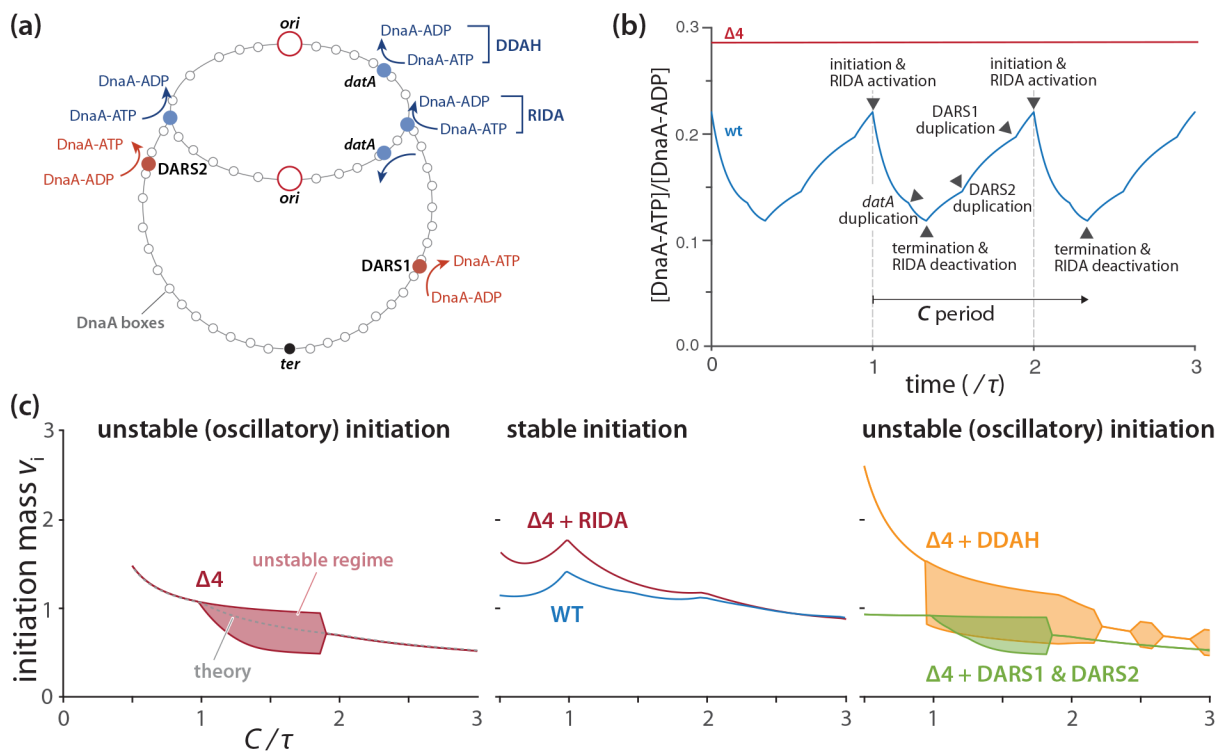
Initiation control. (a) Hypothetical minimal cell. (b) Initiator-titration model v1 [19]. (c) Initiator-titration model v2 (this work).



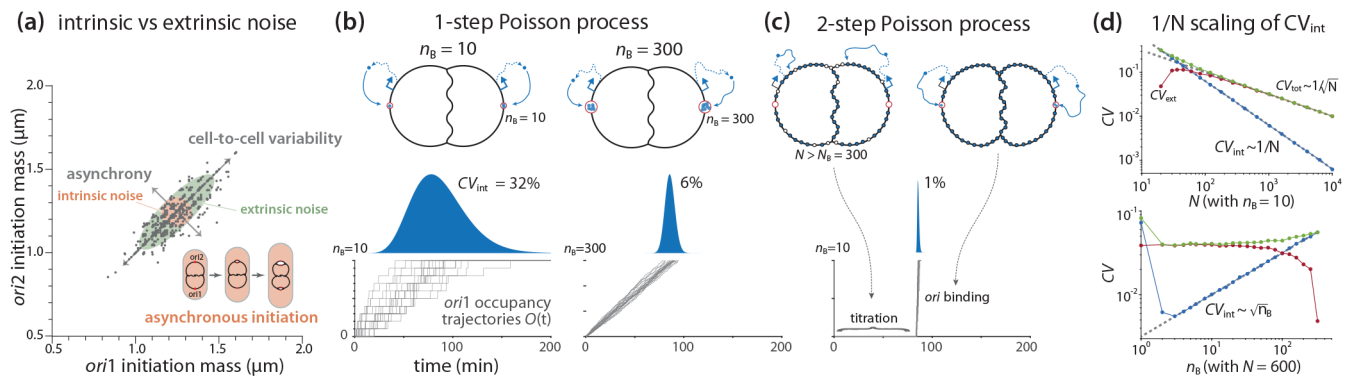
**FIG. 3.** Initiation control of the protocell by initiator protein counting. (a) Model sequence of titration and initiation. (b) Change in the copy numbers of initiators and initiator binding sites during the cell cycle under the condition of two overlapping cell cycles ( $C < \tau < C + D$ ). The initiation condition is  $I(t = t_{ini}) = B(t = t_{ini})$ . (c) Predicted initiation mass in different growth conditions ( $C/\tau$ ) by assuming that  $c_1$  is a constant [33,34].



**FIG. 4.** Dynamical stability analysis for initiation of the protocell. (a) Multifork replication tracker.  $i = 1$  represents the group of replication forks closest to *ori*. (b) Linear (in time) progression of the replication forks from *ori* to *ter* on the circular chromosome. *ter* is on the opposite end of *ori* on the chromosome. (c) Stability phase diagram ( $n_b/N_b$  vs  $C/\tau$  space). Below: initiation mass vs  $C/\tau$  in the condition of  $n_b/N_b = 1/30$  and constant  $c_1$  [33,34]. Inset: stable initiation events vs unstable (oscillatory) initiation events.

**FIG. 5.**

Initiator-titration model v2 predictions (see Sec. II D for more details). (a) External DnaA-ATP  $\leftrightarrow$  DnaA-ADP conversion elements in *E. coli*. RIDA is the only component that depends on the active replication forks. (b)  $[DnaA-ATP]/[DnaA-ADP]$  varies during the cell cycle predicted by computer simulations in the wild-type cells. By contrast, the  $\Delta 4$  mutant that lacks all extrinsic DnaA-ATP  $\leftrightarrow$  DnaA-ADP conversion elements in (a) shows a constant ratio. (c) Predicted initiation mass in different growth conditions ( $C/\tau$  with fixed  $C$ ) by assuming a constant  $[DnaA]$  [33,34]. RIDA, the replication-dependent DnaA-ATP  $\rightarrow$  DnaA-ADP mechanism alone can restore stability as long as titration is present. None of the other extrinsic DnaA-ATP  $\leftrightarrow$  DnaA-ADP conversion elements can restore the initiation stability.

**FIG. 6.**

Initiation precision and the reduction in asynchrony by  $1/N$  scaling in the two-step Poisson process by titration. (a) Asynchrony (intrinsic noise), extrinsic noise, and cell-to-cell variability in initiation control. Gray dots are single-cell data of wild-type *E. coli* from [13]. (b) Initiation by the first-passage-time model based on a simple Poisson process with  $n_b$  the threshold at each *ori*. Left:  $n_b = 10$  vs right:  $n_b = 300$ . (c) Synchronized initiation by titration in the two-step Poisson process.  $N$  is the mean total number of initiator proteins required for triggering initiation at both *ori*'s. (d) Simulation of the scaling behavior of the intrinsic noise ( $CV_{int}$ ), the extrinsic noise ( $CV_{ext}$ ), and the total CV ( $CV_{tot}$ ) (see Appendix G). Top: varying  $N$  with fixed  $n_b$ ; bottom: varying  $n_b$  with fixed  $N$ . The gray dashed lines are from Eq. (17).