# Models of microbial cell cycles

The Harvard community has made this article openly available. **Please share** how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th>Citation</th>
<th>Ho, Po-Yi. 2019. Models of microbial cell cycles. Doctoral dissertation, Harvard University, Graduate School of Arts &amp; Sciences.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:42029521">http://nrs.harvard.edu/urn-3:HUL.InstRepos:42029521</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Models of microbial cell cycles

a dissertation presented
by
Po-Yi Ho
to
The Harvard John A. Paulson School of Engineering and Applied Sciences

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Applied Physics

Harvard University
Cambridge, Massachusetts
May 2019
Models of microbial cell cycles

Abstract

Cell division is a fundamental process of life, yet how the timing of cell division is regulated in microorganisms remain unclear. In this dissertation, I use mathematical models to investigate the problem at the single-cell and molecular, as well as the phenomenological and mechanistic levels. First, I show how stochastic models similar to those used to describe random walks can describe the timing of divisions for diverse microbes. Together with analysis of single-cell data, the models showed that microbes in all three domains of life follow the same regulation strategy, despite drastic differences in their cell cycles. In particular, I show how different modes of coupling between DNA replication and cell division can coordinate the two processes when they occur in parallel to generate the same strategy for timing divisions in different microbes. I then proceed to the molecular level, and provide a progress report on the construction of a biophysical model for the mechanism underlying the regulation of the initiation of DNA replication in bacteria. Finally, I conclude by proposing a mechanistic model for how the circadian clock affects division timing in cyanobacteria. As a whole, the dissertation provides a step towards a quantitative and multi-scale understanding of microbial cell cycles.
## Contents

1. Introduction
2. Modeling cell size regulation: from single-cell statistics to molecular mechanisms
3. Archaeal cells share common size control with bacteria despite noisier growth and division
4. Simultaneous regulation of cell size and DNA replication in bacteria
5. Interrogating the *Escherichia coli* cell cycle by cell dimension perturbations
6. A parallel adder coordinates mycobacterial cell cycle progression and cell size homeostasis in the context of asymmetric growth and organization
7. Models of the molecular mechanism underlying cell cycle regulation in bacteria
8. A mechanistic model for the regulation of the timing of cell division by the circadian clock in the cyanobacterium *Synechococcus elongatus*
9. Outlook

References
Foremost, I thank my advisor Prof. Ariel Amir. Thank you for your support. I look forward to future interactions and collaborations.

I thank my committee members Prof. Michael Brenner and Prof. David Nelson.

I thank my collaborators who made much of the work in this thesis possible: Zheng Hai, Felix Barber, Dr. Ye-Jin Eun, Michelle Logsdon, and Prof. Bree Aldridge.

I am greatly indebted to my collaborators and mentors Prof. Somenath Bakshi, Prof. Chenli Liu, and Prof. Nancy Kleckner.

I am deeply grateful to be accompanied by my best friends and colleagues Dr. Felix Wong and Dr. Jie Lin.
Cell division, in which a cell splits into two smaller cells, is a fundamental process of life\textsuperscript{1}. For unicellular microorganisms, or microbes, cell division corresponds to reproduction, and marks the beginning and end of a generation of the cell cycle. An important determinant of the duration of a generation, or the generation time, of individual microbial cells is the mode of growth of cell size, since cells growing and dividing under a constant environment maintain an average cell size that is constant over time, implying that the generation time depends on the rate of cell growth\textsuperscript{2}. Moreover, like many biological processes, cell division is stochastic. For example, noise in gene expression or fluctuations in the environment can impart randomness onto the generation time. The stochasticity in timing cell divisions in turn adds to the heterogeneity between genetically identical cells, which underlies important phenomena such as persistence under stress\textsuperscript{3}. Despite its importance, the basic problem of what determines the generation time and its stochasticity remains unclear.
Recent advances in microfluidics and microscopy have allowed high throughput measurements of single-cell growth and division. Using these methods, it was found that the model bacterium *Escherichia coli*, while growing with a constant exponential rate under nutrient rich media, adds a constant size from birth to division, a strategy for timing divisions known as an adder\textsuperscript{4,5,6,7}. Intriguingly, the adder is prevalent among microbes in all three domains of life, despite drastic differences in growth morphologies and cell cycle processes\textsuperscript{8,9,10}. For example, a key cell cycle process that affects division timing is DNA replication, since a complete round of DNA replication must precede cell division in order for division to generate viable progenies. In eukaryotes such as the budding yeast *Saccharomyces cerevisiae*, the cell cycle proceeds sequentially and completes only one round of DNA replication before cell division\textsuperscript{11}. In contrast, fast growing bacteria such as *E. coli* may double in size faster than their chromosomes can replicate, in which case the cells maintain parallel, overlapping rounds of DNA replication\textsuperscript{12}. Despite these differences in DNA replication, *S. cerevisiae* and *E. coli* both appear to obey the adder strategy from birth to division\textsuperscript{8,13}. In addition to the coupling between DNA replication and cell division, the timing of the initiation of DNA replication also differs among microbes, and the underlying mechanisms remain unclear for some of the most studied model organisms *S. cerevisiae* and *E. coli*\textsuperscript{14}.

In this dissertation, we construct mathematical models and compare them with single-cell level experiments to investigate the above phenomena and problems. First, we developed a modeling framework that can describe the distributions of and correlations between cell size and generation time produced by the adder, as well as other modes of regulation. The framework can be extended to incorporate various sources of stochasticity to describe the diverse microbes that obey the adder. Second, we constructed models that couple DNA replication and cell division, and showed how different couplings may give rise to the emergent adder from birth to division. We then investigated various molecular level models for mechanisms that underlie
the regulation of the initiation of DNA replication. Finally, we applied the tools and insights gleamed from the above endeavors to construct a mechanistic model for how the circadian clock affects division timing in cyanobacteria. Taken together, our works show how mathematical models can unify single-cell observations in diverse microbes, distinguish among different modes of cell cycle regulation, and guide the search for a molecular level understanding of the microbial cell cycle.

1.1 Outline

Chapter 2 is based on a published review article, Ref. 15. Chapters 3-7 are based on published articles, Refs. 9,10,13,14,16. Chapter 8 is based on an article in preparation. In addition, Chapter 7 contains various unpublished results.

Chapter 2 introduces the problem of how microbes achieve a narrow distribution of cell sizes, or cell size homeostasis. The chapter summarizes how single-cell data and mathematical models revealed that several microbes, including the well studied model organisms E. coli, Bacillus subtilis, and S. cerevisiae are adders. The chapter first reviews two types of models for the problem. The first is based on deterministic dynamics on top of which stochasticity is introduced. The second specifies the instantaneous probability to divide, or the probability to divide per unit time, as a function of cellular physiological parameters such as cell size. The chapter describes several methods for analyzing the behavior of the models. The chapter concludes by introducing directions investigated in the rest of the dissertation, including the nature of the coupling between DNA replication and cell division and the molecular mechanism underlying the initiation of DNA replication.

Chapter 3 extends the modeling framework introduced in Chapter 2 to investigate division timing in an archaeon, which resembles prokaryotes in morphology but eukaryotes in DNA
replication. It was found that the archaeon faces larger stochasticity in processes related to cell growth and cell division, which leads to distributions of and correlations between cell size and generation time incompatible with the modeling framework introduced in Chapter 2, obscuring the underlying regulation of division timing. We extended the modeling framework to incorporate the larger stochasticity, and showed that the archaeon obeys the same regulation of division timing as *E. coli* at the phenomenological level. Together, Chapters 2 and 3 establishes a phenomenological level understanding of when and at what size microbes divide. They also show how mathematical models similar to those for random walks can be applied to analyze single-cell data to generate biological insights.

Chapter 4 investigates the microbial cell cycle beyond only cell division events, and introduces the concept of a tight coupling between DNA replication and cell division. It shows how an adder model at the level of initiations of DNA replication can simultaneously regulate the number of ongoing replication forks within a cell and the size of the cell. The model generates single-cell correlations for cells grown in a constant environment and population level scalings among cells grown in different environments that are in agreement with experiments in *E. coli* and *B. subtilis*. The model, referred to as the adder-per-origin model, supposes that a constant size per origin is added between initiation events and that initiation triggers division after a constant time. The validity of these two assumptions were tested experimentally in Chapters 5 and 6, and their origins explored theoretically in Chapter 7.

Chapter 5 explores and clarifies the nature of the two assumptions of the adder-per-origin model via population level experiments on genetically perturbed *E. coli*. Two genes, *ftsZ* and *mreB*, involved in septum formation and cell wall synthesis, respectively, were placed under inducible control by our collaborators. The perturbations to the expression level of the two genes led to changes in cell dimensions, as well as to the time from the termination of DNA replication to the corresponding cell division. Invariant despite these changes was the average
cell volume per chromosomal origin of replication at initiation. This result suggests that the regulation of initiation can remain unchanged despite perturbations to the coupling between initiation and division, and that volume, and not other dimensions such as surface area, is the key phenomenological variable determining the timing of initiation.

Chapter 6 applies the line of investigation introduced in Chapters 4 and 5 to a clinically relevant microbe *Mycobacterium smegmatis*, a close relative of the tuberculosis pathogen. Single-cell data demonstrated that *M. smegmatis* exhibits larger heterogeneity among cells than *E. coli*, but still obeys the adder strategy from birth to division. However, the relative timing between initiations and divisions cannot be explained by the adder-per-origin model. Instead, we developed a new model in which initiation triggers division after a constant size increment, or a parallel adder model, to capture the data. This results demonstrates that there are multiple ways to couple initiation and division, and that more than one of these couplings can generate the adder strategy from birth to division as an emergent behavior. Together, Chapters 4 to 6 establishes a phenomenological level understanding of how DNA replication and cell division are coupled.

Chapter 7 proceeds beyond the phenomenological level, and provides a progress report on the construction of a mathematical model at the molecular level for the mechanism underlying the timing of initiation. The report first develops a conceptual model without specifying the identity of the molecular players involved, and shows that the model robustly generates adder behavior in face of biological noise. The report then proposes a model for the dynamics of DnaA, a protein required for the initiation of DNA replication in bacteria, as a preliminary working model for the mechanism timing initiation.

Finally, Chapter 8 applies the tools and methods gleamed from the previous chapters to investigate how the circadian clock, or daily oscillations in cellular activity that are relatively robust to environmental perturbations, affects the timing of divisions in cyanobacteria. A mechanistic
model is proposed where a protein limiting for division is accumulated at a rate proportional
to volume growth and modulated by the clock, and shows that the model can describe existing
experiments. The result suggests that the signal from the clock is integrated over time, rather
than acting instantaneously as often assumed, to affect division timing. Together, Chapters
7 and 8 demonstrate how the models developed in this dissertation may guide experiments in
clarifying molecular mechanisms.
2

Modeling cell size regulation: from single-cell statistics to molecular mechanisms

2.1 Introduction

Most microorganisms regulate their cell size, as evidenced by their narrow cell size distributions. In particular, all known species of bacteria have cell size distributions with small coefficient of variations (CV, standard deviation divided by the mean), which can be as low as 0.1^2. For cells that grow exponentially, a small CV for size implies a small CV for interdivision times. However, a small CV for interdivision times is not sufficient to regulate cell size, as we will show that a simple “timer” strategy cannot regulate cell size in face of fluctuations. Cells must therefore have a way to effectively measure size.

The physiological implications of cell size remain under debate. In the context of bacteria, this is discussed in detail in a recent, excellent review^17, which also stresses the intimate connection
between the problem of cell size regulation and that of cell cycle regulation. For instance, cell division, which mechanistically determines cell size, is coupled to DNA replication. In this chapter, we will not focus on the rich biology behind this problem, which will be discussed in Chapters 4 to 7. Instead, we will elaborate on the various phenomenological models developed to study this problem over the last several decades. These are typically coarse-grained models, which consider the cell as a whole and describe cell volume at various stages of the cell cycle. They often seek to capture the statistics of the random process underlying cell size regulation. For example, what is the relation between cell size and interdivision time, and what distributions characterize the fluctuations in these variables?

A devil’s advocate or a biologist may ask why one would care about these questions. Quantitatively describing the distributions and finding scaling relations between variables are worthy goals from a physicist’s statistical mechanical point of view, but can such phenomenological modeling shed light on the biology? Three distinct examples support that the answer to this question is affirmative.

First, for several bacterial model species, including *E. coli* and *B. subtilis*, cell volume scales exponentially with growth rate, and proportionally with, loosely speaking, chromosome copy number (see Chapter 5). The scaling constant is in fact equal to the time from the initiation of DNA replication to cell division. It was shown fifty years ago that this observation can be rationalized within a model in which the regulation of cell size does not occur via controlling the timing of cell divisions, but rather via controlling the timing of the initiation of DNA replication. In this way, a quantitative pattern on the phenomenological level, with the aid of mathematical modeling, led to an important insight regarding bacterial physiology. The same empirical observation helped to address whether cell size regulation occurs over cell volume, surface area, or other dimensions. Experiments in rod-shaped bacteria often measure cell length, which cannot distinguish between these possibilities since cell width in these bacteria is very
narrowly distributed (CV < 0.05)\(^7\). As a result, in addition to cell volume, both cell surface area and length have been proposed to set cell size\(^6,20\). However, recent experiments in \textit{E. coli} showed that the same scaling relation holds, but only for cell volume and not surface area or width, under genetic perturbations to cell dimensions\(^{16}\). This result supports that volume is the key phenomenological variable controlling cell size. Below, we use the term cell size for generality while keeping the above discussion in mind. We further discuss the coupling between DNA replication and cell division in Chapters 4 to 7.

Second, a naive proposal for cell size regulation is a timer strategy, in which cells control the timing of their cell cycles so that, on average, cell size doubles from birth to division. However, it can be shown by theoretical arguments alone that this mode of regulation is incompatible with the small CVs of cell size distributions if cell volume grows exponentially in time at the single-cell level, as seen in experiments\(^{21}\). This is because the cumulative effect of noise will cause the variance in cell size to diverge. Explicitly, consider exponentially growing cells with a constant growth rate \(\lambda\) and stochastic interdivision time \(t_d\). A cell born at size \(v_b\) will generate a progeny of size \(v'_b = v_b e^{\lambda t_d/2}\), assuming perfect symmetric division. Let \(x = \ln(v_b/v_0)\) be the log-size, where \(v_0\) is a constant that sets the mean cell size, and \(x'\) the log-size at the next generation, then

\[
x' = x + \lambda t_d - \ln 2.
\] (2.1)

Uncorrelated fluctuations in \(t_d\) will then lead to a random walk in log-sizes with fluctuations accumulating as the square root of the number of divisions. Thus, the cell size distribution in a growing population will not reach stationarity via a timer strategy (Figure 2.1a), and a different strategy is needed to achieve narrow distributions (Figure 2.1b). Note that without fluctuations, Eq. 2.1 becomes \(x' = x\) if \(t_d = \ln 2/\lambda\), so that cell size is maintained. This example therefore shows that it is necessary to introduce stochasticity to models of cell size regulation.
as the failure of a timer strategy cannot be revealed otherwise.

As a third example, some of us recently investigated the properties of the resulting size regulation strategy from models of molecular mechanisms that do not specify the identity of the molecular players, but nonetheless propose concrete molecular network architectures. Two models were considered, one proposing that cell division is triggered by the accumulation to a threshold number of an initiator protein\textsuperscript{22}, and another that the dilution of an inhibitor triggers an event in cell cycle progression\textsuperscript{23}. It was shown theoretically that in the context of budding yeast \textit{S. cerevisiae}, both of these seemingly reasonable size regulation strategies fail to regulate cell size in the case of symmetric division\textsuperscript{14}. While there could be other explanations, this appears to be a strong constraint that may have contributed to the evolution of asymmetric division in budding yeast. We further discuss mathematical models for the molecular mechanisms underlying microbial cell cycles in Chapters 7 and 8.

These very different examples show how phenomenological modeling, in combination with single-cell or bulk-level level experiments quantifying cell growth, can lead to biologically relevant conclusions and constrain biological mechanisms. Furthermore, this approach allows to construct a theoretical “phase diagram” (e.g. Ref. \textsuperscript{24}), showing not where biology lies, but where biology may exist. In this vein, we proceed with the following aphorism in mind, “All models are wrong; some models are useful.”\textsuperscript{25}

In this chapter, we describe various existing phenomenological models for cell size regulation, some dating decades back and many very recent, and also present several novel results. First, we introduce discrete stochastic maps (DSM) to model cell size regulation and the various, approximate methods of solving for the distributions and correlations they generate. We discuss the connection between the problem of cell size regulation and that of diffusion in a confining potential and autoregressive modeling in time-series analysis. We systematically show, for the first time to our knowledge, that cell size regulation in \textit{E. coli} can be approximated well by a
stochastic model where the cell size at the next generation depends only on the cell size at the present generation. Next, we review continuous rate models and their mapping to DSMs. We then review recent works that analyzed DSMs at higher precision. They revealed that while the qualitative stability regions can be found via approximate methods, detailed statistical results are more nuanced, and specifically, power-law tails may often be generated by DSMs. Finally, we review recent results beyond the phenomenological level, including molecular implementations of different strategies for cell size regulation, the problem of protein number regulation, and the effects of cell size regulation on population growth.

2.2 Models for cell size regulation

To resolve the problem of an unconfined random walk in log-sizes in Eq. 2.1, feedback must be introduced so that larger cells divide sooner than average. Some intuition for the problem of cell size regulation can be gained by considering the familiar scenario of overdamped Brownian motion in a confining potential. This scenario can be described by the Langevin equation describing the dynamics of position $x$,

$$\frac{dx}{dt} = -\frac{1}{\gamma} V'(x) + \sigma \xi. \quad (2.2)$$

Here, $\gamma$ is a drag coefficient that relates the force to the velocity in the overdamped limit, $V(x)$ is the confining potential, and $\sigma$ is the magnitude of the fluctuations described by the stochastic variable $\xi$, which has correlations $\langle \xi(t') \xi(t' + t) \rangle = \delta(t)$. In this review, $\langle \cdot \rangle$ denotes the ensemble average. In the absence of a potential $V(x) = 0$, Eq. 2.2 reduces to unconfined diffusion, whose hallmark is the linear dependence of the mean-squared-displacement $\langle x^2 \rangle$ on time. In a quadratic potential $V(x) = kx^2/2$, Eq. 2.2 corresponds to diffusion confined by a linear restoring force. In this case, Eq. 2.2 is known as an Ornstein-Uhlenbeck (OU) process.
and is useful in describing a plethora of physical phenomena. It can be written as

$$\frac{dx}{dt} = -\frac{k}{\gamma} x + \sigma \xi, \quad (2.3)$$

where $k$ is the strength of the restoring force.

The probability density $p(x, t)$ corresponding to Eq. 2.3 satisfies the Fokker-Planck equation that describes its temporal dynamics

$$\frac{\partial p}{\partial t} = \frac{k}{\gamma} \frac{\partial}{\partial x} (xp) + \frac{\sigma^2}{2} \frac{\partial^2 p}{\partial x^2}. \quad (2.4)$$

The stationary $\partial p/\partial t = 0$ solution is a Gaussian distribution

$$p(x) = \sqrt{\frac{k}{\pi \gamma \sigma^2}} \exp \left( -\frac{kx^2}{\gamma \sigma^2} \right). \quad (2.5)$$

Indeed, Eq. 2.5 is equal to the Boltzmann distribution $p(x) \propto \exp \left( -V(x)/k_B T \right)$, where $k_B$ is the Boltzmann constant and $T$ is the temperature, since $\sigma^2 = 2D = 2k_B T/\gamma$ by the Einstein relation for the diffusion coefficient $D$. As $k \to 0$, the strength of the confining potential weakens. At $k = 0$, the variance of $x$ diverges. However, for any $k > 0$, the variance of $x$ will be finite. The autocovariance $\langle x(t') x(t + t') \rangle$ can be obtained via integration of Eq. 2.3. At stationarity, the autocovariance is exponentially decaying

$$\langle x(t') x(t + t') \rangle = \frac{\gamma \sigma^2}{2k} \exp \left( -\frac{k}{\gamma} |t| \right). \quad (2.6)$$

The familiar example of an OU process turns out to be similar to the problem of cell size regulation, but with the variable $x$ now representing cell size. We now review the formulation of the problem of cell size regulation as a discrete analogue of an OU process.
2.2.1 Discrete stochastic maps

Figure 2.1c shows single-cell data obtained via microfluidic devices that trap single cells in micro-channels to allow measurements of physiological properties such as cell size for many generations\(^4\). The problem of cell size regulation may be investigated initially by considering only division events. The data in this case consist of cell size at birth, division, and interdivision time over many generations. What are the distribution and correlations of cell size at birth and division, and what size regulation strategies lead to such statistics?

At a phenomenological, coarse-grained level, a size regulation strategy can be specified as a map that takes cell size at birth \(v_b\) to a targeted cell size at division \(v_a\) with a deterministic strategy \(f(v_b)\),

\[
v_a = f(v_b). \tag{2.7}
\]

In face of biological stochasticity, the actual cell size at division \(v_d\) is \(v_a\) subject to some coarse-grained noise term. For example, the noise term can be size-additive, so that \(v_d = v_a + \xi_v\), where \(\xi_v\) is uncorrelated between generations. The noise term can also be time-additive. In this case, the stochastic interdivision time \(t_d\) can be written as \(t_d = t_a + \xi_t\), where \(\xi_t\) is the noise term. The deterministic component \(t_a\) can be determined by assuming a constant exponential growth rate \(\lambda\). The deterministic size regulation strategy in Eq. 2.7 then leads to

\[
t_a = \ln \left( \frac{f(v_b)}{v_b} \right) / \lambda. \tag{2.8}
\]

In the case of time-additive noise, if division is perfectly symmetric so that \(v'_b = v_d/2\), then the cell size at birth at the next generation is

\[
v'_b = f(v_b) e^{\lambda \xi_t} / 2. \tag{2.9}
\]
Figure 2.1: Cell size at birth in simulations (a, b) and in experiments (c) as a discrete time-series, with insets showing a zoomed in view of one particular trial of simulation (b) and the data (c). (a, b) Multiple trials (different colors) of numerical simulations of the DSM in Eq. 2.9 with the simple, one-line pseudo-code: $v_{i+1} = (2 (1 - \alpha) v_i + v_0) 2^{\xi t}/2$, where $v_i$ denote cell size at birth in the $i$-th generation, $\xi_t$ is a normally distributed random variable with zero mean and variance $\sigma_t^2$, and $v_0$ is a constant that sets the mean cell size. Here, $\sigma_t = 0.22$ and $\langle v_0 \rangle = 1$. (a) $\alpha = 0$ leads to unconfined diffusion and a divergent distribution. (b) Any $0 < \alpha < 2$, here $\alpha = 0.5$, has the necessary feedback to achieve a stationary distribution. (c) Data from Ref. 27. $v_b$ in this case represents cell length at birth and is normalized so that $\langle v_b \rangle = 1$.

The two forms of noise lead to distributions of different shapes. Experiments have shown that distributions of cell sizes at birth are skewed and can be approximated as a log-normal but that interdivision time distributions can be approximated as normal$^6, 28$. These are consistent with a normally distributed time-additive noise, which we use below.

2.2.2 Approximate solution via first order expansion

The DSM described in Section 2.2.1 is in general difficult to solve for an arbitrary size regulation strategy $f(v_b)$. One method makes the approximation to focus on the behavior of $f$ near the mean size $\langle v_b \rangle$, since the size distribution has a small CV. A size regulation strategy can be linearized by expanding about $\langle v_b \rangle$, $f(v_b) \approx f(\langle v_b \rangle) + f'(\langle v_b \rangle) (v_b - \langle v_b \rangle)$. In this approximation, all regulation strategies that agree to first order will lead to similar distributions near $\langle v_b \rangle$. The following is a convenient choice$^5$,

$$f(v_b) = 2 v_b^{1-\alpha} v_0^\alpha, \quad (2.10)$$
where \( v_0 \) is an arbitrary constant. As shown below, \( \langle v_b \rangle \approx v_0 \), and hence, the slope has value \( f' (\langle v_b \rangle) = 2 (1 - \alpha) \). The value of \( \alpha \) therefore determines the strength of regulation. \( \alpha = 1 \) corresponds to the strongest regulation, a “sizer” strategy where cells attempt to divide upon reaching \( f(v_b) = 2v_0 \). \( \alpha = 0 \) represents no regulation and corresponds to the timer strategy where cells attempt to divide upon reaching \( f(v_b) = 2v_0 \). Recent works have shown that the statistics of cell size can be generated by a regulation strength \( \alpha = 1/2 \) that is between the two extremes\(^6,7,28\). In this case, the slope has value \( f (\langle v_b \rangle) = 1 \) and so is an approximation to the “adder” strategy where cells attempt to divide upon reaching \( f(v_b) = v_b + v_0 \). Several microorganisms in all three domains of life have been shown to approximately follow an adder strategy, or less prescriptively, to exhibit adder correlations. We discuss the prevalence of adder correlations later.

Let \( x = \ln (v_b/v_0) \) be the log-size and \( x' \) denote \( x \) at the next generation. Eqs. 2.9-2.10 then lead to the simple stochastic equation

\[
x' = (1 - \alpha) x + \lambda \xi_t.
\]

(2.11)

At the \( n \)-th generation,

\[
x_n = (1 - \alpha)^n x_0 + \sum_{j=0}^{n-1} (1 - \alpha)^{n-1-j} \lambda \xi_t^{(j)},
\]

(2.12)

where \( x_i \) and \( \xi_t^{(i)} \) respectively denote the value of \( x \) and \( \xi_t \) at the \( i \)-th generation. The first term approaches zero as \( n \to \infty \) if \( 0 < \alpha < 2 \). If \( \xi_t \) is normally distributed with variance \( \sigma_t^2 \), then the variance \( \sigma_x^2 \) of \( x \) will be the sum of the variances in the series in the second term. The geometric
series converges for $0 < \alpha < 2$, and can readily be evaluated to give the variance $\sigma_x^2$ as

$$\sigma_x^2 = \frac{\lambda^2 \sigma_t^2}{\alpha (2 - \alpha)}. \quad (2.13)$$

Furthermore, since $x_n$ is a sum of normal variables, it will also be normally distributed. If $\alpha \leq 0$ or $\alpha \geq 2$, the sum of the series diverges, and hence there is no stationary distribution. The case $\alpha = 0$ produces unconfined diffusion and is analogous to the case where the strength of the restoring force is zero ($k = 0$) in an OU process, as seen in Eq. 2.5. Figure 2.1ab demonstrates the difference between time-series generated by $\alpha = 0$ and by $0 < \alpha < 2$. The variance of $t_d$ can be obtained similarly.

It is not obvious a priori whether the widths of the distributions of interdivision time and cell size are related. It turns out that the two CVs (denoted by $CV(\cdot)$) are related by a dimensionless quantity. The log-size is related to the actual size by $x = \ln (v_b/v_0) \equiv \ln (1 + \delta v_b)$. Since $\delta v_b = v_b/v_0 - 1$ is small, $x \approx \delta v_b = v_b/v_0 - 1$. Therefore, $CV(v_b) \approx \sigma_x$. Calculating $CV(t_d)$ in a similar manner leads to

$$\frac{CV(v_b)}{CV(t_d)} \approx \frac{\ln 2}{\sqrt{2\alpha}}. \quad (2.14)$$

Eq. 2.14 allows to extract the parameter $\alpha$ from CVs that can be accurately measured. Since $x$ and $t_d$ are both distributed normally, the model predicts that these distributions can be collapsed after normalizing by the mean and scaling according to Eq. 2.14, as seen in experiments.

The Pearson correlation coefficients (CC) between two variables (denoted by $C(\cdot, \cdot)$) can also be obtained. Since CCs are not affected by addition or multiplication by a constant, $v_b$ can be replaced by $x$ in the following calculations. The CC between cell size at birth of a mother cell
and that of the daughter cell is therefore\(^5\)

\[ C(v_b, v'_b) = C(x, x') = \frac{\langle xx' \rangle - \langle x \rangle^2}{\sigma_x^2}. \tag{2.15} \]

Substituting in Eq. 2.11,

\[ C(v_b, v'_b) = 1 - \alpha. \tag{2.16} \]

Importantly, the value of \(\alpha \approx 1/2\) extracted via the ratio of CVs in Eq. 2.14 also predicts the CC between size at birth and at division, in agreement with experiments\(^6,7\).

Similarly, using Eq. 2.8 and 2.10, the interdivision time can be written as

\[ t_d = \frac{\ln 2 - \alpha x}{\lambda} + \xi_t. \tag{2.17} \]

The CC between the interdivision times of a mother-daughter pair can then be shown to be\(^7,29\)

\[ C(t_d, t'_d) = -\alpha/2. \tag{2.18} \]

That this CC is non-zero has implications for the population growth rate, which we review later.

2.2.3 Autoregressive models and extensions to incorporate biological details

Eq. 2.11, obtained after linearization of the generically nonlinear DSM Eq. 2.9, is mathematically known as an autoregressive (AR) model, often used in time-series analysis and economics forecasting\(^39\). An AR model of order \(m\) (denoted by AR\((m)\)) takes the form

\[ x_i = b + \sum_{j=1}^{m} c_j x_{i-j} + \xi_i, \] where \(b\) and \(c_j\) are constants and \(\xi_i\) is a noise term uncorrelated for different \(i\). The model describes how previous values of the stochastic variable \(x\) influence
In an AR(1) model, the CC $C(x_i, x_{i+2})$ is non-zero because they are related via the intermediate variable $x_{i+1}$. However, the partial CC $C_{x_{i+1}}(x_i, x_{i+2})$ removes the effects of the intermediate variable and is zero. Experimentally determined values of $C(v_i, v_{i+1})$ is as predicted by Eq. 2.11, which is also a discrete analogue of an OU process.
In the same data set, \( C(v_i, v_{i+2}) \) is non-zero but \( C_{v_{i+1}}(v_i, v_{i+2}) \) is zero (Figure 2.2bc). Indeed, a vanishing \( C_{v_{i+1}}(v_i, v_{i+2}) \) implies that \( C(v_i, v_{i+2}) = C^2(v_i, v_{i+1}) \), which is the case here. This novel check systematically shows that cell size at birth in \( E. coli \) can be described by an AR(1) model. This result is a fortunate simplification, since for example, in certain mammalian cells, the CCs in the interdivision times between cousin cells (\( C_{cc} \)) cannot be determined from those between sister cells (\( C_{ss} \)) and between mother-daughter pairs (\( C_{md} \)). Indeed, experiments observe that \( C_{cc} > C_{md} \), contrary to the expected relation \( C_{cc} = C_{md}^2 C_{ss} \) in an AR(1) model\(^{31-32} \). The expected relation is violated also for cyanobacteria, which possess circadian clocks that affect division timing, which we discuss further in Chapter 8.

Extracting the regulation strength \( \alpha \) via Eq. 2.16 is analogous to estimating the parameters in AR models via the Yule-Walker equations that relate theoretical values of the parameters to theoretical values of the autocorrelation function (ACF)\(^{30} \). The ACF \( \rho(t) \) is the CC between variables separated by \( t \) time points,

\[
\rho(t) = C(x_i, x_{i+t}). \quad (2.20)
\]

As can be seen by Eq. 2.12, the ACF for the AR(1) model of Eq. 2.11 is simply

\[
\rho(t) = (1 - \alpha)^{|t|}. \quad (2.21)
\]

In this case, the ACF decays exponentially as in an OU process as seen in Eq. 2.6. The ACF of cell size at birth indeed decays exponentially (Figure 2.3a)\(^{33} \). Importantly, the estimated ACF is only meaningful after sufficient averaging to eliminate spurious fluctuations. This can
be done most clearly by computing the power spectral density (PSD)

\[ S(f) = \lim_{T \to \infty} \frac{1}{T} \left| \sum_{j=1}^{T} x_j e^{-i2\pi fj} \right|^2, \tag{2.22} \]

where \( T \) is the total number of observations in the time-series with data points \( x_j \). The PSD can also be calculated as the Fourier transform of the ACF according to the Wiener-Khinchin theorem. For the AR(1) model of Eq. 2.11, the PSD turns out to be\(^3\)

\[ S(f) = \frac{\alpha (2 - \alpha)}{1 - 2(1 - \alpha) \cos (2\pi f) + (1 - \alpha)^2 \sigma_x^2}. \tag{2.23} \]

This is again analogous to an OU process, since the Fourier transform of an exponential function is a Lorentzian function. There are significant oscillations only in the case \( \alpha \lesssim 2 \), for which the PSD peaks at high frequencies (Figure 2.3cd). The case of \( E. coli \), where \( \alpha \approx 1/2 \), is far from this regime (Figure 2.3ab). Therefore, experimentally observed fluctuations should not be confused for oscillations\(^3\).

The AR(1) model in Eq. 2.11 can be extended to incorporate details that are relevant to a variety of microorganisms. These include asymmetric and noisy divisions (e.g. in mycobacteria\(^9,34\)), noisy growth rates (e.g. in slow growing \( E. coli\)\(^28\) and in the archaeon \( H. salinarum\)^10), and diverse growth morphologies (e.g. the budding mode of growth of \( S. cerevisiae\)^8). First, noisy divisions and noisy growth rates can be incorporated by modeling the division ratio (daughter cell size at birth divided by mother cell size at division) and growth rate as \( 1/2 + \xi_r \) and \( \lambda + \xi_\lambda \), respectively, at each generation. If the fluctuations \( \xi_r \) and \( \xi_\lambda \) are small and uncorrelated, Eq. 2.11 becomes to first order in small variables

\[ x' \approx (1 - \alpha) x + \lambda \xi_t + 2 \xi_r. \tag{2.24} \]
Figure 2.3: Fluctuations versus oscillations. The ACF $\rho(t)$ (a,c) and the PSD $S(f)$ (estimated via the Welch method) (b,d) of cell size at birth in *E. coli* (a,b) and a simulated AR(1) model described by Eq. 2.11 with $\alpha = 1.9$ (c,d). Blue lines show experimentally determined ACF and PSD from the same data set as that in Figure 2.1. Green lines show simulation results. Dashed red lines show Eq. 2.21 (a,c) and Eq. 2.23 (b,d), for $\alpha = 0.49$ (a,b) and $\alpha = 1.9$ (c,d).
The fluctuation $\xi_\lambda$ enters as a first order correction to interdivision time $t_d$. The CVs and CCs can be calculated as before for Eq. 2.24 to show that the different fluctuations typically affect the CVs and CCs in different ways. For example, the CC between cell size at birth and at division, $C(v_b, v_d)$, is sensitive to fluctuations in division ratios, and is increased by large fluctuations in division ratios. However, the CC in cell size at birth between mother-daughter pairs, $C(v_b, v'_b)$, remains the same as in Eq. 2.16, and is independent of all noise terms. It is thus a robust detector of the underlying regulation strategy even in face of multiple sources of complicating stochasticity\textsuperscript{10}. The idea is applied to understand the cell cycle of the archaeon \textit{H. salinarum} in Chapter 3 and of the mycobacterium \textit{M. smegmatis} in Chapter 6. We also discuss later several models that incorporate other biological details and move beyond AR models.

2.2.4 Continuous rate models and higher order effects

Cell size regulation can also be modeled using continuous rate models (CRM)\textsuperscript{7,35-36,37}. A CRM has also been used recently to model how the clock affects division timing in cyanobacteria, as discussed in Chapter 8. In contrast to DSMs, CRMs consider not just discrete division events, but the continuous cell cycle. They specify the instantaneous division rate $h$, or the probability to divide per unit size increment, as a function of physiological parameters such as the current size $v$, size at birth $v_b$, growth rate $\lambda$, or the time $t$ since division. A simple choice of parametrization is the sloppy sizer model, $h = h(v)$. In this case, the probability for a cell of size $v$ to divide between the size interval $v$ and $v + dv$ is $h(v) \, dv$. Hence if $F(v_d|v_b)$ is the cumulative probability to have not divided at size $v_d$ given $v_b$, then $F(v_d|v_b)$ satisfies $F(v_d + dv|v_b) = F(v_d|v_b) (1 - h(v_d) \, dv)$. In the continuum limit, $dF(v_d|v_b)/dv = -h(v_d) \, F(v_d|v_b)$, so that

$$F(v_d|v_b) = \exp \left( - \int_{v_b}^{v_d} h(v) \, dv \right). \quad (2.25)$$
Eq. 2.25 can be written as
\[ h(v) = -\frac{d}{dv} \ln F(v|v_b), \]  
(2.26)
allowing to extract \( h(v) \) via single-cell experiments that measure \( F(v_d|v_b) \). The division rate can be formulated as a probability to divide per unit time increment as well, using the change of variables between size and time given by exponential growth. Analyses using CRMs have demonstrated that the current size is not the only determinant of the division rate because the sloppy sizer model fails to capture measured distributions of interdivision time and size increment from birth to division\(^3\). This implies that there exists a feedback on the time since birth, or equivalently the size at birth\(^3\). Specifically, a division rate in the form \( h = h(v - v_b) \) can simultaneously describe measured distributions of size at birth, interdivision time, and size increment from birth to division\(^7\).

A CRM can be approximately reduced to a DSM with the target size at division equal to the expectation value of the size at division given the size at birth\(^3\),
\[ f(v_b) = \int_0^\infty p(v|v_b) v dv, \]  
(2.27)
where \( p(v|v_b) = -dF(v|v_b)/dv \) is the probability density for a cell born at size \( v_b \) to divide at size \( v \). The nature and magnitude of the noise term can be determined by inverting the steps described below to map a DSM to a corresponding CRM. To do so, \( p(v|v_b) \) can be calculated from \( f(v_b) \) and a specified coarse-grained noise, then the division rate can be obtained using Eq. 2.26. For example, for a time-additive, normally distributed noise with variance \( \sigma_t^2 \), the division probability density for log-size \( x = \ln (v/v_0) \) is \( p(x|x_b) \propto \exp \left( -\frac{(x - g(x_b))^2}{2\lambda^2\sigma_t^2} \right) \), where \( g(x_b) = \ln (f(v_0 e^{x_b})/v_0) \). Since the typical \( x_b \) is much smaller than \( g(x_b) \), integration leads to

23
the division rate\textsuperscript{37}
\begin{equation}
    h(v, v_b) \approx \frac{\sqrt{2}}{v \sqrt{\pi \lambda \sigma_t}} H \left( \frac{1}{\sqrt{2 \lambda \sigma_t}} \ln \left( \frac{v}{f(v_b)} \right) \right),
\end{equation}
where $H(z) = \exp \left( -z^2 \right) / (1 - \text{Erf}(z))$ and Erf$(\cdot)$ is the error function. For the regulatory function Eq. 2.10, the division rate Eq. 2.28 becomes a function of only the instantaneous size when $\alpha = 1$, corresponding to a sizer strategy.

Although the CRM is generic and may capture complex behavior such as filamentation\textsuperscript{35}, it is not obvious a priori how to parametrize the division rate. On the other hand, the DSM has only a few parameters, is amenable to analytical treatment in several cases, and describes existing measurements well. The complexity sacrificed by DSMs and their first order approximate solutions may become important, for instance, when second and higher order terms become significant. However, higher order effects are difficult to detect unless the number of cells measured is large enough to suppress the confounding effects of fluctuations in the cell cycle. No existing experiments have achieved this regime, perhaps justifying the success of DSMs as models of cell size regulation\textsuperscript{37}.

2.2.5 More precise analysis of DSMs

The approximate first order solution in Section 2.2.2 predicts a log-normal size distribution for a regulatory function $f(v_b) = 2 \left( 1 - \alpha \right) v_b + v_0$ and a time-additive, normally distributed noise. However, closer inspection reveals that the size distribution has a power-law tail instead. This can be seen by analyzing which moments exist for a given regulation strength $\alpha$. Calculations similar to those in Section 2.2.2 show that if the $j$-th moment exists, so too do all the lower moments, but that for any $\alpha > 0$, there always exists an integer $j^*$ past which all moments cease to exist. This suggests that the size distribution has a power-law tail $p(v_b) \sim 1/v_b^{1+\beta}$ with $j^* < \beta \leq j^* + 1$, as confirmed by numerical simulations\textsuperscript{24}. 

24
The value of $\beta$ can be obtained precisely. The evolution of size distributions from one generation to the next can be written as an integral equation $p(v_b') = \int_0^\infty K(v_b', v_b) p(v_b) dv_b$, where the kernel $K(v_b', v_b)$ can be derived from the regulatory strategy $f(v_b)$. For a regulatory function in the form $f(v_b) = 2 (1 - \alpha) v_0 (v_b/v_0)^\eta + v_0$, it can be shown via an asymptotic analysis of the integral equation that a distribution with a power-law tail $1/v_b^{1+\beta}$ evolves to one with a power-law tail $1/v_b^{1+\beta/\eta^2}$\textsuperscript{24}. This implies that for $\eta = 1$, the size distribution indeed has a power-law tail, with

$$\beta = \frac{-2 \ln (1 - \alpha)}{\lambda^2 \sigma_t^2},$$

(2.29)

where $\sigma_t^2$ is the variance of the time-additive noise.

An alternative approach also led to the same power-law tail\textsuperscript{38}. In this approach, a DSM is approximated as a Langevin equation continuous in generations. Let $x = \ln (v_b/v_0)$ be the log-size at birth and let $n$ denote the generation number, then Eq. 2.9 can be written

$$x_{n+1} = x_n + \tilde{g}(x_n) + \lambda \xi_t,$$

(2.30)

where $\tilde{g}(x_n) = \ln (f(\exp(x_n) v_0)/v_0) - x_n$. To lowest order, Eq. 2.30 can be approximated by a Langevin equation continuous in $n$ as

$$\frac{dx}{dn} = \tilde{g}(x) + \lambda \xi_t.$$

(2.31)

As seen before in the context of an OU process described by Eq. 2.3, Eq. 2.31 leads to an equilibrium distribution of log-sizes $p(x) \propto \exp \left( -2 V(x) / (\lambda^2 \sigma_t^2) \right)$, where $V(x) = \int \tilde{g}(x') dx'$ is the effective potential. For the same regulatory function as above, the effective potential diverges linearly as $V(x) \sim -2x \ln (1 - \alpha)$. The equilibrium distribution therefore has a power-law tail $1/v_b^{1+\beta}$ with the same $\beta$ as in Eq. 2.29\textsuperscript{38}. Even further precision can be obtained via
a second order approximation which modifies the effective potential, but leaves the behavior of the power-tail unchanged\textsuperscript{38}.

2.3 Beyond phenomenological models of cell size regulation

As we previously alluded, the formalism of DSMs developed for the problem of cell size regulation can lead to insights on related problems at the molecular, single-cell, and the population level. Below, we discuss these in turn.

2.3.1 Molecular mechanisms to implement cell size regulation

How does a bacterial cell molecularly implement a size regulation strategy? The initiator accumulation model is a network architecture proposing that an initiator protein accumulates during cell growth to trigger cell division upon reaching a threshold copy number $\theta$\textsuperscript{22}. While experiments have suggested that the upstream control occurs over initiation of DNA replication rather than cell division in various microorganisms\textsuperscript{8,16,19,39}, we first review a simpler model where the accumulation of initiators triggers cell division. The model leads to the adder correlations observed in several species of bacteria and other microorganisms\textsuperscript{5,13,14}.

One possible molecular implementation of the initiator accumulation model is as follows\textsuperscript{40}. If the transcription rate of the initiator is assumed to be proportional to the cell volume, which grows exponentially in time, and if each transcript leads to a burst of protein production with mean burst size $b$, then the distribution of added cell size $\Delta v = v_d - v_b$ from birth to division has width\textsuperscript{40}

$$CV^2(\Delta v) = \frac{b^2 + 2b\theta + \theta}{(b + \theta)^2}. \quad (2.32)$$

Furthermore, the resulting distribution has only one characteristic size, the mean added cell
size $\langle \Delta v \rangle$, and therefore can be written as

$$p(\Delta v) = \frac{1}{\langle \Delta v \rangle} \tilde{p} \left( \frac{\Delta v}{\langle \Delta v \rangle} \right).$$  \hspace{1cm} (2.33)

Indeed, experiments showed that distributions of cell sizes with different means collapse after normalizing by the mean\textsuperscript{7,36,41}. The collapse suggests that $b$ and $\theta$ are constant within the implementation here. The same experiments also saw that the distributions of interdivision times collapse after normalizing by the mean doubling time, which is again captured by this model\textsuperscript{40}. There are also additional models that show such scaling collapse, such as an autocatalytic network subject to a threshold criterion for division\textsuperscript{42}, and the coarse-grained “adder-per-origin” model described below.

As discussed in the Introduction, control at other cell cycle events may lie upstream of cell division in various microorganisms. DSMs similar to those reviewed so far can be extended to describe cell cycle regulation. These models can not only produce emergent strategies of cell size regulation identical to those described by the division-centric models reviewed so far, but also describe additional statistics such as the correlations between cell size and various cell cycle timings\textsuperscript{43-44}.

As an example, we review below a model of cell cycle regulation in \textit{E. coli}, whose cell divisions appear to follow a constant time $T$ after the initiation of DNA replication for a broad range of mean growth rates\textsuperscript{12,47}. The time $T$ can be larger than the mean doubling time $\tau$, in which case the cells maintain multiple ongoing rounds of DNA replication. The tight coupling between initiation and division implies that the cell size at birth $v_d$ is

$$v_d = v_i e^{\lambda(T+\xi \tau)},$$  \hspace{1cm} (2.34)
where $v_i$ is the cell size at initiation, $\lambda = \ln 2/\tau$ is the growth rate, and $\xi_T$ describes fluctuations with magnitude $\sigma_T$ in the time between initiation and division. At a coarse-grained level, the initiator accumulation model can be described as

$$\tilde{v}_i' = (v_i + Ov_0) e^{\lambda \xi_t},$$

(2.35)

where $\tilde{v}_i'$ is the total cell size of the daughter cells (typically two) at the next initiation, $O$ is the number of origins of replication (i.e. the site along the chromosome at which DNA replication initiates), and $v_0$ is a constant. As in the division-centric model, regulation is subject to a time-additive noise $\xi_t$ with magnitude $\sigma_t$.

Analysis and simulations of the initiation-centric model of Eqs. 2.34-2.35 show that it produces emergent adder correlations at division$^{13,14}$, as long as the magnitude of the fluctuations in the coordination between initiation and division is much less than that in the control of initiation ($\sigma_t \gg \sigma_T$). This is indeed the case in experiments for fast-growing bacteria, although the picture appears different for slow-growing bacteria$^{28}$, which we discuss later. The model also generates cell size and interdivision time distributions whose CVs only depend on the magnitudes $\lambda \sigma_t$ and $\lambda \sigma_T$ of the fluctuations, and the regulation strength $\alpha$. The distributions therefore collapse after scaling by the mean if these parameters are constant across growth conditions. At the bulk-level, the initiation-centric model produces the observed exponential scaling of mean cell size with mean growth rate, as discussed in the Introduction, without requiring parameters to depend on mean growth rate$^{13}$. These results, together with previous results regarding the universality of cell size distributions, suggest that the initiator accumulation model may be a robust molecular mechanism that produces adder correlations, and that models of cell cycle regulation can continue to shed light on the underlying biology.

The coupling between DNA replication and cell division is discussed further in Chapters 3-6,
and the mechanism underlying the initiation of DNA replication is discussed in Chapter 7.

2.3.2 Regulation of protein numbers

Recent works have begun investigating the statistics of the copy numbers of proteins at the single-cell level in the same spirit as the problem of cell size regulation\textsuperscript{33,45,46,47}. In fact, for a constitutively expressed protein, the distributions of protein numbers at birth can be described by a DSM\textsuperscript{40}. Analysis analogous to that in Figure 2.2, but for the copy number of a constitutively expressed protein in \textit{E. coli} in the same data set\textsuperscript{27}, reveals that the partial CC is also zero in this case. However, it is unclear how protein number and cell size are simultaneously regulated.

One way to investigate this question is via a multi-dimensional, or vector, AR model. An AR(1) vector model in \(M\) dimensions can be written as

\[
\vec{x}' = A\vec{x} + \vec{b} + \vec{\xi},
\]

(2.36)

where \(\vec{x}\) is a vector of the abundances at birth of the \(M\) cellular components, which can include cell size, and \(\vec{x}'\) is the vector at the next generation. \(A\) is a \(M \times M\) matrix representing the regulatory interactions between components, \(\vec{b}\) is a vector representing the basal synthesis level between generations, and \(\vec{\xi}\) is a vector of noise terms uncorrelated between generations but may be cross-correlated at the same generation.

For the one-dimensional case, the condition for stationarity is that \(2 > \alpha > 0\) so that the variance of \(x\) in Eq. 2.13 is finite. This condition is equivalent to that the zero of \(1 - (1 - \alpha) z\) lie outside the unit circle. In the multi-dimensional case, the condition is similarly that all the zeros of \(\det(I - Az)\) lie outside the unit circle\textsuperscript{39}. Given a stable AR(1) vector model, the multi-dimensional analogue of the Yule-Walker equations can be used to estimate by maximum
likelihood the regulatory matrix $A$ from measurements. For the data set discussed above, this method results in

$$A = \begin{pmatrix}
0.50 \pm 0.02 & -0.02 \pm 0.01 \\
-0.16 \pm 0.02 & 0.60 \pm 0.02
\end{pmatrix},$$

where the first and second components are respectively cell size and protein number at birth (both normalized by their means), and plus-minus shows the standard error in the estimate. This novel result suggests that the copy number of this constitutively expressed protein does not affect cell size regulation, while cell size does affect the regulation of this protein number. It is unknown whether this result holds for all constitutively expressed proteins, and how this result will change for proteins that are not constitutively expressed.

To better understand cross-correlations between cell size and protein numbers from a mechanistic perspective, recent works have investigated a dynamical model in the form $\frac{d\vec{x}}{dt} = A\vec{x}$, where $\vec{x}$ is now the abundances of the cellular components during the cell cycle, and $A$ now describes the regulatory interactions in time. This model leads to the components growing as a sum of exponentials that can be approximated as a single exponential function during one generation, in agreement with experimentally observed exponential growth. Describing the statistics generated by dynamical models, and relating a dynamical model to a DSM and vice versa remain important open questions.

2.3.3 Effects of cell size regulation on population growth rate

At the single-cell level, genetically identical cells in the same clonal populations may have different interdivision times and growth rates. How does such variability at the single-cell level affect population growth? Models often assume that the interdivision times $t_d$ are uncorrelated between generations and independent of other variables. In this case, a simple relation connects the asymptotic population growth rate $\Lambda = \frac{(dN/dt)}{N}$, where $N$ is the number of
cells in the population, to the interdivision time distribution \( p(t_d) \),

\[
2 \int_0^\infty p(t_d) \exp(-\Lambda t_d) = 1. \tag{2.37}
\]

Importantly, given a fixed mean interdivision time, a larger variability in \( t_d \) increases \( \Lambda \). However, cell size regulation leads to negative correlations in \( t_d \) between generations, as seen in Eq. 2.18. In this case, recent results obtained by some of us showed that in an asynchronous, exponentially growing population - in which each cell is subject to variability in its single-cell growth rate, as well as to time-additive and size-additive noise in its cell size regulation by the regulatory function \( f(v_b) = 2(1 - \alpha) v_b + 2\alpha v_0 \) - the population growth rate is dependent only on the distribution of single-cell growth rates. In the limit of small correlations in growth rates between generations, variability in single-cell growth rates does not increase, but rather decreases the population growth rate\(^{29}\),

\[
\frac{\Lambda}{\langle \lambda \rangle} = 1 - \left(1 - \frac{\ln 2}{2}\right) CV^2(\lambda), \tag{2.38}
\]

Eq. 2.38 predicts that a population can enhance its population growth rate by suppressing the variability in single-cell growth rates given a fixed mean, which is consistent with the smaller CV of single-cell growth rates than that of interdivision times observed in experiments (Figure 2.4ab)\(^{7,28}\). Eq. 2.38 holds for any size regulation strategy \( 1 > \alpha > 0 \), implying that cell size regulation, as long as it exists (in particular, \( \alpha \neq 0 \) leads instead to Eq. 2.37), does not affect population growth rate within the models studied here (Figure 2.4a).
Figure 2.4: Cell size regulation, as long as it exists, does not affect population growth rate. (a) Population growth rates obtained from simulations (symbols) of an exponentially growing population subject to variability in single-cell growth rates agree with Eq. 2.38 (dashed line). Inset shows population growth rates do not vary with the regulation strength as long as $1 > \alpha > 0$. (b) Variability in single-cell growth rates decreases population growth rate. Blue squares show data (details can be found in Ref. 29). Red dashed line shows Eq. 2.38 for $CV(\lambda)$ measured by experiments. Black solid line shows $\Lambda = \langle \lambda \rangle$ as a guide. Both axes have units $\text{min}^{-1}$. Adapted from Ref. 29.

2.4 Discussion

In this chapter, we summarized the mathematical formulations of the problem of cell size regulation, with a focus on coarse-grained, discrete models. As an example, we showed that a first order autoregressive model can describe the statistics of cell size in *E. coli*. We discussed how detailed analyses of such models led to several biologically relevant insights at the molecular, single-cell, and population level. The same approach may shed light on several outstanding questions.

First, the prevalence of adder correlations in all three domains of life (e.g. the prokaryote *E. coli*<sup>6,7,28</sup>, the eukaryote *S. cerevisiae*<sup>8</sup>, and the archaeon *H. salinarum*<sup>10</sup>) suggest that it may be simpler to implement or may be evolutionarily advantageous compared to other size regulation strategies. An explanation of the prevalence of adder correlations remains missing,
however, since cell size regulation was found not to affect population growth rate within the class of phenomenological growth models reviewed here\textsuperscript{29}.

In contrast, the mean single-cell growth rate affects cell size regulation, since \textit{E. coli} in slow growth conditions no longer exhibits adder correlations\textsuperscript{28}. This observation may be explained by introducing stochasticity in single-cell growth rates into the initiation-centric model discussed in Section 2.3.1, which can lead to a size regulation strategy that varies with mean growth rates\textsuperscript{28}. Indeed, cell size regulation may potentially be an emergent property of cell cycle regulation\textsuperscript{39}. This view is further supported by several models that describe cell division as a downstream effect of another cell cycle event (e.g., initiation of DNA replication in \textit{M. smegmatis}\textsuperscript{9}, the onset of budding in \textit{S. cerevisiae}\textsuperscript{8,14}, and septum constriction in \textit{C. crescentus}\textsuperscript{51}), and nonetheless reproduces the observed statistics at divisions.

Models of cell size regulation may also incorporate diverse growth morphologies. For example in \textit{S. cerevisiae}, division asymmetry depends on the duration of the budded phase, during which all cell growth occurs for the budded daughter cell\textsuperscript{8}. In \textit{M. smegmatis}, cell growth occurs at the two poles of the cell: the old pole grows faster than the new pole, and on average, the daughter that inherits the old pole is larger\textsuperscript{34}. In both cases, the subpopulations formed by the larger and smaller daughter cells exhibit different emergent size regulation strategies\textsuperscript{8,9}. Models incorporating these details move beyond AR models but remain straightforward to simulate numerically, allowing the statistics they generate to be compared to experiments to distinguish between competing models.

At the molecular level, the behavior of molecular network architectures require further analysis. The particular implementation of an initiator accumulation model discussed in Section 2.3.1 made the strong assumption that transcription rate is proportional to cell volume\textsuperscript{49}. It would be interesting to study more detailed network architectures, where this would be a result rather than a model assumption. Models at the molecular level may also begin to investigate
the problem of protein number regulation. Since proteins are made by ribosomes, an important problem is how ribosomes are allocated towards translating different types of proteins. Quantitative patterns at the bulk level have emerged regarding ribosome allocation, but the picture at the dynamical, cell cycle level is less clear. Models of stochastic gene expression that extend existing ones, which often consider a fixed cell volume, to incorporate cell cycle regulation could shed light in this aspect.

Incorporating cell cycle regulation can in turn help understand cell size regulation in organisms with a circadian clock, such as mammalian cells and cyanobacteria. Recent works have begun to examine cell size regulation in these organisms, and some have suggested that the circadian clock may affect cell size regulation in these organisms. How to model these processes at a molecular, coarse-grained, and population level remain intriguing questions.

2.4.1 Which model should we use?

Since the various classes of models of cell size regulation reviewed here are quite different fundamentally, we conclude by discussing the question of which model to use to analyze what data or to elucidate what phenomenon. First, as discussed in Section 2.2.4, continuous rate models (CRMs) differ from discrete stochastic maps (DSMs) in that CRMs take as parameter an entire function that describes the instantaneous division rate, whereas DSMs take, for example, only the strength of regulation and the magnitude of the coarse-grained stochasticity (although the form of the stochasticity must be assumed). Moreover, CRMs assume a priori whether regulation depends only on the current size or also on the size at birth, whereas DSMs can be used to determine the mode of regulation. These could be reasons for the increasing visibility of DSMs as models of cell size regulation. We will see an example of the advantages and disadvantages of the two formalisms in Chapter 8.

Another fundamental distinction is the difference between division-centric models and models
that place control at an upstream event. Division-centric models, such as those described by Eq. 2.9, makes the strong assumption that all information relevant for determining division timing is stored in the current size and the size at birth. This does not have to be the case. For example, it is widely accepted that in S. cerevisiae, control occurs over the Start transition and the duration of the budded phase is uncorrelated with size\textsuperscript{8}. As reviewed briefly in Section 2.3.1, Eq. 2.9 can be adapted to place control at various cell cycle events, which may then lead to additional predictions that can illuminate the coupling between different cell cycle events.

The above distinction does not imply that we should always use the most detailed description. To cite Levins, “All models leave out a lot and are in that sense false, incomplete, inadequate. The validation of a model is not that it is ‘true’ but that it generates good testable hypotheses relevant to important problems\textsuperscript{57}.” It is often helpful to sacrifice details not pertinent to the phenomenon under consideration. For example, the initiator accumulation model can be considered to trigger division rather than initiation. Yet the simplified model may still provide mechanistic insights into the statistical properties, as discussed in Section 2.3.1. These mechanistic models are altogether different from the phenomenological DSMs and CRMs, and we explore some of them in Chapters 7-8.

In summary, the appropriate model to use depends not only on the organism or system in question, but also on the phenomena explored within the model. We have sketched here a map of the existing models. Technological advances now enable collection of more accurate and larger data sets. These will likely stimulate further development of models, which in turn will influence experimental directions.
Archaeal cells share common size control with bacteria despite noisier growth and division

3.1 Adapted abstract

Microorganisms exhibit different volumes spanning six orders of magnitude, but a clonal population in a given environment maintains a relatively narrow distribution of individual cell sizes. Recent studies in the eukaryotic budding yeast and several species of bacteria showed that a narrow distribution in cell size can be accomplished by growing a constant size between two cell cycle events, a phenomenon known as an adder. Here, we wondered whether archaean cells also use the adder strategy. To this end, a soft-lithography method of growing archaean
cells to enable quantitative time-lapse imaging at the single-cell level was developed. Using this method, we demonstrated that *Halobacterium salinarum*, a hypersaline adapted archaeal organism, grows exponentially at the single-cell level and maintains a narrow size distribution by adding a constant length between cell division events. We also found that the archaeal cells exhibited greater variability in cell division placement and exponential growth rate across individual cells in a population relative to those observed in *Escherichia coli*. We present a theoretical framework that explains how these larger fluctuations in archaeal cell cycle events contribute to cell size variability and control.

3.2 Introduction

As discussed in the previous Chapter, individual cells within clonal populations generate narrow distributions of cell sizes, with typical coefficients of variation (CV, the standard deviation divided by the mean) of 0.10-0.20. Three prominent models provide a phenomenological explanation of the narrow distributions (Section 3.4): (1) ‘sizer’, in which cells grow to a threshold size at division; (2) ‘timer’, in which cells grow for a constant amount of time from birth to division; and (3) ‘adder’, in which cells add a constant size from birth to division. Recent studies have demonstrated that the adder model explains cell size control in several different species of bacteria, including *Escherichia coli*, *Bacillus subtilis*, *Pseudomonas aeruginosa*, and *Desulfovibrio vulgaris*. It was shown that diploid daughter cells of the budding yeast *Saccharomyces cerevisiae* follow the adder model. Together, these studies suggest that several evolutionarily divergent organisms have evolved the adder strategy, despite distinct underlying molecular mechanisms. However, control of cell size in archaea, the third domain of life, remains unknown.

In this chapter, we address cell size control in archaea. Archaeal cells resemble bacterial
cells in terms of size and shape, but possess cell cycles with features that are hybrid between eukaryotes and prokaryotes\textsuperscript{61,62}. Development of a phenomenological model of the cell cycle for this domain might therefore inform the intimate connection between cell size regulation and cell cycle regulation. To this end, microfluidic devices were specifically developed for this publication to enable single-cell observations of growth and division under the extreme conditions that support archaeal growth. The details of the device can be found in the published article\textsuperscript{10}.

We studied the hypersaline adapted model archaeon \textit{Halobacterium salinarum}. The rod shape of \textit{H. salinarum} makes it possible to use length and two-dimensional area as a proxy for volume. Using the microfluidic device developed for this publication, we tracked hundreds of cells with high temporal and spatial resolution for up to four generations and up to eight total cell divisions per agarose chamber (Figure 3.1a). Cells in agarose chambers maintained grew at a rate similar to cells in batch culture. We also found that there were no systematic changes in the cell cycle that depend on the generations when grown in the chambers\textsuperscript{10}. Furthermore, we verified that there was no variation between the independent trials in the chambers\textsuperscript{10}. Lastly, we found the diameter CV to be small (0.12), which allowed us to use the length as a proxy for cell volume. While we report our main findings here using measurements of cell length, duplicate analysis using the 2D area data provides identical conclusions\textsuperscript{10}.

3.3 Results

The growth mode of a bacterium is an important determinant of its cell size regulation. We first investigated whether cells grow at a constant rate (linear) or at a rate that is proportional to their size (exponential). One method to determine exponential growth is to measure and plot the size of individual cells over time, and then fit an exponential function to hundreds of individual plots of single-cell growth\textsuperscript{21}. This approach requires measurements of higher
**Figure 3.1**: *H. salinarum* cells grow exponentially, and their lengths at birth and division are narrowly distributed.  
(a) Montage of a typical time-lapse taken using microchambers (video taken every 5 min; not all frames are shown). This montage represents three experiments. Scale bar, 5 µm.  
(b) Single *H. salinarum* cells grow exponentially. We pooled data from three independent trials. The natural log of the ratio of $l_{\text{birth}}$ and $l_{\text{division}}$ was plotted against $t_d$ normalized by $\lambda$ (0.117 h$^{-1}$), which was independently measured by bulk culture growth measurements rather than being a fitted parameter. Unless otherwise noted, experimental data acquired using microchambers are visualized as combinations of scatter plots and probability density heatmaps. The color bar inset shows the densities for the heatmap. Data points after binning along the x-axis are overlaid: the mean and standard deviation of each bin are plotted as pink circles and error bars, respectively. Linear regression of raw data (red solid line) yields a slope of 1.018 ± 0.007 (s.e.; $n = 418$, $P < 0.001$ for the null hypothesis that the slope is zero). A black dashed line with a slope of 1 is overlaid for comparison.  
(c) Histogram of *H. salinarum* cell lengths show narrow distributions (pooled from three experiments) around the mean of 3.5 µm (CV = 0.16) and 6.5µm (CV = 0.13) for birth and division, respectively.
precision than enabled by our methods. We instead considered the correlations between cell
growth and cell cycle timing to determine the growth mode (Figure 3.1, and below for details).
The approach uses statistics to overcome the need for high precision measurements.

For an exponentially growing single cell, the time taken between divisions (that is, the inter-
division time, \( t_d \)) and birth size (\( l_{\text{birth}} \)) are related to the division size (\( l_{\text{division}} \)) by

\[
l_{\text{division}} = l_{\text{birth}} e^{\lambda t_d}
\]

(3.1)

where \( \lambda \) is the exponential growth rate. Rearranging Eq. 3.1, we obtain

\[
\ln \left( \frac{l_{\text{division}}}{l_{\text{birth}}} \right) = \lambda t_d,
\]

(3.2)

yielding a linear relationship with a slope of 1 and an intercept of 0. Fitting the data for cell
length with Eq. 3.2 representing exponential growth (Figure 3.1b), we detected a slope of
1.018 \( \pm \) 0.007 (standard error, or s.e., \( R^2 = 0.978 \)). Similarly, fitting the 2D area data gave
a slope of 0.992 \( \pm \) 0.008\(^10\). Conversely, we found that a linear growth model does not fit the
data well (Figure 3.9b and Section 3.4). Overall, these results demonstrated that single \( H. \)
salinarum cells grow exponentially.

We then investigated the cell size control strategy of \( H. \) salinarum. We found that cell length
distributions at birth and division have a CV of 0.16 and 0.13, respectively (Figure 3.1c).
Linear regression of the cell length data (Figure 3.2a) yielded a slope of 0.88 \( \pm \) 0.06 (s.e.), and
regression of the 2D area data gave a slope of 0.98 \( \pm \) 0.07\(^10\). Furthermore, we found that birth
length and interdivision time are negatively correlated (Figure 3.2b). In fact, the adder model
predicts this negative relationship between birth size and interdivision time (Figure 3.2b). We
also found that the average added size between birth and division is approximately independent
Figure 3.2: *H. salinarum* cells effectively add a constant length between generations, consistent with the adder model. (a) Cell lengths at birth and division are linearly related, and linear regression of the raw data (red solid line) yields a slope of $0.88 \pm 0.06$ (s.e.). A black dashed line with a slope of 1 represents the theoretical line for the adder model. Grey dashed lines with a slope of 2 and 0 are shown to represent the theoretical lines for the timer and sizer models, respectively. (b) Birth length and interdivision time are negatively correlated, and the binned data closely match the theoretical prediction for the adder model, which plots $t_d = \ln \left(1 + \Delta/l_{\text{birth}}\right)/\lambda$ (black dashed line) without adjustable parameters. Theoretical lines for the timer and sizer models are shown as grey dashed lines. The line for the timer model has a slope of 0. For the sizer model, $t_d = \ln \left(\langle l_{\text{division}} \rangle / l_{\text{birth}}\right)/\lambda$ was plotted. (c), A constant length is added regardless of the birth length. Pearson correlation coefficient of birth length and added length was $-0.1 \pm 0.1$. The small magnitude of the coefficient indicates that there were no significant relationships between these parameters. As expected, the binned data closely match the theoretical prediction for the adder model (black dashed line). Theoretical lines for the timer and sizer models are shown as grey dashed lines. The line for the timer model has a slope of 1, assuming that the cell doubles in size to keep a constant doubling time. For the sizer model, $\Delta = \langle l_{\text{division}} \rangle - l_{\text{birth}}$ was plotted. For all panels, data are visualized as combinations of scatter plots and probability density heatmaps. The color bar inset shows the densities for the heatmap. Data points after binning along the x-axis are overlaid: the mean and standard deviation of each bin are plotted as pink circles and error bars, respectively. We pooled data from three independent trials.

of the birth size (Figure 3.2c), further supporting the adder model. These results suggest that the adder model best describes cell size control in *H. salinarum*.

While the adder model best describes the *H. salinarum* data (Figure 3.2), the data also revealed important biological fluctuations in the archaeal cell cycle that affect size control (Figure 3.3a). In particular, we found that the cell division placement is noisier in *H. salinarum* cells than in *E. coli*. That is, the standard deviations of division ratio distributions are 0.03 and 0.01, for *H. salinarum* and *E. coli*, respectively (Figure 3.3bc). The distribution of exponential
growth rates of individual *H. salinarum* cells was also broader than that of *E. coli* cells (Figure 3.3d). The CV of the exponential growth rate distribution was 0.17 for *H. salinarum*, whereas the CV for *E. coli* is reported to be as low as 0.08 in fast growth environments. Together, these data demonstrate that the relative magnitudes of the noise in cell division ratio and growth rate in *H. salinarum* are 2-3-fold larger than those in *E. coli*.

To determine how noisy division site placements and growth rates affect cell size control, we computed Pearson correlation coefficients between different cell cycle parameters (Table 3.1). Previously, a theoretical framework for cell size control was developed assuming that cells divide symmetrically without noise in division site placements and grow at a noiseless exponential rate \(^5\). The framework accounted only for stochasticity in interdivision time. Within this framework, the adder model is equivalent to a value of 0.5 for the correlations between \(l_{\text{birth}}\) and \(l_{\text{division}}\), which is the case for *E. coli*\(^6,7\). Thus, for *E. coli*, the variances in the growth rate and division ratio are small enough that the simple framework requiring only noise in interdivision time was sufficient to capture experimentally observed correlations. By contrast, the correlations between \(l_{\text{birth}}\) and \(l_{\text{division}}\) and other correlations for *H. salinarum* were found to deviate from the predictions of the simple model (Table 3.1), indicating that the previously proposed framework is insufficient to account for the effects of the larger noise sources in *H. salinarum*.

Thus, we developed a theoretical framework (Section 3.4) that includes two additional sources of stochasticity, namely, the noise in division ratio and the noise in exponential growth rate (Figure 3.4a). Within the new framework, the division site placement deviates from one half by a noise term with magnitude \(\sigma_{\text{ratio}}\), and the growth rate of each cell deviates from the mean growth rate by a noise term with magnitude \(\sigma_{\lambda}\) (Figure 3.3). The third and last noise variable in the model describes the stochasticity in the interdivision time in addition to those generated by the other sources of noise. That is, the actual interdivision time of a given cell is modeled
Figure 3.3: Distributions of division ratio and exponential growth rate of *H. salinarum* cells are broader than those of *E. coli*. (a) Noisy symmetric division of mother cells often leads to two daughter cells of different lengths. An example montage of division events is shown. Longer daughter cells are indicated by orange arrows. Scale bar, 5 \( \mu m \). (b) Schematic explaining how noisy divisions were quantified using the division ratio. A perfectly symmetric division would give a ratio of 0.5. (c,d) Histograms depict the division ratio (c) and exponential growth rate (d) distributions of *H. salinarum* cells (blue; pooled from three trials) and *E. coli* cells (red; data from Ref. 63). (c) Gaussian fit of each distribution gave standard deviations of 0.03 and 0.01, for *H. salinarum* (light blue dashed line) and *E. coli* (orange dashed line), respectively. (d) Gaussian fit of each distribution gave standard deviations of 0.16 and 0.08, for *H. salinarum* (light blue dashed line) and *E. coli* (orange dashed line), respectively.
as the sum of a target interdivision time \( t_{\text{target}} \) necessary to precisely add a constant volume and a noise term with magnitude \( \sigma_{\text{time}} \). In theoretical frameworks published previously, the noise in interdivision time was the only noise term that was required to capture experimental data for \( E. \ coli \).

Simulated correlation coefficients based on our framework matched well with experimental observations (Figure 3.4b, Table 3.1, Section 3.4). The dramatically improved agreement relative to the previous framework suggests that our extended framework can be used to describe other species with similarly large noise sources. In addition, this result indicates that noise in division ratio, interdivision time, and exponential growth rate appreciably influence the adder model and cell size distributions. Importantly, the correlation between the birth size of mother and the birth size of daughter cells was the only relationship unaffected by the presence of any of the noise terms (Section 3.4).

To further examine the validity of our framework, we tested the prediction that the standard deviations of cell size distributions are equivalent to the standard deviation in interdivision time after appropriate scaling (Section 3.4). Indeed, these distributions for \( H. \ salinarum \) data collapsed after accounting for \( \sigma_{\text{time}}, \sigma_{\text{ratio}}, \) and \( \sigma_{\lambda} \) (Figure 3.4c). This prediction is in contrast to the experimental and theoretical evidence for \( E. \ coli \) cells, whose interdivision and cell size distributions could be scaled using \( \sigma_{\text{time}} \) alone. Thus, our coarse-grained framework showed that the adder model could accommodate several different biological fluctuations to maintain cell size homogeneity. This framework could be used as a diagnostic tool not only to recognize the implementation of the adder model in other organisms but also to help understand which noise terms prominently affect the correlations between different cell cycle parameters for the adder model.

In conclusion, our study demonstrates that an archaeal organism controls its cell size by adding a constant length between two cell cycle events, but with higher variance in cell cycle
parameters relative to bacteria. This finding expands the list of organisms that implement the adder model for size control across all domains of life, despite the differences in terms of cell shape, cell wall, membrane, DNA replication, transcription, translation, cell division and cell cycle control. It remains to be determined which molecular players implement the adder strategy in \textit{H. salinarum}. Several studies in \textit{E. coli} have suggested that cell cycle control occurs at the initiation of DNA replication rather than at cell division, which was recently shown to be consistent with the adder model. It is possible that initiation events are also important in \textit{H. salinarum}. Given the unique evolutionary position of archaea blending characteristics of both bacteria and eukaryotes, our results serve as a useful foundation for understanding the molecular mechanisms and evolution of cell cycle control.

3.4 Methods

We developed an analytical framework for the adder model with noisy symmetric divisions and noisy growth rates. The framework postulated that cells attempt to add a constant length from birth to division

\[ l_{\text{division}} = l_{\text{birth}} + \Delta. \]  

(3.3)

The framework considered three sources of stochasticity: a time-additive noise in the time to division with standard deviation, \( \sigma_{\text{time}} \tau \), noisy symmetric divisions with standard deviation \( \sigma_{\text{ratio}} \), and noisy growth rates with standard deviation, \( \sigma_{\lambda} \log 2/\tau \), where \( \tau \) is the average doubling time. We analytically obtained approximate distributions given by the model to the lowest order in the noise variables (below). In particular, the CVs of cell size at birth, cell size at
Figure 3.4: Noise in interdivision time, division placement and exponential growth rate significantly affect the archaeal cell size distribution. (a) Schematic illustrating our theoretical framework. The cell growth cycle for two generations is depicted, with time on the x-axis and the log of cell length at any given time on the y-axis. The total interdivision time is the sum of an ideal interdivision time that is necessary to precisely add a constant length plus some time noise. Noise terms in the division ratio, exponential growth rate, and interdivision time have magnitudes of $\sigma_{\text{ratio}}$, $\sigma_{\lambda}$, and $\sigma_{\text{time}}$, respectively. Red double-headed arrows indicate how each noise term leads to individual cells having different division ratio, exponential growth rate and interdivision time. We have drawn a rod-shaped cell and its division placement to illustrate how the noise in the division ratio affects how noisy the cell divisions are. We assumed that every dividing cell has its division ratio randomly drawn from a distribution like the one in Figure 3.3c. Because of the fluctuations in division placement, the CV of birth length is larger than that of the division length (Figure 3.1c). We also show two cells with different growth rates (that is, the slope from $l_{\text{birth}}$ to $l_{\text{division}}$ is steeper for the cell in generation 2), which contribute to different interdivision times. (b) We used correlation coefficients as a sensitive metric to evaluate our theoretical framework. Different pairs of interdivision time and cell length parameters used for correlation calculations are listed on the y-axis. Subscripts b and d indicate birth and division, respectively. Superscripts m, d and s indicate generational relationships: mother, daughter and sister, respectively. Correlation coefficients from stochastic simulations (in black squares) incorporating all three noise terms match well with the values derived from experimental data ($n = 418$, after pooling from three trials; error bars indicate 95% CI). Coefficients from a simulation that took only $\sigma_{\text{time}}$ into account are shown in grey circles for comparison. The exact values for simulation and experimental correlation coefficients are tabulated in Table 3.1. (c) Length distributions can be collapsed with the generation time distribution after accounting for all three noise terms, as predicted by the theory. The $t_d$ distribution was fitted using a Gaussian fit.
division and interdivision time were given by:

\[ \sigma_x^2 = \frac{4}{3} \left( \sigma_{time}^2 + \sigma_{\text{ratio}}^2 \right) \]  
\[ \sigma_y^2 = \frac{4}{3} \sigma_{time}^2 + \frac{1}{3} \sigma_{\text{ratio}}^2 \]  
\[ \sigma_{td}^2 = \frac{4}{3} \sigma_{time}^2 + \frac{1}{3} \sigma_{\text{ratio}}^2 + \sigma_\lambda^2 \]

where \( \sigma_{\text{ratio}} = \frac{\sigma_{\text{ratio}}}{2/\ln 2} \), \( x = \log_2 \left( \frac{l_{\text{birth}}}{\Delta} \right) \), and \( y = \log_2 \left( \frac{l_{\text{division}}}{(2\Delta)} \right) \); and \( \sigma_x \) and \( \sigma_y \) are the standard deviations of \( x \) and \( y \), respectively. Pearson correlation coefficients between different cell cycle variables were also analytically approximated (below). We extracted \( \sigma_{\text{ratio}} \), \( \sigma_x \), \( \sigma_y \), and \( \sigma_{td} \). We used the CV of cell length at division, together with \( \sigma_{\text{ratio}} \) to extract \( \sigma_{time} \). Doing so produced a consistent value for this noise term. We extracted \( \sigma_\lambda \) with the measured CV of generation times.

We used numerical simulations of the stochastic model above to find \( H. \text{salinarum} \) cell length distributions and their CV. Simulations tracked \( l_{\text{birth}}, l_{\text{division}}, \) and \( t_d \) of a growing and dividing cell for 10,000 generations. The starting cell had an initial cell length of 1, but the initial condition did not matter because the simulation reaches and maintains cell size homeostasis. A simulated cell divided after growing for a time, \( t_d = \ln \left( 1 + \Delta/l_{\text{birth}} \right) / \lambda + \sigma_{time} \xi \), where \( \xi \) is a normally distributed noise term with zero mean and unit variance. At division, the cell divided into two with a division ratio, \( 1/2 + \sigma_{\text{ratio}} \xi \). Note that the noise terms are assumed to be independent for different processes, an assumption whose validity was checked against data (below). Only one cell was kept for the next generation, which grew with a rate, \( \ln 2/\tau + \sigma_\lambda \xi \).

We set \( \Delta = \tau = 1 \) because we were interested in only dimensionless quantities, such as CVs and correlation coefficients that do not depend on the absolute mean values.
Figure 3.5: Analysis of 2D cell area data also supports exponential growth of single cells and implementation of the adder model for cell size control for *H. salinarum*. Figure legends are the same as that in Figure 3.2, but with 2D area data rather than cell length data.
Table 3.1: Correlation coefficients between various cell cycle parameters of *H. salinarum*. Pearson correlation coefficients and their 95% confidence intervals are shown for the experimental data. For the correlations from simulated data, we present and compare simulation results using only the time-additive noise term and using all three noise terms described in Section 3.3. As the Table shows, the simulation with all three terms produced correlations that better matched the experimental results.

<table>
<thead>
<tr>
<th>Correlation coefficient</th>
<th>Measured (cell length)</th>
<th>Measured (2D area)</th>
<th>Simulated with three noise terms</th>
<th>Simulated with one noise term</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l_b, l_d$</td>
<td>0.59 ± 0.06</td>
<td>0.58 ± 0.06</td>
<td>0.55</td>
<td>0.47</td>
</tr>
<tr>
<td>$l_b, t_d$</td>
<td>-0.47 ± 0.07</td>
<td>-0.41 ± 0.08</td>
<td>-0.47</td>
<td>-0.47</td>
</tr>
<tr>
<td>$l_d, t_d$</td>
<td>0.19 ± 0.09</td>
<td>0.32 ± 0.08</td>
<td>0.30</td>
<td>0.47</td>
</tr>
<tr>
<td>$t_b^{mother}, t_b^{daughter}$</td>
<td>0.52 ± 0.05</td>
<td>0.53 ± 0.04</td>
<td>0.51</td>
<td>0.45</td>
</tr>
<tr>
<td>$t_d^{mother}, t_d^{daughter}$</td>
<td>-0.1 ± 0.1</td>
<td>-0.1 ± 0.1</td>
<td>-0.15</td>
<td>-0.25</td>
</tr>
<tr>
<td>$l_b, t_sister$</td>
<td>0.5 ± 0.1</td>
<td>0.5 ± 0.1</td>
<td>0.66</td>
<td>0.91</td>
</tr>
</tbody>
</table>

3.4.1 Extended introduction of sizer, timer, and adder

Here, we compare the sizer, timer, and adder models for an extended introduction. These are three phenomenological models widely considered in literature. Each model describes when an exponentially growing cell divides, barring the noise, which is then added explicitly into the model. In the sizer model, a cell divides upon reaching a critical size. In the timer model, a cell divides after growing for a constant time. In the adder model, a cell divides after accumulating a constant increment in size.

Previously, Ref. 5 developed a general framework that describes all three models succinctly. In the framework, the target size at division $f(v_b)$ is a function of the size at birth. An analytically tractable model for $f(v_b)$ is

$$f(v_b) = 2v_b^{1-\alpha}v_0^{\alpha},$$  \hspace{1cm} (3.7)

where $v_0$ is the typical cell size, and $\alpha$ is a parameter that captures the strength of regulation as follows. Since measured cell size distributions are stationary and narrow, we can approximate
the behavior of $f(v_b)$ by expanding around $v_0$. Taking the derivative of Eq. 3.7 and evaluating at $v_b = v_0$, we find

$$f'(v_b) = 2(1 - \alpha).$$

(3.8)

Eq. 3.8 highlights two extreme strategies for cell size control: when $\alpha = 0$, the linear slope $f'(v_b) = 2$; and when $\alpha = 1$, the slope becomes zero. The former case captures the timer model, in which cells attempt to double in size before dividing by growing for a constant time. The latter captures the sizer model, in which cells attempt to divide upon reaching a critical size $2v_0$. In between the two extremes, when $\alpha = 1/2$, Eq. 3.7 describes the adder model, in which $f(v_b) \approx v_b + v_0$.

3.4.2 Incorporating additional sources of stochasticity in cell size control

To Eq. 3.7, Ref. 5 incorporated one coarse-grained source of stochasticity in cell cycle timing and derived the resulting cell size distributions. However, measurements of *H. salinarum* show that there are also significant fluctuations in division ratios and growth rates (Figure 3.3). Below, we incorporate these additional sources of stochasticity into the model and derive the resulting distributions and correlations. As explained in Section 3.3, we have used cell lengths $l_{\text{birth}}$ and $l_{\text{division}}$ as measures of size. Here, we will denote $l_{\text{birth}} \equiv l_b$ and $l_{\text{division}} = l_d$, as well as $\sigma_{\text{time}} \equiv \sigma_t$ and $\sigma_{\text{ratio}} \equiv \sigma_r$ for notational simplicity.

Explained in brief in Section 3.3, we consider the same model as above, but normalize cell size and time variables $\langle l_b \rangle = 1$ and $\langle t_d \rangle = 1$ since we are interested in dimensionless quantities such as the coefficient of variation (CV, the standard deviation divided by the mean) and the Pearson correlation coefficient (CC). Here, $l_b$ is the size at birth, and $t_d$ is the corresponding generation time. Eq. 3.7 becomes

$$f(l_b) = 2l_b^{1-\alpha}.$$ 

(3.9)
To Eq. 3.9, we add three sources of stochasticity: generation times, division ratios, and growth rates. Now, the size at division is \( f(l_b) \) multiplied by the exponential of the noisy growth rate times the time-additive noise,

\[
l_d = f(l_b) 2^{(1 + \xi_\lambda / \ln 2)} \xi_t. \tag{3.10}
\]

The size at birth in the next generation is the size at division multiplied by the noisy division ratio,

\[
l_b' = l_d \left( \frac{1}{2} \xi + \xi r \right), \tag{3.11}
\]

where \( \xi_t, \xi_r \), and \( \xi_\lambda \) are random variables representing stochasticities in generation times, division ratios, and growth rates. They are well approximated as normal random variables with zero mean and small standard deviations \( \sigma_t, \sigma_r, \) and \( \sigma_\lambda \), respectively.

We first calculate the CV of the random size and time variables in our model. Let \( x = \log_2 (l_b) \). We find from Eq. 3.11,

\[
x' = (1 - \alpha) x + \xi_t + \xi_r \xi_\lambda / \ln 2 + \log_2 (1 + 2 \xi_r). \tag{3.12}
\]

Since \( l_b \approx 1 \) by normalization, \( x \) is small. The three noise terms are also small by assumption, so we can approximate Eq. 3.12 to first order in small parameters as

\[
x' \approx (1 - \alpha) x + \xi_t + \frac{2}{\ln 2} \xi_r. \tag{3.13}
\]

Since the variance of the sum of independent random variables is the sum of the variances, we find from Eq. 3.13 that the variance of \( x \) is

\[
\sigma_x^2 = \sigma_t^2 + \left( \frac{2}{\ln 2} \sigma_r \right)^2. \tag{3.14}
\]
For cell size at division, we find from Eq. 3.10,

\[ y \approx (1 - \alpha) x + \xi_t, \]  

(3.15)

for \( y = \log_2 (l_d/2) \). The variance in this case is

\[ \sigma_y^2 = (1 - \alpha)^2 \sigma_x^2 + \sigma_t^2. \]  

(3.16)

Similarly, the time to division can be written as, from Eq. 3.13 and exponential growth,

\[ t_d \approx 1 - \alpha x - \xi \lambda / \ln 2 + \xi_t. \]  

(3.17)

Therefore, the variance is

\[ \sigma_{t_d}^2 = \sigma_t^2 + \sigma_t^2 / \ln^2 2 + \alpha^2 \sigma_x^2. \]  

(3.18)

Eqs. 3.14, 3.16, and 3.18 for the adder model (\( \alpha = 1/2 \)) are cited in Section 3.3 and used in numerical simulations to obtain Figure 2.4.

3.4.3 Calculation of correlation coefficients

Having calculated the CVs, we now calculate the CCs between size and time variables. Note that since \( \nu_b \approx 1 \) by normalization, we can write \( x = \log_2 (l_b) = \log_2 (1 + \delta l_b) \), where \( \delta l_b = l_b - 1 \) is small. Therefore, \( x \approx \delta l_b / \ln 2 = (l_b - 1) / \ln 2 \). However, averages and correlations are not affected by addition or multiplication by a constant. Hence, we can replace \( \nu_b \) by \( x \) in our calculations.
The CC between size at birth and size at division is defined as

\[ C(l_b, l_d) = C(x, y) = \frac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\sigma_x \sigma_y}. \]  

(3.19)

From Eq. 3.15, we find that

\[ \langle xy \rangle = (1 - \alpha) \langle x^2 \rangle. \]  

(3.20)

We have already calculated \( \sigma_x \) (Eq. 3.14) and \( \sigma_y \) (Eq. 3.16). We also know that \( \langle x \rangle = \langle y \rangle = 0 \).

So combining the expressions and simplifying, we find the CC to be

\[ C(l_b, l_d) = \frac{1}{\sqrt{1 + \frac{\sigma_t^2}{(1-\alpha)\sigma_r^2}}}. \]  

(3.21)

For the adder model \( \alpha = 1/2 \),

\[ C(l_b, l_d) = \frac{1}{\sqrt{1 + 3 \frac{\sigma_t^2}{\sigma_i^2 + \left( \frac{2}{m} \sigma_r \right)^2}}} \]  

(3.22)

From Eq. 3.22, if the noise in generation time is much larger than that in division ratio \( (\sigma_t \gg \sigma_r) \) then \( C(l_b, l_d) \approx 1/2 \). This is the case in \textit{E. coli}, and the CC there is indeed close to half. However, the noise in division ratio is larger in \textit{H. salinarum} than in \textit{E. coli}. If the noise in division ratio is much larger instead \( (\sigma_r \gg \sigma_t) \) then \( C(l_b, l_d) \approx 1 \). In other words, noisy divisions increase the CC between size at birth and size at division. We find that the CC in \textit{H. salinarum} is indeed larger than half (Figure 3.4).

The CC between size at birth and time to division is also different between \textit{E. coli} and \textit{H. salinarum}. 

53
Following the above procedure, we find that it is

\[ C(l_b, t_d) = \frac{\langle xt_d \rangle}{\sigma_x \sigma_{t_d}} = -\frac{\alpha \sigma_x}{\sigma_{t_d}}. \]  

(3.23)

Inserting the expressions for \( \sigma_x \) (Eq. 3.14) and \( \sigma_{t_d} \) (Eq. 3.18), and setting for the adder model \( \alpha = 1/2 \), we find

\[ C(l_b, t_d) = -\frac{1}{2} \sqrt{\frac{1 + \frac{4\sigma_y^2}{\ln^2 2\sigma_r^2}}{1 + \frac{3\sigma_y^2}{4\ln^2 2\sigma_r^2} + \frac{\sigma_y^2}{\ln^2 2\sigma_r^2}}}. \]  

(3.24)

Eq. 3.24 again reproduces the results in \( E. coli \): if the noise in generation time is larger than that in other sources - \( \sigma_t \gg \sigma_r, \sigma_{\lambda} \) - then \( C(l_b, t_d) \approx -1/2 \), in agreement with measurements\(^5\).

Eq. 3.24 also shows that noisy divisions exaggerate the negative correlation between \( v_b \) and \( t_d \), whereas noisy growth rates abrogate the correlation. In \( H. salinarum \), Eq. 3.24 is able to explain the measured \( C(l_b, t_d) \approx -0.47 \) given the measured magnitudes of the three sources of stochasticity (Figure 3.4).

In general, CCs are sensitive to the ratios of stochasticities. In the special regimes where some stochasticities are much larger than others, we can take these ratios to be small quantities, expand around them, and compare the resulting sensitivities. For example, consider the CC between sister sizes at birth \( C(l_b, l_{b, \text{sister}}) \). Note that \( l_b = (l_d/2)(1 + 2\xi_r) \) and \( l_{b, \text{sister}} = (l_d/2)(1 - 2\xi_r) \). Following the same procedure as above, we find that

\[ C(l_b, l_{b, \text{sister}}) = \frac{\sigma_y^2 - \left(\frac{2}{\ln 2\sigma_r}\right)^2}{\sigma_x^2}. \]  

(3.25)
In the regime where $\sigma_r \ll \sigma_t$, Eqs. 3.22 and 3.25 can be approximated, respectively, as

\begin{align}
C(l_b, l_d) & \approx \frac{1}{2} + \frac{3\sigma_r^2}{4 \ln^2 2 \sigma_t^2}, \\
C(l_b, l_{\text{sister}}) & \approx 1 - \frac{6\sigma_r^2}{\ln^2 2 \sigma_t^2}.
\end{align}

The CC between sister sizes at birth is eight times more sensitive to noisy divisions than $C(v_b, v_{d})$. This calculation explains the differences between the measured CCs in *E. coli* and *H. salinarum* (Figure 3.4).

Lastly, we can also find the CC between mother size at birth and daughter size at birth, which turns out to be simply

\[ C\left(l_{\text{mother}}^b, l_{\text{daughter}}^b\right) = \frac{\langle x' x \rangle}{\sigma_x^2} = (1 - \alpha). \]

Importantly, $C\left(l_{\text{mother}}^b, l_{\text{daughter}}^b\right)$ is independent of all sources of stochasticity. This CC is thus a robust hallmark of an adder size regulation strategy (Figure 3.4).

Below is a compilation of results derived above for the adder model.
Figure 3.7: Numerical simulations verify analytical derivations. Left and right compare Eqs. 3.33 and 3.35, respectively. Symbols are numerical simulations as described in Section 3.4. Lines are the corresponding analytical expressions. Orange symbols have $\sigma_r = 0$, and color gradient shows different values of $\sigma_\lambda$. (Right) Blue symbols have $\sigma_\lambda = 0$, and color gradient shows different values of $\sigma_r$. Blue symbols are partially obscured by the orange symbols given that $C(v_b^m, v_d^b)$ is independent of $\sigma_\lambda$ and $\sigma_r$.

Variance of log$_2$ size at birth

$$\sigma_b^2 = \frac{4}{3} \sigma_t^2 + \frac{16}{3 \ln^2 2} \sigma_r^2. \quad (3.29)$$

Variance of log$_2$ size at division

$$\sigma_y^2 = \frac{4}{3} \sigma_t^2 + \frac{4}{3 \ln^2 2} \sigma_r^2. \quad (3.30)$$

Variance of generation time

$$\sigma_t^d = \frac{4}{3} \sigma_t^2 + \frac{4}{3 \ln^2 2} \sigma_r^2 + \frac{1}{\ln^2 2} \sigma_\lambda^2. \quad (3.31)$$

CC between size at birth and size at division

$$C(l_b, l_d) = \frac{1}{2} \sqrt{\frac{\sigma_t^2 + \frac{4}{\ln^2 2} \sigma_r^2}{\sigma_t^2 + \frac{1}{\ln^2 2} \sigma_\lambda^2}}. \quad (3.32)$$
CC between size at birth and generation time

\[
C(l_b, t_d) = -\frac{1}{2} \sqrt{\frac{\sigma^2_t + \frac{4}{\ln^2 2} \sigma^2_r}{\sigma^2_t + \frac{3}{4 \ln^2 2} \sigma^2_\lambda + \frac{1}{\ln^2 2} \sigma^2_r}}.
\] (3.33)

CC between sister sizes at birth

\[
C(l_b, l_b^{\text{sister}}) = \frac{1}{4} + \frac{3 \sigma^2_t - \frac{12}{\ln^2 2} \sigma^2_r}{4 \sigma^2_t + \frac{16}{\ln^2 2} \sigma^2_r}.
\] (3.34)

CC between mother size at birth and daughter size at birth

\[
C(l_m \ b, l_d \ b) = \frac{1}{2}.
\] (3.35)

Figure 3.6 and 3.7 verify several of these results with numerical simulations.

3.4.4 Independence of noise terms

We can check that the noise terms are independent in the data by extracting the different noise terms as functions of measured variables,

\[
\xi_r = r - 1/2
\] (3.36)

\[
\xi_t = t_d - \ln (1 + \langle l_b \rangle / l_b) / \lambda
\] (3.37)

\[
\xi_\lambda = \lambda - \langle \lambda \rangle
\] (3.38)

where \( \langle l_b \rangle \) is the measured average cell size at birth, \( \lambda = \ln (l_d / l_b) / t_d \) is the inferred growth rate, and \( \langle \lambda \rangle \) its average. The Pearson correlation coefficient between two noise terms is zero if the noise terms are independent. As shown Figure 3.8, the correlation coefficients between \( \xi_r \) and \( \xi_t \), and that between \( \xi_r \) and \( \xi_\lambda \) are both close to zero. This shows that the division
Figure 3.8: Pearson correlation coefficients between noise terms extracted from measurements. $\xi_r$, $\xi_t$, and $\xi_\lambda$ are the noise terms in the division ratio, interdivision time, and growth rate, respectively.

<table>
<thead>
<tr>
<th>Correlation coefficient</th>
<th>Measured</th>
<th>Measured (division ratio &gt; 0.53)</th>
<th>Adder</th>
<th>Sizer</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l_{\text{division}}, l_{\text{sister}}$</td>
<td>0.25</td>
<td>0.23</td>
<td>0.19</td>
<td>0.00</td>
</tr>
<tr>
<td>$l_{d}, l_{d_{\text{sister}}}$</td>
<td>0.17</td>
<td>0.17</td>
<td>0.14</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Table 3.2: Analysis of sister-sister correlation coefficients. The second column contains the respective measured correlation coefficients indicated in the first column. The third column contains the measured correlation coefficients for sister pairs following an especially asymmetric division (division ratio is more than one standard deviation away from average). The fourth and fifth column contain simulated correlation coefficients of the adder and sizer models, respectively, with the same parameters as those used to generate Figure 3.4.

ratio noise is indeed independent of the other two other noise terms. However, we cannot check the independence of the noise terms in growth rate and interdivision time because we did not measure growth rates directly.

3.4.5 Additional analysis of sister-sister correlations supports the adder model

Two daughter cells show more similar properties to one another compared to two random cells. This can be seen since the correlation coefficient of the size at birth of two random cells is zero, whereas that of sister sizes at birth is non-zero. Consequently, the correlation coefficient of sister sizes at division or generation times will also be non-zero. Simulations with the same set of parameters as in the main text reproduce these sister-sister correlation
Figure 3.9: Linear growth model does not fit the experimental data well. Fitting the experimental data with a forced intercept of zero (the black dotted line; representing the linear growth model) resulted in $R^2 = 0.26$. The fit was improved without forcing the intercept to be zero (the red solid line), yielding $R^2 = 0.39$ and an intercept of $1.1 \pm 0.2$ (95% CI). Data is visualized as combinations of scatter plots and probability density heat maps. The color bar inset shows densities for the heat map. Data points after binning along the x-axis are overlaid: the mean and standard deviation of each bin are plotted as pink circles and error bars, respectively. For this analysis, we pooled data from 3 independent experiments.

coefficients. In addition, sister-sister correlation coefficients demonstrate that the adder growth mode better describes measurements than sizer. This is most clearly seen in the non-zero correlations between sister sizes at division. Lastly, sister cells that are more than one standard deviation away from symmetric division give comparable correlation coefficients as all sister pairs - both in experiments and within the model - indicating that all cells follow the same regulation strategy regardless of the asymmetry of the preceding division. These observations, summarized in the table below, strengthen the distinction between models and further support the adder model.
3.4.6 A linear growth model does not explain *H. salinarum* growth well

We have provided experimental evidence that *H. salinarum* cells grow exponentially at the single-cell level (Figure 3.1). On the other hand, we found that a linear growth model does not fit the data well (Figure 3.9). A linear mode of growth can be described as, \( k_{\text{division}} = k_{\text{birth}} + gt_d \), where \( g \) is the linear growth rate constant. This equation for linear growth can be rearranged to give, \( k_{\text{division}} - k_{\text{birth}} = \Delta = gt_d \). The rearranged equation yields a linear relationship between \( t_d \) and \( \Delta \), with a slope of \( g \) and an intercept of zero. Fitting the experimental data with a forced intercept of zero resulted in a \( R^2 \) of 0.26. The fit was improved without forcing the intercept to be zero, yielding a \( R^2 \) of 0.39 and an intercept of 1.1. Furthermore, the 95\% CI of the intercept did not include zero, demonstrating that the non-zero intercept is statistically significant, and that the linear growth model does not fit the data well.
Simultaneous regulation of cell size and DNA replication in bacteria

4.1 Abstract

Bacteria are able to maintain a narrow distribution of cell sizes by regulating the timing of cell divisions. In rich nutrient conditions, cells divide much faster than their chromosomes replicate. This implies that cells initiate multiple rounds of DNA replication per cell division by regulating the timing of DNA replications. Here, we show that both cell size and DNA replication may be simultaneously regulated by the long-standing initiator accumulation strategy. The strategy proposes that initiators are produced in proportion to the volume increase and is accumulated at each origin of replication, and DNA replication is initiated when a critical amount per origin has accumulated. We show that this model maps to the adder model of size control, which was previously shown to reproduce experimentally observed correlations between various events in
the cell cycle and explains the exponential dependence of cell size on the growth rate of the cell. Furthermore, we show that this model also leads to the efficient regulation of the timing of initiation and the number of origins consistent with existing experimental results.

4.2 Introduction

Bacterial cells are extremely proficient in regulating and coordinating the different processes of the cell cycle. The Cooper-Helmstetter model proposes a molecular mechanism that couples two such processes, the replication of the chromosome and the division of the cell\textsuperscript{12}. In the model, cell division occurs a constant duration after the initiation of DNA replication. The model implies a tight coordination between replication initiation and cell division such that in cells able to double faster than their chromosomes can replicate, multiple rounds of replications proceed simultaneously\textsuperscript{12,64}. To answer how cells regulate the timing of initiation, it was proposed that “replication initiation factors” accumulate to a critical amount per origin of replication to trigger the initiation of replication\textsuperscript{65}. Since the conception of the above model, many experiments and models have attempted to capture the molecular mechanisms responsible for the initiation of multiple rounds of replication. However, no model has been completely satisfactory\textsuperscript{66}.

As a result of the coupling between replication and division, the average cell size per origin is approximately a constant independent of the growth rate of the cell\textsuperscript{19}. Furthermore, it is now understood that a common size regulation strategy for organisms including bacteria and budding yeast is the adder model in which division occurs upon the addition of a constant size dependent on the growth rate of the cell\textsuperscript{5,6,7,8}. However, the molecular mechanisms responsible for the adder model of size control remain in question.

Our main result in this work is to show that the initiator accumulation strategy not only regulates size according to the adder model, but also regulates simultaneously the timing of
initiation and the number of origins of replication. The strategy says that replication initiates upon the accumulation of a critical amount of replication initiation factors per origin. We emphasize the importance of the partitioning of replication initiation factors amongst origins, which we show is essential in order for the multiple replication forks to be adequately regulated. We show, analytically and numerically, that this strategy robustly regulates both cell size and the number of origins. Agreement between existing experiments and predictions of the above model reveals essential features that must be captured in any molecular mechanisms coordinating replication initiation and cell division. Finally, we make distinct predictions regarding the distribution of cell sizes at initiation of replication.

4.3 Methods

4.3.1 Multiple origins accumulation model

We consider the regulation strategy in which replication initiates upon the accumulation of a critical amount of replication initiation factors, or "initiators", per origin of replication\textsuperscript{65}. We assume that the initiators are expressed via an autorepressor model (Figure 4.1)\textsuperscript{22}. In this model, a protein is expressed such that its concentration $c$ remains constant and independent of the growth rate of the cell, which is plausible to achieve through autorepression. Therefore, an increase in the volume of the cell corresponds to a proportional increase in the copy number of this autorepressing protein. A second protein is the initiator and is expressed under the same promoter as the first, but in contrast to the first protein, it is localized at the origins of replication. For simplicity, we assume that the initiators are equally partitioned amongst the origins. Initiation then occurs when a critical copy number per origin $N_{\text{critical}}$ of the localized initiators is reached, after which the initiators are assumed to degrade. Under these assumptions, the copy number of the initiator effectively measures the increase in volume since initiation.
More precisely, if a cell initiated a round of replication at volume \( v_i \) into \( O \) number of origins, the amount of initiators \( N_{\text{initiators}} \) immediately after initiation is zero. To initiate the next round of replication, the cell must accumulate \( ON_{\text{critical}} \) initiators, but because the initiator is expressed under the same promoter as the autorepressor, the cell must also accumulate \( ON_{\text{critical}} \) autorepressors. Because the concentration of the autorepressor is constant, this implies that the cell must accumulate a corresponding volume \( \Delta = N_{\text{critical}}/c \) per origin, independent of the growth rate, to trigger the next initiation.

Thus, on a phenomenological level, the above biophysical model maps to the following regulation strategy for initiation,

\[
v_i^{\text{total,next}} \approx v_i + O\Delta. \quad (4.1)
\]

Eq. 4.1 says that if a cell initiated a round of replication at cell volume \( v_i \) into \( O \) number of origins, then the cell will attempt to initiate another round of replication at total volume \( v_i^{\text{total,next}} \), which is the sum of the volumes of all cells in the lineage since the initiation event at \( v_i \) (typically two cells). This is not to be confused with the threshold model in which cells initiate upon reaching a threshold volume proportional to the number of origins, \( v_i^{\text{next}} \propto O \). For the rest of this manuscript, \( O \) will denote the number of origins after initiation at cell volume \( v_i \) but before initiation at total cell volume \( v_i^{\text{total,next}} \).

We assume an exponential mode of growth for cell volume with a constant doubling time \( \tau \) and a corresponding constant growth rate \( \lambda = \ln 2/\tau \). From Eq. 4.1 and the exponential mode of growth, durations between initiations are

\[
t_i = \frac{1}{\lambda} \ln \left( 1 + \frac{O\Delta}{v_i} \right) + \xi, \quad (4.2)
\]

where \( \xi \) represents some noise in the initiation process. An initiation event will trigger a division event after a constant duration \( C + D \), where \( C \) and \( D \) are respectively the constant
duration required to replicate the chromosome and the constant duration between replication termination and division\textsuperscript{65}. We will refer to Eq. \ref{eq:tau} as the multiple origins accumulation model (i.e. initiators are accumulated per origin). Figure \ref{fig:4.2} illustrates this regulation strategy.

Finally, we will not take into account additional biological mechanisms that act at the level of the initiation of DNA replication, such as oriC sequestration, Dam methylation, and the 'eclipse' phenomenon\textsuperscript{67,68,69}. While these mechanisms are important to prevent rapid re-initiations, by themselves they are insufficient in ensuring an appropriately coordinated coupling between DNA replication and cell division, which is the main focus of our work.

### 4.3.2 Numerical simulations

We can numerically simulate the multiple origins accumulation model given \(C+D, \tau, \Delta\) as experimentally measurable parameters. First, we initialize a population of \(N\) cells with uniformly distributed cell ages. Durations between initiations are calculated as Eq. \ref{eq:tau} and the noise in the initiation process is assumed to be normally distributed with standard deviation \(\sigma_{\tau}\), though the precise nature of the noise does not affect any of our conclusions. It is assumed that in an initiation event, the number of origins in a cell is doubled. The corresponding division event occurs after a constant time \(C+D\). In a division event, the number of origins in a cell, along with the size of the cell, is halved, and two identical cells are generated. We neglect the stochasticity arising from asymmetric divisions, which do not significantly affect any of the results. There are no division events without the corresponding initiation events. Following this procedure,
a population of cells will robustly reach stationarity regardless of initial conditions, as seen in
Figures 4.3 and 4.4.

4.4 Results

4.4.1 Multiple origins accumulation robustly and efficiently regulates the number of
origins of replication

An important measurable consequence of the tight coupling between replication initiation and
cell division is the average number of origins of replication per cell. It has been theoretically
shown that the average number of origins per cell is

\[ \langle O \rangle = 2^{(C+D)/\tau}. \tag{4.3} \]

The derivations leading up to Eq. 4.3, summarized below, hinges on assuming an efficient
process in which the population is growing exponentially and reaches a stationary distribution
of cell ages, implying that there are no delays due to DNA replication. These assumptions
are independent of any molecular mechanisms for initiation and thus should be fulfilled by any
efficient mechanism. We will show that the multiple origins accumulation model reproduces
Eq. 4.3.

The multiple origins accumulation model regulates initiation via negative feedback on the
volume at initiation described by Eqs. 4.1 and 4.2. The feedback enables cells to maintain a
stationary average volume at initiation and a stationary average duration between initiations
despite noise in the initiation process. Specifically, if a cell initiated replication at volume
Figure 4.2: Schematic of the regulation strategy of the multiple origins accumulation model. See text for the details of the model. Slow growth denotes $0 < (C + D) / \tau < 1$. Faster growth denotes $1 < (C + D) / \tau$. In the above example, $(C + D) / \tau < 2$.
\[ v_i = O\Delta, \text{ then the duration to the next initiation event is} \]

\[ t_i(O) \approx \frac{1}{\lambda} \ln \left( 1 + \frac{O\Delta}{O\Delta} \right) = \tau. \tag{4.4} \]

But if a cell initiated replication at a slightly larger volume \( v_i = O\Delta + \delta v \), the duration to the next initiation event is

\[ t'_i(O) \approx \frac{1}{\lambda} \ln \left( 1 + \frac{O\Delta}{O\Delta + \delta v} \right) \approx \tau. \tag{4.5} \]

Eqs. 4.4 and 4.5 say that a cell that initiated at a slightly larger volume than average tend to initiate again faster than average so that its volume at next initiation is again near the average. Similar reasoning says that cells that initiated at slightly smaller volumes tend to initiate again slower than average. In this way, cells maintain a stationary average volume at initiation and a stationary average duration between initiations.

Furthermore, the feedback enables cells to maintain a balanced cell cycle, in which there is on average one and only one initiation event per cell cycle. In the case of negligible noise, a balanced cell cycle implies that cells will initiate at cell age

\[ a_i = 1 + \left\lfloor \frac{C + D}{\tau} \right\rfloor - \frac{C + D}{\tau}, \tag{4.6} \]

where \( a = 0 \) represents cell birth, \( a = 1 \) represents cell division, and \( \lfloor \rfloor \) is the mathematical floor operator (largest integer smaller or equal to the argument). But in the case of realistic noise, a cell may initiate an extra round of replication if the noise is negative enough, \( \xi/\tau \lesssim \left\lfloor (C + D)/\tau \right\rfloor - (C + D)/\tau \), which corresponds to an extra initiation at volume \( v'_i = 2O\Delta - \delta v \).

The multiple origins accumulation model is robust to these stochastic events because a cell that
Figure 4.3: $t_i$, $O$, and $v_b$ approaching stationary distributions in numerical simulations of the multiple origins accumulation model. First, we initialize a population of $N$ cells with uniformly distributed cell ages. Durations between initiations of replication are calculated as Eq. 4.2 and the noise in the initiation process is assumed to be normally distributed. In an initiation event, the number of origins in a cell is doubled. The corresponding division event occurs after a constant time $C+D$. In a division event, the number of origins in a cell, along with the size of the cell, is halved, and two identical cells are generated. There are no division events without the corresponding initiation events. Following this procedure, a population of cells will robustly reach a stationary distribution of cell sizes and number of origins per cell regardless of initial conditions. The plots here track one lineage of cells. Here, $C+D$ is 70 minutes, $\tau = 20$ minutes, and $\sigma_\tau = 4$ minutes. These are biologically realistic choices. We set $\Delta = 1/2^{(C+D)/\tau}$ so that $\langle v_b \rangle \approx 1$. 
initiated an extra round of replication will initiate again after

$$t_i(2O) \approx \frac{1}{\lambda} \ln \left(1 + \frac{2O\Delta}{2O\Delta - \delta v}\right) \approx \tau.$$  \hspace{1cm} (4.7)

In other words, cells with extra rounds of replication will initiate slower than those without so that the stationary average duration between initiations is maintained. The cell cycle following the extra initiation will typically not have any initiations, so that the initiation following the extra initiation will occur at approximately the average volume at initiation. A cell that missed a round of replication will return to a balanced initiation process in the analogous manner. In this way, the multiple origins accumulation model is able to efficiently maintain a balanced cell cycle in fast growth conditions. In contrast, the model simulated by Ref. 6 is not robust to the noise in the initiation process, because in their model, the added volume needed to trigger initiation is not partitioned between origins, which we discuss below.

The multiple origins accumulation model is therefore able to robustly regulate the timing of initiations in face of extra initiations. Extra initiations can occur not only because of noise in the initiation process, but also because of a shift in the growth rate of the cell such as that found in a shift-up experiment, in which a population of cells is abruptly switched from one nutrient condition to a richer nutrient condition allowing for faster growth. The increase in growth rate corresponds to a decrease in the duration between initiations. Therefore, cell in a shift-up experiment will initiate extra rounds of replication in the cycle immediately following the shift-up, but as we have seen, the multiple origins accumulation model is able to appropriately regulate the timing of initiations to reflect the new growth rate. Simulations of the multiple origins accumulation model reached stationary distributions of cell ages, durations between initiations, cell sizes, and number of origins per cell, regardless of initial conditions or the magnitude of the noise $\xi$ in Eq. 4.2, as seen in Figures 4.3 and 4.4. Simulations also show that
Figure 4.4: Stationary exponential distribution of cell ages. Simulations are the same as Figure 4.3. The line plots $p(a) = \ln(2) \cdot 2^{1-a}$.

the number of origins is regulated as in Eq. 4.3, as seen in Figure 4.5. The above considerations show that the multiple origins accumulation model regulates the number of origins robustly and efficiently in face of noise in the initiation process.

4.4.2 Multiple origins accumulation robustly regulates cell size

It was recently shown that the multiple origins accumulation model of replication initiation reduces to the adder model of size regulation

$$v_d \approx v_b + v_0.$$  \hspace{1cm} (4.8)

Eq. 4.8 says that if a cell is born with volume $v_b$, then the cell will attempt to divide at volume $v_d$, where $v_0$ is the constant added volume from birth to division. In fact, $v_0$ can be expressed in terms of known parameters. First, if a cell initiated replication at volume $v_i$, then the corresponding division event will occur at total volume $v_d^{\text{tot}} = v_i 2^{(C+D)/\gamma}$. But there will have been $\log_2 O$ division events since initiation at $v_i$, so that the corresponding volume at birth

71
\[ v_b = \frac{v_{d}^{\text{total}, \text{next}}}{O} = \frac{v_i 2^{(C+D)}/\tau}{O}. \] (4.9)

In a balanced cell cycle, the next initiation event will occur at total volume \( v_{i}^{\text{total}, \text{next}} \approx v_i + O \Delta \) and the corresponding division event will occur at total volume \( v_{d}^{\text{total}, \text{next}} = v_i 2^{(C+D)/\tau}. \)

Similarly, there will have been \( \log_2 O \) division events since initiation at \( v_{i}^{\text{next}} \), so that the corresponding volume at division is

\[ v_d = \frac{v_{d}^{\text{total}, \text{next}}}{O} = \frac{v_i 2^{(C+D)/\tau}}{O} + \Delta 2^{(C+D)/\tau}. \] (4.10)

Therefore

\[ v_d - v_b \approx \Delta 2^{(C+D)/\tau}. \] (4.11)

Within the multiple origins accumulation model, this derivation is valid for any \( C+D \) and \( \tau \).

The adder model of size regulation predicts distributions, correlations, correlation coefficients, and scalings consistent with existing measurements\(^5,6,7,8\). In particular, the average cell volume at birth

\[ \langle v_b \rangle \approx \Delta 2^{(C+D)/\tau}. \] (4.12)

Eq. 4.12 says that the average cell volume at birth is exponentially dependent on the growth rate, a well-known and well-tested result for \textit{E. coli} and \textit{B. subtilis}\(^7,18\). Simulations of the multiple origins accumulation model also confirm this result, as seen in Figure 4.5. Thus, the multiple origins accumulation model robustly regulates cell size.

4.4.3 Master accumulation predictions are inconsistent with existing experiments

Consider the regulation strategy

\[ v_{i}^{\text{total}, \text{next}} \approx v_i + \Delta \] (4.13)
Eq. 4.13 says that if a cell initiated a round of DNA replication at cell volume $v_i$ with $O$ number of origins of replication, then the cell will attempt to initiate another round of replication at total volume $v_i^{\text{total, next}}$, where $\Delta$ is a constant volume independent of the growth rate. This is an adder model of size control applied at initiation. If we assume the same mode of initiator expression as before, then the regulation strategy described by Eq. 4.13 corresponds to replication initiation upon accumulation of a critical amount of initiators without partitioning of initiators between origins. Instead, a plausible molecular picture is that of initiators accumulating at a "master" origin, whose initiation triggers the cascade initiation of other origins$^{71}$. We will therefore refer to Eq. 4.13 as the master accumulation model. In contrast to the multiple origins accumulation model described above, here in the master accumulation model, the total volume at next initiation does not depend on the number of origins present in the cell.

As before, we assume an exponential mode of growth with a constant doubling time $\tau$. The
durations between initiations are therefore

\[ t_i(O) = \frac{1}{\lambda} \ln \left( 1 + \frac{\Delta}{v_i} \right) + \xi, \]  

(4.14)

where \( \xi \) represents some noise in the initiation process. Again as before, an initiation event will trigger a division event after a constant duration \( C+D \). Eq. 4.14 differs from Eq. 4.2 by a missing factor of \( O \). As we show below, the factor of \( O \) is essential in regulating appropriately the timing of initiations.

However, the master accumulation model does not reproduce the well-known exponential scaling of cell size with growth rate. The derivation follows and is similar to that of the multiple origins accumulation model. First, if a cell initiated replication at volume \( v_i \), then the corresponding division event will occur at total volume \( v_{i,\text{total}} = v_i 2^{(C+D)/\tau} \). But there will have been \( \log_2 O \) division events since initiation at \( v_i \), so that the corresponding volume at birth is

\[ v_b = \frac{v_{i,\text{total}}}{O} = \frac{v_i 2^{(C+D)/\tau}}{O}. \]  

(4.15)

In a balanced cell cycle, the next initiation event will occur at total volume \( v_{i,\text{total,next}} \approx v_i + \Delta \) and the corresponding division event will occur at total volume \( v_{i,\text{total,next}}^d = v_{i,\text{total,next}} 2^{(C+D)/\tau} \). Similarly, there will have been \( \log_2 O \) division events since initiation at \( v_{i,\text{next}} \), so that the corresponding volume at division is

\[ v_d = \frac{v_{i,\text{total,next}}}{O} = \frac{v_i 2^{(C+D)/\tau}}{O} + \frac{\Delta 2^{(C+D)/\tau}}{O}. \]  

(4.16)

Therefore,

\[ v_d - v_b = \frac{\Delta 2^{(C+D)/\tau}}{O}. \]  

(4.17)
This derivation is valid for any C+D and $\tau$. However, from Eq. 4.3, $O$ should scale exponentially with the growth rate like $2^{(C+D)/\tau}$ so that

$$\langle v_d \rangle \approx v_d - v_o \sim \Delta.$$  \hspace{1cm} (4.18)

Eq. 4.18 says that the average cell size is a constant roughly independent of the growth rate, a prediction contradicting the well-tested exponential scaling with growth rate for the model organisms mentioned above.

Furthermore, the above reasoning assumes that the master accumulation model can maintain a balanced cell cycle. But the master accumulation model cannot robustly maintain a balanced cell cycle in face of noise as Eq. 4.14 demonstrates. Specifically, if a cell initiated replication at volume $v_i = \Delta$, then the duration to the next initiation event is

$$t_i(O) \approx \frac{1}{\lambda} \ln \left( 1 + \frac{\Delta}{\Delta} \right) = \tau.$$  \hspace{1cm} (4.19)

But the cell may proceed to initiate an extra round of replication if the noise is negative enough, $\xi/\tau \leq \lfloor (C + D)/\tau \rfloor - (C + D)/\tau$, which corresponds to an extra round of initiation at volume $v'_i = 2\Delta - \delta v$. The next initiation will then occur after

$$t_i(2O) \approx \frac{1}{\lambda} \ln \left( 1 + \frac{\Delta}{2\Delta - \delta v} \right) \geq \log_2 \left( \frac{3}{2} \right) \tau,$$  \hspace{1cm} (4.20)

implying that

$$t_i(2O) \leq t_i(O).$$

That is, cells with more origins will initiate faster than those with less, giving rise to cells with average durations between birth and division not equal to $\tau$. The master accumulation
model therefore does not robustly regulate the initiation process to maintain a balanced cell cycle. Indeed, simulations of the master accumulation model do not converge to a balanced cell cycle. Likewise, Ref. 6 carried out simulations of the master accumulation model and obtained "widely abnormal cell size distributions." Given the above inconsistent predictions, the master accumulation model can be ruled out as a possible regulation strategy for replication initiation.

4.4.4 Multiple origins accumulation suggests that variations in C+D are small

In claiming that the master accumulation model gives incorrect correlations between growth rate dependent variables, Ref. 6 simulated the master accumulation model and reported negative correlations between cell size at birth $v_b$ and cell size differences between birth and division $\Delta v$, whereas none is observed experimentally. In contrast to claims in Ref. 6, the negative correlations do not provide evidence against the multiple origins accumulation model nor the master accumulation model. Instead, the negative correlations provide evidence that variability in the durations from initiation to division C+D should be small. Indeed, the multiple origins accumulation model, because of its reduction to the adder model of size control, predicts no correlations between $v_b$ and $\Delta v$, given that variations in C+D are small compared to variations in $\tau$. Simulations assuming that durations from initiation to the corresponding division are normally distributed with mean C+D and standard deviation $\sigma_{C+D}$ show that the correlations between $v_b$ and $\Delta v$ become increasingly negative as $\sigma_{C+D}/\sigma_\tau$ increases (Figure 4.6). Figure 4.6 shows that as long as $\sigma_{C+D}/\sigma_\tau < 0.3$, the correlations between $v_b$ and $\Delta v$ will be close to zero. This is intuitive because when $\sigma_{C+D}$ is small compared to $\sigma_\tau$, fluctuations in cell sizes at birth arise due to variations in cell sizes at initiation, but these variations are negatively fed back into the multiple origins accumulation model as explained above. On the other hand, when $\sigma_{C+D}$ is comparable to $\sigma_\tau$, some fluctuations in cell sizes at birth arise due to variations in durations between initiation and division, but these variations are not accounted for by

76
the multiple origins accumulation model. Variations of this nature give rise to the negative correlations between $v_b$ and $\Delta v$.

4.4.5 Multiple origins accumulation predicts proportionality between cell size and the number of origins per cell

The simultaneous regulation of cell size and the number of origins per cell in the multiple origins accumulation model gives rise to a strict relationship between the two variables. In the multiple origins accumulation model, the average cell volume at birth Eq. 4.12 is exponentially dependent on the growth rate, while the average number of origins per cell Eq. 4.3 also scales exponentially with the growth rate. Therefore we have that

$$\langle v_b \rangle \approx \Delta \langle O \rangle. \quad (4.21)$$

That is, the multiple origins accumulation model predicts that given a fixed volume increment per origin $\Delta$, the average volume at birth and the average number of origins per cell will scale appropriately with respect to a varying $(C+D)/\tau$ to give rise to the above approximate proportionality. The critical size regulation strategy proposed by assumed this proportional-
Figure 4.7: $\langle O \rangle$ against $\langle v_b \rangle$. Simulations are as in Figure 4.3, with a varying $\tau = 20 - 100$ mins, a fixed C+D of 70 minutes, and $\sigma_\tau/\tau = 0.2$. Dashed line plots Eqs. 4.3 and 4.12.

ity\textsuperscript{19}, but is inconsistent with measured correlations in variables in \textit{E. coli}; on the other hand, multiple origins accumulation predicts this proportionality, and is consistent with measured correlations\textsuperscript{5,6,7,8}. Simulations of the multiple origins accumulation model confirm that cell size is indeed approximately proportional to the number of origins per cell (Figure 4.7). We emphasize that the approximate proportionality is a property predicted by the multiple origins accumulation model. In contrast, other strategies that do not regulate the number of origins, such as the master accumulation model, would not predict the approximate proportionality.

4.4.6 Multiple origins accumulation predicts bimodal cell sizes at initiation

In addition to the approximate proportionality between cell size and the number of origins per cell, the multiple origins accumulation model predicts that the distribution of cell sizes at initiation will be approximately bimodal because cells will initiate extra rounds of replication when the noise is large enough. Simulations show that the distribution of cell sizes at initiation is indeed bimodal, with one large peak corresponding to a subpopulation whose cells initiated the
expected number of rounds of replication and a smaller subpopulation whose cells initiated extra rounds of replication (Figure 4.8). Naively, because the distribution of cell sizes is lognormal in the multiple origins accumulation model, the distribution of cell sizes at initiation should be approximately the sum of two lognormal distributions with means $O_0\Delta$ and $2O_0\Delta$, where $O_0 = 2^{\lfloor(C+D)/\tau\rfloor}$, and the ratio between the frequencies of the two peaks equal to the probability that $\xi/\tau \lesssim [(C + D) / \tau] - (C + D) / \tau$. However, the value $O_0\Delta$ overestimates the average sizes at initiation of cells that initiated extra rounds of replication because of correlations between the volumes at initiation and the probability for extra rounds of replication. The correlations arise from Eq. 4.2, which says that a smaller volume at initiation correlates with a larger probability for extra rounds of replication during the current cell cycle. The bimodal distribution of cell sizes at initiation highlights how the multiple origins accumulation model, without invoking other mechanisms, can robustly maintain a balanced cell cycle despite noise in the initiation process and is an experimentally testable prediction of our model.
4.4.7 Multiple origins accumulation predictions are consistent with experiments on mutants

Existing experiments on mutants of *E. coli* and *B. subtilis* produced results consistent with the predictions of the multiple origins accumulation model. Existing experiments have shown by manipulating the cell size of *E. coli* via mutations that a decrease in \( v \) is correlated with a decrease in \( C+D \) and that a decrease in \( v \) is also correlated with a decrease in \( \langle O \rangle \). In the language of the multiple origins accumulation model, \( v \) is controlled by \( \Delta \) and \( (C+D)/\tau \). We assume that the mutations left unchanged the replication initiation mechanism, so that \( \Delta \) is a constant throughout. But \( \tau \) remained approximately constant with changes in \( v \), so a decrease in \( v \) must correspond to a decrease in \( C+D \). One particular measurement reported for *E. coli* cells an average doubling time \( \tau \approx 25 \) minutes for wildtype and mutant cells, \( C+D \approx 40+20 \) minutes, respectively, for wildtype cells, and \( C+D \approx 30+20 \) minutes, respectively, for mutant cells\(^7\). Given these values and Eq. 4.12, the relative change in cell sizes corresponding to the reported difference in the durations from initiation to division should be \( 2^{((C+D)’-(C+D))/\tau} \approx 0.75 \). That is, the mutant cells should be 0.75 times the size of the wildtype cells. This is in excellent agreement with the 25% decrease in volume reported. Moreover, the size at initiation did not change for these smaller mutants, consistent with our model which predicts the size at initiation to depend only on \( \Delta \) and not on \( C+D \). In contrast, in the case of *B. subtilis*, \( C+D \) in smaller mutant cells remained constant, suggesting that \( \Delta \) is the quantity which was changed. Based on this interpretation, our model would predict that both size at initiation and at birth would change proportionally. Indeed, it was found that both the average mutant cell size and the mutant cell size at initiation both decreased by approximately 35%\(^7\). The above experiments also observed the predicted approximate proportionality between cell size and the number of origins per cell in both *E. coli* and *B. subtilis\(^7\). It remains to be shown that cell sizes at
initiation fall into an approximate bimodal distribution. The agreement between experimental results and the predictions made by the multiple origins accumulation model speaks to the importance of regulating simultaneously cell size and the number of origins.

4.4.8 Derivations of the average number of origins per cell

The average number of origins per cell has been calculated previously in two distinct derivations\textsuperscript{12,73}. The two derivations seemed to make different assumptions to arrive at the same conclusions, bringing into question the necessity of the underlying assumptions. Here, we reproduce the two derivations and show that both derivations in fact make the same assumptions.

The model of the cell cycle under consideration is due to Ref.\textsuperscript{12}. In this model, replication initiation occurs on average every doubling time $\tau$. An initiation event then triggers a division event after a constant duration $C+D$, where $C$ and $D$ are respectively the constant duration required to replicate the chromosome and the constant duration between replication termination and division. Given $C$, $D$, and $\tau$, we want to find the average number of origins per cell. The average number of origins per cell is defined as $\langle O \rangle = \langle O_{\text{total}}/N \rangle$, where $O_{\text{total}}$ is the total number of origins in a population of cells, $N$ is the number of cells in that population, and brackets denote the ensemble average.

First, we reproduce the derivation due to Ref.\textsuperscript{12}. To calculate the average number of origins per cell, we must first define the probability distribution underlying the ensemble average. In an asynchronous population of exponentially growing cells, the cells must be exponentially distributed in the cell cycle for ensemble averages to be stationary with respect to time\textsuperscript{49}. Defining cell age $a = 0$ at birth and $a = 1$ at division, the exponential distribution of cell ages is

\begin{equation}
    p(a) = (\ln 2) 2^{1-a}.
\end{equation}
We can now calculate the desired ensemble averages. For example, if $0 < (C + D) / \tau < 1$, then a cell younger than $(\tau - (C + D)) / \tau$ will not be replicating its chromosome and will have only one origin, whereas a cell older than $(\tau - (C + D)) / \tau$ will be replicating its chromosome and will have two origins. We have assumed that the amount of time a cell spends with more than two origins is negligible, which is plausible for weak noise in the initiation process. The average number of origins per cell in an asynchronous, exponentially growing population is then

$$\langle O \rangle = \ln 2 \left[ \left( \int_0^{\tau - (C + D)} \frac{2}{\tau} da \right) + 2 \left( \int_{\tau - (C + D)}^1 \frac{2}{\tau} da \right) \right]$$

$$= \left[ \left( 2 - 2^{(C + D) / \tau} \right) + 2 \left( 2^{(C + D) / \tau} - 1 \right) \right]$$

$$= 2^{(C + D) / \tau}.$$  (4.23)

Similarly if $1 < (C + D) / \tau < 2$, then a cell must initiate replication not only for its daughter cells, but also for its granddaughter cells. In this case, a cell with cell age less than $(2\tau - (C + D)) / \tau$ will not be replicating its chromosome for its granddaughters and will have only two origins, whereas a cell with cell age more than that will be replicating its chromosome for its granddaughters and will have four origins. Again, we have assumed weak noise.

Thus, we see that

$$p (O = O_0) = \ln 2 \int_0^{\Delta T} 2^{1-a} da,$$  (4.24)

$$p (O = 2O_0) = \ln 2 \int_{\Delta T}^1 2^{1-a} da,$$  (4.25)

where $O_0 = 2^{[(C + D) / \tau]}$ and $\Delta T = [(C + D) / \tau] + 1) \tau - (C + D)$. Simplification gives

$$\langle O \rangle = 2^{(C + D) / \tau},$$  (4.26)
which is valid for any C+D and \( \tau \). The two assumptions made in this derivation are that the population is growing exponentially and that the population has reached a stationary distribution of cell ages.

Next, we reproduce the derivation due to Ref. 73. Assuming exponential growth, the number of cells must grow exponentially as \( N \propto 2^{t/\tau} \). Similarly, the total number of origins must grow at the same exponential rate so that \( O_{\text{total}} \propto 2^{t/\tau} \). But an initiation event triggers a division event after a constant duration C+D, so the number of cells must on average lag behind the total number of origins by \( 2^{(C+D)/\tau} \). The average number of origins per cell must then be \( \langle O \rangle = (O_{\text{total}}/N) = 2^{(C+D)/\tau} \). Although the distribution of cell ages was not explicitly involved in this derivation, the assumption of a stationary ensemble average under exponential growth is satisfied if and only if the distribution of cell ages is exponential

The exponential distribution of cell ages is not always realized in experimental setups. For example, single-cell experiments that track a lineage of cells, such as those in7, will follow a different distribution, as discussed in Ref. 74. Experiments that track a single cell will follow a uniformly distributed cell age. In that case, Eq. 4.3 is replaced by

\[
\langle O \rangle = O_0 \left( 1 + \frac{C + D}{\tau} - \left\lfloor \frac{C + D}{\tau} \right\rfloor \right). 
\]

(4.29)

Simulations tracking a population of cells with uniformly distributed cell ages confirm this result. The differences between Eq. 4.3 and the above do not significantly change the predictions of the multiple origins accumulation model.

4.5 Discussion

The multiple origins accumulation model proposes that replication initiates upon the accumulation of a critical amount of initiators per origin. If the initiators are expressed as in the
autorepressor model, this strategy corresponds to Eq. 4.1, which in turn reduces to the adder model of size control, which predicts distributions, correlations, and scalings consistent with existing measurements. Specifically, the average cell size scales exponentially with the growth rate of the cell, Eq. 4.12, as does the average number of origins per cell, Eq. 4.3. The model robustly regulates both cell size and the number of origins per cell such that cell size is approximately proportional to the number of origins per cell. These predictions are consistent with existing experiments on \textit{E. coli} and \textit{B. subtilis}\textsuperscript{72}. A proportionality between ploidy and cell size has also been observed in other organisms, including yeast\textsuperscript{75}. The multiple origins accumulation model is a general regulation strategy that may illuminate the source of the approximate proportionality between cell size and the number of origins across organisms.

An essential feature of the multiple origins accumulation model is the tight coupling between DNA replication and cell division. The differences between the multiple origins accumulation model and the master accumulation model emphasizes this coupling and the importance of regulating the timing of initiation. By negatively regulating cell size in response to the number of origins via Eq. 4.1, the multiple origins accumulation model is able to maintain a balanced cell cycle and achieve robustness in face of noise in the initiation process. However, the master accumulation model described by Eq. 4.13 is a regulation strategy without such a feedback mechanism. The master accumulation model is unable to maintain a balanced cell cycle and does not predict the exponential scaling of cell size. This suggests that regulation strategies must account for the number of origins per cell in order to regulate appropriately the frequency of division.

The coupling between the number of origins to the division frequency could be demonstrated via a shift-up experiment. It was found for \textit{E. coli} that cells maintain their rate of division for a duration of C+D after a shift-up, a phenomenon known as rate maintenance\textsuperscript{76,77}. The multiple origins accumulation model naturally accounts for rate maintenance, because division always
occurs at time $C+D$ after initiation (DNA replication rate is independent of growth rate). Furthermore, the model offers a robust mechanism to regulate, after a transient, the number of origins per cell appropriately with the new growth rate via Eq. 4.3. The existence of a rate maintenance period implies that division is coupled to replication initiation, and the adder model applied at birth and division is a valid phenomenological description only at stationarity. Instead, it is the underlying molecular mechanism of replication initiation that dictates the frequency of division.

Although the multiple origins accumulation model captures many aspects of the coupling between replication and division, experiments with minichromosomes suggest that the molecular mechanism is more complicated. Minichromosomes are plasmids containing the oriC sequence coding for chromosomal origins. In general, minichromosomes initiate replications in coordination with chromosomes and do not affect the growth properties of the cell, such as the doubling time or the average cell size\textsuperscript{78}. However, if more than $\sim 40$ minichromosomes are present in a cell, replication initiation is no longer synchronous, the doubling time increases, the average number of origins per cell decreases, and the average cell size decreases\textsuperscript{79}. Another experiment inserted a second origin into \textit{E. coli} chromosome and observed again that the extra origin does not affect the growth properties of the cell\textsuperscript{80}. These result points to a more complicated molecular mechanism than accumulation of initiators per origin. Several mechanisms have been suggested, but none has been completely satisfactory. For example, the master accumulation mechanism is ruled out for being unable to robustly regulate the number of origins. Another plausible regulation strategy is one in which replication initiates when a critical ratio of active to inactive initiators is reached\textsuperscript{66}. The validity of this strategy remains to be tested.

The molecular mechanism underlying the regulation of replication initiation is yet to be unraveled, but here we have given significant constraints regarding the potential mechanisms. Specifically, this work and previous works have shown that the molecular mechanism in ques-
tion should satisfy both the adder model of size control and the mathematical form of the multiple origins accumulation model described by Eq. 4.1, so that the predicted distributions, correlations, and scalings remain intact and consistent with existing experiments. In Chapter 7, we provide a progress report on this endeavor.
Interrogating the *Escherichia coli* cell cycle by cell dimension perturbations

5.1 Abstract

Bacteria tightly regulate and coordinate the various events in their cell cycles to duplicate themselves accurately and to control their cell sizes. Growth of *Escherichia coli*, in particular, follows a relation known as Schaechter’s growth law. This law says that the average cell volume scales exponentially with growth rate, with a scaling exponent equal to the time from initiation of a round of DNA replication to the cell division at which the corresponding sister chromosomes segregate. Here, we sought to test the robustness of the growth law to systematic perturbations in cell dimensions achieved by varying the expression levels of *mreB* and *ftsZ*. We found that decreasing the *mreB* level resulted in increased cell width, with little change in cell length, whereas decreasing the *ftsZ* level resulted in increased cell length. Furthermore, the time from
replication termination to cell division increased with the perturbed dimension in both cases. Moreover, the growth law remained valid over a range of growth conditions and dimension perturbations. The growth law can be quantitatively interpreted as a consequence of a tight coupling of cell division to replication initiation. Thus, its robustness to perturbations in cell dimensions strongly supports models in which the timing of replication initiation governs that of cell division, and cell volume is the key phenomenological variable governing the timing of replication initiation. These conclusions are discussed in the context of our recently proposed “adder-per-origin” model, in which cells add a constant volume per origin between initiations and divide a constant time after initiation.

5.2 Significance

How bacteria regulate cell division to achieve cell size homeostasis, with concomitant coordination of DNA replication, is a fundamental question. Currently, there exist several competing models for cell cycle regulation in *Escherichia coli*. We performed experiments where we systematically perturbed cell dimensions and found that average cell volume scales exponentially with the product of the growth rate and the time from initiation of DNA replication to the corresponding cell division. Our data support a model in which cells initiate replication on average at a constant volume per origin and divide a constant time thereafter.

5.3 Introduction

Bacteria can regulate tightly and coordinate the various events in their cell cycles to accurately duplicate their genomes and to homeostatically regulate their cell sizes. This is a particular challenge under fast growth conditions where cells are undergoing multiple concurrent rounds of DNA replication. Despite much progress, we still have an incomplete understanding of
the processes that coordinate DNA replication, cell growth, and cell division. This lack of understanding is manifested, for instance, in discrepancies among various recent studies that propose different models for control of cell division in the bacterium Escherichia coli.

One class of models suggests that cell division is triggered by the accumulation of a constant size (e.g., volume, length, or surface area) between birth and division\textsuperscript{6,7,20}. Such models are supported by experiments measuring correlations between cell size at birth and cell size at division, which showed that, when averaged over all cells of a given birth size \( v_b \), cell size at division \( v_d \) approximately follows

\[
v_d = v_b + v_0, \tag{5.1}
\]

where the constant \( v_0 \) sets the average cell size at birth. This is known as the “incremental” or “adder” model, and cells following this behavior are said to exhibit “adder correlations”\textsuperscript{5,6,7,8,20,36,58}. Importantly, these models postulate that cell division is governed by a phenomenological size variable, with no explicit reference to DNA replication.

A second class of models for control of cell division postulates that cell division is governed by the process of DNA replication, which can be described as follows. The time from a replication initiation event to the cell division that segregates the corresponding sister chromosomes can be split into the C period, from initiation to termination of replication, and the D period, from termination of replication to cell division\textsuperscript{65,81}. Both the C and D periods remain constant at 40 min and 20 min, respectively, for cells grown in various growth media supporting a range of doubling times between 20 min and 60 min\textsuperscript{12,28}. We refer to growth rates within this range as fast. All experiments described here are carried out under such fast growth conditions. Note that C+D is 60 min and larger than the time between divisions at fast growth. This situation is achieved by the occurrence of multiple ongoing rounds of replication. That is, under these conditions, a cell initiates a round of replication simultaneously at multiple origins that
ultimately give rise to the chromosomes of their granddaughters or even great-granddaughters. Extending the basic definition of the C and D periods, Cooper and Helmstetter specifically proposed that an initiation event triggers a division after a time C+D, thereby ensuring that cells divide only after the completion of a round of DNA replication. This Cooper–Helmstetter (CH) formulation, hereafter the CH model, belongs to the second class of models for control of cell division.

The CH model is supported by the phenomenon of rate maintenance\textsuperscript{77}: After a change from one growth medium to a richer one (a shift up), cells continue to divide at the rate associated with the poorer medium for a period of 60 min. According to the CH model, in which division is triggered by initiation, all cells that have already initiated replication before a shift up will have also already committed to their ensuing divisions, and thus the rate of division will remain unchanged for a time C+D following a shift up.

The same value of 60 min also emerged in a seemingly different context a decade earlier, in the seminal study of Schaechter et al.\textsuperscript{18}. In their work, cell volumes, averaged over an exponentially growing population, were measured for culture growing under dozens of different growth media supporting fast growth. It was found that average cell volume was well described by an exponential relation with growth rate $V = \Delta e^{\lambda T}$, where $V$ is the average cell volume, $\Delta$ is a constant with dimensions of volume, $\lambda$ is the growth rate, $\tau = \log(2)/\lambda$ is the doubling time, and $T \approx 60$ minutes.

Donachie\textsuperscript{19} showed that this exponential scaling of average cell volume with growth rate can also be explained by the CH model if it is further assumed that cells initiate replication on average at a constant volume $\Delta_I$ per origin of replication at initiation. Because cells grow exponentially at the single-cell level\textsuperscript{21}, cells will then divide on average at a volume $\Delta_I$ per
origin times a scaling factor $S = 2^{(C+D)/\tau}$. The average cell volume then follows

$$V = \Delta 2^{(C+D)/\tau} = S\Delta,$$

with $\Delta = \log(2) \Delta_1$, because the cell volume averaged over an exponentially growing population is the average cell volume at birth times $2 \log(2)$ \cite{19}. In Schaechter's experiments, the C and D periods were approximately constant, giving rise to the exponential scaling observed. However, the derivation for Eq. 5.2 holds regardless of the values of the C and D periods, and in cases where they are not constant, average cell volume is not expected to scale exponentially with growth rate. Eq. 5.2 is known as Schaechter's growth law, but is referred to simply as the growth law for the rest of this paper.

Recent single-cell analyses found that cells indeed initiate replication on average at a constant volume per origin or per some locus close to the origin \cite{28}. Although further experiments are required, the fact that introduction of an origin onto a plasmid does not affect cell cycle timings or cell size suggests that the latter possibility is correct \cite{28,80,82}. Below, for simplicity, we use the phrase “per origin,” while keeping this complexity in mind.

Clearly, the two classes of models for control of cell division differ fundamentally. In the first class, division depends only on the accumulation of size from birth, and DNA replication plays no explicit role. In the second class, division is downstream of the preceding initiation of DNA replication. Importantly, also, the experiments leading to the first class of models defined cell size differences by measuring cell length. Because for a constant growth environment the widths of bacterial cells are very narrowly distributed with a coefficient of variation (CV, standard deviation divided by the mean) less than 0.05 \cite{7}, these analyses cannot distinguish whether cell size in a given environment is set by a constant volume, surface area, or length. This ambiguity raises the question of what is the key phenomenological variable governing cell
cycle progression.

Here, we sought to test these models by perturbing cell dimensions in E. coli and assaying the effects of those perturbations on both replication events and cell division. In our study, shape perturbations were achieved by systematically varying expression levels of the protein MreB, an actin homologue involved in cell wall synthesis, and the protein FtsZ, a tubulin homologue involved in the formation of the division septum\textsuperscript{72,83}. Our approach is indicated schematically in Figure 5.1. It extends and complements Schaechter’s experiments, in which growth rate was perturbed, and is reminiscent of the work of Harris et al.\textsuperscript{20}, in which cell dimensions were perturbed, but with the important addition to both studies that we also measured the cell cycle periods C and D because they play an important role in the CH model. For this paper, we define division as completion of septation.

5.4 Results

5.4.1 Decreased \textit{mreB} level resulted in increased cell width, with little change in cell length

Cell length and cell width are two major characteristics of a rod-shaped cell. As a cell grows, cell length increases exponentially, whereas cell width remains constant. How cell width is determined and maintained is largely unknown. However, several mutations of \textit{mreB} were reported to result in altered cell dimensions\textsuperscript{83,84,85}, thus raising the possibility that alterations in \textit{mreB} expression level would alter cell width. To continuously and systematically vary cell width, we constructed a strain in which the level of \textit{mreB} could be experimentally controlled. We used a system in which a \textit{P}_{\text{tet}} – \textit{tetR} feedback loop triggered \textit{mreB} expression\textsuperscript{16}. The modulated copy of \textit{mreB} was the sole version of the gene in the genome as the native copy of \textit{mreB} was replaced by a kanamycin-resistance gene. In this construct, expression of \textit{mreB} could
Figure 5.1: (a) Schematic illustration of the experiment. MreB and FtsZ are involved in cell wall synthesis and septum formation, respectively. Using \textit{mreB}- or \textit{ftsZ}-titratable trains, we are able to tune their expression levels continuously and perturb cell dimensions. In both experiments, the D period increased with cell width and length. The C period and doubling time remained constant. (b) Schematic illustration of our model. The perturbed D period sets the average number of origins per cell, which is equal to the scaling factor $S$ because replication initiation triggers cell division. The average number of origins per cell then sets the average cell volume, following the growth law. For titrated \textit{mreB} levels, cell volume changes manifested mostly as cell width changes. For titrated \textit{ftsZ} levels, cell length changed instead, because FtsZ did not affect cell width.
be tightly controlled by adjusting the concentration of an appropriate inducer of \( P_{\text{tet}} - \text{tetR} \), anhydrotetracycline (aTc).

We found that cell size increased with decreasing inducer concentration until, at very low \( mreB \) levels, the cells eventually lysed (at an aTc concentration below 1 ng mL\(^{-1} \) in rich defined medium (RDM) + glucose). Above this minimum threshold, the expression level of \( mreB \) varied linearly with the concentration of inducer\(^{16} \). We also found that within a certain range of \( mreB \) expression levels, the volume growth rates, or \( \text{OD}_{600} \) doubling rates, of the titratable strain remained approximately constant (CV of 0.03) (Figure 5.2d). Together, these results suggest that the titratable system is suitable for characterizing the functions of MreB in a quantitative manner with no need to consider complications due to differing growth rates.

We next measured cell dimensions by phase contrast microscopy (Figure 5.2a) and found that with decreasing \( mreB \) expression level, the cell width increased (Figure 5.2b). The cell length changed slightly, which we discuss below (Figure 5.2c). We verified that neither the presence of inducer nor the presence of the genetic circuit construct has any effect on WT cells within the ranges of inducer concentrations studied here (Figure 35.2bc). Similar results were obtained in all other growth media, supporting various fast growth rates, tested in this study. All of our experiments, and thus the resultant conclusions, concern fast growth conditions as defined above.

5.4.2 Decreased \( ftsZ \) level resulted in increased cell length and no change in cell width

We constructed and characterized, using the same method as above, an \( ftsZ \)-titratable strain to allow perturbation of cell length. We found, in agreement with previous work\(^{72} \), that cell length increased with decreased \( ftsZ \) expression levels, but that both cell width and growth rate remained relatively constant within the ranges of inducer concentration studied here (CV of 0.01 and 0.03, respectively, across different experiments, Figure 5.2e-h).
Figure 5.2: Titratable mreB or ftsZ expression to systematically perturb cell width or cell length, respectively, without affecting the volume growth rates. (a) Representative phase contrast images of the mreB-titratable and wild-type strains. (b,c) Scatter plot presents the average, mean cell width (averaged along the long axis of a cell) (b) or average cell length (c) of the individual wild-type (WT) or mreB-titratable cells. (d) Volume growth rates in bulk culture vs. mean cell width for mreB-titratable cells. (e-h) The same as (a-d) but for the ftsZ-titratable strain. Error bars represent the SEM of three replicates.
5.4.3 Correlations between perturbed cell dimensions and cell cycle timings

We next investigated how the *mreB* and *ftsZ* expression levels affect the C and D periods. For a bulk culture in steady state exponential growth, the CH model predicts that the average number of origins per cell, $\langle O \rangle$, scales exponentially with growth rate$^{13,70}$. This is because in steady-state exponential growth, the number of cells and the total number of origins in the population must both grow exponentially at the same rate. However, because the CH model postulates that division occurs only after a time C+D following the corresponding initiation, the total number of origins will be larger than the number of cells by the scaling factor $S$, defined above. Therefore, $\langle O \rangle = S$. This relation holds regardless of the value of C+D.

An expression for the average number of copies $X$ of a gene per cell as a function of the location $m$ of the gene along the chromosome ($m = 0$ for *oriC* and $m = 1$ for *terC*) can be derived similarly. Under the assumption that replication forks travel at a constant speed, the expression is $X = 2^{(C(1-m)+D)/70}$. We used this relation to extract the lengths of the C and D periods from quantitative PCR (qPCR) data of the copy numbers of different chromosome loci (including *oriC*, *terC*, and a series of different loci between them). We also measured $\langle O \rangle$, using replication run-out experiments.

These analyses revealed that, over the analyzed ranges in the levels of both *mreB* and *ftsZ* expression, the C period remained unchanged (CV of 0.09 and 0.05 for *mreB*- and *ftsZ*-titratable strains, respectively), whereas the D period increased with increasing cell width or length, respectively (Figure 5.3). Note that the relation between the D period and cell length predicted by our model below is not linear, but appears approximately linear given the particular values of the relevant parameters under the conditions of these experiments.
Figure 5.3: Changes in the cell cycle parameters as cell dimensions are perturbed. (a) The D period increased monotonically with cell width in mreB-titratable strains. The line is the best linear fit. (b) The C period remained approximately constant as cell width changed in response to titrated mreB expression levels. The line is the mean value of C averaged over mreB expression levels. (c) The D period increased monotonically with cell length in ftsZ-titratable trains. (d) The C period remained approximately constant as cell length changed in response to titrated ftsZ expression levels. The circle, triangle, and square indicate mreB-titratable, ftsZ-titratable, and WT strains, respectively. Different colors denote growth media: Red is RDM + glucose, and blue is RDM + glycerol. The SEMs of three replicates were smaller than the size of the symbols.
5.4.4 The growth law holds in the face of perturbations to cell dimensions

We find that the growth law, Eq. 5.2, holds in our experiments, both across a range of growth media (Figure 5.3) and across the two titratable perturbations that drastically affected cell width and cell length (Figure 5.4). The best-fit proportionality constant is $\Delta = 0.55 \pm 0.04 \ \mu m^3$. The plus-or-minus symbol indicates the 95% confidence interval of the fit. From the derivation of the growth law in the Introduction, we find the average cell size per origin at initiation to be $\Delta_I = 0.79 \pm 0.06 \ \mu m^3$. In contrast, the average cell area, cell length, and cell width are not proportional to the scaling factor $S$ (Figures 5.4 and 5.4).

We further tested the validity of a constant $\Delta$ by fixing $\Delta$ to $0.55 \ \mu m^3$ and calculating the product of $\log_2 (V/\Delta)$ and $\tau$, which is equal to $C+D$ according to Eq. 5.2 (Figure 5.6a). We found that the values of $C+D$ obtained in this way agree well with independent measurements in both the titratable and the WT strains (Figure 5.6b).
Figure 5.5: The growth law does not apply for cell length or cell width. The average cell length (a) and average cell width (b) were not proportional to the scaling factor. The black line shows the best fit with intercept forced to zero. The coefficients of determination of the fits are −0.62 and −1.09, respectively. A negative coefficient of determination implies that the mean is a better predictor than the fit. The circle, triangle, and square indicate mreB-titratable, ftsZ-titratable, and WT strains, respectively. Different colors denote growth media: Red is RDM + glucose, and blue is RDM + glycerol. The SEMs of three replicates were smaller than the size of the symbols.

5.4.5 Average cell volume at initiation is constant per some locus close to the origin

As discussed in the Introduction, the growth law follows from (i) the CH model and (ii) the constancy of average cell volume \( \langle \Delta I \rangle \) per origin at initiation. The second assumption says that the average volume at initiation is \( V_I = \bar{O} \Delta I \), where \( \bar{O} = 2^{\lfloor (C+D)/\tau \rfloor} \) is the number of origins in a cell at the time of initiation (e.g., 1 at slow growth and higher powers of 2 at fast growth\(^{13}\)).

The floor operator \( \lfloor x \rfloor \) finds the greatest integer less than or equal to \( x \). The first assumption then specifies a division at total cell volume (summed over all offspring of the mother cell corresponding to this initiation) \( V_D = V_I S \). Immediately after the specified round of division, there will have been a total of \( \log_2 (\bar{O}) + 1 \) rounds of divisions since the original initiation. Hence, the average cell volume at birth is \( V_B = \Delta I S / 2 \). Then in terms of the cell volume averaged over an exponentially growing population, \( V = \log (2) \Delta I S \). A similar reasoning follows if the second assumption is modified from “per origin” to “per some locus” a distance \( m \) away from
Figure 5.6: The average cell volume per origin at initiation is constant in the face of perturbations in cell dimensions. (a) Colored lines connect \( \log_2 (\Delta) \) to symbols in boldface type as examples. The respective slopes of the lines are computed as the ratio of \( \log_2 (V/\Delta) \) over \( 1/\tau \) and shown as numbers (minutes). (b) Measured \( C+D \) values are plotted against the ratios calculated in a. The circle, triangle, and square indicate \( \text{mreB-titratable}, \text{ftsZ-titratable}, \text{and WT strains, respectively. Different colors denote growth media: Red is RDM + glucose, and blue is RDM + glycerol. The SEMs of three replicates were smaller than the size of the symbols.}

the origin. In this case, the modified second assumption gives \( V_I = \bar{X} \Delta_I \) is the number of the locus in a cell immediately following initiation. Following the same reasoning, we find that the growth law becomes \( V = \log (2) (\bar{X}/\bar{O}) \Delta IS \). However, this expression cannot in general explain the constancy of \( \Delta \) across our experiments because the quantity \( \bar{X}/\bar{O} \) explicitly depends on the growth rate. The supposed locus must therefore be close enough to the origin such that \( \bar{X}/\bar{O} \) remains approximately constant at fast growth.

5.4.6 Cell length changes in \textit{mreB}-titratable strains

Under the assumption that \textit{mreB} expression level specifies both cell width and the D period, average cell volume is determined by the growth law, \( V = \Delta S \), with \( \Delta \), the C period, and \( \tau \) all relatively constant across \textit{mreB} expression levels. The cell volume of a rod shaped bacterium can be approximated as the volume of a cylinder with hemispherical caps, \( V \approx \frac{2}{3} W^2 (L - W) + \frac{1}{6} \pi W^3 \). The growth law therefore specifies cell length to be \( L \approx \frac{4}{\pi} (\Delta S + \pi W^3/12) / W^2 \). The
Figure 5.7: Cell length changes in mreB-titratable strains are small but consistent with the predictions by the growth law. Circles plot the same data as Figure 5.3C for mreB-titratable trains. Different colors denote growth media: Red is RDM + glucose, and blue is RDM + glycerol. Black symbols plot predictions of the growth law given cell width and the D period. Error bars represent 10% propagated errors estimated from uncertainties in cell width, the C and D periods, the doubling time, and the proportionality constant.

Cell lengths calculated in this way compare favorably to those measured via phase contrast microscopy (Figure 5.7). Importantly, although the variations in cell length are small, they are consistent with the predictions by the growth law.

5.5 Discussion

Here, we perturbed cell dimensions and then observed the effects of these perturbations on the cell cycle to interrogate the mechanism of cell cycle regulation in E. coli. The most important finding of this work is that the growth law, that average cell volume is proportional to the scaling factor \( S = 2^{(C+D)/\tau} \), remained valid across large perturbations in cell dimensions. As discussed in the Introduction, the growth law can be quantitatively derived under two assumptions: (i) the CH model, that replication initiation triggers cell division after a constant time \( C+D \), and (ii) that the average cell volume at initiation of DNA replication is proportional to the number of origins at initiation. The robustness of the growth law documented above suggests that both
assumptions hold in face of the perturbations studied here.

Correspondingly, given the first assumption, the presented results support the second class of models for control of cell division, in which the timing of initiation governs that of division, and oppose the first class of models in which the timing of division is governed by the accumulation of cell size with no explicit reference to DNA replication. The presented findings do not rule out models in which, opposite to the CH model, division triggers initiation. However, these models are heavily challenged by the finding that C+D is essentially constant (CV of ~0.1) on the single-cell level at fast growth. To be consistent with this finding, these models must somehow coordinate the triggered initiation event with the division event in the next generation. Further studies will be required to investigate this possibility.

Our findings also provide information about the question of what key phenomenological variable governs cell cycle progression. We find that only cell volume, and not surface area, length, or width, is proportional to the scaling factor $S$ (Figures 5.4 and 5.5). Together with the second assumption, the robustness of the growth law then supports the hypothesis that volume per origin, rather than other geometric features, is the key phenomenological variable that is invariant at initiation. Hence, cell volume governs the timing of initiation and, by the CH model, the timing of division.

Figures 5.4 and 5.6 also show that the proportionality constant $\Delta$, which links average cell volume to the scaling factor $S$ in the growth law (Eq. 5.2), remained constant across the perturbations studied here. In the derivation of the growth law, $\Delta$ is proportional to $\Delta_I$, or the average volume per origin required for initiation. The constancy of therefore suggests that none of the perturbations studied (MreB, FtsZ, and various growth media) affected the molecular mechanism underlying the regulation of initiation. Despite drastic changes in cell shape, the titratable strains still initiated replication on average at a constant $0.79 \pm 0.06 \, \mu m^3$ per origin just as in WT. This value for the average cell volume per origin at initiation is similar to the
recently reported value of $0.9 - 1.0 \, \mu m^3$.

Although the present study examined altered genetic conditions that result in altered cell length and cell width, our analysis implies that these changes in cell length and cell width did not affect cell volume directly. Rather—because neither genetic perturbation affected growth rate, the C period, or $\Delta$ - changes in the D period alone were responsible for the observed changes in cell volume, which in turn are manifested as changes in cell length and cell width.

Figure 5.1b illustrates schematically how the two genetic perturbations might exert their effects. In one case, reduced $ftsZ$ expression increases the D period. Because FtsZ is a direct mediator of septum formation, this effect could result directly from a prolongation of the septation process. The increased D period dictates an increased average cell volume, which, in this situation, happens to manifest as an increase in length alone, with no change in width. This asymmetric change in length, vs. width, matches the fact that cell width is maintained normally even in filamentous cells where septation is completely eliminated$^{72,86}$.

Reduced $mreB$ expression also increases the D period. In one possibility, MreB would have two direct roles: both in septation via local effects at the division site$^{87}$ and in determining cell width. Here, at reduced MreB levels, delayed septation would increase the D period. Due to the second role of MreB, the corresponding increase in cell volume in this case is implemented mostly by the increase in cell width. Alternatively, MreB might play only the role of determining cell width. In this scenario, the increased cell width then results in a prolonged septation process and thus a longer D period. In both scenarios, however, given experimentally determined values for the D period and cell width, the cell volume predicted by the growth law also dictates small changes in cell length. The magnitudes of these predicted changes are just at the level of detection of the current study but are consistent with observed values (Figure 5.7).

In general, the growth law specifies cell volume without reference to the aspect ratio of the rod-shape morphology. As a result, a given change in cell volume due to a change in the
parameters of the growth law can be manifested as diverse combinations of changes in cell length and cell width.

Our analysis has shown that, in analyzing size-related measurements, the growth law and its underlying tenets imply that any perturbations to the C or D periods or growth rate will affect cell volume - even if the perturbations are not affecting the core mechanism of size regulation that determines the value of the invariant average cell volume per origin at initiation. Thus, with respect to determining cell size, an important distinction can be made between “primary” and “secondary” regulators\textsuperscript{88}. In the context of the above analysis, MreB and FtsZ appear to be secondary regulators in \textit{E. coli} because remained constant across titrated \textit{mreB} and \textit{ftsZ} levels.

All of the considerations above build on the classical works of Schaechter et al.\textsuperscript{18}, Cooper–Helmstetter\textsuperscript{12,65}, and Donachie\textsuperscript{19}, which consider population average behaviors. Donachie further proposed a single-cell interpretation of his idea in which initiation occurs in a cell when the cell reaches a constant volume per origin\textsuperscript{19}. However, the proposal is not compatible with experimental single-cell measurements showing adder correlations between cell sizes at births and at divisions because it predicts no such correlations under conditions where growth rate is essentially constant\textsuperscript{13}.

As a resolution to this conflict, some of us have recently proposed a phenomenological adder-per-origin model in which a constant volume is added between two rounds of initiations, rather than between two rounds of divisions\textsuperscript{13}. We have shown that this adder-per-origin model leads both to adder correlations and to a constant average cell volume per origin at initiation. Furthermore, adder per origin can also explain rate maintenance and the growth law.

We conclude by discussing several unsolved questions. First, we have ignored single-cell fluctuations in the C and D periods in our discussion, but recent works show that they are important to understanding single-cell correlations\textsuperscript{13-44}. We have also not discussed \textit{E. coli}
in slow growth conditions, but several works raise the possibility that *E. coli* might behave qualitatively differently there. For instance, previous works suggested that average cell volume per origin at initiation is not constant in such conditions\textsuperscript{89,90}. Recent works also suggested that at slow growth, *E. coli* does not show adder correlations\textsuperscript{28}. In light of these observations, it will be important to further study the single-cell physiology of *E. coli* under cell dimension perturbations at slow growth.

There has recently been much interest in the question of cell size homeostasis across all domains of life (Chapter 3), and we note that adder per origin may be applicable to other organisms as well. For example, it is known that the bacterium *Bacillus subtilis* also exhibits the growth law and adder correlations. Repeating the experiment here in *B. subtilis* may help probe the relations between cell dimensions and cell cycle timings in a Gram-positive bacterium. Adder correlations in cell volume were also found recently in budding yeast diploid daughter cells\textsuperscript{8}. Given the different morphology of these cells, which changes dramatically throughout the cell cycle and particularly at budding, it is plausible that cell volume, and not cell shape, is the key phenomenological variable governing cell cycle regulation also in this case.

Importantly, although our study here has suggested a coarse grained, phenomenological model on the level of cell dimensions, it has not alluded to the molecular players involved. Several hypothetical molecular mechanisms, such as the accumulation of a threshold amount of an initiator protein per origin\textsuperscript{22} or the dilution of an inhibitor protein\textsuperscript{23}, were previously shown to implement molecularly the phenomenological model discussed here\textsuperscript{8,13}. However, despite decades of work, the molecular mechanism for cell cycle regulation in bacteria remains a fundamental unresolved question.
A parallel adder coordinates mycobacterial cell cycle progression and cell size homeostasis in the context of asymmetric growth and organization.

6.1 Adapted summary

In model bacteria, such as *E. coli* and *B. subtilis*, regulation of cell cycle progression achieves consistency in cell size and replication dynamics\(^3\). Mycobacteria elongate and divide asymmetrically, giving rise to significant variation in cell size and elongation rate among closely related cells\(^3\). Given the physical asymmetry of mycobacteria, the models that describe coordination
of cellular organization and cell cycle progression in model bacteria are not directly translatable\textsuperscript{6,7,12}. Here, we used time-lapse microscopy and fluorescent reporters of DNA replication to examine the coordination of growth, division, and DNA replication at a single-cell level in \textit{Mycobacterium smegmatis} and \textit{Mycobacterium bovis Bacillus Calmette-Guerin} (BCG). Using measurements and stochastic modeling of mycobacterial cell size and cell cycle timing in both slow and fast growth conditions, we found that well-studied models of cell size control are insufficient to explain the mycobacterial cell cycle. Instead, we showed that mycobacterial cell cycle progression is regulated by an unprecedented mechanism involving parallel adders (i.e., constant growth increments) that start at replication initiation. Together, these adders enable mycobacterial populations to regulate cell size, growth, and heterogeneity in the face of varying environments.

6.2 Results

The growth mode of \textit{M. smegmatis} differs significantly from that of \textit{E. coli}\textsuperscript{34}. First, \textit{M. smegmatis} cells accrue cell volume at the poles, rather than uniformly throughout the cell. Cell division is asymmetric. At division, the daughter inheriting the older pole is termed the accelerator cell because it is born larger and elongates faster than the sister inheriting the new pole, termed the alternator cell because it is born smaller and elongates more slowly. The average division ratio for accelerator and alternator sister cells is 44\%/56\%\textsuperscript{9}. We wondered how mycobacteria control cell size in face of such asymmetric growth and division. We used time-lapse imaging to measure cell size and cell cycle timing in \textit{M. smegmatis} cells. To characterize cell cycle progression, we measured the timing of DNA replication using a \textit{M. smegmatis} strain expressing a fluorescently tagged single-stranded binding protein reporter SSB-GFP\textsuperscript{9}. Using time-lapse microscopy, we observed one to two SSB-GFP foci during DNA replication (\textsuperscript{\textdagger}C\textdagger)
but none during pre- (“B”) and post- (“D”) replication periods (Figure 6.1). We note that the experiments were conducted under slow growth conditions, as defined in Chapter 5. We and others have observed *M. smegmatis* subpopulations that begin new rounds of replication after completing the first round but before division\textsuperscript{91-93}. We identified this additional replication period (“E”) in 54% of cells by observing SSB foci near poles pre-division, a result we discussed in Chapter 4 and will discuss in detail below.

Several studies recently developed a division adder model of cell size control, in which bacteria add a constant length ($\Delta l_{bd}$) from birth to division, regardless of birth size (Figure 6.2a)\textsuperscript{6,7}. Because $\Delta l_{bd}$ is not correlated to birth length, we found that *M. smegmatis* is consistent with this aspect of the division adder model (Figure 6.2b). However, *M. smegmatis* is inconsistent with other aspects of the adder model. For example, there is a proportional relationship between birth length and interdivision time expected in exponentially growing cells that are division adders: $t_d \propto \log_2 (1 + \Delta l_{bd}/l_b)$ (Figure 6.2c). Using a method described in Chapter 3, we found that *M. smegmatis* data are consistent with exponential growth (Figure 6.5ab), but
systematically deviate from the expected relationship between birth length and interdivision
time, in contrast to the result in *E. coli* (Figure 6.2c). These data suggest a more detailed
model is needed to capture the coupling between DNA replication and cell division, and the
resulting statistics of cell size and division timing.

We therefore considered whether the coupling between DNA replication and cell division in *M.
smeğmatis* is consistent with existing models that couple division to initiation timing.\(^{12,13,19,28}\) The recent adder-per-origin model, an adaption of the Cooper-Helmstetter model, postulates that (1) the time from initiation to division is constant, regardless of cell size, and (2) initiation occurs after a constant growth increment measured from the previous initiation (Figure 6.2d, Chapter 5). Models that include initiation timing must account for the number of origins, because cells may initiate multiple rounds of replication before division. In the adder-per-origin model, the growth increment scales with the number of origins to ensure cell size homeostasis when there are re-initiation events.\(^{13}\) Thus, the adder-per-origin model achieves size homeostasis by implementing the adder mechanism at initiation. In this and previous studies, we used the SSB-GFP reporter to characterize the timing of DNA replication in *M. smeğmatis* (Figure 6.2e). We found that the frequency and duration of the E period correlate positively with birth length (Figure 6.2e). We also found a negative correlation between cell size at initiation and time from initiation to division (slope of 0.38 ± 0.09 95% confidence interval [CI]; Figure 6.2f). Because timing from initiation to division is not independent of cell size at initiation, the coupling between DNA replication and cell division in *M. smeğmatis* cannot be described by the adder-per-origin model.
Figure 6.2: *M. smegmatis* growth and cell cycle timing in rich medium. (a) Schematic representation of the division adder model of cell size control in cells with (bottom) or without (top) E period. $\Delta l_{bd}$ is the growth increment between birth and division events. (b) Scatterplot of birth length compared to growth from birth to division. Light gray dots are individual data points, and gray squares are averages of data binned in 1-mm increments with SEM bars. The linear regression line is plotted in gray. Inset shows normalized birth size versus division size with Pearson correlation $r = 0.65 \pm 0.06$ 95% CI. (c) Scatterplot as in (b) of normalized birth size $l_b/\langle l_b \rangle$ versus normalized generation time $t_d/\langle t_d \rangle$. Red diamonds are binned *E. coli* data from8. (d) Schematic representing the adder-per-origin model of cell size control in cells with (bottom) or without (top) E period. $\Delta l_{ii}$ is the growth increment between replication initiation events in mother and daughter cells and C+D or C+D+E is the time between initiation and division. Open circles represent number of origins ($O$) present in different phases of the cell cycle. Open circles represent number of origins present in different phases of the cell cycle. (e) Pie chart of average time spent in each phase of the *M. smegmatis* cell cycle. Violin plots of cell phase timing are binned by cell size at birth (bin ranges span 1 mm, centered as indicated on the axes); red bars denote mean and SEM. The E period violin plot also displays the percentage of cells per bin that experience E period. (f) Scatterplot as in (b) of initiation to division timing ($((C + D + E) / t_d)$) versus cell length at initiation ($l_i/\langle l_b \rangle$). Pearson correlation $r = 0.35 \pm 0.09$. (G) Normalized initiation to division increment versus normalized cell length at initiation, plotted for accelerator (Acc; green) and alternator (Alt; red) subpopulations; error bars represent SEM. Linear regression lines are fit to data of each cell type with Acc Pearson correlation $r = 0.06 \pm 0.09$ and Alt Pearson correlation $r = 0.08 \pm 0.07$. 
Figure 6.2 (Continued).
Instead of a constant time, we observed a constant growth increment between initiation and division. This increment is maintained in both accelerator and alternator subpopulations, despite their distinct size and growth properties (Figure 2g). We sought to reconcile this initiation-to-division adder with the birth-to-division adder in *M. smegmatis*. Inspired by the success of the adder-per-origin model, i.e. an initiation-to-initiation adder, in describing the emergent birth-to-division adder in *E. coli*\textsuperscript{13}, we hypothesized that cell cycle regulation in *M. smegmatis* might be described by a parallel adder model: one from initiation to division and another from initiation to initiation (Figure 6.3a).

The difference between the parallel adder and the adder-per-origin models is how division is coupled to initiation. In the adder-per-origin model, a constant time elapses between initiation and division. In the parallel adder model, cells grow a constant length from initiation to division. The models assume the same form of regulation at initiation, namely that a constant length per origin of replication is added between initiation events. The parallel adder therefore achieves cell size homeostasis in face of the E period, as well as allowing for multifork replication, a phenomenon recently reported in *M. smegmatis*\textsuperscript{94}.

We numerically simulated the parallel adder model and compared the results to experimental measurements, similar to our approach in Chapter 3 (Figure 6.3a). Because most parameters could be extracted from our data, only two free parameters in the parallel adder model were fit to our data (standard deviation of interdivision timing and growth rate; details below). We numerically calculated thirteen statistics of stochastic variables related to cell cycle timing, including coefficients of variation (CVs, or the standard deviation divided by the mean) and correlation coefficients (Figure 6.3b). To compare simulated and measured statistics, we performed an analysis of the sensitivity of the various statistics to errors in size measurements, following the approach discussed in Chapter 2 and executed in Chapters 3 and 7. We observed that many coefficients are relatively insensitive to noisy measurements whereas others (such as

112
growth rate) exhibit a broad range of values with a modest 10% error in size measurements. The results indicate that some CVs and correlation coefficients are sensitive to various sources of stochasticity (Figure 6.3b). Our approach identified the CVs and correlation coefficients that are most robust to noise, and therefore appropriate to be compared to experimental data.

Having identified robust indicators of the underlying cell cycle regulation strategy, we numerically simulated the division adder, adder-per-origin, and parallel adder models. All three models account for the morphological differences between accelerator and alternator cells. Comparison of the numerical results with our measurements in *M. smegmatis* demonstrate that the thirteen CVs and correlation coefficients are consistent with the parallel adder model and inconsistent with both the division adder and the adder-per-origin model (Figure 6.3b). We verified that the time resolution used for imaging does not affect the comparison between measurements and simulations. Importantly, the parallel adder model simultaneously captures the correlations between Δld and initiation size and between birth length and interdivision time, which cannot be explained by the adder-per-origin model (Figure 6.3cd). The parallel adder model leads to a convergence in the population averaged cell size even for hypothetical cells born very large or small (Figures 6.3e and 6.5ef), as well as for accelerator and alternator subpopulations, despite larger accelerator cells adding larger increments than smaller alternator cells (Figures 6.2g, 6.3e, 6.4C, and 6.5i).

We next evaluated whether the parallel adder could describe cell cycle regulation in *M. smegmatis* in carbon-limited slow growth conditions and with the slow-growing species BCG (Figure 6.6). BCG exhibited an interdivision time of 15–20 hr whereas the average carbon-limited *M. smegmatis* interdivision time was 4.9 hr, in contrast to 3.1 hr in rich medium (Figures 6.2e, 6.4a, and 6.6cd). We identified several differences in BCG and carbon-limited *M. smegmatis* cell cycle timing compared to rapidly growing *M. smegmatis*. B period was disproportionately extended in slow growth, particularly in small cells (Figures 6.4a and 6.3d).
Figure 6.3: Parallel adder model of mycobacterial cell size control. (a) Schematic of parallel adder model of cell-size control in cells with (bottom) or without (top) E period. $\Delta l_{ii}$ is the growth increment between two initiation events, and $\Delta l_{id}$ is the growth increment between initiation and division. Open circles represent number of origins ($O$) present in different phases of the cell cycle. (b) Chart of cell-cycle correlation coefficients and CVs comparing *M. smegmatis* measurements normalized to population mean (blue diamonds with 95% confidence interval bars) to parallel adder model simulations (light blue squares), adder-per-origin model simulations (red triangles), and division adder model simulations (green triangles). Parallel adder simulations with 0% measurement error prediction are represented with open light blue squares, and simulations with 10% measurement error prediction are represented with light blue bars. Variables represented are $l_b$, birth length; $l_d$, length before division; $l_i$, length at initiation of DNA replication; $t_d$, interdivision time; $\Delta l$, length increment between the b, birth, i, initiation, and d, division events indicated in the subscripts; C+D+E, total time spent initiation to division; and $\lambda$, growth rate. (c) Scatterplot of normalized initiation length ($l_i/\langle l_i \rangle$) versus normalized initiation to division increment ($\Delta l_{id}/\langle l_b \rangle$). Light blue dots are individual data points, and blue squares are binned data with SEM bars. Simulated correlations from parallel adder and adder-per-origin model simulations are plotted as blue (parallel adder) and red (adder per origin) lines. (d) Scatterplot as in (c) of normalized birth length ($l_b/\langle l_b \rangle$) versus normalized generation time ($t_d/\langle t_d \rangle$). (e) Plot showing parallel adder simulations for progeny of hypothetical cells born extremely large (blue) or small (green). The average cell birth length with SEM bars over 8 generations is plotted for a hypothetical cell (accelerator) born 2.53 the population average and a hypothetical cell (alternator) born 0.33 the population average. Average sizes of accelerator and alternator cell progeny from each hypothetical progenitor cell are also plotted with SEM bars over 7 generations.
BCG and carbon-limited *M. smegmatis* birth lengths were more variable than rapidly growing *M. smegmatis*, with CVs of 20% and 22%, respectively, compared to 19% for rapidly growing *M. smegmatis* (Figures 6.3b, 6.4b, 6.6ae) and 12% for *E. coli*.\textsuperscript{6}
Figure 6.4: Cell cycle timing and growth of *M. smegmatis* under carbon limitation. (a) *M. smegmatis* cell-cycle timing during carbon limitation, as in Figure 6.2. (b) Chart of carbon-limited *M. smegmatis* cell-cycle correlation coefficients and CVs comparing measurements normalized to population mean to model simulations, as in Figure 6.3. (c) Average normalized accelerator and alternator $\Delta l_{cd}$ and $\Delta l_{ci}$ growth increments for *M. smegmatis* in rich and carbon-limited media and BCG in rich medium.
Figure 6.4 (Continued).
We compared numerical simulations of the parallel adder, adder-per-origin, and division adder models to measurements from slow growth. We observed that the parallel adder model describes these data well, in contrast to the adder-per-origin and division adder models (Figures 6.4b and 6.6e). Whereas the length increments to division and initiation are similar in rich growth conditions, \( \Delta l_{id} \) is much longer than \( \Delta l_{ii} \) for *M. smegmatis* in carbon-limited medium and BCG (Figure 6.4c). Within the context of the parallel adder model, this difference in increments explains the observation that the E period is rare, and B extended, during slow growth (Figures 6.2e, 6.4a, and 6.6d). This provides one example of how the parallel adder model offers flexibility to shift and capture cell cycle coordination in varying growth conditions.

We conclude that mycobacteria utilize a parallel adder at initiation to control cell size in the context of deterministic growth and size variation. We have yet to fully understand mechanisms by which mycobacteria implement a parallel adder, but our study informs the search for the molecular basis of these processes. Given the important role of mycobacterial growth and variation on disease and treatment outcome, we anticipate the further understanding of these fundamental processes may give rise to new therapeutics against mycobacterial diseases, such as tuberculosis.

6.3 Methods

6.3.1 Mode of growth at the single-cell level

Most models of bacterial cell size control are based on the assumption that cells grow exponentially in volume over time. However, single cell traces cannot distinguish between linear and exponential modes of growth in *M. smegmatis* (Figure 6.5a). We therefore evaluated cell growth parameters from numerous cells to determine whether normalized cell generation time (i.e., generation time divided by the mean generation time) was proportional to the natural log
of the ratio of division length to birth length (\(\ln (l_d/l_b)\)), as is expected for exponential growth or comparable to the linear growth expectation of total elongation \((l_d - l_b)\) (Figure 6.5a). Overall, we found that *M. smegmatis* measurements were consistent with the exponential model of cell growth in bulk compared to a linear growth model (Figure 6.5b). Cells with the longest generation times (the largest two binned data points) deviate slightly from exponential growth, however the effect on the overall fit of the data to the \(\ln (l_d/l_b) = \lambda t_d\) line and subsequent comparison of models is not significant because they contain very few data points (6/391 cells or 1.5% of the data) (Figure 6.5b). Determining the growth mode directly at the single cell level requires more advanced techniques\(^{21}\). Nevertheless, the agreement with exponential growth is strong enough that we modeled cells growing exponentially with growth constant \(\lambda\).
Figure 6.5: *M. smegmatis* growth mode. (a) Three representative single cell traces of *M. smegmatis* growth fit to linear and exponential growth models. The mean squared error (MSE) of both models fit to single cell growth data is similar, suggesting that growth data from imaging is insufficient to differentiate between linear and exponential growth models (b) Light blue dots are individual data points and blue squares are binned data with SEM bars. Data fit to $\ln(l_d/l_b) = \lambda t_d$ (blue line) are compared to simulation of linear growth (green diamonds). (c) Schematic of cell size convergence simulations corresponding with panels d-i. (top) Schematic demonstrates convergence over several generations by following the average birth size of multiplying cell population. Schematic shows progeny of hypothetical cells born extremely large (purple, starting at the purple star) or small (red, starting at the red star). The number of cells in the total population of which the average is plotted is shown below the x-axis. (bottom) Simulation of a model to demonstrate convergence over several generations of the pure accelerator cell lineage (blue) and pure alternator lineage (green) from individual large and small accelerator and alternator cells (e.g. four different simulations are shown as indicated with stars). For the accelerator lineage, the simulation begins with either a large or a small accelerator cell and then considers only what happens with the accelerator daughter cell (alternator daughter cells are disregarded in this case, as diagramed below the x-axis). Plotted is the birth length of the accelerator progenitor and its accelerator daughters for 8 total generations. Only one of the daughters is an accelerator so cell lengths are reported for only one cell per generation. The alternator progenitor cell and alternator daughters are similarly considered in the alternator lineage. The population-averaged value at stationarity (black) is obtained from simulations. Dashed light-colored lines plot the values of model predictions from the expressions for each model. (d-f) Simulation of population cell size convergence to the predicted mean birth length without added noise (top) and with added noise in growth rate, division ratio, and cell cycle timing (inset) for the division adder (d), adder-per-origin (e), and parallel adder (f) models. The average cell birth length with SEM bars over 8 generations is plotted for a hypothetical cell (accelerator) born 2.5x the population average and a hypothetical cell (alternator) born 0.3x the population average. Average sizes of accelerator and alternator cell progeny with SEM bars over 7 generations for each hypothetical cell are also plotted. See (c, top) for a schematic description. (g-i) Simulation of pure accelerator and alternator lineages to demonstrate convergence of the longest and shortest cell subpopulations in the division adder (g), adder-per-origin (h), and parallel adder (i) models. See (c, bottom) for a schematic description.
Figure 6.5 ( Continued ).
6.3.2 Model details

We consider three models of cell size regulation: the division adder, the adder-per-origin, and the parallel adder. In all three models, we distinguish between alternator and accelerator cells. Accelerator cells are, barring noise, the larger cells upon asymmetric division with ratio \( r \) (\( r = 0.5 \) corresponds to symmetric division), and may have different requirements than alternator cells for replication initiation and division, as discussed below. The three models each maintain cell size homeostasis, as illustrated with simulation in Figures 6.3e and 6.5c-i, and also explained with derivations below.

Division adder model  In the division adder, cells attempt to add a constant size increment from birth to division. The increment may be different between alternator and accelerator cells. This model does not consider replication initiation, and can be summarized as

\[
\ell_d^z = \ell_b^z + \Delta \ell_{bd}^z,
\]

where \( \ell \) is the cell length at the event denoted by the subscripts b and d, for birth and division. The superscript \( z \) denotes alternator (alt) or accelerator (acc). Given exponential growth (\( \ell_d = \ell_b e^{\lambda t_d} \) where \( \lambda \) is the growth constant and \( t_d \) is the generation time), a division adder (\( \ell_d = \ell_b + \Delta \ell_{bd} \)) may be rewritten in terms of \( t_d \), by a log transformation and substitution, as follows and described also in Chapter 3, \( t_d = \ln (1 + \Delta \ell_{bd}/\ell_b) \). The convergence property of the “basic” division adder without distinction between accelerators and alternators has been rigorously investigated and is well understood\(^{15}\). In summary, the division adder converges because a cell adds a constant size regardless of its size at birth. Following this strategy, cells smaller than average will increase in size and cells larger than average will decrease in size until they reach the average size over several generations.
With the additional consideration of accelerators and alternators, a similar reasoning applies. In this model, the growth increment may be different for the accelerator and alternator sub-populations. We show by model simulation that even when accelerator and alternators cells do have different requirements for the growth increment between division events, the birth length of the population converges (Figure 6.5d). In this simulation, a large accelerator cell (purple) and a small alternator cell (red) each seed a population in which the average cell lengths at birth converge over several generations.

To understand the theoretical basis for convergence, it is helpful to consider what happens to the birth lengths of “pure” subpopulations of accelerator cells (from a lineage of all accelerator cells) and “pure” subpopulation of alternator cells (from a lineage of all alternator cells) separately, as they represent the extremes of large and small cell subpopulations, respectively. We note that these pure subpopulations do not exist naturally because they consider only one of two daughter cells at each division. In reality, every division gives rise to one accelerator and one alternator cell. Consider an accelerator cell from this artificial pure subpopulation lineage. Its size at division is on average

\[
\langle l_{bd}^{acc} \rangle = \langle l_{bd}^{acc} \rangle + \Delta l_{bd}^{acc}.
\]  \hspace{1cm} (6.2)

Hence for a division ratio \( r \), we find that

\[
\langle l_{bd}^{acc} \rangle = \frac{r}{1-r} \Delta l_{bd}^{acc}.
\]  \hspace{1cm} (6.3)

Cells in the pure accelerator subpopulation are the largest cells on average, and their sizes converge to the above expression. Similarly, the average size of a pure alternator cell is

\[
\langle l_{bd}^{alt} \rangle = \frac{1-r}{r} \Delta l_{bd}^{alt}.
\]  \hspace{1cm} (6.4)
We illustrate the convergence of both pure lineages by simulation (Figure 6.5g). In the longest cell subpopulation possible (the pure accelerator cells), the lineages from a large cell and a small cell each converge to \((r/(1-r)) \Delta l_{bd}^{acc}\). Similarly, the lineages of alternator cells (the shortest cell subpopulation) converge to \(((1-r)/r) \Delta l_{bd}^{alt}\). Because the accelerator and alternator subpopulations each converge, the entire population must also converge. An expression for the population averaged size at birth can be derived by considering the target size at division \(l_{zd}\) of a cell of type \(z\), not necessarily from a pure lineage. Summing the target sizes for the two types

\[
\langle l^\prime_d \rangle_{total} = l^{acc}_b + l^{alt}_b + \Delta l^{acc}_{bd} + \Delta l^{alt}_{bd}.
\]

(6.5)

Because the two types exist in equal numbers in a population, the population averaged size at birth and division can be approximated as \(\langle l_b \rangle = (\langle l^{alt}_b \rangle + \langle l^{acc}_b \rangle) / 2\) and \(\langle l_d \rangle = (\langle l^\prime_d \rangle + \langle l^\prime_{d'} \rangle) / 2\), respectively. Because cell size homeostasis requires that \(\langle l_d \rangle = 2 \langle l_b \rangle\), the above expression reduces to \(\langle l_b \rangle = (\Delta l^{acc}_{bd} + \Delta l^{alt}_{bd}) / 2\). We verified this convergence with stochastic population simulation (Figure 6.5d).

Adder-per-origin model  In the adder-per-origin model, cells attempt to add a constant size increment per origin of replication between replication initiations, and a round of replication initiation triggers cell division after a constant time initiation to division. The model can be summarized as

\[
(\ell)_{total}^{i'} = \ell_i + O \Delta \ell_{ii}^{i'}
\]

\[
l_d = \ell_i e^{(C+D)}.
\]

(6.6)

(6.7)

Here, \((\ell)_{total}^{i'}\) is the total cell length for the next (indicated by 'i') round of initiation. If the cell divides before this next initiation, the total length is the summation of lengths of both
Figure 6.6: Mycobacterial growth and division in fast and slow growth conditions. (a) Histograms of cell birth lengths of *M. smegmatis* growing in rich medium (7H9), carbon limited medium, and BCG in rich medium. Mean birth length in each condition is displayed as a black line. (b) Box plots of *M. smegmatis* growth rate in rich medium (7H9), carbon limited medium, and BCG in rich medium. (c) Box plots of *M. smegmatis* interdivision times in rich medium (7H9), carbon limited medium, and BCG in rich medium. (d) BCG cell cycle timing. (e) Chart of BCG cell cycle correlation coefficients and coefficients of variation comparing BCG measurements normalized to population mean to model simulation, as in Figure 3B.
daughters. $z$ (for Acc or Alt) corresponds to the cell type at the first initiation (e.g., the mother and not each daughter). $O$ is the number of origins in the cell. The number of origins $O$ is not set by hand, but is a dynamic variable. For example, in cases without E period where there are no ongoing replication forks after division, the number of origins is equal to one before initiation and is equal to two after initiation (Figure 6.2d).

Note that because the number of origins per cell is halved at division, the growth required until initiation must be recalculated as follows. Each daughter cell must accumulate a size increment $O \Delta l_i / 2$. But each daughter will inherit a different cell length accumulated after the mother’s initiation because of asymmetric division so the initiation length for the sister accelerator and alternator cells will be adjusted according to the division ratio. Consider the accelerator daughter: it has already accumulated a length $r(l_d - l_i)$ toward the requirement, and will initiate once it has accumulated the remaining length. Therefore, the accelerator daughter will initiate replication at size

$$ (t_i)^{acc} = (t_i)^{total} / 2 + (r - 1/2) l_i. \quad (6.8) $$

Similarly, the alternator daughter will initiate replication at size

$$ (t_i)^{alt} = (t_i)^{total} / 2 - (r - 1/2) l_i. \quad (6.9) $$

For simulations, it is useful to express $(t_i)^{acc}$ in terms of total requirement and the previous size at initiation.

It is important that a cell attempts to add a constant size increment per origin. This is because cells may initiate an extra round of replication due to stochasticity in cell cycle timings. In these cases, adding not just a constant size increment, but a size increment per origin allows
the cell to maintain size homeostasis (Chapter 4). Due to these extra rounds of replication initiating before division (i.e., E period) a sequential adder model where cells grow a constant increment from birth to initiation and another constant increment from initiation to division was also ruled out. In a sequential adder model, every initiation is uniquely coupled to a division event. Thus, a sequential adder model does not allow a cell cycle in which two initiations occur, in contradiction to our experimental findings that cell cycles with an “E period” do occur.

Expressions for the average size at birth of cells in pure lineages can be found as follows. Because the regulation at initiation in the adder-per-origin model is the same as that in the parallel adder model, the result at initiation is the same as above. The average size at division of a pure accelerator/alternator cell is simply

$$
\langle l_z^d \rangle = e^{(C+D)} \langle l_z^i \rangle,
$$

where $\langle l_z^i \rangle$ is the same as in the next section.

The adder-per-origin model achieves size homeostasis by implementing the adder mechanism between replication initiations. Its convergence has been shown theoretically and experimentally\textsuperscript{13}. Briefly, the per-origin requirement is necessary for convergence because of the possibility of extra rounds of initiation within a division cycle or E periods. The per-origin requirement allows the model to recover from these stochastic events. We show by simulation that cell size converges even when the growth increment between initiations is larger for accelerator cells than alternator cells (Figure 6.5e). As in the division adder model, we also show in Figure 6.5h that the largest and smallest subpopulations (the pure lineages of accelerator and alternator cells) converge. Because the accelerator and alternator subpopulations each converge, the entire population must also converge.

The population averaged cell size at birth can be derived as follows. As will be derived in
the next section, the target size at initiation $L_i^y$ of a cell of type $y$ whose mother is of type $z$ is

$$L_i^y = s(z) L_i^z + \Delta l_{ii}^z,$$  \hspace{1cm} (6.11)

where $s(z) = r$ or $1 - r$ for accelerator and alternator cells, respectively. The four types of target size at initiation then lead to four types of target size at division via multiplication by a factor $e^{\lambda(C+D)}$. Because these four types of cells exist in equal numbers in the population, the population averaged cell size at birth can be approximated as $\langle l_b \rangle = (\Sigma_{iy} L_i^y)/8$. After simplification, $\langle l_b \rangle = (\Delta l_{ii}^{alt} + \Delta l_{ii}^{acc}) e^{\lambda(C+D)}/2$. Simulations confirm this result (Figure 6.5e).

Parallel adder model  In the parallel adder model, cells attempt to add a constant size increment per origin between initiations, and from initiation to division. The increments may be different between initiations and from initiation to division. The model can be summarized as

\begin{align*}
(\ell'_i)^{\text{total}} & = \ell_i^z + O\Delta l_{ii}^z \hspace{1cm} (6.12) \\
\ell_d^z & = \ell_i^z + O\Delta l_{id}^z, \hspace{1cm} (6.13)
\end{align*}

where $l_i$ is the cell size at initiation, $(\ell'_i)^{\text{total}}$ is the total combined cell size of the daughter cells at the next initiation as described above for the adder-per-origin model, $l_d$ is the cell size at division, and $O$ is the number of origins of replication. $O$ is typically two at the beginning of C period, because C+D is often less than the average doubling time, but is four if the cell experiences re-initiation at E period. $\Delta l_{ii}^z$ is the size increment required per origin between successive initiations, while $\Delta l_{id}^z$ is the increment required from initiation to division. There is a cell type dependence on increments in the parallel model such that the superscript $z = \text{acc}; \text{alt}$ denotes alternator (Alt) or accelerator (Acc) cell types and these different increments
capture the asymmetric growth rates and different lengths of accelerator and alternator cells. The increments are set in part by measured population averaged values, because

\[ \langle l_b \rangle \approx (\Delta_{id} + \Delta_{ii}) / 2, \]  
\[ \langle l_i \rangle \approx \Delta_{ii} \]  

(6.14)  

(6.15)

where \( \Delta_{ii} = \Delta_{ii}^{acc} + \Delta_{ii}^{alt} \) and \( \Delta_{id} = \Delta_{id}^{acc} + \Delta_{id}^{alt} \). This result is derived below. Although the ratios between the accelerator and alternator increments can be chosen arbitrarily, numerical simulations described below show that the parallel adder best describe the data when the increments are either equally distributed or distributed in proportion to the asymmetry ratio \( r \). Simulations of the model best capture data from \( M. \) smegmatis and BCG growing in rich medium when the increments are proportioned between alternator and accelerator cell types per the asymmetry ratio \( r \) and best capture data from \( M. \) smegmatis growing in carbon limited medium when the increments are equally distributed between alternator and accelerator cell types.

When increments are proportioned between the alternator and accelerator cell types per the asymmetry ratio \( r \), the average size of pure accelerator and pure alternator cells can be derived as follows. To derive \( \langle l_{b}^{acc} \rangle \) for a pure accelerator cell, write the average size at birth in terms of the average size at initiation

\[ \langle l_{b}^{acc} \rangle = r \langle l_{i}^{acc} \rangle + 2r\Delta_{id}^{acc}. \]  

(6.16)

For simplicity, consider the case without an extra round of initiations so that \( O = 2 \). As discussed above, an accelerator cell inherits a fraction \( r \) of the size accumulated between initiation
and division toward the size $\Delta_{i_{ii}}^{acc}$ required for the next initiation. In other words,

$$2r\Delta_{i_{id}}^{acc} + (\Delta_{i_{id}}^{acc} - \Delta_{i_{b}}^{acc}) = \Delta_{i_{ii}}^{acc}. \quad (6.17)$$

The size at next initiation is therefore

$$l_{i_{next}}^{acc} = rl_{i_{i}}^{acc} + \Delta_{i_{ii}}^{acc}. \quad (6.18)$$

Stationarity then implies that the average size at initiation of a pure accelerator cell is

$$\langle l_{i_{i}}^{acc} \rangle = \Delta_{i_{ii}}^{acc} / (1 - r). \quad (6.19)$$

This expression allows us to express $\langle l_{b_{i}}^{acc} \rangle$ in terms of model parameters $\Delta_{i_{id}}^{acc}$, $\Delta_{i_{ii}}^{acc}$, and $r$.

The same reasoning leads to the average size at birth for a pure alternator cell, but with $r$ replaced by $1 - r$ and accelerator increments replaced by alternator increments, giving

$$\langle l_{b_{i}}^{acc} \rangle = 2r\Delta_{i_{id}}^{acc} + \frac{r}{1 - r}\Delta_{i_{ii}}^{acc}, \quad (6.20)$$

and similarly for alternator cells.

The parallel adder converges because of the same principle as the adder-per-origin model (e.g., convergence occurs because of the constant growth increment between initiation events per origin). We show by simulation that cell size converges even when the growth increment between initiations is larger for accelerator cells than alternator cells (Figures 6.33 and 6.5f). We illustrate the convergence of both pure accelerator and alternator lineages by simulation (Figure 6.5i). In the largest subpopulation possible (the pure accelerator cells), the lineages from a large cell and a small cell each converge to $2r\Delta_{i_{id}}^{acc} + (r / (1 - r)) \Delta_{i_{ii}}^{acc}$. Similarly, the lineages of
alternator cells (the smallest subpopulation) converge to \(2(1 - r) \Delta l_{id}^{alt} + ((1 - r) / r) \Delta l_{ii}^{alt}\). These simulations did not include any fitting parameters and verify convergence the pure accelerator and alternator subpopulations. Because the accelerator and alternator subpopulations each converge and are the subpopulations with the longest and shortest cells, respectively, the entire population must also converge.

The population averaged cell size at birth can be derived as in the adder-per-origin model. In this case, each of the four types of target size at initiation leads to the following target size at division

\[
l^y_d = l^y_i + \Delta l^y_{id}.
\]

Here, the superscripts \(xy\) denote a cell of type \(y\) whose mother is of type \(z\). Because these four types exist in equal numbers in the population, the population averaged cell size at initiation and at birth can be approximated as \(\langle l_i \rangle = \sum_y \langle l^y_i \rangle / 4\) and \(\langle l_b \rangle = \sum_y \langle l^y_d \rangle / 8\). Simplification leads to the result quoted above \(\langle l_b \rangle = (\Delta l_{id} + \Delta l_{ii}) / 2\) and confirmed in simulations (Figure 6.5f).

The parallel adder model also predicts that the mean growth increment from birth to initiation \(\langle \Delta l_{bi} \rangle\) can be calculated from both of the constant growth increments

\[
\langle \Delta l_{bi} \rangle = \langle l_i \rangle - \langle l_b \rangle = \Delta l_{ii} - \frac{\Delta l_{ii} + \Delta l_{id}}{2} = \frac{\Delta l_{id} - \Delta l_{ii}}{2}.
\]

Because the B period is so short in rich growth conditions (< 4% of the length of the cell cycle, on average), we evaluated the increment during the B period in slow (carbon limited) conditions. The average normalized growth increment in B, \(0.24 \pm 0.03\) 95% CI, was very similar to the value (0.22) calculated using the parallel adder model. The growth increment during B is correlated with birth length because the longer accelerator cells inherit more of the mother’s cell body, and therefore more of the accumulated length since the last initiation.
Simulation of the parallel adder model predicts a slope of 0.28 (normalized to average $l_b$) for $\Delta l_b$ versus $l_b$, in agreement with (and within the 95% confidence interval of) our measurement of $0.24 \pm 0.09$ 95% CI (Figure 6.4b).

6.3.3 Numerical simulations

We use numerical simulations to compare division adder, adder-per-origin, and parallel adder models to measurements. We consider sources of stochasticity that are important to describe measurements: a time-additive noise between successive initiations with standard deviation $\sigma_t$, an asymmetric and noisy division ratio with mean $r$ and standard deviation $\sigma_r$, and a noisy growth rate with mean $\lambda$ and standard deviation $\sigma_\lambda$. Noise in cell cycle timings is implemented as a time-additive noise onto the size required at initiation. In the division adder, a size-additive noise gives size and time distributions of slightly different shapes, but does not change the correlation coefficients compared to a time-additive noise. We therefore focus on time-additive noise onto the size required at initiation, so that

$$\left(l_i^\text{total}\right) = \left(l_i^t + O\Delta l_i^t\right) e^{\lambda\xi_t}.$$

At division, a cell divides according to a noisy ratio $r + \xi_r$, where $r$ is the average asymmetry ratio. The newborn cell is also given a noisy growth rate $\lambda + \xi_\lambda$, where $\lambda = \ln 2/\tau$ is the average growth rate. $\xi_t$, $\xi_r$, $\xi_\lambda$ are random variables with zero mean and standard deviations $\sigma_t$, $\sigma_r$, $\sigma_\lambda$. Because all other parameters can be extracted, there are only two free parameters in the parallel adder model: $\sigma_t$ and $\sigma_\lambda$.

Simulations begin with one accelerator and one alternator cell, although the initial condition does not matter because the average cell size achieves stationarity. The simulations advance in time for twelve generations and track all cells in the population. Initiation and division
events are dictated by the models. The simulations also mimic measurement errors by adding to the recorded cell size at birth a random noise with standard deviation $\sigma_e$. Measurement errors are assumed to be multiplicative in size, so that the noise has magnitude $2\sigma_e$ at division. The single-cell statistics of the population are used to calculate the coefficients of variation and Pearson correlation coefficients, which we compare to measurements. In particular, because $\sigma_t$, $\sigma_\lambda$, and $\sigma_e$ are fitting parameters, we can determine the best fit by minimizing the residue between the measured and simulated coefficients of variation of cell size at birth, at division, and interdivision time.

Because measurements were taken at discrete time intervals (every 15 min for *M. smegmatis*, every 45 min for BCG), the average cell size at division is slightly less than twice the average size at birth. This affects the mean behavior when comparing the measured correlation between $v_d - v_i$ and $v_i$. Hence, the simulated size at divisions are multiplied by a corrective factor equal to $\langle v_d \rangle / \langle v_b \rangle / 2$. Note that this correction does not affect the correlation coefficient.

To test if experimental imaging frequencies were too long to compare to model simulations, we sampled model simulations in which correlation coefficients were collected at limited time resolution, ranging from 0%-30% interdivision time (Figures S3F–S3H). We observe noticeable changes in correlation coefficients when time resolution is over 25% of the cell cycle, whereas *M. smegmatis* experiments in rich medium have a time resolution of 8% interdivision time, *M. smegmatis* experiments in carbon limited medium have a time resolution of 5% interdivision time, and BCG experiments have a time resolution of 4% interdivision time. Because experimental resolutions are much shorter than the resolution threshold obtained from model simulations, we conclude that experimental imaging frequencies do not affect the evaluation of the parallel adder model against experimental data.
In this chapter, I present published and unpublished results on the molecular mechanism underlying cell cycle regulation in bacteria. I first introduce the problem, then present published and unpublished results on conceptual molecular models. I then provide a progress report of unpublished results on constructing a realistic molecular level model.

7.1 Introduction

Bacteria can divide faster than their chromosomes replicate by maintaining multiple ongoing rounds of DNA replication. The parallel nature of the bacterial cell cycle is in direct contrast to the sequential checkpoints of the eukaryotic cell cycle, and complicates the search for a molecular understanding of the phenomenon. However, it is timely now to seek a molecular
level understanding because recent advances in microfluidics and microscopy enable accurate measurements of single-cell quantities across the cell cycle. For *E. coli* growing under conditions where cells divide faster than their chromosomes replicate, single-cell measurements suggest that cells on average add a constant size per origin of replication from one event of initiation of DNA replication to the next (Chapter 4)\(^3\). This “adder” behavior must be generated by the molecular mechanisms underlying initiation, severely constraining possible solutions. Below, we outline my preliminary results and next steps to construct a molecular model for the bacterial cell cycle.

To construct a realistic molecular level model for the initiation of DNA replication, we first investigated conceptual molecular models, in which the identity of the molecular players are not specified\(^9\). For example, one simple model we considered is the initiator accumulation model, in which the accumulation of a threshold number of an initiator protein triggers initiation. Another plausible model is the inhibitor dilution model, in which the dilution past a threshold concentration of an inhibitor protein triggers initiation. We do not consider the molecular identity of the initiator or the inhibitor within conceptual molecular models. Within conceptual molecular models of the bacterial cell cycle, we found that the initiator accumulation model robustly generates adder behavior, while the inhibitor dilution model does not\(^14\).

We then proceeded to construct realistic molecular level models for putative molecular players. We considered two important proteins, DnaA and SeqA. A threshold amount of the former is necessary for initiation, while a high concentration of the latter blocks initiation. However, neither the initiator accumulation nor the inhibitor dilution model alone can describe the many existing experiments that perturb DnaA or SeqA concentrations. First, the initiator accumulation model with DnaA as the initiator predicts that the average cell size at initiation, or initiation size, should be inversely proportional to the DnaA concentration. However, experiments that increased the DnaA concentration found that the initiation size remained
unchanged. Second, experiments that perturbed SeqA concentrations also found that the initiation size barely changed. These and other results from the past five decades suggest that initiators and inhibitors work simultaneously to regulate the cell cycle.

Intriguingly, the molecule DnaA might act as both an initiator and an inhibitor. This is because DnaA exists in two forms, either ATP- or ADP-bound, and while both forms bind to the origin of initiation, only the ATP bound form triggers initiation. We therefore sought to construct a realistic molecular level model to describe how the levels of the two forms of DnaA are regulated during the cell cycle to capture both the puzzling perturbation experiments at the bulk level and the adder behavior at the single-cell level. The model must take as input the growth rate of the cell, which is specified by the environment, and must produce as output oscillations in the initiator and inhibitor concentrations to trigger initiation at a frequency matching exactly the growth rate. This must hold under perturbations to the expression levels of initiators and inhibitors, the number of origins in the cell, and the growth rate. Below, we present preliminary results towards constructing a realistic molecular level model for replication initiation.

7.2 Conceptual molecular models

In this section, I present some of the results in the published paper, Ref. 14. The article discusses both bacteria and budding yeast. Here, I focus on the bacterial aspects.

7.2.1 Adapted abstract

Organisms across all domains of life regulate the size of their cells. However, the means by which this is done is poorly understood. We study two abstracted “molecular” models for size regulation: inhibitor dilution and initiator accumulation. We apply the models to bacteria like
*E. coli*, which grow fully before they set a division plane and divide into two equally sized cells. We investigated whether the models robustly led to cells that add a constant size before they divide, a behavior known as the “adder” that is experimentally observed for several species of bacteria. We find that an inhibitor dilution model produces adder correlations that are not robust to noise in the timing of DNA replication initiation or in the timing from initiation of DNA replication to cell division (the C+D period). In contrast, an initiator accumulation model yields robust adder correlations in the regime where noise in the timing of DNA replication initiation is much greater than noise in the C+D period. We therefore conclude that it is unlikely that bacteria displaying adder behavior use an inhibitor dilution strategy to regulate their cell size, since such a strategy does not generate adder correlations that are robust to noise. Instead, an initiator accumulation model is consistent with the experimentally observed adder behavior, provided that cells grow in the regime where noise in their timing of DNA replication initiation is much greater than noise in the time from initiation of DNA replication to cell division.

### 7.2.2 Background

Organisms across all domains of life regulate their cell size, as reviewed in Chapter 2. However, the molecular mechanisms of cell size control, and the intimately related cell cycle control, remain unclear despite decades of research. Two longstanding models which connect cell size with cell cycle progression are the initiator accumulation and the inhibitor dilution models. These models occur at the conceptual level, where the molecular identity of the initiators and inhibitors are not specified. The initiator accumulation model considers a protein whose activity triggers the initiation of DNA replication. We consider a model where a threshold amount of initiators triggers initiation. The inhibitor dilution model considers instead a protein whose activity inhibits initiation. We consider the case where initiation is triggered below a threshold
concentration of inhibitors. If the production and degradation of initiators and inhibitors are
coupled to cell size, the two models bring together cell size and cell cycle control at the single-
cell level. For example, we consider an inhibitor dilution model where the inhibitor is produced
only within one part of the cell cycle, and where DNA replication in the subsequent cell cycle
initiates only when the inhibitor concentration has been diluted through new growth to the
threshold concentration.

Compared to when the above models were first proposed, recent advances in microfluidics
and microscopy have enabled measurements of single-cell statistics that allow better exploration
between cell size regulation models at the phenomenological level and cell cycle regulation
models at the molecular level. In particular, recent works have shown that microbes from all
domains of life share a common size regulation strategy where a constant cell size is added
from birth to division, known as the “adder”\textsuperscript{6,7}. More precisely, an organism is an adder if
its single-cell correlations between the size increment from birth to division and the size at
birth is zero. The adder behavior is found is several species of bacteria, in budding yeast,
and in an archaeon, as discussed in Chapters 2-3. In this work, we aim to use the generality
and robustness, elaborated below, of the adder phenomenon to investigate which conceptual
molecular models are more likely to describe the underlying molecular mechanism of cell cycle
regulation in bacteria. Although the initiator accumulation model has recently been shown to
simultaneously regulate of both cell size and the number of origins of replication, there is no
known evidence that definitively excludes either an initiator accumulation or inhibitor dilution
model from consideration.

We considered specific implementations of the initiator accumulation and inhibitor dilution
models and investigate the single-cell correlations they generate. As noted in Chapters 2-3, the
adder phenomenon is a robust observation, in the sense that multiple labs, using a variety of
environments, have observed the phenomenon. We therefore sought to construct a model where
the adder phenomenon will be generated regardless of the sources or magnitudes of biological noise within the model. Moreover, since there can be multiple relevant noise terms, we evaluated our models in a large region of parameter space relevant to experimental conditions. Within the space, we calculated the slope of a linear regression between the cell size at birth and at division as a test for deviations from adder behavior. Measurements reported a variation of ±0.1 in the slope between size at birth and division across a selection of different growth media. We therefore defined a model to be consistent with the adder phenomenon if it generates a slope of $1 \pm 0.1$. We further defined a model to be robust if it generates the adder behavior across broad ranges of parameter space rather than in fine-tuned pockets. In other words, we assume that if cells required strongly coupled noise strengths for diverse noise sources to behave as adders, it would be unlikely to observe adder behavior over a variety of growth conditions. Our approach is consistent with previous studies such as that presented in Chapter 4, which showed that an initiator accumulation model generates adder behavior provided that noise in initiation of DNA replication is much greater than noise in the duration of the C+D period. Our approach allowed us to infer the viability of different strategies of size regulation as follows.

7.2.3 Models

The inhibitor dilution model assumes that the initiation of DNA replication occurs upon the dilution of an inhibitor molecule below a critical concentration. The inhibitor is assumed to be synthesized only in the period following the initiation of DNA replication. The model can be described as

$$V_i = \frac{(I_b + \xi)}{c_i}$$  \hspace{1cm} (7.1)

$$I_d = I_b + \Delta.$$  \hspace{1cm} (7.2)
Here, $V_i$ is the volume at initiation, $I_b$ and $I_d$ are the number of inhibitors at birth and division, respectively, and $\Delta$ is the amount of inhibitor synthesized during that cell cycle. $\xi$ is a noise term that captures the stochasticity in the initiation of DNA replication, with standard deviation $\sigma_\xi$. The subscripts $b$, $i$, $d$ indicate the value of the variable at birth, initiation, and division, respectively. We note that $c_1$ has the effect of setting the scale of the average cell size in combination with $\langle \Delta \rangle$, but will not affect the correlation between volume at birth and division. We assume that at division, the inhibitors are distributed to the two progenies according to their relative volumetric fractions. To describe the distribution of $\Delta$, we will consider two variants for the synthesis of the inhibitor. The “noisy synthesis rate” synthesis model assumes that the inhibitor is produced at a rate $K$ drawn from a normal distribution with mean $\langle K \rangle$ and standard deviation $\sigma_K$. The inhibitor is produced between initiation and termination of replication. Since replication terminates a time $C$ after initiation, $\Delta = KC$ in this case. The “noisy integrator” synthesis model assumes that inhibitor production is uncorrelated with growth, so that $\Delta$ is drawn from a normal distribution with mean $\langle \Delta \rangle$ and standard deviation $\sigma_\Delta$.

The initiator accumulation model assumes that a cell initiates DNA replication upon the accumulation of a sufficient amount of some initiator protein. The initiator is synthesized during cell growth, such that a volume increment of $\delta V$ leads to a newly synthesized amount $\delta A = \delta V/c_2$ of initiator protein. Here, $c_2$ is a scaling factor with units of concentration that sets the scale of the size distribution in a similar manner to $c_1$. As in the inhibitor dilution model, we assume that at cell division the initiator protein is distributed to both progenies according to their relative volumetric fractions. This process can be described by

$$A_d = c_2 (V_d - V_i)\quad (7.3)$$
$$V_i = V_b - A_b/c_2 + A_c/c_2,\quad (7.4)$$
where \( A_b \) and \( A_d \) are the initiator abundance at birth and division, respectively, and \( A_c \) is the critical amount of initiator required for initiation. We assume that \( A_c \) is drawn from a normal distribution with mean \( \langle A_c \rangle \) and standard deviation \( \sigma_t \). To derive Eq. 7.3, we have made the assumption that initiators are degraded entirely after initiation. The behavior of the model depends on this assumption, an unpublished result that we discuss below. Eq. 7.4 comes from setting the number of initiators at the subsequent initiation event (i.e., the sum of the number of initiators at birth and new initiators produced through growth, \( A_b + c_2 (V_i - V_b) \)) equal to \( A_c \). This model is a simplified case of that previously proposed for fast-growing bacteria, where we now restrict the maximum number of DNA replication forks and initiation events per cell cycle to one.

For both models, we assume that cells grow exponentially with rate \( \lambda \) and that cells divide symmetrically to generate two progenies of equal size. We further assume that volume at division is related to volume at initiation by

\[
V_d = V_i \exp(\lambda T),
\]

and that the cell cycle period \( T \equiv C + D \) is uncorrelated with the generation time \( t_d \). In fast-growing bacteria the presence of multiple replication forks means that the division event prompted by an initiation of DNA replication may not take place until later cell cycles. In this case, Eq 7.5 refers to the total volume at the division event prompted by that round of DNA replication initiation, but we do not consider this scenario here.

7.2.4 Results

The dilution models, with either the noisy synthesis rate or the noisy integrator assumptions, do not give robust adder correlations. For simplicity, we focus on the case of perfectly sym-
metric division in slow growing bacteria, in which the C+D period is shorter than the doubling
time. The inhibitor dilution model has the additional requirement that initiation and division
must occur alternatively. This restriction complicates analytical calculations. Nevertheless,
simulations of two variants of the inhibitor dilution model show that the slope is sensitive to
both noise in the initiation threshold and noise in the C+D period, implying that the inhibitor
dilution models considered do not produce robust adder behavior (Figure 7.1).

On the other hand, the accumulation model can give robust adder behavior. Since rapidly
growing bacteria maintain multiple ongoing rounds of DNA replication, we consider an accumu-
lation model in which a constant volume per origin of replication is added between replication
initiation events. This model can allow for an extra round of replication initiation late in the
cell cycle, through stochastically accumulating a threshold number of initiators before division.
The simultaneous regulation of DNA replication and cell division allows the model to robustly
recover from these stochastic events\(^{13}\).

Here, we derive an analytical expression for the slope \(S(V_b, V_d)\) between size at birth and at
division, under the simplifications that cells undergo perfectly symmetric division and that cells
do not undergo extra rounds of replication initiation. The slope can be written as a normalized
covariance,

\[
S(V_b, V_d) = \frac{\langle V_b V_d \rangle - \langle V_b \rangle \langle V_d \rangle}{\langle V_b^2 \rangle - \langle V_b \rangle^2}.
\] (7.6)

The size at birth can be written in terms of the size at the previous DNA replication initiation,
\(V_b = V_i \exp(\lambda(T + \xi_T))/2\), where \(\lambda\) is the noiseless growth rate, \(T = C + D\), and \(\xi_T\) is a
Gaussian random variable with standard deviation \(\sigma_T\). We can then write \(S(V_b, V_d)\) in terms
of \(V_i\) as

\[
S(V_b, V_d) = 2 \left( \frac{\langle V_i V'_i \rangle - \langle V'_i \rangle^2}{a^2 \langle V_i^2 \rangle - \langle V_i \rangle^2} \right),
\] (7.7)

where \(V'_i\) is the size at the next DNA replication initiation, and \(a = \langle \exp(\lambda\xi_T) \rangle\). Since
Figure 7.1: The inhibitor dilution model is not robust for the symmetrically dividing bacterial mode of growth. Heat maps of linear regression slopes from fitting $V_b$ vs $V_d$ for symmetrically dividing bacterial cells. The models simulated are variants of an inhibitor dilution model in which the amount of inhibitor synthesized is (a,b) uncorrelated with the time spent post DNA replication initiation (noisy integrator), or (c,d) equal to a constant rate multiplied by the time spent post DNA replication initiation (noisy synthesis rate, with $\sigma_K = 0$). Black outlines provide a guide to the eye for regions in which adder-like behavior is observed (slope = 1.0 ± 0.1). Adder behavior is seen to be sensitive to noise strength. This indicates that the dilution model is unlikely to be implemented as a means of size regulation in symmetrically dividing bacteria which display adder behavior.
\(\xi_T\) is a Gaussian random variable, \(a = \exp\left(\lambda^2 \sigma^2_T/2\right)\). Note that factors of \(\exp(\lambda T)\) in the numerator and in the denominator cancel. Similarly, we can relate \(V_i'\) to \(V_i\) by writing \(2V_i' = V_i + (\langle A_c \rangle + \xi_i)/c\), where \(\xi_i\) is a Gaussian random variable with standard deviation \(\sigma_i\). We have used the simplifications that cells undergo perfect symmetric division and that cells do not undergo extra rounds of replication initiation. Substituting into the expression for the slope, we find after simplification

\[
S(V_b, V_d) = \frac{1}{a^2 + 3\left(\frac{a^2-1}{b}\right)}, \tag{7.8}
\]

where \(b = \sigma_i/\langle A_c \rangle\). To lowest order in small variables \(\sigma_i/\langle A_c \rangle\) and \(\lambda \sigma_T\), the expression becomes \(S(V_b, V_d) \approx 1/(1 + 3\lambda^2 \sigma^2_T/b^2)\). Hence if \(\sigma_i/\langle A_c \rangle \gg \lambda \sigma_T\), the slope approaches one, as confirmed by simulations. For comparison with the inhibitor dilution case, Figure 7.2 considers the case of slow-growth \((C+D < \tau)\). However, Eq. 7.8 and its derivation both hold for the fast-growth case \((C+D > \tau)\) as well. The approximate Eq. 7.8 deviates from numerical results only when the fraction of cells undergoing extra initiations becomes significant at \(\sigma_i/\langle A_c \rangle \gtrsim 0.3\) (Figure 7.3). This is a biologically unrealistic regime since experiments show that \(E. coli\) has \(\sigma_i/\langle A_c \rangle \approx 0.1\) and \(\sigma_T/\tau \approx 0.1\), where both Eq. 7.8 and simulations predict the observed adder behavior.

In summary, we observed that achieving adder behavior in a symmetrically dividing inhibitor dilution model requires fine-tuning of noise in the C+D period. This leads us to conclude that such a model is unlikely to be biologically relevant, in light of the robust adder correlations observed. In contrast, a symmetrically dividing initiator accumulation model is robust provided that noise in the C+D period of the cell cycle is smaller than noise in DNA replication initiation. An initiator accumulation model also allows simultaneous regulation of the number of origins of replication. Since there currently exists no definitive demonstration of a particular
Figure 7.2: The initiator accumulation model is robust for the symmetrically dividing bacterial mode of growth, provided that $\sigma_i/\langle A_c \rangle \gg \sigma_T/\tau$. Heat maps of linear regression slopes from fitting $V_b$ vs $V_d$ for symmetrically dividing bacterial cells. The model simulated is the initiator accumulation model. Black outlines provide a guide to the eye for regions in which adder-like behavior is observed (slope = 1.0±0.1). In the limit of $\sigma_i/\langle A_c \rangle \gg \sigma_T/\tau$ the observed slopes approach 1, consistent with experimental observations.

molecular mechanism of size control in bacteria, the finding that an inhibitor dilution model in bacteria is not robust to noise may prove useful in narrowing the range of possible molecular size control mechanisms. As such, future studies should therefore be focused around determining the molecular candidates for an initiator activation mechanism. This finding showcases the efficacy of cell cycle correlations in providing a connection between phenomenological and molecular models of size regulation.

7.3 Additional analyses of the initiator accumulation model

In this section, I present unpublished results on additional analyses of conceptual level model. First, I discuss an important consideration in investigating the scaling relation between cell size and growth rate at the single-cell level. I then compare the adder-per-origin model to an existing experiment to distinguish it from other competing models. Finally, I discuss how
Figure 7.3: Comparison of the approximate expression (Eq. 7.8) and numerical results for the linear regression slopes of $V_b$ vs $V_d$ in the initiator accumulation model for symmetrically dividing bacterial cells. Circles plot numerical results. Dashed lines plot Eq. 7.8. Colors denote $\sigma_T/\tau$ as indicated in the legend.

resetting the initiator affects the behavior of the initiator accumulation model.

7.3.1 Single-cell correlations between cell size and growth rate

As discussed in Chapter 5, classic, bulk level experiments showed that the average cell size in an exponentially growing population scales exponentially with growth rate $\lambda$ in different media,

$$\langle v \rangle \propto e^{\lambda T}, \quad (7.9)$$

where the scaling exponent $T$ is equal to the time between initiation and replication$^{12,18}$. This observation led to the hypothesis that cells initiated replication on average at a constant size per origin of replication$^{19}$. Eq. 7.9 follows since if cells initiated replication on average at a constant size $\Delta$ per origin, then they will divide at size $\Delta e^{\lambda T}$. Hence, the average cell size in an exponentially growing population will be $\langle v \rangle = \log(2) \Delta e^{\lambda T}$ (see Chapter 5 for a detailed discussion).

The same reasoning shows that the exponential scaling of cell size with growth rate should also hold at the single-cell level with slight modifications. More precisely, the size at birth of a cell will scale with its growth rate in the generation during which the corresponding round of
replication was initiated. This is because at the single-cell level, cell size at division is

\[ v_d = v_i \exp \left( \lambda_i \left( T - \sum_j^n t_j \right) + \sum_j^n \lambda_j t_j \right), \tag{7.10} \]

where the subscript \( i \) denotes the generation during which corresponding round of replication was initiated, and \( t_j \) is the generation time for the \( j \)-th of \( n = \lfloor T/\tau \rfloor \) succeeding generation until division. Here, the floor operator, \( \lfloor \rfloor \), returns the largest integer less than the input. Eq. 7.10 concerns the coupling between initiation and division, and holds for any mode of regulation between initiations. As an example of Eq. 7.10, if \( 2\tau > T > \tau \), then there is typically one generation between a round of initiation and the corresponding division. Averaging Eq. 7.10 over growth rates \( \lambda_i \), the scaling relation at the single-cell level is

\[ \langle v_b \rangle = \Delta \exp (\lambda_i T) \exp (\nu \tau (\langle \lambda \rangle - \lambda_i)). \tag{7.11} \]

Testing Eq. 7.11 requires experiments that follow a cell for several generations, and could be a sensitive test of the nature of the coupling between DNA replication and cell division at the single-cell level\(^{100}\).

### 7.3.2 Comparison of the initiator accumulation model to existing experiments

As discussed in Chapter 5, bulk level experiments led to the following model for cell cycle regulation at replication initiation\(^{19, 28}\). The model supposes that cells initiate replication at a constant size per origin, and divided a constant time after initiation. We will refer to this hypothesis as sizer-per-origin, which can be summarized as

\[ v_i \approx O\Delta, \tag{7.12} \]
where \( v_i \) is the cell size at initiation, \( O \) is the number of origins, and \( \Delta \) is the size-per-origin required for initiation. Notably, the thresholding mechanism in the sizer-per-origin model abrogates correlations between a round of initiation and preceding events. It is therefore unlikely that sizer-per-origin can generate the observed adder behavior at divisions. To reconcile the results at the single-cell and the bulk level, the adder-per-origin model was recently proposed, as discussed in Chapter 4. The model supposes that cells add a constant size per origin of replication from one round of initiation to the next. It can be summarized as

\[
v'_i \approx v_i + O\Delta,
\]

where \( v'_i \) is the total size of the daughter cells, typically two, at the next initiation event. Intuitively, if the fluctuations in growth rates and in the intervals from initiations to the corresponding divisions are both small, then the adder-per-origin model produces \emph{emergent} adder behavior at divisions. Below, we compare the two competing models to recent experiments published in Ref. 28. In Ref. 28, initiation events were determined via fluorescently labeled DnaQ (a DNA polymerase) for cells in three media corresponding to slow, intermediate, and fast growth. Table 7.1 summarizes the single-cell statistics obtained.

The best-fit models to all the listed statistics can be obtained as follows. First, both the sizer- and adder-per-origin models take as parameters the ratio between the average doubling time \( \tau \) and the average interval \( T \) from initiation to the corresponding division, as well as the magnitudes \( \sigma_x \) of fluctuations in cell cycle parameters, where \( x \) can be \( \lambda \), \( T \), or \( t \), representing respectively the growth rate, the interval from initiation to the corresponding division, and the interval between initiations. All of these except \( \sigma_t \) were measured, and \( \sigma_t \) can be extracted from the coefficient of variation of generation times or of cell size at birth. Both models were extended to incorporate the additional detail that \( T \) depends on \( \lambda \) at the single-cell level\textsuperscript{28}. 

148
This observation has significant implications, particularly at slow growth, regarding the coupling between cell division and DNA replication, discussed in the Outlook section below. In summary, both models must explain seven statistics with no fitting parameters.

Table 7.1 shows that adder-per-origin captures measurements better than sizer-per-origin. This is particularly evident in the correlation coefficient $C(v_b, v_i)$ between cell size at birth and at initiation. Adder-per-origin predicts a higher correlation coefficient $C(v_b, v_i)$ than sizer-per-origin. This is because in adder-per-origin, the size $v_i^{\text{previous}}$ at the previous initiation determines both $v_b$ and $v_i$ through the interval $T$ and the size-per-origin requirement $\Delta$, respectively. On the other hand, $v_i^{\text{previous}}$ determines only $v_b$ in sizer-per-origin since $v_i$ is determined solely by a threshold mechanism. The positive correlation between $v_b$ and $v_i$ in slow growth even in sizer-per-origin is due to the scaling of $T$ with $\lambda$ so that the ratio $T/\tau$ is close to one at slow growth. This increases the probability of a second round of initiation late in the cell cycle, at about twice the typical cell size at initiation, and artificially increases the correlation between $v_b$ and $v_i$. Nevertheless, the claim that adder-per-origin leads to a larger $C(v_b, v_i)$ than does sizer-per-origin remains true.
7.3.3 Resetting the initiator

The initiator accumulation model proposes that an initiator protein accumulates during cell growth to trigger the initiation of DNA replication upon reaching a threshold copy number. Here, we consider the model in a simplified picture without the coupling between DNA replication and cell division, and instead, the accumulation of a threshold copy number \( \theta \) of some divisor protein triggers division (Chapter 8). At division, a cell will have accumulated up to \( y_d = \theta + \xi_\theta \) divisors, where \( \xi_\theta \) is a coarse-grained noise term with magnitude \( \sigma_\theta \). A biologically relevant question is whether divisors are degraded following division. For example in \textit{E. coli}, the protein DnaA, a threshold number of which is necessary for the initiation of DNA replication, is inactivated (i.e. hydrolyzed, see section below) following initiation to prevent immediate re-initiations. If divisors are symmetrically segregated at division, and a fraction \( 0 \leq s \leq 1 \) of them are not degraded following division, then the number \( y'_b \) of divisors following division is

\[
y'_b = sy_d/2.
\]  

(7.14)

Furthermore, if divisors are accumulated at the same rate as volume growth, then the change in the number of divisors \( \Delta y \) is equal to the change in volume \( \Delta v \), up to some proportionality constant that can be set to unity. Hence, if divisions are perfectly symmetric, the cell in question will be born next with size

\[
v'_b = (v_b + \theta + \xi_\theta - y_b + \xi_v)/2,
\]  

(7.15)

where \( \xi_v \) is a noise term with magnitude \( \sigma_v \) that captures fluctuations in volume growth.

Analysis of Eq. 7.14-7.15 reveals that as long as the magnitude of the fluctuations in volume growth is larger than that in divisor threshold \( (\sigma_v \gg \sigma_\theta) \), the details of the deactivation
of divisors does not affect the resulting cell size regulation. For example, if $s = 1$, there is no deactivation of divisors. This results in a sizer strategy ($C(v_b, v'_b) \approx 0$) when $\sigma_v \ll \sigma_\theta$. However, when $\sigma_v \gg \sigma_\theta$, the size regulation strategy becomes an adder ($C(v_b, v'_b) \approx 1/2$) (Figure 7.4).

In fact, when $\sigma_v \gg \sigma_\theta$, $C(v_b, v'_b) \approx 1/2$ for any $0 \leq s \leq 1$. The intuition is that when $\sigma_v \gg \sigma_\theta$, the correlations in cell size induced by fluctuations in the divisor threshold are overwhelmed by fluctuations in volume growth, and the only correlations that remain are due to the fact that volume is added from birth to division. In other words, the broad regulatory architecture of the accumulation of a required protein up to a threshold can produce robust adder behavior, regardless of model details such as the dynamics of accumulation and degradation of the division limiting protein.

### 7.4 Molecular level models

In bacteria, the DNA binding protein DnaA exists in either the ATP or ADP bound form\textsuperscript{98-99}. Both forms can bind to the origin of replication, or the $oriC$ locus, but only the ATP bound
form can trigger the initiation of DNA replication. Both forms also bind to other DnaA binding sites along the chromosome. In particular, the *datA* locus near the origin can bind many DnaA molecules with high affinity. There are approximately two thousand DnaA molecules in an *E. coli* cell growing under conditions supporting a doubling time of approximately thirty minutes. Also, there are about ten times more ATP than ADP in an *E. coli* cell, so that newly synthesized DnaA are more likely to bind to ATP than to ADP. DnaA-ATP is converted to DnaA-ADP via several processes, the most important of which is the catalysis of the hydrolysis of DnaA-ATP to DnaA-ADP via enzymes recruited by ongoing replication forks. This process is known as the regulatory inactivation of DnaA, RIDA. On the other hand, DnaA-ADP is converted to DnaA-ATP in a process dependent on DnaA reactivation sequences, *dars*. The *dars* loci can bind DnaA-ADP molecules and promote the dissociation of ADP from DnaA, which then binds to the abundant free ATP molecules in the cell. Genetic perturbations of these processes affect the average cell size at initiation as well as the synchrony of initiations from the multiple origins within the cell, suggesting that the dynamics of the interconversion between the two forms of DnaA regulate the timing of initiation. For a detailed review, see Ref. 99.

We sought to construct a mathematical model with realistic features to describe the interconversion dynamics of DnaA. To do so, we took the approach of constructing the model step by step, beginning from the simplest working model, namely the initiator accumulation model discussed above. The initiator accumulation model makes the following assumptions. First, it assumes that the initiator molecule is produced at a rate proportional to volume growth. Second, it assumes that initiation is triggered upon the accumulation of a threshold amount of initiator per origin. Finally, it assumes that following initiation, the initiator is degraded instantaneously to zero. Our approach is to relax the assumptions one by one towards more realistic portrayals of the underlying processes. At each step, we will consider both existing experiments on the biochemical details of the various processes and the robustness of the models.
in face of various sources of stochasticity and perturbations.

The model we seek must agree with existing experiments in several ways. First, the model must achieve cell size and cell cycle homeostasis. That is, the model must result in an average time between initiations that is equal to the doubling time of volume growth. Second, the model must generate single-cell correlations similar to adder correlations. Although it remains unclear whether the adder-per-origin model precisely describes the correlations between initiation events in *E. coli*, a recent work suggests that a constant size increment is indeed added between initiation events in *E. coli*\(^{100}\). Analysis of existing experiments, as discussed in the section above, also supports this conclusion. Finally, the model must satisfy the above two criteria while generating interconversion dynamics in line with existing measurements. Ref. 102 showed that approximately 80\% of DnaA molecules exist in the ATP bound form at initiation, and that the percentage decreases to approximately 50\% some time after, before rising again to trigger the next round of initiation. We do not require the model to reproduce exactly the measured trajectories of interconversion, but rather we require the model to generate trajectories where the two forms of DnaA are approximately of the same abundance at some point during the cell cycle.

We first relaxed the second assumption that a threshold amount of initiator per origin is required for initiation. Although the per origin requirement leads to cell size and cell cycle homeostasis, it is not consistent with experiments that showed that doubling the number of origins does not change the average cell size at initiation\(^8\). Furthermore, the hypothesis that the locus *datA* may be the “origin” in question, as discussed in Chapter 5, is not consistent with experiments that showed that deletion of *datA* does not affect the average cell size at initiation except under special conditions\(^9\). We therefore considered other possibilities for the threshold trigger for initiation. Since only the ATP bound form triggers initiation, we considered DnaA-ATP to be the initiator molecule. Moreover, since neither the number of origins nor the number
of the datA locus are the sole determinant of initiation, we considered all the DnaA binding sites along the chromosome to play the role of a titrating mechanism. Taken together, we considered that a threshold number of DnaA-ATP per DNA triggers initiation\textsuperscript{103}.

We note that it might be illustrative to further relax the threshold trigger assumption by constructing a biophysical model including the dynamics of the binding and unbinding of DnaA to and from the chromosome. Within such a model, initiation would be triggered once the origin is mostly bound by the ATP bound form. Such a model can be used to investigate how multiple origins within the cell can initiate synchronously and how the interconversion dynamics might affect simultaneously the stochasticity between rounds of initiations and the asynchrony within a round of initiation\textsuperscript{99}. Here, we focus first on constructing a model for the interconversion dynamics.

To specify the dynamics of DnaA-ATP, we relaxed the assumption of immediate degradation of initiators to zero following initiation by modeling the RIDA process. We assumed that DnaA-ATP is converted to DnaA-ADP at a rate proportional to the number of replication forks as the simplest approximation. That is, if $A$ and $B$ denotes the number of DnaA-ATP and -ADP molecules, respectively, then

\[
\frac{dA}{dt} = r_A V - \frac{r_B [A]}{d_B + [A]} F \tag{7.16}
\]

\[
\frac{dB}{dt} = \frac{r_B [A]}{d_B + [A]} F. \tag{7.17}
\]

Here, $V$ is the cell volume, $F$ is the number of replication forks in the cell, square brackets $[\cdot]$ denote concentrations, $r_A$ and $r_B$ are the rates of production of DnaA and conversion of DnaA-ATP to -ADP, respectively, and $d_B$ is the concentration at which the RIDA rate begins to saturate with respect to $[A]$. $F$ is also equal to the difference between the number of origins
and the number of termini of replication. Eqs. 7.16-7.17 are supplemented by

\begin{align}
\frac{dV}{dt} &= \lambda V \\
\frac{dL}{dt} &= \frac{1}{C}F \\
\frac{A_i}{L_i} &= \theta,
\end{align}

(7.18) (7.19) (7.20)

where \(\lambda\) is the exponential growth rate of the cell, \(C\) is the time required to complete DNA replication, and \(\theta\) is the threshold level of \(A/L\) that triggers initiation. The subscript \(i\) denotes the value of the variable at initiation. Finally, division is triggered a time \(D\) after the termination of the corresponding round of replication. Together, Eqs. 7.16-7.20 describes the deterministic dynamics DnaA interconversion, as well as volume growth and DNA replication, that should lead to a “steady-state” where the cellular variables form a limit cycle with a period equal to the doubling time \(\tau = \ln 2/\lambda\).

We calibrated the values of the various rates and saturating constants as follows. There are on the order of 2000 DnaA molecules in a cell growing in rich nutrient conditions supporting a doubling time of approximately 30 minutes. Therefore, since the majority of DnaA molecules exist in the ATP bound form at initiation, \(A_i \sim 2000\). This result calibrates the value of \(\theta\), since the amount of DNA at initiation \(L_i\) is known given \(\tau\), \(C\), and \(D\), and can be written as \(L_i = \left(\sum_j F_j t_j\right)/C\), where \(j\) denotes segments of the cell cycle with differing numbers of replication forks. Furthermore, within Eqs. 7.16-7.17, the total number of DnaA grows at a rate proportional to volume, so that at steady-state, \(r_A = \lambda ([A] + [B])\). This result calibrates the value of \(r_A\), if we assume that volume is scaled to the order of unity. \(d_B\) is first assumed to be small, so that the RIDA rate is simply linear in \(F\) unless \(A\) is close to zero, in which case the RIDA rate becomes small, ensuring that \(A\) does not become negative. Finally, \(r_B\) can be treated as a free parameter and the behavior of the model can be explored as a function of \(r_B\).
We found that although Eqs. 7.16-7.20 can lead to a steady cell cycle, with the time between initiations equal to the doubling time, they cannot generate the experimentally observed interconversion dynamics, namely that the percentage of ATP bound DnaA varies from approximately eighty percent to approximately fifty percent. This can be seen by solving Eqs. 7.16-7.17 assuming that \([A]\) is always larger than \(d_B\). Since the DnaA molecules are not degraded, only interconverted, \(B_i\) is the sum of the linear RIDA rates, or \(B_i = r_B \sum_j F_j t_j\). In this case, \(B_i\) is proportional to \(L_i\). Therefore, if we require \(A_i > f_B B_i\), we must have \(\tilde{r}_B < (1/f_1)\), where \(\tilde{r}_B = C r_B / \theta\) is a dimensionless RIDA rate, and \(f_1\) describes how many more fold of DnaA molecules are ATP bound than ADP bound at initiation. On the other hand, if we require a fraction \(f_2\) of \(A_i\) to be deactivated after initiation, we must also have \(\tilde{r}_B > f_2 \left( \sum_j F_j t_j / \sum_j F_j t_j \right)\), where \(j'\) denotes segments of the cell cycle during which the percentage of the ATP bound form is decreasing. Since the quantity in the parenthesis on the RHS must be less than one, we must have \(f_1 f_2 < 1\), which cannot generate the experimentally observed interconversion dynamics that requires \(f_1 \sim 4\) and \(f_2 \sim 1/2\). Intuitively, the model fails because \(\dot{B}\) is always positive, so that the percentage of the ADP bound form can only decrease via dilution through the production of new DnaA molecules, which is not fast enough to generate the sharp reactivation seen in experiments.

Reactivation of the ADP to the ATP bound form is therefore required. We introduced the effects of the \(dars\) loci by modifying Eqs. 7.16-7.17 to include a reactivation rate analogous to the RIDA rate,

\[
\frac{dA}{dt} = r_A V - \frac{r_B [A]}{d_B + [A]} F + \frac{r_C [B]}{d_C + [B]} G \tag{7.21}
\]

\[
\frac{dB}{dt} = \frac{r_B [A]}{d_B + [A]} F - \frac{r_C [B]}{d_C + [B]} G \tag{7.22}
\]
Here, $G$ is the number of $dars$ loci within a cell, $r_C$ is the reactivation rate, and $d_C$ is a saturation constant. The $dars$ locus is assumed to be some fraction $f_C$ of the chromosome away from the origin, so that $G$ is doubled a time $f_CC$ after initiation. We assumed $f_C$ to be 0.5 to approximate the case in $E. coli$. The specific value of $f_C$ does not affect our qualitative conclusions. Experimentally, the values of the RIDA and reactivation rates are not known, due to the difficulty of measuring the number of ATP versus ADP bound forms in vivo. We therefore scanned the parameter space of $r_B$ and $r_C$ to find the region of parameter space that leads to a correct steady-state (Figure 7.5a). Using the same reasoning as above for Eqs. 7.16-7.17, it can also be shown that for Eqs. 7.21-7.22, the requirement $A_i > B_i$ implies $r_C \gtrsim r_B$. Together, these two criteria locates a region of parameter space in which the model leads to the correct cell cycle regulation and generates experimentally observed interconversion dynamics (Figure 7.5ab).
Figure 7.5: A working model at the realistic molecular level for the timing of the initiation of DNA replication in *E. coli*, Eqs. 7.18-7.20 and 7.21-7.22. (a) A parameter space scan for the model with \( \tau = 40 \), \( C = 45 \), and \( D = 25 \) minutes. For the \((r_B, r_C)\) pair marked, the model achieves cell cycle homeostasis (brown), and also generates interconversion dynamics approximating those observed (pink). The pair in blue marks the values used in panels b-e. (b) An example of the deterministic interconversion dynamics generated by the model. Blue and red lines denote the number of DnaA-ATP and -ADP in the cell. Black solid line marks the time of initiation. Purple, yellow, and green lines mark the time of termination of replication, replication of the *dars* loci, and cell division. (C) An example of the interconversion dynamics under a time-additive noise to the time of initiation. The time of initiation is the deterministic time plus a normally distributed random variable with standard deviation \( \sigma / \tau = 0.15 \), a magnitude that generates size and generation time distributions with realistic widths. The second and third initiation cycle shown have fluctuations in the order of cell cycle events, from which the model robustly recovers. (d) The model generates correlations between the cell size at birth and that at the next generation similar to those in an adder. (e) The model at slow growth, with \( \tau = 80 \) minutes and the rest of the parameters the same as for panels (a-d), generates no correlations between the cell size at birth and that at the next generation. In panels d-e, blue squares show the average values binned according to the x-axis.
Figure 7.5 (Continued).
We next examined the behavior of the model under biological noise, implemented as a time-additive, coarse-grained noise on top of the deterministic time between initiations, as introduced in Chapter 2. In this case, the model generates single-cell correlations approximately like adder correlations, suggesting that the resetting mechanism can be described at a more detailed level by deactivation and reactivation (Figure 7.5cd). Indeed, like the adder-per-origin model, the model here is robust to fluctuations in the order of cell cycle events, such as initiation and termination of replication, caused by the time-additive noise (Figure 7.5c). It is instructive to consider the case where $\tau > C$. In this case, there exists a period after the termination of replication without ongoing replication, and thus, without RIDA. During this period, all DnaA will be in the ATP bound form, in which case our model reduces to an initiator accumulation model with a constant concentration of the initiator. In other words, our model predicts that cell size at initiation will obey a sizer strategy, and that the correlation coefficient between the cell sizes of a cell and its progeny is zero (Figure 7.5e). Experimentally, the correlations at birth is indeed closer to zero at slow growth, as discussed in the section above. However, since the correlations at birth may be affected by other sources of stochasticity, including noise in the C+D period, the exact prediction at initiation remains to be tested experimentally.

The model proposed above is therefore a preliminary working model, since it can capture some of the existing experiments. However, the model also fails in several ways. First, the model may not capture the experimental result that doubling the gene copy number of dnaA does not affect the average cell size at initiation\(^{96}\). Within the model, doubling the gene copy number of dnaA corresponds to doubling the value of $r_A$, which would lead to a decreased cell size at initiation. Although existing experiments concerning manipulating DnaA expression levels differ in results\(^{97}\), the above discrepancy between experiments and our model suggests that the model may be oversimplified. For example, it is known that DnaA represses its own expression\(^{99}\). The effect of such autorepression could be accounted for by replacing $r_A$ with
\( r_A/(d_A + [A]) \), where the constant \( d_A \) describes the concentration at which DnaA begins to autorepress. Second, as in Ref. 14, we require the model here to be robust in a large region of the parameter space, rather than discrete pockets. However, the criteria for \( r_B \) and \( r_C \) depend sensitively on cell cycle timings, and only a small pocket of the parameter space allow the model to capture existing experiments. We therefore suspect that a more robust mechanism is at play, which might be implemented by biochemical features of the processes that we ignored. For example, RIDA is known to depend on the protein Hda, whose expression is correlated with that of DnaA. This effect might be described by a higher-order term in the RIDA rate. Similarly, reactivation by the \( dars \) loci is known to depend on the protein Fis, whose activity appears to somehow depend on cell cycle progression. Together, these features might confer a robust adaptation mechanism to overcome the need to fine-tune the parameters of the model.
A mechanistic model for the regulation of the timing of cell division by the circadian clock in the cyanobacterium *Synechococcus elongatus*

8.1 Summary

The circadian clock regulates the timing of cell divisions in cyanobacteria such that cells tend to divide away from darkness\(^5^6\), but the mechanism underlying how the clock regulates division timing remains unclear. Here, we developed a mechanistic model in which a protein limiting for division accumulates at a rate proportional to cell volume growth but modulated by the clock.
Our “modulated rates” model, in which the clock signal is integrated over time to affect division timing, differs fundamentally from the previously proposed “gating” concept, where the clock is assumed to act instantaneously to affect division timing\textsuperscript{104,105}. We found that the modulated rates model better reproduce experimentally observed single-cell statistics of division timing in the cyanobacterium \textit{Synechococcus elongatus}. We also found that the modulated rates model robustly places divisions away from darkness during changes in the environment, whereas the gating model does not. Finally, existing data analyzed within the framework of the modulated rates model is consistent with the simple mechanism that the accumulation of a division limiting protein in phase with the promoter activity of the core clock proteins regulates division timing.

8.2 Results

8.2.1 The growth and division of \textit{S. elongatus} cells

\textit{S. elongatus} cells possess a circadian clock in the form of a group of core clock proteins whose concentrations and phosphorylation states oscillate with a period of approximately 24 hours under constant conditions. The activity of the core clock proteins affects many cellular processes, including volume growth and cell division\textsuperscript{106}. Furthermore, \textit{S. elongatus} cells grow and divide only under light, and do not grow nor divide under darkness\textsuperscript{56,107}. However, it remains unclear how the clock mechanistically affects the timing of divisions so that cells do not divide during darkness. We therefore sought to construct a mathematical model for the mechanism that takes as inputs the environmental light conditions and the internal circadian clock to give the division timing as output (Figure 8.1).

To do so, we analyzed data from Ref. \textsuperscript{56}, which observed the growth and division of single cells for a wildtype strain and a strain without the core clock proteins, referred to here as the clock-deletion strain. Cells were grown and imaged under constant light (LL) or periodic
cycles of light and darkness (12:12 LD, i.e. 12 hours of light with a graded intensity profile, below, followed by 12 hours of darkness, and 16:8 LD). For each cell, its length at birth $l_b$ and division $l_d$ and its generation time (the time between birth and division) $t_d$ were measured. Before imaging, the cells were grown under 12:12 LD to entrain and synchronize the activity of the clock to the environmental light conditions. The circadian phase $\theta$ corresponding to the internal subjective time of day encoded by the clock can then be assumed to be set to the environmental light-dark cycle. We therefore defined $\theta = 0$ to be dawn, or the beginning of the period under light. Each cell can then be assigned a circadian phase at birth $\theta_b$. We analyzed the distributions of and correlations among the four stochastic variables ($l_b$, $l_d$, $t_d$, $\theta_b$), and compared these statistics of division timing to those generated by our models.

The growth mode of a bacterium is an important determinant of its division timing. The growth mode of $S.\ elongatus$ can be approximated to be exponential, with a rate dependent on the environmental light intensity\textsuperscript{56,108}. We therefore model the growth of volume $V$ as

$$\dot{V} = \lambda(\theta) \, V. \quad (8.1)$$

The dot denotes the time derivative. The growth rate $\lambda(\theta)$ may depend on the light intensity, which is a function of $\theta$ for the periodic environments under consideration. Under LL, $\lambda(\theta)$ can be approximated as constant for our purposes, although in reality it varies up to $\approx 5\%$ with the circadian phase\textsuperscript{56}. Under LD, $\lambda(\theta)$ is approximately proportional to the environmental light intensity. The experimental light intensity profile is sinusoidal during the period under light. We therefore model $\lambda(\theta)$ as

$$\lambda(\theta) = \begin{cases} 
\lambda_0 \sin \left( \frac{\pi \theta}{T_L} \right) & \theta < T_L \\
0 & \theta \geq T_L 
\end{cases}. \quad (8.2)$$
where \( \lambda_0 \) is the maximum growth rate and \( T_L \) is the duration of the period under light. For our analyses, we use cell volume and cell length interchangeably since volume can be well approximated as proportional to cell length in rod-shaped bacteria that grow by elongation such as \( S. \ elongatus \). Experimentally, cells divide symmetrically with small fluctuations in the division ratio. We assume perfectly symmetrical divisions in our models.

8.2.2 Divisor accumulation with degradation can describe division timing in a clock-deletion strain

The clock-deletion strain under LL behave, with minor deviations, like several other bacteria whose cells appear to add a constant size from birth to division (Figure 8.2a). Inspired by models that sought to describe such single-cell correlations, we considered the following model. The basic component of our model is the accumulation of a protein limiting for division, whose amount is denoted by \( X \), at a rate proportional to volume growth,

\[
\dot{X} = \dot{V}.
\]  

(8.3)

A threshold amount of these divisors triggers division. Divisors are consumed during division so that the amount of divisors is zero at birth, denoted by \( t = 0 \). That is,

\[
X(t = 0) = 0, \quad (8.4)
\]

\[
X(t = t_0) = 1. \quad (8.5)
\]
Figure 8.1: Two models for the regulation of division timing by the circadian clock. Both models take as inputs (a) the environmental light-dark cycles (yellow shade) and a coupling function (green line) dependent on the internal circadian phase to give as outputs (b) the single-cell distributions of and correlations among cell length at birth $l_b$ and division $l_d$, the circadian phase at birth $\theta_b$, and the generation time $t_d$. Shown is an experimentally observed distribution of $\theta_b$ for *S. elongatus* under periodic conditions, showing that divisions occur away from dawn and dusk. (c) (Middle) The basic model without the clock, Eq. 8.3. The divisor is accumulated at a rate proportional to volume growth. (Left) The modulated rates model, Eq. 8.8. The divisor accumulation rate is modulated by the clock. (Right) The gating model, Eq. 8.10. The divisor accumulation rate is not affected by the clock, but only a fraction of divisors, determined by the clock, is active towards reaching the threshold.
Here, $t_0$ is the noiseless generation time. We model the stochasticity in division timing as a coarse-grained time-additive noise, so that the actual generation time $t_d$ is

$$t_d = t_0 + \sigma \xi,$$  \hspace{1cm} (8.6)

where $\xi$ is a normal random variable with zero mean and unit standard deviation, and $\sigma$ is the magnitude of the time-additive noise. The above approach of simple deterministic dynamics plus simple noise terms has been successfully applied to understand cell cycle regulation in other microbes\textsuperscript{5,9,10,15}. In Eq. 8.3 and Eq. 8.5, we chose the proportionality constant between $\dot{X}$ and $\dot{V}$ as well as the threshold amount to be one because we are interested in the statistics of division timing independent of the absolute numbers.

The model is numerically simulated as follows. The deterministic generation time was determined by numerically integrating the equations for the accumulation of divisors, Eq. 8.3 or its various extensions below. The generation time is obtained via Eq. 8.6. Cell volume is calculated according to Eqs. 8.1-8.2 and is divided in half at division. The process is repeated for at least $10^5$ generations, following only one of the newborn cells at division.

The basic model predicts no correlations between $l_d - l_b$ and $l_b$, but experiments showed a slight negative correlation (Figure 8.2a). This discrepancy does not affect the rest of the results because the effects of the clock on division timing are more significant. The basic model does not contain a clock, and as such, $t_d$ is independent of $\theta_b$, as observed in experiments (Figure 8.2b).

The basic model presented above with the experimental light profile predicts that the distribution of circadian phases at birth $p(\theta_b)$ is non-zero near dawn, in contrast to the experimental observation that $p(\theta_b)$ appears to vanish there (Figure 8.2ce). A simple explanation within our framework is that the divisors degrade, so that cells do not have enough divisors to divide.
immediately after dawn. This explanation could be describing, for example, FtsZ, a protein required for cell division that degrades rapidly in *E. coli*. Degradation is assumed to occur at a rate $\delta$, so that

$$\dot{X} = \dot{V} - \delta X. \quad (8.7)$$

The best fit value of $\delta$ was determined by fitting to $p(\theta_b)$ and the distribution of generation times $p(t_d)$ (Methods). We found that a degradation rate $\delta \approx \ln 2 / (28 \text{ h}) \approx 0.025 \text{ h}^{-1}$ best describes the clock-deletion strain under 16:8 LD (Figure 8.2c-f). However, degradation rates corresponding to half-lives on the timescale of a day still lead to qualitatively similar distributions (Section 8.4). Existing experiments therefore do not allow a precise determination of $\delta$.

One confounding factor is that the shape of $p(\theta_b)$ is sensitive to the details of the ensemble, such as whether statistics were collected over a lineage of cells or over a growing population of cells (Section 8.4). Nonetheless, the model describes well the qualitative features of the statistics, such as when divisions first begin to occur after dawn (Figure 8.2ce).

$p(t_d)$ under LD are bimodal because some cells divide before reaching a period of darkness (short-generation cells), whereas other cells must wait through a period of darkness before division (long-generation cells). The average volume doubling time compared to $T_L$ (Eq. 8.2) largely determines the fractions of short- and long-generation cells, while degradation increases the difference between the mean generation times of short- and long-generation cells (Figure 8.2df). This is because, compared to the case without degradation, the long-generation cells now have even longer generation times due to the degradation of accumulated divisors during darkness. The long-generation cells must now compensate for the degraded divisors, and therefore will be larger at division (Section 8.4). The model with degradation better describes the data than one without degradation with a significance level smaller than 0.01 by likelihood analysis (Methods). Overall, degradation is a simple mechanism that can capture with one
parameter the qualitative features of division timing in the clock-deletion strain.

8.2.3 Divisor accumulation with modulated rates can describe division timing under a circadian clock

For the wildtype strain under LL, the clock generates correlations between $\theta_b$, $l_b$, and $t_d$ that cannot be captured by the above model (Figure 8.3ab). To describe these correlations, we considered a modulated rates model where the rate of accumulation of the divisor is modulated by a periodic function $y(\theta)$,

$$\dot{X} = \dot{V}y(\theta) - \delta X. \quad (8.8)$$

$y(\theta)$ could be describing, for example, the approximately sinusoidal promoter activity of FtsZ\textsuperscript{110}. We therefore assumed the following sinusoidal form,

$$y(\theta) = 1 + A \left( \cos \left( \omega t - \frac{\pi \varphi}{12} \right) - 1 \right), \quad (8.9)$$

where $\omega = 2\pi / (24 \text{ h})$, and $A$ and $\varphi$ are the magnitude and phase offset of the modulation. We chose $y(\theta)$ to have a maximum of one, since only the variations, and not the mean value, of $y(\theta)$ affect the statistics of division timing. The best fit values of $A$ and $\varphi$ can be determined by fitting to the various distributions and correlations (Methods). We determined $A$ and $\varphi$ for each condition, reflecting the fact that the molecular players that implement $y(\theta)$ may depend on the environmental light conditions, as discussed in detail below. We also assumed that clocks are entrained quickly relative to the duration of the experiments, so that the measured statistics approximate those for cells entrained under imaging conditions. In other words, we assumed that the data for the 16:8 LD imaging conditions contained even though the actual
Figure 8.2: Divisor accumulation with degradation can describe division timing in a clock-deletion strain. The correlations (a,b) and distributions (c-f) of the stochastic variables as defined in Figure 8.1. \langle \cdot \rangle denotes the average over all single-cells. Blue denotes data from Ref. 56. Red lines denote predictions of the basic model with degradation, Eq. 8.7. Yellow lines denote predictions of the basic model without degradation, Eq. 8.3. (a,b) The clock-deletion strain under LL. Small transparent points represent single-cell data. Large solid points are averages binned according to the x-axis. (c-f) The clock-deletion strain under (c,d) 16:8 LD or (e,f) 12:12 LD. Yellow shade shows the shape of the light intensity profile. Table 8.2 contains the parameter values used.
protocol entrained the cells under 12:12 LD for even for the 16:8 LD imaging condition.

Despite its simplicity, the model can capture the correlations between $t_d$ and $\theta_b$ (Figure 8.3a). The model without further adjustments also captures the correlations between $l_d - l_b$ and $l_b$ (Figure 8.3b), which is more negative than in the clock-deletion strain. This is because, for the doubling time relevant here, cells that are larger at birth often have just grown through periods of repressed accumulation, and will therefore tend to grow through periods of derepressed accumulation. The size increments between birth and division of larger cells will therefore be smaller than average (Section 8.4). The model also captures the other statistics of division timing. In particular, because the statistics were not collected over lineages but over growing populations, $p(\theta_b)$ is not the same as the distribution of circadian phases at division $p(\theta_d)$, where $\theta_d$ is defined as $(\theta_b + t_d) \mod 24$. The model captures both distributions after taking into account the experimental details of the ensemble (Section 8.4). Moreover, the model also captures correlations between distantly related cells such as the cousin-cousin correlations between generation times (Section 8.4). The model can also describe the division timing of the wildtype strain under LD (Figure 8.3c-f). In summary, divisor accumulation with modulated rates, Eq. 8.8, is a model with two free parameters ($A$ and $\varphi$) that can describe the statistics for the division timing of wildtype S. elongatus under both constant and periodic environments (Figure 8.3).

8.2.4 Distinguishing between the modulated rates and the gating models

The modulated rates model, in which the clock signal is integrated over time to affect division timing, differs fundamentally from the widely considered gating hypothesis, which assumes that the clock acts instantaneously to affect division timing$^{104, 105}$. The difference leads to qualitatively different predictions for how the two models respond to changing environments as follows.
Figure 8.3: Divisor accumulation with modulated rates or gating can describe division timing. Figure styles and axes labels are the same as in Figure 8.2. Here, red lines denote predictions of the modulated rates model, Eq. 8.8, and purple the gating model, Eq. 8.10. (a,b) The wildtype strain under LL. (c-f) The wildtype strain under (c,d) 16:8 LD or (e,f) 12:12 LD. Table 8.2 contains the parameter values used.
We formulated the gating model on top of the structure of divisor accumulation with degradation, Eq. 8.7. However, instead of Eq. 8.8, the gating model supposes that only a fraction \( y(\theta) \) of the accumulated divisors is active in contributing to reaching the threshold. That is,

\[
\tilde{X} = Xy(\theta),
\]  

(8.10)

where \( \tilde{X} \) is the amount of active divisors. A threshold amount of active divisors triggers division,

\[
\tilde{X}(t = t_0) = 1.
\]  

(8.11)

All other aspects of the gating model are the same as the modulated rates model. In the gating model, \( y(\theta) \) could be describing, for example, the approximately sinusoidal concentrations of the core clock proteins. We therefore considered again the case where \( y(\theta) \) is sinusoidal as in Eq. 3.15. The gating model can also capture the qualitative features of the statistics of division timing (Figure 8.3). However, the modulated rates model better describes data compared to the gating model with a significance level of less than 0.01 by likelihood analysis (Methods). Qualitatively, the two models differ in several aspects.

First, the best fit values of \( \varphi \) suggest that the effect on division timing by the clock is implemented by different molecular players in the two models. For the modulated rates model, a simple interpretation is that \( y(\theta) \) describes the promoter activity of the divisor. In this case, the concentration of the divisor lags some time, depending on \( \lambda \) and \( \delta \), behind the promoter activity (Section 8.4). For the values of \( \lambda \) and \( \delta \) in the experiments we analyzed, the concentration lags approximately 5 hours behind. The best fit value of \( \varphi \) therefore predicts that the divisor concentration peaks approximately 12 hours after dawn under 12:12 LD (Figure 8.4a). Experiments have observed that the concentrations of the core clock proteins peaks 14 ± 1 hours after dawn under similar conditions\textsuperscript{108,111,112}. The approximate agreement between the two
suggests that within the modulated rates model, \( y(\theta) \) is implemented by molecular players that are in synchrony with the core clock proteins. For the gating model, a simple interpretation is that \( y(\theta) \) describes the concentration of some effector since its effects are instantaneous. In this case, the best fit values of \( \varphi \) in the gating model imply that the concentrations peak 8 hours after dawn under 12:12 LD (Figure 8.4a). Therefore, within the gating model, \( y(\theta) \) would be implemented by molecular players that are not in synchrony with the core clock proteins, in contrast with the modulated rates model.

Second, the best fit values of \( \varphi \) are more parsimoniously interpreted in the modulated rates model. Experiments have shown that for different values of \( T_L \) (Eq. 8.2), the concentrations of the core clock proteins shift in circadian phase such that the phase at the peak increases by \( T_L/2 \). Consistent with this observation, the best fit value of \( \varphi \) under 16:8 LD is two hours more than that under 12:12 LD in the modulated rates model. We note the caveat that the experiments in Ref. 112 were done with on-off light intensity profiles without the sinusoidal dependence used in Ref. 56. Also in the modulated rates model, the best fit value of \( \varphi \) under LL is the same as that under 12:12 LD, consistent with the fact that the clock was entrained under 12:12 LD (Figure 8.4a). In contrast, the best fit value of \( \varphi \) in the gating model under LL is five hours different from that under 12:12 LD, inconsistent with the fact that the clock was entrained under 12:12 LD.

The differences between the two models in predictions involving \( \varphi \) may arise from the difference between integrating a signal over time, and acting on the signal instantaneously. To see this, the gating model can be rewritten as

\[
\dot{\tilde{X}} = \dot{\tilde{V}}y - \delta \tilde{X} + \tilde{X}y, 
\]

(8.12)

which is equivalent to the modulated rates model with an extra term \( \tilde{X}y \). When the effects of
degradation are small compared to that of the clock, as is the case here, the extra term $X \dot{y}$ approximately scales like $\dot{V} y$. In this case, the rate of divisor accumulation is modulated by both $y$ and the derivative of $y$. The form of the extra modulation may explain how both models can describe existing experiments, albeit with qualitatively different predictions involving the best fit values of $\varphi$ (Section 8.4).

Finally, the modulated rates model does not predict divisions during darkness without growth, whereas the gating model can lead to divisions during darkness without growth. The latter case occurs when enough divisors have accumulated, but not enough are active according to the gating function $y(\theta)$. Divisions can then occur just by the passage of time, without cell growth, and the consequent activation of divisors with increasing $y(\theta)$. The above scenario can be demonstrated in a numerical simulation using the best fit parameters for 16:8 LD, and tracking the division events for cells entrained under 16:8 LD but imaged for one cycle where the light is turned off abruptly during the day. The gating model predicts that a noticeable fraction of cells will divide during darkness in this scenario, whereas the modulated rates model predicts no divisions during darkness (Figure 8.4b). This difference could be relevant for cells in nature facing fluctuations in environmental light intensity\textsuperscript{107}, for example caused by clouds blocking the sun. The experimental realization of the scenario could be one way to directly differentiate the two models.

8.3 Discussion

How cyanobacteria regulate division timing has been studied for decades, but how and why the clock regulates division timing remain unclear\textsuperscript{32,56,104,108}. One widely considered mechanism is that of gating, where the signal from the clock is assumed to affect division timing instantaneously\textsuperscript{104,105}. Here, we proposed a different mechanism of modulated rates, where the
Circadian phase at peak (h)

<table>
<thead>
<tr>
<th>Related experiments</th>
<th>Modulated rates</th>
<th>Gating</th>
</tr>
</thead>
<tbody>
<tr>
<td>LL</td>
<td>14 ± 1</td>
<td>12</td>
</tr>
<tr>
<td>16:8 LD</td>
<td>16 ± 1</td>
<td>14</td>
</tr>
<tr>
<td>12:12 LD</td>
<td>14 ± 1</td>
<td>12</td>
</tr>
</tbody>
</table>

Figure 8.4: Distinguishing between the modulated rates and the gating models. (a) The models predict different molecular players to implement the effects on division timing by the clock. The Table shows the circadian phase at the peak of the concentrations of core clock proteins measured in related experiments, and determined from best fit values of \( \varphi \) in the two models. See main text for more details. (b) Predictions of the modulated rates (red) and gating (purple) models entrained under LD and imaged for one cycle where the light is turned off abruptly during the day. Yellow shade shows the light profile during the imaging cycle, and the yellow dashed line shows the profile during entrainment. Green dotted line shows the \( y(\theta) \) used in the gating model. Table 8.2 contains the parameter values used.
signal from the clock is integrated over time to affect division timing. The gating model could correspond to a post-translational mechanism while the modulated rates model could correspond to a transcriptional mechanism, for example. We formulated a simple framework that describes how cell volume growth, the environmental light profile, and the internal circadian clock together determine division timing. We showed that the modulated rates model appears to be more consistent both quantitatively and qualitatively with existing experiments. Our analysis suggests the simple mechanism that the accumulation of a division limiting protein in phase with the promoter activity of the core clock proteins regulates division timing. Suggestively, FtsZ is approximately synchronous with the core clock proteins. Our models differ from existing ones in both formalism and structure. Ref. 108 modeled the relation between the progression of the circadian phase and that of division timing with a general nonlinear map. Ref. 32 studied a kicked cell cycle model in which the generation time is determined by a linear combination of the previous generation time and an oscillatory function of the circadian phase. The above approaches did not consider the feedback of cell size on division timing. However, for exponentially growing cells such as those of *S. elongatus*, timing divisions without feedback from cell size fails to maintain a homeostatic average cell size. Ref. 56 accounted for the effects of cell size regulation by modeling the instantaneous probability to divide as a function of cell size multiplied by a periodic coupling function of the circadian phase. The approach of Ref. 56 can accurately describe the measured statistics of division timing, but it is difficult to parametrize the coupling function to gain an understanding of the underlying molecular mechanism. Our models specify division timing via simple deterministic dynamics and implement stochasticity via a coarse-grained noise term. The simplicity provides mechanistic insights by describing how the clock affects division timing via only two parameters.

To obtain more insights into the mechanism, one direction is to experimentally characterize
the mechanistic underpinnings for the few parameters in our models. For example, what is the mechanistic basis for the magnitude of the modulation function, and what are its effects on the average cell size\textsuperscript{105} and division timing? Single-cell statistics of division timing in more environmental conditions may help illuminate how the modulation function responds to environmental light intensity.

We also seek to understand the fitness benefits and costs of regulating division timing by the clock. Serial dilution experiments have shown that wildtype \textit{S. elongatus} fixates in the population over clock-deletion mutants only under LD, and that neither fixates over the other under LL\textsuperscript{113}. The latter result implies that their biomass growth rates are the same\textsuperscript{29}, at least under LL, and suggests that the clock-deletion mutants lost under LD partly because of the regulation of division timing by the clock. Another potential fitness effect is that cells may face penalties to divide near darkness\textsuperscript{107}. It may therefore be constructive to investigate what mechanisms, in addition to the concept of modulated rates, robustly place divisions away from darkness during changes in the environment\textsuperscript{114}. In summary, our work might offer a simple and illustrative modeling framework to investigate how and why the clock regulates division timing.

8.4 Methods

8.4.1 Determination of the best fit values for model parameters

The best fit value for $\delta$ in Eq. 8.7 was determined as follows. For a given $\delta$, $\sigma$ was chosen to match the standard deviation of cell lengths at birth. Then, the best fit value for $\delta$ for the clock-deletion strain under LD was chosen to minimize the sum of squared residues between model predictions and experimental observations for the distributions of circadian phases at birth $p(\theta_b)$ and generation times $p(t_d)$. Minimizing similar distance measures, such as the Bhattacharyya distance and the Jensen-Shannon divergence, gave similar results. The above
procedure was chosen to capture the important qualitative features of the statistics of division timing, namely when during the day divisions begin to occur and stop occurring, and the mean generation times of short- and long-generation cells. The best fit values for $A$ and $\varphi$ for the wildtype strain under LD were determined similarly. For the wildtype strain under LL, the best fit values were chosen to minimize the sum of squared residues for the correlations between $t_d$ and $\theta_b$, binned according to $\theta_b$. This procedure was chosen because the shape of $p(\theta_b)$ is particularly sensitive to the details of the statistical ensemble in this case, whereas the correlations are not (Section 8.4). Table 8.2 summarizes the best fit values of all parameters obtained.

The goodness of fit of the models and the errors on the best fit values of the model parameters can be compared and estimated, respectively, by calculating likelihood functions. To do so, we assumed that each division event is independently distributed, so that the likelihood of obtaining the observed data points given a model is proportional to the product of the probabilities to obtain each point. For each comparison, we calculated the likelihoods of obtaining the observed $p(\theta_b)$ and $p(t_d)$ given the models. The ratio than gives the significance level, which is the probability of incorrectly choosing the better model. For comparisons between models (degradation versus no degradation, and modulated rates versus gating), the best fit values of the model parameters were used. To compare the model with degradation and without, we accounted for the extra parameter that is the degradation rate using the Akaike information criterion. To estimate the errors on the best fit values of the model parameters, we scanned values away from the best fit values for the parameter in question, while holding all other parameters constant, until a significance level smaller than 0.01 is obtained. We determined error estimates to the tenth decimal place for $A$, and to the hour for $\varphi$ and for the half-life corresponding to $\delta$. 

179
Figure 8.5: The deterministic behavior of the models. Model predictions with best fit values for all parameters (Table 8.2) under LL unless otherwise specified. All curves were obtained by setting $\sigma = 0$. (a,b) Comparison of the deterministic behavior of the modulated rates (red) and gating (purple) models. (c,d) The deterministic behavior of the modulated rates model with different values of $A$ (light, medium, and dark red).

8.4.2 Properties of the model

In this section, we summarize how the magnitude $A$ of the modulation function, the degradation rate $\delta$, and the modulation function $y(\theta)$ captures the various features of how the clock affects division timing. First, the best fit $y(\theta)$ for the two models produces similar deterministic behaviors, obtained by setting $\sigma = 0$ (Figure 8.5ab). The similar behaviors arise because the gating model can be approximated as a modulated rates model with a similar modulation function, as shown in the Discussion. Increasing $A$, or the coupling from the clock to division, leads to sharper correlations (Figure 8.5cd).

Degradation increases the difference between the mean generation times of short- and long-generation cells as follows. Compared to the case without degradation, the long-generation
cells now have even longer generation times due to the degradation of accumulated divisors during darkness. The long-generation cells must now compensate for the degraded divisors, and therefore will be larger at division. Cells with the largest \( t_d \) are therefore also the largest cells at division (Figure 8.7a). Although degradation is necessary to describe existing experiments within our basic framework, the experiments can be described by a large range of degradation rates that correspond to half-lives on the timescale of a day (Figure 8.6).

Coupling division timing to the clock introduces correlations between cell size and division timings. This is because, for the doubling time relevant here, cells that are larger at birth often have just grown through periods of repressed accumulation, and will therefore tend to grow through periods of derepressed accumulation. The size increments between birth and division of larger cells will therefore be smaller than average (Figure 8.7b).

Finally, within the modulated rates model, the concentration of the divisor, \( x = X/V \), lags some time behind the promoter activity described by \( y(\theta) \). The lag duration can be estimated by ignoring divisions and calculating the resulting concentration, which gives

\[
x \propto (\lambda + \delta) \cos(t') + \omega \sin(t') ,
\]

(8.13)
where $t' = \omega t + \pi \phi/12$. Therefore, if $(\lambda + \delta) \gg \omega$, then $x \propto y$ and the promoter activity and the concentration are approximately synchronous. If $(\lambda + \delta) \ll \omega$, then the concentration lags a quarter of a period behind. For the values of $\lambda$ and $\delta$ here, the concentration lags approximately 5 hours behind the promoter activity.

8.4.3 The effects of the statistical ensemble

The details of the experimentally recorded ensemble may significantly affect the reported statistics of division timing. The recorded cells may be sampled from the entire population tree or from a few lineages, from experiments that begin with synchronized or asynchronized cells, or from experiments that end at different times during the day. Such details significantly affect the shape of $p(\theta_b)$ under constant light conditions. For statistics from a single lineage, the two distributions $p(\theta_b)$ and $p(\theta_d)$, where $\theta_d = (\theta_b + t_d) \mod 24$, are exactly the same, since the circadian phase at division is simply the circadian phase at birth for the next generation in the lineage. However, for synchronized cells in a growing population, the two distributions could be drastically different depending on when the experiments end, since the cells are synchronized...
Figure 8.8: The details of the statistical ensemble affect the distribution of circadian phases at birth. The measured (blue histograms) and predicted (red lines) distributions of circadian phase at birth (dark) and division (light) under LL (a) and 16:8 LD (b). The predictions were obtained by simulating the modulated rates model, using the best fit values in Table 8.2, for a growing population of synchronized cells for the duration of the experiment.

<table>
<thead>
<tr>
<th></th>
<th>Wildtype strain</th>
<th>Modulated rates</th>
<th>Model without clock</th>
<th>Clock-deletion strain</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_{md}$</td>
<td>$-0.25 \pm 0.12$</td>
<td>$-0.46$</td>
<td>$-0.27$</td>
<td>$-0.07 \pm 0.13$</td>
</tr>
<tr>
<td>$\rho_{ss}$</td>
<td>$0.62 \pm 0.10$</td>
<td>$0.68$</td>
<td>$0.31$</td>
<td>$0.41 \pm 0.11$</td>
</tr>
<tr>
<td>$\rho_{cc}$</td>
<td>$0.46 \pm 0.10$</td>
<td>$0.56$</td>
<td>$0.21$</td>
<td>$0.29 \pm 0.13$</td>
</tr>
</tbody>
</table>

Table 8.1: Generation time correlations between mother-daughter, sister-sister, and cousin-cousin pairs ($\rho_{md}$, $\rho_{ss}$, $\rho_{cc}$, respectively). The errors on the experimental values report 95% confidence intervals. Table 8.2 contains the parameter values used.

and will tend to divide nearby in time. The modulated rates model, when simulated at the population level while taking into account the recording protocol, reproduces $p(\theta_b)$ and $p(\theta_d)$ in qualitative agreement with experimental observations (Figure 8.8a). The above effect is not present in the qualitative features under LD that are important for our purposes, namely when during the day divisions begin to occur and stop occurring (Figure 8.8b).

Cousin-cousin correlations

Correlations in generation times along cell lineages have been used to study systems from cyanobacteria to cancer cells$^{31,32,115}$. In particular, the relations between the mother-daughter,
sister-sister, and cousin-cousin correlations in generation times ($\rho_{md}$, $\rho_{ss}$, $\rho_{cc}$, respectively) inform how division timing might be inherited. If the generation time of a cell is determined solely from properties inherited from its mother cell, then $\rho_{cc}$ will be equal to $\rho_{md}^2 \rho_{ss}$\footnote{\label{note3}}. This is not the case for \textit{S. elongatus}, whose $\rho_{cc}$ is larger than $\rho_{md}$ in magnitude, suggesting that the ongoing circadian clock might affect division timing in addition to simple mother-daughter inheritance. The modulated rates model captures these correlations (Table 8.1).

### Summary of the parameter values used and comparisons to data

For the purpose of analyzing the statistics of division timing, there are only two free parameters ($A$ and $\varphi$) in the models, whereas the other parameters ($\lambda_0$, $\delta$, $\sigma$) are already determined and fixed or extracted from the data. The two free parameters are also the important parameters that characterize how the clock affects division timing within the model. The maximum growth rate $\lambda_0$ is extracted from the measured instantaneous growth rates. The degradation rate $\delta$
is assumed to be the same as inferred from the experiments with the clock-deletion strain. The magnitude of the coarse-grained stochasticity $\sigma$ is chosen to match the measured standard deviation of cell lengths at birth. The two remaining parameters describe the modulation function $y(\theta)$ must then describe the statistics of division timing, including those shown in Figures 8.2, 8.3, and Table 8.1.
This dissertation provides a step towards a more quantitative understanding of the microbial cell cycle, from the molecular to the single-cell, and at a phenomenological or a mechanistic level. Throughout, many closely related open problems were discussed, such as the origins and implications of the prevalence of adders in all domains of life. We conclude the dissertation by outlining additional outstanding problems that may be amenable to the modeling approach taken here.

As we have seen, cell division is intimately related to cell growth, or the increase in cell volume. With the exception of Chapter 8, all other chapters considered experiments in which the cells are growing exponentially at a rate that is constant throughout the experiments. Having developed some understanding of how the cell cycle is regulated under constant exponential growth, it is now imperative to consider the case outside exponential growth. For example, *E. coli* cells exiting stationary phase upon the introduction of fresh media reach near the
average cell size of cells growing exponentially in that media before the first division (Somenath Bakshi, manuscript in submission). Preliminary analysis shows that the adder-per-origin model discussed in Chapter 4 fails to capture this observation. The result is one of several results suggesting that the coupling between DNA replication and cell division may not be as simple as described in Chapters 4-6. For example, previous experiments have shown that cells in stationary phase may contain multiple completed chromosomes\textsuperscript{116}. How then do stationary state cells with different numbers of chromosomes exit stationary phase? Suggestively, single-cell measurements show that width increases faster than exponentially during exit from stationary phase (Somenath Bakshi, manuscript in submission), suggesting that other processes such as cell wall synthesis may affect the coupling between DNA replication and cell division outside exponential growth.

Another challenge is to understand the strikingly diverse modes of cell cycle regulation in nature. For example, the protein DnaA is required for the initiation of DNA replication in both \textit{E. coli}, as discussed in Chapter 7, and in \textit{B. subtilis}. However, the cellular processes surrounding the production and interconversion of the two forms of DnaA are drastically different between the two bacteria\textsuperscript{98}. In both cases, the mechanisms must take as input an environmentally specified volume doubling time to give as output an average time between initiations equal to the input doubling time, but must do so with different cellular processes. Our modeling efforts in Chapter 7 might therefore inform how gene circuits in general might be coupled to externally specified timescales. Another example of diverse modes of cell cycle regulation is the unicellular flagellate \textit{Chlamydomonas reinhardtii}, a cell of which grows many-fold its size at birth before dividing several times rapidly to return near its birth size\textsuperscript{117}. Under this exotic cell cycle, how are DNA replication, cell growth, and cell division coupled? The modeling approach taken in this dissertation will continue to aid our quest to understand the wonderful mechanisms of microbial cell cycles.
References


[69] A. Zaritsky, N. Vischer, and A. Rabinovitch, “Changes of initiation mass and cell dimensions by the 'eclipse',' Molecular Microbiology, 2007, 63(1).


