TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

Jason Lloyd-Price

# Dynamics of Genetic Circuits with Molecule Partitioning Errors in Cell Division and RNA-RNA Interactions

Julkaisu 1310 • Publication 1310

Tampere 2015

Tampereen teknillinen yliopisto. Julkaisu 1310
Tampere University of Technology. Publication 1310

Jason Lloyd-Price

# Dynamics of Genetic Circuits with Molecule Partitioning Errors in Cell Division and RNA-RNA Interactions

Thesis for the degree of Doctor of Science in Technology to be presented with due permission for public examination and criticism in Tietotalo Building, Auditorium TB109, at Tampere University of Technology, on the 7th of August 2015, at 12 noon.

Tampereen teknillinen yliopisto - Tampere University of Technology
Tampere 2015

ISBN 978-952-15-3546-8 (printed)
ISBN 978-952-15-3609-0 (PDF)
ISSN 1459-2045

# Abstract

Many signaling and regulatory molecules within cells exist in very few copies per cell. Any process affecting even limited numbers of these molecules therefore has the potential to affect the dynamics of the biochemical networks of which they are a part. This sensitivity to small copy-number changes is what allows stochasticity in gene expression to introduce a degree of randomness in what cells do. While this randomness can be suppressed, it does not appear to be so in many biological systems, at least not to the maximum degree possible. This suggests that this randomness is not necessarily detrimental to cell populations, as it can produce qualitatively new behaviours in genetic networks which may be utilized by cells.

In this thesis, two other mechanisms are investigated which, through their interaction with low copy-number molecules, are able to produce qualitatively different dynamics in genetic networks: the stochastic partitioning of molecules in cell division, and the direct interaction of two low copy-number molecules. For this, a novel simulator of chemical kinetics is first presented, designed to simulate the dynamics of genetic circuits inside growing populations of cells. It is then used to study a genetic switch where one repressive link is formed by direct interaction between RNA molecules. This arrangement was found to decouple the stability of the two noisy attractors of the network and the speeds of the state transitions. In other words, it allows the network to have two equally-stable noisy attractors, but differing state transition speeds.

Next, the cell-to-cell diversity in RNA numbers (as quantified by the normalized variance) of a single gene over time in a growing model cell population was studied as a function of the division synchrony. In the model, synchronous cell divisions introduce transient increases in the cell-to-cell diversity in RNA numbers of the population, a prediction which was verified using single-molecule measurements of RNA numbers. Finally, the effects of the stochastic partitioning of regulatory molecules in cell division on the dynamics of two genetic circuits, a switch and a clock, were studied. Of these two circuits, the switch has the most dramatic changes in its dynamics, brought on by the inevitable negative correlation in molecule numbers that sister cells inherit. This negative correlation can allow a cell population to partition the phenotypes of the individual cells with less variance than a binomial distribution.

These results advance our understanding of the different behaviours that can be produced in genetic circuits due to these two mechanisms. Since they produce unique behaviours, these mechanisms, and combinations thereof, are expected to be used for specialized purposes in natural genetic circuits. Further, since the downstream effects of these mechanisms may be more predictable than, e.g., modifying promoter sequences, they may also be useful in the design and implementation of future synthetic genetic circuits with specific behaviours.

# Preface

This study was carried out at the Department of Signal Processing, Faculty of Computing and Electrical Engineering, Tampere University of Technology, under the supervision of Associate Professor Andre S. Ribeiro.

I would like to express my sincere gratitude to my supervisor and friend, Associate Professor Andre S. Ribeiro, for his diligent supervision and relentless desire to impart his wealth of knowledge and life experiences. I would also like to thank all of the members of the Laboratory of Biosystem Dynamics, past and present. This thesis is born out of the numerous discussions (and arguments) with all of you, and would be quite different without the individual imprints each has left. I would especially like to thank Antti Häkkinen and Abhishekh Gupta for the many discussions on theoretical aspects, and Meenakshisundaram Kandhavelu and Jarno Mäkelä for the practical aspects.

I am grateful to the department secretary, Virve Larmila, and the coordinators Elina Orava, Ulla Siltaloppi, and Sari Salo, for all their excellent help during my studies at TUT. Further, I would like to thank the TUT President's graduate programme for supporting my studies during my time at TUT.

I am also grateful to the two pre-examiners, James J. Collins and Ilya Shmulevich, for their insightful suggestions which have helped improve this thesis.

I would finally like to thank my family, for all their love and encouragement over the years, and my beloved wife Lingjia for her unlimited love, support, and patience during the completion of this work.

Tampere, June 2015

Jason Lloyd-Price

# Contents

# List of Abbreviations

| | |
|---|---|
| **CFPE** | Chemical Fokker-Planck Equation |
| **CLE** | Chemical Langevin Equation |
| **CME** | Chemical Master Equation |
| **DM** | Direct Method |
| **DNA** | Deoxyribonucleic Acid |
| **DSSA** | Delayed SSA |
| **FRM** | First Reaction Method |
| **ISSA** | Inhomogeneous SSA |
| **KDE** | Kernel Density Estimation |
| **mRNA** | Messenger RNA |
| **NRM** | Next Reaction Method |
| **ODE** | Ordinary Differential Equation |
| **RBS** | Ribosome Binding Site |
| **RNA** | Ribonucleic Acid |
| **RRE** | Reaction Rate Equation |
| **srRNA** | Small Regulatory RNA |
| **SSA** | Stochastic Simulation Algorithm |
| **TF** | Transcription Factor |
| **TSS** | Transcription Start Site |

# List of Publications

I Jason Lloyd-Price, Abhishekh Gupta and Andre S. Ribeiro, “SGNS2: a compartmentalized stochastic chemical kinetics simulator for dynamic cell populations,” *Bioinformatics*, vol. 28, no. 22, pp. 3004−3005, 2012.

II Jason Lloyd-Price and Andre S. Ribeiro, “Bistability in a stochastic RNA-mediated gene network,” *Physical Review E*, vol. 88, pp. 032714, 2013.

III Jason Lloyd-Price, Maria Lehtivaara, Meenakshisundaram Kandhavelu, Sharif Chowdhury, Anantha-Barathi Muthukrishnan, Olli Yli-Harja and Andre S. Ribeiro, “Probabilistic RNA partitioning generates transient increases in the normalized variance of RNA numbers in synchronized populations of *Escherichia coli*,” *Molecular Biosystems*, vol. 8, pp. 565−571, 2012.

IV Jason Lloyd-Price, Huy Tran, Andre S. Ribeiro, “Dynamics of small genetic circuits subject to stochastic partitioning in cell division,” *Journal of Theoretical Biology*, vol. 356, pp. 11−19, 2014.

The author of this thesis contributed to these publications as follows. In **Publication I**, the author designed and implemented the simulator, and participated in the writing of the manuscript. This publication will also appear in Abhishekh Gupta’s doctoral thesis. In **Publication II**, the author conceived the study, designed the model, carried out the simulations, analyzed the results, and wrote the manuscript. In **Publication III**, the author performed the simulations, aided in the execution of the experiments, analyzed the results of the experiments, and participated in the writing of the manuscript. In **Publication IV**, the author designed the models and tests, and participated in the writing of the manuscript. This publication will also appear in Huy Tran’s doctoral thesis.

# 1 Introduction

Many RNAs and regulatory proteins exist within cells in very low copy-numbers (Paulsson 2004; Kaern et al. 2005). Since these molecules are discrete entities, any process that affects their numbers and involves some randomness will inevitably introduce a level of noise in their numbers. Though this noise can, to some extent, randomize the actions of a cell, and thus disrupt cellular functioning, it is not always detrimental to cell populations. This noise can be exploited to produce new and interesting behaviours in biochemical networks (Arkin et al. 1998; Elowitz and Leibler 2000; Kaern et al. 2005; Bratsun et al. 2005; Lipshtat et al. 2006), and cell populations have been shown to use this source of diversity to their advantage. Stochastic switching between phenotypes serves as a means to cope with fluctuating environments (Kussell and Leibler 2005; Kussell et al. 2005; Acar et al. 2008), or to maintain a random subset of a population in a particular phenotype (Süel et al. 2006). Stochastic differentiation underlies the retinal mosaic behind colour vision in *Drosophila melanogaster* (Wernet et al. 2006). Random, though coordinated decisions between infecting $\lambda$ phages determine a new host *Escherichia coli*'s fate (Arkin et al. 1998; Zeng et al. 2010).

While stochasticity in the processes underlying gene expression is sufficient to explain these observations, it is not the only mechanism acting on low copy-number molecules with the potential to change the dynamics of circuits. One other such mechanism is the random partitioning of molecules during cell division (Huh and Paulsson 2011b; Huh and Paulsson 2011a). In bacteria, the unequal partitioning of macromolecules such as plasmids clearly has the potential to create significant differences between sister cells (Huh and Paulsson 2011b; Reyes-Lamothe et al. 2013). Further, the unequal partitioning of damaged and/or non-functional proteins has been implicated in the aging process of *E. coli* (Lindner et al. 2008; Gupta et al. 2014b; Gupta et al. 2014d), which can result in a significant difference in population vitality (Ackermann et al. 2003; Gupta et al. 2014c). However, whereas non-functional proteins merely reduce the vitality of the cells, RNAs and regulatory proteins can have a myriad of other downstream effects in their respective genetic networks. This leads to the first question motivating this thesis: in what way does the stochastic partitioning of the RNAs and regulatory proteins that compose a genetic network affect its dynamics?

Another means by which noise, from any source, can be amplified to produce phenotypic variation is to have two or more low copy-number molecules interact directly. Such a scheme is ubiquitous, and is found in all kingdoms of life: gene silencing or activation by small non-coding RNA molecules. There are at least 2000 microRNAs in humans (Friedländer et al. 2014), which is comparable to the amount of Transcription Factors (TFs) (Vaquerizas et al. 2009). Prokaryotes employ a similar regulatory mechanism, whereby a Small Regulatory RNA (srRNA) can tightly bind to its complementary Messenger RNA (mRNA) to either silence or stabilize it (Gottesman and Storz 2011). This form of interaction can lead to a highly non-linear function: a threshold-linear regulation function (Levine et al. 2007; Levine and Hwa 2008). This regulatory function has non-trivial noise characteristics, capable of both suppressing and amplifying noise (Levine et al. 2009). Thus, the second question motivating this thesis: in what way does the direct interaction between RNA molecules affect the dynamics of a stochastic genetic network (specifically, a switch)?

Though the stochastic dynamics of even simple genetic networks is often too complex to describe analytically, it can be studied with the aid of stochastic simulation (Gillespie 2007). Thus, in order to study the effects of the above mechanisms, a simulation method that accurately captures the sources and effects of this noise must be employed. The Stochastic Simulation Algorithm (SSA) is the gold standard simulation algorithm for such systems, providing statistically exact samples of trajectories from the distribution prescribed by the Chemical Master Equation (CME), which in turn can be rigorously derived from microphysical arguments (Gillespie 1992). This has been extended to incorporate delayed reactions which have enabled the time taken by the individual processes in gene expression to be efficiently modelled without explicitly representing each individual step (Roussel and Zhu 2006; Ribeiro 2010).

## 1.1 Objectives

In this thesis, the impact on the dynamics of genetic circuits of the two above mechanisms was examined. First, a new stochastic simulation tool is presented, constructed to enable the simulation of the models that were used in the remainder of this thesis. This simulator was built based on the SSA, with the ability to dynamically create and destroy interlinked compartments at runtime to introduce transient spatial restrictions on the interactions between molecules. Next, the behaviour of a genetic switch when one of the links is mediated by RNA-RNA interactions was studied. Finally, the effects of deviations from a perfectly symmetric partitioning of RNA and other low copy molecules during cell division were studied. These effects were first characterized at the single-gene level, using both simulations and single-cell measurements in live cells. This was then studied at the network level for two genetic circuits: a switch and a clock.

The thesis has three objectives:

**I** To construct a simulation tool capable of simulating the above mechanisms. Specifically, this simulator must account for the inherent stochasticity in gene expression, the time taken by the involved processes, while also supporting the creation of new cells during which selected molecule partitioning schemes can be applied to different molecules.

**II** To study the dynamics of a stochastic genetic circuit utilizing direct interaction between RNA molecules as a regulatory connection, and to identify new dynamical features that can result from this kind of interaction.

**III** To characterize the differences that arise in the dynamics of single genes and genetic circuits when placed in a growing and dividing population of cells with stochastic partitioning of molecules in division.

**Objective I** was completed in **Publication I**. **Objective II** was completed in **Publication II**. Finally, **Objective III** was completed in **Publications III** and **IV**.

## 1.2 Thesis Outline

This thesis is organized as follows. Chapter 2 briefly introduces the necessary biological background, as well as the *in vivo* single-molecule measurement techniques against which model predictions were tested. Chapter 3 introduces the modelling strategies and simulation algorithms employed in the publications of the thesis. Chapter 4 presents the necessary background on the genetic networks, focusing on the Toggle Switch and the Repressilator. Finally, the conclusions and final discussion are presented in chapter 5.

# 2 Biological Background and Methods

## 2.1 Bacterial Growth and Division

The publications in this thesis focus on *E. coli*, a common rod-shaped bacterium found in the guts of many warm-blooded organisms (Alberts et al. 2002). It is also ubiquitous in molecular biology labs, and a wealth of knowledge about its structure and behaviour has accumulated over many years of study, making it the most intensively studied prokaryotic model organism. Bacteria have the advantage of being somewhat simpler than eukaryotic systems - they are unicellular organisms with no discernible organelles apart from the nucleoid. Their gene expression systems are also simpler, lacking the physical separation afforded by the eukaryotic nucleus as well as the complex RNA processing that occurs in eukaryotes. It is for these reasons that the first synthetic genetic circuits have been constructed in these cells (see, e.g. (Elowitz and Leibler 2000; Gardner et al. 2000)), and it is in these organisms that studies of gene expression at the single-event level are being conducted (see e.g. (Lutz et al. 2001; Golding and Cox 2004; Golding et al. 2005; Muthukrishnan et al. 2012; Kandhavelu et al. 2012a)). An example image of *E. coli* cells, taken with phase contrast microscopy is shown in Figure 2.1.

In suitable media, *E. coli* cells grow exponentially by repeatedly elongating, and then dividing in two (Alberts et al. 2002; Scott et al. 2010; Osella et al. 2014). During elongation, the chromosome is duplicated, and the two copies are segregated to the quarter points of the cell in structures known as nucleoids (Fisher et al. 2013). The division septum is then formed from the protein FtsZ (Weiss 2004), which is positioned in the center of the cell by a combination of nucleoid occlusion and an oscillatory protein-based system called the Min system (Margolin 2006). The septum then constricts the cell wall, dividing the cell into two new cells.

This division process results in a remarkably precise division point in the center of the cell, though there does exist some variance in this point (Männik et al. 2012; Gupta et al. 2014d). Molecules in the cytoplasm of the cell are therefore frequently assumed to be partitioned into the new cells equally and independently, resulting

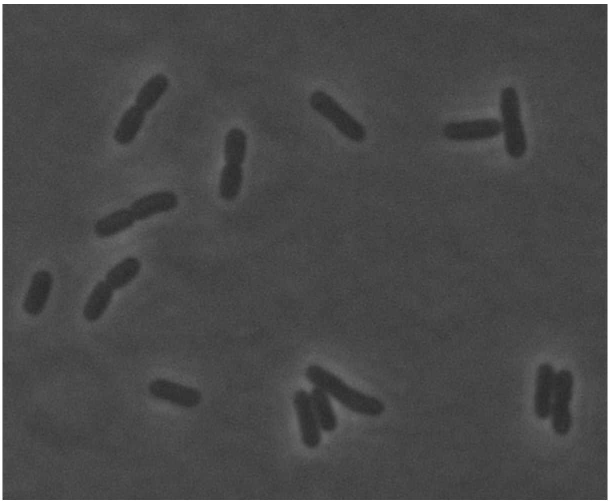


**Figure 2.1:** Image of *E. coli* cells taken by phase contrast microscopy.

in a binomial molecule partitioning distribution upon division (Berg 1978; Rigney 1979). However, this may be affected by a number of factors including, but not limited to, the following. Additional variance in the division point (e.g. due to stress (Männik et al. 2012)) will bias the partitioning of molecules towards the larger cell (Huh and Paulsson 2011a; Huh and Paulsson 2011b; Gupta et al. 2014d). That is, one cell will likely receive more of the contents of the parent cell than the other. The limited diffusion of macromolecules can further bias their partitioning towards the cell inheriting the region where they were synthesized or trapped (Lindner et al. 2008; Montero Llopis et al. 2010; Gupta et al. 2014b). Lastly, clustering of the partitioned molecules will bias the partitioning towards the cell inheriting the largest clusters. All these effects increase the variability in the numbers of molecules inherited by daughter cells.

In general, to decrease this variability, energy must be spent by the cell (Huh and Paulsson 2011a). For example, molecules may form pairs which are segregated evenly into the daughter cells, as is the case with the genome (Fisher et al. 2013), and other large single-copy structures within cells such as F-plasmids (Schumacher et al. 2010). Molecules may also bind to a central structure, which is partitioned evenly between the daughters, such as the spindle apparatus employed during mitosis in eukaryotes (Alberts et al. 2002; Huh and Paulsson 2011b). Finally, cells may rely on the sheer size of macromolecules to distribute them evenly between the cells (Huh and Paulsson 2011b).

The population-level effects of events which occur in division, such as asymmetric partitioning of cellular components, will change based on the timing of the divisions in the population. In particular, if all cell divisions occur synchronously, the added cell-to-cell diversity will be introduced simultaneously. Thus, an experiment measuring the cell-to-cell diversity at that moment will overestimate the variability between the cells. Likewise, if this variability affects the way the population interacts with its environment, the population's behaviour will drastically change at that point, whereas an asynchronous population will not. Division synchrony can be induced by a number of different mechanisms in *E. coli*, which generally relate to stressful conditions such as heat shock (Smith and Pardee 1970) and nutrient deprivation (Cutler and Evans 1966). Once synchronized, cell populations can maintain division synchrony for numerous generations (Hoffman and Frank 1965).

In **Publication IV**, the effects of the aforementioned partitioning schemes on the dynamics of genetic networks were studied. In **Publication III**, this partitioning was additionally studied as a function of cell division synchrony.

## 2.2 Gene Expression

Genes are the unit of heredity of living organisms (Alberts et al. 2002). In general, this refers to stretches of DNA containing the information necessary to produce proteins, the functional components within cells. The process by which this information is read to produce these proteins is called gene expression. There are two main steps involved in gene expression, transcription and translation, which together form the Central Dogma of Molecular Biology, and are depicted in Figure 2.2.

Structurally, each gene is composed of two functionally distinct sequences in the DNA. The first is the "promoter", a region of DNA upstream from the coding region where transcription is initiated. Since gene expression begins in this region, it is also the point where many regulatory molecules bind to alter the expression level of the gene. The second consists of the sequence coding for the protein itself. Briefly, transcription begins with the binding of an RNA polymerase enzyme to the promoter of a gene. The polymerase then copies the coding part of the DNA molecule into a complementary RNA molecule, until the terminator is reached. The resulting protein-coding RNA molecule, called a Messenger RNA (mRNA), is then bound by a Ribosome to initiate translation. Note that prokaryotes lack a delimited nucleus, and therefore translation can initiate immediately after the ribosome binding site on the mRNA has been transcribed by the RNA polymerase. During translation, the mRNA's nucleotides are read in triplets, called codons, which each correspond to a particular amino acid to append to the new protein. When the stop codon, a special codon denoting the end of the protein, is reached, the new protein is released. The new protein will then fold into its active shape to finally perform its function within the cell.

The DNA within each cell of an organism contains the information required to produce all proteins needed by the organism at any point of its life. At a given point in time, it is the particular subset of proteins which are expressed by a given cell which determines what behaviour it will have, i.e. its phenotype. To

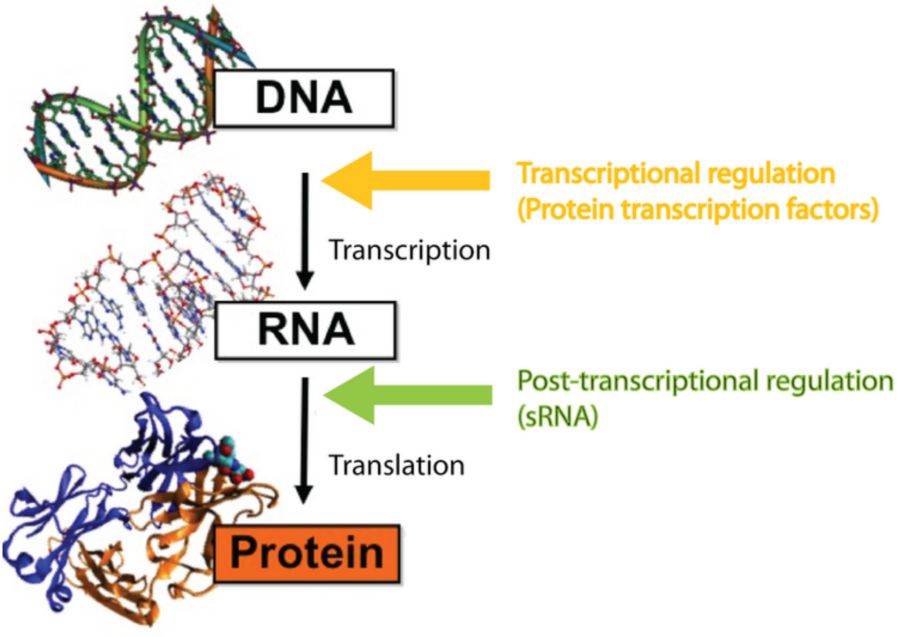
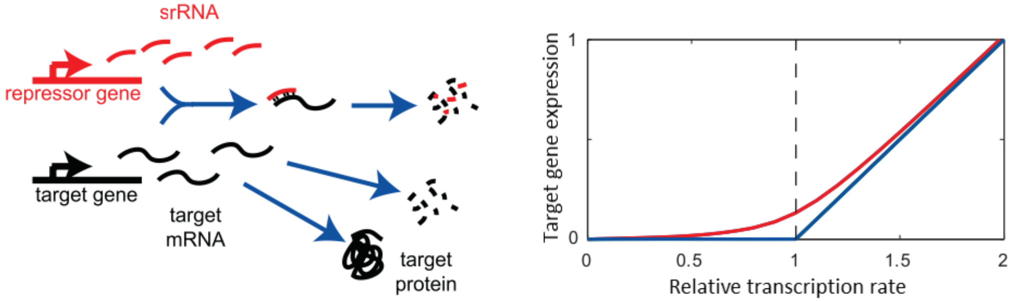

DNA
Transcriptional regulation
(Protein transcription factors)
Transcription
RNA
Post-transcriptional regulation
(sRNA)
Translation
Protein

**Figure 2.2:** The Central Dogma of Molecular Biology: DNA is transcribed into RNA, which is translated into proteins. Also shown are two points of regulation: transcriptional regulation (DNA-TF interactions) and post-transcriptional regulation (small RNAs). The image is modified from http://2011.igem.org/Team:DTU-Denmark/Background_sRNA, under the CC BY 3.0 license.

understand why a cell behaves in a certain way, then, we must understand why that particular set of genes was expressed. Gene expression can be regulated by several means, which behave differently based on which step during gene expression they affect. This thesis focuses on two levels: transcriptional (i.e. regulation at the promoter by transcription factors), and post-transcriptional regulation.

Transcription Factors (TFs) are proteins which affect the expression of their target gene by binding at specific sites at or near the target gene's promoter. Such regulation is presented in Figure 2.2 as Transcriptional regulation. The simplest form of interaction is when a TF bound at the promoter region blocks the RNA polymerase from initiating transcription, such as the regulation of the Lac operon by LacI in *E. coli* (Schlax et al. 1995). Since the gene cannot be transcribed, no protein is produced, and the gene is said to be repressed or turned off. Other transcription factors can increase the rate of transcription by bending the DNA such that the RNA polymerase is more likely to recognize and bind to the promoter, such as the AraC protein when bound to arabinose, which regulates the araBAD operon in *E. coli* (Schleif 2000). This kind of TF-based regulation is found in the models in all four publications in this thesis.

Post-transcriptional regulation takes place at the level of the mRNA, as pictured in Figure 2.2. Though regulation at this level is more common and complex in eukaryotes, it is not absent in prokaryotes. In *E. coli*, numerous genes produce small non-coding RNA molecules which are complementary to a stretch of the mRNA of another gene, their target. These small RNA molecules are first bound by a chaperone protein HfQ, which is thought to both protect the srRNA from degradation, and to increase the chances that it meets its target mRNA (Gottesman and Storz 2011). Upon binding to the target, the srRNA can either up- or down-regulate the translation of the target protein, depending on where in the target the srRNA binds (Gottesman and Storz 2011). Up-regulation is achieved by active recruitment of Ribosomes to the Ribosome Binding Site (RBS) of the target. Down-regulation is achieved by either blocking the RBS of the target, or by utilizing the cell's double-stranded RNA degradation machinery to degrade both RNA molecules. An example of this mechanism in *E. coli* is the regulation of the iron storage regulator *fur* by the srRNA *ryhB* (Massé and Gottesman 2002; Massé et al. 2003).



**(a)** Illustration of the regulation of a gene by an srRNA. When the srRNA (red) is produced, it binds to the target mRNA (black), and both molecules are subsequently degraded. When the srRNA is not produced, the target mRNA is translated into the target protein. Image is modified from (Levine et al. 2007) under the CC BY 3.0 license.

**(b)** Threshold-linear regulatory function. The blue line represents an ideal regulation function when srRNA-mRNA binding is very fast (Levine et al. 2007). Target expression is completely suppressed when the srRNA is produced at a greater rate than the mRNA. Expression increases linearly beyond that. The red line represents a more realizable function. Image is modified from (Levine et al. 2007) under the CC BY 3.0 license.

**Figure 2.3:** Gene regulation by a small regulatory RNA.

The down-regulation of a gene by srRNA is illustrated in Figure 2.3a. This interaction results in a highly non-linear gene regulation function: a threshold-linear function, pictured in Figure 2.3b. This regulation function can be exploited to reduce (Levine and Hwa 2008) or increase (Elf et al. 2003; Levine et al. 2009) the amount of noise in gene expression, or to sharpen spatial patterns in gene expression (Levine et al. 2007). These interesting noise properties result from the facts that, when the srRNA is produced in greater abundance than the target mRNA, expression of the target protein will approach Poissonian; meanwhile, when the mRNA is produced in greater abundance, expression of the target protein will be as noisy as without the srRNA, approaching a constant based on the number of proteins produced per mRNA (Levine et al. 2009). In between these two regimes, the noise is dramatically increased due to critical phenomena (Elf et al. 2003; Levine et al. 2009). Altogether, these properties make the resulting dynamics of any circuit containing this type of regulation non-trivial.

The dynamics of a network utilizing this kind of RNA-mediated regulation was studied in **Publication II**.

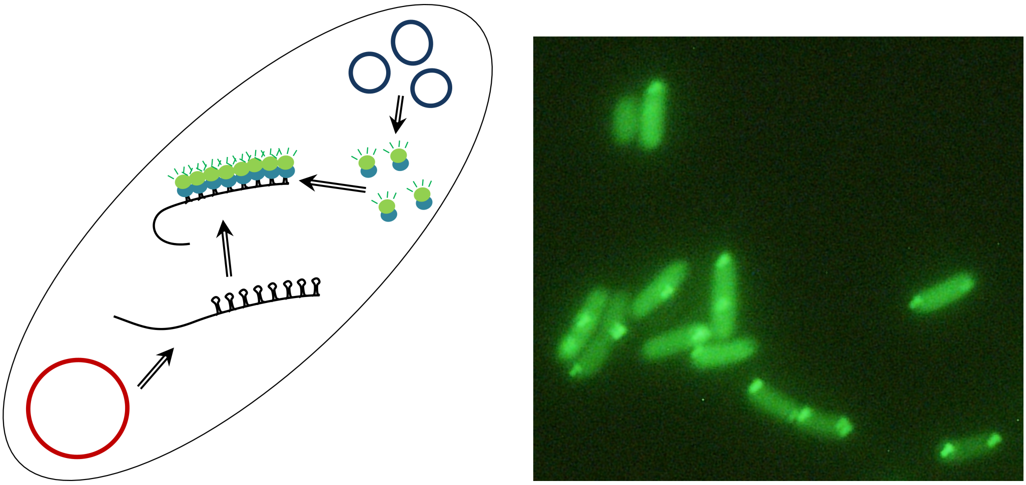
## 2.3 Single-molecule Measurements of mRNA

Advances in microscopy and fluorescent reporters have given rise to a number of RNA visualization techniques with single-molecule precision, which can be used to study the dynamics of the processes mentioned in the previous sections. In **Publication III**, one such method, invented for use in *Saccaromyces cerevisiae* (Fusco et al. 2003) and adapted for use in *E. coli* (Golding and Cox 2004), was used to characterize the cell-to-cell diversity in the number of produced mRNA molecules in a synchronous population of cells. This method is described here.

### 2.3.1 MS2 System

Single RNA molecules can be detected *in vivo* by a method that utilizes the MS2 bacteriophage's coat protein's ability to specifically bind to specific sequences of RNA (Fusco et al. 2003; Golding and Cox 2004). In this system, a multi-copy plasmid carrying a fusion protein MS2-GFP is inserted into the cells. An array of MS2 binding sites is then placed downstream of the promoter of interest. When the two constructs are co-expressed, the MS2-GFP proteins rapidly bind to the array of binding sites on the RNA molecules transcribed from the promoter of interest (Golding and Cox 2004), drastically increasing the local concentration of fluorescent molecules. This system is depicted in Figure 2.4a. The result is a bright "spot" when seen with a fluorescence microscope. An example image of cells with fluorescently labelled RNA molecules within is shown in Figure 2.4b.

By imaging the same cells over time and tracking the total spot fluorescence inside each cell, this system can be used to study the dynamics of transcription (see e.g. (Golding and Cox 2004; Golding et al. 2005; Kandhavelu et al. 2011; Kandhavelu et al. 2012a)). Further, once the RNA has been wrapped by a sufficient number of MS2-GFP molecules, it becomes immune to degradation (Golding and Cox 2004; Montero Llopis et al. 2010), resulting in a large fluorescent molecule diffusing through the cytoplasm of the cell. This property has been exploited to study



**(a)** Illustration of the MS2-based system to detect RNA *in vivo* with single-molecule precision. MS2-GFP molecules (blue/green balls) are produced from a high-copy plasmid (blue). Target RNA carrying 96 MS2 binding sites (black) is produced from a single-copy F-plasmid (red). When a target RNA is produced, MS2-GFP molecules bind to it, forming a bright spot when imaged with a fluorescence microscope.

**(b)** Example image from an epifluorescence microscope of *E. coli* cells co-expressing MS2-GFP and target RNA. Cells are visible due to being flooded uniformly with MS2-GFP. Individual RNA molecules are visible as fluorescent spots.

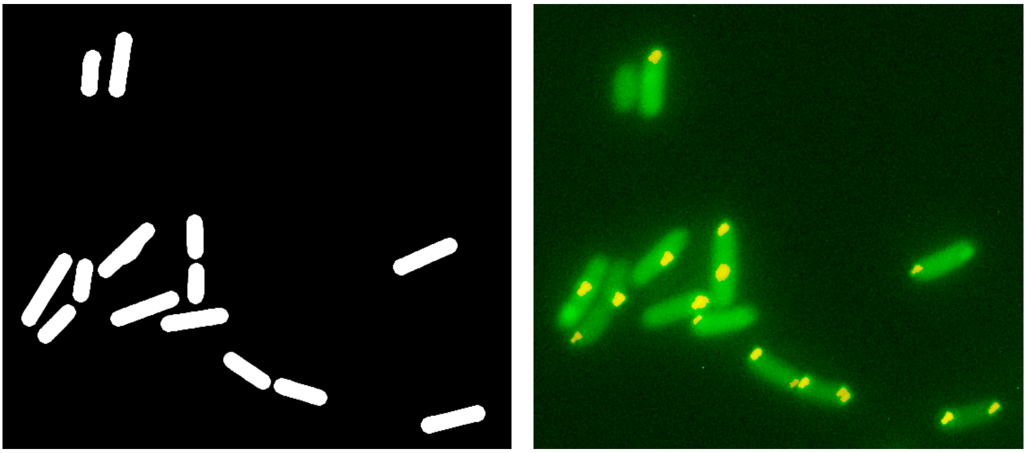
**Figure 2.4:** *In vivo* detection of RNA molecules using MS2-GFP.

the physical properties of the cytoplasm by way of the movement of these large fluorescent particles (Golding and Cox 2006; Gupta et al. 2014b; Gupta et al. 2014d).

### 2.3.2 Image Analysis

**Publication III** examined the diversity of behaviours within a population of cells. For this, it was necessary to examine the behaviours of many different cells over time. To gather sufficient data, a semi-automated image analysis pipeline was used to analyze many images of cells, taken at different timepoints after synchronization. This section describes these methods.

The first step in any single-cell analysis pipeline is to segment the cells from the background. While several automated methods exist, for snapshots of cells (i.e. only a single moment in time), this step can be rapidly performed by the use of image manipulation software. This was favoured in **Publication III** due to the high level of background noise present in the images, visible in Figure 2.4b. An example segmentation used in **Publication III** for the image shown in Figure 2.4b is shown in Figure 2.5a.



**(a)** Segmentation of the cells from the image in Figure 2.4b.

**(b)** Segmentation of the fluorescent spots from the image in Figure 2.4b by KDE (red) superimposed on the image from Figure 2.4b. The green and red in the spots mixes to produce yellow.

**Figure 2.5:** Image analysis of cells expressing MS2-GFP and target RNA.

Fluorescent spots can be detected from the image by a method based on Kernel Density Estimation (KDE) (Ruusuvuori et al. 2010). This method begins by applying the following transformation to the image:

$$H(i,j) = \frac{1}{\mathcal{C}(N)} \sum_{m,n \in N} K\left(\frac{I(i,j) - I(i+m, j+n)}{\alpha}\right) \tag{2.1}$$

where $N$ is the set of neighbour pixels to include, $\mathcal{C}$ is the cardinality of the set, $K$ is the chosen kernel, $\alpha$ is the bandwidth, and $I(i,j)$ is the intensity of the image at coordinates $(i,j)$.

$H(i,j)$ can informally be thought of as the local smoothness of the image, and ranges from 0 to 1. Spots are features with low local smoothness, i.e. the intensities of the pixels in the local neighbourhood of a spot are distinct from the intensities within the spot. Thus, spots can be segmented from the transformed image by defining a threshold $t$, and labelling areas with $H(i,j) < t$ as spots.

In **Publication III**, $N$ was set to a circular neighbourhood with a radius of $r$, and set $K$ to a Gaussian kernel. The parameters $\alpha$, $r$ and $t$ were tuned by eye to produce a good segmentation, an example of which is shown in Figure 2.5b.

# 3 Modelling and Simulation of Stochastic Gene Expression

All publications in this thesis include models of gene expression and genetic circuits, and analyses thereof. Here, the construction of these models is described, along with the simulation algorithms used to simulate their dynamics.

## 3.1 Chemical Master Equation

The models presented here account for the non-negligible amount of noise in the biochemical processes underlying gene expression. This noise originates from the randomness in the timings of the individual births and deaths of the molecules involved. When this noise affects the numbers of molecules in low copy-number, it has the potential to impact the dynamics of downstream regulatory circuits. For example, if a reaction happens to occur due to the random collision of two molecules within a cell, which produces a molecule of which there were only two before, this single random event has just increased the population of that molecule by 50%. This drastic change will have downstream repercussions if this molecule is involved in other reactions, since those reactions will (at least temporarily) have 50% increased propensity to occur.

The processes which we are interested in, namely gene expression and regulation, involve such low-copy molecules. Specifically, RNA molecules are only present in limited quantities per cell (Gillespie 2007), and there is only one copy of the genome. Thus modelling strategies and simulation techniques which ignore this biochemical noise will miss important features of the dynamics. One way to accurately simulate the dynamics of such a noisy system would be to model it at a painstaking level of detail: model the space of the system explicitly, track the positions and momenta of every single molecule, and detect and react to collisions between them. While technically correct, this approach is extremely computationally demanding (Gillespie 2007).

Instead, we make the assumption that for each reaction $\mu$, we can write a function $a_\mu(\mathbf{x})$ of the state of the system $\mathbf{x}$ at the current time $t$, such that $a_\mu(\mathbf{x})dt$ is the probability that a combination of its reactants will meet and react in the next infinitesimal time interval $(t, t + dt)$ (Gillespie 2007). Here, the elements of $\mathbf{x}$ are the current numbers of each of the molecular species, and $a_\mu(\mathbf{x})$ is called the *propensity* of reaction $\mu$. From this assumption alone, it is possible to write the time-evolution of the probability $P(\mathbf{x})$ that the system is in state $\mathbf{x}$, as a master equation called the Chemical Master Equation (CME):

$$\frac{dP(\mathbf{x})}{dt} = \sum_\mu \left(a_\mu(\mathbf{x} - \nu_\mu)P(\mathbf{x} - \nu_\mu) - a_\mu(\mathbf{x})P(\mathbf{x})\right) \tag{3.1}$$

where $\nu_\mu$ is the stoichiometry of reaction $\mu$, i.e. the vector representing the difference in molecule numbers when reaction $\mu$ occurs.

The justification behind the existence of the propensity function depends on the type of reaction it represents. For unimolecular reactions, this justification is often from quantum mechanics (Gillespie 2007), which dictates that such a probability should exist for each molecule which can react via that channel. Therefore, there exists some constant $c_\mu$ for which the propensity function can be written as (Gillespie 2007):

$$a_\mu(\mathbf{x}) = c_\mu X \tag{3.2}$$

where $X$ is the number of molecules which can react via this reaction.

For bimolecular reactions, additional assumptions must be made. Specifically, the molecules must be in thermal equilibrium at a constant temperature, and must be uniformly distributed within the reaction volume. The latter can be achieved either by direct stirring or if the number of non-reactive collisions between molecules outnumber the reactive collisions (Gillespie 2007). Given these assumptions, it is possible to rigorously derive a constant $c_\mu$ from microphysical arguments for the bimolecular reaction propensity (Gillespie 1992):

$$a_\mu(\mathbf{x}) = c_\mu X_1 X_2 \tag{3.3}$$

where $X_1$ and $X_2$ are the populations of the two reacting molecule species. Note that if two of the same molecular species react, the propensity function changes, since a molecule cannot react with itself:

$$a_\mu(\mathbf{x}) = \frac{c_\mu X(X-1)}{2} \tag{3.4}$$

Lastly, while they do not represent a “real” reaction, zero-order reactions are also extremely useful in models. For example, these can be used to represent reactions where the reactants are not explicitly represented in the modelled system, such as water or other molecules assumed to be pervasive and in constant abundance. They can also represent the entry of molecules into the reaction volume from an outside source. Since these reactions do not depend on the population of any of the modelled molecules within the reaction volume, their propensity function is simply:

$$a_\mu(\mathbf{x}) = c_\mu \tag{3.5}$$

Since the CME for a model is often too complex to present, let alone solve, it is often simpler and more intuitive to present models built in the stochastic formulation by the set of reactions which compose them. In the following sections, various sets of reactions representing the different processes involved in gene expression and gene regulation will be presented, which form the basis of the models used in the publications of this thesis. The reactions are presented in the following form:

$$\mathrm{A} + \mathrm{B} \xrightarrow{k} \mathrm{C} \tag{3.6}$$

Here, a molecule of species A reacts with a molecule of species B to form a molecule of species C, with stochastic constant $c_\mu = k$.

## 3.2 Modelling Gene Expression

Gene expression is the process by which a protein is constructed based on the amino acid sequence encoded in DNA (for details, see 2.2). To remind the reader, this process begins when an RNA polymerase binds to the promoter sequence of a gene, and initiates transcription of that gene. This produces a complementary RNA molecule, to which a Ribosome binds to produce the final proteins.

The above process can be summarized in a very compact, high-level reaction (Ribeiro et al. 2006):

$$\mathrm{Pro} + \mathrm{RNAp} \xrightarrow{k_t} \mathrm{Pro} + \mathrm{RNAp} + n\mathrm{P} \tag{3.7}$$

Here, Pro is the promoter of the gene, RNAp is the RNA polymerase, P is the produced protein, and $n$ is the mean number of proteins produced per mRNA. Though this reaction does not change the amount of Pro, its presence as a reactant in this reaction allows other reactions to change the production rate of P, e.g. those in section 3.2.2.

The next step is to model both transcription and translation explicitly, accurately recreating measured protein burst distributions (Zhu et al. 2007):

$$\mathrm{Pro} + \mathrm{RNAp} \xrightarrow{k_t} \mathrm{Pro} + \mathrm{RNAp} + \mathrm{RBS} \tag{3.8}$$

$$\mathrm{RBS} + \mathrm{Rib} \xrightarrow{k_{tr}} \mathrm{RBS} + \mathrm{Rib} + \mathrm{P} \tag{3.9}$$

where RBS is the Ribosome Binding Site (RBS) on the mRNA, and Rib represents a Ribosome. Note that RNAp and Ribosomes are high-copy housekeeping molecules in the cell, and are often considered to be in constant concentration. They are thus sometimes dropped from these reactions (e.g. (Zhu et al. 2007; Loinger and Biham 2007)). Note that here, we have represented the RNA molecule by its RBS, and not by the complete molecule, since in prokaryotes, translation of an mRNA can initiate before the RNA has been fully transcribed.

RNA and protein turnover rates are also extremely important to the dynamics of genetic circuits, if not more than the production rates since these determine how quickly the system will approach a steady-state. In both eukaryotes and prokaryotes, many proteins have been shown to exhibit exponential degradation (Belle et al. 2006). In prokaryotes, mRNA also degrades exponentially (Bernstein et al. 2002), though in eukaryotes, this is altered by polyadenylation (Pedraza and Paulsson 2008). In the publications in this thesis, degradation of RNA and proteins is therefore modelled with first-order reactions:

$$\mathrm{RBS} \xrightarrow{k_{rd}} \varnothing \tag{3.10}$$

$$\mathrm{P} \xrightarrow{k_{pd}} \varnothing \tag{3.11}$$

### 3.2.1 Delays

The above reactions assume that the processes involved in gene expression are instantaneous. For example in reaction (3.9), the protein produced by translation appears immediately upon initiation of the reaction. However, translation takes a non-negligible amount of time to complete, requiring the stepwise elongation of the nacent polypeptide chain. Further, the new protein is not immediately functional upon the addition of the final amino acid - it must fold into its final conformation, a process that can take minutes to hours (Cormack et al. 1996).

Such delays are commonly modelled in two ways. First, it is possible to explicitly represent every individual step required for the process to complete. This approach is fruitful when studying the effects of events that can occur during those steps (Rajala et al. 2010; Ribeiro 2010; Mäkelä et al. 2011). However, this approach results in the need to simulate a greatly increased number of reaction events, and can only be applied to smaller systems (Potapov et al. 2011). It becomes impractical when the number of intermediate steps is large or when a larger network of interacting genes is simulated.

The second approach is to introduce “delayed reactions” - reactions where the products are not immediately released into the system (Roussel and Zhu 2006). Such reactions have the additional advantage that the nature of the intermediate steps do not need to be known; only the statistics of the delay, such as the mean and variance, are required. The drawback is that the delay cannot be affected by events that occur after the reaction has initiated. We write such a delayed reaction as follows:

$$\mathrm{A} + \mathrm{B} \longrightarrow \mathrm{C}(\tau) \tag{3.12}$$

This represents the reaction between an A molecule and a B molecule, but while the reacting molecules are immediately removed from the system, the produced C molecule is not available to react with other molecules until $\tau$ time has elapsed. In the context of gene expression, delays are introduced to model the time taken by transcription, translation, and protein maturation.

During transcription, there are at least two important steps which take a non-negligible amount of time (McClure 1985; Hsu 2009; Mäkelä et al. 2011). First, to read the DNA template, after the RNA polymerase binds to the promoter region, it must unwind the DNA double helix. This process, called the Open Complex Formation, is non-trivial and requires a number of isomerization steps to occur before completion (McClure 1980; McClure 1985; Saecker et al. 2011), and can take on the order of 100 s to 1500 s to occur (McClure 1980; Bertrand-Burggraf et al. 1984; Kandhavelu et al. 2012b; Muthukrishnan et al. 2012). Second, the RNA polymerase must initiate elongation and clear the promoter region. Some promoters, however, do not allow the polymerase to escape easily, causing a large number of "abortive transcripts" to be produced as the polymerase transcribes the initial nucleotides of the gene, but then aborts and returns to the Transcription Start Site (TSS), releasing the short initial transcript (Hsu 2009). If this effect is strong enough, it can introduce another delay before the polymerase initiates a successful production. During both of the above steps, the polymerase is situated at the start site of the promoter, blocking any other polymerase from initiating transcription.

Since these steps must occur in sequence, they are potentially rate-limiting steps in the production of mRNA, and can thus significantly alter the dynamics of gene networks. Further, since the time taken by a sequence of reactions has less variability than a single reaction with the same rate (Ribeiro et al. 2010), the regulation of the durations of these steps can also alter the amount of noise resulting from transcription initiation (Kandhavelu et al. 2012a).

The final step in transcription is the elongation of the RNA molecule, which takes on the order of several minutes (Davenport et al. 2000; Golding and Cox 2004). However, since transcription and translation are coupled in prokaryotes, this delay will only introduce dynamical differences if transcription elongation occurs slower than translation elongation. In the absence of long sequence-dependent transcriptional pauses (Herbert et al. 2006), these two processes proceed at roughly the same rate (Mäkelä et al. 2011). If these are present, a more detailed model such as the one presented in  (Mäkelä et al. 2011) is required. In the absence of such special conditions, the time to produce the complete RNA can be ignored, and the representation of the RNA molecule by the RBS (presented in section 3.2), is sufficient.

Including the above delays into the transcription reaction (3.8) results in the following reaction (Ribeiro and Lloyd-Price 2007):

$$\mathrm{Pro} + \mathrm{RNAp} \xrightarrow{k_t} \mathrm{Pro}(\tau_p) + \mathrm{RBS}(\tau_p) + \mathrm{RNAp}(\tau_{rna}) \tag{3.13}$$

Here, $\tau_p$ represents the sum of the open complex formation and promoter escape, and $\tau_{rna}$ represents the time to complete the elongation of the new mRNA.

There are non-negligible delays in translation as well, and though it proceeds in a manner reminiscent of transcription, the dynamically-relevant delays are different.

Translation begins with the binding of a Ribosome to the mRNA's RBS. Unlike in transcription, however, the Ribosome does not need to open a double-helix, and can initiate elongation of the new Protein almost immediately (after $\sim 3$ s (Mitarai et al. 2008)). After initiation, the new protein must be elongated, after which it must fold into its active conformation. Both of these processes take time to complete, and introduce a delay between the initiation of gene expression and the appearance of the first active proteins.

Including the above delays into the translation reaction (3.9) results in the following (Ribeiro et al. 2006):

$$\mathrm{RBS} + \mathrm{Rib} \xrightarrow{k_{tr}} \mathrm{RBS}(\tau_{rbs}) + \mathrm{Rib}(\tau_{rib}) + \mathrm{P}(\tau_{protein}) \tag{3.14}$$

where $\tau_{rbs}$ is the time to release the RBS after initiating translation, $\tau_{rib}$ is the time to complete translation of the protein, and $\tau_{protein}$ is $\tau_{rib}$ plus the additional time for the protein to fold.

The above model of transcription and translation has been used to investigate, among others, the importance of the open complex formation in gene expression (Ribeiro et al. 2010), and the dynamics of several genetic circuits with delays, including the Toggle Switch and the Repressilator (Zhu et al. 2007; Ribeiro 2007a; Ribeiro 2007b; Ribeiro 2008). Delayed reactions, including the above model of transcription and translation, can be found in all publications in this thesis.

### 3.2.2 Regulation

The previous sections have dealt with modelling the expression of a single gene. To form a network of genes, we must model gene regulation. As described in section 2.2, many genes are regulated by one or more TFs which bind to specific operator sites in the promoter region of a gene. For example, the Lac promoter in *E. coli* can be bound by two TFs: LacI and CAP. The former is a repressor which unbinds from the promoter in the presence of lactose, while the latter is an activator in the presence of cAMP. Combined, these two regulators allow the expression of the *lacZYA* operon under appropriate conditions.

Regulation of a gene by a TF can be modelled as follows (Roussel and Zhu 2006; Ribeiro and Lloyd-Price 2007):

$$\mathrm{Pro} + \mathrm{TF} \xrightarrow{k_r} \mathrm{Pro} \cdot \mathrm{TF} \tag{3.15}$$

$$\mathrm{Pro} \cdot \mathrm{TF} \xrightarrow{k_u} \mathrm{Pro} + \mathrm{TF} \tag{3.16}$$

Here, the TF repeatedly binds and unbinds from the operator site at the promoter region of the gene. While in the Pro · TF state, the RNAp cannot bind to the promoter to initiate transcription with reaction (3.8). These reactions will thus reduce the expression rate of the gene by the fraction of time the TF is bound to the promoter, making this a repressive interaction. To model activation, an additional reaction is required (Ribeiro and Lloyd-Price 2007):

$$\mathrm{Pro} \cdot \mathrm{TF} + \mathrm{RNAp} \xrightarrow{k_t^{act}} \mathrm{Pro} \cdot \mathrm{TF} + \mathrm{RBS} + \mathrm{RNAp} \tag{3.17}$$

where $k_t^{act} > k_t$.

Generally, the effects of TF-based interactions such as these are well-approximated by a Hill function, a function which smoothly interpolates from full production to full repression (or vice-versa) as the number of TFs is increased. In particular, provided that the bind-unbind reactions are fast, the reactions given above result in a Hill function with a hill coefficient of 1 (see section 3.4.1 for a derivation):

$$P(\mathrm{Pro} = 1) = \frac{K_d}{K_d + [\mathrm{Rep}]}, K_d = \frac{k_u}{k_r} \tag{3.18}$$

Using the above modelling strategy, multiple TF binding sites can be modelled for a single promoter (for an example, see the supplementary information in **Publication III**). Further, any combinatorial effects between them can also be modelled (see e.g. (Arkin et al. 1998)). Genetic networks using these kinds of interactions were studied in **Publication IV**, and were used as examples in **Publication I**.

In **Publication II**, however, the dynamics was investigated of a network which utilized a post-transcriptional regulatory mechanism: direct RNA-RNA interaction resulting in the silencing of the target gene (described in section 2.2). To summarize, the RNA molecule produced by one gene does not code for a TF. Instead, it is the complement of the mRNA of its target gene. When the two bind, they are both degraded, thus silencing the target. This interaction can be modelled with the following reactions (Levine and Hwa 2008):

$$\mathrm{Pro}_{\mathrm{srRNA}} + \mathrm{RNAp} \xrightarrow{k_t^{\mathrm{srRNA}}} \mathrm{Pro}_{\mathrm{srRNA}} + \mathrm{srRNA} + \mathrm{RNAp} \tag{3.19}$$

$$\mathrm{RBS} + \mathrm{srRNA} \xrightarrow{k_s} \varnothing \tag{3.20}$$

where RBS is that of the target gene, and $\mathrm{Pro}_{\mathrm{srRNA}}$ is the promoter of the srRNA gene. Here, reaction (3.19) represents the transcription of the srRNA, and reaction (3.20) represents the silencing of its target. In contrast to the TF-based regulation described above, these reactions result in a threshold-linear regulation function, as depicted in Figure 2.3b.

## 3.3 Simulation Algorithms

The preceding models are all built within the stochastic formulation of chemical kinetics. As such, their dynamics is described by the CME built from the set of reactions in the model. However, even for the simplest models involving only a few interacting molecular species, directly solving the CME becomes an intractable problem, having only been solved analytically for systems containing only monomolecular reactions (Jahnke and Huisinga 2007). Instead, we opt to *simulate* the dynamics of the model by sampling trajectories from the distribution described by the CME. The algorithm underlying these simulations is the Stochastic Simulation Algorithm (SSA) (Gillespie 1976; Gillespie 1977; Gillespie 1992; Gillespie 2007).

### 3.3.1 Stochastic Simulation Algorithm

The SSA produces a sample from the distribution of trajectories through the chemical system's state space. It does so by repeatedly answering two questions: when does the next reaction occur? and which reaction is it? Formally, this means selecting a time $\tau$ until the next reaction occurs, and the reaction to occur $\mu$, from suitable probability distributions described by the CME.

To answer these questions, first consider only a single reaction $\mu$. At the current time $t$ and state $\mathbf{x}$, this reaction has propensity $a_\mu(\mathbf{x})$. We can then ask: assuming that no other reactions occur before $\mu$, how long must we wait until this reaction occurs? This question is answered by the product of $P_0(\tau_\mu)$, the probability that reaction $\mu$ does *not* occur between $t$ and $t+\tau_\mu$, and the probability that it then *does* occur in the next infinitesimal time interval $(t+\tau_\mu, t+\tau_\mu+d\tau_\mu)$. From the definition of $a_\mu(\mathbf{x})$, $\tau_\mu$ can be shown to follow an exponential distribution (Gillespie 1976):

$$P(\tau_\mu)d\tau_\mu = P_0(\tau_\mu)\cdot a_\mu(\mathbf{x})d\tau_\mu = a_\mu(\mathbf{x})e^{-a_\mu(\mathbf{x})\tau_\mu}d\tau_\mu \tag{3.21}$$

It would then be possible to construct a simulation algorithm by sampling a "next reaction time" for all reactions, and then answering the two questions from the reaction with the earliest next reaction time. This approach is called the First Reaction Method (FRM) (Gillespie 1976). In this method, however, all next reaction times must be regenerated every time these questions are answered, since we assumed above that no other reaction occurs before the next reaction time. This therefore requires a significant amount of work to be done every iteration of the algorithm.

Instead, in the original formulation of the SSA, another approach was proposed based on a direct sampling the distributions of the two answers, and is therefore called the Direct Method (DM) (Gillespie 1976; Gillespie 1977). An informal derivation of this method is as follows (see (Gillespie 1976) for details). The distribution of the earliest next reaction time is the distribution of the minimum of all next reaction times. The minimum of a set of independent exponential distributions with different rates happens to itself be an exponential distribution with a rate equal to the sum of the individual exponentials' rates (Gillespie

1976). Therefore, the distribution of $\tau$, the time until the next reaction occurs, independent of which reaction it happens to be, is:

$$\begin{aligned} P(\tau)d\tau &= a_0(\mathbf{x})e^{-a_0(\mathbf{x})\tau}d\tau, \\ a_0(\mathbf{x}) &= \sum_{\mu} a_{\mu}(\mathbf{x}) \end{aligned} \tag{3.22}$$

Further, the probability that a given exponential is the minimum is proportional to the rate of the exponential (Gillespie 1976). Thus, the next reaction to occur, $\mu$, follows a multinomial distribution:

$$P(\mu) = \frac{a_{\mu}(\mathbf{x})}{a_0(\mathbf{x})} \tag{3.23}$$

Samples for $\tau$ and $\mu$ can therefore be generated from two uniformly distributed random numbers in $(0, 1)$, $U_1$ and $U_2$, as follows:

$$\tau = \frac{-\ln U_1}{a_0(\mathbf{x})} \tag{3.24}$$

$$\sum_{m<\mu} a_m(\mathbf{x}) \leq U_2 a_0(\mathbf{x}) < \sum_{m\leq\mu} a_m(\mathbf{x}) \tag{3.25}$$

Using these formulas, we can sample from the *exact* distribution that is described by the CME, i.e. we have made no approximations in their derivation. In that sense, since both the CME and the SSA are derived from the same set of theorems, they can be considered to be logically equivalent (Gillespie 1992).

**Algorithm 1** Stochastic Simulation Algorithm (Direct Method)

```
$t \leftarrow t_0$
$\mathbf{x} \leftarrow \mathbf{x}_0$
while $t < t_{stop}$ do
    $a_0 \leftarrow \sum_{\mu} a_{\mu}(\mathbf{x})$
    $U_1, U_2 \leftarrow$ Independent uniform random numbers in $(0, 1)$
    $\tau \leftarrow -a_0^{-1} \ln U_1$
    $\mu \leftarrow \mu$ such that $\sum_{m<\mu} a_m(\mathbf{x}) \leq U_2 a_0 < \sum_{m\leq\mu} a_m(\mathbf{x})$
    $t \leftarrow t + \tau$
    $\mathbf{x} \leftarrow \mathbf{x} + \nu_{\mu}$
```

Given a starting time $t_0$, a stopping time $t_{stop}$, and an initial vector of species populations $\mathbf{x}_0$, the DM of the SSA is given in Algorithm 1 (Gillespie 1977).

### 3.3.2 Next Reaction Method

Both the DM and the FRM must perform an $O(R)$ operation for every reaction actually performed, where $R$ is the number of possible reaction channels in the system being simulated. In the case of the FRM, this is the generation of all the next reaction times and selection of the earliest, while in the DM, this is the calculation of $a_0(\mathbf{x})$ (hereafter abbreviated as $a_0$), and the selection of $\mu$ (lines 4 and 7 of Algorithm 1). Clearly, this will cause the simulation to run slowly when $R$ is large. Several alternative algorithms have thus been suggested to accelerate the simulation without compromising the exactness of the algorithm. The simulator in **Publication I** uses an optimization of the FRM called the Next Reaction Method (NRM) (Gibson and Bruck 2000).

The NRM is based firstly on the observation that not all of the next reaction times generated by the FRM need to be discarded when a particular reaction occurs. If the propensity of a reaction $m$ does not change due to the occurrence of reaction $\mu$, then the distribution of the next reaction time of $m$, given that $\tau_\mu$ time has passed since it was last generated *and* this reaction has not occurred yet, is the same as the distribution before that time had passed. This is due to the memoryless property of the exponential distribution from which $\tau_m$ was drawn. That is, from equation (3.21), for reaction $m \neq \mu$, the distribution of time until reaction $m$ occurs after reaction $\mu$ is:

$$\begin{aligned}
P(\tau_m - \tau_\mu | \tau_m > \tau_\mu) d\tau_m &= \frac{P(\tau_m - \tau_\mu) d\tau_m}{P(\tau_m > \tau_\mu)} \\
&= \frac{a_m(\mathbf{x}) e^{-a_m(\mathbf{x})(\tau_m - \tau_\mu)} d\tau_m}{e^{-a_m(\mathbf{x})\tau_\mu}} \\
&= a_m(\mathbf{x}) e^{-a_m(\mathbf{x})\tau_m} d\tau_m \\
&= P(\tau_m) d\tau_m
\end{aligned} \tag{3.26}$$

Therefore, if we store the putative next reaction times for each reaction as absolute times $t_\mu$, rather than relative times $\tau_\mu$, the majority will not need to be regenerated from one iteration of the algorithm to the next. The NRM therefore proceeds as follows (Gibson and Bruck 2000). In a priority queue, we store the absolute putative reaction times of all reactions. The reaction at the front of the priority queue, which can be found in $O(1)$ time, is then always the next reaction to occur. Upon execution of this reaction, it and any reaction whose propensity has been changed by the occurrence of this reaction then have their putative next reaction times regenerated, and reinserted into the priority queue. To accelerate this process, the reaction dependency graph, i.e. the list of reactions which will require their propensities to be updated for every possible occurring reaction, is calculated and stored beforehand.

Since this algorithm requires the possibility to remove an arbitrary reaction from the priority queue, a simple implementation such as a binary heap will not suffice. Instead, we must maintain the mapping between the reactions and their present location in the priority queue. The resulting data structure is called an indexed priority queue (Gibson and Bruck 2000), an example of which is illustrated in Figure 3.1.

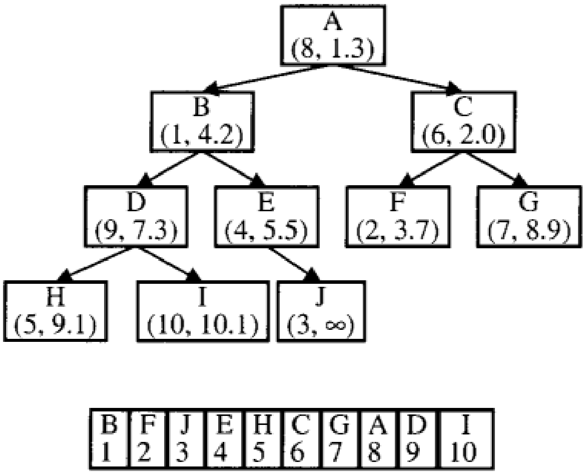


**Figure 3.1:** Example indexed priority queue used by the Next Reaction Method, partway through a simulation with 10 reactions. Top: Binary heap structure with nodes are labelled with letters A-J. Each node contains the index of the reaction to occur ($\mu$), and the putative reaction time of that reaction ($t_\mu$). Bottom: Index structure, containing the mapping from reaction indices to the current position in the heap of that reaction. Reprinted with permission from (Gibson and Bruck 2000). Copyright (2000) American Chemical Society.

Finally, even in the case where the propensity of a reaction does change due to the occurrence of another reaction, it is possible to reuse the previously-generated putative reaction time. Since the time until the next occurrence of a given reaction, after the occurrence of the currently executing reaction, is exponentially-distributed (from equation (3.26)), and the distribution from which we generate the time until the next putative reaction time is also exponentially-distributed, we can simply scale the old exponential distribution to match the rate of the new exponential distribution, rather than generating an entirely new putative reaction time. The scaling factor required is the ratio between the new propensity and the old propensity (Gibson and Bruck 2000). This can be seen by first transforming the old exponential distribution from equation (3.26) to a uniform distribution, and using that as $U_1$ in equation 3.24. After scaling, rather than removing and reinserting the reaction into the priority queue, the reaction is simply moved up/down in the priority queue to its appropriate position, an operation termed "bubbling up/down". When the ratio between the old and new propensities is near 1, this has the advantage that the reaction will not need to be bubbled far from its current location, thus reducing the cost of this operation in practice (Gibson and Bruck 2000). Combining all of the above, we get the NRM, presented in Algorithm 2.

If the NRM's indexed priority queue is implemented based on a binary heap, as depicted in Figure 3.1, insertion, deletion, and bubbling operations all take

**Algorithm 2** Stochastic Simulation Algorithm (Next Reaction Method)

```
$t \leftarrow t_0$
$\mathbf{x} \leftarrow \mathbf{x}_0$
$Q \leftarrow$ Empty indexed priority queue
for each reaction $\mu$ do
    $U \leftarrow$ Uniform random number in $(0, 1)$
    $t_\mu \leftarrow t - a_\mu(\mathbf{x})^{-1} \ln U$
    Insert reaction $\mu$ into $Q$ with putative reaction time $t_\mu$
while $t < t_{stop}$ do
    $\mu, t_\mu \leftarrow$ Earliest reaction in $Q$
    Pop the earliest reaction from $Q$
    $t \leftarrow t_\mu$
    $\mathbf{x} \leftarrow \mathbf{x} + \nu_\mu$
    $U \leftarrow$ Uniform random number in $(0, 1)$
    $t_\mu \leftarrow t_\mu - a_\mu(\mathbf{x})^{-1} \ln U$
    Insert reaction $\mu$ into $Q$ with putative reaction time $t_\mu$
    for each reaction $m$ for which $m \neq \mu$ and $a_m(\mathbf{x}) \neq a_m(\mathbf{x} - \nu_\mu)$ do
        $t_m \leftarrow t + a_m(\mathbf{x} - \nu_\mu) a_m(\mathbf{x})^{-1} (t_m - t)$
        Bubble up/down reaction $m$ in $Q$ as necessary
```

$O(\log R)$ time. Thus, so long as the reaction dependency graph is sparse, i.e. the maximum number of reactions that must have their putative reaction time updated when a given reaction is executed is independent of $R$, the inner loop of the NRM runs in $O(\log R)$ time (Gibson and Bruck 2000). When $R$ is large, this results in a significant boost to the speed of the simulation compared to the DM and the FRM. Further, since the NRM utilizes a general priority queue, it is has the additional advantage of being able to simulate events with non-exponential waiting times. For these reasons, along with the ease of adding and removing reactions from the system, the NRM was used as the basis of the simulation in **Publication I**.

### 3.3.3 Delayed Stochastic Simulation Algorithm

The SSA does not allow for explicit delays to be simulated (i.e. reactions like those presented in section 3.2.1). Delays can be incorporated into the SSA with the use of a "wait list", a priority queue maintained in parallel to the SSA where the products from delayed reactions in the past are stored until the appropriate release time (Roussel and Zhu 2006). The Delayed SSA (DSSA) proceeds by repeatedly executing a reaction or releasing a delayed product from the wait list, whichever event is earlier. The DSSA is presented in Algorithm 3.

Using a simple binary heap to implement the DSSA's wait list results in an $O(\log W)$ addition to the runtime of the simulation's main loop, where $W$ is

**Algorithm 3** Delayed Stochastic Simulation Algorithm

```
$t \leftarrow t_0$
$\mathbf{x} \leftarrow \mathbf{x}_0$
$L \leftarrow$ Empty wait list
while $t < t_{stop}$ do
    $\tau, \mu \leftarrow$ SSA
    if $L$ is empty then
        $t_L \leftarrow \infty$
    else
        $t_L \leftarrow$ Earliest time in $L$
    if $t_L < t + \tau$ then
        Pop the earliest molecule in $L$ and add it to $\mathbf{x}$
        $t \leftarrow t_L$
    else
        $t \leftarrow t + \tau$
        $\mathbf{x} \leftarrow \mathbf{x} + \nu_\mu$
        if Reaction $\mu$ has delayed products then
            Add them to $L$
```

the number of elements on the wait list. However, except for pathological wait list-heavy models, the reaction events will generally outweigh the wait list events, whatever implementation of the SSA is chosen. Thus the SSA's runtime will eclipse the DSSA's additional overhead. The NRM is particularly well-suited to be paired with the DSSA, since both the reactions and wait list events can share the same priority queue, unifying the two algorithms into one.

## 3.4 Approximate Simulation

Several publications in this thesis employ approximate simulation techniques. These techniques make simplifying assumptions about the model to reduce the computational complexity of the simulation, in the same manner as the CME can be derived from a full molecular dynamics simulation by making the simplifying assumptions given in section 3.1, e.g. that all molecules are uniformly distributed within the reaction volume.

One such technique has already been applied informally in the preceding sections. In section 3.2, it was noted that it is possible to remove one of the species from consideration in a reaction if it can be considered to be in constant concentration in the cell. In general, simplification of models in this way uses information known by the modeller that some feature of the detailed model is not relevant (or has little relevance) to the features under study. An approximation of this feature is then made, resulting in a new, simpler model.

Two pieces of information frequently used in model reduction for models of genetic circuits are timescale separation, and the knowledge that a particular molecule will not significantly change concentration during the simulation. Such techniques were applied in **Publication III** to produce a “reduced” model of gene expression to simplify the parameter fitting procedure (see the supplementary information), and in publication **Publication IV** to produce a simple enough model of a Toggle Switch so that extremely large populations of cells could be simulated simultaneously.

### 3.4.1 Timescale Separation

The different subsystems of a genetic network do not operate on the same timescale. For example, TFs can interact with the promoter region of their target gene very rapidly - on the order of tens of seconds (Dunaway et al. 1980) - while transcription initiation occurs on a timescale of several minutes (McClure 1980). When two subsystems operate on sufficiently different timescales, the system is amenable to simplification by timescale separation (Gunawardena 2014). In this, the faster process is assumed to always be in steady state, and downstream slow components are affected by only considering the expected behaviour of the fast components. In principle, this means that the faster components, and the many reaction events they create in the SSA, do not need to be simulated explicitly, and the model can be simulated considerably faster.

A common application of timescale separation is the example given above: the regulation of a target gene by a transcription factor. Recall that the reactions for repression are as follows (equivalent to reactions (3.15) and (3.16)):

$$\text{Pro} + \text{Rep} \xrightarrow{k_r} \text{Pro} \cdot \text{Rep} \tag{3.27}$$

$$\text{Pro} \cdot \text{Rep} \xrightarrow{k_u} \text{Pro} + \text{Rep} \tag{3.28}$$

These two reactions interact with transcription by intermittently removing the gene’s promoter from from the system. If the repressor bind-unbind reactions occur far more frequently than the transcription reaction (i.e. $k_u \gg [\text{RNAp}]k_t$), then timescale separation can be applied, and the system can be simplified as follows (Cao et al. 2005). First, we write the CME for the Rep-Pro subsystem for a fixed population of Rep, which only contains two states:

$$\frac{dP(\text{Pro} = 1)}{dt} = -k_r[\text{Rep}]P(\text{Pro} = 1) + k_u P(\text{Pro} = 0) \tag{3.29}$$

$$\frac{dP(\text{Pro} = 0)}{dt} = k_r[\text{Rep}]P(\text{Pro} = 1) - k_u P(\text{Pro} = 0) \tag{3.30}$$

where [Rep] is the population of Rep molecules.

Solving for steady state, and using the fact that the total probability must be 1 (i.e. $P(\text{Pro} = 0) + P(\text{Pro} = 1) = 1$), we get:

$$P(\text{Pro} = 1) = \frac{K_d}{K_d + [\text{Rep}]}, K_d = \frac{k_u}{k_r} \tag{3.31}$$

The fast species (Pro and Pro · Rep) are then removed from the rest of the system. This is done by replacing all reactions involving them, such as transcription, with reactions with the stochastic constant modified to be the expectation of the stochastic constant of the original reaction, given the steady-state probability distribution of the fast species (Cao et al. 2005). In this case, the transcription reaction (3.8) becomes:

$$\text{RNAp} \xrightarrow{\frac{k_t K_d}{K_d + [\text{Rep}]}} \text{RBS} + \text{RNAp} \tag{3.32}$$

The above is the justification behind the use of Hill functions to represent the effects of a TF on its target gene. For this reason, reactions with a Hill function term are made available in the simulator presented in **Publication I**, and this facility is used to simplify the models employed in **Publication IV**.

### 3.4.2 Constant Concentrations

In addition to taking advantage of timescale separation, models can also be simplified if it is known that the concentration of a molecule will be approximately constant for the duration of the simulation. This implies that the molecule is in high copy number, and thus fluctuations in its numbers as it reacts with the other molecules in the system will be negligible. It can then be removed from explicit consideration in the system by factoring its contribution into the propensities of each reaction it takes part in.

This technique has already been informally mentioned earlier when referring to the RNA polymerase and Ribosome concentrations in section 3.2. Both of these are housekeeping molecules which are constitutively expressed by cells, and can therefore be considered to be in roughly constant concentration. For example, if the RNAp concentration is considered to be constant, then the transcription reaction (3.8) becomes:

$$\text{Pro} \xrightarrow{k_t[\text{RNAp}]} \text{Pro} + \text{RBS} \tag{3.33}$$

This was applied in all publications to remove non-dynamically-relevant molecules from explicit representation in the simulation. As an example, see the derivation of the “reduced model” in the supplementary material of **Publication III**.

### 3.4.3 Approximate Simulation Algorithms

Another means to simplify the simulation procedure is to make approximations in the simulation algorithm itself. That is, we compromise the exactness of the SSA using a simplifying assumption, in order to gain a payoff in speed. If the assumption is valid for the model being simulated, then the resulting speedup frees computational resources which might be spent, e.g. providing a more exhaustive exploration of the parameter space, or performing more simulations to gain more certainty of the results. An optimal strategy, then, would be to select a simulation method that makes as many simplifying assumptions as possible, yet accurately captures all the dynamics of interest. Often, however, a simulation strategy is chosen without regard to the dynamics that may be relevant. Specifically, the traditional means to model and simulate systems like those presented here is to set up a system of Ordinary Differential Equations (ODEs), called Reaction Rate Equations (RREs), without regard for the assumptions that are implicit in their use (Gillespie 1992). For an example of the differences observed in a model built with RREs compared to a full simulation with the SSA, see **Publication II**, in which the RREs for the model of the srRNA-mediated switch are referred to as the "deterministic model". The relationship between the RREs and the CME (Gillespie 2007; Gillespie 2009) is thus described here.

In the derivation of the RREs from the CME, the first simplifying assumption to make is that there is some time $\tau$ over which the propensities of all reactions do not change significantly (Gillespie 2001). The number of times each reaction occurs in this time window will therefore follow a Poisson distribution with rate $a_\mu(\mathbf{x})\tau$. It is then possible to "leap" over $\tau$-sized blocks of time by generating a random Poisson-distributed number for each reaction, which are then performed that many times. Meanwhile, the SSA would have had to perform one iteration for every occurrence of every reaction. In this manner, this method, called $\tau$-leaping, lumps together many occurrences into a single operation, accelerating the simulation for highly-propense reactions.

Simplifying further, if we assume that the number of times each reaction occurs within the time $\tau$ is large, i.e. $a_\mu(\mathbf{x})\tau \gg 1$, then the Poisson distribution can be well-approximated by a Normal distribution (Gillespie 2007). The result is a set of coupled stochastic differential equations known as Chemical Langevin Equations (CLEs) (Gillespie 2000). In this, the molecule populations become continuous, though the dynamics remains stochastic. Similar to the relationship between the SSA and the CME, the evolution of the CLE can be described by a partial differential equation called the Chemical Fokker-Planck Equation (CFPE), describing the evolution of the probability density over all the state space (Gillespie 1996; Gillespie 2000).

The final simplification is termed the "thermodynamic limit", wherein both the reaction volume and all molecular populations are taken to infinity, but in a way such that the concentrations of the molecules remain constant (Gillespie 2007; Gillespie 2009). Since the deterministic term of the CLE scales linearly with the molecular populations, this term approaches a non-zero value in this limit. However, the stochastic term scales as the square root of the populations. In this limit, therefore, this term disappears and we are left with a set of coupled ODEs describing the time-evolution of the concentrations of all molecules, called the RREs.

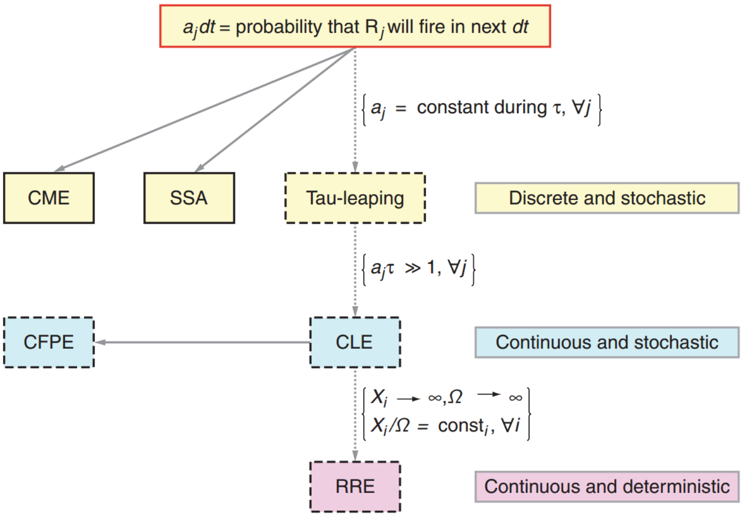

$a_j dt$ = probability that $R_j$ will fire in next $dt$
$\{a_j$ = constant during $\tau, \forall j\}$
CME
SSA
Tau-leaping
Discrete and stochastic
$\{a_j\tau \gg 1, \forall j\}$
CFPE
CLE
Continuous and stochastic
$X_i \to \infty, \Omega \to \infty$
$X_i/\Omega$ = const$_i$, $\forall i$
RRE
Continuous and deterministic

**Figure 3.2:** Logical relationship between the various descriptions of chemical kinetics and their simulation algorithms. Dotted arrows represent approximations, with their simplifying assumptions beside them. Solid arrows represent exact derivations. Methods in solid boxes are exact, i.e. derived from the fundamental assumption that a function $a_\mu(\mathbf{x})$ exists, while dashed boxes are approximate methods. Here, $\Omega$ represents the reaction volume. Reproduced with permission from (Gillespie 2007).

The chain of assumptions required to derive the RREs from the CME, and the simulation methods resulting from each assumption, are depicted in Figure 3.2. Methods higher in the hierarchy make fewer assumptions about the model, and are thus more computationally-demanding than those lower in the hierarchy. Since they produce the most physically-meaningful results, it may be tempting to always attempt to use a simulation method that makes the minimum number of assumptions, regardless of the potential speed increases they can bring. However, beyond showing that these assumptions are justifiable for the given model, forcing oneself to work at a higher level of detail than necessary can be detrimental.

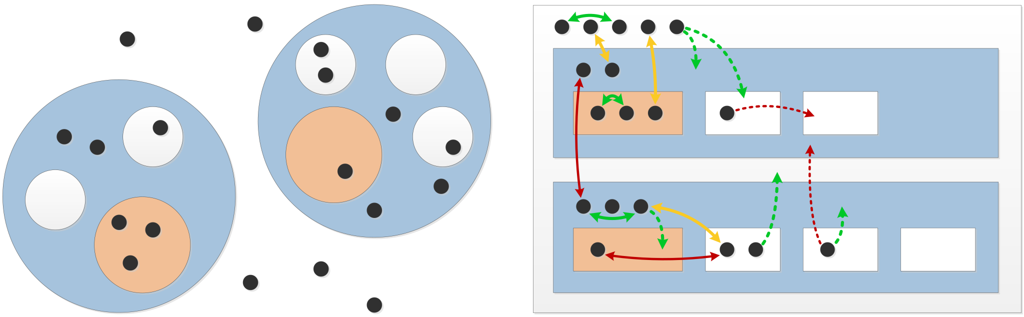
First, fewer assumptions also require the modeller to provide more information to the simulation, in the form of additional parameters and initial conditions, which may be difficult to acquire and which is not relevant to the dynamics under study. Further, these methods produce a correspondingly larger amount of output which must be properly analysed. Second, when the assumptions are reasonable, the time taken to simulate and analyse an excessively detailed model would be better spent providing a more in-depth study of the simplified model, such as the study of the high-level model of the Toggle Switch in populations of growing and dividing cells in **Publication IV**. The optimal strategy would therefore be to select the maximum number of assumptions that can be justified for a given model, and to then simulate it with the corresponding method.

Unfortunately, none of the simulation methods are easily parallelizable in general, and cannot be since it is always possible that a change in a given molecular species rapidly propagates to other molecular species. Thus, the simulation of large reaction systems cannot be efficiently split across multiple computers for any but the most loosely-coupled systems. This limits the use of parallel computational resources such as clusters or computational grids to running large amounts of independent simulations. For some system size, therefore, it will always be impractical to perform the simulation at the highest levels of detail, and some simplifying assumptions will become necessary to achieve any results. For example, we cannot simulate an entire cell using the level of the CME/SSA, but by applying the above assumptions where appropriate, it is starting to be possible to perform whole-cell simulations which yield useful predictions (Karr et al. 2012).

## 3.5 Compartments

Real cells are not homogeneous, and instead have an intricate internal spatial structure and organization with volumes delineated by membranes and with molecules tethered together by macromolecules (Alberts et al. 2002). This organization can affect the probabilities that certain reactions occur, and violates the assumption of spatial homogeneity made by the CME and SSA. However, as noted earlier, simulating every particle explicitly in space is too computationally demanding. A commonly-taken middle ground is to divide the space into compartments, with a CME-based simulation running in each compartment. This partitioning of space can be done based on the actual physical separations in the cell, resulting in stochastic P-systems (Păun 2001; Spicher et al. 2008), or a more refined partitioning can be performed, resulting in the Inhomogeneous SSA (ISSA) for reaction-diffusion systems (Lampoudi et al. 2009).

In P-systems, compartments are organized in a hierarchical manner (Păun 2001; Spicher et al. 2008). Within each compartment, the usual CME assumptions are made, allowing the SSA to be used. Further, all communication between compartments occurs by diffusion reactions between a parent compartment and a child compartment, i.e. there is no direct child-child communication. This implies that both the molecules within a compartment, as well as all compartments contained within a compartment are always uniformly distributed. As examples, in a model of a eukaryotic cell, molecules within a Nucleus compartment might be able to diffuse out into the containing Cell compartment (representing the cell cytoplasm), which then diffuse into a Mitochondrion compartment. However, molecules cannot diffuse directly from the Nucleus compartment to the Mitochondrion compartment. An example hierarchically compartmentalized system representing cells with structures within is shown in Figure 3.3a. Examples of allowed reactions between molecules in compartments, as described above, are shown in Figure 3.3b.



**(a)** Example of a hierarchically compartmentalized system: two cells (blue) with different structures inside (white and orange), all containing interacting and diffusing molecules (black).

**(b)** Compartment hierarchy for the example cells pictured in Figure 3.3a. Examples of some possible reactions (solid green arrows) and diffusion reactions (dotted green arrows) in P-systems are shown. Examples of additional reactions made possible in SGNS2 are shown in yellow. Invalid reactions and diffusion reactions are shown in red.

**Figure 3.3:** Example of a compartment hierarchy.

On the other hand, in the ISSA, an explicit spatial model of the reaction volume is divided into subcompartments. As with P-systems, molecules are assumed to be homogeneously distributed within each compartment, however here there is no compartment hierarchy. Molecules are instead allowed to diffuse from one compartment to adjacent compartments. In theory, as more subcompartments are used to represent the reaction volume, the simulation becomes more accurate, however the diffusion reactions will quickly dominate the time taken to simulate the system (Lampoudi et al. 2009).

While the ISSA is more physically accurate, P-systems are considerably easier to set up and reason about. Further, since they do not explicitly model space, the division of compartments in P-systems is simple to implement - a new compartment is simply created in the parent compartment of the dividing compartment. For these reasons, the simulator presented in **Publication I** is based on P-systems, with one important limitation removed. Bimolecular reactions are allowed to occur between molecules in a child compartment and molecules in a parent compartment (the yellow interactions in Figure 3.3b). This enhancement allows compartments to be used to simulate the spatial restrictions created by interactions with a single macromolecule. For example, this ability was used to simulate coupled transcription and translation in (Mäkelä et al. 2011), where the global compartment contained the DNA and individual nacent RNA molecules were each contained within their own subcompartments. Such cross-compartment reactions were used in the transcription elongation reaction, Ribosome-RNAp interaction, and when Ribosomes in the global compartment interacted with the RNA, while intra-compartment reactions ensured that Ribosomal traffic on individual RNA molecules was correctly modelled.

### 3.5.1 Partitioning of Molecules

P-systems provide a formalism in which dynamic cell populations can be simulated. As mentioned, compartments can be divided by creating new compartments at the same level in the compartment hierarchy as the original compartment. To correctly model cell partitioning using such a methodology, one additional rule must be defined: how the molecules are partitioned between the two new compartments upon division. The partitioning schemes mentioned in section 2.1, and summarized in Figure 3.4, are presented here as "mock processes" that resemble the process resulting in the partitioning, and generate the appropriate distribution, as in (Huh and Paulsson 2011b).

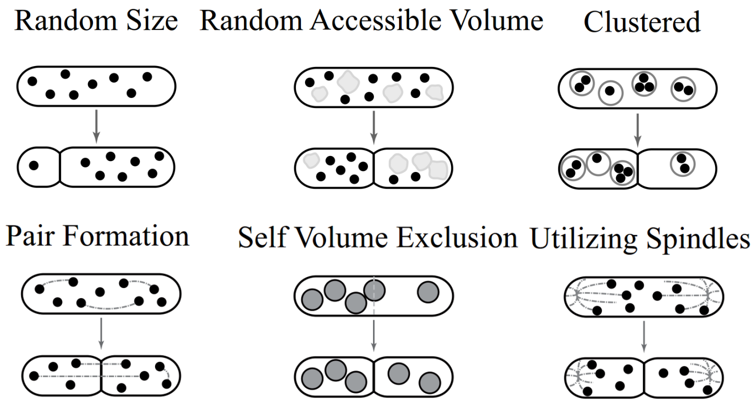


**Figure 3.4:** Molecule partitioning schemes in cell division presented in (Huh and Paulsson 2011b). Top: "Disordered" partitioning schemes, resulting in greater-than-binomial partitioning error. Bottom: "Ordered" partitioning schemes, resulting in more-even-than-binomial partitioning. Reprinted with permission from (Huh and Paulsson 2011b). Copyright Dann Huh, 2011.

These mock processes are presented here as reaction systems where the molecule M refers to the molecule being partitioned. To determine how the molecules are partitioned, the mock processes are run in isolation from the rest of the system until $t = \infty$, and the number of L and R molecules at that point are then taken to be the number of molecules partitioned into the "Left" and "Right" daughter compartments, respectively. For example, consider the mock process for an independent partitioning:

$$\mathrm{M} \xrightarrow{1} \mathrm{L} \tag{3.34}$$

$$\mathrm{M} \xrightarrow{1} \mathrm{R} \tag{3.35}$$

Every M molecule has an independent and equal chance of being partitioned into either new compartment. After this mock process has been run to $t = \infty$, the number of L molecules will thus follow a binomial distribution with $N$ set to the number of molecules which were partitioned and $p = 0.5$.

The Random Size partitioning scheme, in which the daughter cells inherit a different amount of the parent cell's cytoplasm, can be simulated by a mock process which biases the molecule partitioning towards the larger cell. If $\Omega$ is the volume of the cell, and $v_L$ is the volume of the Left daughter in this particular division, then this partitioning scheme's mock process is as follows:

$$\mathrm{M} \xrightarrow{v_L} \mathrm{L} \tag{3.36}$$

$$\mathrm{M} \xrightarrow{\Omega - v_L} \mathrm{R} \tag{3.37}$$

Large molecules or structures such as vacuoles reduce the space available for other molecules in the cell. If these structures are partitioned randomly between the two daughter cells (likely by the Volume Exclusion scheme below), this will be an additional source of variability in the partitioning of other molecules. This partitioning scheme, labelled as Random Accessible Volume in Figure 3.4, is identical to the Random Size partitioning scheme, with $\Omega$ replaced with the total accessible volume in both cells, and $v_L$ replaced with the accessible volume in the Left daughter cell, after the vacuoles and other large structures have been partitioned by their partitioning schemes.

If the molecules cluster before partitioning, the cell inheriting the larger clusters will likely inherit more molecules than its sibling. Molecule clustering can be simulated as follows. If the number of clusters formed by the molecules is $C \geq 1$, then we first run a mock process partitioning the clusters into the two daughter cells:

$$\mathrm{C} \xrightarrow{1} \mathrm{C_L} \tag{3.38}$$

$$\mathrm{C} \xrightarrow{1} \mathrm{C_R} \tag{3.39}$$

The partitioned molecules are then partitioned into the cells, biased towards the cell which gained the most clusters:

$$\mathrm{M} \xrightarrow{\mathrm{C_L}} \mathrm{L} \tag{3.40}$$

$$\mathrm{M} \xrightarrow{\mathrm{C_R}} \mathrm{R} \tag{3.41}$$

This order of operations (partitioning clusters first, followed by partitioning of molecules into clusters), is equivalent to first partitioning the molecules into clusters, and then partitioning the clusters into cells (Huh and Paulsson 2011b). Note that all of the above partitioning schemes are variants of a biased (i.e. Preferential) independent partitioning scheme, where the bias is drawn from a distribution rather than fixed.

Molecules which are partitioned non-binomially due to limited diffusion must be handled separately, since their location before partitioning, a feature lost in these mock processes, is required. These can be simulated, for example, by tracking where these molecules are in the cell, and partitioning them appropriately, as was done in  (Gupta et al. 2014c).

The Pair Formation partitioning scheme is the first partitioning scheme considered here which can result in lower than binomial variance in the partitioning. This scheme is parametrized by the probability $r$ that a given pair of molecules will form a pair, and by $p$, the probability that a pair will partition evenly into the two daughter cells. An independent partitioning scheme is therefore realized when $r = 0$. Further, for $r > 0, p < 1$, it is possible for this scheme to result in greater variance than binomial, similar to clustered partitioning. This partitioning scheme can be simulated with the following mock process:

$$2\,\mathrm{M} \xrightarrow{\infty} \mathrm{ProtoPair} \tag{3.42}$$

$$\mathrm{M} \xrightarrow{1} \mathrm{I} \tag{3.43}$$

$$\mathrm{I} \xrightarrow{1} \mathrm{L} \tag{3.44}$$

$$\mathrm{I} \xrightarrow{1} \mathrm{R} \tag{3.45}$$

$$\mathrm{ProtoPair} \xrightarrow{1} 2\,\mathrm{I} \tag{3.46}$$

$$\mathrm{ProtoPair} \xrightarrow{r(1-r)^{-1}} \mathrm{Pair} \tag{3.47}$$

$$\mathrm{Pair} \xrightarrow{1} 2\,\mathrm{L} \tag{3.48}$$

$$\mathrm{Pair} \xrightarrow{1} 2\,\mathrm{R} \tag{3.49}$$

$$\mathrm{Pair} \xrightarrow{2p(1-p)^{-1}} \mathrm{L} + \mathrm{R} \tag{3.50}$$

Partitioning by a spindle apparatus, which has $\mathrm{S_L}$ binding sites for the Left cell, and $\mathrm{S_R}$ binding sites for the Right cell, can be simulated by first assigning each molecule to a spindle binding site:

$$\mathrm{S_L} + \mathrm{M} \xrightarrow{1} \mathrm{L} \tag{3.51}$$

$$\mathrm{S_R} + \mathrm{M} \xrightarrow{1} \mathrm{R} \tag{3.52}$$

Any remaining molecules are then partitioned independently using reactions (3.34) and (3.35).

Volume Exclusion, i.e. the partitioning of molecules so large that only a limited number fit into a given cell, results in a partitioning scheme very similar to the spindle binding sites. The primary difference is that the different “binding sites” represent the limited possible locations within the new cells that each macromolecule can occupy, and thus $\mathrm{S_L} + \mathrm{S_R}$ must be at least as large as M.

The simulator in **Publication I** was designed with the capability to generate these partitioning schemes, based on the mock processes presented here. These capabilities were used in **Publication III** and **Publication IV** to study the effects of molecule partitioning on the dynamics of single genes and on genetic circuits, as well as in  (Gupta et al. 2014c; Gupta et al. 2014a) to study the effects of asymmetric partitioning of non-functional protein aggregates on population vitality in *E. coli.*

# 4 Genetic Networks

It has been proposed that the “programming” of cells is encoded in the network of interactions between genes (Waddington 1957; Kauffman 1969). This section first presents the current view as to how a single genetic network can give rise to its diverse behaviours, and introduces theoretical concepts which appear frequently in **Publication II** and **Publication IV**. Subsequently, two specific genetic networks and their dynamics are presented: the Toggle Switch and the Repressilator. The behaviour of these two networks was studied in **Publication II** and **Publication IV**, where their behaviour was modified by stochastic partitioning in cell division and direct RNA-RNA interaction.

## 4.1 Noisy Attractors and Ergodic Sets

Boolean networks were one of the first dynamical representations of genetic networks studied (Kauffman 1969). In a Boolean network, each gene is represented by a Boolean variable, which is True when the gene is expressing (i.e. its product is present), and False when it is not. Connections between genes are represented by a Boolean function for each gene which determines what value that gene should have based on the states of all genes which might influence it. Time is discrete in this model, and at each time moment, each gene’s state is set to the state prescribed by its Boolean function based on the states of its inputs in the preceding time moment.

Though very simple, this model allowed several key insights to be made about how genetically identical cells might give rise to different phenotypes. First, notice that a given Boolean network with $N$ nodes has a large, though finite state space of $2^N$ states. Therefore, if the network is run long enough, at some point it must revisit a state which has already been visited. From this point on, it will continue to repeat the same sequence of states, since the update function is deterministic. Since there are many possible starting states which will eventually lead to a given repeating sequence of states, these loops in state space are called “attractors”.

Since a given Boolean network can have multiple attractors, these have been proposed as a model of how the interactions between genes in a pluricellular organism’s genetic network can give rise to many different cell types, each with its own set of expressed proteins. This interpretation has, however, been called into question, given the importance of stochasticity in gene expression. Boolean networks with noisy dynamics do not have attractors (Aldana et al. 2003). A related concept was therefore invoked to generalize this hypothesis to stochastic networks: Ergodic sets (Ribeiro and Kauffman 2007).

Ergodic sets are the sets of states which, once entered, cannot be left given some level of noise (Ribeiro and Kauffman 2007). Under the right conditions, Boolean networks subject to a given level of internal noise can have multiple such regions of the state space (Ribeiro and Kauffman 2007). Thus, Ergodic sets recapture the features necessary to explain how a single network can give rise to multiple behaviours, even in a stochastic setting. However, the definition of an Ergodic set is very harsh: a single noisy transition from one attractor of the non-noisy network to another suffices to merge both regions of the state space into a single Ergodic set. Nevertheless, noisy networks can remain in restricted regions of the state space for long periods of time. If this length of time is, on average, longer than the lifetime of a cell, then for all *practical* purposes, this region is an Ergodic set. Such regions of the state space, which exhibit long-term stability in the face of noise, are equivalently called “noisy attractors” (Dai et al. 2009) or “metastable states”.

In this thesis, noisy attractors feature prominently in the analysis of stochastic genetic circuits in **Publication II** and **Publication IV**. In the former, a network with two noisy attractors was analyzed, and the stabilities of the two attractors was quantified. In the latter, the effects of stochastic partitioning of molecules in cell division were found to differ in networks with differing numbers of noisy attractors: networks with one noisy attractor did not change their behaviour significantly, whereas networks with two noisy attractors gained qualitatively new features.

## 4.2 Toggle Switch

The Toggle Switch is an extensively studied genetic network (see e.g. (Arkin et al. 1998; Gardner et al. 2000; Atkinson et al. 2003; Lipshtat et al. 2006; Loinger et al. 2007; Zhu et al. 2007; Ribeiro 2007b)), comprised of two genes which mutually repress one another, as illustrated in Figure 4.1. The Toggle Switch is bistable, that is, it has two noisy attractors in which the system will tend to remain unless forced to change by an external signal or by spontaneous fluctuations in the RNA and transcription factors that compose the switch.

The canonical example of a Toggle Switch in nature is the “$\lambda$-switch” in the bacteriophage $\lambda$, composed of the TFs CI and Cro. In a landmark study which demonstrated the importance of stochasticity in gene expression, this phage was shown to exploit this stochasticity to make a randomized decision early during infection, between lysing the cell and turning it into a lysogen (Arkin et al. 1998).

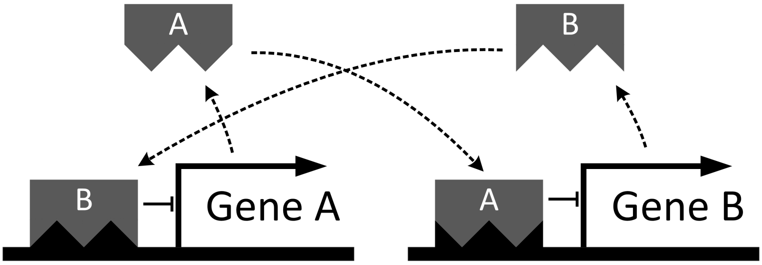


**Figure 4.1:** Illustration of the structure of the genetic Toggle Switch. Gene A produces transcription factor A, which binds to the promoter of gene B and represses it. Likewise, Gene B produces transcription factor B, which represses gene A.

If the virus's lytic pathway is disabled, populations of *E. coli* can be cultured for long periods of time with the switch in either state (Neubauer and Calef 1970), demonstrating how this circuit can also be used to store one bit of heritable epigenetic information.

To understand the dynamics of this circuit, first consider the following deterministic model (Gardner et al. 2000):

$$\frac{d[\mathrm{A}]}{dt} = \frac{\alpha_A K_B^\gamma}{K_B^\gamma + [\mathrm{B}]^\gamma} - \beta_A[\mathrm{A}] \tag{4.1}$$

$$\frac{d[\mathrm{B}]}{dt} = \frac{\alpha_B K_A^\gamma}{K_A^\gamma + [\mathrm{A}]^\gamma} - \beta_B[\mathrm{B}] \tag{4.2}$$

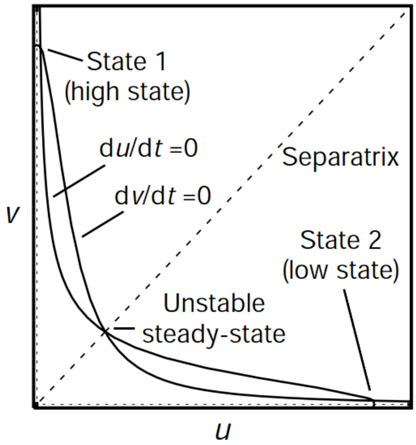


**Figure 4.2:** Phase space of the Toggle Switch. $u$ and $v$ refer to the populations of the two TFs. Solid lines represent the nullclines of the model. The switch has two steady states, labelled State 1 and State 2, separated by a separatrix. A third, unstable steady state lies between the two stable states. Reproduced with permission from (Gardner et al. 2000).

In this model, the parameter $\alpha_x$ controls the production rate of the protein for gene $x$, $\beta_x$ controls the protein's degradation rate, $K_x$ controls the number of repressors required to reduce the other gene's expression by half, and $\gamma_x$ controls the cooperativity of that interaction (explained below). The phase space of this model is shown in Figure 4.2 for a parameter set which results in bistability. As visible, there are two stable steady states, where one of the genes is expressing and the other gene is repressed. If the system is initialized anywhere in the upper triangle of the state space, the system will asymptotically approach the upper steady state, and vice-versa for the lower triangle and lower steady state. Note that in this case, the switch is symmetric, though this does not have to be the case in the likely scenario where $\alpha_A \neq \alpha_B$, $K_A \neq K_B$, or $\beta_A \neq \beta_B$. A separatrix lies in between the two stable states. In this deterministic model, initial conditions lying on this line will lead the system to the unstable steady state in the middle of the diagram.

The parameter $\gamma_x$ controls the cooperativity between TFs when repressing the target gene. When $\gamma_x = 1$, this corresponds to a single repressor protein binding/unbinding from the promoter (this function was derived in section 3.4.1). When $\gamma_x \neq 1$, TFs interact at the promoter regions of their target genes to produce a non-linear gene regulation function. In the limit of highly-cooperative binding, $\gamma_x$ will equal the number of binding sites for the TF, but can, in practice, take non-integer values. To achieve bistability in the deterministic model, $\gamma_x$ must be greater than 1 (Gardner et al. 2000).

In a discrete stochastic simulation, however, bistability can be achieved without cooperative interactions (Lipshtat et al. 2006). Further, while the deterministic model accurately predicts that the circuit can be bistable, it cannot predict how stable the steady states will be to noise since, if a steady state is reached, the system will remain there forever.

The delayed stochastic model of the Toggle Switch using the modelling strategy given in section 3.2 is as follows, where $i$ represents either gene A or gene B, and $j$ represents the other gene:

$$\mathrm{Pro}_i + \mathrm{RNAp} \xrightarrow{k_t} \mathrm{Pro}_i(\tau_{pro}) + \mathrm{RBS}_i(\tau_{pro}) + \mathrm{RNAp}(\tau_{rna}) \tag{4.3}$$

$$\mathrm{RBS}_i + \mathrm{Rib} \xrightarrow{k_{tr}} \mathrm{RBS}_i(\tau_{rbs}) + \mathrm{Rib}(\tau_{rib}) + \mathrm{P}_i(\tau_{protein}) \tag{4.4}$$

$$\mathrm{RBS}_i \xrightarrow{k_{rbsd}} \varnothing \tag{4.5}$$

$$\mathrm{P}_i \xrightarrow{k_{proteind}} \varnothing \tag{4.6}$$

$$\mathrm{Pro}_i + \mathrm{P}_j \xrightarrow{k_r} \mathrm{Pro}_i \cdot \mathrm{P}_j \tag{4.7}$$

$$\mathrm{Pro}_i \cdot \mathrm{P}_j \xrightarrow{k_u} \mathrm{Pro}_i + \mathrm{P}_j \tag{4.8}$$

Typical stochastic dynamics of a Toggle Switch is shown in Figure 4.3. The system is seen to stably remain in one of two states, with spontaneous switching events

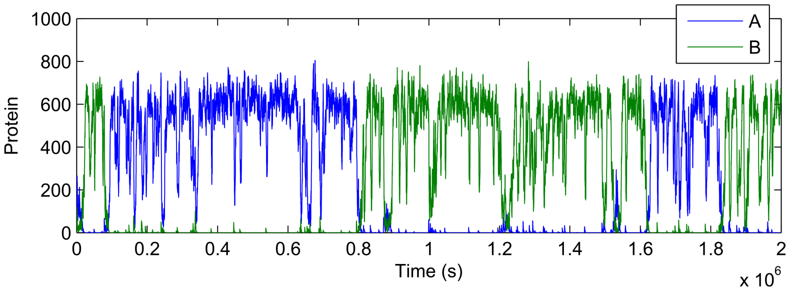


**Figure 4.3:** Typical dynamics of the stochastic model of a Toggle Switch given by reactions (4.3)-(4.8). Stochastic switching events can be seen (e.g. at $t = 0.8 \times 10^6$ s), in addition to failed switching events (e.g. at $t = 1.2 \times 10^6$ s).

between them. Further, some "failed" switching events can be seen. Thus, unlike in the deterministic model, the two possible states of the switch have a long, but limited lifetime. The stability $S$ of the stochastic switch is defined as the mean time that the switch can remain in one of its two noisy attractors, quantified as follows (Ribeiro 2007b):

$$S = \frac{T_{obs}}{W + 1} \tag{4.9}$$

where $T_{obs}$ is the total observation time and $W$ is the number of times $\mathrm{P_A} - \mathrm{P_B}$ changed sign in that observation time. However, a large amount of switching events (i.e. sign changes) are frequently generated when the switch's state lies near the unstable fixed-point attractor, e.g. at $0.88 \times 10^6$ s in Figure 4.3. Thus, switches that occur too soon after another switch should not be counted in $W$. Note that this definition of stability assumes a symmetric switch. If not, the state space of the switch must be characterized first in order to classify which noisy attractors the switch is in at a given point in time.

In general, stronger repressive interactions increase the stability of the switch (Loinger et al. 2007). Similarly, cooperative repressive interactions can greatly increase the stability of the switch (Loinger et al. 2007). Meanwhile, the various delays (see section 3.2.1) have differing effects, with the promoter open complex formation having the most complex interaction with the stability of the switch. When this delay is large, the mean TF population will decrease, thus decreasing the repression strength and destabilizing the switch. However, if $k_t$ is compensated so as to produce the same mean production rate (and thus the same mean protein level when unrepressed), the stability of the switch is still reduced. This is due to the weaker *relative* repression strength. That is, reaction (4.7) becomes less competitive with reaction (4.3). Lastly, coupling between Toggle Switches, either within the same cell, or due to communication between cells, will increase the stability of the switch (Ribeiro 2007b). This would therefore be one viable way to build a stable switch out of unstable switches.

The Toggle Switch can be seen as an example of epigenetic memory, which, when stable, can "store" one bit of information (Wolf and Arkin 2003). As such, circuits such as this have been proposed to underlie cell differentiation, where cells commit to one pathway over another (Gardner et al. 2000). This circuit can also be used to make randomized decisions, such as the "$\lambda$ switch" (Arkin et al. 1998). Lastly, a switch does not need to be perfectly stable to be useful. Unstable switches, i.e. those with a stability on the order of the length of the cell cycle or shorter, can be used as a survival strategy for a population of cells in a fluctuating environment (Acar et al. 2008). In this case, it is advantageous to have a stability such that the cell's phenotype switches at the same frequency as the environment.

In this thesis, the Toggle Switch appears in two publications: the $\lambda$ switch is used as an example application in **Publication I**, and it is one of the two networks studied in **Publication IV**.

### 4.2.1 RNA-Mediated Toggle Switch

A variation of the Toggle Switch, using RNA-RNA interactions rather than TF-DNA interactions, is the srRNA-mediated Double Negative Feedback Loop (MDNFL in (Zhou et al. 2012)). In this variant, one gene does not produce a TF, instead producing a srRNA which binds to, and silences the other gene's mRNA. This network motif can be found in both prokaryotes and eukaryotes (Zhou et al. 2012). The circuit is illustrated in Figure 4.4.

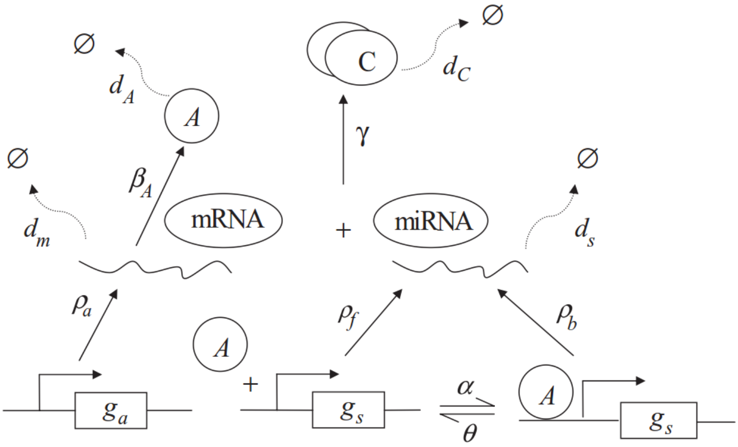


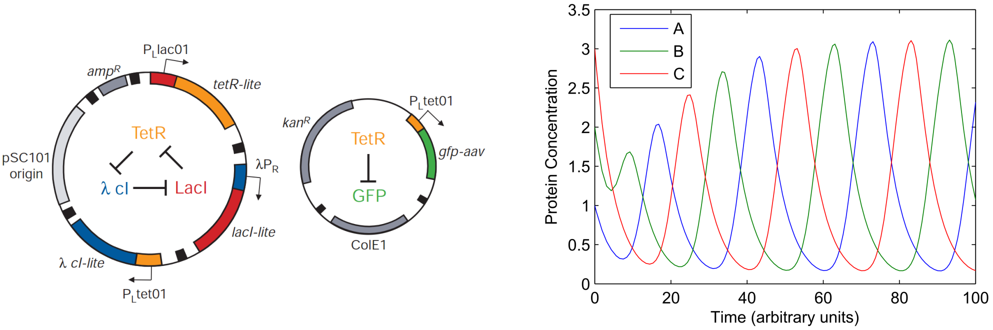
**Figure 4.4:** Illustration of the model of the srRNA-mediated double feedback loop from (Zhou et al. 2012). A TF-coding gene $g_a$'s product interacts with the srRNA-coding gene $g_s$, which represses $g_a$'s mRNA. The TF-coding gene can either repress or activate the srRNA gene, depending on the sign of $\rho_b - \rho_f$. Reproduced with permission from (Zhou et al. 2012).

When the TF in the circuit behaves as a transcriptional repressor, this circuit exhibits bistability, and can produce dynamics very similar to the Toggle Switch (Zhou et al. 2012). Interestingly, in a deterministic model, this network does not require cooperative interactions between repressors to exhibit bistability, due to the nonlinearity in the srRNA regulation function (seen in Figure 2.3b). Lastly, if the TF acts as a transcriptional activator rather than repressor, this circuit exhibits oscillatory dynamics.

In **Publication II**, the stochastic dynamics of this circuit was investigated, using realistic copy numbers for all molecules.

## 4.3 Repressilator

Circuits such as the Toggle Switch are capable storing one bit of memory - more if the circuit has more stable states. Logic and information processing can be performed by repressive and activatory interactions such as those presented in sections 2.2 and 3.2.2. One final control component necessary to produce an information processing machine is a clock (Hasty et al. 2002). The Repressilator is a synthetic genetic circuit with oscillatory dynamics (Elowitz and Leibler 2000), which can therefore function as such a clock.



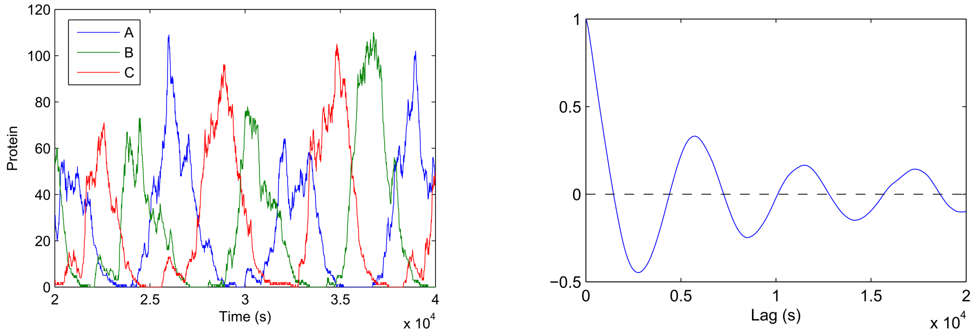
**(a)** Structure of the Repressilator. Three repressors are arranged in a loop such that each represses the next in turn (left). A reporter plasmid (right) was used in (Elowitz and Leibler 2000) to observe the dynamics of the circuit. Reproduced with permission from (Elowitz and Leibler 2000).

**(b)** Timeseries of a deterministic model of the Repressilator for parameters which produce sustained limit-cycle oscillations.

**Figure 4.5:** Structure and Deterministic Dynamics of the Repressilator

The term “Repressilator” is a combination of “Repression” and “Oscillator”, due to its structure: three genes repressing each other in a ring, as shown in Figure 4.5a. The Repressilator from (Elowitz and Leibler 2000) used the LacI, CI, and TetR transcription factors. When one of these genes is expressed, it represses the next gene in the ring. Since this next gene is repressed, it cannot prevent the gene responsible for repressing the first gene from expressing. In this way, the three genes will be repeatedly expressed in sequence. The RRE model of the Repressilator, built in a similar manner to the Toggle Switch’s RRE in equations

(4.1) and (4.2), has a single fixed point attractor which becomes unstable in suitable parameter ranges, leading to the sustained limit cycle oscillations described above. This behaviour is shown in Figure 4.5b.



**(a)** Example timeseries of a stochastic model of the Repressilator. Oscillations are visible, but the time between rises of each protein is no longer constant.

**(b)** Autocorrelation function for the timeseries in Figure 4.6a. By examining the distance between the zeros of this function, we can determine that the period of this Repressilator is $\sim 5800$ s.

**Figure 4.6:** Stochastic dynamics of the Repressilator

When stochasticity in gene expression is included in the model, both the amplitude of each rise of a TF, as well as the delay between rises vary from one oscillation to the next. The net result is a decrease in the precision with which the Repressilator can keep track of time. A timeseries of a model of the Repressilator, built by extending the model of the Toggle Switch from the preceding section (with one additional alteration, mentioned below), is shown in Figure 4.6a.

Ideally, to quantify the period of the Repressilator in the stochastic setting, we would use the Power Spectral Density (PSD) (e.g. as in (Garcia-Ojalvo et al. 2004)) - the Fourier transform of its autocorrelation function. Due to difficulty in measuring the PSD for real timeseries, the distance between the zeros of its autocorrelation function is often used instead (Chandraseelan et al. 2013). This function is shown in Figure 4.6b for the timeseries in Figure 4.6a. By far, the most important parameter governing the period of oscillation is the protein decay rate (Loinger and Biham 2007). For this reason, in the Repressilator synthesized in (Elowitz and Leibler 2000), the TF decay rates were accelerated, and made more uniform, by attaching a tag to each of the TFs which is recognized by the proteases in the cell.

One additional feature of the model has a significant impact on the Repressilator's dynamics: the possibility of a TF to degrade while bound to its target promoter. If it cannot, as in the model of the Toggle Switch presented above, then the bound TF is 'protected' from degradation, and will likely take considerably longer to finally disappear from the system. If this single molecule event takes a non-negligible amount of time, it will delay the rise of the next gene in the Repressilator by a long, exponentially-distributed amount of time, ultimately destroying the periodicity of the dynamics (Loinger and Biham 2007). For this reason, the model which produced the timeseries in Figure 4.5b included the following reaction representing the degradation of a bound repressor for each TF:

$$\mathrm{Pro}_i \cdot \mathrm{P}_j \xrightarrow{k_{proteind}} \mathrm{Pro}_i \tag{4.10}$$

An example of a fluorescence timeseries from cells containing the Repressilator as depicted in Figure 4.5a is shown in Figure 4.7. First, note that an upward trend is visible in the timeseries. This is likely simply due to the accumulation of the reporter protein, which has a longer half-life than the TFs which compose the circuit (Elowitz and Leibler 2000). Despite this trend, oscillations are clearly observed in the fluorescence, indicating that the underlying circuit is functioning.

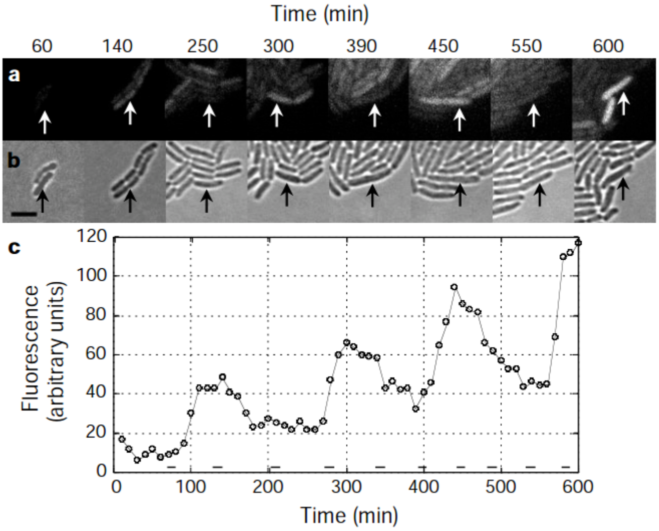


**Figure 4.7:** Cells containing the constructs depicted in Figure 4.5a. Top: Fluorescence and phase contrast images. Bottom: Fluorescence timeseries of one cell (marked with an arrow in the upper images). Reproduced with permission from (Elowitz and Leibler 2000).

In this thesis, the Repressilator is one of the two networks studied in **Publication IV**, where the effects of partitioning of its regulatory molecules in division were explored.

# 5 Conclusions and Discussion

This thesis has focused on two mechanisms that can qualitatively change the dynamics of genetic networks: the stochastic partitioning of regulatory molecules during cell division, and the direct interaction between low copy-number regulatory molecules. The four publications work towards this by first presenting the tool to be used in these studies (**Publication I**), followed by a study of a bistable circuit with a link composed of the direct interaction of two low-copy molecules (**Publication II**), a study of the expression of a single gene with stochastic partitioning of its mRNA molecules in cell division (**Publication III**), and a study of a Toggle Switch and a Repressilator subject to this partitioning (**Publication IV**).

The new simulator presented in **Publication I**, SGNS2, is based on the NRM of the SSA (see Section 3.3.2), which has been augmented to efficiently simulate stochastic reaction systems within dynamic, hierarchically-linked compartments (see Section 3.5). In this thesis, the primary use of these compartments is to properly simulate the dynamics of stochastic genetic circuits within growing cell populations, however three additional use cases were also considered during its development:

**Single-nucleotide transcription and translation:** The ability of molecules within compartments at a lower level of the compartment hierarchy to interact with molecules at a higher level was used to simulate a model of coupled transcription and translation at the single nucleotide level. A preliminary version of the simulator was used for this purpose in (Mäkelä et al. 2011; Potapov et al. 2011; Martins et al. 2012).

**Infection by $\lambda$ phage:** A compartmentalized model of the infection of *E. coli* cells by the $\lambda$ phage, was presented in **Publication I**. In this, the individual phages infecting a cell each existed in their own compartments within each cell's compartment, allowing them to make their own lysis/lysogeny decisions, as was shown in (Zeng et al. 2010).

**Asymmetric disposal of protein aggregates:** *E. coli* have been shown to accumulate protein aggregates in the older pole of the cell (Lindner et al. 2008). SGNS2's ability to partition molecules in division based on different partitioning schemes has thus been used to model and study this process (Gupta et al. 2014c).

SGNS2 was built using the NRM since it allows the reactions for each compartment to be grouped into their own Indexed Priority Queues, simplifying their creation and destruction. Further, it forms the basis of a general discrete event simulation, allowing both the delayed products from reactions and the reactions themselves to coexist within the same framework. Nevertheless, in the future, it would be advantageous to implement the Markovian subset of the reaction system (i.e. the part that does not include delays) using the Composition-Rejection method (Slepoy et al. 2008), which scales better to larger system sizes.

Though SGNS2 was designed to simulate compartmentalized systems, it does not explicitly model the spatial relationships between the compartments. This is simultaneously a drawback and a benefit − a drawback since effects deriving from the exact spatial relationship between compartments cannot be studied, such as quorum sensing in a population of bacteria (Waters and Bassler 2005). However, it can be advantageous since it frees the modeller from having to explicitly give these often extraneous details for all models. Models in SGNS2 are therefore simpler to set up and reason about. This design choice makes the simulator a valuable contribution to a field filled with explicitly spatial simulations (Loew and Schaff 2001; Hattne et al. 2005; Andrews et al. 2010) and more fixed simulators of unchanging chemical interactions (Sanft et al. 2011; Ramsey et al. 2005; Hoops et al. 2006; Ribeiro and Lloyd-Price 2007). The source code of the simulator has been released under an open source license, the New BSD License, such that other researchers can improve on it, and/or modify it for their own use.

In **Publication II**, the stochastic dynamics of an srRNA-mediated Toggle Switch (presented in section 4.2.1) was investigated. First, it was found that in order to achieve long-term bistability, a switch in this configuration requires the repressive interactions to be rather strong, to compensate for the sensitivity of the circuit to noise. Nevertheless, these were well within the realistic range of interaction strengths. Additional features such as cooperative binding, which were not considered, make this repression strength easily achievable in real cells. Second, for realistic copy-numbers, a deterministic representation of the system using RREs was found to greatly overestimate the region in parameter space where long-term bistability is achieved.

Next, the initiation dynamics at the promoter was found to have a strong influence on the dynamics of the switch, as would be expected given the srRNA link's susceptibility to any extra noise in this dynamics. More noisy than Poissonian (super-Poissonian) dynamics disrupts the srRNA's ability to silence the target. Nevertheless, low-noise production does not negate the need for strong repression strengths. Finally, it was shown that the use of the srRNA in one of the repressive interactions allows the network to rapidly switch from one of the two states to the other in response to an environmental signal, but the same is not true in the other direction. This property exists independent of the normal stability of the two noisy attractors of the network, i.e. both states can be equally resistant to stochastic switching.

This property could be used by organisms to make a switch which is highly sensitive to a specific external input, but requires prolonged exposure to the opposite environmental signal to switch back to the first state. This property may be used, for example, in several bacterial species to regulate iron storage genes since the major iron storage regulator *fur* is arranged in this network motif (Zhou et al. 2012). In this case, the srRNA RyhB represses the *fur* gene, whose protein represses RyhB in the presence of $Fe^{2+}$, and which regulates the downstream iron storage genes. The model in **Publication II** therefore predicts that iron storage genes will activate rather quickly when the bacteria are placed into an iron-rich environment, while they will take longer to deactivate after transitioning to an iron-deficient environment.

In **Publication III**, the effects of stochastic partitioning of RNA molecules in cell division were examined using a delayed stochastic model of gene expression coupled with SGNS2's ability to randomly segregate molecules when creating new compartments. In a synchronously dividing population of cells, stochastic partitioning was found to cause transient increases in the phenotypic diversity of the population. The length of this transient is dependent on the degradation rate of the RNA, and for long enough RNA lifetimes (or short enough cell division times), this diversity can accumulate over generations. Meanwhile, in asynchronously dividing populations, partitioning errors manifest themselves as a simple increase in the phenotypic diversity at all time points. Finally, the amplitude of the transient increase can be controlled if the RNA is partitioned in a biased manner, i.e. one of the two daughter cells is more likely to inherit more RNA. The dynamic range in normalized variance which is realistically achievable by the combination of the above mechanisms was found to be on the order of $\sim$16 fold, of which the contribution from cell synchrony was $\sim$3 fold.

The predictions of this model were then compared to measurements with single-molecule precision in live *E. coli* cells. In a population of cells synchronized by heat shock, the distribution of the number of mRNA tagged with MS2-GFP was measured (see section 2.3), before and after the expected division point. After division, a significant increase in the normalized variance across the population was found. Further, when measuring this value over time in a synchronized population of cells, transient increases were observed where expected given the division rate of the cells, which were not observed in a population which had not been synchronized. Evidence for a bias in the partitioning of the RNA molecules was discovered, which exacerbated the size of the observed transient increases.

Though in this case, the bias is likely an artifact of the immortalization of the mRNA molecules by a large mass of MS2-GFP molecules (Lindner et al. 2008; Montero Llopis et al. 2010), it is still possible that untagged mRNA is partitioned asymmetrically if, for example, it diffuses slowly when being translated by a large amount of ribosomes and thus they tend to remain near their site of transcription (Montero Llopis et al. 2010).

Cell synchrony can be induced by a number of different conditions, mainly related to stress such as starvation or heat shock (Cutler and Evans 1966). Curiously, it is in these periods of stress where population diversity is most advantageous (Kirschner and Gerhart 1998). This suggests that cell synchrony, and a bias in partitioning of RNA molecules may be a form of reproductive bet hedging, used by bacterial populations to increase the amount of phenotypic diversity in times of hardship.

Finally, in **Publication IV**, the different behaviours that two genetic networks exhibited when subject to the random partitioning of their molecules in cell division were studied. Several interesting results were found for the Toggle Switch. First, though the stability of the switch decreased with increasing partitioning errors, anti-correlations between sister cells, an inevitable by-product of the partitioning errors, increase the chances that two daughter cells will end up in different noisy attractors. The result is that a particular balance between the two states in the population is more reliably achieved. For the Repressilator, increasing the partitioning errors was found to decrease the robustness of the period of oscillation. However, the rate of desynchronization of a population of cells was remarkably slow, only significantly accelerating for the strongest errors in partitioning.

These results show that the effects of partitioning of low copy-number molecules in division are not trivial to predict at the population level. The differences observed between the effects that it had in the two genetic circuits show how the interplay between the topology of the state space of a network and high-variance partitioning can result in qualitatively different behaviour. In the switch, a network with two noisy attractors, the increased variability resulted in a counter-intuitive *decrease* of the variance in the phenotype distribution. Meanwhile the clock, a network with only one noisy attractor, did not show any new features – merely an acceleration of its desynchronization over generations. This means that any new effects that are considered, such as the inclusion of more spatial information, must be considered from the point of view of many different network types, since the details of the added mechanism may result in qualitatively new features in different networks.

Biological systems have evolved not only the ability to cope with the stochasticity inherent in gene expression, but also the ability to use it to their own advantage. Both of the mechanisms studied in this thesis modify the noise within gene networks. It is therefore likely that cells also utilize these mechanisms to control or enhance diversity, especially if these are easy to regulate and/or evolve.

Even if this is not the case in real cells, it will be advantageous to make use of these mechanisms in synthetic biology, where the design of circuits with desired dynamics is complicated by the difficulty of predicting those dynamics. In the past, this problem has been overcome by constructing large libraries of different parts (e.g. promoters), and measuring the dynamical properties of each to find one suitable for the present application (Ellis et al. 2009). Creating targeted mutations to produce desired behaviour is still difficult since, for example, predicting the dynamics of DNA-protein interactions involved in TF-based regulation is complicated by the myriad of interactions that can occur between the TF and any nearby molecules, including other proteins and other features of the DNA (Kim et al. 2013). On the other hand, it is simpler to predict RNA-RNA interactions from sequence alone (Wright et al. 2014). It may also be possible to control, at least to some degree, the variance in partitioning by relying on e.g. limited diffusion (Montero Llopis et al. 2010), clustering of molecules (e.g. membrane receptors (Sourjik and Berg 2004)), or the positioning of macromolecules within cells (Gupta et al. 2014b). It may therefore be advantageous in synthetic biology to utilize these more predictable and controllable mechanisms to generate the desired behaviours.

It will also be of interest to study whether there are behaviours that can be achieved only with the combination of the studied mechanisms. For example, if srRNA molecules are partitioned in a high-variance manner, what new behaviours can this confer in a circuit? When there is an abundance of either the target mRNA or the srRNA, this partitioning is unlikely to cause large differences, however when both the target mRNA and the srRNA are produced at approximately the same mean rate, critical phenomena result in a significant increase in the level of noise (Elf et al. 2003). In this scenario, stochastic partitioning of the srRNA has the potential to cause drastic phenotypic differences between sister cells. This might thus be a means to construct a circuit which is sensitive to partitioning errors only in a narrow set of circumstances.

The random partitioning of molecules in division poses an additional interesting problem for genetic circuits where noise is not advantageous, since this source of noise is unavoidable. Correcting for both the natural fluctuations in gene expression and errors in partitioning is energetically expensive, involving negative feedback loops (Becskei and Serrano 2000) and complex partitioning schemes (Huh and Paulsson 2011b; Huh and Paulsson 2011a), respectively. However, if one of these sources of noise is not compensated for, then it will render any work spent reducing the other moot. Thus, we expect that noise reduction mechanisms will only be present for molecular species for which cell-to-cell diversity is extremely disadvantageous and, when this is the case, there ought to be multiple mechanisms at play to reduce the variability.

Many small-scale mechanisms can significantly alter the behaviour of genetic circuits, and thus cells, due to their interaction with the molecules in low copy-number that compose them. The studies in this thesis extend our knowledge to include the effects at the gene network level of two such mechanisms. Since small changes in the population counts of these molecules have large changes in the phenotype of the cells, these low-copy molecules are prime targets for other low-energy mechanisms to change the behaviour of cells. We therefore predict that many more such mechanisms will be found in nature, which are utilized by cells to produce specific behaviours, e.g. when interacting with the environment or to optimize a specific cellular function. These molecules and mechanisms will also be of use in future synthetic circuits, where they will be employed to produce entirely new behaviours in cells.

# Bibliography

Acar, M., J. T. Mettetal, and A. van Oudenaarden. “Stochastic switching as a survival strategy in fluctuating environments.” *Nature Genetics* 40.4 (2008), pp. 471–5. DOI: 10.1038/ng.110.

Ackermann, M., S. C. Stearns, and U. Jenal. “Senescence in a bacterium with asymmetric division”. *Science* 300.5627 (2003), p. 1920. DOI: 10.1126/science.1083532.

Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell.* 4th ed. New York: Garland Science, 2002.

Aldana, M., S. Coppersmith, and L. P. Kadanoff. “Boolean Dynamics with Random Couplings”. In: *Perspectives and Problems in Nolinear Science.* Ed. by E. Kaplan, J. E. Marsden, and K. R. Sreenivasan. Springer, 2003, pp. 23–89.

Andrews, S. S., N. J. Addy, R. Brent, and A. P. Arkin. “Detailed simulations of cell biology with Smoldyn 2.1.” *PLoS Computational Biology* 6.3 (2010), e1000705. DOI: 10.1371/journal.pcbi.1000705.

Arkin, A. P., J. Ross, and H. H. McAdams. “Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage $\lambda$-Infected Escherichia coli Cells”. *Genetics* 149 (1998), pp. 1633–48.

Atkinson, M. R., M. A. Savageau, J. T. Myers, and A. J. Ninfa. “Development of Genetic Circuitry Exhibiting Toggle Switch or Oscillatory Behavior in Escherichia coli”. *Cell* 113 (2003), pp. 597–607.

Becskei, A. and L. Serrano. “Engineering stability in gene networks by autoregulation”. *Nature* 405 (2000).

Belle, A., A. Tanay, L. Bitincka, R. Shamir, and E. K. O’Shea. “Quantification of protein half-lives in the budding yeast proteome.” *Proceedings of the National Academy of Sciences of the United States of America* 103.35 (2006), pp. 13004–9. DOI: 10.1073/pnas.0605420103.

Berg, O. G. “A model for the statistical fluctuations of protein numbers in a microbial population”. *Journal of Theoretical Biology* 71.4 (1978), pp. 587–603. DOI: 10.1016/0022-5193(78)90326-0.

Bernstein, J. A., A. B. Khodursky, L. Pei-Hsun, S. Lin-Chao, and S. N. Cohen. “Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays”. *Proceedings of* *the National Academy of Sciences of the United States of America* 99.15 (2002), pp. 9697–9702.

Bertrand-Burggraf, E., J. F. Lefèvre, and M. Daune. “A new experimental approach for studying the association between RNA polymeras and the tet promoter of pBR322”. *Nucleic Acids Research* 12.3 (1984), pp. 1697–1706.

Bratsun, D., D. Volfson, L. S. Tsimring, and J. Hasty. “Delay-induced stochastic oscillations in gene regulation”. *Proceedings of the National Academy of Sciences of the United States of America* 102.41 (2005), pp. 14593–14598.

Cao, Y., D. T. Gillespie, and L. R. Petzold. “The slow-scale stochastic simulation algorithm.” *The Journal of Chemical Physics* 122.1 (2005), p. 14116. DOI: 10.1063/1.1824902.

Chandraseelan, J. G., S. M. D. Oliveira, A. Häkkinen, H. Tran, I. Potapov, A. Sala, M. Kandhavelu, and A. S. Ribeiro. “Effects of temperature on the dynamics of the LacI-TetR-CI repressilator.” *Molecular Biosystems* 9.12 (2013), pp. 3117–23. DOI: 10.1039/c3mb70203k.

Cormack, B. P., R. H. Valdivia, and S. Falkow. “FACS-optimized mutants of the green fluorescent protein (GFP)”. *Gene* 173 (1996), pp. 33–38.

Cutler, R. G. and J. E. Evans. “Synchronization of bacteria by a stationary-phase method”. *Journal of Bacteriology* 91.2 (1966), pp. 469–476.

Dai, X., O. Yli-Harja, and A. S. Ribeiro. “Determining noisy attractors of delayed stochastic gene regulatory networks from multiple data sources”. *Bioinformatics* 25.18 (2009), pp. 2362–2368. DOI: 10.1093/bioinformatics/btp411.

Davenport, R. J., G. J. L. Wuite, R. Landick, and C. Bustamante. “Single-Molecule Study of Transcriptional Pausing and Arrest by E. coli RNA Polymerase”. *Science* 287.5462 (2000), pp. 2497–2500. DOI: 10.1126/science.287.5462.2497.

Dunaway, M., J. S. Olson, J. M. Rosenberg, O. B. Kallai, R. E. Dickerson, and K. S. Matthews. “Kinetic Studies of Inducer Binding to lac Repressor-Operator Complex”. *The Journal of Biological Chemistry* 255.21 (1980), pp. 10115–10119.

Elf, J., J. Paulsson, O. G. Berg, and M. n. Ehrenberg. “Near-critical phenomena in intracellular metabolite pools.” *Biophysical Journal* 84.1 (2003), pp. 154–70. DOI: 10.1016/S0006-3495(03)74839-5.

Ellis, T., X. Wang, and J. J. Collins. “Diversity-based, model-guided construction of synthetic gene networks with predicted functions.” *Nature Biotechnology* 27.5 (2009), pp. 465–471. DOI: 10.1038/nbt.1536.

Elowitz, M. B. and S. Leibler. “A synthetic oscillatory network of transcriptional regulators.” *Nature* 403.6767 (2000), pp. 335–8. DOI: 10.1038/35002125.

Fisher, J. K., A. Bourniquel, G. Witz, B. Weiner, M. Prentiss, and N. Kleckner. “Four-dimensional imaging of E. coli nucleoid organization and dynamics in living cells”. *Cell* 153.4 (2013), pp. 882–95. DOI: 10.1016/j.cell.2013.04.006.

Friedländer, M. R., E. Lizano, A. J. S. Houben, D. Bezdan, M. Báñez Coronel, G. Kudla, E. Mateu-Huertas, B. Kagerbauer, J. González, K. C. Chen, E. M. LeProust, E. Martí, and X. Estivill. “Evidence for the biogenesis of more than 1,000 novel human microRNAs." *Genome Biology* 15.4 (2014), R57. DOI: 10.1186/gb-2014-15-4-r57.

Fusco, D., N. Accornero, B. Lavoie, S. M. Shenoy, J.-M. Blanchard, R. H. Singer, and E. Bertrand. "Single mRNA Molecules Demonstrate Probabilistic Movement in Living Mammalian Cells". *Current Biology* 13.2 (2003), pp. 161–167. DOI: 10.1016/S0960-9822(02)01436-7.

Garcia-Ojalvo, J., M. B. Elowitz, and S. H. Strogatz. "Modeling a synthetic multicellular clock: repressilators coupled by quorum sensing." *Proceedings of the National Academy of Sciences of the United States of America* 101.30 (2004), pp. 10955–10960. DOI: 10.1073/pnas.0307095101.

Gardner, T. S., C. R. Cantor, and J. J. Collins. "Construction of a genetic toggle switch in Escherichia coli." *Nature* 403.6767 (2000), pp. 339–42. DOI: 10.1038/35002131.

Gibson, M. A. and J. Bruck. "Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels". *The Journal of Physical Chemistry A* 104.9 (2000), pp. 1876–1889. DOI: 10.1021/jp993732q.

Gillespie, D. T. "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions". *Journal of Computational Physics* 22 (1976), pp. 403–34.

– "A rigorous derivation of the chemical master equation". *Physica A: Statistical Mechanics and its Applications* 188.1-3 (1992), pp. 404–425. DOI: 10.1016/0378-4371(92)90283-V.

– "Approximate accelerated stochastic simulation of chemically reacting systems". *The Journal of Chemical Physics* 115.4 (2001), pp. 1716–1733.

– "Deterministic limit of stochastic chemical kinetics." *The Journal of Physical Chemistry B* 113.6 (2009), pp. 1640–4. DOI: 10.1021/jp806431b.

– "Exact Stochastic Simulation of Coupled Chemical Reactions". *The Journal of Physical Chemistry* 81.25 (1977), pp. 2340–2361.

– "Stochastic simulation of chemical kinetics." *Annual Review of Physical Chemistry* 58 (2007), pp. 35–55. DOI: 10.1146/annurev.physchem.58.032806.104637.

– "The chemical Langevin equation". *The Journal of Chemical Physics* 113.1 (2000), pp. 297–306.

– "The multivariate Langevin and Fokker–Planck equations". *American Journal of Physics* 64.10 (1996), p. 1246. DOI: 10.1119/1.18387.

Golding, I. and E. C. Cox. "Physical Nature of Bacterial Cytoplasm". *Physical Review Letters* 96.9 (2006), p. 098102. DOI: 10.1103/PhysRevLett.96.098102.

– "RNA dynamics in live Escherichia coli cells." *Proceedings of the National Academy of Sciences of the United States of America* 101.31 (2004), pp. 11310–5. DOI: 10.1073/pnas.0404443101.

Golding, I., J. Paulsson, S. M. Zawilski, and E. C. Cox. "Real-time kinetics of gene activity in individual bacteria." *Cell* 123.6 (2005), pp. 1025–36. DOI: 10.1016/j.cell.2005.09.031.

Gottesman, S. and G. Storz. “Bacterial small RNA regulators: versatile roles and rapidly evolving variations.” *Cold Spring Harbor Perspectives in Biology* 3.12 (2011). DOI: 10.1101/cshperspect.a003798.

Gunawardena, J. “Time-scale separation – Michaelis and Menten’s old idea, still bearing fruit.” *The FEBS Journal* 281.2 (2014), pp. 473–88. DOI: 10.1111/febs.12532.

Gupta, A., J. Lloyd-Price, and A. S. Ribeiro. “In silico analysis of division times of Escherichia coli populations as a function of the partitioning scheme of non-functional proteins”. *In Silico Biology* (2014), in press. DOI: 10.3233/ISB-140462.

Gupta, A., J. Lloyd-Price, R. Neeli-Venkata, S. M. D. Oliveira, and A. S. Ribeiro. “In Vivo Kinetics of Segregation and Polar Retention of MS2-GFP-RNA Complexes in Escherichia coli”. *Biophysical Journal* 106.9 (2014), pp. 1928–1937. DOI: 10.1016/j.bpj.2014.03.035.

Gupta, A., J. Lloyd-Price, and A. S. Ribeiro. “Modelling Polar Retention of Complexes in Escherichia coli”. In: *Computational Methods in Systems Biology*. Ed. by P. Mendes, J. O. Dada, and K. Smallbone. Springer, 2014, pp. 239–243. DOI: 10.1007/978-3-319-12982-2_17.

Gupta, A., J. Lloyd-Price, S. M. D. Oliveira, O. Yli-Harja, A.-B. Muthukrishnan, and A. S. Ribeiro. “Robustness of the division symmetry in Escherichia coli and functional consequences of symmetry breaking.” *Physical Biology* 11.6 (2014), p. 066005. DOI: 10.1088/1478-3975/11/6/066005.

Hasty, J., D. McMillen, and J. J. Collins. “Engineered gene circuits.” *Nature* 420.November (2002), pp. 224–230. DOI: 10.1038/nature01257.

Hattne, J., D. Fange, and J. Elf. “Stochastic reaction-diffusion simulation with MesoRD.” *Bioinformatics* 21.12 (2005), pp. 2923–4. DOI: 10.1093/bioinformatics/bti431.

Herbert, K. M., A. La Porta, B. J. Wong, R. A. Mooney, K. C. Neuman, R. Landick, and S. M. Block. “Sequence-resolved detection of pausing by single RNA polymerase molecules.” *Cell* 125.6 (2006), pp. 1083–94. DOI: 10.1016/j.cell.2006.04.032.

Hoffman, H. and M. E. Frank. “Synchrony of Division in Clonal Microcolonies of Escherichia coli”. *Journal of Bacteriology* 89.2 (1965), pp. 513–517.

Hoops, S., S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer. “COPASI – a COmplex PAthway SImulator.” *Bioinformatics* 22.24 (2006), pp. 3067–74. DOI: 10.1093/bioinformatics/btl485.

Hsu, L. M. “Monitoring abortive initiation.” *Methods* 47.1 (2009), pp. 25–36. DOI: 10.1016/j.ymeth.2008.10.010.

Huh, D. and J. Paulsson. “Non-genetic heterogeneity from stochastic partitioning at cell division.” *Nature Genetics* 43.2 (2011), pp. 95–100. DOI: 10.1038/ng.729.

– “Random partitioning of molecules at cell division.” *Proceedings of the National Academy of Sciences of the United States of America* 108.36 (2011), pp. 15004–9. DOI: 10.1073/pnas.1013171108.

Jahnke, T. and W. Huisinga. “Solving the chemical master equation for monomolecular reaction systems analytically.” *Journal of Mathematical Biology* 54.1 (2007), pp. 1–26. DOI: 10.1007/s00285-006-0034-x.

Kaern, M., T. C. Elston, W. J. Blake, and J. J. Collins. “Stochasticity in gene expression: from theories to phenotypes.” *Nature Reviews Genetics* 6.6 (2005), pp. 451–64. DOI: 10.1038/nrg1615.

Kandhavelu, M., H. Mannerström, A. Gupta, A. Häkkinen, J. Lloyd-Price, O. Yli-Harja, and A. S. Ribeiro. “In vivo kinetics of transcription initiation of the lar promoter in Escherichia coli. Evidence for a sequential mechanism with two rate-limiting steps.” *BMC Systems Biology* 5.1 (2011), p. 149. DOI: 10.1186/1752-0509-5-149.

Kandhavelu, M., J. Lloyd-Price, A. Gupta, A.-B. Muthukrishnan, O. Yli-Harja, and A. S. Ribeiro. “Regulation of mean and noise of the in vivo kinetics of transcription under the control of the lac/ara-1 promoter.” *FEBS Letters* 586.21 (2012), pp. 3870–5. DOI: 10.1016/j.febslet.2012.09.014.

Kandhavelu, M., A. Häkkinen, O. Yli-Harja, and A. S. Ribeiro. “Single-molecule dynamics of transcription of the lar promoter.” *Physical Biology* 9.2 (2012), p. 026004. DOI: 10.1088/1478-3975/9/2/026004.

Karr, J. R., J. C. Sanghvi, D. N. MacKlin, M. V. Gutschow, J. M. Jacobs, B. Bolival, N. Assad-Garcia, J. I. Glass, and M. W. Covert. “A whole-cell computational model predicts phenotype from genotype”. *Cell* 150.2 (2012), pp. 389–401. DOI: 10.1016/j.cell.2012.05.044.

Kauffman, S. A. “Metabolic Stability and Epigenesis in Randomly Constructed Genetic Nets”. *Journal of Theoretical Biology* 22 (1969), pp. 437–467.

Kim, S., E. Broströmer, D. Xing, J. Jin, S. Chong, H. Ge, S. Wang, C. Gu, L. Yang, Y. Q. Gao, X.-d. Su, Y. Sun, and X. S. Xie. “Probing allostery through DNA.” *Science* 339 (2013), pp. 816–9. DOI: 10.1126/science.1229223.

Kirschner, M. and J. Gerhart. “Evolvability”. *Proceedings of the National Academy of Sciences of the United States of America* 95 (1998), pp. 8420–8427.

Kussell, E. and S. Leibler. “Phenotypic diversity, population growth, and information in fluctuating environments.” *Science* 309.5743 (2005), pp. 2075–8. DOI: 10.1126/science.1114383.

Kussell, E., R. Kishony, N. Q. Balaban, and S. Leibler. “Bacterial persistence: a model of survival in changing environments.” *Genetics* 169.4 (2005), pp. 1807–14. DOI: 10.1534/genetics.104.035352.

Lampoudi, S., D. T. Gillespie, and L. R. Petzold. “The multinomial simulation algorithm for discrete stochastic simulation of reaction-diffusion systems.” *The Journal of Chemical Physics* 130.9 (2009), p. 094104. DOI: 10.1063/1.3074302.

Levine, E. and T. Hwa. “Small RNAs establish gene expression thresholds.” *Current Opinion in Microbiology* 11.6 (2008), pp. 574–9. DOI: 10.1016/j.mib.2008.09.016.

Levine, E., M. Huang, Y. Huang, and T. Kuhlman. “On noise and silence in small RNA regulation”. *Proceedings of the National Academy of Sciences of the United States of America* (2009).

Levine, E., P. McHale, and H. Levine. “Small regulatory RNAs may sharpen spatial expression patterns.” *PLoS Computational Biology* 3.11 (2007), e233. DOI: 10.1371/journal.pcbi.0030233.

Lindner, A. B., R. Madden, A. Demarez, and E. J. Stewart. “Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation”. *Proceedings of the National Academy of Sciences of the United States of America* 105.8 (2008), pp. 3076–3081.

Lipshtat, A., A. Loinger, N. Q. Balaban, and O. Biham. “Genetic Toggle Switch without Cooperative Binding”. *Physical Review Letters* 96.18 (2006), p. 188101. DOI: 10.1103/PhysRevLett.96.188101.

Loew, L. M. and J. C. Schaff. “The Virtual Cell: a software environment for computational cell biology.” *Trends in Biotechnology* 19.10 (2001), pp. 401–6. DOI: 10.1016/S0167-7799(01)01740-1.

Loinger, A. and O. Biham. “Stochastic simulations of the repressilator circuit”. *Physical Review E* 76.5 (2007), p. 051917. DOI: 10.1103/PhysRevE.76.051917.

Loinger, A., A. Lipshtat, N. Q. Balaban, and O. Biham. “Stochastic simulations of genetic switch systems”. *Physical Review E* 75.2 (2007), p. 021904. DOI: 10.1103/PhysRevE.75.021904.

Lutz, R., T. Lozinski, T. Ellinger, and H. Bujard. “Dissecting the functional program of Escherichia coli promoters: the combined mode of action of Lac repressor and AraC activator”. *Nucleic Acids Research* 29.18 (2001), pp. 3873–81.

Mäkelä, J., J. Lloyd-Price, O. Yli-Harja, and A. S. Ribeiro. “Stochastic sequence-level model of coupled transcription and translation in prokaryotes.” *BMC Bioinformatics* 12.1 (2011), p. 121. DOI: 10.1186/1471-2105-12-121.

Männik, J., F. Wu, F. J. H. Hol, P. Bisicchia, D. J. Sherratt, J. E. Keymer, and C. Dekker. “Robustness and accuracy of cell division in Escherichia coli in diverse cell shapes.” *Proceedings of the National Academy of Sciences of the United States of America* 109.18 (2012), pp. 6957–62. DOI: 10.1073/pnas.1120854109.

Margolin, W. “Bacterial Division: Another Way to Box in the Ring”. *Current biology* 16.20 (2006), R881–4. DOI: 10.1016/j.cub.2006.09.025.

Martins, L., J. Mäkelä, A. Häkkinen, M. Kandhavelu, O. Yli-Harja, J. M. Fonseca, and A. S. Ribeiro. “Dynamics of transcription of closely spaced promoters in Escherichia coli, one event at a time.” *Journal of Theoretical Biology* 301 (2012), pp. 83–94. DOI: 10.1016/j.jtbi.2012.02.015.

Massé, E. and S. Gottesman. “A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli”. *Proceedings of the National Academy of Sciences of the United States of America* 99.7 (2002).

Massé, E., N. Majdalani, and S. Gottesman. “Regulatory roles for small RNAs in bacteria”. *Current Opinion in Microbiology* 6 (2003), pp. 120–124. DOI: 10.1016/S1369-5274(03)00027-4.

McClure, W. R. “Mechanism and control of transcription initiation in prokaryotes.” *Annual Review of Biochemistry* 54 (1985), pp. 171–204. DOI: 10.1146/annurev.bi.54.070185.001131.

– “Rate-limiting steps in RNA chain initiation.” *Proceedings of the National Academy of Sciences of the United States of America* 77.10 (1980), pp. 5634–8.

Mitarai, N., K. Sneppen, and S. Pedersen. “Ribosome collisions and translation efficiency: optimization by codon usage and mRNA destabilization.” *Journal of Molecular Biology* 382.1 (2008), pp. 236–45. DOI: 10.1016/j.jmb.2008.06.068.

Montero Llopis, P., A. F. Jackson, O. Sliusarenko, I. Surovtsev, J. Heinritz, T. Emonet, and C. Jacobs-Wagner. “Spatial organization of the flow of genetic information in bacteria.” *Nature* 466.7302 (2010), pp. 77–81. DOI: 10.1038/nature09152.

Muthukrishnan, A.-B., M. Kandhavelu, J. Lloyd-Price, F. Kudasov, S. Chowdhury, O. Yli-Harja, and A. S. Ribeiro. “Dynamics of transcription driven by the tetA promoter, one event at a time, in live Escherichia coli cells.” *Nucleic Acids Research* 40.17 (2012), pp. 8472–83. DOI: 10.1093/nar/gks583.

Neubauer, Z and E Calef. “Immunity Phase-shift in Defective Lysogens: Non-mutational Hereditary Change of Early Regulation of $\lambda$ Prophage”. *Journal of Molecular Biology* 51 (1970), pp. 1–13.

Osella, M., E. Nugent, and M. C. Lagomarsino. “Concerted control of Escherichia coli cell division”. *Proceedings of the National Academy of Sciences of the United States of America* 111.9 (2014), pp. 3431–5. DOI: 10.1073/pnas.1313715111.

Paulsson, J. “Summing up the noise in gene networks.” *Nature* 427.6973 (2004), pp. 415–8. DOI: 10.1038/nature02257.

Păun, G. “From cells to computers: computing with membranes (P systems)”. *BioSystems* 59 (2001), pp. 139–158.

Pedraza, J. M. and J. Paulsson. “Effects of molecular memory and bursting on fluctuations in gene expression.” *Science* 319.5861 (2008), pp. 339–43. DOI: 10.1126/science.1144331.

Potapov, I., J. Lloyd-Price, O. Yli-Harja, and A. S. Ribeiro. “Dynamics of a genetic toggle switch at the nucleotide and codon levels”. *Physical Review E* 84.3 (2011), p. 031903. DOI: 10.1103/PhysRevE.84.031903.

Rajala, T., A. Häkkinen, S. Healy, O. Yli-Harja, and A. S. Ribeiro. “Effects of transcriptional pausing on gene expression dynamics.” *PLoS Computational Biology* 6.3 (2010), e1000704. DOI: 10.1371/journal.pcbi.1000704.

Ramsey, S., D. Orrell, and H. Bolouri. “Dizzy: Stochastic Simulation of Large-Scale Genetic Regulatory Networks”. *Journal of Bioinformatics and Computational Biology* 3.2 (2005), pp. 415–436.

Reyes-Lamothe, R., T. Tran, D. Meas, L. Lee, A. M. Li, D. J. Sherratt, and M. E. Tolmasky. “High-copy bacterial plasmids diffuse in the nucleoid-free space, replicate stochastically and are randomly partitioned at cell division.” *Nucleic Acids Research* 42.2 (2013), pp. 1042–1051. DOI: 10.1093/nar/gkt918.

Ribeiro, A. S. “A Model of Genetic Networks with Delayed Stochastic Dynamics”. In: *Analysis of Microarray Data: A Network-Based Approach.* 2007, pp. 1–30. DOI: 10.1002/9783527622818.ch7.

– “Dynamics and evolution of stochastic bistable gene networks with sensing in fluctuating environments”. *Physical Review E* 78.6 (2008), p. 061902. DOI: 10.1103/PhysRevE.78.061902.

– “Effects of coupling strength and space on the dynamics of coupled toggle switches in stochastic gene networks with multiple-delayed reactions”. *Physical Review E* 75.6 (2007), p. 061903. DOI: 10.1103/PhysRevE.75.061903.

– “Stochastic and delayed stochastic models of gene expression and regulation.” *Mathematical Biosciences* 223.1 (2010), pp. 1–11. DOI: 10.1016/j.mbs.2009.10.007.

Ribeiro, A. S. and S. A. Kauffman. “Noisy attractors and ergodic sets in models of gene regulatory networks”. *Journal of Theoretical Biology* 247 (2007), pp. 743–755.

Ribeiro, A. S. and J. Lloyd-Price. “SGN Sim, a Stochastic Genetic Networks Simulator”. *Bioinformatics* 23.6 (2007), pp. 777–779.

Ribeiro, A. S., R. Zhu, and S. A. Kauffman. “A General Modeling Strategy for Gene Regulatory Networks with Stochastic Dynamics”. *Journal of Computational Biology* 13.9 (2006), pp. 1630–1639.

Ribeiro, A. S., A. Häkkinen, H. Mannerström, J. Lloyd-Price, and O. Yli-Harja. “Effects of the promoter open complex formation on gene expression dynamics”. *Physical Review E* 81.1 (2010), p. 011912. DOI: 10.1103/PhysRevE.81.011912.

Rigney, D. R. “Stochastic model of constitutive protein levels in growing and dividing bacterial cells”. *Journal of Theoretical Biology* 76.4 (1979), pp. 453–480. DOI: 10.1016/0022-5193(79)90013-4.

Roussel, M. R. and R. Zhu. “Validation of an algorithm for delay stochastic simulation of transcription and translation in prokaryotic gene expression”. *Physical Biology* 3.4 (2006), pp. 274–84. DOI: 10.1088/1478-3975/3/4/005.

Ruusuvuori, P., T. Aijö, S. Chowdhury, C. Garmendia-Torres, J. Selinummi, M. Birbaumer, A. M. Dudley, L. Pelkmans, and O. Yli-Harja. “Evaluation of methods for detection of fluorescence labeled subcellular objects in microscope images”. *BMC Bioinformatics* 11 (2010), p. 248. DOI: 10.1186/1471-2105-11-248.

Saecker, R. M., M. T. Record, and P. L. Dehaseth. “Mechanism of bacterial transcription initiation: RNA polymerase - Promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis”. *Journal of Molecular Biology* 412 (2011), pp. 754–771. DOI: 10.1016/j.jmb.2011.01.018.

Sanft, K. R., S. Wu, M. Roh, J. Fu, R. K. Lim, and L. R. Petzold. “StochKit2: software for discrete stochastic simulation of biochemical systems with events.” *Bioinformatics* 27.17 (2011), pp. 2457–8. DOI: 10.1093/bioinformatics/btr401.

Schlax, P. J., M. W. Capp, and M. T. Record. “Inhibition of Transcription Initiation by lac Repressor”. *Journal of Molecular Biology* 245 (1995), pp. 331–350.

Schleif, R. “Regulation of the L-arabinose operon of Escherichia coli”. *Trends in Genetics* 16.12 (2000), pp. 559–565.

Schumacher, M. A., K. M. Piro, and W. Xu. “Insight into F plasmid DNA segregation revealed by structures of SopB and SopB-DNA complexes.” *Nucleic Acids Research* 38.13 (2010), pp. 4514–26. DOI: `10.1093/nar/gkq161`.

Scott, M., C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa. “Interdependence of cell growth and gene expression: origins and consequences.” *Science* 330.6007 (2010), pp. 1099–102. DOI: `10.1126/science.1192588`.

Slepoy, A., A. P. Thompson, and S. J. Plimpton. “A constant-time kinetic Monte Carlo algorithm for simulation of large biochemical reaction networks.” *The Journal of Chemical Physics* 128.20 (2008), p. 205101. DOI: `10.1063/1.2919546`.

Smith, H. S. and A. B. Pardee. “Accumulation of a protein required for division during the cell cycle of Escherichia coli”. *Journal of Bacteriology* 101.3 (1970), pp. 901–909.

Sourjik, V. and H. C. Berg. “Functional interactions between receptors in bacterial chemotaxis”. *Nature* 428.March (2004), pp. 1–4. DOI: `10.1038/nature02371.1`.

Spicher, A., O. Michel, M. Cieslak, J.-L. Giavitto, and P. Prusinkiewicz. “Stochastic P systems and the simulation of biochemical processes with dynamic compartments.” *BioSystems* 91.3 (2008), pp. 458–72. DOI: `10.1016/j.biosystems.2006.12.009`.

Süel, G. M., J. Garcia-Ojalvo, L. M. Liberman, and M. B. Elowitz. “An excitable gene regulatory circuit induces transient cellular differentiation.” *Nature* 440.7083 (2006), pp. 545–50. DOI: `10.1038/nature04588`.

Vaquerizas, J. M., S. K. Kummerfeld, S. A. Teichmann, and N. M. Luscombe. “A census of human transcription factors: function, expression and evolution”. *Nature Reviews Genetics* 10 (2009), pp. 252–263. DOI: `10.1038/nrg2538`.

Waddington, C. H. *The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser.* London: George Allen & Unwin, Ltd., 1957.

Waters, C. M. and B. L. Bassler. “Quorum sensing: cell-to-cell communication in bacteria.” *Annual Review of Cell and Developmental Biology* 21 (2005), pp. 319–46. DOI: `10.1146/annurev.cellbio.21.012704.131001`.

Weiss, D. S. “Bacterial cell division and the septal ring.” *Molecular Microbiology* 54.3 (2004), pp. 588–97. DOI: `10.1111/j.1365-2958.2004.04283.x`.

Wernet, M. F., E. O. Mazzoni, A. Celik, D. M. Duncan, I. Duncan, and C. Desplan. “Stochastic spineless expression creates the retinal mosaic for colour vision.” *Nature* 440.7081 (2006), pp. 174–80. DOI: `10.1038/nature04615`.

Wolf, D. M. and A. P. Arkin. “Motifs, modules and games in bacteria”. *Current Opinion in Microbiology* 6.2 (2003), pp. 125–134. DOI: `10.1016/S1369-5274(03)00033-X`.

Wright, P. R., J. Georg, M. Mann, D. a. Sorescu, A. S. Richter, S. Lott, R. Kleinkauf, W. R. Hess, and R. Backofen. “CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains.” *Nucleic Acids Research* 42 (2014), W119–23. DOI: 10.1093/nar/gku359.

Zeng, L., S. O. Skinner, C. Zong, J. Sippy, M. Feiss, and I. Golding. “Decision making at a subcellular level determines the outcome of bacteriophage infection.” *Cell* 141.4 (2010), pp. 682–91. DOI: 10.1016/j.cell.2010.03.034.

Zhou, P., S. Cai, Z. Liu, and R. Wang. “Mechanisms generating bistability and oscillations in microRNA-mediated motifs”. *Physical Review E* 85.4 (2012), p. 041916. DOI: 10.1103/PhysRevE.85.041916.

Zhu, R., A. S. Ribeiro, D. Salahub, and S. A. Kauffman. “Studying genetic regulatory networks at the molecular level: delayed reaction stochastic models.” *Journal of Theoretical Biology* 246.4 (2007), pp. 725–45. DOI: 10.1016/j.jtbi.2007.01.021.

# Publications

# Publication I

Jason Lloyd-Price, Abhishekh Gupta, and Andre S. Ribeiro "SGNS2: A compartmentalized Stochastic Chemical Kinetics Simulator for Dynamic Cell Populations", *Bioinformatics*

BIOINFORMATICS APPLICATIONS NOTE Vol. 28 no. 22 2012, pages 3004–3005
doi:10.1093/bioinformatics/bts556

Systems biology Advance Access publication September 26, 2012

# SGNS2: a compartmentalized stochastic chemical kinetics simulator for dynamic cell populations

Jason Lloyd-Price*, Abhishekh Gupta and Andre S. Ribeiro

Department of Signal Processing, Tampere University of Technology, 33101 Tampere, Finland

Associate Editor: Martin Bishop

**ABSTRACT**

**Motivation:** Cell growth and division affect the kinetics of internal cellular processes and the phenotype diversity of cell populations. Since the effects are complex, e.g. different cellular components are partitioned differently in cell division, to account for them in silico, one needs to simulate these processes in great detail.

**Results**: We present SGNS2, a simulator of chemical reaction systems according to the Stochastic Simulation Algorithm with multi-delayed reactions within hierarchical, interlinked compartments which can be created, destroyed and divided at runtime. In division, molecules are randomly segregated into the daughter cells following a specified distribution corresponding to one of several partitioning schemes, applicable on a per-molecule-type basis. We exemplify its use with six models including a stochastic model of the disposal mechanism of unwanted protein aggregates in *Escherichia coli*, a model of phenotypic diversity in populations with different levels of synchrony, a model of a bacteriophage's infection of a cell population and a model of prokaryotic gene expression at the nucleotide and codon levels.

**Availability**: SGNS2, instructions and examples available at www.cs.tut.fi/~lloydpri/sgns2/ (open source under New BSD license).

**Contact**: jason.lloyd-price@tut.fi

**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

Received on July 31, 2012; revised on August 30, 2012; accepted on September 6, 2012

## 1 INTRODUCTION

Recent evidence suggests that even in cellular organisms whose division is morphologically symmetric, there are a number of asymmetries between daughter cells. These arise, among other things, from the stochasticity in the partitioning of components in division (Huh and Paulsson, 2011) and from biased partitioning schemes for some components. For example, in *Escherichia coli*, unwanted protein aggregates follow biased partitioning schemes dependent on the age of the daughter cells' poles (Lindner *et al*., 2008).

These and other recent findings suggest that the phenotypic diversity of cell populations, among other factors, depends on errors and biases in the partitioning of RNA, proteins and other molecules. This is of relevance since most RNAs exist in small numbers (Bernstein *et al*., 2002) and small fluctuations in these numbers can alter the behavior of genetic circuits (Ribeiro and Kauffman, 2007) and trigger visible phenotype changes (Choi *et al*., 2008).

These sources of phenotypic heterogeneity are difficult to distinguish from, e.g. noise in gene expression (Huh and Paulsson, 2011). Although some effects can be assessed analytically (Huh and Paulsson, 2011), others are too complex and must be assessed numerically. A simulator is thus needed that accounts for noise and delays (Kandhavelu *et al*., 2012) in gene expression and for compartmentalization of processes and components.

Presently, simulators of the dynamics of noisy biochemical systems rely on the Stochastic Simulation Algorithm (SSA) (Gillespie, 1977), e.g. (Blakes *et al*., 2011; Hattne *et al*., 2005; Hoops *et al*., 2006; Lok and Brent, 2005). Some support compartmentalization, simulating reaction-diffusion systems in either static (Hattne *et al*., 2005) or dynamically sized compartments (Blakes *et al*., 2011; Versari and Busi, 2008). Others support rule-based creation of reactions at runtime (Lok and Brent, 2005; Spicher *et al*., 2008), and thus can simulate a dynamic cell population. Very few support delays on the release into the system of one or more products of a reaction (Roussel and Zhu, 2006). These delays are essential to accurately model the kinetics of some processes, e.g. transcription, as RNA production is mostly regulated by the duration of events in transcription initiation (Muthukrishnan *et al*., 2012).

Here, we present SGNS2, an extension of SGN Sim (Ribeiro and Lloyd-Price, 2007) that incorporates dynamic compartments and multiple partitioning distributions at cell division, applicable on a per-molecule-type basis.

## 2 METHODS

SGNS2 is an extension of SGNS, the stochastic simulator of SGNSim (Ribeiro and Lloyd-Price, 2007). It contains all the features of SGNS, such as reactions with multi-delayed events. The two key additions in SGNS2 are (i) it supports dynamic, interlinked, hierarchical compartments and (ii) it supports multiple molecule and compartment partitioning schemes, applicable on a per-molecule-type basis. The novel features considerably extend the class of models that can be simulated.

SGNS2 uses a modified version of the Next Reaction Method (NRM) (Gibson and Bruck, 2000). Namely, the NRM was adapted to stochastic P-systems (Spicher *et al*., 2008) by using a hierarchy of indexed priority queues (IPQ, an ordered list of elements that keep track of their position in the list) and further modified to allow multiple delays in reactions. The IPQ data structure, implemented with a binary heap, is described in Gibson and Bruck (2000). We use a separate IPQ for each compartment, which publish a 'tentative next event time' to an overall IPQ which determines the next event time in the entire simulation. We optimize the update step when molecule populations in a parent compartment change by using a hierarchical refinement of the IPQs with appropriate scaling of tentative firing times (see Supplementary Material). Delayed events were implemented by creating wait lists, implemented by binary heap-based priority queues, whose earliest event is published to each compartment's indexed priority queue. The simulation's elementary SSA steps scale logarithmically with

*To whom correspondence should be addressed.

the number of reactions, compartments and delayed events, allowing complex models to be simulated in reasonable time.

To simulate cell division, we introduced a special reaction event, whose timing follows the SSA rules. When executed, instead of subtracting substrates from the system, a random number is generated based on one of the several partitioning distributions available, including some of those listed in Huh and Paulsson (2011). Each of these mimics a specific molecule partitioning process during cell division. SGNS2 allows both biased and unbiased partitioning of molecules and sub-compartments. The results of these events can be instantaneous or be placed on the wait list. Compartment division and molecule partitioning are represented in the following form:

$$\mathrm{split}(p) : \mathrm{Protein@Cell} \xrightarrow{c_\mu} \mathrm{@Cell+} : \mathrm{Protein@Cell}$$

When this reaction occurs, a new cell compartment is created (@Cell in the product list). Proteins in the original cell are partitioned according to a biased binomial partitioning scheme. In this, each protein is independently partitioned into the new cell with probability $p$. Other common partitioning distributions include the independent partitioning of molecules into daughter cells with random (beta-distributed) sizes and the binding of molecules to spindle binding sites which are segregated evenly between daughter cells such as during mitosis. Available distributions are listed in the manual.

SGNS2 is a command line utility, designed to fit into a toolchain, supporting various input and output formats. Input can be specified in two formats: SBML (Hucka *et al.*, 2003) and SGNSim's native format (Ribeiro and Lloyd-Price, 2007). A subset of SBML Core level 3 version 1 is supported, allowing simulation of most SBML models. Output can be in csv, tsv or in binary format. A text editor may be used to write models in SGNSim format. SBML-based graphical interfaces such as CellDesigner (Funahashi *et al.*, 2008) or Cytoscape (Smoot *et al.*, 2011) may be used to manage SBML models. The results of simulations are interpretable by programs like MATLAB, R or Excel. An example of running a model in SGNSim format of a growing cell population is shown in Supplementary Figure S1.

## 3 DISCUSSION

SGNS2 is the first stochastic simulator that includes multi-delayed events, dynamic compartments and molecule partitioning schemes in division. To test its correctness, we simulated models from the Discrete Stochastic Model Test Suite (Evans *et al.*, 2008). All showed the expected behavior (Supplementary Figs S2 and S3).

SGNS2, though making use of existing and slightly modified versions of existing algorithms, can simulate an array of biological processes not previously possible. For example, it is ideal for simulating gene expression at the nucleotide and codon levels (see 'Availability' section) and study features such as how events in transcription elongation affect protein production kinetics (Mäkelä *et al.*, 2011).

SGNS2 is also suited to study partitioning in cell division, which affects aging, among other processes, and is of particular relevance when modeling populations over multiple generations. To exemplify this, we modeled the biased partitioning of protein aggregates in *E. coli*, known to accumulate in cells with older poles, reducing vitality (Lindner *et al.*, 2008). The results in Supplementary Figure S4 agree with measurements (Stewart *et al.*, 2005). We further studied how cell-cycle synchrony affects the population-level statistics of RNA numbers [Supplementary Fig. S5, in agreement with measurements in Lloyd-Price *et al.* (2012)]. As a side note, we expect the partitioning of RNA and proteins to affect the dynamics of genetic circuits, particularly the stability of their noisy attractors (Ribeiro and Kauffman, 2007). To further demonstrate the simulator's utility, we modeled the viral infection of a dynamic bacterial population.

In conclusion, SGNS2 provides novel functionalities to model and simulate cellular processes not previously possible, as seen from the examples. In general, SGNS2 enables the modeling of stochastic processes in live cells that require compartmentalization, multi-delayed complex processes and complex stochastic partitioning schemes at a per-molecule type in cell division. These features are necessary to study *in silico*, among other phenomena, phenotypic diversity in cell populations.

*Funding*: Work supported by Academy of Finland (126803).

*Conflict of Interest*: none declared.

## REFERENCES

Bernstein,J.A. *et al.* (2002) Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl Acad. Sci. USA*, **99**, 9697–9702.

Blakes,J. *et al.* (2011) The infobiotics workbench: an integrated in silico modelling platform for systems and synthetic biology. *Bioinformatics*, **27**, 3323–3324.

Choi,P.J. *et al.* (2008) A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science*, **322**, 442–446.

Evans,T. *et al.* (2008) The SBML discrete stochastic models test suite. *Bioinformatics*, **25**, 285–286.

Funahashi,A. *et al.* (2008) Celldesigner 3.5: a versatile modeling tool for biochemical networks. *Proc. IEEE*, **96**, 1254–1265.

Gibson,M.A. and Bruck,J. (2000) Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A*, **104**, 1876–1889.

Gillespie,D.T. (1977) Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, **81**, 2340–2361.

Hattne,J. *et al.* (2005) Stochastic reaction-diffusion simulation with mesord. *Bioinfomatics*, **21**, 2923–2924.

Hoops,S. *et al.* (2006) Copasi a complex pathway simulator. *Bioinformatics*, **22**, 3067–3074.

Hucka,M. *et al.* (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–531.

Huh,D. and Paulsson,J. (2011) Random partitioning of molecules at cell division. *Proc. Natl Acad. Sci. USA*, **108**, 15004–15009.

Kandhavelu,M. *et al.* (2012) Random partitioning of molecules at cell division. *Phys. Biol.*, **9**, 026004.

Lindner,A.B. *et al.* (2008) Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. *Proc. Natl Acad. Sci. USA*, **105**, 3076–3081.

Lloyd-Price,J. *et al.* (2012) Probabilistic RNA partitioning generates transient increases in the normalized variance of RNA numbers in synchronized populations of Escherichia coli. *Mol. Biosys.*, **8**, 565–571.

Lok,L. and Brent,R. (2005) Automatic generation of cellular reaction networks with moleculizer 1.0. *Nat. Biotech.*, **23**, 131–36.

Mäkelä,J. *et al.* (2011) Stochastic sequence-level model of coupled transcription and translation in prokaryotes. *BMC Bioinf.*, **12**, 121.

Muthukrishnan,A.-B. *et al.* (2012) Dynamics of transcription driven by the tetA promoter, one event at a time, in live Escherichia coli cells. *Nucleic Acids Res.*, **40**, 8472–8483.

Ribeiro,A.S. and Kauffman,S. (2007) Noisy attractors and ergodic sets in models of gene regulatory networks. *J. Theor. Biol.*, **247**, 743–755.

Ribeiro,A.S. and Lloyd-Price,J. (2007) SGN sim, a stochastic genetic networks simulator. *Bioinformatics*, **23**, 777–779.

Roussel,M.R. and Zhu,R. (2006) Validation of an algorithm for delay stochastic simulation of transcription and translation in prokaryotic gene expression. *Phys. Biol.*, **3**, 274–284.

Smoot,M. *et al.* (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, **27**, 431–432.

Spicher,A. *et al.* (2008) Stochastic P systems and the simulation of biochemical processes with dynamic compartments. *Biosystems*, **91**, 458–472.

Stewart,E.J. *et al.* (2005) Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol.*, **3**, e45.

Versari,C. and Busi,N. (2008) Efficient stochastic simulation of biological systems with multiple variable volumes. *Elec. Notes Theor. Comp. Sci.*, **194**, 165–180.

# Supplement to “SGNS2: A Compartmentalized Stochastic Chemical Kinetics Simulator for Dynamic Cell Populations”

Jason Lloyd-Price, Abhishekh Gupta, and Andre S. Ribeiro

## Implementation Details

SGNS2 uses the Next Reaction Method[1] (NRM) to simulate the dynamics according to the Stochastic Simulation Algorithm[2] (SSA). This method is an efficient implementation of the SSA, which begins by randomizing a ‘next firing time’ for each possible reaction in the system and storing these tentative reaction times in an indexed priority queue (IPQ). The reaction with the soonest tentative firing time is then taken from the queue, performed, and its next firing time is re-randomized. Any reaction whose propensity depends on the set of molecule species affected by this reaction then have their tentative firing times transformed to follow the new distribution of firing times prescribed by the Chemical Master Equation. Then, their positions in the priority queue are updated. These reactions are determined by pre-generating the graph depicting which reactions potentially affect the propensities of other reactions (the reaction dependency graph). We implement the NRM’s IPQs using array-based binary heaps, which provide logarithmic scaling of the runtime with the number of reactions for the SSA steps in a sparsely-coupled model (i.e. a model whose reaction dependency graph is sparse).

To allow compartments to be quickly created and destroyed, a separate IPQ is created for each compartment. These IPQs are inserted into a higher-level IPQ which acts as a “Next Compartment Method”, allowing us to determine which compartment the next reaction will occur in, in logarithmic time with the number of compartments. Creating/destroying compartments is then done by constructing/destructing these IPQs and inserting/removing them from the overall IPQ. In this arrangement, compartment creation takes O(logC + RlogR) time, while compartment destruction takes O(log C) time, where R is the number of reactions in the new compartment and C is the current number of compartments in the simulation.

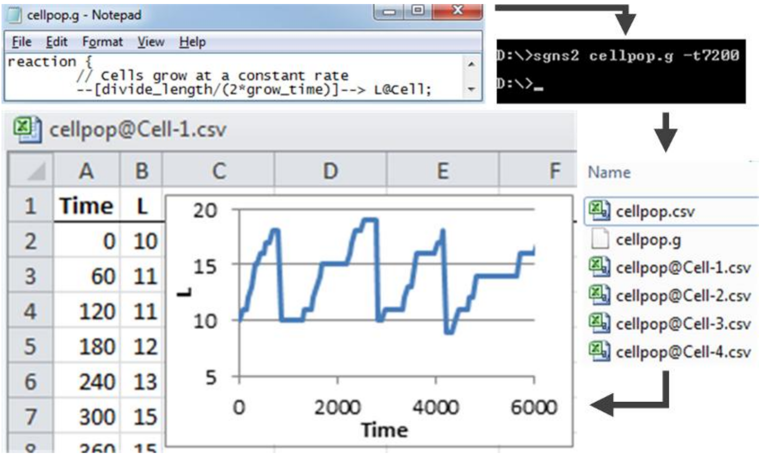
Communication between compartments is accomplished by reactions that affect molecules in both a ‘parent’ and a ‘child’ compartment. Since the propensity of each instance of such a reaction depends on the population of the reactant in the parent compartment, O(C) propensities must be recalculated when this quantity changes, an O(ClogR) operation. Since each reactant of a reaction factors independently into the propensity of the reaction, the reactant in the parent compartment can be factored out from all of the instances of the reactions in the sub-compartments. This calculation is similar to the partial propensity methods [3]. To accomplish this without requiring an O(C) operation, we create a separate IPQ for the sub-compartment’s reaction instances in which the local simulation time, $t_{\text{sub}}$, is advanced such that $dt_{\text{sub}} = Xdt$, where X is the current population size of the reactant in the parent compartment and t is the global simulation's time variable. This sub-simulation then publishes a next firing time to the parent compartment's IPQ, adjusted according to the NRM’s propensity update formula. When the parent compartment reactant's population changes, only the adjusted next firing time must be recalculated and only one element of an IPQ may change position, reducing the cost of this operation to O(logR). SGNS2 assumes that there are no direct interactions between compartments at the same level of the hierarchy.

To include multi-delayed reactions as well, which are simulated according to the Delayed SSA[4], we implement a wait list using a binary heap-based priority queue. The transient nature of compartments makes it necessary for each to contain its own wait list. The earliest event in a compartment's wait list is then inserted into the compartment's IPQ. All operations on the wait lists are therefore O(logW + logR + logC), where W is the number of delayed events on that wait list. When a compartment is destroyed, all delayed events in that compartment are forgotten, assuring that no delayed molecules of that compartment are released following this event.

## References

1. M. A. Gibson and J. Bruck, J. Phys. Chem. A 104, 1876, 2000.
2. D. T. Gillespie, J. Phys. Chem. 81, 2340, 1977.
3. R. Ramaswamy, N. González-Segredo, and I. F. Sbalzarini, J. Chem. Phys. 130, 244104, 2009.
4. M. R. Roussel and R. Zhu, Phys. Biol. 3, 274, 2006.

## Supplementary Figures



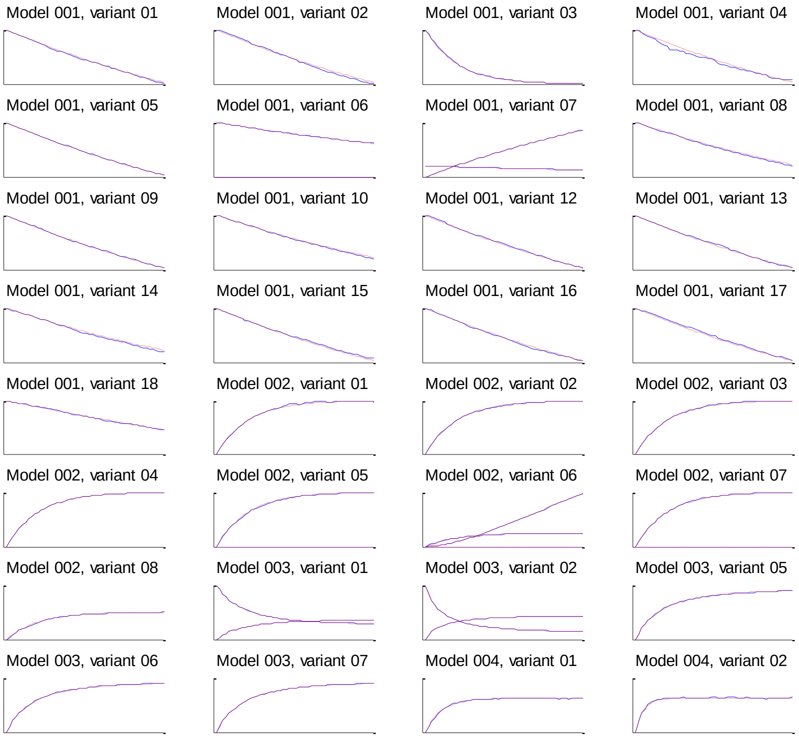
```
reaction {
        // Cells grow at a constant rate
        --[divide_length/(2*grow_time)]--> L@Cell;
```

```
D:\>sgns2 cellpop.g -t7200
D:\>_
```

| | A | B |
|---|---|---|
| 1 | Time | L |
| 2 | 0 | 10 |
| 3 | 60 | 11 |
| 4 | 120 | 11 |
| 5 | 180 | 12 |
| 6 | 240 | 13 |
| 7 | 300 | 15 |
| 8 | 360 | 15 |

Name
- cellpop.csv
- cellpop.g
- cellpop@Cell-1.csv
- cellpop@Cell-2.csv
- cellpop@Cell-3.csv
- cellpop@Cell-4.csv

Fig S1: Example of SGNS2 in use. A model is created in a text editor, here Notepad (upper left), and is simulated with SGNS2 (upper right). The csv files output (lower right) are loaded and analyzed in Excel (lower left).



Model 001, variant 01
Model 001, variant 02
Model 001, variant 03
Model 001, variant 04
Model 001, variant 05
Model 001, variant 06
Model 001, variant 07
Model 001, variant 08
Model 001, variant 09
Model 001, variant 10
Model 001, variant 12
Model 001, variant 13
Model 001, variant 14
Model 001, variant 15
Model 001, variant 16
Model 001, variant 17
Model 001, variant 18
Model 002, variant 01
Model 002, variant 02
Model 002, variant 03
Model 002, variant 04
Model 002, variant 05
Model 002, variant 06
Model 002, variant 07
Model 002, variant 08
Model 003, variant 01
Model 003, variant 02
Model 003, variant 05
Model 003, variant 06
Model 003, variant 07
Model 004, variant 01
Model 004, variant 02

Fig S2: Means of molecule populations over time for the models in the Discrete Stochastic Model Test Suite which do not use Rules or Events. Solid blue lines are the means of the results from 500 runs of SGNS2, while dashed red lines show the analytical solutions. The overlap of these lines results in a purple-like line.

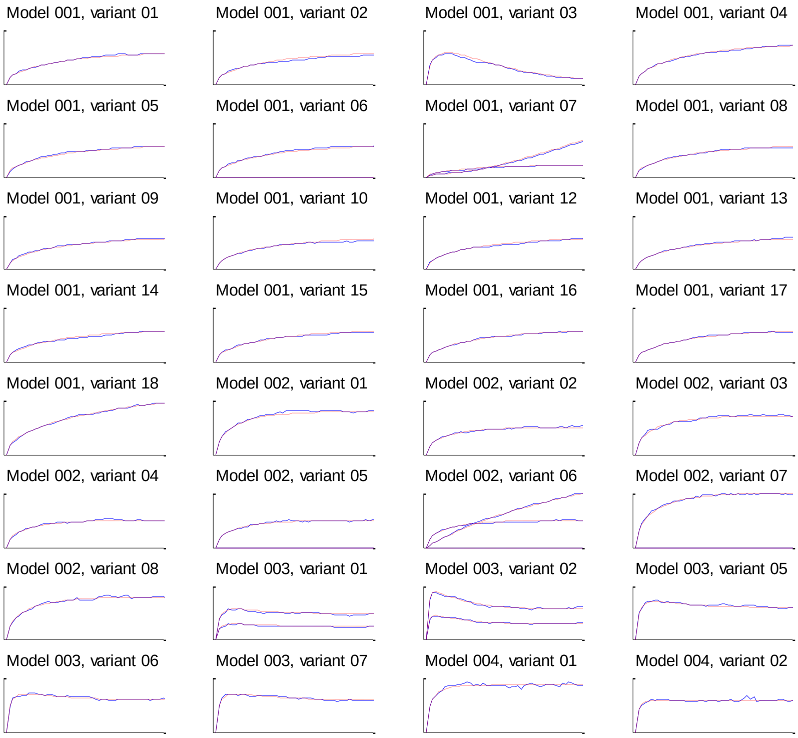


Fig S3: Standard deviations of molecule populations over time for the models in the Discrete Stochastic Model Test Suite which do not use Rules or Events. Solid blue lines are the standard deviations of the results from 500 runs of SGNS2, while dashed red lines show the analytical solutions. The overlap of these lines results in a purple-like line.

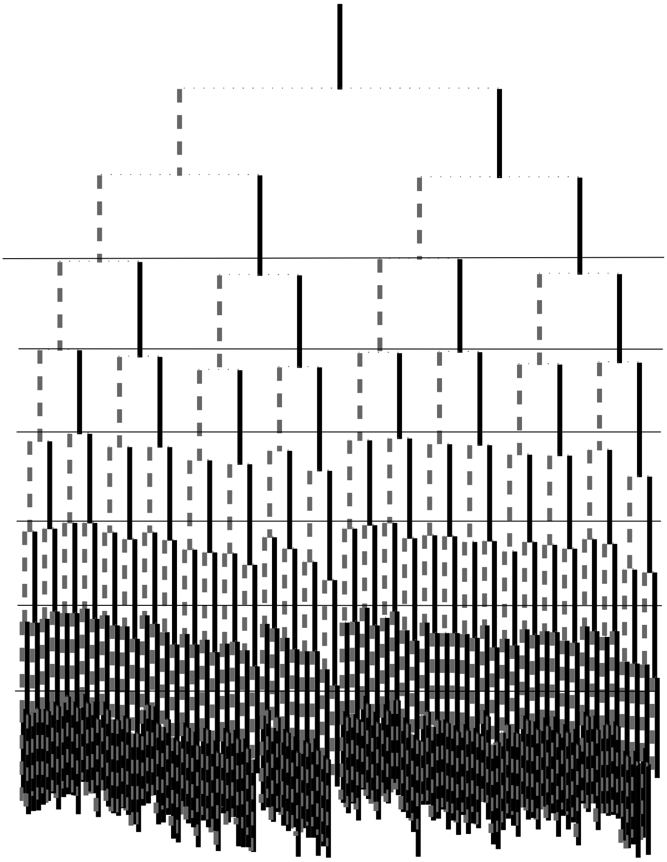


Fig S4: Cell lineage with biased partitioning of vitality-reducing protein aggregates. Cells with older poles are placed on the right. The length of each cell's line is proportional to its lifetime.

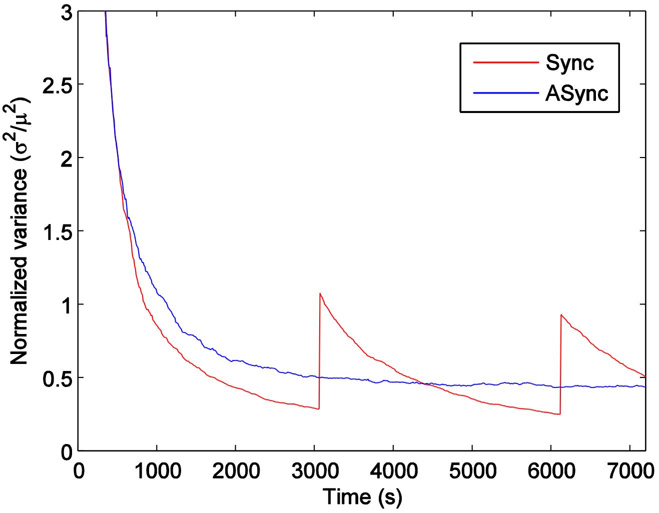


Fig S5: Normalized variance of RNA numbers over time in perfectly synchronous (red) and asynchronous (blue) cell populations. 500 cells were simulated in each population.

# Publication II

Jason Lloyd-Price and Andre S. Ribeiro "Bistability in a stochastic RNA-mediated gene network", *Physical Review E*

# Bistability in a stochastic RNA-mediated gene network

Jason Lloyd-Price and Andre S. Ribeiro*

*Laboratory of Biosystem Dynamics, Department of Signal Processing, Tampere University of Technology, P.O. Box 527, FI-33101 Tampere, Finland*

(Received 3 May 2013; revised manuscript received 10 July 2013; published 23 September 2013)

Small regulatory RNAs (srRNAs) are important regulators of gene expression in eukaryotes and prokaryotes. A common motif containing srRNA is a bistable two-gene motif where one gene codes for a transcription factor (TF) which represses the transcription of the second gene, whose transcript is a srRNA which targets the first gene's transcript. Here, we investigate the properties of this motif in a stochastic model which takes the low copy numbers of the RNA components into account. First, we examine the conditions for stability of the two "noisy attractors." We find that for realistic low copy numbers, extreme, but within realistic intervals, mutual repression strengths are required to compensate for the variability of the RNA numbers and thus, achieve long-term bistability. Second, the promoter initiation kinetics is found to strongly influence the bistability of the switch. Super-Poissonian RNA production disrupts the ability of the srRNA to silence its target, though sub-Poissonian RNA production does not rule out the need for strong mutual repression. Finally, we show that asymmetry between the two interactions forming the switch allows an external input to induce the transition from "high srRNA" to "'high TF" more easily (i.e., with a shorter input) than in the opposite direction. We hypothesize that this asymmetric switching property allows these circuits to be more sensitive to one external input, without sacrificing the stability of one of the noisy attractors.

DOI: 10.1103/PhysRevE.88.032714 PACS number(s): 87.18.Cf, 87.18.Tt, 87.10.Mn, 87.14.gn

## I. INTRODUCTION

Small noncoding regulatory RNAs (srRNAs) have been found targeting the majority of eukaryotic genomes [1], and are abundant in prokaryotes as well [2,3]. In bacteria, srRNAs generally modify the expression of their target genes by binding to the 5′ region of the messenger RNA, and inhibit translation by blocking the ribosome binding site [3], usually resulting in the degradation of the target mRNA and often also the srRNA [3]. This regulation scheme differs from transcription factor (TF) based regulation in several aspects, the most important being that when the target is expressed at a lower rate than the srRNA, it is nearly fully silenced [4], while above the srRNA production rate, the target's expression increases linearly. This regulatory function is highly nonlinear and is believed to be responsible for several complex behaviors in genetic circuits [5].

TF-based and srRNA-based regulatory mechanisms function together in gene regulatory networks, and a number of such mixed motifs have been identified including various feedforward and feedback loops [6]. Of interest is the srRNA-mediated double feedback loop (SMDFL), which is present in both eukaryotic and prokaryotic organisms (see, for example, [6–12]). In this motif, a srRNA represses a gene, whose protein is a TF which represses transcription of the srRNA. This network has been shown to exhibit bistability [7], and can thus operate as a switch, similar to the genetic Toggle Switch motif in which two TFs mutually repress each other [13].

Some bistable circuits are involved in cell fate decisions [14,15], including the SMDFL [9–12]. Such switches must remain in a state for long periods of time [16]. On the other hand, cell populations can take advantage of unstable switches to generate phenotypic diversity and increase fitness in unpredictable environments [17]. One source of instability is noise in gene expression.

srRNA-mediated repressive interactions have interesting noise properties [4]. When the srRNA production rate is significantly below the target mRNA production rate, the noise in the protein numbers over time, as measured by the Fano factor, is as in the unrepressed system. On the other hand, when the srRNA production rate is significantly above the production rate of the target mRNA, the protein Fano factor decreases to 1, since the srRNA decreases the protein burst size from each mRNA and, consequently, protein production becomes Poissonian [4]. When the two rates are approximately equal, near-critical phenomena increase the noise in the protein numbers beyond the level in the nonrepressed case [18]. This noise is, in turn, dependent on the initiation kinetics at the promoter, for which evidence exists for a range of kinetics, from bursty [19,20], to sub-Poissonian [21,22]. Given the above, it is nonobvious how low RNA copy numbers affect the dynamics of the SMDFL.

Here, we study the behavior of a stochastic model of the SMDFL within realistic parameter ranges. We focus on how the behavior of the switch is affected by low copy numbers, TF repression strengths, srRNA production rates, and different promoter initiation kinetics. Finally, we study how asymmetries between the two interactions forming the switch affect its sensitivity to external inputs.

## II. METHODS

### A. Stochastic model of the srRNA-mediated double feedback loop

We use a stochastic version of the srRNA-mediated double feedback loop model presented in [7]. This model, depicted in

[*] Author to whom correspondence should be addressed: andre.ribeiro@tut.fi

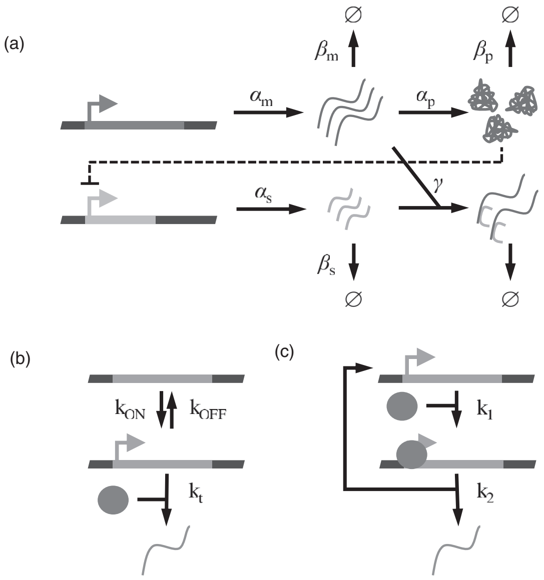


FIG. 1. Cartoon of the model. (a) mRNA (dark gray) and srRNA (light gray) are transcribed from the DNA with mean rates $\alpha_m$ and $\alpha_s$, respectively, degrade via first-order reactions with rates $\beta_m$ and $\beta_s$, respectively, and irreversibly bind to one another with rate $\gamma$. The mRNA is translated into transcription factors with mean rate $\alpha_p$, which degrade as a first-order reaction with rate $\beta_p$, and can bind to the srRNA gene's promoter, repressing it. (b) Telegraph model of transcription regulation [19]. The gene stochastically switches between OFF and ON states. RNA polymerase (ball) can transcribe the gene only when it is ON. (c) Multistep model of transcription initiation [25]. The RNA polymerase (ball) must perform a series of time-consuming steps (here two) to initiate transcription. The parameters are explained in Table I.

Fig. 1(a), consists of reactions (1)–(9):

$$\mathrm{Pro}_m \xrightarrow{\alpha_m} \mathrm{Pro}_m + R_m, \tag{1}$$

$$R_m \xrightarrow{\beta_m} \varnothing, \tag{2}$$

$$R_m \xrightarrow{\alpha_p} R_m + P, \tag{3}$$

$$P \xrightarrow{\beta_p} \varnothing, \tag{4}$$

$$\mathrm{Pro}_s \xrightarrow{\alpha_s} \mathrm{Pro}_s + R_s, \tag{5}$$

$$R_s \xrightarrow{\beta_s} \varnothing, \tag{6}$$

$$R_m + R_s \xrightarrow{\gamma} \varnothing, \tag{7}$$

$$\mathrm{Pro}_s + P \xrightarrow{k_r} \mathrm{Pro}_s \bullet P, \tag{8}$$

$$\mathrm{Pro}_s \bullet P \xrightarrow{k_u} \mathrm{Pro}_s + P. \tag{9}$$

Here, $m$ and $s$ are the genes producing the TF and the srRNA, respectively. $\mathrm{Pro}_x$ and $R_x$ are the promoter of gene $x$ and its transcript, respectively. $P$ is the TF and $\mathrm{Pro}_s \bullet P$ is the repressed promoter.

It is worth mentioning that srRNA regulation is achieved in a number of ways. Firstly, srRNA can bind to the target and actively promote degradation of both the target and regulatory RNAs [reaction (7)]. Alternatively, the srRNA may bind to the target and prevent translation, but not promote degradation. If this binding is strong and the srRNA cannot dissociate from the mRNA, this is dynamically equivalent to the first scenario since, in both cases, the mRNA is unable to produce proteins after the srRNA-mRNA binding event. The weak-binding scenario is not considered here, due to its requiring modifications in the model that are beyond the scope of this work. A third option exists, also not considered here for similar reasons, whereby the srRNA promotes the degradation of the target, but is not itself consumed [23].

Most parameters of the model were set to realistic values (Table I). The remaining ones were reparametrized to introduce three dynamically relevant, but not necessarily physically relevant parameters: $\theta$, $R$, and $\lambda$. $\theta$ controls the system size, which scales the mean copy numbers of RNA and proteins in the model, and was arbitrarily chosen to represent the mean number of mRNA molecules if there were no srRNA regulation (specifically, $\alpha_m = \theta\beta_m$). Decreasing $\theta$ increases low-copy-number effects, while increasing $\theta$ makes the system more similar in behavior to the deterministic solution (see Supplemental Material [24]). $R$ controls the strength of the TF's repression of the srRNA gene's promoter by setting the dissociation constant TF-promoter interactions to $\mu_P R^{-1}$, where $\mu_P$ is the mean amount of TF produced with no srRNA interaction (specifically, $K_d = \mu_P R^{-1} = \alpha_m B \beta_p^{-1} R^{-1}$). Thus, a value of 2 sets $K_d$ to one half of the unrepressed TF mean. Lastly, $\lambda$ controls the srRNA repression efficiency, and is equal to the ratio between srRNA production and mRNA production, in the absence of TF regulation (specifically, $\alpha_s = \lambda\alpha_m$). Higher $\lambda$ increases the repression strength of the srRNA. $\lambda$ must be at least 1 in order to fully silence the gene. Note that since $\alpha_s$ is a multiple of $\alpha_m$, the mean srRNA production rate also scales with $\theta$.

In this model, only a single TF represses the srRNA promoter [reactions (8) and (9)]. Since it does so as a monomer, the repression does not introduce nonlinear effects. Nonlinear mechanisms, such as cooperativity and multimerization, can greatly enhance the stability of a switch, though they are not necessary [25]. If bistability is observed in the present model, it should also be observable and enhanced in a model with these properties [25]. We therefore do not consider these cases here.

The deterministic kinetic equations corresponding to the reactions given above are presented in the Supplemental Material [24], along with the analysis methods to determine the regions of bistability.

### B. Promoter initiation dynamics

To test the effects of different RNA production dynamics, we employ two extra models of initiation. We characterize the amount of noise that these alternate RNA initiation models will produce in the RNA time series by the coefficient of variation ($\eta$, defined as the variance over the squared mean) of the distribution of time intervals between RNA production events. Except for very narrow, near-deterministic distributions, which is not the case here, the $\eta^2$ of this distribution captures the

TABLE I. Model parameters, values, and sources.

| Parameter | Meaning | Value | Source |
|---|---|---|---|
| $\theta$ | Mean mRNA numbers | $10^{-3}$–1 | Reference [20] |
| $R$ | TF repression strength | | |
| $\lambda$ | srRNA repression strength | | |
| $\alpha_m$ | mRNA production rate | $\theta\beta_m$ s$^{-1}$ | See Methods |
| $\alpha_s$ | srRNA production rate | $\lambda\alpha_m$ s$^{-1}$ | See Methods |
| $\beta_m$ | mRNA degradation rate | $600^{-1}$ s$^{-1}$ | Reference [34] |
| $\beta_s$ | srRNA degradation rate | $3000^{-1}$ s$^{-1}$ | Reference [41] |
| $\gamma$ | mRNA-srRNA binding rate | $0.1\theta^{-1}$ s$^{-1}$ | Reference [42] |
| $\alpha_p$ | Protein production rate | $B\beta_m$ s$^{-1}$ | Set to match $B$ |
| $B$ | Protein burst size per mRNA | 4.2 | Reference [43] |
| $\beta_p$ | Protein degradation rate | $36\,000^{-1}$ s$^{-1}$ | References [35,36] |
| $k_u$ | Unrepression rate | $25^{-1}$ s$^{-1}$ | Reference [44] |
| $k_r$ | Repression rate | $k_{\mathrm{unrep}}K_d^{-1}$ s$^{-1}$ | Set to match $K_d$ |
| $K_d$ | TF-promoter dissociation constant | $\alpha_m B\beta_p^{-1}R^{-1}$ | See Methods |

contribution of the initiation dynamics to the fluctuations in the numbers of RNA and protein molecules over time [21].

The default reactions modeling RNA production are reactions (1) and (5). These result in an exponential distribution of time intervals between RNA production intervals, and thus have a $\eta^2$ of 1. Since this distribution produces a Poisson-distributed number of production events in a fixed time window, this initiation dynamics is termed Poissonian. Noisier-than-Poissonian production kinetics are achieved by a promoter that can randomly transition between an OFF and an ON state, and which only allows transcription when ON [19], producing bursts of RNA production. The reactions modeling this promoter are depicted in Fig. 1(b), and are as follows, where $x$ is replaced by $m$ or $s$ when replacing reactions (1) or (5), respectively:

$$\mathrm{Pro}_x^{\mathrm{ON}} \xrightarrow{k_{\mathrm{OFF}}} \mathrm{Pro}_x^{\mathrm{OFF}}, \tag{10}$$

$$\mathrm{Pro}_x^{\mathrm{OFF}} \xrightarrow{k_{\mathrm{ON}}} \mathrm{Pro}_x^{\mathrm{ON}}, \tag{11}$$

$$\mathrm{Pro}_x + \mathrm{Pro}_x^{\mathrm{ON}} \xrightarrow{k_t} \mathrm{Pro}_x + \mathrm{Pro}_x^{\mathrm{ON}} + R_x. \tag{12}$$

Here, reaction (12) should not be confused for a bimolecular reaction. The notation only implies that the promoter must be ON and unrepressed for transcription to occur. We assume that bursts take a very short amount of time compared to the interburst time (i.e., $k_{\mathrm{ON}} \ll k_{\mathrm{OFF}}$), and thus set $k_{\mathrm{OFF}}$ to 1 s$^{-1}$. It can be shown (see Supplemental Material [24]) that reactions (10)–(12) produce a $\eta^2$ of $2S + 1$, where $S = k_t/k_{\mathrm{OFF}}$ is the mean number of RNA molecules produced in each burst. To obtain a specific $\eta^2$ in Fig. 5(a), we therefore set $k_t = (\eta^2 - 1)/2$, and $k_{\mathrm{ON}} = \alpha_x/S$ to match the mean production rate.

Sub-Poissonian dynamics is achieved with a promoter model that requires a series of Poissonian steps to be completed before an RNA is produced [26,27]. The reactions modeling this promoter dynamics are depicted in Fig. 1(c), and are as follows, where $N > 1$ is the total number of steps involved, $1 < n < N$, and $x$ is replaced by $m$ or $s$ when replacing reactions (1) or (5), respectively:

$$\mathrm{Pro}_x \xrightarrow{k_1} \mathrm{Pro}_x^1, \tag{13}$$

$$\mathrm{Pro}_x^{n-1} \xrightarrow{k_n} \mathrm{Pro}_x^n, \tag{14}$$

$$\mathrm{Pro}_x^{N-1} \xrightarrow{k_N} \mathrm{Pro}_x + R_x. \tag{15}$$

It can be shown (see Supplemental Material [24]) that reactions (13)–(15) produce a $\eta^2$ of $1/N$. To obtain a specific $\eta^2$ in Fig. 5(a), we therefore set $N = 1/\eta^2$, and $k_i = N\alpha_x$ for $1 \leqslant i \leqslant N$ to match the mean production rate.

### C. Characterization of noisy attractors

Stable states do not technically exist in the stochastic model above, since the probability that the system will leave any state after reaching it approaches 1 as time goes to infinity. We therefore use the term "noisy attractor" to refer to a set of microstates from which the system is unlikely to leave for a physiologically relevant time frame [28]. These noisy attractors correspond roughly, but not always, to the stable states found in the corresponding, deterministic model. For example, unstable steady states of the deterministic model will vanish in the stochastic model while stable steady states either remain the same or can vanish or settle around different mean molecule concentrations.

Since the system is not symmetric as in a toggle switch of two mutually repressing TF-coding genes, it is not immediately clear how to group the microstates of the system into noisy attractors. Here, the categorization was performed by examining the overall joint distribution of TF and srRNA populations for each value of $\theta$ and selecting a threshold in this plane that separated the two modes. States for which $P - 10R_s - 10\theta > 0$ were classified as part of the TF-high noisy attractor, while other microstates were categorized as part of the srRNA-high noisy attractor.

The stability of a noisy attractor is defined as the mean time that the system will remain in that region of the state space before stochastically leaving it (and in this case, traveling to the other noisy attractor). For both noisy attractors, this quantity was measured by initializing a simulation with RNA and/or protein populations set to the mean amount that would be produced with no repression (i.e., $R_m = \theta$ and $P = \alpha_m B\beta_p^{-1}$ for TF-high and $R_s = \alpha_s/\beta_s$ for srRNA-high) and simulating until it switched to the other noisy attractor, sampling every hour, limited to 1 month of simulation time. Simulations were conducted in SGNS2 [29], a stochastic molecular dynamics simulator based on the stochastic simulation algorithm [30].

## III. RESULTS

The bistable regions of the parameter space of the deterministic version of the SMDFL have been studied previously [7]. The regions of bistability found in the deterministic solution are recovered in the high-copy-number limit of the present model (Fig. S1 in Supplemental Material [24]), i.e., in the high-$\theta$ limit. Since $\theta$ was chosen to represent the mean RNA numbers of the unrepressed TF-encoding gene, we can use genome-wide measurements in cell populations of *Escherichia coli* to place it within a realistic range, measured to be $\sim 10^{-3}$–1 [20].

### A. Robust bistability

We first study what TF and srRNA repression strengths (parameters $R$ and $\lambda$, respectively) are required to achieve robust bistability in the stochastic model with $\theta = 1$, at the higher end of the realistic range. We define "robust bistability" as when the system can remain in either noisy attractor for at least 1 month of simulation time, on average. Results are shown in Fig. 2(a). In this case, robust bistability is achieved when $\lambda > 2.75$ and $R$ is sufficiently strong for the chosen $\lambda$. This is shown in Fig. 2(b), where the TF population from two independent runs holds its initial state (high or low) for the duration of the simulation. Meanwhile, Fig. 2(d) shows monostability, where the system is unable to stay in the srRNA-high, TF-low noisy attractor due to insufficient $\lambda$, despite lying within the parameter region of deterministic bistability. Figure 2(c) shows the classic stochastic toggle switch behavior where the switch stochastically jumps between the two noisy attractors.

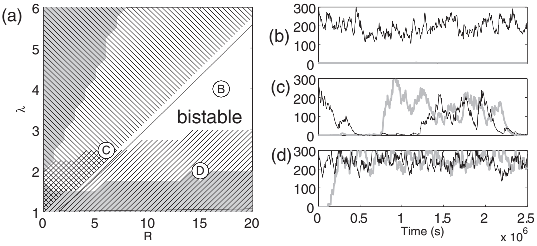


FIG. 2. (a) Bifurcation diagram with $\lambda$ and $R$ as control parameters and $\theta = 1$. Hatched areas indicate where a noisy attractor is not robustly stable (i.e., stable for less than 90% of 1 month of simulation time, on average). Upwards and downwards hatching indicates the srRNA-high and TF-high noisy attractors are unstable, respectively. Shaded areas indicate that the unstable noisy attractor is less stable than 5% of 1 month of simulation time (i.e., the switch is monostable or unstable). This diagram and all subsequent ones are from 500 runs per tested parameter pair and initial state. The solid line indicates the extent of the region where the deterministic model is bistable (see Supplemental Material [24]). Example time series of TF populations alone are shown from two independent simulations with (b) $R = 17$, $\lambda = 4$ (robustly bistable), (c) $R = 6$, $\lambda = 2.5$ (weak bistability), and (d) $R = 15$, $\lambda = 2$ (monostable), with initial conditions set to start in the TF-high (black line) and in the srRNA-high, TF-low (gray line) noisy attractors. Note that in (b), the gray line remains very low for the duration of the simulation.

The region of robust bistability appears to be a subset of the region of deterministic bistability in Fig. 2(a). Interestingly, outside the region of deterministic bistability, it is possible for one or both of the noisy attractors to be only transiently stable. That is, the system remains in a noisy attractor for 5%–90% of 1 month of simulation time. When both noisy attractors are only transiently stable, the switch stochastically transitions unbiasedly between them [double-hatched region in Fig. 2(a)].

The highest value of $R$ shown, 20, corresponds to a dissociation constant between the promoter and the TF of approximately $K_d = \theta\beta_m B\beta_p^{-1}R^{-1} = 12$ molecules, which is within realistic ranges for TF-promoter interactions [31]. Thus, the TF repression strength required to achieve robust bistability is within realistic bounds. We are not aware of measurements of srRNA production rates. Nevertheless, using transcription rates of protein-coding genes [20] as a guide, the values of $\lambda$ required to achieve robust stability are high, since the high value of $\theta$ already places the mRNA production rate at the upper limit of the range observed in *E. coli* [20]. Thus, we next study how the bifurcation diagram changes for lower $\theta$.

### B. Low copy numbers

Low-copy-number effects, i.e., when $\theta$ is lowered, are expected to significantly affect the stability of the noisy attractors of the switch. This is shown in Fig. 3, as $\theta$ is lowered from 1 to 0.5, 0.2, and finally 0.1. Robust bistability (i.e., existence of two distinct noisy attractors) is observable within realistic ranges for a limited range for $\theta = 0.5$. The likely cause for this is that srRNA-based repression is based on the interaction between two species with few copies in the cell at any given time. Consistent with this explanation, the strength of TF repression required to stabilize the TF-high noisy attractor is largely unchanged from the $\theta = 1$ case.

For $\theta = 0.2$ and $\theta = 0.1$, robust bistability is lost for the same parameter range. Worse, the same value of $R$ corresponds to increasing repression strength as $\theta$ is decreased, and with $R = 20$ and $\theta = 0.1$ this corresponds to a TF-promoter dissociation constant of less than ∼ two molecules, which is extreme but realizable [32]. The region where both noisy attractors are

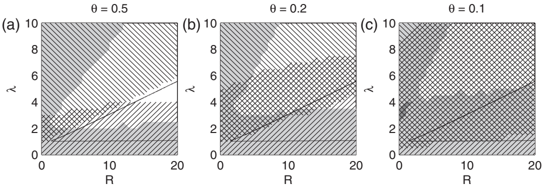


FIG. 3. Bifurcation diagram with $\lambda$ and $R$ as control parameters, with (a) $\theta = 0.5$, (b) $\theta = 0.2$, (c) $\theta = 0.1$.

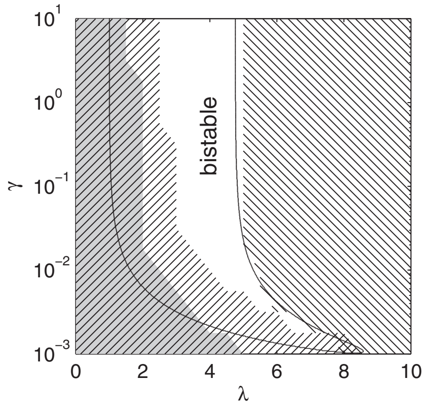


FIG. 4. Bifurcation diagram using $\gamma$ and $\lambda$ as control parameters, with $\theta = 1$ and $R = 17$.

stable on an intermediate time scale grows as $\theta$ is reduced, and begins to cover most of the tested parameter space. Thus, short-term bistability remains possible in this regime.

### C. mRNA-srRNA binding

One parameter for which we could not find measurements is $\gamma$, which was set following a previous model of srRNA regulation. This parameter controls the time it will take for an mRNA to bind to an srRNA, and therefore affects the effectiveness of srRNA repression. To understand how this parameter can affect the switch, we generated the bifurcation diagram of the switch using $\lambda$ and $\gamma$ as control parameters, shown in Fig. 4. Above a certain critical value, here ~0.01 $\text{s}^{-1}$ per mRNA-srRNA pair, the dynamics does not change significantly. Below this value, the region of bistability shrinks rapidly to a point where small changes in $\gamma$ can move the switch from monostable TF-high to monostable srRNA-high.

We note that $\gamma$ scales inversely with $\theta$. This scaling allows the stochastic model to converge to the deterministic solution in the high-$\theta$ limit. We tested whether the changes observed in Fig. 3 resulted from this scaling. Setting $\gamma$ to 0.1 (Fig. S2, Supplemental Material [24]), we found no appreciable change.

### D. Promoter kinetics

Since regulation by srRNA has been shown to have nontrivial noise characteristics [4], it is of interest to study how a network involving srRNA interactions behaves with different noise properties. To this end, we varied the level of noise introduced by transcription initiation and observed how it affects the stability of the switch, starting from a parameter set where robust bistability is observed when both promoters are Poissonian ($\lambda = 4$ and $R = 17$). The level of noise in the production of an RNA species was adjusted by replacing the appropriate RNA production reaction [reaction (1) or (5)] with a set of reactions producing a distribution of intervals between production events with a given $\eta^2$ [reactions (10)–(12) for $\eta^2 > 1$ or reactions (13)–(15) for $\eta^2 < 1$; see Methods]. The results are shown in Fig. 5(a).

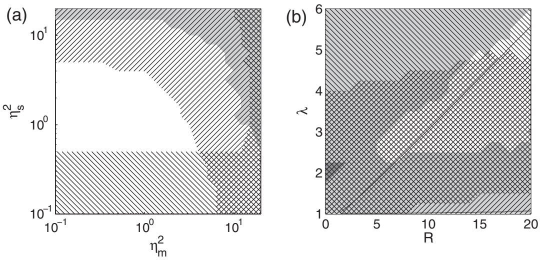


FIG. 5. (a) Bifurcation diagram with different promoter kinetics. Control parameters are the $\eta^2$ (variance over the squared mean) of the time interval distribution between transcription initiations for the srRNA production ($\eta_s^2$) and for the TF production ($\eta_m^2$), assuming no regulation. These were modified by replacing reactions (1) or (5), or both with reaction sets (10)–(12) or (13)–(15) to obtain a given $\eta^2$ for both mRNA and srRNA production intervals (see Methods). For reference, the least noisy interval distribution has $N = 10$ steps, while the most noisy has a burst size of $S = 9.5$. $\theta = 1$, $\lambda = 4$, $R = 17$. (b) Bifurcation diagram as in Fig. 3(b), with $\theta = 0.2$, and a four-step promoter for srRNA.

When either promoter is bursty, the srRNA-high noisy attractor loses stability. This is expected, since srRNA regulation involves interaction between two RNA species, which are both in low copy number. Consistent with this, the noise in srRNA production ($\eta_s^2$) has the strongest overall effect on the stability, determining whether it is bistable or monostable in either noisy attractor. Since it is monostable in the low-noise srRNA production region, it appears that this allows it to repress its target more consistently. Given the loss of bistability in the low-$\theta$ region of the parameter space [Figs. 3(a)–3(c)], it is plausible that this increase in regulation strength with more deterministic production might allow the switch to operate more effectively for low $\theta$. We therefore repeated Fig. 3(c) with four-step, sub-Poissonian srRNA production, shown in Fig. 5(b). This change was not sufficient to restore robust bistability in the parameter range tested. Instead, although the $\lambda$ required to stabilize the srRNA-high noisy attractor has decreased, this change came at the cost of the stability of the TF-high noisy attractor. That is, the $R$ required to stabilize the TF-high noisy attractor is considerably greater. Note that no deterministic region of bistability is displayed in Fig. 5(a), since no parameters affecting deterministic bistability were varied in this figure.

### E. Asymmetric switching

Robust bistability is achievable in noncooperative TF-based toggle switches as well [25]. Under what circumstances then would an srRNA-mediated switch be preferable to use in a real genetic circuit rather than a purely TF-mediated switch? One difference between the two motifs is that one of the regulatory molecules (the srRNA) has a much smaller half-life than most natural proteins, despite its extended lifetime due to the binding of Hfq [33] in comparison to mRNA [34]. This allows its level to decrease more quickly in response to regulation. We therefore expect that the switch is able to change from the srRNA-high noisy attractor to the TF-high noisy attractor

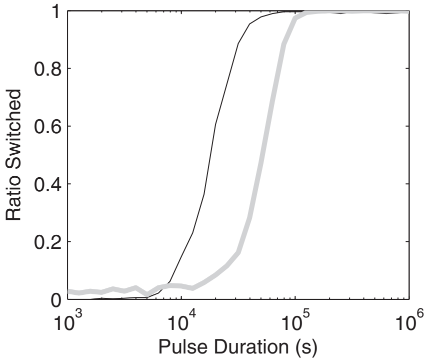


FIG. 6. Fraction of SMDFL switches that changed noisy attractor and remained in the new one after an input pulse of varying duration. The simulation was started in either the srRNA-high (black line) or the TF-high noisy attractor (thick gray line), with $\theta = 1$, $R = 15$, and $\lambda = 4$. The simulation was first run for 100 h, after which an input pulse of the given duration was applied, where $R$ or $\lambda$ was scaled by 0.2, to push the switch into the opposite noisy attractor. After the pulse, the simulation was run for another 100 h to allow the switch to settle into its new noisy attractor, and the final state was measured. Data are from 500 simulations for each tested pulse duration.

much faster than vice versa. To test this, we simulated the switch, starting in one of the two noisy attractors, in a robustly bistable region of the parameter space ($R = 15$, $\lambda = 4$, $\theta = 1$) for 100 h, and applied an input pulse of varying durations. This pulse moved the system into a region of the parameter space which is monostable in the other noisy attractor by scaling $R$ or $\lambda$ by $1/5$. The switch was then simulated for another 100 h and the final state was recorded.

The fraction of times the switch was found in the other noisy attractor at that stage is shown in Fig. 6. From the figure, the switch displays a strong asymmetry in the duration of the input pulse required to switch noisy attractor. Specifically, half of srRNA-high simulations ended in the TF-high noisy attractor after applying a pulse of 15 000 s, while switches in the TF-high attractor take a much longer input pulse of 50 000 s for half to change noisy attractors.

## IV. CONCLUSIONS AND DISCUSSION

Using a stochastic model of the SMDFL with most parameters taken from the literature, we showed that, within realistic parameter ranges, this model can exhibit robust bistability. That is, the stochastic fluctuations of RNA and protein numbers cannot, on average, cause the switch to change noisy attractor within 1 month of simulation time. Reducing the mean RNA and protein numbers (i.e., increasing the finite-size effects) limited the regions of robust bistability, owing to the inability of the srRNA to reliably repress its target at such low mean levels. Realistically realizable regions of long-term bistability exist down to $\theta = 0.5$, despite the absence of cooperative repression by the TF. Below this, robust bistability is lost for realistic repression strengths. Similarly, for highly noisy srRNA production and for noisy mRNA production, the srRNA loses effectiveness, with the srRNA production variability having the largest impact. Lower levels of noise in srRNA production increase the effectiveness of the srRNA regulation, but decrease the stability of the TF-high noisy attractor, and thus cannot be used to compensate for low-copy-number effects to regain bistability in the low-RNA parameter range. Thus, such switches must operate at the higher end of the mean RNA number spectrum in order to function reliably, or must have some additional machinery to strengthen the regulation such as cooperative repression by the TFs.

One of the parameters for which we could not find measurements in the literature is $\gamma$, which controls the mRNA-srRNA binding rate. Examining the dependence of the dynamics on this parameter reveals that there is a point, here $\sim 0.01\theta^{-1}$ s$^{-1}$ per mRNA-srRNA pair, beyond which further increases do not change the dynamics of the switch. Below this point, the bistability of the switch is sensitive to changes in $\gamma$. This parameter controls the rate of a bimolecular reaction, and is therefore likely to be diffusion limited, and will change with, for example, temperature. Having a slow binding rate and placing the switch in the lower-right portion of the parameter space shown in Fig. 4 might therefore be a way to create a temperature-dependent switch, without the need for a specific sensing apparatus.

By applying input pulses of varying time length to determine how quickly the switch can change to another noisy attractor under external control, we determined that the change from the srRNA-high noisy attractor to the TF-high noisy attractor is considerably faster than the reverse, due to the higher degradation rate of the srRNA. This asymmetry allows rapid changes from one noisy attractor to the other based on a single, short input, but requires a much longer, sustained input to change the other way. Most proteins have much longer half-lives [35,36], making this hard to achieve in switches relying on TFs alone, though it could be accomplished by active degradation of the proteins [37], which thus requires a larger number of interactive players in the system (and, most likely, additional energy expenditure).

We note that the present model does not include the effects of cellular growth and division, which act as an increased degradation rate of all cellular components. This will affect proteins more than RNAs since they have a longer mean lifetime. This should cause the asymmetry in switching times to decrease, but remain, in fast-growing bacterial populations under optimal growth conditions. However, in natural environments, cell populations are not commonly under optimal conditions meaning that the mean division rates are much slower. It is also worthwhile to note that, similar to results from measurements in live cells, our results are expected to depend, to some extent, on the values of some of the parameters not varied in the present study. Our choice of parameters to vary was based on our observations of which were more prone to cause behavior modifications. Nevertheless, future studies may provide additional insight into the currently unknown relevance of some of the untested parameters. Another interesting study would be to investigate how the kinetics of the model changes with cell growth phase. Finally, we note that, when considering the effects of cell division, we also expect that it is necessary to account for the effects of asymmetries in the partitioning of cellular components, including RNA and proteins, as well as for the effects of cellular aging [38].

Asymmetry in the switching between noisy attractors may be of use, particularly given the physiologically relevant difference in the duration of the pulses required to switch between the noisy attractors. For example, the iron storage regulator Fur and srRNA RyhB are arranged in the SMDFL motif in several bacterial species [8]. Since Fur only represses RyhB transcription when $Fe^{2+}$ is present, we predict that the transition from the RyhB-high state (with no iron storage genes active) to the Fur-high state will require a relatively short time in an iron-rich environment. Conversely, it will take much longer to disable the iron storage genes when transitioning to an iron-deficient environment.

Finally, we note that the model employed here makes a number of simplifying assumptions, which may limit the applicability of the results. First, transcription and translation are assumed to take no time. These processes introduce delays, which can be non-negligible in the dynamics of a switch [39]. These delays are expected to be longer in eukaryotes, where several additional processes such as pre-mRNA processing and nuclear export must take place to produce the TFs and repress the target [40]. However, these delays have been shown to generally have smaller effects on the dynamics of a switch than the delay caused by the open complex formation at the promoter [39], which was modeled here in the less noisy promoter model. We thus believe that the results are reasonably applicable to eukaryotes and to prokaryotes in stationary phase.

## ACKNOWLEDGMENTS

Work was supported by TUT President's Doctoral Programme (J.L.-P.), and Academy of Finland (A.S.R.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

[1] A. S. Flynt and E. C. Lai, Nat. Rev. Genet. **9**, 831 (2008).
[2] G. Storz, S. Altuvia, and K. M. Wassarman, Annu. Rev. Biochem. **74**, 199 (2005).
[3] S. Gottesman, Trends Genet. **21**, 399 (2005).
[4] E. Levine, M. Huang, Y. Huang, T. Kuhlman, H. Shi, Z. Zhang, and T. Hwa [Proc. Natl. Acad. Sci. USA (to be published)].
[5] S. Mukherji, M. S. Ebert, G. X. Y. Zheng, J. S. Tsang, P. A. Sharp, and A. van Oudenaarden, Nat. Genet. **43**, 854 (2011).
[6] Y. Shimoni, G. Friedlander, G. Hetzroni, G. Niv, S. Altuvia, O. Biham, and H. Margalit, Mol. Syst. Biol. **3**, 138 (2007).
[7] P. Zhou, S. Cai, Z. Liu, and R. Wang, Phys. Rev. E **85**, 041916 (2012).
[8] B. Večerek, I. Moll, and U. Bläsi, EMBO J. **26**, 965 (2007).
[9] F. Fazi, A. Rosa, A. Fatica, V. Gelmetti, M. L. De Marchis, C. Nervi, and I. Bozzoni, Cell **123**, 819 (2005).
[10] X. Li and R. W. Carthew, Cell **123**, 1267 (2005).
[11] C. P. Bracken, P. A. Gregory, N. Kolesnikoff, A. G. Bert, J. Wang, M. F. Shannon, and G. J. Goodall, Cancer Res. **68**, 7846 (2008).
[12] A. H. Juan, R. M. Kumar, J. G. Marx, R. A. Young, and V. Sartorelli, Mol. Cell **36**, 61 (2009).
[13] A. Arkin, J. Ross, and H. H. McAdams, Genetics **149**, 1633 (1998).
[14] U. Alon, Nat. Rev. Genet. **8**, 450 (2007).
[15] L. Wang, B. L. Walker, S. Iannaccone, D. Bhatt, P. J. Kennedy, and W. T. Tse, Proc. Natl. Acad. Sci. USA **106**, 6638 (2009).
[16] S. Huang, Y.-P. Guo, G. May, and T. Enver, Dev. Biol. **305**, 695 (2007).
[17] M. Acar, J. T. Mettetal, and A. van Oudenaarden, Nat. Genet. **40**, 471 (2008).
[18] J. Elf, J. Paulsson, O. G. Berg, and M. Ehrenberg, Biophys. J. **84**, 154 (2003).
[19] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox, Cell **123**, 1025 (2005).
[20] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, Science (NY) **329**, 533 (2010).
[21] A.-B. Muthukrishnan, M. Kandhavelu, J. Lloyd-Price, F. Kudasov, S. Chowdhury, O. Yli-Harja, and A. S. Ribeiro, Nucleic Acids Res. **40**, 8472 (2012).
[22] M. Kandhavelu, J. Lloyd-Price, A. Gupta, A.-B. Muthukrishnan, O. Yli-Harja, and A. S. Ribeiro, FEBS Lett. **586**, 3870 (2012).
[23] Y. Hao, L. Xu, and H. Shi, J. Mol. Biol. **406**, 195 (2011).
[24] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.88.032714 for the analysis of the deterministic model, explanation of $\eta^2$ for the promoter kinetics, and supplemental figures.
[25] A. Lipshtat, A. Loinger, N. Q. Balaban, and O. Biham, Phys. Rev. Lett. **96**, 188101 (2006).
[26] H. Buc and W. R. McClure, Biochemistry **24**, 2712 (1985).
[27] M. Kandhavelu, H. Mannerström, A. Gupta, A. Häkkinen, J. Lloyd-Price, O. Yli-Harja, and A. S. Ribeiro, BMC Syst. Biol. **5**, 149 (2011).
[28] A. S. Ribeiro and S. A. Kauffman, J. Theor. Biol. **247**, 743 (2007).
[29] J. Lloyd-Price, A. Gupta, and A. S. Ribeiro, Bioinformatics **28**, 3004 (2012).
[30] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[31] M. Merika and S. H. Orkin, Mol. Cell. Biol. **13**, 3999 (1993).
[32] A. K. Vershon, S. Liao, W. R. McClure, and R. T. Sauer, J. Mol. Biol. **195**, 311 (1987).
[33] T. Møller, T. Franch, P. Højrup, D. R. Keene, H. P. Bächinger, R. G. Brennan, and P. Valentin-Hansen, Mol. Cell **9**, 23 (2002).
[34] J. A. Bernstein, A. B. Khodursky, S. Lin-Chao, and S. N. Cohen, Proc. Natl. Acad. Sci. USA **99**, 9697 (2002).
[35] K. L. Larrabee, J. O. Phillips, G. J. Williams, and A. R. Larrabee, J. Biol. Chem. **255**, 4125 (1980).
[36] R. D. Mosteller, R. V. Goldstein, and K. R. Nishimoto, J. Biol. Chem. **255**, 2524 (1980).
[37] S. Göttesman, Annu. Rev. Genet. **30**, 465 (1996).
[38] E. J. Stewart, R. Madden, G. Paul, and F. Taddei, PLoS Biol. **3**, e45 (2005).
[39] A. S. Ribeiro, A. Häkkinen, H. Mannerström, J. Lloyd-Price, and O. Yli-Harja, Phys. Rev. E **81**, 011912 (2010).

[40] M. J. Moore and N. J. Proudfoot, Cell **136**, 688 (2009).

[41] I. Moll, RNA **9**, 1308 (2003).

[42] E. Levine, Z. Zhang, T. Kuhlman, and T. Hwa, PLoS Biol. **5**, e229 (2007).

[43] J. Yu, J. Xiao, X. Ren, K. Lao, and X. S. Xie, Science (NY) **311**, 1600 (2006).

[44] M. Dunaway, J. S. Olson, J. M. Rosenberg, O. B. Kallai, R. E. Dickerson, and K. S. Matthews, J. Biol. Chem. **255**, 10115 (1980).

# Supplement to “Bistability in a stochastic RNA-mediated gene network”

Jason Lloyd-Price and Andre S. Ribeiro

## Deterministic Model

Here, we present the analysis of the deterministic counterpart to the stochastic model presented in the main text. The ODEs describing the mean dynamics of reactions (1)-(9) are:

$$\frac{d[\mathrm{Pro}_s]}{dt} = \overbrace{k_u(1-[\mathrm{Pro}_s])}^{(9)} - \overbrace{k_r[\mathrm{Pro}_s][\mathrm{P}]}^{(8)} \tag{S1}$$

$$\frac{d[\mathrm{R}_s]}{dt} = \overbrace{\alpha_s[\mathrm{Pro}_s]}^{(5)} - \overbrace{\beta_s[\mathrm{R}_s]}^{(6)} - \overbrace{\gamma[\mathrm{R}_m][\mathrm{R}_s]}^{(7)} \tag{S2}$$

$$\frac{d[\mathrm{R}_m]}{dt} = \overbrace{\alpha_m}^{(1)} - \overbrace{\beta_m[\mathrm{R}_m]}^{(2)} - \overbrace{\gamma[\mathrm{R}_m][\mathrm{R}_s]}^{(7)} \tag{S3}$$

$$\frac{d[\mathrm{P}]}{dt} = \overbrace{\alpha_p[\mathrm{R}_m]}^{(3)} - \overbrace{\beta_p[\mathrm{P}]}^{(4)} \tag{S4}$$

The reaction from which each term originates is shown above the term. These equations correspond to equations (1)-(4) of [1]. Setting the right hand sides of equations (S1)-(S4) to zero gives the following constraint on the concentration of R$_m$ at equilibrium in terms of the kinetic constants of the model:

$$0 = (\alpha_m - \beta_m[\mathrm{R}_m])(\beta_m k_u + k_r\alpha_p[\mathrm{R}_m])(\beta_s + \gamma[\mathrm{R}_m]) - \gamma\alpha_s\beta_p k_u[\mathrm{R}_m] \tag{S5}$$

Rewriting (S5) in terms of the dynamic parameters introduced in the text (see Table 1), and writing $\tilde{\gamma} = \gamma\theta$ and $x = [\mathrm{R}_m]/\theta$, we get:

$$0 = (1-x)(1+Rx)(\beta_s + \tilde{\gamma}x) - \tilde{\gamma}\lambda x \tag{S6}$$

This is a cubic polynomial whose roots correspond to the values of $[\mathrm{R}_m]/\theta$ at equilibrium. When (S6) has three positive real roots, the system has three equilibria (two stable and one unstable), and is therefore bistable. Note that the system size parameter $\theta$ does not affect the number of roots of (S6), and thus does not affect bistability. Parameter sets which produce three positive real roots must satisfy the following conditions [1,2]:

$$\begin{gathered} b<0,\ c>0,\ d<0, \\ 27d^2 + 4b^3d - 8bcd - b^2c^2 + 4c^3 < 0 \end{gathered} \tag{S7}$$

where $b = (\tilde{\gamma}R - \beta_s R - \tilde{\gamma})/a$, $c = (\beta_s R + \tilde{\gamma} - \beta_s - \tilde{\gamma}\lambda)/a$, $d = \beta_s/a$, and $a = -\tilde{\gamma}R$. Deterministic regions of bistability were found by numerically evaluating the constraints in (S7).

The system size parameter, as shown above, does not affect whether the deterministic solution is bistable or not. However, the region of robust bistability found in the main text does not exactly correspond to the region of bistability of the deterministic model. The differences arise due to low copy number effects. To show this, we recreated Fig. 2A, except with $\theta$ = 100, shown in Fig. S1. In this limit, the regions of bistability of the stochastic and deterministic models overlap considerably.

## RNA Production Kinetics

The three RNA production models used differ in the variability of the time intervals between production events, which we quantify with the squared coefficient of variation ($\eta^2$, defined as the variance over the squared mean). Under certain assumptions, it can be shown that the expected amount of noise in the RNA or protein timeseries, in terms of either the Fano factor or $\eta^2$, is monotonic with the $\eta^2$ of the production interval distribution (see eq. (1) of [3]).

The simplest production model is described by a single reaction (reaction (1) for mRNA production and (5) for srRNA production), which results in exponentially-distributed intervals between events, giving a $\eta^2$ = 1. This model produces a Poisson-distributed number of RNA molecules in a fixed time window, and is thus called ‘Poissonian’.

In the following sections, the random variable $T$ follows the distribution of intervals between productions for a specific promoter model, and has mean $\mu_T$ and variance $\sigma_T^2$.

*Super-Poissonian initiation kinetics*

Interval distributions resulting in RNA counts with higher variance than Poissonian were generated using a bursting promoter (reactions (10)-(12)). This promoter has long, exponentially-distributed OFF periods with mean duration $\mu_{\text{OFF}}$, with short ON periods during which it produces RNAs. During this ON period, there is a chance $p$ that an RNA is produced before turning OFF again, producing geometrically-sized bursts of new RNAs. This is achieved by setting the kinetic constant to produce $k_t = pk_{\text{OFF}}/(1-p)$, where $k_{\text{OFF}}$ is the inverse of the mean burst duration.

For the purposes of this analysis, we make the simplifying assumption that ON periods take very little time compared to the OFF periods, and are thus negligible. To get the $\eta^2$ of the interval distribution, we first make the distinction between productive and non-productive ON periods. An ON period has a chance of turning OFF without producing any RNA with probability $1-p$. Since OFF periods are exponentially distributed, the time between productive ON periods is also exponentially-distributed with mean $\mu_{\text{OFF}}/(1-p)$. Further, the probability of producing $k$ RNAs in an ON period, and therefore producing $k-1$ intervals of length 0, is $p^k(1-p)$, making $p$ of the intervals non-zero. The first and second moments of $T$ can then be written as:

$$\mathbf{E}T = p\mu_{\text{OFF}}$$
$$\mathbf{E}T^2 = 2p\mu_{\text{OFF}}^2$$

Giving the $\eta^2$ of $T$ as:

$$\frac{\sigma_T^2}{\mu_T^2}=\frac{\mathbf{E}T^2-(\mathbf{E}T)^2}{(\mathbf{E}T)^2}=\frac{2}{p}-1$$

We therefore set $p=2/(\eta^2+1)$ to achieve a given $\eta^2$.

*Sub-Poissonian initiation kinetics*

Reactions (13)-(15) describe a series of $N$ steps to be completed by the RNA polymerase before an RNA can be produced. This reduces the variability of the inter-RNA interval distribution, and thus the distribution of the number of RNA molecules produced in a fixed time window has less variance than a Poisson distribution with the same mean. The $\eta^2$ of the sum of $N$ independent exponentially-distributed variables, each with mean $\mu_X$, is given by:

$$\frac{\sigma_T^2}{\mu_T^2}=\frac{N\mu_X^2}{(N\mu_X)^2}=\frac{1}{N}$$

We therefore set $N=1/\eta^2$ to achieve a given $\eta^2$.

## References

[1] P. Zhou, S. Cai, Z. Liu, and R. Wang, Physical Review E **85**, 041916 (2012).

[2] B. D. Aguda and B. L. Clarke, The Journal of Chemical Physics **87**, 3461 (1987).

[3] J. M. Pedraza and J. Paulsson, Science (New York, N.Y.) **319**, 339 (2008).

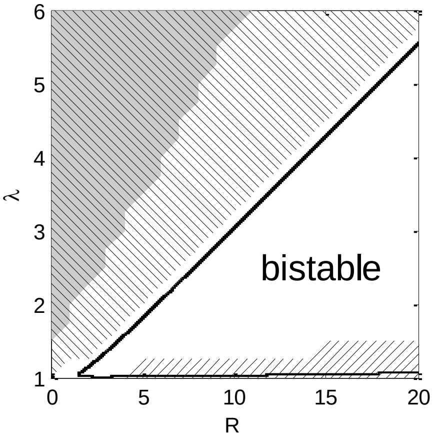


**Figure S1:** Bifurcation diagram with *λ* and *R* as control parameters and *θ* = 100. Hatched areas indicate where a noisy attractor is not robustly stable (i.e. stable for less than 90% of one month of simulation time, on average). Upwards and downwards hatching indicates the srRNA-high and TF-high noisy attractors are unstable, respectively. Shaded areas indicate that the unstable noisy attractor is less stable than 5% of one month of simulation time (i.e. the switch is monostable/unstable). Data is from 500 runs per tested parameter pair and initial state.

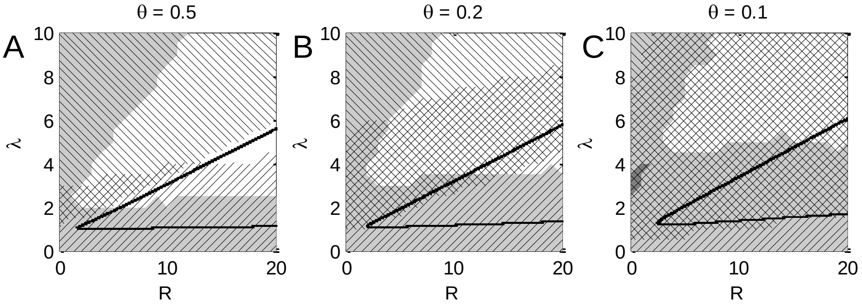


**Figure S2:** Recreation of Figure 3 from the main text, with $\gamma = 0.1$.

# Publication III

Jason Lloyd-Price, Maria Lehtivaara, Meenakshisundaram Kandhavelu, Sharif Chowdhury, Anantha-Barathi Muthukrishnan, Olli Yli-Harja, and Andre S. Ribeiro "Probabilistic RNA partitioning generates transient increases in normalized variance in RNA numbers in synchronized populations of *Escherichia coli*", *Molecular Biosystems*

Molecular BioSystems

Dynamic Article Links

Cite this: *Mol. BioSyst.*, 2012, **8**, 565–571

www.rsc.org/molecularbiosystems

**PAPER**

# Probabilistic RNA partitioning generates transient increases in the normalized variance of RNA numbers in synchronized populations of *Escherichia coli*†

**Jason Lloyd-Price,[a] Maria Lehtivaara,[a] Meenakshisundaram Kandhavelu,[a] Sharif Chowdhury,[a] Anantha-Barathi Muthukrishnan,[a] Olli Yli-Harja[ab] and Andre S. Ribeiro*[a]**

***Received 11th March 2011, Accepted 13th October 2011***
**DOI: 10.1039/c1mb05100h**

We explore the effects of probabilistic RNA partitioning during cell division on the normalized variance of RNA numbers across generations of bacterial populations. We first characterize these effects in model cell populations, where gene expression is modeled as a delayed stochastic process, as a function of the synchrony in cell division, the rate of division, and the RNA degradation rate. We further explore the additional variance that arises if the partitioning is biased. Next, in *Escherichia coli* cells expressing RNA tagged with MS2d–GFP, we measured the normalized variance of RNA numbers across several generations, with cell divisions synchronized by heat shock. We show that synchronized cell populations exhibit transient increases in normalized variance following cell divisions, as predicted by the model, which are not observed in unsynchronized populations. We conclude that errors in partitioning of RNA molecules generate diversity between the offspring of individual bacteria and thus constitute a form of reproductive bet-hedging.

## Introduction

Phenotypic diversity aids bacterial populations in coping with environmental fluctuations.[1] Evidence suggests that the diversity of a monoclonal cell population can change under different environmental conditions.[2,3] Noise in gene expression is a major source of this diversity since it generates cell to cell variability in RNA and protein numbers.[4] To some extent, this noise is sequence dependent, as it varies from gene to gene.[5]

There are other sources of cell to cell diversity in RNA and protein numbers. A recent work[6] mathematically demonstrated that stochastic partitioning of RNA and proteins in cell division contributes to the variance of these molecules in a population. However, it is still uncertain if the partitioning of these molecules is subject to any means of internal control. Experimental verification of their results[6] may shed light on this question, but is not yet available.

The extent to which deviations from a perfectly equal partitioning of RNA and protein molecules can affect the diversity of their numbers in a cell population depends on several parameters including, but not limited to, the degradation rate of the RNA, the generation time of the cells, and the level of synchronization of the cell cycles. Some of these parameters may be sequence dependent, causing the effects to differ between RNAs.

Here, we study the cell to cell diversity in RNA numbers in populations of dividing *Escherichia coli* cells, as a function of the variables listed above. An additional factor may be relevant in generating diversity during cell division. Internally, *E. coli* cells are not spatially homogeneous. For example, different plasmids preferentially localize in different regions of the cell,[7] and RNA molecules do not diffuse considerably from their point of transcription, leading to the spatial organization of translation and RNA decay within the cell.[8] Additional factors may contribute to asymmetry in divisions.[9] We thus also consider the possibility that the RNA partitioning, while probabilistic, may be biased.

We address the following question: to what extent does the probabilistic nature of the RNA partitioning affect the variance of RNA numbers of populations of dividing cells?

Using a delayed stochastic model of the expression dynamics of the $P_{lac/ara}$ promoter, we first characterize the normalized variance in RNA numbers that can realistically arise from stochastic partitioning during cell division as a function of the RNA degradation rate, the mean division time, the level of synchrony, and the strength of the bias in partitioning.

[a] *Computational Systems Biology Research Group, Tampere University of Technology, P.O. Box 553, 33101 Tampere, Finland. E-mail: andre.ribeiro@tut.fi*
[b] *Institute for Systems Biology, 1441N 34th St, Seattle, WA 98103-8904, USA*

† Electronic supplementary information (ESI) available: Supplementary document describing the model. See DOI: 10.1039/c1mb05100h

Next, we measure absolute RNA quantities *in vivo* in DH5α-PRO *E. coli* cells that have been synchronized by heat shock. We compare with measurements from a cell population not subject to heat shock and with the predictions from the model.

## Methods

### Cells, plasmids and chemicals

The method of RNA detection and quantification was proposed by Fusco *et al.*[10] and characterized in *E. coli* by Golding and Cox.[11] The *E. coli* strain DH5α-PRO (identical to DH5α-Z1) contains two constructs. The first is the bacterial expression vector PROTET-K133 carrying a single chain MS2 dimer (MS2d) fused with green fluorescent protein (MS2d–GFP). This vector has a promoter, $P_{LtetO-1}$, inducible by anhydrotetracycline (aTc; IBA GmbH, Göttingen, Germany). The second construct is a pIG–BAC ($P_{lac/ara}$–mRFP1–MS2-96bs) vector, a bacterial artificial chromosome based on F factor replication with an array of 96 MS2d binding sites under the control of the $P_{lac/ara}$ promoter. The constructs were generously provided by Dr Ido Golding, University of Illinois, USA. The mRNA target is inducible by isopropyl β-D-1-thiogalactopyranoside (IPTG) (Fermentas, Finland) and/or L-arabinose (Sigma-Aldrich, Schnelldorf, Germany).

### *In vivo* measurements of tagged RNA molecules in *E. coli*

Cells were grown overnight at 37 °C in LB supplemented by the appropriate antibiotics. The next day, cells were diluted in fresh medium plus antibiotics. To induce production of MS2d–GFP, 100 ng $mL^{-1}$ aTc and 0.1% L-arabinose were added to the diluted bacterial culture. Cells were then incubated with these inducers at 37 °C with shaking for 45 min to a final optical density (OD-600 nm) of ∼0.4. Afterwards, expression of the target RNA was induced by 1 mM IPTG. For imaging, 8 μL of culture were placed on a microscopic slide between a cover slip and 0.8% LB-agarose gel pad set at specific points in time after induction by IPTG. Epifluorescence microscopy was used to minimize the risk of not detecting spots. Measurements were done with a B-2A filter (EX 450-490, DM 505, BA 520), Nikon DS-Fi1 camera and NIS-Elements F software (version 2.20, Nikon Corp).

Cells were imaged 30, 45, 60, 75, 90, 105 and 120 minutes after induction. Cells were taken from the liquid culture at these moments and immediately placed under the microscope and imaged over a period of ∼10 minutes. In this way, we aimed to measure the RNA numbers and spatial distributions in individual cells as if they were in liquid culture.

### Synchronizing cell divisions by heat shock

In one experiment, cell division times were synchronized by heat shock as described by Lomnitzer and Ron.[12] The cells were grown overnight as described above, and then subjected to 45 °C for 15 min prior to induction by IPTG.

Division times of heat shocked synchronous cells were determined by the OD of the liquid culture. The OD was measured from at least two samples every ten minutes from 30 to 120 minutes after addition of the inducers. The OD was then averaged over the samples. Dilutions were used so that the OD remained smaller than 0.4. From OD measurements it is possible to estimate the mean division time.[13]

### Detecting cells and quantifying tagged RNA molecules from the images

We detected cells from raw images using the method proposed by Wang *et al.*[14] This method divides a grayscale image into three classes: background, cell border and cell region. It then exploits an iterative cell segmentation process that identifies and segments clumped cells based on the size and edge information (Fig. 1). Cell detection performance degrades if several cells are clumped together. This can be avoided by a threshold based on the cell size and discarding the "cells" whose size is beyond the threshold.

After detecting the cells, we detect RNA molecules tagged with MS2d–GFP. We segment the RNA spots with the kernel density estimation method for spot detection.[15] This method estimates the probability density function over the image from local information. The method processes an image *f* by filtering it with a desired kernel as follows:

$$\hat{f}(i,j) = \frac{1}{\mathrm{card}(C(i,j))h} \sum_{(k,l)\in C(i,j)} K\left(\frac{f(i,j) - f(k,l)}{h}\right) \quad (1)$$

where $h$ is the smoothing parameter or bandwidth, $(k, l)$ represents pixel location inside the kernel, card is the cardinality of the set, and $K(u)$ is the kernel. We used a Gaussian kernel and applied Otsu's thresholding method[16] to segment RNA spots from the kernel density estimated image, highlighting the spots.

Finally, the number of RNA molecules in each spot is quantified by assuming that the first peak in the distribution of intensities of many RNA spots from cells on the same slide corresponds to individual RNA molecules.[17] Subsequent peaks in the distribution of intensities correspond to spots of multiple RNA molecules. A sample intensity distribution is shown in Fig. S5 (ESI†).

The MS2 binding sites of each RNA molecule may not be saturated at all times. This is likely one of the sources of variance of each peak (along with the movements of the tagged RNA molecules along the *z*-axis, for example). Nevertheless, these sources of noise are not strong enough to prevent a clear distinction between individual peaks (see Fig. S5, ESI†), and thus it does not significantly affect the accuracy of the quantification of the number of mRNAs in each spot.

Another possible source of error in the RNA quantification method would be the occurrence of recombination events that would lower the number of MS2 binding site repeats.

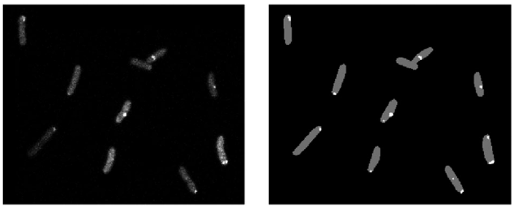


**Fig. 1** Unprocessed image of tagged RNA molecules in *E. coli* cells from fluorescence microscopy (left) and the corresponding segmented image with the detected cells (grey) and RNA spots (white) (right).

The variability in the number of binding sites in different cells used for the measurements was tested by amplifying the target gene from several colonies after many divisions, and determining its length by electrophoresis. For this, we designed primers to amplify the region containing MS2 binding sites (ESI†). The results (Fig. S9, ESI†) show that there is no significant variation between samples and within samples, as the bands are all in the same region and the width of each band is equal to the width of the bands of the ladder. Given the length of each MS2 binding site (∼45 nucleotides long, see supplementary material from ref. 11), a significant diversity in the number of binding sites ought to be detectable. This allows us to conclude that errors in quantifying RNA numbers due to variability in light intensity arising from errors in recombination are negligible in our measurements.

The copy number and partitioning of F-plasmids are stringently controlled by internal cellular mechanisms. However, a fraction of inaccurate F-plasmid distributions have been reported. For example, in the case of the tetO-array derivative (pDAG480) during growth at 20 °C in LB medium, 5.7 ± 1.0% per generation (standard error of three measurements) are usually lost.[18] In the case of ΔsopC mini-F (pDAG115), this percentage is 4.8 ± 0.6%.[18] In both cases, ∼5% of the cells in the next generation did not contain the plasmid.[18] Following these measurements, we considered 5% of all cells as outliers in each experiment, since this quantity is in good agreement with the observed outliers in the number of RNAs per cell.

It is noted that the MS2d–GFP tagging proteins are expressed from a strong promoter ($P_{\text{LtetO-1}}$) on a high-copy number plasmid (PROTET-K133). Within the duration of our measurements we observed that there was always enough MS2d–GFP in the cells to properly detect all target RNA molecules. This was assessed by measuring the background fluorescence over time (Section S5, ESI†). This also shows that if partitioning errors in the MS2d–GFP exist, they are negligible for our measurements. Our assessment is in agreement with previous reports.[11]

### Modelling populations of cells with stochastic gene expression and probabilistic RNA partitioning during cell division

The delayed stochastic modelling strategy of gene expression[19] accounts for the stochasticity of the chemical interactions and the duration of complex steps in gene expression. It was shown to match gene expression dynamics at the single molecule level.[20] We use it to model the expression of the target RNA.

In *E. coli* cells DH5α-PRO, transcription is regulated by a repressor (LacI) and inducers (IPTG and arabinose).[21] We designed a delayed stochastic model[19] of the gene expressing the target RNA (see ESI†). The model explicitly represents the promoter and the binding/unbinding of the activators and repressors that modify the probability that an RNAP will bind to the promoter and initiate transcription. The model includes the effects of the promoter open complex formation,[22] and the time required for the polymerase to produce the final RNA. The simulation of multiple cells, subject to divisions, is implemented in CellLine.[23]

In perfectly synchronized cell populations we imposed that all cells divided simultaneously, while in asynchronous populations the time until the first division was scaled by a uniform random number in the range [0,1) (half-open interval including 0, but excluding 1). Subsequent divisions occur at regular intervals, as determined by the division time. When a division occurs, all RNA molecules in the mother cell are partitioned between the daughter cells by generating a random number $N_1$ following the binomial distribution $B(N,p)$, where $N$ is the number of RNA molecules in the mother cell and $p$ is the partitioning bias. One daughter cell inherits $N_1$ RNA molecules, while the other inherits $N - N_1$.

## Results

### Behaviour of the model

The effect of probabilistic RNA partitioning on the cell-to-cell diversity in RNA numbers in a population depends on several factors. It depends on the rate by which RNA molecules degrade, which ''dissipates'' the effects of errors in partitioning between daughter cells. It depends on the mean lifetime of the cells since each division can contribute to the diversity in RNA numbers of the population.[6] The degree of synchrony of cell divisions affects the evolution of the diversity in RNA numbers in the population. Finally, we consider the effects of having a bias in the probabilistic partitioning towards one of the daughter cells.

The simulations allow us to explore what diversity in RNA numbers is achievable by varying these four parameters in realistic ranges (present methods to detect individual RNA molecules *in vivo* do not allow some of these parameters to be varied, such as the degradation rate of the target RNA). To quantify cell to cell diversity in RNA numbers, we use the normalized variance ($CV^2$, variance over the mean squared) of the RNA numbers in each cell.[6]

The model of transcription was tuned so that the mean number of target RNAs at 60 min matched the measured RNA production. From measurements of 1000 cells, 1 hour after full induction, we observed that each cell contained, on average, 3.36 tagged RNA molecules (data not shown).

Interestingly, we also observed that tagged RNA molecules tended to be located in the first or third quarter of the cell (Fig. 1; Fig. S6, ESI†). They were also distributed asymmetrically along the major axes of the cells. The absolute difference between the numbers of RNA molecules in each side was consistent with a binomial distribution with $p$ equal to 0.85. Finally, in *E. coli*, mean division time and mean RNA lifetimes vary, respectively, from 20 to 60 min and from 3 to 20 min.[24]

Given the above, we model populations of dividing cells and measure the $CV^2$ of RNA numbers 60 min after induction, as we vary each of the aforementioned parameters within realistic intervals. Unless stated otherwise, we set the RNA degradation rate to 0.1 $\text{min}^{-1}$, the division time to 30 min, the bias in RNA partitioning to 0.85, and asynchronous divisions (Table 1, case 1).

In each row of Table 1, besides the parameter values, there is the mean number of RNA molecules per cell and the $CV^2$ of RNA numbers at 60 min. The results of each row are from model populations with 10 000 cells per generation, simulated for 60 min. From Table 1, the $CV^2$ of RNA numbers 60 min

**Table 1** Effects of RNA probabilistic partitioning in division, for different mRNA degradation rates, division times, degrees of synchrony in division, and partitioning biases, on the mean and $CV^2$ of RNA numbers of model cell populations. Each case represents a different set of model parameters. Parameters that are blank in the table were set to their value in case 1

| Case | RNA degradation rate/$min^{-1}$ | Division time/min | Synchrony | Partitioning bias | Mean no. RNAs at 60 min | $CV^2$ at 60 min |
|---|---|---|---|---|---|---|
| 1 | 1/10 | 30 | Async | 0.85 | 3.36 | 0.29 |
| 2 | | | Sync | | 1.93 | 0.94 |
| 3 | | | | 0.5 | 3.35 | 0.27 |
| 4 | | | | 1 | 3.36 | 0.33 |
| 5 | 1/20 | | | | 5.61 | 0.23 |
| 6 | 1/3 | | | | 1.15 | 0.77 |
| 7 | | 20 | | | 3.07 | 0.33 |
| 8 | | 60 | | | 3.70 | 0.25 |
| 9 | 1/20 | 60 | | 0.5 | 6.53 | 0.15 |
| 10 | 1/3 | 20 | Sync | 1 | 0.59 | 2.42 |

after induction of the model cell populations varies widely as the parameters are varied. In some cases this change does not imply significant changes in the mean RNA level.

Case 2 (synchronous divisions) of the table has a notably higher $CV^2$ than case 1 (asynchronous divisions). This is because the measurement of $CV^2$ occurs immediately after a synchronized division in case 2. A more detailed temporal analysis shows this.

In Fig. 2, we plot the $CV^2$ of RNA numbers over time for the cells in cases 1 and 2 (until the second synchronized division of cells of case 2). For comparison, we also plot the $CV^2$ of RNA numbers over time for cells with synchronous divisions but unbiased RNA partitioning. The results show that synchronized divisions generate transient increases in $CV^2$ of RNA numbers that can be strongly enhanced if there is a biased partitioning of RNA molecules at cell division. Asynchronous divisions cause the $CV^2$ to not fluctuate with time and to be slightly higher than the $CV^2$ of synchronous populations prior to a division. Comparing cases 3 and 4 (Table 1), we observe that, in the absence of synchronization in divisions, the stronger the bias in RNA partitioning when cells divide, the higher is the $CV^2$.

The rate of RNA degradation has an interesting effect. Decreasing it leads to an increase in mean RNA levels, which decreases $CV^2$. However, if the RNA lifetime is of the same order of magnitude as that of the cell lifetime, $CV^2$ can "accumulate" across generations. To show this, in Fig. 3, we plot the value of $CV^2$ over time, for synchronous divisions, of three cell populations differing in RNA degradation rates.

Two of the rates are within realistic intervals,[24] while one rate (900 $min^{-1}$) is set abnormally weak so as to ease the visualization of the effect (note the inversions between the $CV^2$ of the three populations following the divisions). Such an abnormally weak rate of degradation could be accomplished artificially by, *e.g.*, coating the RNA with viral coat proteins.[25]

From Fig. 3 it is obvious that decreasing cell division time has two effects. First, there are more frequent partitions of RNA molecules, thus contributing more frequently to $CV^2$. When the mean division time is small enough to be of the same order of magnitude as the mean RNA lifetime, then cumulative effects in $CV^2$ appear (increasing $CV^2$ from one generation to the next). This is shown in Fig. 3, for the two populations with slower RNA degradation rates.

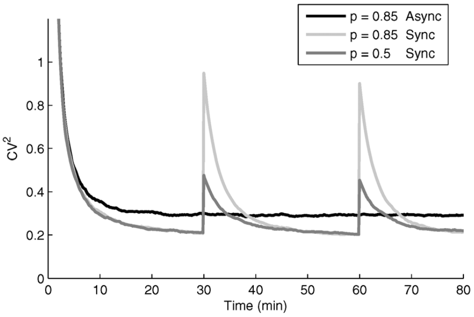


**Fig. 2** $CV^2$ of RNA numbers over time of three cell populations (cases 1 and 2 in Table 1) and a population similar to case 2 but with unbiased partitioning (not in Table 1). Population 1 has synchronous divisions and biased partitioning, population 2 has asynchronous divisions and biased partitioning, and population 3 has synchronous divisions and unbiased partitioning. In all cases the mean division time is 30 min. Data are from 10 000 model cells per generation of each population.

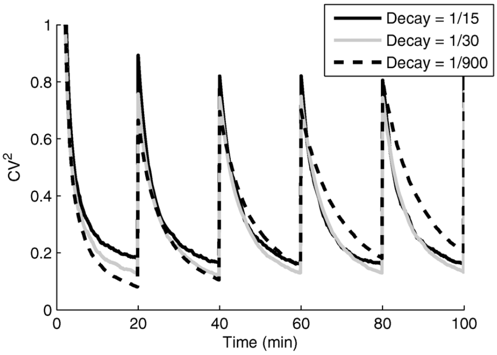


**Fig. 3** $CV^2$ of RNA numbers over time of three cell populations that differ in RNA degradation rates (in $min^{-1}$).

Finally, we investigated the range of the effects of probabilistic RNA partitioning on the $CV^2$ of RNA numbers of a population. Within realistic intervals, case 9 has all parameters set so as to minimize $CV^2$, while in case 10, all parameters are set to maximize the $CV^2$. These cases show that a dynamic range of ~16 fold is achievable by varying these parameters within realistic intervals.

### Experimental validation of the model

To verify if the probabilistic RNA partitioning at cell division, under specific conditions, such as when divisions are synchronous,

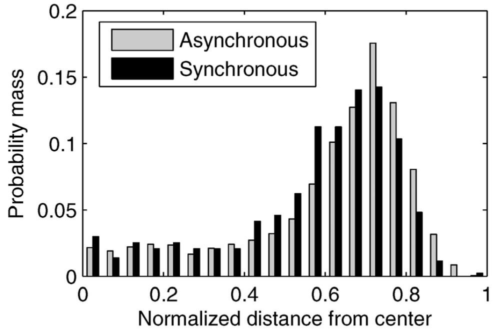


**Fig. 4** Distribution of the normalized distance of RNA locations from the center of the cells from the synchronous (heat shocked, data from 435 RNA spots in 490 cells) experiment and the asynchronous experiment (data from 1989 RNA spots in 1756 cells).

causes an observable effect on real cell populations, we imaged *E. coli* cells from a population whose cell divisions were synchronized by heat shock (see Methods). These cells express RNA target for MS2d–GFP, allowing the detection of individual RNA molecules.[17] The $CV^2$ of RNA numbers was measured every 15 minutes from multiple cells. Results are shown in Fig. 5 (circles). For comparison, we repeated the experiment without the heat shock (crosses in Fig. 5). Each data point was obtained by observing on average 140 cells.

From OD measurements (data not shown) we calculated the mean division time to be approximately 51.5 minutes in the first two divisions in the synchronous population. This long division time is likely due to the high metabolic costs of the processes induced by aTc, IPTG, and arabinose.[11] We verified the mean division time and the degree of synchrony by recording division times of cells under the microscope for two hours following the heat shock (data in Fig. S7, ESI†). Asynchronous cells were found to have a similar mean division time.

To verify that the synchronization by heat shock does not introduce abnormal RNA distributions, we measured the RNA positions along the major axes of the cells. The distributions for the synchronous and asynchronous cases are shown in Fig. 4. The distributions are identical, indicating that the heat shock does not significantly affect the RNA localization in the cells. Also, we verified that there are no significant changes in the RNA spatial distribution over time (data not shown).

Finally, we also verified that the absolute difference between the numbers of RNA molecules on each side of the cells is similar in the synchronous and asynchronous cases. We found them to be consistent with binomial distributions with $p = 0.92$ and 0.87 for the synchronous and asynchronous cases, respectively.

The normalized variance in RNA numbers of the synchronized cell population varies over time (Fig. 5), strongly increasing at specific time points, in particular between 45 and 60 min, and between 90 and 105 min. These increases occur where they would be expected if there are synchronized divisions roughly every 50 min. These transient increases would not be possible if cell divisions were not synchronized or if the equal partitioning of RNA molecules was perfect.

To verify that this trend in $CV^2$ is enhanced by the biased RNA partitioning, we compared the distributions of RNA numbers in the cell population at 45 min and at 60 min after the heat shock (Fig. 6). Between these two time points, an event (synchronized divisions) occurred that reshaped the distribution, from unimodal to bimodal. The cells were partitioned into two subpopulations: one with low numbers of RNA molecules, the other with high numbers.

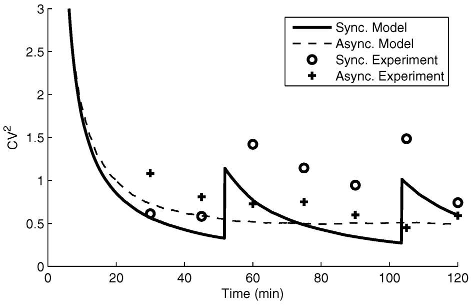


**Fig. 5** $CV^2$ of the number of RNA molecules in cells and model cells at various moments. Cell divisions in two populations were synchronized by heat shock. Cell divisions occur every 51.5 min. Clear increases in $CV^2$ after each division are visible for cells synchronized by heat shock (circles) and corresponding model cells (solid line). These transient increases are absent in cells not subjected to heat shock (crosses) and corresponding model cells (dashed line).

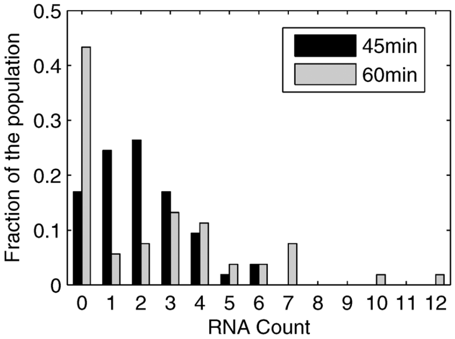


**Fig. 6** Fractions of cells with a given number of RNA molecules at 45 min and 60 min after a heat shock. Cell divisions occurred at 51 min. The effects of biased RNA partitioning at that event are visible in the distribution of RNA numbers at 60 min. The distributions are from the same dataset as Fig. 5 (circles). Each distribution was built from 53 cells.

To determine the degree of agreement between model and measurements, we modelled two cell populations under the same conditions as the two measurements. Divisions are set to occur at the same rate as in the measurements (every 51.5 min). RNA molecules are not subject to degradation since, in the experimental setting, RNAs are virtually immortalized by the tagging molecules.[17] We simulated two populations of 10 000 model cells for 3 generations. The $CV^2$ of RNA numbers of the two model cell populations are shown in Fig. 5. The model cell populations differ in that in one case the divisions are synchronous (solid line), while in the other they are fully asynchronous (dashed line). As in the measurements, the variation of $CV^2$ over time differs significantly. Namely, in the synchronous cell populations, the $CV^2$ of RNA numbers varies over time, sharply increasing after each division event. This is then followed by a decrease until the next synchronized division event occurs. These model cells behave identically to those whose dynamics is shown in Fig. 2, except that in the former, the RNA does not degrade.

Comparing the two measurements with their respective models, we find that the normalized variance in RNA numbers in the model cells accurately follows the trend of the measurements in both cases. Models and measurements only differ in that the $CV^2$ is slightly lower in the model cells, both in the synchronous and the asynchronous cases. This is expected given that the measurements have additional sources of diversity such as cell death, errors in counting of RNA molecules, and variability in the amount of inducers absorbed by each cell.

## Conclusions and discussion

We studied the normalized variance of RNA numbers in model cell populations of *E. coli* across generations. In addition to the inherent stochasticity in RNA transcription and degradation, the diversity is found to depend on the probabilistic nature of RNA partitioning. The contribution from this source was found to depend on the mean cell division time, the degree of synchrony in cell division, and it can be further enhanced if there is bias in the partitioning. RNA degradation limits the extent to which the probabilistic partitioning affects the overall diversity in RNA numbers, as it tends to reduce the effects of errors in partitioning.

These results confirm a previous study[6] on the relevance of the probabilistic nature of RNA partitioning at cell division on the overall diversity of RNA numbers of a bacterial population. We add to this study in that the numerical simulations of the realistic models of gene expression and dynamics of cell divisions allowed us to quantify the dynamic range (~16 fold) of normalized variance that is realistically obtainable by populations of *E. coli* cells. By inducing synchrony alone, we found from the simulations that the variation should be on the order of 3-fold, which is expected to be detectable in measurements of cell-to-cell diversity in RNA numbers, *e.g.*, using the MS2d–GFP system for tagging RNA molecules.[11]

We further extend the results of Huh and Paulsson[6] by showing with the simulations that synchronous divisions allow transient increases in normalized variance, which can be enhanced by any bias in the partitioning. The amplitude of these transient increases is also tunable in a wide range, with the parameters described above within realistic ranges.

We provide experimental verification of this result. Our measurements show that the degree of synchrony in cell division is non-negligible and causes transient increases in the normalized variance of RNA numbers, irrespective of the existence of a bias in partitioning, which can enhance the amplitude of the transient increase. Comparing the distributions in RNA numbers of the synchronous cell population, before and after the division, allows the contribution of the partitioning events to the variance in RNA numbers to be estimated.

Cell synchrony can be induced by many types of stress such as starvation, and can be stably maintained through several generations.[26] Interestingly, cell-to-cell variability is likely to be most advantageous under stress conditions,[27] and it was under these conditions where we observed the strong transient increases in the normalized variance of RNA numbers. In our case, the amplitude of these transient increases was enhanced by the observed bias in RNA partitioning.

We note that the comparison between the distributions of RNA numbers before and after divisions in the synchronous cell population shows that the bias in partitioning observed in our measurements is a phenomenon that occurred in most cells, rather than a rare event, as it caused the population to split into two subpopulations, distinct in the number of RNA molecules in each cell. This suggests that we are observing a phenomenon of biased partitioning towards one of the daughter cells, or a phenomenon of unbiased partitioning with high variance (*e.g.* if the RNA molecules clumped together and the clumps were partitioned binomially).

Either case could be explained by the mechanism described by Lindner and colleagues[28] of an asymmetric strategy, whereby dividing cells segregate damage at the expense of aging individuals. Since the tagged RNA complexes are not naturally occurring in *E. coli*,[11] it may be that the cell recognizes these complexes as an undesirable substance. If so, it is possible that we are observing, for the first time at the single molecule level, a mechanism by which dividing cells accumulate unwanted substances at the older poles.

Our results cannot be used to show that RNA is partitioned in a biased fashion in *E. coli*, and it was not our intention to do so since the spatial kinetics of the target RNA in our measurements are likely altered by the tagging. Our goal here was to quantitatively investigate the effects of probabilistic partitioning, whether biased or not, on cell to cell diversity in RNA numbers in a cell population over time. The experimental results confirmed the model's predictions of the effects of synchrony and bias in partitioning, which is independent of whether the bias is, in this case, natural or artificial.

The ability of *E. coli* to spatially organize RNA molecules was recently demonstrated by Llopis and colleagues.[8] This organization may cause some RNAs to be partitioned in a biased fashion. Our study provides the means to predict the consequences of such partitioning mechanisms in cell to cell diversity in RNA numbers.

In any case, our measurements are an example of the ability of bacteria to spatially organize macromolecules (RNA tagged with MS2d–GFP). Further, we show that this process is discriminative, since MS2d–GFP is homogeneously distributed in the absence of the target RNA. From an evolutionary point of view, what is relevant is that errors[6] and biases in partitioning of RNA molecules generate diversity between the offspring of individual bacteria. It thus constitutes a form of reproductive bet-hedging.

## Acknowledgements

This work was supported by Academy of Finland and Finnish Funding Agency for Technology and Innovation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors thank Daniel Cloud for valuable comments, as well as Eero Lihavainen, Antti Häkkinen, and Shannon Healy.

## Notes and references

1 E. Kussel and S. Leibler, *Science*, 2005, **309**, 2075–2078.

2 R. Losick and C. Desplan, *Science*, 2008, **320**, 65–68.

3 J.-W. Veening, W. K. Smits and O. P. Kuipers, *Annu. Rev. Microbiol.*, 2008, **62**, 193–210.

4 A. Arkin, J. Ross and H. McAdams, *Genetics*, 1998, **149**, 1633–1648.

5 Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili and X. S. Xie, *Science*, 2010, **329**, 533–538.

6 D. Huh and J. Paulsson, *Nat. Genet.*, 2010, **43**, 95–100.

7 T. Quoc, Z. Zhong, S. Aung and J. Pogliano, *EMBO J.*, 2002, **21**, 1864–1872.

8 P. M. Llopis, A. Jackson, O. Sliusarenko, I. Surovtsev, J. Heinritz, T. Emonet and C. Jacobs-Wagner, *Nature*, 2010, **466**, 77–81.

9 E. J. Stewart, R. Madden, G. Paul and F. Taddei, *PLoS Biol.*, 2005, **3**, e45.

10 D. Fusco, N. Accornero, B. Lavoie, S. M. Shenoy, J. M. Blanchard, R. H. Singer and E. Bertrand, *Curr. Biol.*, 2003, **13**, 161–167.

11 I. Golding and E. C. Cox, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 11310–11315.

12 R. Lomnitzer and E. Ron, *J. Bacteriol.*, 1972, **109**, 1316–1318.

13 E. O. Powell, *J. Gen. Microbiol.*, 1956, **15**, 492–511.

14 Q. Wang, J. Niemi, C. M. Tan, L. You and M. West, *Cytometry, Part A*, 2010, **77**, 101–110.

15 T. B. Chen, H. Lu, Y. S. Lee and H. J. Lan, *J. Biomed. Inf.*, 2008, **41**, 1021–1027.

16 N. Otsu, *IEEE Trans. Syst. Man Cybern.*, 1979, **9**, 62–66.

17 I. Golding, J. Paulsson, S. M. Zawilski and E. C. Cox, *Cell*, 2005, **123**, 1025–1036.

18 G. S. Gordon, J. Rech, D. Lane and A. Wright, *Mol. Microbiol.*, 2004, **51**, 461–469.

19 A. S. Ribeiro, R. Zhu and S. A. Kauffman, *J. Comput. Biol.*, 2006, **13**, 1630–1639.

20 R. Zhu, A. S. Ribeiro, D. Salahub and S. A. Kauffman, *J. Theor. Biol.*, 2007, **246**, 725–745.

21 R. Lutz and H. Bujard, *Nucleic Acids Res.*, 1997, **25**, 1203–1210.

22 W. R. McClure, *Annu. Rev. Biochem.*, 1985, **54**, 171–204.

23 A. S. Ribeiro, D. Charlebois and J. Lloyd-Price, *Bioinformatics*, 2007, **23**, 777–779.

24 J. Bernstein, A. Khodursky, P. Lin, S. Lin-Chao and S. Cohen, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 9697–9702.

25 D. S. Peabody, *EMBO J.*, 1993, **12**, 595–600.

26 R. G. Cutler and J. E. Evans, *J. Bacteriol.*, 1966, **91**, 469–476.

27 M. Kirschner and J. Gerhart, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 8420–8427.

28 A. B. Lindner, R. Madden, A. Demarez, E. J. Stewart and F. Taddei, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 3076–3081.

# Supplementary material for "Probabilistic RNA partitioning generates transient increases in the normalized variance of RNA numbers in synchronized populations of *Escherichia coli*"

Jason Lloyd-Price, Maria Lehtivaara, Meenakshisundaram Kandhavelu, Sharif Chowdhury, Anantha-Barathi Muthukrishnan, Olli Yli-Harja, and Andre S. Ribeiro

## Table of Contents

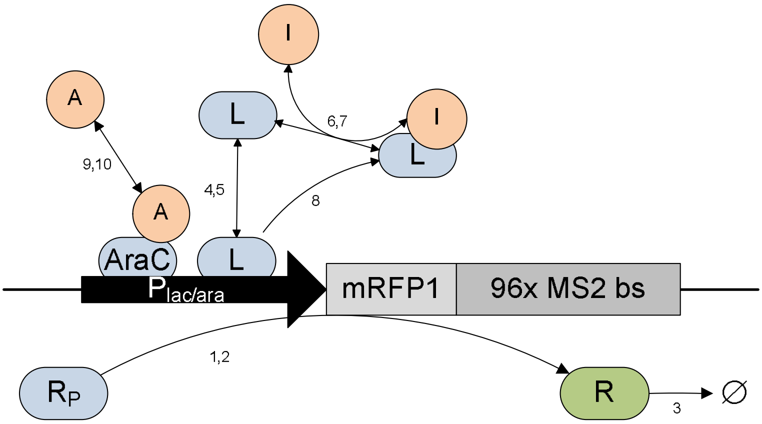
## 1. Introduction

This document describes the delayed stochastic model of the dynamics of the $P_{lac/ara}$ promoter (Golding and Cox, 2004) as well as the tuning procedure to match the mean expression levels measured experimentally. Our model follows the delayed stochastic modeling strategy proposed in (Ribeiro et al, 2006), and is implemented in SGNSim (Ribeiro and Lloyd-Price, 2007). An initial model of this genetic system at the single cell level is first presented where the relevant chemical components are represented explicitly (section 2). From that complete model, we then introduce a reduced version of the model by approximating some components in order to facilitate to tuning and prediction, without affecting the relevant dynamics of the system (section 3). The complete set of reactions of the reduced model is presented in section 3.4, while the values used for the parameters are presented in section 4.5.

Afterwards, we present results concerning the spatial distribution of free MS2d-GFP molecules in live *E. coli* cells (section 5), and details on the method of quantification of mRNA target for MS2d-GFP in life cells (section 6). Finally, we present an example image of cells taken by epifluorescence microscopy (section 7) and supporting information regarding the degree of synchronization of cell division following heat shock (section 8).

## 2. Explicit Model

The chemical components and interactions are depicted in Supplementary Fig. 1. The production of RNA molecules (and thus fluorescent spots in the cells) is accomplished when an RNA polymerase binds to the promoter region and transcribes the DNA into RNA. Transcription is regulated by a repressor protein (LacI) and inducers (IPTG and Arabinose). IPTG can bind to LacI, causing a conformational change in LacI which results in the protein falling off the promoter, allowing transcription to occur (Lutz and Bujard, 1997). Another protein (AraC) can also bind to the promoter. This protein does not modify the transcription initiation rate on its own, but when bound to Arabinose, it increases the affinity of the RNA polymerases to the promoter region, thereby promoting transcription. GFP molecules are not explicitly represented since they always exist in sufficient amounts so as not to be a limiting factor of RNA detection.



**Supplementary Fig. 1**: Components and reactions in the model of the $P_{lac/ara}$ promoter and RNA production. Numbers indicate the corresponding reactions. Molecule labels are defined in section 2.1.

### 2.1. Notation

Hereafter, we use *P* to denote the promoter, $R_P$ to denote RNA polymerase, *R* to denote RNA, *L* to denote LacI, *I* to denote IPTG, and *A* to denote Arabinose in reactions and equations.

The model contains several reversible bimolecular reactions. In general, rates of unbinding are denoted by $k_{uXY}$, where X and Y denote the interacting species. Dissociation constants ($K_{dXY}$) are denoted similarly. For example, $K_{dLI}$ is the dissociation constant between LacI and IPTG.

The promoter can be in several 'states', depending on which substances are bound to it. In the following reactions, the promoter with LacI bound to it is denoted $P_L$, whereas $P_{L;-}$ denotes the promoter with no LacI bound. This notation also applies to Arabinose, where $P_A$ and $P_{A;-}$ denote the presence or absence of an Arabinose molecule bound to an AraC protein which is in turn bound to the promoter. For simplicity, we assume that an AraC is always bound to the promoter, given its small dissociation constant (on the order of $10^{-8}$ M (Timmes et al, 2004)). Similarly, $P_{L;-A}$ denotes a promoter with no LacI bound, but with Arabinose bound to it.

Following standard chemical notation, the number of molecules of a chemical species present in a cell is denoted $[X]$. For example, $[L]$ denotes the current number of LacI proteins in the cell. Additionally, we use $[X]_0$ to denote the number of a given molecule in the cell at time 0 (the start of the experiment).

## 2.2. Transcription

Transcription is initiated when an RNAP binds to the promoter region, forms the open complex, and begins elongating the RNA. This process is modeled in reaction (1) (Ribeiro et al, 2006).

$$P_{\bar{L}\bar{A}} + R_P \xrightarrow{k_t} P_{\bar{L}\bar{A}}(\tau_P) + R_P(\tau_R) + R(\tau_R) \tag{1}$$

Here, $\tau_P$ and $\tau_R$ denote time delays used to model the time it takes for this highly complex, multi-step reaction to occur (Ribeiro et al, 2006). This notation denotes that, for example, the RNA ($R$) is fully transcribed and visible in the cell $\tau_R$ seconds after the transcription reaction began.

When Arabinose binds to the AraC molecule bound to the promoter, it induces transcription by actively recruiting RNA polymerases. This is modeled by a reaction which differs from reaction (2) in the value of its rate constant ($k_{tA}>k_t$).

$$P_{\bar{L}A} + R_P \xrightarrow{k_{tA}} P_{\bar{L}A}(\tau_P) + R_P(\tau_R) + R(\tau_R) \tag{2}$$

Normally, the RNA molecules are assumed to degrade via a first-order chemical reaction. However, the MS2-coated RNA molecules have been shown to have a considerably longer lifetime than normal RNA molecules, and cell division was shown to be the largest term in the (Golding et al, 2005).

## 2.3. Decay

The degradation of transcripts is modeled as a first-order process in reaction (3) (Ribeiro et al, 2006):

$$R \xrightarrow{k_d} \varnothing \tag{3}$$

## 2.4. Interactions of the promoter with LacI and IPTG

LacI's binding/unbinding to/from the promoter is modeled by reactions (4) and (5). When LacI is bound to the promoter, transcription cannot occur.

$$P_{\bar{L}} + L \xrightarrow{\frac{k_{uLP}}{K_{dLP}}} P_L \tag{4}$$

$$P_L \xrightarrow{k_{uLP}} P_{\bar{L}} + L \tag{5}$$

IPTG can bind to LacI (reactions (6) and (7)). The complex *IL* represents LacI with a different conformation that has much weaker affinity to the promoter than LacI. In reality, the stability of the bond between this complex and the promoter is much weaker. For simplicity, this is modeled by an immediate dissociation from the promoter (reaction (8)).

$$I + L \xrightarrow{\frac{k_{uLI}}{K_{dLI}}} IL \tag{6}$$

$$IL \xrightarrow{k_{uLI}} I + L \tag{7}$$

$$P_L + I \xrightarrow{\frac{k_{uLI}}{K_{dLI}}} P_{\bar{L}} + IL \tag{8}$$

### 2.5. Interactions of the promoter with AraC and Arabinose

Since an AraC is assumed to be bound at all times to the promoter, we only model the binding and unbinding of the Arabinose to AraC. This event changes the rate of transcription initiation. The binding and unbinding of Arabinose to AraC is modeled in reactions (9) and (10), respectively.

$$P_{\bar{A}} + A \xrightarrow{\frac{k_{uAP}}{K_{dAP}}} P_A \tag{9}$$

$$P_A \xrightarrow{k_{uAP}} P_{\bar{A}} + A \tag{10}$$

## 3. Reduced Model

The explicit model contains several parameters that are currently difficult to measure *in vivo*, and some reactions which, under normal conditions, do not significantly affect the dynamics of gene expression. To reduce the complexity of the model, we implement some approximations and justify why they are appropriate and do not compromise the realism of the simulation. Generally, these approximations consist of removing a reaction species which is in sufficient abundance to be considered constant.

### 3.1. RNA Polymerase

Under normal conditions, the amount of $R_P$ available for transcription events is approximately constant in an *E. coli* at all times (McClure, 1983). For that reason, instead of representing $R_P$ explicitly, the stochastic rates of reactions (1) and (2) can be multiplied by 20, the known quantity of free RNA polymerases per gene in *E. coli*.

### 3.2. Arabinose

Similar to $R_P$, when Arabinose is present in the cell, it is present sufficient amount so as to be assumed as constant. With 6.67 mM (1%) of Arabinose, and given the mean volume of an *E. coli* ($10^{-15}$ L, from http://redpoll.pharmacy.ualberta.ca/CCDB/cgi-bin/STAT_NEW.cgi), we estimate that the number of molecules of Arabinose in the cell is on the order of $10^{10}$. Reactions (9) and (10) can then be simplified to:

$$P_{\bar{A}} \xrightarrow{\frac{[A]k_{uAP}}{K_{dAP}}} P_A \qquad (11)$$

$$P_A \xrightarrow{k_{uAP}} P_{\bar{A}} \qquad (12)$$

This simplification has the additional important benefit that the Arabinose concentration and $K_{dAP}$ no longer need to be expressed in units of molecules per cell. Instead, both can be expressed as a ratio of 6.67 mM, decreasing the number of parameters required in the model (the cell volume is removed).

### 3.3. LacI/IPTG

Just as Arabinose and $R_P$ are present in the system in abundance, so are LacI and IPTG. These species can be removed with the same benefits as removing Arabinose, by the equilibrium point of reactions (6) and (7). The equilibrium values can be then used in simplified versions of reactions (4), (5) and (8). From reactions (6) and (7), this equilibrium point is reached when:

$$[I][L] = [IL]K_{dLI} \qquad (13)$$

Assuming that are no IPTG·LacI complexes initially present in the system, the amount of the complex can be written in terms of the initial concentrations of LacI (denoted by $[L]_0$). Similarly, the amount of IPTG can also be written as a function of LacI.

$$[IL] = [L]_0 - [L] \qquad (14)$$

$$[I] = [L] + [I]_0 - [L]_0 \qquad (15)$$

The solution for $[L]$ is:

$$[L] = \frac{-b \pm \sqrt{b^2 + 4K_{dLI}[L]_0}}{2}, b = [I]_0 - [L]_0 + K_{dLI} \qquad (16)$$

These calculations remove the need for reactions (4) and (5), and the equilibrium concentrations of $[L]$ and $[I]$ can be inserted into the reaction rates of reactions (6), (7) and (8).

This approximation yields two advantages. First, as with Arabinose, it is no longer necessary to translate IPTG and LacI concentrations and $K_{dLP}$ into molecules per cell. Second, we have sped up the simulation considerably by removing two high-frequency reactions, since the runtime of the SSA largely depends on the propensity of the reactions.

The validity of this approximation depends on the actual rate of $k_{uLI}$ (and therefore the time it takes to reach equilibrium), and the amounts of LacI and IPTG. In vitro studies have measured $k_{uLI}$ to be 0.2 $s^{-1}$ (Dunaway et al, 1980), implying that the system should be sufficiently close to equilibrium within one minute. The amount of LacI proteins in these cells has been measured to be on the order of 5000

(Lanzer and Bujard, 1988). Based on the average volume of an *E. coli* ($10^{-15}$ L, from http://redpoll.pharmacy.ualberta.ca/CCDB/cgi-bin/STAT_NEW.cgi), we estimate that the amount of molecules of IPTG in the cell when induced by 1mM of IPTG is on the order of $10^9$. All the conditions for the applicability of this approximation are therefore met by at least one order of magnitude.

### 3.4. The Final Model

The final delayed stochastic model consists of reactions (17) to (24):

$$P_{\bar{L}\bar{A}} \xrightarrow{[R_p]k_t} P_{\bar{L}\bar{A}}(\tau_P) + R(\tau_R) \tag{17}$$

$$P_{\bar{L}A} \xrightarrow{[R_p]k_{tA}} P_{\bar{L}A}(\tau_P) + R(\tau_R) \tag{18}$$

$$R \xrightarrow{k_d} \varnothing \tag{19}$$

$$P_{\bar{L}} \xrightarrow{\frac{[L]k_{uLP}}{K_{dLP}}} P_L \tag{20}$$

$$P_L \xrightarrow{k_{uLP}} P_{\bar{L}} \tag{21}$$

$$P_L \xrightarrow{\frac{[I]k_{uLI}}{K_{dLI}}} P_{\bar{L}} \tag{22}$$

$$P_{\bar{A}} \xrightarrow{\frac{[A]k_{uAP}}{K_{dAP}}} P_A \tag{23}$$

$$P_A \xrightarrow{k_{uAP}} P_{\bar{A}} \tag{24}$$

## 4. Constants and Tuning

### 4.1. Promoter Occupancy

Reaction pairs (20)-(21) and (23)-(24) determine the occupancy state of the promoter. It is useful in this case to precisely define the quantity ‘Promoter Occupancy’ as the expected fraction of time that the promoter is bound by LacI, or Arabinose (by the AraC/Arabinose complex). This fraction can be calculated from the propensities of the reactions as:

$$O(X, K_d) = \frac{X}{X + K_d} \tag{25}$$

In (25), $X$ denotes the concentration of the binding molecule, and $K_d$ is the dissociation constant. For example, $O([L]_0, K_{dLP})$ is the fraction of time that the lac repressor is bound to the promoter in the absence of IPTG.

### 4.2. Accounting for degradation and the promoter open complex formation duration ($\tau_P$)

To calculate the values of the transcription rate constants ($k_t$ and $k_{tA}$) necessary to obtain a given mean level of RNA after a transient time $t$, we first estimate the mean field behavior of a system producing RNA with rate $k_p$, which can degrade with rate $k_d$, without accounting for the effects of $\tau_P$ as a limiting factor of transcription initiation. The differential equation describing such mean field behavior is:

$$\frac{d[R]}{dt} = k_p - [R]k_d \tag{26}$$

Solving for $k_p$ yields:

$$k_p([R]) = \frac{[R]k_d}{1-e^{-tk_d}} \tag{27}$$

Finally, the values of the transcription rate constants ($k_t$ and $k_{tA}$) are not exactly equal to $k_p$ due to the effects of the promoter open complex formation. This effect can be accounted for by discounting the expected time that the promoter spends in this state, giving the mean rate at which the transcription reaction must occur to reach a mean RNA count of $[R]$ after $t$ seconds:

$$k([R]) = \frac{k_p([R])}{1-\tau_p k_p([R])} \tag{28}$$

We can now use (28) to calculate the necessary transcription rate to reach a given mean amount of RNA molecules after a transient.

### 4.3. Tuning $k_t$, $k_{tA}$, $K_{dLP}$

We now tune some of the free parameters (not yet experimentally measured) of the model to match the experimental results. The parameters $k_t$ and $k_{tA}$ determine the maximum possible RNA production rate, while $K_{dLP}$ determines how strongly LacI represses the system.

First, we deduce the production rate of RNA, according to reactions (17) and (18). The fraction of time that the promoter is, on average, not repressed by LacI (and thus is free for these reactions) is given by (1-$O([L],K_{dLP})$). Similarly, the fraction of time that the promoter is in state $P_A$ is given by $O([A],K_{dAP})$). The mean production rate of RNA is a mixture of the propensities of the two reactions that lead to RNA production, weighted by the fraction of times each is expected to occur:

$$k([R]) = [R_P](1-O([L],K_{dLP}))[k_t(1-O([A],K_{dAP}))+k_{tA}O([A],K_{dAP})] \tag{29}$$

We can now write formulas for the mean production rates of three cases: full repression of the promoter ($k_1 = k([R]_1)$, where $[R]_1$ was measured to be 0.532 RNAs), activation by Arabinose alone ($k_A = k([R]_A)$, where $[R]_A$ was measured to be 0.612 RNAs), and full activation ($k_M = k([R]_M)$, where $[R]_M$ was measured to be 3.36 RNAs).

$$k_1 = [R_P]k_t(1-O([L],K_{dLP})) \tag{30}$$

$$k_A = [R_P](1-O([L]_0,K_{dLP}))[k_t(1-O(1,K_{dAP}))+k_{tA}O(1,K_{dAP})] \tag{31}$$

$$k_M = [R_P][k_t(1-O(1,K_{dAP}))+k_{tA}O(1,K_{dAP})] \tag{32}$$

Solving the system of equations, we get:

$$k_t = \frac{k_M k_1}{[R_P]k_A} \tag{33}$$

$$K_{dLP} = \frac{[R_P]k_t[L]_0}{[R_P]k_t - k_1} - [L]_0 \tag{34}$$

$$k_{tA} = \frac{k_M - k_t(1 - O(1, K_{dAP}))}{[R_P]O(1, K_{dAP})} \tag{35}$$

### 4.4. Cell Division

The simulation of multiple cells, subject to cell division, is modeled by the CellLine simulator (Ribeiro et al, 2007), which can model cell division and impose desired distributions of partitioning of RNA molecules between the daughter cells at cell division. However, all calculations thus far have assumed that there is no cell division. To obtain the correct mean production rate with asynchronous division, the value of $k_d$ in section 4.2 is increased by adding $\frac{\ln 2}{g}$, where $g$ is the generation time.

### 4.5. Values of Constants and Parameters

| Constant | Value | Source |
|---|---|---|
| $[R_P]$ | 20 molecules | McClure (1983) |
| $t$ | 3600 – $E(\tau_R)$ = 3465.9 s | Ourselves |
| $\tau_P$ | 32* s | Lutz and Bujard (1997) |
| $\tau_R$ | $\tau_P$ + Γ(Gene Length, Elongation Rate$^{-1}$) s | |
| Gene Length | mRFP1 Length + 96 BS Length = 4287 bp | |
| mRFP1 Length | 654 bp | Zhang et al (2002) |
| 96 BS Length | 3633 bp | Golding and Cox (2004) |
| Elongation Rate | 42 bp·s$^{-1}$ | Gotta et al (1991) |
| $[L]_0$ | 5000 molecules = 8.3 × 10$^{-3}$ × 1x $[I]$ | Lanzer and Bujard (1988) |
| $k_{uLP}$ | 0.04 s$^{-1}$ | Dunaway et al (1980) |
| $K_{dLI}$ | 0.1† × 1x $[I]$ | Lutz and Bujard (1997) |
| $k_{uLI}$ | 0.2 s$^{-1}$ | Dunaway et al (1980) |
| $K_{dAP}$ | 0.1† × 1x $[A]$ | Lutz and Bujard (1997) |
| $k_{uAP}$ | 1.5 s$^{-1}$ | Miller et al (1983) |

**Supplementary Table 1:** Constants.

| Parameter | Value | Source/Section |
|---|---|---|
| $k_d$ | 0 s$^{-1}$ | Golding et al (2005) |
| $k_t$ | 4.191 × 10$^{-5}$ $[R_P]^{-1}$s$^{-1}$ | Section 4.3 |

* Since the transcription events are rare (a few per cell lifetime), we do not model the promoter delay as a distribution, but rather as a constant value. We observed no significant difference when $\tau_P$ followed a distribution with realistic variance (Lutz and Bujard, 1997).

† Approximated from the induction curve from Lutz and Bujard (1997)

| | | |
|---|---|---|
| $k_{tA}$ | $4.886 \times 10^{-5}\ [R_P]^{-1}s^{-1}$ | Section 4.3 |
| $K_{dLP}$ | $1.879 \times 10^{-3} \times 1x\ [I]$ | Section 4.3 |

**Supplementary Table 2:** Model tuning in Fig. 5.

| Parameter | Value | Source/Section |
|---|---|---|
| $k_d$ | $1/600\ s^{-1}$ | Bernstein (2002) |
| $g$ | 1800 s | Section 4.4 |
| $k_t$ | $3.727 \times 10^{-4}\ [R_P]^{-1}s^{-1}$ | Section 4.3 |
| $k_{tA}$ | $4.370 \times 10^{-4}\ [R_P]^{-1}s^{-1}$ | Section 4.3 |
| $K_{dLP}$ | $1.492 \times 10^{-3} \times 1x\ [I]$ | Section 4.3 |

**Supplementary Table 3:** Model tuning in Table 2 and subsequent figures.

## 5. Uniformity of the MS2d-GFP distribution

The MS2d-GFP tagging proteins are expressed from a strong promoter ($P_{\text{LtetO-1}}$) on a high-copy number plasmid (PROTET-K133). Within the duration of our measurements we observed that there was always enough MS2d-GFP in the cells to properly detect all target RNA molecules. This can be assessed by measuring the uniformity of the background fluorescence. This assessment further shows that if partitioning errors in the MS2d-GFP exist, they have negligible effects on the detection of target RNA in daughter cells.

The simplest way to make this assessment is to quantify the uniformity of the fluorescence background of cells with no target RNA molecules. We first establish a measure of uniformity based on the local spatial entropy of the fluorescence in the image each cell and then show that it is able to detect clumping and gradients in model cells with clumps and gradients. The measure is then applied to cells expressing MS2d-GFP without the target RNA to determine if the MS2d-GFP clumps and/or tends to be localized in any particular region of the cell.

### 5.1. Clumping and spatial distribution of MS2d-GFP in the cells

To determine if the MS2d-GFP molecules form clumps and/or tend to be co-localized in any particular region of the cell, we compute the local spatial entropy of the intensity of the pixels composing a bacterium, which allows us to quantify the degree of randomness of a set of variables (Shannon 1948). Here, we aim to show that all neighborhoods of pixels within the cell have similar distributions of fluorescence intensities among the pixels within each neighborhood, that is, that they have no detectable gradients or clumps of MS2d-GFP. The entropy $H_k$ in a neighborhood of $k$ pixels is defined as:

$$H_k = \sum_{\mathbf{y}\in\Omega^k} p(\mathbf{y})\log(p(\mathbf{y})) \tag{36}$$

where **y** is the vector of pixel intensities in the neighborhood, Ω is the domain of the elements of **y**, and $p(\mathbf{y})$ is its probability measure.

Entropy $H_1$ informs us of how diverse the intensities of the pixels composing the cell are, but not whether there are correlations between the intensities of neighboring pixels. If MS2d-GFP molecules clump or tend to be preferentially located in any particular region of the cell, these correlations ought to exist. If such correlations do not exist, then the entropy of the joint distribution of the intensities of neighboring pixels should equal $kH_1$ (since the pixel intensities are independent), otherwise, it is smaller than this value. The minimum possible (totally correlated pixel intensities) $H_k$ equals the entropy of a single pixel $H_1$. A simple way to quantify effects of possible spatial correlations or gradients in pixel intensities is then the ratio between $H_k$ and $H_1$. To have a normalized measure of correlation, we define $J_k$, ranging from $k^{-1}$ to 1, as:

$$J_k = \frac{H_k}{kH_1} \tag{37}$$

We measure the two-pixel neighborhood entropy $H_2$ from both vertical and horizontal pairs of adjacent pixels. For each cell segmented from the images from the microscope, we first subtract the mean pixel intensity from each pixel, and scale the resulting intensities by the variance of the distribution of pixel intensities. The scaled pixel intensities were then binned into bins of size 0.2 per unit variance. The aforementioned scaling is required for the entropies of different cells to be comparable, due to the effects of the binning.

### 5.2. Generating spatial patterns and clumpiness in model cells

Model cells with various degrees of clumpiness are generated from cell shapes taken from real cells with the following algorithm. Let the *x* axis correspond to the major axis of the cells, and I(x,y) be the gray level of the pixel at (x,y). The algorithm proceeds as follows:

```
1. Set all I(x,y) = 0
2. Repeat N times:
    a. Select $x_c$, $y_c$ uniformly from the pixels in the real cell.
    b. Set $I(x_c, y_c) = I(x_c, y_c) + 1$
3. Convolve I with a Gaussian kernel with standard deviation $\sigma$
```

The degree of clumpiness of model cells is then determined by the choice of parameters N and σ. Larger N and/or larger σ produce less clumpiness. We use two sets of values of (N, σ), namely, (25, 1) and (100, 3), as these produce different, both detectable, degrees of clumpiness.

We also model cells with gradients, where the gradients aim to mimic what would be observed if the MS2d-GFP molecules were preferentially localized near, e.g., the cell poles or approximately along the cell border. Gradients are generated as follows:

$$I(x, y) = c(x - x_0)^p + (y - y_0)^p \tag{38}$$

where $x_0$ and $y_0$ are the coordinates of the cell center and, *p* determines the order of the gradient. To attain a linear gradient we set *p* to 1. For a quadratic gradient, we would set *p* to 2. By changing *c*, the eccentricity of the gradient can be varied.

### 5.3. Null model cells with no spatial correlations between pixel intensities

Null model cells without spatial correlations between pixel intensities can be generated by, for each pixel, generating an intensity value, drawn from the distribution of pixel intensities of a real cell.

However, a procedure is necessary prior to this, since real cells have two external sources of spatial correlations in the intensities of neighbor pixels, regardless of the existence or not of clumping or accumulation of MS2d-GFP molecules at any region of the cell. One source is the point spread function of the microscope. Since its effects cannot be removed, we expect slightly higher local correlations in real cells than in null model cells. The other source is the rod shape of the cells, whose effect can be accounted for in null model cells for proper comparison.

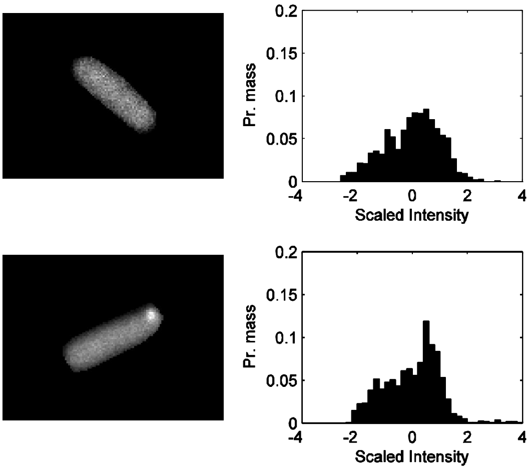
To generate null model cells from real ones that account for the shape of a cell, first, we remove the effect of the rod-shape from the pixels of a real cell by dividing by a scaling factor (described below). From the distribution of resulting pixel intensities, we generate the intensity of each pixel of the model cell. Next, we reintroduce the effect of the rod shape into the pixel intensities of the model cell by multiplying each pixel by the scaling factor. This allows null model cells to be generated that lack spatial correlations in pixel intensities except due their shape. The scaling factor to account for the rod shape of the cells for the pixel at (x,y) is:

$$s(x,y) = \sqrt{\frac{d_{\min}(x,y)}{\max(d_{\min})}} \qquad (39)$$

where $d_{min}(x,y)$ is the Euclidean distance to the nearest pixel that is not in the cell.

### 5.4. Clumping of MS2d-GFP

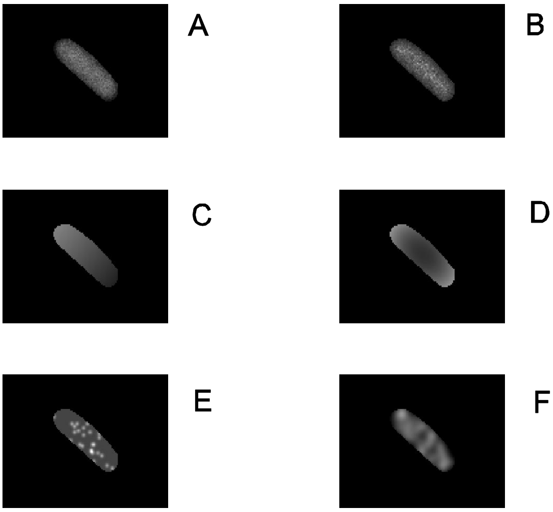
We now study whether there is a tendency of MS2d-GFP molecules to accumulate at any particular region of the cell or to clump in vivo. We start by performing a test on the method of detection of spatial correlations from images of cells taken by confocal microscopy. If the measure of $J_k$ of a cell is accurate enough to detect spatial correlations due to clumping of MS2d-GFP, its value should differ measurably between a cell with no target RNA and a cell with a target RNA, to which ~60 MS2d-GFP molecules are bound at any moment (Golding and Cox, 2004). Supplementary Fig. 2 shows the image of a cell with no visible RNA-MS2d-GFP spot (top left). Also shown is its distribution of scaled pixel intensities (top right). The image of a cell with one RNA-MS2d-GFP spot is also shown (bottom left) along with its distribution of scaled pixel intensities (bottom right).



**Supplementary Fig. 2**: A cell with no visible RNA-MS2d-GFP spot (top left) and its scaled pixel intensities (top right). Also shown is a cell with one RNA-MS2d-GFP spot (bottom left) and its distribution of scaled pixel intensities (bottom right).

Comparing the distributions in Supplementary Fig. 2, it is visible that the RNA-MS2d-GFP spot creates a small peak in the highest intensities in comparison to the cell with no spot. More spots would further increase the height of this peak and thus the difference between the two distributions. The values of $J_2$ are 0.91 for the cell with no spot and 0.84 for the cell with one spot. As shown below, and given that this measure varies from 0.5 to 1, this difference is significant, allowing the detection of MS2d-GFP clumps or gradients, if these exist (regardless of their origin).

In Supplementary Fig. 3 we show the image of a real cell (5A) and of a null model cell generated from this real cell (5B). Also shown are model cells, one with a linear gradient along the major axis (5C), and another with an eccentric quadratic gradient that results in stronger pixel intensities near the cell poles (5D). Two model cells with different degrees of artificial clumpiness and their distributions of pixel intensities are also shown (5E and 5F).



**Supplementary Fig. 3**. Images of a real cell (A) and of a null model cell (B) generated from the real cell. Also shown are model cells, one with a linear gradient along the major axis (C) and another with a quadratic gradient that creates stronger pixel intensities near the poles (D). Two model cells with different degrees of artificial clumpiness are also shown (E and F).

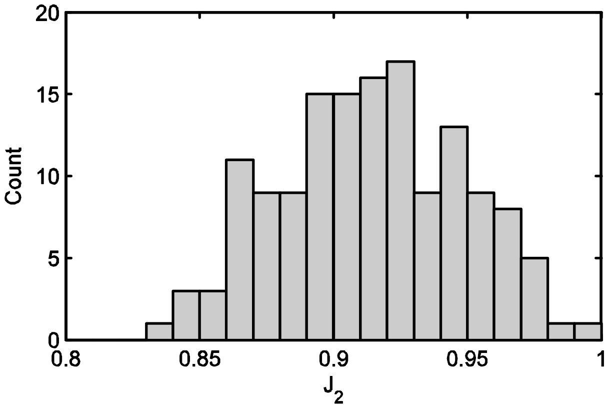
Each model cell is used as a null model to test for the existence of a type of pattern in the spatial distribution of MS2d-GFP molecules in real cells. The cells generated by randomly choosing the pixel intensities from the original distribution of pixel intensities are used to determine, by comparison, if there are local correlations between the intensities of neighbor pixels that are not detectable by eye. If no such local pixel intensity correlations exist, cells and model cells with random pixels intensities ought to have identical values of $J_k$.

The model cells with gradients are used as a null model for possible accumulation of MS2d-GFP molecules at any particular location of the cell. If such preferential locations were to exist, they would result in gradients of pixel intensities in the real cells and cause $J_k$ to be lower than in the cells with randomized pixel locations.

Finally, the model cells with artificial clumps are used as a null model for possible accumulation of MS2d-GFP molecules at certain locations in the cell. If these exist, they would cause $J_k$ to be lower than in the cells with randomized pixel locations. The values of $J_2$ of the cells A to F depicted in Supplementary Fig. 3 are shown in Supplementary Table 4. For this particular cell, it is possible to conclude that there are no spatial correlations between neighboring pixel intensities, as its value of $J_2$ is identical to that of the model cell with randomized pixel locations. Since $J_2$ is much higher than the $J_2$ of the model cells with artificial gradients and clumps, we can also conclude that this cell has no gradients or clumps.

| Cell | $H_2$ | $H_1$ | $J_2$ |
|---|---|---|---|
| Real Cell, no spot | 5.44 | 2.98 | 0.91 |
| Real Cell, one RNA-MS2d-GFP spot | 4.88 | 2.92 | 0.84 |
| Null model cell | 5.44 | 2.97 | 0.92 |
| Linear gradient | 3.58 | 2.92 | 0.61 |
| Quadratic gradient | 3.71 | 2.79 | 0.66 |
| Small artificial spots | 2.76 | 1.78 | 0.77 |
| Large artificial spots | 4.74 | 2.98 | 0.79 |

**Supplementary Table 4**: Values, for the cell and the null-model cells depicted in Supplementary Fig. 3, of their entropy in a two-pixel neighborhood ($H_2$), their entropy of individual pixels ($H_1$) and the value of $J_2$, a measure of spatial correlations, attained from the ratio between $H_2$ and $H_1$.



**Supplementary Fig. 4**: Distribution of $J_2$ for cells induced only with aTc (data is from 145 cells).

Supplementary Fig. 4 shows the distribution of $J_2$ values for 145 cells induced with only aTc for one hour. The mean value of $J_2$ of each of these cells is 0.914. We generated 10 models cells, from each of these 145 cells, with pixel intensities randomly drawn from the distribution of pixel intensities of the real cell. The mean value of $J_2$ of the 1450 model cells is 0.93, identical to that of the real cells. This demonstrates that there is no indication of accumulation of MS2d-GFP molecules at any particular region of the cells, or formation of clumps, one hour after induction by aTc.

Further note that the value of J2 for the cell with a target RNA (spot), shown in Table 2 is in fact smaller than more than 95% of the values in the distribution shown in Supplementary Fig. 4.

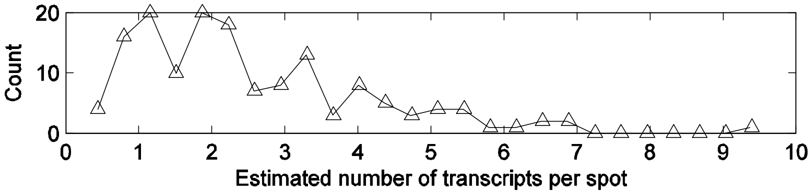
A final test can be made to further verify these conclusions. If MS2d-GFP molecules do not form clumps or accumulate at any region of the cells, then the value of $J_2$ of cells measured over a long period of time ought to be constant (aside from small stochastic fluctuations). We imaged cells at $t$ = 326 s, $t$ = 1791 s, and $t$ = 3591 s after placed under the microscope, and measured $J_2$ (Table 2). The results show that this quantity does not change significantly over time in any cell, further verifying that MS2d-GFP molecules neither tend to accumulate at any particular region of the cell, nor aggregate.

| Cell index | $J_2$ at $t$ = 326 s | $J_2$ at $t$ = 1791 s | $J_2$ at $t$ = 3591 s |
|---|---|---|---|
| 1 | 0.89 | 0.88 | 0.89 |
| 2 | 0.87 | 0.87 | 0.87 |
| 3 | 0.88 | 0.88 | 0.87 |
| 4 | 0.89 | 0.88 | 0.89 |
| 5 | 0.89 | 0.88 | 0.88 |
| 6 | 0.87 | 0.86 | 0.87 |
| 7 | 0.89 | 0.89 | 0.89 |

**Supplementary Table 5**: Values of $J_2$ of cells at $t$ = 326, $t$ = 1791, and $t$ = 3591 s.

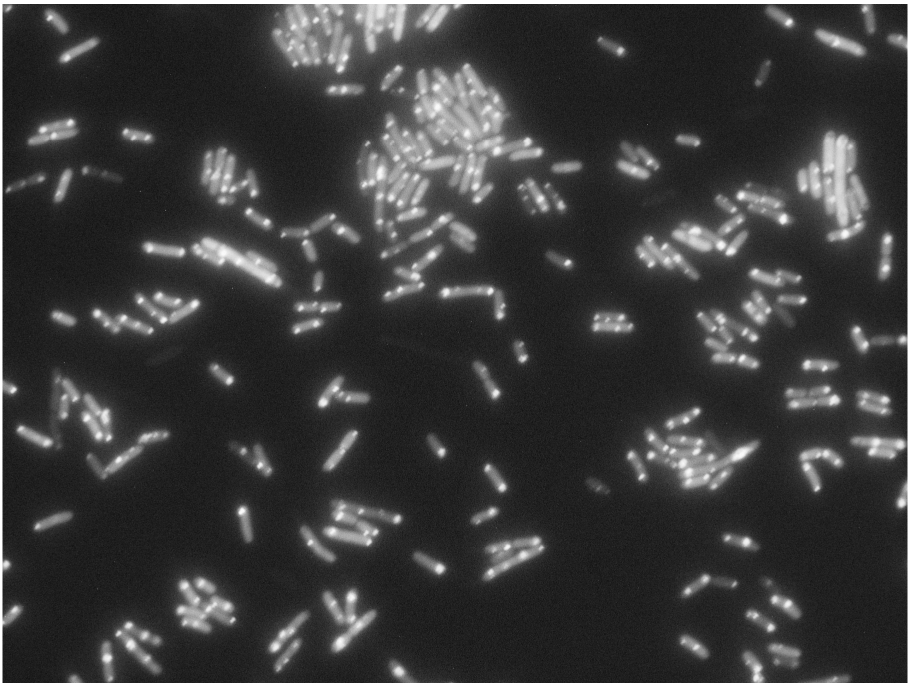
## 6. Quantification of mRNA in cells

The RNA quantification method used here was proposed in (Golding et al, 2005). The number of RNA molecules in each spot is quantified by assuming that the first peak in the distribution of intensities of many RNA spots from cells on the same slide corresponds to individual RNA molecules. The intensities are then normalized by the intensity of this peak to obtain the number of RNA molecules in each spot. This is possible due to the discrete nature of the peaks and the approximately uniformity of the distance between consecutive peaks. An example of such a distribution of intensities is shown in Supplementary Fig. 5.



**Supplementary Fig 5**: Example distribution of spot intensities obtained from a single slide, normalized by the mean intensity of the first peak in the distribution which corresponds to a single tagged RNA molecule.

## 7. Example image of cells expressing MS2d-GFP

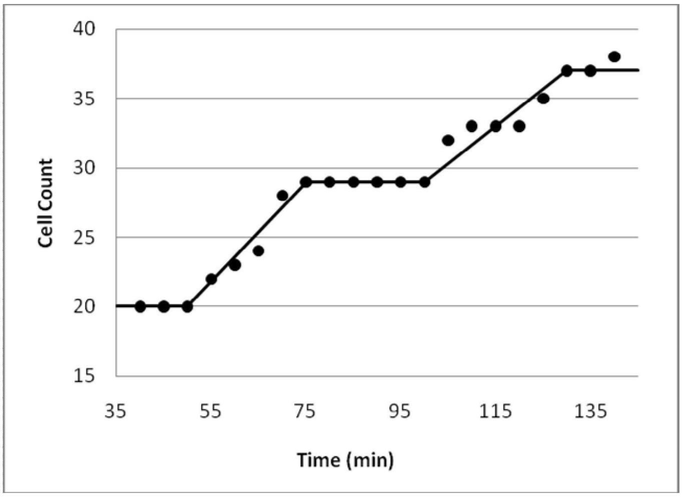


**Supplementary Fig 6**: Cells expressing MS2d-GFP and RNA target. Bright spots in each cell correspond to RNA molecules tagged with ~60-100 MS2d-GFP molecules. Cellular background is also fluorescent due to the freely diffusing MS2d-GFP molecules.

## 8. Assessing the degree of synchrony of cells following a heat shock

To determine the degree of synchrony in division of cells subject to heat shock, we redid the experiment as described in Materials and Methods, except that 30 minutes after induction by IPTG, 8 μL of culture was placed between a 1% LB agarose gel and a microscope cover slip. Starting from 40 minutes after induction, images of a set of cells were taken by DIC every 5 minutes for the following 100 minutes. Cells were held at 37$^{o}$C while under the microscope.

The number of cells visible at each point in time is shown in Supplementary Fig. 7. Two approximately synchronous divisions were observed (between 55 minutes and 75 minutes and between 105 minutes and 130 minutes), indicating that the heat shock successfully synchronized the divisions with the same efficiency as reported in (Lomnitzer and Ron, 1972). The divisions appear slightly later than the jumps in CV of RNA numbers observed in the synchronous experiment (see main document). This is expected because the imaging procedure for each time point of the synchronized experiment took approximately 10 minutes (this includes obtaining the cells from the liquid culture, placing them under the microscope, etc…). From the figure, it is observable that the fraction of cells that do divide under the microscopy at each generation is in line with this type of experiments (Lomnitzer and Ron, 1972)(Laskin and Lechevalier, 1984).



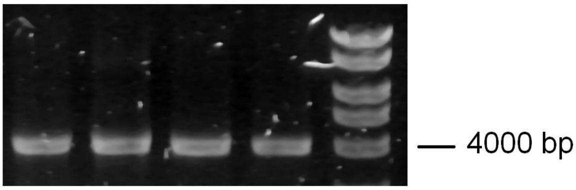
**Supplementary Fig 7**: Number of cells observed under the microscope over time. Dots are data points, the line shows the trend. Two approximately synchronous divisions can be seen.

## 9. PCR Analysis of mRFP1-96bs

To characterize the heterogeneity of the plasmids in the cells and to determine whether recombination errors cause the loss of binding sites or some region of the BAC clone (due to the tandem repeats of the binding sites), we have isolated several colonies from the transformant and purified the BAC clone after several cell divisions (overnight cell culturing at 37˚C at 250 rpm). These cells carried the BAC clone with the mRFP1-96bs target gene. To amplify the target gene, the plasmid was isolated and purified using a plasmid purification kit (Fermentas).

The target gene was amplified with Forward primer 5' GACGTCTGTGTGGAATTGTGAGCGG 3' and Reverse primer 5' ACGCGTTCGAAGCTTCGGACGCTA 3' (Thermo Scientific) from the purified BAC clone. Standard PCR protocol was used to amplify the target which was then run using 1% agarose gel electrophoresis, shown in Supplementary Fig. 8. The target gene is clearly visible at ≈ 4kb in each independent colony, in agreement with the expected length of the original target gene reported in (Golding and Cox, 2004).

From Supplementary Fig 8, it is possible to state that there is no significant variance within each sample since the width of each band is equal to the width of the bands of the ladder. Also, there are no significant differences between bands from different colonies. This indicates that our quantification of RNA numbers per cell is not affected by heterogeneity in the number of MS2-GFP binding sites of the target RNA in different cells, given the observed homogeneity of the results from the PCR analysis of the target.



**Supplementary Fig 8**: 1% agarose gel electrophoresis of the amplified target gene (mRFP1-96bs) from four different colonies (lanes 1-4), and a 10kb ladder (lane 5).

As a side note, we do not rule out the possibility that recombination events in the 96 binding site region may have occurred and lowered the number of binding sites of the original construct. What can be stated from the results here reported is that if this has occurred, it is a rare event, as it did not introduce diversity in the length of the target RNA of cells of the various colonies used in this study.

## References

Bernstein, J, Khodursky A, Lin P, Lin-Chao S, and Cohen S (2002) Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci USA* **99**: 9697–9702.

Dunaway M et al (1980) Kinetic Studies of Inducer Binding to *lac* Repressor-Operator Complex. *Journal of Biological Chemistry* **255**: 10115-10119.

Golding I, and Cox EC (2004) RNA dynamics in live *Escherichia coli* cells. *Proc Natl Acad Sci USA* **101**: 11310–11315.

Golding I, Paulsson J, Zawilski SM, and Cox EC (2005) Real-Time Kinetics of Gene Activity in Individual Bacteria. *Cell* **123**: 1025-1036.

Gotta SL, Miller OL, and French SL (1991) rRNA transcription rate in *Escherichia coli*. *Journal of Bacteriology* **173**: 6647-6649.

Lanzer M, Bujard H (1988) Promoters largely determine the efficiency of repressor action. *Proc Natl Acad Sci USA* **85**: 8973-8977.

Laskin AI and Lechevalier HA (1984) Handbook of Microbiology: Growth and Metabolism, Vol. 6. CRC Press.

Lomnitzer R and Ron E (1972) Synchronization of Cell Division in Escherichia coli by Elevated Temperatures: a Reinterpretation. *Journal of Bacteriology* **109**: 1316-1318.

Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/$I_1$-$I_2$ regulatory elements. *Nucleic Acids Research* **25**: 1203-1210.

Miller DM, Olson JS, Pfluhrath JW, Quiocho FA (1983) Rates of Ligand Binding to Periplasmic Proteins Involved in Bacterial Transport and Chemotaxis. *Journal of Biological Chemistry* **258**: 13665-13672.

McClure WR (1983) A biochemical analysis of the effect of RNA polymerase concentration on the in vivo control of RNA chain initiation frequency. In *Biochemistry of Metabolic Processes*, Lennon DLF, Stratman FW, and Zahlten RN (ed) pp. 207-217. New York: Elsevier Science Publishing Co. Inc.

Ribeiro AS, Zhu R, Kauffman SA (2006) General modeling strategy for gene regulatory networks with stochastic dynamics. *Journal of Computational Biology* **13**: 1630–1639.

Ribeiro AS, Lloyd-Price J (2007) SGN Sim, a Stochastic Genetic Networks Simulator. *Bioinformatics* **23**: 777-779.

Ribeiro AS, Charlebois DA, and Lloyd-Price J (2007) *CellLine*, a stochastic cell lineage simulator. *Bioinformatics* **23**: 3409-3411.

Timmes A, Rodgers M, and Schleif R (2004) Biochemical and Physiological Properties of the DNA Binding Domain of AraC Protein. *Journal of Molecular Biology* **340**: 731-738.

Zhang J, Campbell RE, Ting AY, and Tsien RY (2002) Creating new fluorescent probes for cell biology. *Nature Reviews Molecular Cell Biology* **3**: 906-918.

# Publication IV

Jason Lloyd-Price, Huy Tran, Andre S. Ribeiro "Dynamics of small genetic circuits subject to stochastic partitioning in division", *Journal of Theoretical Biology*

Journal of Theoretical Biology 356 (2014) 11–19

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

# Dynamics of small genetic circuits subject to stochastic partitioning in cell division

Jason Lloyd-Price, Huy Tran, Andre S. Ribeiro*

*Laboratory of Biosystem Dynamics, Computational Systems Biology Research Group, Department of Signal Processing, Tampere University of Technology, PO Box 527, FI-33101 Tampere, Finland*

## HIGHLIGHTS

- We study effects of partitioning errors on the dynamics of genetic circuits.
- Effects of partitioning errors differ widely with network topology and behavior.
- In switches, errors reduce the phenotype distribution's variance across generations.
- The synchrony of a population with clocks is robust to the majority of errors.
- Errors produce qualitatively different effects than noise in gene expression.

## ARTICLE INFO

*Article history:*
Received 21 October 2013
Received in revised form
20 February 2014
Accepted 15 April 2014
Available online 23 April 2014

*Keywords:*
Partitioning errors
Toggle switch
Repressilator
Cell-to-cell diversity
Synchrony

## ABSTRACT

In prokaryotes, partitioning errors during cell division are expected to be a non-negligible source of cell-to-cell diversity in protein numbers. Here, we make use of stochastic simulations to investigate how different degrees of partitioning errors in division affect the cell-to-cell diversity of the dynamics of two genetic circuits, a bistable switch and a clock. First, we find that on average, the stability of the switch decreases with increasing partitioning errors. Despite this, anti-correlations between sister cells, introduced by the partitioning errors, enhance the chances that one of them will remain in the mother cell's state in the next generation, even if the switch is unstable. This reduces the variance of the proportion of phenotypes across generations. In the genetic clock, we find that the robustness of the period decreases with increasing partitioning errors. Nevertheless, the population synchrony is remarkably robust to most errors, only significantly decreasing for the most extreme degree of errors. We conclude that errors in partitioning affect the dynamics of genetic circuits, but the effects are network-dependent and qualitatively different from noise in gene expression.

## 1. Introduction

Phenotypic diversity is a feature of all cell populations, including monoclonal ones, that significantly affects their survival chances, particularly in fluctuating environments (Kussell and Leibler, 2005; Samoilov et al., 2006). The stochastic nature of the biochemical reactions involved in the dynamics of gene regulatory networks is one well-known contributing source of phenotypic diversity (Kaern et al., 2005; McAdams and Arkin, 1999).

Recently, it has been recognized that the partitioning of plasmids, RNAs, proteins and other macromolecules during cell division can also be a non-negligible source of phenotypic diversity (Huh and Paulsson, 2011a, 2011b; Lloyd-Price et al., 2012). Similar to noise in gene expression, this source generates diversity that can propagate through reaction networks to high-copy number components, even in organisms with a morphologically symmetric division process, such as *Escherichia coli* (Huh and Paulsson, 2011a). After establishing a mathematical framework with which to characterize this source of noise (Huh and Paulsson, 2011b), it was shown that the random errors in partitioning result in cell-to-cell diversity in RNA and protein numbers that is difficult to distinguish from the diversity arising from gene expression noise, when observing cell populations at a single time moment (Huh and Paulsson, 2011a). Nevertheless, while noise in gene expression continuously generates diversity, noise from partitioning only occurs sparsely, when cells divide. Thus, the effects of these two sources should be readily distinguishable from a temporal perspective (Lloyd-Price et al., 2012). So far, it is unknown how these two sources of noise differ in regards to their effects on the dynamics of genetic circuits.

* Corresponding author. Tel.: +358331153928; fax: +358331154989.
*E-mail addresses:* jason.lloyd-price@tut.fi (J. Lloyd-Price), huy.tran@tut.fi (H. Tran), andre.ribeiro@tut.fi (A.S. Ribeiro).

http://dx.doi.org/10.1016/j.jtbi.2014.04.018
0022-5193/© 2014 Elsevier Ltd. All rights reserved.

Most cellular processes are regulated by small genetic networks, named motifs (Wolf and Arkin, 2003; Alon, 2007). It is conceivable that the noise in the process of partitioning of the products of gene expression affects the cell-to-cell diversity of behaviors of these motifs. Here, we study the effects of errors in partitioning on the behavior of two such motifs, the Toggle Switch (Gardner et al., 2000) and the Repressilator (Elowitz and Leibler, 2000). These two circuits differ widely in behavior. While the former is able to switch between two noisy attractors (Ribeiro et al., 2006; Ribeiro and Kauffman, 2007; Zhu et al., 2007), the latter only has one noisy attractor, a limit cycle (Elowitz and Leibler, 2000; Zhu et al., 2007; Loinger and Biham, 2007). Due to their dynamic properties, these circuits are likely candidates to serve as master regulators of future synthetic genetic circuits. Also, similar circuits have evolved in natural cells to perform similar tasks (Wolf and Arkin, 2002; Arkin et al., 1998; Lahav et al., 2004; Nelson et al., 2004). Thus, understanding the effects of random partitioning of RNA and proteins in cell division on the dynamics of these two circuits may aid in understanding how cells maintain robust phenotypes across generations.

The Toggle Switch is a two-gene motif, where each gene expresses a transcription factor that represses the expression of the other gene. As this circuit has two noisy attractors (Gardner et al., 2000; Arkin et al., 1998), it can store one bit of information. It can thus be used to make decisions (Arkin et al., 1998), or to store the results of one (Wolf and Arkin, 2003). The level of gene expression noise determines the frequency at which the Toggle Switch changes between its noisy attractors (Loinger et al., 2007; Potapov et al., 2011). A well-studied Toggle Switch is the "λ-switch", a decision circuit of the λ phage (Arkin et al., 1998), which determines whether an infecting phage will lyse the cell or, instead, integrate itself into the bacterial genome, forming a lysogen. The lytic cycle can be activated in lysogens either stochastically (Neubauer and Calef, 1970), or due to environmental cues such as irradiation by UV light (Baluch and Sussman, 1978). Meanwhile, the Repressilator is a synthetic three-gene motif which exhibits oscillatory behavior (Elowitz and Leibler, 2000), as each gene represses the next gene in the loop. In the Repressilator, gene expression noise determines the robustness of the period of oscillation (Häkkinen et al., 2013).

We study the effects of errors in partitioning on the behavior of these two circuits, focusing on their ability to 'hold state' (i.e. on the stability of their noisy attractors) across cell lineages, when subject to different partitioning schemes. Namely, we explore a wide range of magnitudes of partitioning errors, since in *E. coli* the process of partitioning of gene expression products ranges from highly symmetric (Di Ventura and Sourjik, 2011) to heavily asymmetric, e.g. due to spatially organized protein production (Montero Llopis et al., 2010). For this, we first examine the switching dynamics of the Toggle Switch in cell lineages. In this context, we further consider two biologically motivated scenarios: the phenotypic diversity in a continuous cell culture, and the population dynamics when one state of the switch is lethal to the cells, as in the case of λ lysogens. We then study the effects of errors in partitioning on the behavior of the Repressilator. Specifically, we study the robustness of the period of oscillations, and the rate of desynchronization across cell lineages of an initially synchronous population.

## 2. Methods

The models used here contain three main components. The first is the genetic circuit within each cell. The second is cell growth and division, and the last is the partitioning scheme of the proteins and RNA molecules in division. For simulations, we used the SGNS2 stochastic simulator (Lloyd-Price et al., 2012), which utilizes the Stochastic Simulation Algorithm (Gillespie, 1977).

### 2.1. Stochastic model of gene expression

The model of gene expression, illustrated in Fig. 1A, consists of the following set of reactions (Häkkinen et al., 2013):

$$\mathrm{Pr} \xrightarrow{k_c \times f(R,V)} \mathrm{Pr}_c \tag{1}$$

$$\mathrm{Pr}_c \xrightarrow{k_o} \mathrm{Pr} + \mathrm{M} \tag{2}$$

$$\mathrm{M} \xrightarrow{k_P} \mathrm{M} + \mathrm{P} \tag{3}$$

$$\mathrm{M} \xrightarrow{d_M} \varnothing \tag{4}$$

$$\mathrm{P} \xrightarrow{d_P} \varnothing \tag{5}$$

The model includes transcription (Reactions (1) and (2)), translation (reaction (3)), and degradation of mRNA (M, Reaction (4)) and proteins (P, reaction (5)). Transcription initiation is a two-step process, consisting of the closed and open complex formations (Buc and McClure, 1985; Ribeiro et al., 2010). The free promoter is represented by Pr and the promoter-RNAP complex is represented by $\mathrm{Pr}_c$. Here, the closed complex formation can be repressed by a transcription factor produced by another gene by blocking access to the transcription start site. The repression function is a hill function with hill coefficient 2, as in (Zhu et al., 2007). Specifically, it is

$$f(R,V) = \frac{K_d^2}{\left(\frac{R}{V}\right)^2 + K_d^2} \tag{6}$$

where $R$ is the number of repressor molecules, $V$ is the normalized volume of the cell ranging from 0.5 to 1 over the cell cycle, and $K_d$ is the dissociation constant. This repression function arises when the promoter has two operator sites, and there is strongly cooperative binding between the two repressors which bind there.

### 2.2. Cell growth and division

Cell division in *E. coli* is remarkably stable, with little variance in division time of sister cells when under optimal growth conditions (squared coefficient of variation of division times $\approx 0.02$ (Hoffman and Frank, 1965)). We therefore divide cells according to a fixed doubling time $T_D$, implying that the population doubles in size synchronously. Cell growth is modeled by increasing $V$ linearly from 0.5 to 1 over the lifetime of the cell.

Each cell division is modeled as an instantaneous process which occurs at regular intervals, wherein the DNA (i.e. the promoter region, Pr) is replicated, and the M and P molecules are randomly partitioned into the daughter cells (see next section). After division, we assume that the promoters in the daughter cells are in the initial state (Pr), since any bound molecules are assumed to have been removed from the DNA by the DNA polymerase during replication (Guptasarma, 1995).

To illustrate the dynamics of the single-gene expression model from the previous section with the growth and division model here, we show several time traces in Fig. 1C, as well as the average behavior. Note that the subtle oscillatory behavior observed in the average behavior is due to the effects of linear cell growth and exponential protein degradation.

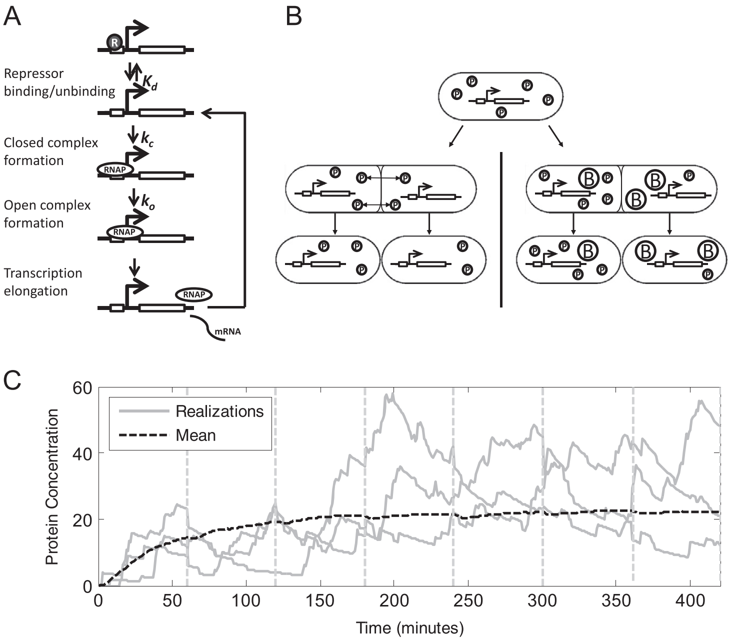


**Fig. 1.** (A) Illustration of the stochastic model given by reactions (1)–(6). (B) Illustration of the partitioning schemes used, pair formation (left) and random accessible volume (right), which result in lower- and higher-than-binomial variance in partitioning, respectively. (C) Protein concentration ($P/V$) over time for three independent realizations of a model with a single gene, and the overall mean from 1000 realizations. The vertical dashed lines show division points.

### *2.3. Molecule partitioning in cell division*

The molecule partitioning in cell division is done according to one of three partitioning schemes, which differ in the amount of variance in the molecule numbers that they introduce in division. In (Huh and Paulsson, 2011b), the partitioning error was quantified by $Q_X^2 = CV_L^2 - CV_X^2$, where $CV_X^2$ and $CV_L^2$ are the squared coefficients of variation ($CV^2$, defined as the variance over the squared mean) of the number of molecules in parent cells immediately before division ($X$), and in daughter cells immediately after division ($L$), respectively. If the molecules are partitioned independently and randomly, this will result in a binomial distribution in the number of molecules which are inherited by a given daughter, and thus $Q_X^2 = \langle X \rangle^{-1}$(Huh and Paulsson, 2011b). We quantify differences between the variances produced by partitioning schemes by the log of the ratio between the $Q_X^2$ produced by that scheme and what would be expected from a binomial partitioning, giving $\lg \tilde{Q}$, defined as

$$\lg \tilde{Q} = \lg(\langle X \rangle Q_X^2) \tag{7}$$

If molecules are partitioned independently, i.e. binomially, $\lg \tilde{Q}$ is 0. “Ordered” partitioning schemes resulting in lower variance have $\lg \tilde{Q} < 0$. For this, we use the ‘Pair Formation’ partitioning scheme (Huh and Paulsson, 2011b), where the segregated molecules first form pairs with probability $k$. These pairs then are equally divided into the daughter cells while the unpaired molecules are segregated independently (left side of Fig. 1B). It can be shown (see Supplementary Material) that to achieve a given $\lg \tilde{Q} < 0$, $k$ must be set as (from Eq. (8) of Huh and Paulsson, 2011b):

$$k = 1 - 10^{\lg \tilde{Q}} \tag{8}$$

“Disordered” partitioning schemes, resulting in greater variance, will have $\lg \tilde{Q} > 0$. For this, we use the ‘Random Accessible Volume' segregation scheme (Huh and Paulsson, 2011b), where large macromolecules in low copy number are independently segregated into the daughter cells. These macromolecules (denoted by B in Fig. 1B) reduce the volume accessible to other molecules, and the error in their partitioning is imparted to the segregated molecules (right side of Fig. 1B). If the same number of B molecules is used to partition all molecules in the cell, this will introduce a correlation in the number of molecules inherited by a given daughter. In all cases, we assessed whether this correlation affected the results by testing both a correlated model and an uncorrelated model, where the B molecules are different for each partitioned molecule. To achieve a given $\lg \tilde{Q} > 0$, we use the following number of B molecules (derived from (Huh and Paulsson, 2011b), see Supplementary Material):

$$B = \frac{\langle X \rangle CV_X^2 + \langle X \rangle - 1}{10^{\lg \tilde{Q}} - 1} \tag{9}$$

The values of $CV_X^2$ and $\langle X \rangle$ in (9) were calculated by running a simulation of the model with the binomial partitioning scheme, and extracting the $CV^2$ and mean of the protein distribution prior to divisions. We verified that Eqs. (8) and (9) produce the desired values of $\lg \tilde{Q}$. Supplementary Fig. S1 shows a good correspondence between the input $\lg \tilde{Q}$ and the value determined by simulation.

Finally, to test the behavior of the model in the limit of disordered partitioning, we use an all-or-nothing scheme, where one daughter always receives all molecules, while the other receives none. The $\lg \tilde{Q}$ for this scheme is labeled as “max” in the figures. As might be expected, greater variance in partitioning (higher $\lg \tilde{Q}$) increases the CV$^2$ of the protein concentration $P/V$

taken over all time (Supplementary Fig. S2), although the mean concentration is unaffected.

### 2.4. Stability of the toggle switch

We quantify the stability of the noisy attractors of the Toggle Switch by $\tau_s$, the mean time for this 2-gene network to change from one noisy attractor to the other, similar to (Potapov et al., 2011). Switching points are defined as the moments when the sign of the difference between the two protein concentrations differs from the previous moment. However, a large amount of small switching intervals are generated when the system is at the border between the two noisy attractors. We discount any intervals shorter than 40 min, so as to only consider 'definitive' switches. To get $\tau_s$, we use a maximum likelihood estimator of the conditioned exponential distribution, i.e. the sample mean subtracted by the threshold of 40 min.

### 2.5. Continuous culture of cells containing a toggle switch

For the study of the Toggle Switch in continuous culture, we use an abstracted model, so as to support the simulation of thousands of cells for many generations. The number of cells in each state is represented by $c_1$ and $c_2$, which evolve according to the following reactions:

$$p_i \xrightarrow{1/\tau_s} p_{3-i} \tag{10}$$

$$p_i \xrightarrow{\text{divide}} p_i^+ + p_i^- \tag{11}$$

$$p_i^+ \xrightarrow{k_\infty} p_i(\tau_P) \tag{12}$$

$$p_i^- \xrightarrow{k_\infty} p_i \tag{13}$$

$$p_i^- \xrightarrow{B_P \times k_\infty} p_{3-i} \tag{14}$$

Cells switch stochastically between states with mean time $\tau_s$ (reaction (10)). Cells divide synchronously every hour into two daughters, one with more molecules, $p_i^+$, and one with less, $p_i^-$ (reaction (11)). The sister cell inheriting more molecules is protected from switching for a given amount of time, proportional to the level of bias in partitioning ($B_P$, ranging from 0 to 1), denoted in reaction (12) by the fixed time delay $\tau_P$. Here, we set $\tau_P$ to $B_P \times T_D$. Meanwhile, the sister cells inheriting less molecules become more prone to switch to the other state with increasing $B_P$ (reactions (13) and (14)). In the above reactions, $k_\infty$ is an arbitrarily fast rate. In order to maintain the number of cells stable, exactly half of each type of cells, $p_i^+$ and $p_i^-$, are removed from the system after each division event.

## 3. Results

We study the effects of partitioning errors in the dynamics across cell generations of, first, the Toggle Switch, and second, the Repressilator. In both cases, we explore a wide range of partitioning error rates per division, given the known diversity of the partitioning schemes of proteins and plasmids in bacteria (Huh and Paulsson, 2011a, 2011b). For that, and by using different partitioning schemes, we vary $\lg \tilde{Q}$ from $-1$ (corresponding to highly symmetric partitioning) to 0 (binomial), up to the maximum allowed by the mean protein number in the mother cell at the moment of division (here labeled 'max').

### 3.1. Toggle switch

We first examined the effects of errors in the partitioning of regulatory molecules on the dynamics of the Toggle Switch. This circuit is constructed by duplicating reactions (1)–(5), and controlling the expression of each gene with the protein concentration of the other gene (full model presented in Supplementary Material, reactions (1)–(10)). Unless stated otherwise, we set the model parameters as described in Table 1, which result in a mean protein number before division $\langle X \rangle$ of approximately 20, if unrepressed. We set the dissociation constant, $K_d$, to a value that makes the noisy attractors stable ($K_d = \langle X \rangle / 3$) by maintaining the protein numbers of the repressed gene close to zero at all times. We simulated the system for $2 \times 10^8$ s, sampling every 2 min, for each $\lg \tilde{Q}$ tested. After each division, only one daughter of each lineage was simulated (randomly selected), to avoid exponential population growth.

In Fig. 2, we show the stability of the Toggle Switch for different levels of error in partitioning (black lines with crosses). As expected, increasing the error in partitioning from $\lg \tilde{Q} = 0$ to max destabilizes the switch on average, with the mean switching

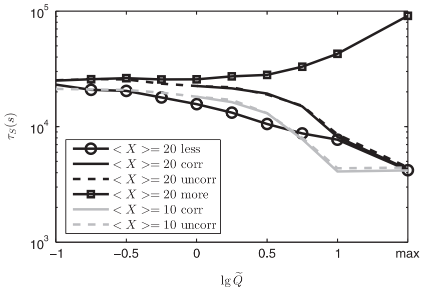


**Fig. 2.** Mean switching interval ($\tau_s$) of the Toggle Switch for different levels of error in partitioning, two mean protein levels, and correlated/uncorrelated disordered partitioning schemes. For $\langle X \rangle = 20$, switching intervals are also shown for the lineage which always inherits more/less molecules. Data is from one $2 \times 10^8$ s simulation for each data point.

**Table 1**
Parameters used in the model, unless stated otherwise.

| Parameter | Description | Value | Source |
|---|---|---|---|
| $T_D$ | Doubling time | 3600 s | Yu et al. (2006) |
| $k_c$ | Closed complex formation (maximum) | 1/300 s$^{-1}$ | Kandhavelu et al. (2011) |
| $k_o$ | Open complex formation | 1/300 s$^{-1}$ | Kandhavelu et al. (2011) |
| $d_M$ | mRNA degradation rate | 1/200 s$^{-1}$ | Bernstein et al. (2002) |
| $k_P$ | Translation rate | 3/200 s$^{-1}$ | Yu et al. (2006) |
| $d_P$ | Protein degradation rate[a] | 1/10,000 s$^{-1}$ | Taniguchi et al. (2010) |

[a] The degradation rate was set to match a mean protein number of ~20 molecules (Taniguchi et al., 2010), assuming no regulation and given the values of the other parameters.

time decreasing by a factor of ~4. Interestingly, the stabilizing effect from partitioning schemes with less variance than binomial does not seem to be as strong as the destabilizing effect from high-variance partitioning. We suspected that this may be due to the already-low variance introduced due to binomial partitioning with $\langle X \rangle = 20$. We therefore halved the translation rate $k_P$ to $1.5/200\ \text{s}^{-1}$ so as to reduce the $\langle X \rangle$ to 10 (gray line). This increases the error in partitioning in the binomial case, and should therefore increase the impact of error correction. Even after this change, however, the stabilizing effects of the lower-variance partitioning schemes appear to be minimal. Note that for the $\langle X \rangle = 10$ case, the maximum value of $\lg \tilde{Q}$ is 1, and thus the mean switching time for $\lg \tilde{Q} = 1$ and $\lg \tilde{Q} =$ max is the same. Finally, we did not observe any significant difference in the stability of the switch when using correlated and uncorrelated disordered partitioning (Fig. 2).

With high errors in partitioning, we noted one interesting phenomenon. Although the added variance destabilizes the Toggle Switch on average, there are many instances where one daughter cell inherits most of the proteins of the gene that was 'ON' in the mother cell at the moment of division. This generates a transient time during which the probability of switching state in that daughter cell is much smaller than otherwise. As such, high errors in partitioning can be a source of robustness of the states of the circuit in some cells, at the cost of loss of robustness in the sister cells. To show this, we simulated the lineage which inherits more molecules from the parent (black line with squares in Fig. 2). For these cells, an increased stability is observed for the high $\lg \tilde{Q}$ cases. Conversely, the lineage which inherits less molecules exhibits reduced stability (black line with circles in Fig. 2). Thus, high-variance partitioning of the proteins of the Toggle Switch leads to the splitting of the population into sub-populations of cells that differ in the degree of stability in their noisy attractors at birth. This has a far from straightforward effect on the phenotypic distribution of the cell population.

To study this, we constructed an abstracted model of a population of cells, each containing a Toggle Switch (see Methods). High-variance partitioning was modeled by protecting the daughter cells which inherit more molecules from switching, and destabilizing the daughter cells which inherit less, by increasing the probability that they change state after division (see Methods). For simplicity, we assume that the partitioning is biased, i.e. one cell always inherits significantly more than the other. We set the mean switching time when there is no error in partitioning to the measured time in Fig. 2 for $\lg \tilde{Q} = -1$ and $\langle X \rangle = 20$, i.e. $\tau_S = 2.5\times\ 10^4 s$. We simulated this model with 1000 cells for $10^8$ s and recorded the fraction of cells in one of the two states at each time moment. The variance-to-mean ratio (VMR) of this number is shown in Fig. 3 for different levels of bias in partitioning $B_P$.

If the phenotype of each cell is randomly, independently and unbiasedly chosen, the phenotypes should follow a binomial distribution with $p=0.5$. This distribution has a VMR of $1-p=0.5$, which is observed for the lower biases in partitioning in Fig. 3. Increasing the variance in partitioning (by increasing the bias in partitioning) has the counterintuitive effect of reducing the variance of the phenotype distribution. In other words, while the frequency of the fluctuations of this distribution is faster, the amplitude of the fluctuations is smaller, and thus the distribution is less broad over time. In the limit of fully biased partitioning ($B_P=1$), the VMR decreases to 1/3, independent of the population size and the switching rate (see Supplementary Material). This lower limit on the VMR is due to the combined effect of the randomization of the states of half of the population after each division ($p_i^-$) and the noise arising from previous events, namely divisions and switches between noisy attractors. Note that these levels of VMR would not be possible to reach without anti-correlations in protein numbers between sister cells (which do not arise from noise in gene expression).

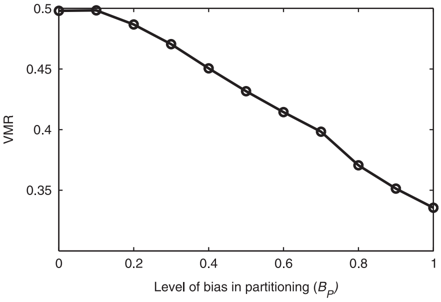


**Fig. 3.** Variance-to-mean ratio (VMR) of the phenotype distribution in a cell population with Toggle Switches in each cell, as a function of the bias in partitioning, $B_P$. Data is from populations of 1000 cells simulated for $10^8$ s.

It is possible to envision realistic scenarios in which reduced variance is advantageous. Assume that, under certain conditions, one of the two noisy attractors is lethal for example. In this case, high-variance partitioning may protect a population from extermination. To show this, we simulated a population of cells containing a switch that either produces a maintenance protein or a protein that leads to the lethal noisy attractor (we refer to this protein as being, in that sense, 'lethal').

For these simulations, we used the model of the Toggle Switch at the molecular level (built from reactions (1)–(5)), along with one extra condition: if the lethal protein exceeds a threshold, here set to eight molecules, the cell dies (full model presented in Supplementary Material, reactions (1)–(10)). The simulations were initialized with one cell in the nonlethal noisy attractor (25 maintenance proteins, and none of the lethal ones). Finally, in contrast to the switch in Fig. 2, we tripled the rate of the closed complex formation for the gene controlling the lethal state and weakened the repression strength to $K_d=14$ for both genes, so as to mimic a lysogenic cell under stress, e.g. due to UV irradiation (Baluch and Sussman, 1978). From the simulations, we obtained the probability that the resulting population survived the stress (here lasting 9.5 generations, or 30,600 s). The results are shown in Fig. 4 for both correlated and uncorrelated disordered partitioning schemes.

Fig. 4A shows that high-variance partitioning increases the chance of survival of a small population of cells. In particular, the survival chance of each cell increased by ~1.3 fold in the correlated case, despite the increased rate at which the switch changes state on average (Fig. 2), and by ~1.7 fold in the uncorrelated case. This increase in the latter case is due to the reduced chance that the cell inheriting the maintenance protein also inherits most lethal proteins. This strategy comes at a cost: the mean number of cells in the surviving population decreases with increasing variance in partitioning (Fig. 4B). In this case, the mean drops from 3.3 cells to 1.5 cells when increasing $\lg \tilde{Q}$ from 0 to max in both the correlated and uncorrelated schemes. Interestingly, the increased survival probability becomes apparent for intermediate values of $\lg \tilde{Q}$, without incurring a large loss in the mean surviving population size (mean of 2.6 and 2.7 for $\lg \tilde{Q} = 1$ cases with correlated and uncorrelated partitioning, respectively). No large differences were observed in the survival chances or in the mean number of cells in surviving populations when the variance in partitioning was decreased below binomial.

We note that the survival chance (Fig. 4A) is not monotonic with $\lg \tilde{Q}$, as it decreases slightly for small positive $\lg \tilde{Q}$, for both the correlated and uncorrelated cases. We expect that this is due

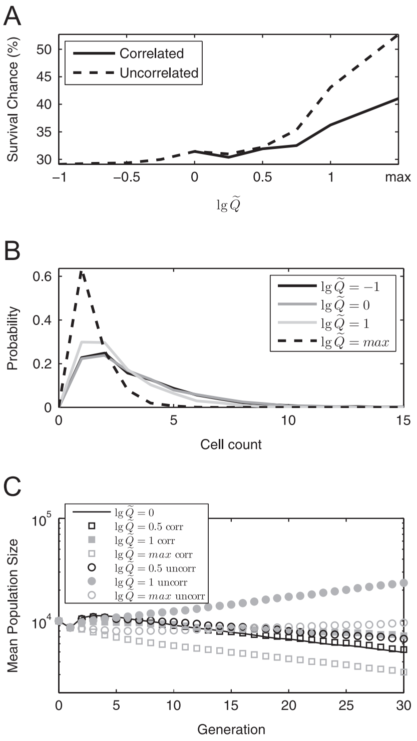


**Fig. 4.** (A) Survival chance of cell populations for different levels of error in partitioning and correlated/uncorrelated disordered partitioning schemes. (B) Distribution of the number cells in surviving population. (C) Mean population size over time for different partitioning schemes. Corr/uncorr refer to the correlated and uncorrelated disordered partitioning schemes. All data is from 10,000 simulations starting from one initial cell.

to the weakness of the aforementioned protection effect for small $\lg\tilde{Q}$. Without it, cells inheriting the majority of the proteins would be less stable, since they also inherit the majority of the lethal proteins. If this is the case, this effect should be exacerbated in the correlated case, where one daughter cell always falls into this category. As expected, Fig. 4A shows a slightly larger drop in survival chance for the correlated case compared to the uncorrelated case.

The above simulations assumed small initial populations. Increasing the initial size of a population increases its survival chances, since the probability that all cells die decreases. Nevertheless, the partitioning scheme affects the mean size of the population over time. To exemplify this, the number of cells in a population starting with 10,000 cells is shown in Fig. 4C for both the correlated and uncorrelated disordered partitioning schemes. After a transient of ~5 generations, all populations enter an exponential phase (linear on the log-linear plot). The slope of this line informs on how quickly the population is growing/shrinking in the different conditions. Table 2 shows the change in population size over 10 generations in this exponential phase. In agreement with the above results, all the disordered partitioning schemes improved populations' numbers, with the uncorrelated disordered schemes providing the best improvements, to the point where the numbers even grow. In addition, the point at which the growth rate is maximized is for an intermediate value of $\lg\tilde{Q}$, consistent with the above observation that the benefits of disordered partitioning appear before the drawbacks.

**Table 2**
Change in population size over 10 generations in exponential phase, for the independent partitioning scheme ($\lg\tilde{Q}=0$) and two disordered partitioning schemes with varying $\lg\tilde{Q}$. Data is from a least-squares linear fit to generations 15–30 in Fig. 4C.

| $\lg\tilde{Q}$ | Uncorrelated | Correlated |
|---|---|---|
| 0 | −27% | |
| 0.5 | −16% | −25% |
| 1 | +37% | −12% |
| *Max* | +8% | −24% |

The increase of survival rates with high-variance partitioning should apply at least so long as the protein lifetime is on the order of, or longer than the cell doubling time, to ensure that the added stability of the state of the cells is not lost during the cell cycle. Nevertheless, we tested several other parameter sets, including protein degradation faster than cell division. Qualitatively, the results hold, except for extreme parameter values. For example, high-variance partitioning decreases the population survival chance when the lethal protein's interaction with the maintenance gene's promoter is extremely cooperative.

We also tested whether resetting the promoter state at division (see Methods) affected the above results. The only significant effect was a reduction by ~1 to 5% in the survival chances in the lethal noisy attractor scenario. Finally, we tested whether, in this context, the use of a hill function (Eq. (6)) is equivalent to using elementary reactions by measuring the stability of Toggle Switches in both cases, as recent studies show that these two modeling strategies can exhibit significant differences (Zhu et al., 2007; Thomas et al., 2012). For this test, we did not allow cells to divide. We found no significant differences between the two models.

### 3.2. *Repressilator*

We next studied how errors in partitioning affect the behavior of the Repressilator across cell generations. The circuit is constructed by triplicating reactions (1)–(5), and controlling the production of each gene (labeled A, B and C) with the protein concentration of the previous gene (the full model is presented in the Supplementary Material, reactions (11)–(25)). We set the model parameters to those in Table 1, and $K_d$ to five proteins. We simulated the system for $10^7$ s, sampling every minute, for each $\lg\tilde{Q}$ tested, and quantified the period by the zeros of the autocorrelation function of the concentration of one gene's product. While the mean of the period (~375 min) does not differ between conditions, the variance does, as it increases with increasing variance in partitioning for both the correlated and uncorrelated disordered partitioning schemes (Fig. 5). The uncorrelated case exhibits lower variance than the correlated case, since there is a higher probability of transmitting some of the phase information to the daughter cells. No significant change in either the mean or the variability of the periods was observed for lower-variance partitioning schemes.

With less robust periods in the higher-variance partitioning case, we predict that an initially synchronized population of cells will desynchronize faster. To test this, we simulated the growth of 500 initially synchronous cells, and measured the mean protein concentration of each protein within the entire population at each moment (an example is shown in Fig. 6A for binomial partitioning). Note that the mean overall protein concentration exhibits a small oscillation, and does not converge to a constant value because of the combined effects of protein degradation and the linear increase of the cell volume over the cell cycle (same as in Fig. 1C). To quantify the loss of synchrony, we measured the

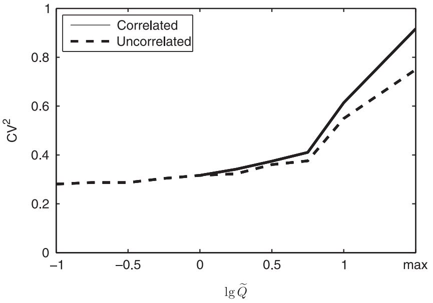


**Fig. 5.** $CV^2$ of the period of oscillation of the Repressilator, subject to differing levels of errors in partitioning and correlation in partitioning. Data is from one $10^7$ s simulation for each data point.

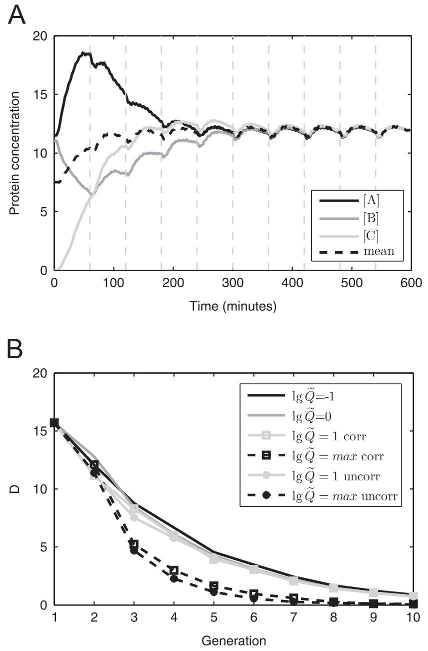


**Fig. 6.** (A) Mean protein concentrations of each protein in the Repressilator (solid lines), when subject to independent partitioning of molecules at division (i.e. $\lg \tilde{Q} = 0$). The overall mean protein concentration is also shown (dashed line). Vertical dashed lines indicate division points. Data is from a population starting with 500 cells in the same state ([A]=12, [B]=12, [C]=0). (B) Difference between the mean protein concentration of each protein and the mean overall protein concentration, averaged over each generation ($D$), for differing levels of error in partitioning and different partitioning schemes. Corr/uncorr refer to the correlated and uncorrelated disordered partitioning schemes. Data is from a population of 500 initial cells for each line.

absolute difference between the mean protein concentration of each protein (solid lines in Fig. 6A) and the mean overall protein concentration of the three genes (dashed line in Fig. 6A). We define $D$ as the average of this value over each generation. If cells are in perfect synchrony, as in the beginning of the simulations, $D$ will be large. Random fluctuations and random partitioning will reduce $D$, since the mean protein concentrations will converge to the overall mean, indicating that the population is less synchronous. The values of $D$ across generations for different levels of error in partitioning are shown in Fig. 6B.

For correlated partitioning, Fig. 6B shows that the synchronization of the Repressilator is remarkably robust to partitioning errors, as there is only a slight change in the rate of desynchronization (i.e. how quickly $D$ approaches 0) for $-1 \le \lg \tilde{Q} \le 1$. This is despite the observed increase in the noise in the lengths of the periods at $\lg \tilde{Q} = 1$ (Fig. 5).

When all-or-nothing partitioning is applied, the cells desynchronize rapidly, with its effects being visible already in the third generation. This is explained as follows: in the second generation, the cell receiving all proteins will oscillate more robustly than in the cases with less variance in partitioning, while the cell that received no proteins remains in an undetermined phase for most of its lifetime.

Finally, despite the decreased variance in the periods (Fig. 5), uncorrelated partitioning decreases the synchrony of the population slightly faster than correlated partitioning. This is due to the highly synchronous subpopulation of cells inheriting the majority of molecules in the correlated case. Overall, our results show that the synchrony of the Repressilator is not affected by moderate partitioning errors, but decreases for more extreme errors.

When relaxing the assumption that the promoter state resets during division, we found no significant effect on either the noise in the oscillation period or the rate of desynchronization. Further, we looked for phase-locking effects when the period of the Repressilator was near an integer multiple of the cell cycle. We were, however, not able to observe phase-locking in our simulations, perhaps due to the magnitude of the noise in the period's length, even in the lowest $\lg \tilde{Q}$ case (Fig. 5). This is in agreement with a lack of change in behavior when changing the ratio between the mean period and the mean length of the cell cycle (data not shown).

## 4. Conclusions and discussion

From stochastic simulations, we studied the effects of errors in partitioning on the behavior across cell generations of two genetic circuits, the Toggle Switch and the Repressilator. Knowledge of the effects is necessary not only to understand their kinetics in long scales across cell lineages but also in the context of synthetic biology, where partitioning schemes may potentially be used as regulatory mechanisms. The results suggest that genetic circuits are far from immune to this source of cell-to-cell variability, although the extent to which they are affected is heavily network-dependent.

We found that increasing partitioning errors not only decreases the stability of the noisy attractors of the Toggle Switch but also decreases the variance of the phenotypic distribution of the population below that of a binomial distribution. Notably, while the former result could be obtained by increasing noise in gene expression, the latter could not. This effect was due to the anti-correlation between the protein numbers inherited by sister cells, which is enhanced with high-variance partitioning and increases the stability of the inherited state of one cell at the cost of the stability of its sister cell. In this context, we considered an extreme case, by assuming that one of the states of a switch led to cell death. We found that finite cell populations, originally heading to extinction when employing binomial partitioning of components, increase their survival chances and may even grow in numbers over time, if they employ disordered partitioning schemes instead. This is due to the increased chances that, following each division, one of the daughter cells will remain in the non-lethal state.

The Repressilator was found to be more robust than the switch to increasing partitioning errors. Though the variance in the periods increased, consistent with an increase in noise in gene expression, the synchrony of a population was remarkably robust to this increase. Only the strongest errors in partitioning (i.e., the all-or-nothing partitioning scheme) were able to significantly affect the degree of synchrony of the population. This is interesting, in that the function of circuits to track time is commonly the maintenance of synchrony between cells in a population. Further, even in the 'all-or-nothing' scenario, we expect that a simple cell to cell communication system will suffice to quickly resynchronize the clocks of sister cells following division.

We also studied the effects of correlations in the partitioning errors of the different proteins of the two circuits above. Such correlations are expected if the division process is morphologically asymmetric. We found that these correlations have no effect on the stability of the genetic switches and only slightly increase the rate of desynchronization of Repressilators in sister cells.

In all conditions tested, we did not observe any significant effect in the behavior of cells when using ordered partitioning schemes, when compared to binomial partitioning. This, combined with the fact that the implementation of such schemes is likely energy-consuming (due to requiring error correction (Huh and Paulsson, 2011b)), may explain why its use, while not absent (Di Ventura and Sourjik, 2011), is seemingly rare in nature, at least for low-to-medium-copy components such as RNA and regulatory proteins.

Non-binomial partitioning errors can arise in a number of different ways. Here, we have used pair formation to achieve $\lg \tilde{Q} < 0$, and random accessible volume to achieve $\lg \tilde{Q} > 0$. Though we believe that $Q_X^2$, and thus $\lg \tilde{Q}$, captures the most important aspect of the partitioning schemes (Huh and Paulsson, 2011b), other partitioning schemes could result in similar values of $\lg \tilde{Q}$, but lead to different behaviors. For example, correlated and uncorrelated disordered partitioning schemes produce slightly different effects for the same $\lg \tilde{Q}$. As more complex networks are analyzed in this context in the future, it will likely become necessary to characterize the various possible partitioning schemes more comprehensively.

One sort of error in partitioning not considered here occurs when the circuit is expressed from a multi-copy plasmid (Gardner et al., 2000; Elowitz and Leibler, 2000), whose numbers are also partitioned stochastically in division (Reyes-Lamothe et al., 2013). We expect that errors in plasmid partitioning will affect the dynamics of the circuits they code for in a manner similar to the correlated disordered partitioning schemes employed here, since the same partitioning error eventually affects all proteins. However, the impact of the division event will be delayed and diluted over time by noise in plasmid replication and gene expression. Other extrinsic noise sources not considered here include cell to cell diversity in RNA polymerase and ribosome numbers, among others. Future studies may assist in quantifying the contribution of these sources on the temporal distributions of cellular phenotypes.

The results above show that the effects of errors in partitioning differ widely from those of noise in gene expression. The differences arise primarily from the unavoidable anti-correlation in the numbers of molecules inherited by sister cells, whereas noise in gene expression affects all cells in the population independently. In other words, unlike noise in gene expression, the division process forces sister cells to move in opposite directions in the network's state space, starting from the location of the mother cell the moment prior to division.

The qualitative differences in the effects of errors in partitioning in the Toggle Switch and the Repressilator derive from the differences in their long-term behaviors. In the Toggle Switch, with two noisy attractors, division can move one of the daughter cells close to the border of the basin of attraction that the mother cell lied on, while moving the other daughter cell further into the basin. That creates a strong possibility that the former cell switches into the neighbor attractor while the chances that the latter remains in the present attractor are enhanced. In other words, there are increased chances that sister cells will exhibit opposite behaviors. The Repressilator, on the other hand, has only one attractor, a state cycle. Regardless where the daughter cells lie on the state space following division, they will both travel towards the same attractor, thus becoming closer in the state space with time. Thus, in this network, the effects of partitioning are hardly distinguishable from those of noise in gene expression.

From all of the above, it is possible to infer general consequences of errors in partitioning on the dynamics of small genetic circuits. In circuits with only one noisy attractor, the effects of these errors are not expected to differ qualitatively from those of noise in gene expression. Meanwhile, in circuits with more than one noisy attractor, large errors in partitioning enhance the chances for sister cells to begin their lifetime in different noisy attractors, with one sister cell deeper into the mother's basin of attraction and the other jumping way from it. There is a process in natural organisms that exhibits some similarity. Namely, multicellular organisms have stem cells which, in division, produce both a renewed stem cell (i.e. on the same noisy attractor as the mother cell) and a differentiated cell (i.e. on another noisy attractor of the gene regulatory network). It would be of interest to assess in the future the degree to which the many asymmetries in these division events are deliberate.

Finally, it is interesting to note that the effects of noise in gene expression and of errors in partitioning must, in one or more aspects, differ for all networks. This is because, first, their effects at the single gene level differ, in that while noise in gene expression enhances fluctuations at all time points, errors in partitioning occur only at specific, rare moments. Second, at the network level, increased noise in gene expression decreases the stability of all noisy attractors at all times. Meanwhile, partitioning errors promote transitions between attractors at specific points in time, without affecting their stability otherwise. As such, it is reasonable to hypothesize that both of these 'perturbation mechanisms' will be of use in present efforts in Synthetic Biology and are likely to be used for different aims in natural organisms.

## Acknowledgments

Work supported by Finnish Funding Agency for Technology and Innovation (ASR), Academy of Finland (HT, ASR), Portuguese Foundation for Science and Technology (ASR) and TUT President's graduate programme (JLP). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.jtbi.2014.04.018.

## References

Alon, U., 2007. Network motifs: theory and experimental approaches. Nat. Rev. Genet. 8, 450–461.

Arkin, A., Ross, J., Mcadams, H.H., 1998. Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in phage. Genetics 149, 1633–1648.

Baluch, J., Sussman, R., 1978. Correlation between UV dose requirement for lambda bacteriophage induction and lambda repressor concentration. J. Virol. 26, 595–602.

Bernstein, J.A., Khodursky, A.B., Lin-Chao, S., Cohen, S.N., 2002. Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. Proc. Natl. Acad. Sci. U. S. A. 99, 9697–9702.

Buc, H., McClure, W.R., 1985. Kinetics of open complex formation between Escherichia coli RNA polymerase and the lac UV5 promoter. Evidence for a sequential mechanism involving three steps. Biochemistry 24, 2712–2723.

Di Ventura, B., Sourjik, V., 2011. Self-organized partitioning of dynamically localized proteins in bacterial cell division. Mol. Syst. Biol. 7, 457.

Elowitz, M.B., Leibler, S., 2000. A synthetic oscillatory network of transcriptional regulators. Nature 403, 335–338.

Gardner, T.S., Cantor, C.R., Collins, J.J., 2000. Construction of a genetic toggle switch in Escherichia coli. Nature 403, 339–342.

Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81, 2340–2361.

Guptasarma, P., 1995. Does replication-induced transcription regulate synthesis of the myriad low copy number proteins of Escherichia coli? Bioessays 17, 987–997.

Häkkinen, A., Tran, H., Yli-Harja, O., Ribeiro, A.S., 2013. Effects of rate-limiting steps in transcription initiation on genetic filter motifs. PLoS One 8, e70439.

Hoffman, H., Frank, M.E., 1965. Synchrony of division in clonal microcolonies of Escherichia coli. J. Bacteriol. 89, 513–517.

Huh, D., Paulsson, J., 2011a. Non-genetic heterogeneity from stochastic partitioning at cell division. Nat. Genet. 43, 95–100.

Huh, D., Paulsson, J., 2011b. Random partitioning of molecules at cell division. Proc. Natl. Acad. Sci. U. S. A. 108, 15004–15009.

Kaern, M., Elston, T.C., Blake, W.J., Collins, J.J., 2005. Stochasticity in gene expression: from theories to phenotypes. Nat. Rev. Genet. 6, 451–464.

Kandhavelu, M., Mannerstrom, H., Gupta, A., Häkkinen, A., Lloyd-Price, J., Yli-Harja, O., et al., 2011. in vivo kinetics of transcription initiation of the lar promoter in Escherichia coli. Evidence for a sequential mechanism with two rate-limiting steps. BMC Syst. Biol. 5, 149.

Kussell, E., Leibler, S., 2005. Phenotypic diversity, population growth, and information in fluctuating environments. Science 309, 2075–2078.

Lahav, G., Rosenfeld, N., Sigal, A., Geva-Zatorsky, N., Levine, A.J., Elowitz, M.B., et al., 2004. Dynamics of the p53-Mdm2 feedback loop in individual cells. Nat. Genet. 36, 147–150.

Lloyd-Price, J., Lehtivaara, M., Kandhavelu, M., Chowdhury, S., Muthukrishnan, A.-B., Yli-Harja, O., et al., 2012. Probabilistic RNA partitioning generates transient increases in the normalized variance of RNA numbers in synchronized populations of Escherichia coli. Mol. Biosyst. 8, 565–571.

Lloyd-Price, J., Gupta, A., Ribeiro, A.S., 2012. SGNS2: a compartmentalized stochastic chemical kinetics simulator for dynamic cell populations. Bioinformatics 28, 3004–3005.

Loinger, A., Biham, O., 2007. Stochastic simulations of the repressilator circuit. Phys. Rev. E. 76, 051917.

Loinger, A., Lipshtat, A., Balaban, N., Biham, O., 2007. Stochastic simulations of genetic switch systems. Phys. Rev. E. 75, 021904.

McAdams, H.H., Arkin, A., 1999. It's a noisy business! Genetic regulation at the nanomolar scale. Trends Genet. 15, 65–69.

Montero Llopis, P., Jackson, A.F., Sliusarenko, O., Surovtsev, I., Heinritz, J., Emonet, T., et al., 2010. Spatial organization of the flow of genetic information in bacteria. Nature 466, 77–81.

Nelson, D.E., Ihekwaba, A.E.C., Elliott, M., Johnson, J.R., Gibney, C.A., Foreman, B.E., et al., 2004. Oscillations in NF-kappaB signaling control the dynamics of gene expression. Science 306, 704–708.

Neubauer, Z., Calef, E., 1970. Immunity phase-shift in defective lysogens: non-mutational hereditary change of early regulation of λ prophage. J. Mol. Biol. 51, 1–13.

Potapov, I., Lloyd-Price, J., Yli-Harja, O., Ribeiro, A.S., 2011. Dynamics of a genetic toggle switch at the nucleotide and codon levels. Phys. Rev. E. 84, 031903.

Reyes-Lamothe, R., Tran, T., Meas, D., Lee, L., Li, A.M., Sherratt, D.J., et al., 2013. High-copy bacterial plasmids diffuse in the nucleoid-free space, replicate stochastically and are randomly partitioned at cell division. Nucleic Acids Res. 42, 1042–1051.

Ribeiro, A.S., Kauffman, S.A., 2007. Noisy attractors and ergodic sets in models of gene regulatory networks. J. Theor. Biol. 247, 743–755.

Ribeiro, A.S., Zhu, R., Kauffman, S.A., 2006. A general modeling strategy for gene regulatory networks with stochastic dynamics. J. Comput. Biol. 13, 1630–1639.

Ribeiro, A.S., Häkkinen, A., Mannerstrom, H., Lloyd-Price, J., Yli-Harja, O., 2010. Effects of the promoter open complex formation on gene expression dynamics. Phys. Rev. E. 81, 011912.

Samoilov, M.S., Price, G., Arkin, A.P., 2006. From fluctuations to phenotypes: the physiology of noise. Sci. STKE 2006, re17.

Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., et al., 2010. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science 329, 533–538.

Thomas, P., Straube, A.V., Grima, R., 2012. The slow-scale linear noise approximation: an accurate, reduced stochastic description of biochemical networks under timescale separation conditions. BMC Syst. Biol. 6, 39.

Wolf, D.M., Arkin, A.P., 2002. 15 min of fim: control of phase variation in E. coli. Omi. J. Integr. Biol. 6, 91–114.

Wolf, D.M., Arkin, A.P., 2003. Motifs, modules and games in bacteria. Curr. Opin. Microbiol. 6, 125–134.

Yu, J., Xiao, J., Ren, X., Lao, K., Xie, X.S., 2006. Probing gene expression in live cells, one protein molecule at a time. Science 311, 1600–1603.

Zhu, R., Ribeiro, A.S., Salahub, D., Kauffman, S.A., 2007. Studying genetic regulatory networks at the molecular level: delayed reaction stochastic models. J. Theor. Biol. 246, 725–745.

# Supplement to “Dynamics of small genetic circuits subject to stochastic partitioning in cell division”

Jason Lloyd-Price, Huy Tran, and Andre S. Ribeiro

## Parameters for partitioning schemes

The formulas used to adjust the $lg\tilde{Q}$ of a given partitioning scheme are presented here.

The partitioning error for the Pair Formation scheme is given by equation 8 of ref. 6 in the main manuscript (where the probability of evenly partitioning a pair is *p* = 1):

$$Q_X^2 = \frac{1-k}{\langle X\rangle}$$

where *k* is the fraction of molecules that form pairs. The value of $lg\tilde{Q}$ is given by:

$$lg\tilde{Q} = lg(\langle X\rangle Q_X^2) = \lg(1-k)$$

To achieve a given $lg\tilde{Q}$, we therefore set $k = 1 - 10^{lg\tilde{Q}}$.

The partitioning error for the Random Accessible Volume segregation scheme, from equation 2 of ref. 6 in the main manuscript, is:

$$Q_X^2 = \frac{1-Q_{vol}^2}{\langle X\rangle} + Q_{vol}^2(CV_X^2+1)$$

where $Q_{vol}^2$ is the partitioning error of the accessible volume. The value of $Q_{vol}^2$ is determined by the number of macromolecules (denoted by B) that reduce the volume accessible to other molecules:

$$Q_{vol}^2 = \frac{1}{\langle B\rangle}$$

The value of $lg\tilde{Q}$ is given by:

$$lg\tilde{Q} = lg(\langle X\rangle Q_X^2) = \lg(1 - Q_{vol}^2 + Q_{vol}^2(CV_X^2+1)\langle X\rangle)$$

$$= \lg\left(1 + \frac{(CV_X^2+1)\langle X\rangle - 1}{B}\right)$$

To achieve a given $lg\tilde{Q}$, we therefore set $B = \frac{\langle X\rangle CV_X^2 + \langle X\rangle - 1}{10^{lg\tilde{Q}} - 1}$, where the values of $CV_X^2$ and $\langle X\rangle$ were calculated by simulating a model with the binomial partitioning scheme, and sampling these values immediately before division events. Figure S1 shows that the above formulas produce the desired values of $lg\tilde{Q}$ when applying the different partitioning schemes.

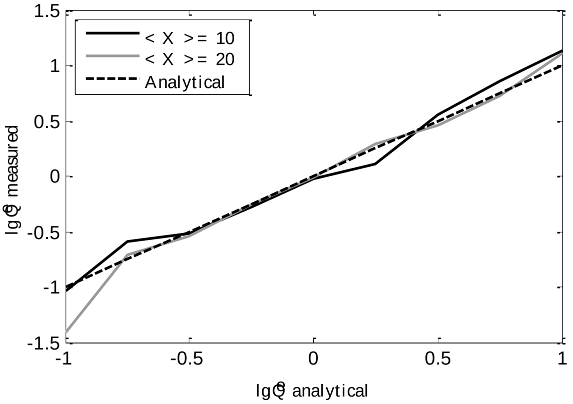


**Figure S1:** Input lg$\tilde{Q}$ and simulated results after applying the partitioning schemes.

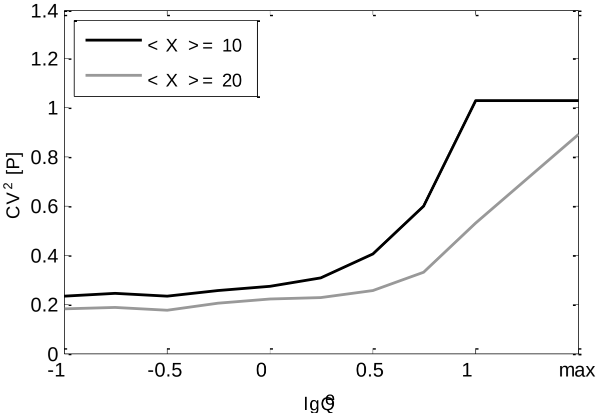


**Figure S2:** CV$^2$ of the protein concentration ([P] = P/V), taken over all time with different errors in partitioning in division, for different mean protein levels before division. Data is from a single simulation of length $10^8$ s, for each level of partitioning error.

## Stochastic model of the Toggle Switch

The Toggle Switch's model comprises two genes A and B, which repress each other via their protein products. The repression function is a hill function, described in equation (6) of the manuscript. Parameters are shown in Table S1.

| Parameter | Description | Value |
|---|---|---|
| $T_D$ | Doubling Time | 3600 s |
| $k_{cA}$ | Gene A's Closed Complex Formation Rate | 1/300 $s^{-1}$ |
| $k_{cB}$ | Gene B's Closed Complex Formation Rate | 1/300 $s^{-1}$ |
| $k_o$ | Open Complex Formation Rate | 1/300 $s^{-1}$ |
| $d_M$ | mRNA Degradation Rate | 1/200 $s^{-1}$ |
| $k_P$ | Translation Rate | 3/200 $s^{-1}$ |
| $d_P$ | Protein Degradation Rate | 1/10000 $s^{-1}$ |
| $K_d$ | Dissociation Constant | 20/3 |

**Table S1:** Parameters used in the single lineage simulation of a Toggle Switch. To vary the mean protein level from 20 to 10, the translation rate $k_P$ is halved from 3/200 $s^{-1}$ to 1.5/200 $s^{-1}$. The dissociation constant $K_d$, set as <X>/3, is 20/3 or 10/3 respectively. For parameter sources, see Table 1 in the main manuscript.

The model consists of the following set of reactions:

$$\mathrm{Pr}_A \xrightarrow{k_{cA}\times f(P_B,V)} \mathrm{Pr}_{Ac} \quad (1)$$

$$\mathrm{Pr}_{Ac} \xrightarrow{k_o} \mathrm{Pr}_A + \mathrm{M}_A \quad (2)$$

$$\mathrm{M}_A \xrightarrow{k_P} \mathrm{M}_A + \mathrm{P}_A \quad (3)$$

$$\mathrm{M}_A \xrightarrow{d_M} \varnothing \quad (4)$$

$$\mathrm{P}_A \xrightarrow{d_P} \varnothing \quad (5)$$

$$\mathrm{Pr}_B \xrightarrow{k_{cB}\times f(P_A,V)} \mathrm{Pr}_{Bc} \quad (6)$$

$$\mathrm{Pr}_{Bc} \xrightarrow{k_o} \mathrm{Pr}_B + \mathrm{M}_B \quad (7)$$

$$\mathrm{M}_B \xrightarrow{k_P} \mathrm{M}_B + \mathrm{P}_B \quad (8)$$

$$\mathrm{M}_B \xrightarrow{d_M} \varnothing \quad (9)$$

$$\mathrm{P}_B \xrightarrow{d_P} \varnothing \quad (10)$$

In the case where one noisy attractor is 'lethal', the model consists of reactions (1)-(10) with one additional condition: if $P_B$ equals or exceeds 8, the simulation of that cell is immediately ended. Parameters are shown in Table S2.

| **Parameter** | **Description** | **Value** |
|---|---|---|
| $T_D$ | Doubling Time | 3600 s |
| $k_{cA}$ | Gene A’s Closed Complex Formation Rate | 1/300 $s^{-1}$ |
| $k_{cB}$ | Gene B’s Closed Complex Formation Rate | 3/300 $s^{-1}$ |
| $k_o$ | Open Complex Formation | 1/300 $s^{-1}$ |
| $d_M$ | mRNA Degradation Rate | 1/200 $s^{-1}$ |
| $k_P$ | Translation Rate | 3/200 $s^{-1}$ |
| $d_P$ | Protein Degradation Rate | 1/10000 $s^{-1}$ |
| $K_d$ | Dissociation Constant | 14 |

**Table S2:** Parameters used in the Toggle Switch simulation in the case where one noisy attractor is ‘lethal’. Gene B, the ‘lethal’ gene, has the rate of closed complex formation $k_{cB}$ 3 times faster than that of gene A ($k_{cA}$), the maintenance gene. For parameter sources, see Table 1 in the main manuscript.

## Stochastic model of the Repressilator

The model of the Repressilator consists of three genes A, B, and C, which repress each other in a ring. The repression function is a hill function, described in equation (6) of the manuscript. The model parameters are shown in Table S3.

| **Parameter** | **Description** | **Value** |
|---|---|---|
| $T_D$ | Doubling Time | 3600 s |
| $k_c$ | Closed Complex Formation Rate | 1/300 $s^{-1}$ |
| $k_o$ | Open Complex Formation Rate | 1/300 $s^{-1}$ |
| $d_M$ | mRNA Degradation Rate | 1/200 $s^{-1}$ |
| $k_P$ | Translation Rate | 3/200 $s^{-1}$ |
| $d_P$ | Protein Degradation Rate | 1/10000 $s^{-1}$ |
| $K_d$ | Dissociation Constant | 5 |

**Table S3:** Parameters used in the Repressilator simulation. For parameter sources, see Table 1 in the main manuscript.

The model consists of the following set of reactions:

$$\mathrm{Pr}_A \xrightarrow{k_c \times f(P_C, V)} \mathrm{Pr}_{Ac} \qquad (11)$$

$$\mathrm{Pr}_{Ac} \xrightarrow{k_o} \mathrm{Pr}_A + \mathrm{M}_A \qquad (12)$$

$$\mathrm{M}_A \xrightarrow{k_P} \mathrm{M}_A + \mathrm{P}_A \qquad (13)$$

$$\mathrm{M}_A \xrightarrow{d_M} \varnothing \qquad (14)$$

$$\mathrm{P}_A \xrightarrow{d_P} \varnothing \qquad (15)$$

$$\mathrm{Pr}_B \xrightarrow{k_c \times f(P_A, V)} \mathrm{Pr}_{Bc} \qquad (16)$$

$$\mathrm{Pr}_{Bc} \xrightarrow{k_o} \mathrm{Pr}_B + \mathrm{M}_B \qquad (17)$$

$$\mathrm{M}_B \xrightarrow{k_P} \mathrm{M}_B + \mathrm{P}_B \qquad (18)$$

$$\mathrm{M}_B \xrightarrow{d_M} \varnothing \qquad (19)$$

$$\mathrm{P}_B \xrightarrow{d_P} \varnothing \tag{20}$$

$$\mathrm{Pr}_C \xrightarrow{k_c \times f(P_B, V)} \mathrm{Pr}_{Cc} \tag{21}$$

$$\mathrm{Pr}_{Cc} \xrightarrow{k_o} \mathrm{Pr}_C + \mathrm{M}_C \tag{22}$$

$$\mathrm{M}_C \xrightarrow{k_P} \mathrm{M}_C + \mathrm{P}_C \tag{23}$$

$$\mathrm{M}_C \xrightarrow{d_M} \varnothing \tag{24}$$

$$\mathrm{P}_C \xrightarrow{d_P} \varnothing \tag{25}$$

## Minimum VMR with fully biased partitioning

In the case of fully biased partitioning in the continuous culture model of a population of cells containing Toggle Switches, it is possible to derive the VMR of the phenotype distribution as follows. Let $N$ be the number of cells before a division, $\sigma^2$ be the variance of the number of cells in state 1 at that time. The variance of the number of cells in state 1 after the division (after reactions (13) and (14) of the manuscript), is then $\sigma^2/4 + Np(1-p)/2$, where $p$ is the probability that one of the cells receiving nothing will end up in state 1 after division, which is 0.5 since the switch is unbiased. We then obtain the variance of the stationary phenotype distribution by setting:

$$\sigma^2 = \frac{\sigma^2}{4} + \frac{Np(1-p)}{2}$$

The VMR is therefore:

$$\frac{\sigma^2}{N/2} = \frac{1}{3}$$

We note that this result does not apply if the population is allowed to grow indefinitely. In this case, the VMR converges to the VMR of a binomial.

Tampereen teknillinen yliopisto
PL 527
33101 Tampere

Tampere University of Technology
P.O.B. 527
FI-33101 Tampere, Finland

ISBN 978-952-15-3546-8
ISSN 1459-2045