12-2016

Spatial Crowding and Confinement Effects on Bursty Gene Expression

Charles Wei-Shing Chin
University of Tennessee, Knoxville, cchin@utk.edu

Recommended Citation

Chin, Charles Wei-Shing, "Spatial Crowding and Confinement Effects on Bursty Gene Expression. " PhD diss., University of Tennessee, 2016.
https://trace.tennessee.edu/utk_graddiss/4130

To the Graduate Council:

I am submitting herewith a dissertation written by Charles Wei-Shing Chin entitled "Spatial Crowding and Confinement Effects on Bursty Gene Expression." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Energy Science and Engineering.

Michael L. Simpson, Major Professor

We have read this dissertation and recommend its acceptance:

Steven M. Abel, Mitchel J. Doktycz, Scott T. Retterer, Eric T. Boder

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Spatial Crowding and Confinement Effects on Bursty Gene Expression

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Charles Wei-Shing Chin

December 2016

Dedication

To my wife Mary Katherine,
and to my loving family.

Acknowledgements

I'd like to acknowledge first and foremost by advisors, Dr. Michael Simpson and Dr. Steve Abel, two people without whom I would not have completed this achievement. Both of these people provided the support, advice, and guidance I needed to make it through this program, and for their contributions I am eternally grateful.

I would also like to acknowledge my doctoral committee, consisting of Drs. Simpson, Abel, Retterer, Doktycz, and Boder, who have helped me through the past few years hone and focus what was a nebulous idea in computational biology into the hypothesis detailed in this dissertation.

The Center for Nanophase Materials Sciences (CNMS) at the Oak Ridge National Laboratory (ORNL) was a unique and invaluable resource which provided me with a world-class environment and equipment. It is here that I also met several colleagues who have helped at various stages of my career, including Jason Fowlkes, Pat Collier, Roy Dar, and Brandon Razooky.

I'd also like to thank the University of Tennessee for the use of their Newton Computational Resource, which turned what could have been months of arduous calculations into a matter of weeks.

I would be remiss to ignore the significant support from Dr. Lee Riedinger and the entire Bredesen Center, who have not only helped support my academic career financially, but also helped through career advice and support through the ups and downs that come with a graduate degree.

I'd also like to thank some of my classmates, including Liz Norred and Patrick Caveney, for keeping me sane during the long boughs of coding and helping discuss ideas which lead to new insights on these research topics.

Finally, I would like to thank my friends and family, who have always believed in me, even when my faith in myself wavered. Thank you for your unending support, and I hope to make all of you proud.

Abstract

Synthetic biology and genetic engineering are valuable tools in the development of new, sustainable energy generation technologies. The characterization of stochastic gene expression is vital to the efficient application of genetic engineering techniques. Transcriptional bursting, in which periods of high expression are punctuated by periods of no expression, is extensively observed in gene expression. While various molecular mechanisms have been hypothesized to be responsible for transcriptional bursting, spatial considerations have largely been neglected. This work uses computational modeling to examine in detail the influence of spatial factors such as macromolecular crowding and confinement on gene expression.

In the first part of the thesis, cell-free expression chambers containing *E. coli* extract were fabricated and analyzed under varying confinement scenarios to explore how resource sharing influences gene expression. Interestingly, fluorescence measurements reveal that expression burst size, but not burst frequency, is highly sensitive to changes in chamber volume and the size of the shared resource pool. Computational models reveal that the timing of initial transcriptional activity strongly influences the acquisition of resources, such that mRNA transcripts produced early in time dominate the burst behavior of a chamber.

In the second part of the thesis, computational models were developed to study the effects of macromolecular crowding and confinement on transcriptional bursting. Spatially resolved gene expression models reveal significant changes in fluctuations and noise in mRNA behavior compared with well-mixed systems. The spatial results were compared to two- and three-state models to determine whether the effects of crowding and confinement could be adequately captured using simpler models. The comparisons reveal that the two- and three-state models, which do not explicitly incorporate spatial features, are unable to capture features of the noise of crowded and confined systems due to differences in the distribution of times between transcriptional events.

The work presented here reveals the importance of spatial influences when analyzing gene expression and transcriptional bursting in cells. Future work will expand on the role of resource sharing on gene expression through spatial considerations, as well as explore the effects of crowding on more complex gene expression systems.

Table of Contents

List of Figures

List of Symbols and Variables

τ	time constant
α or k_m	transcription rate for the Gillespie algorithm
k_p	translation rate for the Gillespie algorithm
γ_m	mRNA decay rate for the Gillespie algorithm
γ_p	protein decay rate for the Gillespie algorithm
k_{ON}	rate for Gillespie algorithm to turn to on state
k_{OFF}	rate for Gillespie algorithm to turn to off state
σ^2	variance
μ	mean
V	volume
x_i	number of molecules of species i
R	chemical reaction
M	number of possible reactions
S	molecular species
c	probability of reaction constant
h	number of distinct R reactant combinations
α_0	total reaction propensity
r_1 and r_2	uniformly distributed random variables
O	On fraction
b	Translational burst size
B	Transcriptional burst size
b_f	Burst frequency
g_m	gain factor

List of Abbreviations

E. coli	*Escherichia coli*
HIV-1	Human Immunodeficiency Virus type 1
ODE	Ordinary differential equations
BD	Brownian dynamics
CME	Chemical master equation
SSA	Stochastic simulation algorithm
CV^2	Coefficient of variance squared
smFISH	Single-molecule fluorescence *in situ* hybridization
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
mRNA	Messenger RNA
P	Protein
CFPS	Cell-free protein synthesis
PDMS	Polydimethylsiloxane
EGFP	Enhanced Green Fluorescent Protein
AU	Arbitrary units
Voxel	Volumetric pixel

1. Introduction

1.1 Biotechnology's Role in the World's Greatest Challenge

One of the greatest challenges to the world today is providing the increasing amount of energy needed to support global demand[1]. However, the energy of the future must be both sustainable and plentiful to avoid the repercussions of global climate change due to energy production. Over the past 10 years, global energy consumption grew at an average of 2% per year[2], currently totaling near 13,000 Million tonnes oil equivalent[3]. As more of the global population reaches for higher standards of living, the need for more resources increases at an even higher rate in underdeveloped regions[2]. Specifically, energy demands are estimated to reach 30 TW_{avg} by the year 2050[4]. However, to bring the entirety of the world's population up to the living standards of western society, that number will need to be doubled[1, 5]. To sustainability produce the needed energy levels and stem the 16 fold increase in global carbon dioxide production since 1900[6], global energy production emissions[7] must be reduced as the world demands more energy.

Current research into sustainable energy practices use biological techniques to reduce greenhouse gas emissions in a variety of different energy related fields. Fossil fuels account for over 80% of the world's energy needs[3], with oil making up 32% of the total[2]. In the U.S., 71% of oil is used solely for transportation[8]. Biofuels are a heavily researched and attractive method for offsetting carbon emissions due to their ability to replace fossil fuels. Biofuel production can take many forms, such as converting waste biomass from existing agricultural industries into ethanol, as Brazil does to generate 5.5 billion gallons of ethanol from sugarcane waste[9], or through the use of switchgrass to replace corn ethanol in the U.S.[10].

Additionally, biological techniques are used to reduce greenhouse gas emissions by developing alternate feedstocks for products which are traditionally petrochemical based. For instance, bioplastics are a class of plastics derived from various sources of biomass, including food waste[11] and cellulose[12], and are designed to both avoid the use of fossil fuel derived chemicals, and biodegrade once discarded[13]. Researchers have designed materials made of lignin, an underused structural polymer present in plant material, to create carbon fiber materials[14, 15]— a 40,000 ton per year industry with applications ranging from automobiles to sports equipment[16]. These techniques, however, often pull biological feedstocks from existing waste streams, limiting the range of feedstocks available for offsetting fossil fuels.

Other researchers have focused on processes that instead rely on microbes, with the use of genetic engineering and synthetic biology. Synthetic biology is the study and redesigning of biological systems through genetic manipulation[17-20]. Synthetic biology has been harnessed to recreate time-delay circuits[21, 22], switches[23], and oscillators[24] with biological components, as well as optimize gene networks through directed evolution[25]. Previous work has used synthetic biology to create strains of *E. coli* that produce isobutanol and isopropanol[26] (products used as biodiesel additives), as well as strains of *S. cerevisiae* that produce ethanol out of plant polymers[27], such as xylose. Other researchers have engineered yeast to produce nanoparticles, such as cadmium sulfide quantum dots[28], that can increase the efficiency of solar cells up to 66%[29].

A deep understanding of the behavior of individual gene circuits within the broader context of global gene networks is critical for the realization of many synthetic biology applications. Individual cells can have hundreds to thousands of different genes which are regulated by intricately controlled pathways[30]. Gene regulation is made up of universal principles across both prokaryotes and eukaryotes[31], and although significant differences in complexity have been shown, the understanding of basic gene regulation structures are applicable across many cell types. Gene expression has been shown to occur primarily through two modes: constitutive gene expression and bursty gene expression[32, 33].

Constitutive gene expression is characterized by the Poisson-like accumulation of gene products[32]. The production of messenger ribonucleic acids (mRNA) and protein molecules is a discrete and stochastic process. Transcription can often be described by a Poisson process, where the time between mRNA productions follows an exponential distribution and times between production events are independent[34]. Additionally, the lifetime of a molecule is typically described by an exponential distribution. A large-scale study of over 400 genes in bacteria has suggested that constitutive gene expression is the predominate method of expression[32].

However, not all genes are constitutively expressed. Several studies have shown that in bacteria and yeast, many genes instead produce mRNA through transcriptional bursting[33, 35-38]. Transcriptional bursting is characterized by episodic periods of transcriptional activity, punctuated by lengths of time with no mRNA production (figure A.1). Recent studies have shown that bursty gene expression may in fact be the predominant form of gene expression[39]. While transcriptional bursting has been found in many genes and can have important consequences in expression, there is no consensus regarding its mechanistic source[40, 41]. Researchers have studied a variety of molecular sources to explain transcriptional bursting, including transcription factor binding and unbinding [42-44], promoter architecture (which can modify transcription kinetics)[36, 45, 46], the buildup of supercoiling[47, 48], and transcriptional re-initiation[33, 49, 50]. Figure A.2 shows many of the different factors shown to influence gene bursting. To characterize the different types of gene expression behaviors, models are used to simplify complex processes into frameworks that can be analytically or computationally examined.

1.2 Modeling Techniques for Gene Expression

The central dogma of gene expression states that genes code for the production of mRNA, which in turn code for the production of proteins[51, 52](figure A.3). The protein can then influence other genes, provide functionality for the cell, or be exported into the environment outside. Genes have a multitude of parts, including regulatory regions which repress or induce expression due to the binding of transcription factors[53, 54]. Most models, however, account for factors other than genes, mRNAs, and proteins in terms of effective rate constants[55].

Understanding and characterizing gene expression requires quantification of the numbers of mRNA and protein produced by the system. Simple modeling approaches for gene expression are deterministic and are formulated in terms of ordinary differential equations (ODEs)[56-58]. A simple deterministic model of gene expression is described by the following coupled differential equations for both transcription and translation,

$$\frac{dr}{dt} = \alpha_r(t) - (\gamma_r + \delta)r$$

$$(1.1)$$

$$\frac{dp}{dt} = k_P r - (\gamma_P + \delta)p$$

where r and p refer to the mRNA and protein concentrations, respectively; γ_r and γ_P are the decay rates for mRNA and protein; α_r and k_P are the production rates for transcription and translation; and δ is the rate of dilution due to cell growth. The steady state values for both of these products are

$$< r >= \frac{\alpha_R}{(\gamma_R + \delta)}$$

$$< p >= \frac{\alpha_R k_P}{(\gamma_R + \delta)(\gamma_P + \delta)} \tag{1.2}$$

These equations are useful in examining systems with large populations on the macroscopic level. ODEs describe the change in the concentration of each species over time in terms of the underlying chemical reactions. The system is assumed to be well-mixed, such that there are no concentration gradients or compartmentalization[59]. ODE models of reaction dynamics are useful in modeling the average behavior of systems because the ODEs use bulk reaction rates to characterize behavior. However, the total number of molecules of a given species is assumed to be sufficiently large so that stochastic effects are not important. These models are less useful for systems involving low molecule copy numbers, diffusion-limiting processes, and spatial inhomogeneities. These features can lead to stochastic fluctuations significantly affecting the dynamics important to gene expression. Due to the shortcomings of deterministic models, various stochastic methods that capture stochastic fluctuations have been developed.

The Brownian dynamics (BD) method incorporates stochastic influences, making it suited to describing low abundance transcription processes[60, 61]. Cellular species are modeled as particles diffusing in a uniform solvent such as water, which is treated implicitly as a random, stochastic force. The movement of a particle is influenced by the stochastic nature of the solvent and results in particles undergoing random walks[61]. In most BD models, molecular reactions can occur upon collision. The BD method is attractive because it accounts for the position of every molecule in the system and is also able to incorporate the stochasticity of reactions between few molecules. Additionally, BD is capable of modeling crowded systems by incorporating excluded volume particles into the system. However, modeling significantly crowded systems increases the computational cost of the model, as the cost scales with the number of collisions that occur and the smaller time steps needed to resolve them.

The chemical master equation (CME) also captures the dynamics of stochastic chemical reactions. The CME describes the time evolution of the probability that a system will be in a state $X(t) = x$ given an initial state $X(t_0) = x_0$ for some $t > t_0$. A state is defined here as the number of molecules of each species in the system. Changes in the state occur in discrete numbers when a reaction occurs. However, unless the reaction system is simple, the CME becomes difficult to solve analytically. Instead, a practical approach is to generate trajectories of the chemical master equation using the stochastic simulation algorithm (SSA), or Gillespie algorithm[62-64]. The Gillespie algorithm is widely used in the simulation of many gene models. Briefly, the algorithm is derived by posing the question: given the system is in a state $X(t) = x$ at time t, at what time $t + \tau$ will the next reaction occur, and which reaction will it be? The algorithm accounts for the inherent stochastic nature of gene reactions by modeling the time to the next reaction as an exponential distribution, and stochastically chooses the next reaction based on a reaction's propensity. The propensity quantifies the rate at which a specific reaction occurs, given the system is in some state $X(t) = x$. Spatial considerations are incorporated into the model by including propensities for molecules to diffuse through space, and the state of the system additionally contains information about the particle locations.

The Gillespie Algorithm is widely used in the field of computational biology[65]. For many biochemical systems, however, the computational cost of the Gillespie Algorithm can be prohibitively high[66]. Further algorithmic efficiency changes have been made to increase the computational speed of the simulation methods. The Next Reaction Method calculates a putative time (time a reaction would occur if no other reaction occurs first) for each reaction propensity and stores it in a dependency graph[67]. Propensities are only recalculated when they change due to a reaction. While storing the putative times for each reaction in a dependency graph incurs added computational costs, the efficiency of the algorithm is increased by carefully recalculating propensities only if they change, and by reusing putative times where appropriate.

Approximate accelerated stochastic methods have also been developed to speed computational times. The tau-leaping method[68] improves computational efficiency by advancing the simulation by predefined time steps. While the original algorithm asks “which reaction will occur in the next time interval τ,” the tau-leaping method instead asks “over the next time interval τ, now many times does each reaction occur?” By using larger τ values and approximating the number of times a given reaction occurs, computational resources are saved at the expense of simulation accuracy. The degree of accuracy tolerated by the method depends on an error parameter, which is a measure of the change in a single propensity over the τ leap. If a propensity changes more than is expected over a long time leap, then a shorter τ is chosen and tested until the error conditions are satisfied. In this thesis, we use the standard stochastic simulation algorithm.

A commonly used framework for modeling bursty gene expression is the random telegraph model, often referred to as the two-state model of gene expression[69]. The two-state model (figure A.4) is a stochastic model that represents the gene as a system that can transition randomly from an ON state, where gene expression occurs, to an OFF state, where no expression occurs. This model is common when describing bursty gene expression because it captures the features of experimental data independent of the mechanisms that cause bursting.

It is often difficult to directly observe the regulatory behavior of a gene in a cell. Experimental measurements are commonly collected from fluorescently tagged mRNA and protein molecules. Previous work has shown that measuring the fluctuations in molecular populations (noise) gives information on the behavior of the underlying gene circuit[70-72].

1.3 Noise Analysis in Gene Expression

Noise analysis is a critical tool useful for understanding gene expression and transcriptional bursting. Noise in gene expression refers to the random fluctuations associated with molecule populations over time. Fluctuations in molecular populations influence how certain genes are expressed[73], how individual cells determine phenotypes[74], and how cells choose directions in motility[75]. Noise originates from the synthesis and decay of molecules occurring in discrete numbers and at random times. Many factors influence noise[76], including cellular size, molecular concentrations, and the distribution of resources. Figure A.5 shows a number of molecular processes that influence the properties of noise in even a simple model of gene expression. Noise can generally be divided into two categories: extrinsic and intrinsic noise.

Extrinsic noise factors are independent of the gene and instead correspond to parameters associated with cells or environments[77]. Isogenic populations of cells are genetically identical, but still show significant cell-to-cell variations reflected by stochastic differences in the number of molecules[78, 79]. One source of extrinsic noise is cell growth, as individual cells at different stages of the cell cycle will have differences in size and expression[80]. Division introduces another source of extrinsic noise, since the division of a cell into daughter cells can lead to different concentrations of molecules in subsequent generations[77]. Resources used globally inside the cell, such as ribosomes, amino acids, and energy molecules, also introduce extrinsic noise due to sharing among many different regions of the cell at any given time.

Intrinsic noise refers to noise associated with the stochastic and discrete production of molecules. Even in a hypothetical system where the number, location, and activity of all molecules in two cells are identical, inherent stochasticity in random microscopic events will still cause noise. Intrinsic noise can be thought of as the extent to which two identical genes in identical environments fail to correlate[79]. Intrinsic noise in transcription influences processes including bacterial quorum sensing[81], eukaryotic cell differentiation[82, 83], and probabilistic fate determination[71]. Noise in gene expression has also been shown to control fate selection between active infection and proviral latency in an immunodeficiency virus type 1 (HIV-1) model system[39, 71, 84-86]. The characterization of extrinsic noise and intrinsic noise is an area of active research, and several methods have been detailed that propose to isolate intrinsic noise[32, 78, 87].

The measurement of noise in gene expression is useful because it allows for the indirect observation of transcriptional bursting. Bursting is often described through two parameters: burst frequency, a measure of how often bursts occur, and burst size, a measure of how many mRNA are produced in a single burst. Simpson et al. have shown in several papers that the parameters of transcriptional bursting can be inferred by performing frequency domain analysis on the time resolved fluorescence values of protein expression[39, 43, 80, 88]. Golding et al. previously estimated the bursting behavior of a gene through the use of fluorescent imaging of mRNA molecules[89]. By quantifying mRNA populations using single-molecule fluorescence *in situ* hybridization (smFISH), Golding et al. counted the number of mRNA in *E. coli* on a single-transcript level and measured the mean and variance of the number of mRNA per cell in order to characterize the burst parameters. Analyses of steady state and time-dependent measurements of both mRNA and protein populations provide insight into transcriptional bursting independent of mechanistic causes. However, the analyses rely on relatively simple models of gene expression. Several recent papers have shown that in some cases, the distribution of burst arrival times is not exponential[45, 90, 91]. Additionally, factors including macromolecular crowding and spatial confinement have been shown to change the stochastic behavior of gene expression systems. Due to these discrepancies, more sophisticated models must be considered.

1.4 Gene Expression and Spatial Effects

Many molecular mechanisms have been probed for their influence on bursting in gene expression. However, spatial considerations associated with crowding and confinement are often neglected, and have been explored only recently[41, 92]. Crowding refers to macromolecules that restrict the movement of other molecules through excluded volume effects, hydrophobic effects, or van der Waal interactions[93]. Macromolecular crowding of the intracellular space has been measured to be up to 30% by volume[94, 95], which has been shown to significantly alter the ability of particles to diffuse and react[96]. Crowding can additionally alter protein solubility[97], affect protein folding and stability[98], and change enzymatic activity[99]. Crowding effects on proteins are important because proteins are often regulatory factors in the expression of other genes. Confinement has been shown to further change transcriptional dynamics, causing bursty transcription through spatial correlations[59]. Confinement refers to the effects of boundaries, such as those that confine systems to small volumes, including organelles, nuclei, or cellular membranes. While crowding and confinement each influence gene expression individually, the combined interactions of crowding and confinement has been shown to collectively lead to larger effects[100].

Rein ten Wolde and coworkers have shown that spatial fluctuations and rapid rebinding of transcription factors significantly increase the measured noise in a gene network[101]. Additionally, macromolecular crowding was shown to enhance binding through changes to equilibrium constants[102]. Separately, work by Meyer et al. has also shown that highly crowded, static environments significantly alter the behavior of gene expression, and are capable of driving what was originally a constitutive processes to behave in a bursty manner[59]. However, these works have either considered crowding only though rescaled rate parameters, or have considered only static spatial crowding effects on gene expression, neglecting dynamic crowding influences that may occur. It has also been shown that confinement and altered protein mobility can markedly influence the stochastic behavior of other biochemical reaction networks[103, 104]. Kulkarni et al. recently detailed an analytical method describing a generalized stochastic model in which the arrival rate of bursts does not need to follow a Poisson distribution[105]. While the generalized model applies to a wider range of gene expression influences, spatial considerations have not yet been directly examined. Work regarding spatial influences on gene expression have addressed some spatial aspects, but specific questions regarding confinement and dynamic crowding have yet to be explored.

1.6 Scope of the Dissertation

This work uses stochastic models to investigate how spatial features such as crowding and confinement impact gene expression. The topic is examined by analyzing models of gene expression under confined and crowded conditions using computer simulations and noise analysis. The following questions are explored. In confined, cell-free expression chambers, how does confinement change transcriptional bursting? Additionally, can the models used to describe the behavior in these cell-free chambers be related to the behavior measured in cellular systems? A spatial model has subsequently been developed that considers a simple transcriptional gene system in which spatial influences of crowding and confinement are explored and analyzed with regards to their behavior in noise space. For these simulations, the question of whether spatial effects are captured using simpler, two or three-state models is considered. Any differences are further analyzed to determine how spatial properties effect noise. Evidence is provided that the consideration of spatial effects are an important component of future models, and that spatial effects are valuable to the understanding of many underlying genetic mechanisms.

1.7 Organization of the Dissertation

Chapter 2 details the background needed to understand the computational design choices, as well as computational information required for the understanding of subsequent chapters. This chapter additionally details the methods used for the subsequent noise analysis methods that are used on the data collected from both experiments and simulations. Chapter 3 focuses on work done with experimental cell-free reaction chambers, where the effects of confinement on resource sharing and expression are examined in detail. Several models are used to explore the details associated with the noise gathered from these experimental chambers, each of which are described in detail to show how resource sharing can have profound changes in the burst behavior within these chambers. The focus of chapter 4 is on the transcriptional burst properties of mRNA populations generated from the spatial simulations, focusing on how well a simple two-state model can reproduce the results from a spatially resolved model. Changes in crowding and confinement geometry influence burst properties of the system, and comparisons between the spatial model and various parameterizations of two-state and three-state models reveal the intricate way spatial effects influence transcriptional bursting through noise measurements. The behavior in a spatially resolved system is shown to be poorly captured using simpler two-state models of gene expression. The bursting dynamics of the spatial system compared to a two-state models indicate that ignoring the spatial properties of crowding and confinement ignores many of the subsequent effects that present themselves in the noise and burst properties of the system. Finally, chapter 5 concludes these findings and provides additional avenues for future work, giving emphasis on specific applications, such as in HIV-1 behavior and treatment and the examination of efficient usage of limited resources in gene expression.

2. Background and Methodology

Several models of gene expression are presented in this work that capture the stochastic nature of transcription and translation. Each is useful in exploring a different facet of gene expression, including resource sharing and allocation, and the effects of spatial crowding and confinement on bursty expression. Here, the fundamentals and background needed to understand the models and how noise is used to probe gene expression and bursting are described. Specifics regarding any particular model or experiment are provided in individual chapters.

2.1 Gillespie Algorithm Background and Overview

The simulation of chemical reactions can be viewed through a simple problem statement: Given a fixed volume containing a spatially uniform mixture of chemical species interacting through a number of reaction channels, and given the number of molecules of each species at an initial time, what is the population at any later time? The stochastic framework for solving this problem is known as the Chemical master equation (CME) and was primarily developed with the work of McQuarrie[106]. Assume a volume *V* contains N chemical species undergoing M possible reactions. The state of the system is defined by specifying $X_i(t)$, the number of molecules of species S_i at time t, where i = 1 . . . N. The state evolves in time through M reaction channels R_μ, where μ = 1 . . . M. The CME is an equation that describes the time-evolution of the probability that $X(t) \equiv (X_1(t),\ldots,X_N(t))$ will be equal to $x = (x_1,\ldots,x_N)$, given $X_0(t_0) = x_0$ for some $t > t_0$. The probability is written as $P(x,t|x_0,t_0)$.

The master equation can be written as

$$\frac{\partial}{\partial t}P(x,t|x_0,t_0) = \sum_{\mu=1}^{M}[\alpha_\mu(x-v_\mu)P\big(x-v_\mu,t|x_0,t_0\big) - \alpha_\mu(x)P(x,t|x_0,t_0)] \qquad (2.1)$$

where α_μ is the propensity function of an R_μ reaction, and $v_\mu \equiv (v_{1\mu}, \ldots , v_{N\mu})$ is the change in the state caused by one R_μ event. The first term in the sum is the probability that the system is one R_μ reaction removed from state *x* at time *t* and then undergoes an R_μ reaction in (*t*, *t* + *dt*). The propensity of a reaction $\alpha_\mu(x)$ is defined as the probability, given X(t) = x, that an R_μ

reaction will occur in the system in the next time interval [t,t+dt). While exact, the CME is not analytically solvable for the vast majority of systems.

The Gillespie algorithm is an exact method of generating simulation trajectories consistent with the chemical master equation[62]. To determine the time evolution of a system, the simulation generates two pieces of information to evolve the system forward in time: when the next reaction will occur, and which reaction it will be. The Gillespie algorithm introduces a new function P(τ, μ | x, t) which is defined as the probability that given a state X(t) = x at time t, the next reaction to occur will occur in the infinitesimally small time interval (t + τ, t + τ + dt), and will be reaction R_μ. With this, P(τ, μ | x, t) is defined as

$$P(\tau,\mu|x,t)=\begin{cases}\alpha_\mu \exp(-\alpha_0\tau) & \text{if } 0\le\tau<\infty \text{ and } \mu=1\ldots M\\ 0 & \text{Otherwise}\end{cases} \tag{2.2}$$

where

$$\begin{gathered}\alpha_\mu \equiv h_\mu c_\mu\\ \alpha_0\equiv\sum_{j=1}^{M}\alpha_j\equiv\sum_{j=1}^{M}h_j c_j\end{gathered} \tag{2.3}$$

Here, α_μ is the propensity (rate) that reaction R_μ will occur over the next time step τ, with α_0 equal to the total propensity of all M reaction channels in the system. To calculate the propensity of each reaction R_μ, a function h_μ is defined as the number of distinct R_μ reactant combinations available in the state x = $x_1 \ldots x_N$. That is to say, if R_2 is of the form $S_1 + S_2 \rightarrow \ldots$, then the number of distinct reactant combinations is equal to x_1x_2. h_μ takes different forms based on the order of the reaction R_μ: for unimolecular reactions ($S_1 \rightarrow \ldots$), h = x_1, for bimolecular reaction of the form $S_1 + S_2 \rightarrow \ldots$, h = x_1x_2, and for bimolecular reaction of the form $2S_1 \rightarrow \ldots$, h = $\frac{1}{2}x_1(x_1 - 1)$. c_μ is the stochastic rate constant of the reaction R_μ, defined such that $c_\mu dt$ is the probability that reaction R_μ will occur in *V* between a particular pair of reactant molecules in the next infinitesimal time interval *dt*. The stochastic rate constant is closely related to the more common reaction rate constant k_μ, the primary difference being the reaction rate constant normally relates to concentrations rather than total numbers of molecules[62]. Additional differences are present for reactions containing 2 or more identical species, but because no reactions of those type are considered here, the differences will not be detailed. The constant *c* has units of inverse time.

The expression for P(τ, μ | x, t) can be determined at all times knowing only the reaction constants and the current number of molecules of each species in the system. The Gillespie simulation method generates two uniformly distributed random numbers to obtain τ and μ:

$$\begin{gathered} \tau = \frac{1}{a_0} \cdot \ln \frac{1}{r_1} \\ \sum_{v=1}^{\mu-1} \alpha_v < r_2 \alpha_0 \leq \sum_{v=1}^{\mu} \alpha_v \end{gathered} \tag{2.4}$$

Here, r_1 and r_2 are two independent, uniformly distributed random numbers over the interval (0, 1]. A proof of why these equations give the proper probability distribution has been given by Nitzan and Ross[107]. The first equation generates τ values that are exponentially distributed and weighted by the total propensity of the system. The second equation generates the next reaction R_μ by specifying μ as the first integer to satisfy the equation.

For a chemical reaction at an initial state and time, the algorithm is iterated for all future times until the time limit set in the simulation is exceeded. Because the algorithm deals in absolute numbers of particles, when storing the results at regular time intervals, instead of at absolute time steps, the population of a species at an arbitrary time point is equal to the population at the time of the last reaction to occur. That is, if the number of molecules is being stored at 10 second intervals, and the last reaction to occur did so at 8 seconds, then the number of molecules at the 10 seconds is the same as the number of molecules at 8 seconds.

The implementation of the Gillespie algorithm is outlined as follows:

- Initialization:
    - The initial populations of all species in the system are set.
    - The total time of the simulation is defined.
    - The reactions of the system are defined.

- Monte Carlo step:
    - The propensities of all reactions are updated based on the current populations and rates.
    - One random variable is generated that determines the time to the next reaction
    - One random variable is generated that determines which reaction takes place.
- Update:
    - The system is updated according to the previous reaction step and the time is moved forward. The number of members in any given species changes based on the reaction that occurs.
- Iterate:
    - Continue running Monte Carlo steps and updating the simulation until the simulation time has been exceeded.

2.3 Noise Analysis Framework

In a Poisson process, the timing between events is independent and exponentially distributed. For mRNA production, it is commonly assumed that constitutive gene expression of mRNA is a Poisson process. However, due to transcriptional and translational bursting, the production of mRNA and protein can deviate significantly from a Poisson process. For this reason, the noise analysis framework is important in examining the underlying behavior of gene expression.

Noise in a given gene expression system is defined as the stochastic fluctuations in molecule abundance values, and can be characterized using the noise framework developed previously by Cox et al.[70]. Noise can be analyzed in noise space, which is a three dimensional space consisting of three components: average species abundance, noise magnitude, and autocorrelation time. Noise magnitude is described using the coefficient of variation squared (CV^2), which is a measure of the dispersion of a probability distribution, and is defined as

$$CV^2 = \frac{\sigma^2}{\mu^2} \tag{2.5}$$

where σ^2 is the variance of the signal and μ is its mean value. CV^2 is useful because it allows for the comparison of noise from signals whose abundance values vary over orders of magnitude. Autocorrelation describes the correlation between values of the trajectory at different times. Noise autocorrelation time is defined as the time when the autocorrelation is half its initial value ($\tau_{1/2}$).

2.3.1 Using Noise to Calculate Burst Frequency and Burst Size

The behavior of a gene is often not measured directly, but is inferred through the measurement of fluorescence values of tagged mRNA molecules or proteins. Measurements of many different cells are taken, either over time through optical microscopy or over large populations through flow cytometry. As described in the previous chapter, noise in transcriptional bursting is characterized primarily through burst frequency, defined as the number of mRNA bursts per unit time, and burst size, defined as the average number of mRNA produced in a single burst. Golding et al. previously showed how burst parameters could be estimated using fluorescent imaging of mRNA molecules[89]. By quantifying mRNA populations using single-molecule fluorescence *in situ* hybridization (smFISH), they were able to count the number of mRNA in *E. coli* on a single-transcript level. From these counts, they were able to measure the mean and variance in the number of mRNA in individual cells in order to generate burst parameters. Suter et al. were able to estimate the bursting parameters for the random telegraph model using abundance measurements of mRNA and proteins over time[45]. Their measurement revealed that the kinetics of bursting were highly gene specific.

The noise analysis framework previously discussed allows for another way of examining transcriptional burst behavior using time-dependent measurements of protein fluorescence values. Described in Dar et al., the three dimensional noise space can be decomposed into three two-dimensional projections[39]. Changes in burst parameters were shown to be visible through the CV^2 and abundance projection of the three-dimensional space, since the distribution of data can give information on bursting in gene expression. Noise analysis assumes the gene can be adequately described by the random-telegraph model (shown previously in figure A.4). Bursting can be visualized as a pulse train, where molecular production occurs only when the pulse is nonzero (figure A.6). The pulse train has three main parameters. The first is the frequency, a measure of how often pulses occur over time. The second is burst length, a measure of how long a pulse is. Finally, burst height measures the rate of production. For measurements of protein populations, the amplitude is a measure of both the transcriptional and translational rate. Transcriptional burst size is defined as the number of mRNA produced during a given burst, while translational burst size is measured as the number of proteins produced from each mRNA. For the analysis, it is assumed the translation rate, protein decay rate, and mRNA decay rate are constant[39].

Transcriptional burst parameters can be calculated by solving for the autocorrelation function and the noise magnitude (CV^2). For a system following the two-state model of gene expression, and under the assumption that the transcriptional burst size is greater than or equal to 1, the autocorrelation of the noise is[39]

$$\Phi(\tau) \approx \frac{k_m O b}{\gamma_d} b e^{(-\gamma_d \tau)} + \left(\frac{k_m O b}{\gamma_d}\right)^2 \frac{(1-O)}{Ok} \left(\frac{\gamma_d}{\left[1-\left(\frac{\gamma_d}{k}\right)^2\right]} e^{(-\gamma_d \tau)} + \frac{k}{\left[1-\left(\frac{\gamma_d}{k}\right)^2\right]} e^{-k\tau} \right) \tag{2.6}$$

where k$_m$ is the transcription rate during a burst, O is the fraction $\left(\frac{k_{ON}}{k_{ON}+k_{OFF}}\right)$ of time the gene is in the on state , b is the translational burst size, γ_d is the dominant decay constant (usually taken as the protein decay constant as for most systems, $\gamma_p \gg \gamma_m$), and k is the sum of k$_{ON}$ and k$_{OFF}$. The average steady state protein population is

$$\langle P_s \rangle = \frac{k_m O b}{\gamma_d} \tag{2.7}$$

and the noise magnitude is[39]

$$CV^2 = \frac{\Phi(0)}{\langle P_s \rangle^2} \approx \frac{b}{\langle P_s \rangle} + \frac{(1-O)}{Ok}\left(\frac{\gamma_d}{\left[1-\left(\frac{\gamma_d}{k}\right)^2\right]} + \frac{k}{\left[1-\left(\frac{k}{\gamma_d}\right)^2\right]}\right) \tag{2.8}$$

The first term on the right hand side of equation 2.19 is referred to as the shot-noise[43], and may dominate at high ON fraction (constitutive expression), low protein population, or if $k \gg \gamma_d$ (fast switching between gene states). On the other hand, the second term, referred to as burst noise, may be dominant at low ON fraction, high protein population, or if $\gamma_d \gg k$ (slow switching between gene states). Under conditions where shot noise is dominant, CV^2 varies inversely with protein abundance and is indistinguishable from constitutive expression in noise space. In contrast, CV^2 shifts upward when burst noise is dominant, varying inversely with on fraction and the kinetics of gene switching (k_{ON} + k_{OFF}).

Kulkarni et al.[105] describe the properties of transcriptional bursting in a more general framework, where burst size and burst frequency can be described through queuing theory without the assumption of Poissonian distributions from the two-state model. Queuing theory is the mathematical relationship associated with waiting lines formed by customers who arrive by some stochastic process and remain in the system until serviced. Gene expression can be described through this theory by assigning mRNA and protein molecules as customers, where bursts of mRNA or protein production are analogous to batches of customers arriving in the system. mRNA and protein molecules leave the system when they degrade.

The Kulkarni model describes the arrival of mRNA in bursts as an arbitrary arrival function f(t). mRNA decay, protein production, and protein decay are subsequently modeled similarly to the two-state model (figure A.7). Because previous work has shown that mRNA burst distributions are geometric[108], the model focuses on geometrically distributed bursts for both mRNA and protein. For this model, the steady state of the mRNA population is given by

$$\langle m_s \rangle = \frac{b_f}{\gamma_m} B \tag{2.9}$$

Additionally, the protein steady state population is given by

$$\langle p_s \rangle = \frac{b_f B b}{\gamma_p} \tag{2.10}$$

where b_f is the mean arrival rate of mRNA bursts, γ_m is the decay rate of mRNA, B is the mean mRNA burst size and b is the mean protein burst size from a single mRNA. Under the general queuing theory framework, the noise in the mRNA steady state population is

$$\frac{\sigma_{m_s}^2}{\langle m_s \rangle^2} = \frac{1}{\langle m_s \rangle} + \frac{\gamma_m}{b_f} + \frac{\gamma_m}{2b_f}\left[K_g(\gamma_m) - 1 + \frac{\sigma_{m_b}^2}{B^2} - \left(1 + \frac{1}{B}\right)\right] \tag{2.11}$$

where $\sigma_{m_s}^2$is the variance in the mRNA steady state population, $\sigma_{m_b}^2$ is the variance in the mRNA burst distribution, and $K_g(\gamma_m)$ is the gestation factor, which is defined as

$$K_g(\gamma_m) = 1 + 2\left[\frac{f_L(\gamma_m)}{1 - f_L(\gamma_m)} - \frac{b_f}{\gamma_m}\right] \tag{2.12}$$

where $f_L(\gamma_m)$ denotes the Laplace transform of arrival time distribution of mRNA bursts. The gestation factor encodes information on the arrival process of the bursts. For Poisson arrivals, $K_g(\gamma_m) = 1$. Under the assumption of Poisson arrivals and geometrically distributed bursts, the equation for CV^2 reduces to

$$\frac{\sigma_{m_s}^2}{\langle m_s \rangle^2} = \frac{1}{\langle m_s \rangle} + \frac{B\gamma_m - \gamma_m}{b_f B} \tag{2.13}$$

Solving for burst size and burst frequency results in the following equations.

$$B = CV^2 \langle m_s \rangle \tag{2.14}$$

$$b_f = \frac{\gamma_m}{CV^2}$$

Relationships in burst parameter changes in noise space are described in figure A.8. The set of equations above provides a concise way to relate the noise and abundance measured in a population with the burst behavior of the system. In the CV^2 and abundance plane, changes in burst frequency correspond with CV^2 changing with mRNA abundance. The change follows an inverse relationship, as the transition to higher abundance follows from more mRNA production due to an increase in bursts over a shorter amount of time. Lateral changes to different abundance values with no change in noise magnitude are instead indicative of changes in burst size.

The translational burst parameter relationships can also be derived for measurements of protein populations. As above, under the assumption that $\gamma_p \gg \gamma_m$, the noise in protein population is defined as[105]

$$\frac{\sigma_{p_s}^2}{\langle p_s\rangle^2} = \frac{1}{\langle p_s\rangle} + \frac{\gamma_p}{b_f} + \frac{\gamma_p}{2b_f}\left[K_g(\gamma_p) - 1 + \frac{\sigma_{m_b}^2}{B^2} - \left(1 + \frac{1}{B}\right) + \left(\frac{\sigma_{p_b}^2}{b^2} - \left(1 + \frac{1}{b}\right)\right)\frac{1}{B}\right] \tag{2.15}$$

where $\sigma_{p_s}^2$ is the variance in the protein population, $\langle p_s\rangle$ is the mean steady state protein value, γ_p is the protein decay rate, $\sigma_{p_b}^2$ is the variance in the protein burst distribution, and $\langle p_b\rangle$ is the protein burst size. Again, assuming Poisson arrivals and geometric distributions for the bursts, solving for the burst size and burst frequency of transcriptional bursting results in

$$B = \frac{CV_p^2\langle p_s\rangle}{b} \tag{2.16}$$

$$b_f = \frac{\langle p_s \rangle \gamma_p}{Bb}$$

The equations relate the noise found in a protein population to the transcriptional burst behavior in the system. In the systems being modeled, unless otherwise noted, it is assumed that the translational burst size b is large (at least 100 proteins per mRNA transcript), and is relatively constant[39, 80]. The relationship between protein noise magnitude and burst frequency is consistent with the relationship between mRNA noise magnitude and burst frequency, with noise following an inverse relationship with abundance under constant burst size. A full derivation of these equations can be found in Kulkarni et al.[105].

While the equations for burst size and burst frequency (eq. 2.25) from Kulkarni et al. are based on a different initial description of the simple genetic expression model, the noise analysis framework set out by Dar et al. assumes that the two-state model of gene expression is sufficiently bursty ($k_{OFF} >> k_{ON}$) such that each individual pulse of activity can be represented as an impulse function[39, 105]. Comparisons between the simplified equations of noise magnitude in protein abundance derived though queuing theory (Kulkarni el al.) and through the random telegraph model (Dar et al.[39]) show that both methods result in the same relationships between noise magnitude and burst parameters.

2.3.2 Noise Processing

A gene expression trajectory (i.e., a time history of fluorescence from cells or cell-free expression chambers) is composed of deterministic components – background fluorescence and gene expression as described by the deterministic ODEs – and a stochastic component that captures the noise in gene expression including those that emerge from bursting[71]. Isolation of the stochastic component of raw measurements requires the determination of: (1) background signals (usually autofluorescence from cell components or cell extract) present in each trajectory; (2) the deterministic general trend of expression across an entire population of cells or cell-free reaction chambers; and (3) how strongly the general trend couples into each individual trajectory. With this information, the noise of each trajectory may be determined. Each of these 3 steps is described below.

For simplicity, throughout this description, the term "cell" will be used but should be interpreted to mean cell, cell-free reaction chamber, or simulated trajectories. Not all steps of the full noise analysis framework are necessary for all experiments or models presented in this work. However, the noise processing steps are presented in their entirety here, and any omissions or modifications used for specific models are detailed in the appropriate section. A more detailed analysis of these noise methods are shown in several other papers[39, 71]. The process is shown in general in figure A.9.

The first step is background suppression, which removes signals not related to expression (i.e. autofluorescence). This is accomplished by removing the background signal such that the expression level is 0 at time 0. Separate experiments have shown that for the measurements considered here, the initial fluorescence values are composed entirely of autofluorescence and not indicative of gene activity. Simulations are simply initialized at 0 abundance.

The deterministic general trend that is seen across the population, A(t), is estimated as the mean value of all traces within an experiment or simulation:

$$A(t) = \frac{1}{M}\sum_{m=0}^{M} I_m(t) \tag{2.17}$$

where M is the number of trajectories in the data set and $I_m(t)$ is the gene expression trajectory which is measured over time for each trace m = 1, . . ., M.. The assumption inherent in this calculation is that all trajectories display similar general trends (i.e. differing only in a gain factor (see below)). For a large enough population of cells, the noise will average out and A(t) will provide a reasonable estimate of the deterministic general trend seen in the population.

The third step requires the generation of a gain factor, which quantifies how strongly the general trend couples into each trace. Because the underlying expression levels among cells can be variable, the gain attempts to remove as much of the deterministic component as possible by minimizing the cross-correlation between the isolated noise trace and the general trend. Several factors could cause an individual trace to scale differently from the average behavior captured in the general trend, including small variations in the interior of a cell, slight differences in volume, or different numbers of reaction components[71]. This process is described by the equation

$$N_m(t) = I_m(t) - g_m \cdot A(t) \tag{2.18}$$

where g_m is the gain factor, I_m is the individual trace, and N_m is the noise associated with that trace. The gain values are calculated by minimizing the cross-correlation value between $N_m(t)$ and A(t), which is

$$\left|\sum_{t=0}^{t=T} N_m(t) \cdot A(t)\right| \tag{2.19}$$

Once the deterministic portion of an individual trace is removed using the gain values calculated, only the noise component remains. The autocorrelation is then calculated from each of the noise traces

$$R_m(\tau) = \frac{1}{T}\sum_{t=0}^{T-\tau}(N_m(t) - \mu_m) \cdot (N_m(t+\tau) - \mu_m) \tag{2.20}$$

where R_m is the biased autocorrelation of the noise of trace m[109], τ is the time-lag, μ is the mean of trace m, and T is the number of samples in the trace. The steady state used for the calculated CV^2 for transient trajectories was taken as:

$$< p_m >= I_m(\max\{t\}) - I_m(0) \tag{2.21}$$

where $I_m(\max\{t\})$ is the steady state value of a given trace (last time point) and $I_m(0)$ is the starting fluorescence of all the traces of a given chamber size. Because experimental data is limited to transient measurements over short times, the last point of a trajectory is used as a substitute for the steady state value. For analysis of cells measured over steady state, the mean value at steady state is used. The variance needed to calculate CV^2 is taken from the autocorrelation at 0 lag time, which is a property of the autocorrelation function.

While it is important that both the simulation data and the experimental data should be subjected to identical noise analysis methods, it is also critical to understand why each step is taken, so that steps that are unnecessary for simulation data are not used, as they may result in unwanted removal of noise. For example, the removal of the general trend is included because data from cell-free chambers is unable to be taken at steady state, meaning the trend present in the population (i.e. due to growth) needs to be removed. The inclusion of the gain factor is also a way of removing unwanted signal components, as it has been shown that in cell populations, the general trend can have different degrees of coupling into the each individual trace[71]. However, simulations have the benefit that many of these external deterministic components are ignored. As such, applying the entire course of noise processing steps on the simulation data overestimates the magnitude of the deterministic component of the signal, removing some portion of the stochastic noise component from simulations. Both the full noise processing method and a simplified noise processing method (which excluded the search for a gain term) have been applied to sample simulation data as a control, showing that the noise is indeed overcompensated for. In short, the simplified method does not calculate a gain factor, and instead measures the abundance trajectory at steady state, removes the mean, and calculates the autocorrelation to calculate the variance and CV^2 values.

2.4 Major Model Assumptions

When models are developed for any cellular or experimental system, assumptions must be made so that resulting simulations are tractable. The two-state, random telegraph model has several inherent assumptions. In many gene circuits, when highly active transcription is not occurring, there is still some amount of basal transcriptional activity[80]. This means that even in the OFF state, some amount of mRNA is being produced, albeit at a much slower rate. The two-state model simplifies transcriptional activity and creates an OFF state where absolutely no transcription takes place. The two-state model also makes no mechanistic assumptions on the underlying structure of the bursty gene, meaning any number of causes could be responsible for the transition from the ON state to the OFF state, and vice versa. Any complicated factors, such as transcription factor binding or chromatin remodeling, are left out in order to keep the model as simple as possible. Transcription and translation are considered instantaneous events in the two-state model, even though the production of mRNA and proteins requires a finite amount of time.

Additional assumptions are inherent in the Gillespie algorithm. Rate constants are based upon experimental measurements, meaning complicating factors such as diffusion are implicitly included in the measured reaction rates. The introduction of spatial considerations introduces additional assumptions, and simplifications must be made such that the complex cellular system being examined can be analyzed in a computationally tractable manner. One simplifications was the choice to represent the space as 3-dimensional cubic lattice instead of continuous space. This is standard practice when simulating spatially resolved systems with the Gillespie algorithm and correctly accounts for diffusive properties over sufficiently long time and length scales. The lattice simplification also removed the issues associated with spatial proximity and particle reactions. In continuous space, the rate at which two particles react can be strongly dependent on their distance from each other. Instead of having to continuously recalculate the reaction rates of particles based on changes in their distance, it is simpler and faster to consider only reactions that occur when the two particles share the same lattice site.

Crowders were assumed to occupy a single lattice site and interact with other particles through excluded volume effects. I.e., no other particles could move into the same lattice site as a crowding molecule. No other interactions, such as hydrostatic or van der Waal interactions, were considered in this model. More complex macromolecular crowding interactions, such as aggregation or polymer chain interactions, were also neglected.

Assumptions were also made during the development and use of the noise analysis methods. One major assumption was that of ergodicity. A process is said to be ergodic if the mean value of the stochastic process is equal for a single sample of an infinite amount of time and an infinite ensemble of samples over limited time. This property is necessary to the noise framework analysis because experimental methods cannot follow a cell over infinite time: they must make the assumption that looking over many cells over a limited time will be equivalent. Secondly, a non-ergodic process cannot be analyzed using the previously described noise analysis framework. It was also assumed that the regime in which the system was observed was within the valid bounds of the noise framework equations, namely, that the system was highly bursty, with well separated bursts ($k_{OFF} >> k_{ON}$).

3. Modeling and Noise Analysis of Confined Cell-free Chambers

The consequences of confining gene expression to various cell-free reaction chamber volumes and the consequences of resource sharing on transcriptional burst dynamics are explored in this chapter. Resources include any molecules required for the production of mRNA or protein in a gene expression system, including polymerases, ribosomes, and amino acids. Cell-free systems are *in vitro* tools that incorporate molecular expression machinery or structures from a cell into synthetic frameworks, allowing the study of specific reactions in the absence of confounding cellular components. Experimental cell-free chambers were fabricated and measured by Patrick Caveney and Sarah Norred. Subsequently, the measured flourescence data was analyzed using the the noise framework outlined previously. Models were then developed to help explain the transcriptional burst behavior measured in the cell-free expression systems.

3.1 Synthetic Exploration of Resource Sharing

Resource sharing is rarely considered when exploring the molecular processes associated with gene expression bursting. Many molecular processes have been shown to control transcriptional bursting (detailed previously in the introduction). However, the focus on molecular mechanism effects is often limiting in terms of scope. In a system in which a common reservoir of resources is shared among many expressing genes, it seems likely that the global activity of the genes would depend on the size of the reservoir and the spatial distribution of the common resources available.

Gene expression has been studied using various experimental techniques[46, 49, 78], both in cells and in bulk cell-free systems. Cell-based platforms are advantageous due to the ability to observe function within its natural context. However, it is difficult to isolate and manipulate specific parameters such as confinement independent of other cellular processes, including growth, cell division, and global gene expression.

In vitro reaction chambers provide a platform to isolate specific mechanistic effects of gene expression from confounding cellular processes [110, 111]. Cell-free protein synthesis (CFPS) systems have been used to observe gene expression bursting[48]. CFPS systems have been used to study noise in gene expression through the use of cellular-scale microfabricated arrays of reaction chambers as well [112, 113]. Here, microfabricated CFPS reactors and gene expression noise measurements are used in combination to explore gene expression bursting and resource sharing in well-controlled and easily manipulated environments. Figure A.10 details the resource-sharing environment in the cell-scale reaction chambers. In a confined system with limited resources, multiple genes pull from the reservoir in a time dependent manner. Bursts of gene activity correspond to short periods of high resource utilization (indicated in the figure as a change in the resource utilization heat map). However, no resources are required for long periods of gene inactivity.

This chapter focuses on the study of cell-free gene expression in synthetic reaction chambers under different resource sharing scenarios (figure A.11). Measurements of gene expression patterns were completed while the number of genes and the size of the resource pool were increased proportionally, either through the compilation of individual chambers (top resource sharing scenario in figure A.11-A), or through an increase in single chamber size (bottom resource sharing scenario in figure A.11-A). Under different resource sharing scenarios, the experiments aimed to determine if the properties of gene bursting were dependent on the global distribution of resources available (figure A.11-B). Cell-free reaction chambers were populated with cell-extract and observed over time using microscopy. Protein fluorescence time traces were collected and analyzed using the noise framework presented in the previous chapter. Modeling methods were then created and used to describe the behavior of the underlying gene regulatory system and resource sharing process. Finally, similar models were applied to transcriptional and translational burst size measurements from *E. coli* to determine whether resource sharing could explain the degree to which coupling between transcriptional and translational burst size occurs.

3.2 Cell-Free Chamber Fabrication and Analysis

Cell-free expression chambers were fabricated using soft lithography techniques, which are described in detail in Norred et al.[113]. In short, chambers were constructed out of polydimethylsiloxane (PDMS) and fabricated on a flexible, actuatable membrane suspended above microfluidic channels (figure A.12-A). All chambers were cylindrical in shape, with a height of 5 µm and a range of diameters from 2 µm to 10 µm. The diameters corresponded to volumes from 15 fL to 400 fL. The CFPS reaction raw extract was mixed with Enhanced Green

Fluorescent Protein (EGFP) coding pET3a plasmid, and was loaded into the reaction platform by flowing it into the microfluidic channel using pressurized nitrogen. The membrane was then actuated with deionized water, which forced the flexible membrane onto the cylinders, sealing the chambers and creating an easily defined reaction volume. Imaging of these chambers began shortly after actuation, and within minutes of the CFPS mixture being activated through the addition of the plasmid, which meant that a well-defined t = 0 could be measured. This property allowed for experimental chambers that were measured on different days to be directly compared.

Measurements were taken every 3 minutes for an hour of the total EGFP fluorescence for each individual chamber (figure A.12-B). The time course average of all 119 chambers showed rapid increase in fluorescence initially, followed by a slower rate of fluorescence at longer times. The measurements were similar to bulk reactions, proceeding at a slightly higher rate as noted elsewhere[114, 115]. Because the protein decay rate was much longer than the measured window of time, the falloff in measured fluorescence was not due to equilibrium expression and decay of the protein, but was instead more consistent with resource limitation. Reduction in expression at long times within synthetic, cell-free systems has been previously noted[112]. The shape of the transient and the variation in the measured fluorescence values was similar to cellular experiments[72]. To test the volume effects on resource sharing and gene expression, 2 µm, 5 µm, and 10 µm diameter chambers were fabricated and tested.

Time traces from each of the experimental chambers were subjected to the noise processing framework described earlier (figure A.12-C). Due to the transient nature of these experimental traces, noise was extracted by removing the gained general trend from each of the individual fluorescence trajectories. The general trend was taken on a per-day bases, as chambers captured on different days showed different trends which needed to be removed individually. Additionally, a gain factor was used to modify the degree to which each general trend coupled into a given experimental trace, and was found by minimizing the cross correlation between the general trend and the individual trace. The magnitude of the noise was quantified using the square of the coefficient of variation (CV^2). Because these chambers were transient in nature and never reached steady state, the “mean fluorescence value” used for the calculation of CV^2 was taken as the final time point fluorescence value of each time trace after background fluorescence was removed. The measured fluorescence for a given time trace was taken as the sum of the intensity for all pixels inside an experimental chamber, and was therefore a measure of the protein abundance for each chamber (as opposed to the mean value of each pixel within a given chamber, which is a measure of concentration). Rigorous testing on the size of the region of interest (ROI) and its effect on the measured fluorescence intensity revealed relative insensitivity to small changes in ROI near the wall of the chamber. As such, a single ROI region was used for all chambers of a given size.

3.3 Chamber Experimental Results and Discussion

The expression of 2 µm chambers are examined first. Figure A.13 shows the CV^2 vs protein abundance plot for individual 2 µm chambers. Each small triangle represents the CV^2 and protein abundance value for a single 2 µm chamber. The mean abundance over the ensemble of individual chambers was ~$2x10^4$ arbitrary units (AU) and the mean CV^2 was ~10^{-3}, as noted by the large triangular marker. Initially, to investigate the effects of resource sharing on expression, composite chamber were created by summing the fluorescence values from a number of individual 2 µm chambers, which varied from combinations of 2 to 6 chambers. For these composite chambers, no set of chambers shared more than 4 chambers in common with any other set, which was meant to minimize correlations between composite chambers that were identical in composition. The composite chambers serve as an illustrative case where resources are manually separated into distinct volumes which are not shared. The results from randomly chosen composites of 2 µm chambers are shown in figure A.13 as empty yellow triangles. Each empty triangle represents the mean value of 40 composites of a given number of chambers. As the number of chambers in a composite is increased, the measured sum value of the CV^2 decreases inversely with abundance (dotted black line), which is consistent with the idea that each chamber is an independent noise source.

To contrast with the composite chambers, larger chambers were measured at higher chamber diameters, such that a larger volume of reagents was subjected to a single proportionally large pool of shared resources. The results are compared in figure A.14-A. The three chambers sizes are presented in different colors (yellow for 2 µm, blue for 5 µm, and green for 10 µm chambers). The larger, darker colored symbols represent the geometric mean of each distribution. Within each chamber size, there is a clear inverse relationship between relative fluorescence and CV^2. Fluorescent abundance for the 2 µm individual chambers, for example, varied in value over about one order of magnitude, or from 10^4 to $8*10^4$ AU, while CV^2 values ranged over 1.5 orders of magnitude, from $3*10^{-4}$ to 10^{-2}.

The dimensions of the chambers were set such that a single 5 µm chamber contained close to the same volume as six 2 µm chambers, which could therefore be directly compared. For the 5 µm chambers, it was shown that the CV^2 of the chambers was not dependent on the volume of the reaction chamber, and was instead an order of magnitude greater in noise compared to the composite 2 µm chambers. The 10 µm chambers continued the trend seen in the 5 µm chambers, showing little change in CV^2 even at much higher abundance values. Comparisons between the composite of 6 2 µm chambers with the 5 µm chambers reveal a significant difference in CV^2 (an order of magnitude) at the same abundance values. This comparison reveals that even in systems with the same number of genes and resources, the delineation of those resources has a significant effect on the behavior of the system. The changes in noise behavior are also apparent when considering the distribution of final protein abundance values reached, as shown in figure A.14-B, where the chambers of larger volume (blue bars) have a much wider final protein abundance distribution compared to the composite chambers (orange bars).

Separate experiments were undertaken to confirm that the flat CV^2 relationship across increasing chamber volumes was not unique to the PDMS reaction chambers, or due to any kind of surface interactions or molecular adsorption to the chamber walls. The experiments encapsulated PURE cell-free reactions in POPC water-in-water vesicles, which are more biologically similar to cells. The vesicles were then imaged using confocal microscopy. The results from these experiments are shown in figure A.14-C. The volume range of these vesicles was larger than the range of volumes of the chamber (4 to 20 µm in diameter), but overlapped, allowing for comparisons between the two systems. Each colored set of points represented vesicles with a similar range of diameters. The comparison of the two colored sections of volume in the vesicles showed that measured protein abundance scaled linearly with volume, but noise magnitude remained invariant to changes in abundance.

The differences between the composite and the large diameter chambers reveals the importance of resource sharing when considering bursting in gene expression. Even when the ratio of DNA and expression resources are kept constant, changes in the degree of isolation result in significant changes in measured noise in expression. An interesting observation from these results comes from the change in burst dynamics within each chamber. When moving to a larger chamber, and subsequently a larger pool of resources, individual genes are more likely to consume a larger proportion of the total resources in an infrequent manner, as opposed to consuming a small proportion of the total resources in a more frequent manner. This suggests that expression bursts are more readily made bigger, rather than more often.

Observations of the noise magnitude at different chamber sizes revealed that while insensitive to changes in reaction chamber size, CV^2 was hypersensitive to random fluctuations in protein abundance within the same sized reaction chambers. For the 2 µm chambers, individual protein abundance values varied less than one order of magnitude (from 10^4 to $8x10^4$ AU), but individual CV^2 values varied more than an order of magnitude (from 10^{-2} to $3x10^{-4}$). This behavior was also observed in the 5 µm and 10 µm chamber populations. Figure A.14-D shows the hypersensitivity of CV^2 to protein abundance, where the dotted lines are fits to each chamber size where CV^2 goes as one over abundance squared.

3.4 Resource Sharing Model

When comparing a 2 µm chamber to a larger diameter chamber at constant concentration, the abundance values for each constituent inside the system increase proportionally with the volume of the chamber. As such, the resources available to each gene, be it ribosomes, polymerases, etc., is the same for each gene, independent of system volume. Although the amount of resources available for a given gene is the same across all genes, not all genes may utilize a proportionate amount of resources. Due to a number of factors including spatial correlations, cooperative binding, and positive feedback, the number of resources an individual gene utilizes may be vastly different from another gene in close proximity. The experimental data shows that larger chambers have a higher transcriptional or translational burst size, but not burst frequency, due to the fact that an abundance change over three orders of magnitude results in only a half order of magnitude change in CV^2. Two different models were considered to explain the behavior seen in the experimental chambers. The first describes a system in which sharing is driven by positive feedback between the binding of resources and genes. The second model describes resource sharing in which changes in burst size are primarily driven by time-dependent production events associated with transcription.

3.4.1 Shared Resource Pool Model

The first model developed assumed that reaction chambers are resource limited such that not all genes are capable of proceeding with transcription at the same time (Figure A.15-A). In order to express protein, each gene must share the available resources in a time-dependent manner. The sharing of resources is similar to a genetic toggle switch, which describes a system in which two genes are related in such a way that activation of one gene deactivates the other[23]. In the shared resource model, negative regulation is introduced implicitly, as a gene that strongly sequesters resources from a common pool leaves few resources available for a second gene. Additionally, positive feedback (due to cooperative binding, for example) causes genes that bind resources to continue to bind resources at a higher rate. The combination of positive and negative feedback suggests that in a system in which resources are shared among many genes, it is possible that a few genes accumulate the vast majority of the resources at any given time, thus sequestering resources from use by other genes. In this scenario, a portion of the genes can only pull from a strongly depleted resource pool, lowering the amount of resources available for binding. The model is expected to increase burst size as volume increases (more concurrent genes and resources) because the combination of positive and negative feedback will allow a few number of genes to become active for longer periods of time, as opposed to becoming active more frequently.

Monte Carlo simulations of the master equation using the Gillespie algorithm were performed on a system in which a set number of genes would pull from a large resource pool. These resources generically represent any molecules required for gene expression in both transcription and translation. The binding rate (k_{Bn}) of the resources to each gene n = 1, … , N in the system was subject to a positive feedback loop, such that bound resource molecules increase the affinity of the gene for further resource binding. A sigmoidal curve was applied to resource binding (Figure A.15-B). The equation took the form

$$k_{Bn} = \frac{k_{Bmax}}{(1 + e^{-\lambda(x_n - \mu)})} + k_{Bmin} \tag{3.1}$$

where k_{Bmax} determines the maximum k_B value, k_{Bmin} is the minimum k_B value, λ determines the slope of the binding curve, μ determines the curve offset (at what bound number does positive feedback begin to take effect), and x_n is the number of resources currently bound to gene n. The unbinding rate was equal for all genes and was not subjected to any kind of feedback.

To determine whether increasing the size of the system changed the behavior of resource allocation in the model in the same manner as inferred in the experimental data, simulations were run for an increasing number of genes in a single system. The number of resource molecules per gene in the system was constant as the number of genes in the system increased, meant to represent the move to higher chamber size at constant concentration in the experiment. It is assumed that the number of resources utilized by a gene directly correlates with the number of proteins it is capable of producing. Simulations were run until steady state, and the variance in the number of resources bound to each gene in a given system was measured. Systems with 1 to 10 simultaneous genes with positive feedback and without positive feedback (k_B constant over all bound resource values) were simulated for 100 trajectories for each case. Parameters were chosen such that no single gene would "lock" into a highly bound state and remove any time dependent sharing dynamics in the system.

Characteristic gene traces are shown in figure A.16. In a system with only one gene, the number of bound resource varies around a constant steady state value. Because of the chosen parameter space, a single gene does not reach the bound resource value required to transition into the high binding regime. However, in a system with more than one gene and positive feedback, a gene begins to stochastically transitions between a high steady state and low steady state bound resource value. The two steady state values are dependent on the high and low binding rates of the sigmoidal curve. Although the system is populated with enough resource molecules for each gene to accumulate a number of bound molecules above the lower steady state value, the positive feedback allows genes that stochastically bind resource molecules first to continue to bind resources at a higher rate. Because of the limited nature of the resource pool, resource sequestration results in implicit negative regulation between genes. Genes that bind a large proportion of resources force genes without bound resources into an environment without access to the same number of free resources, lowering the gene's steady state bound ribosome value. This combination of forces creates a system where a few genes bind proportionally more resources, while the remaining genes are unable to bind as many. Due to stochastic fluctuations, genes that release enough resources to fall into the low binding regime provide an opportunity for another gene to burst into the high binding regime.

As more genes are added into the system, the variance in the number of bound resources per gene in the system increases more than would be expected compared to a system where there is no positive feedback and all genes have an equal number of bound resources on average at any given time. This relationship is shown in figure A.17, which shows the change in the variance in bound resources per gene for systems with and without positive feedback for an increasing number of concurrent genes. With positive feedback, the increase in variance is due to an individual gene accumulating a large proportion of resources for an extended period of time. The accumulation forms a relationship between the number of genes and the time the gene spends in the “active” state. In systems with few genes, an individual gene transitions quickly between the high and low steady state values. In systems that contain many genes, individual genes spend more time in either the high or low state before transitioning. When examining the transitions of a single gene within a system of many concurrent genes, the system increases variance by exhibiting both longer times between transitions and a higher difference between the high and low steady states. Notably, as the number of genes increases past 5, the variance in bound resources per gene levels off, indicating that the addition of more genes is no longer significantly changing the dynamics of the system. In systems that did not have positive feedback, individual genes did not behave in a bursty manner. The variance in the number of bound resources scales linearly with the number of concurrent genes in the system. Instead, the number of resource bound to all genes varied around a common steady state value. Sample traces are shown in figure A.16-D.

This model shows that it is possible to create a system that preferentially increases burst size when moving to larger chamber volumes by using positive feedback to increase the time between transitions as the number of concurrent genes increases. However, a major concern with this model is the propensity to develop anticorrelation between genes. In a system with more than one gene, an increase in the number of resources bound to a gene is coupled with a decrease in the number bound to another gene. Anticorrelations in the simulations would reduce the ability of this model to explain the properties of the experimental system. Because the experimental system measures behavior in an entire experimental chamber and not on a single gene basis, anticorrelations between individual genes would reduce the variance in the sum of bound resources for all genes. Consider a two gene system transitioning between a high and a low steady state in a perfectly anticorrelated manner. While each individual gene would have high bound resource variance, the sum of the two genes would have low bound resource variance, as the total number of bound resources changes little over time.

Anticorrelation measurements were done for each system of increasing concurrent gene number, where two random genes in a given system were compared to each other by cross-correlating the number of bound resources per gene over time (figure A.18). In the system with only two concurrent genes, those two genes were compared to each other. Cross correlations of systems with few genes show strong anticorrelation, denoted by strong negative cross correlation values at zero lag time (figure A.18-A). The strong anticorrelation is due to single genes sequestering the majority of the resources at low concurrent gene number, causing the system to simply "toggle" between active genes. However, as the number of concurrent genes increases, the likelihood that any two genes within the system are perfectly anticorrelated decreases, as the ability to rise into the highly bound state is shared among a greater number of different genes (figure A.18-B). This property can be shown as a decrease in the anticorrelation value as the number of concurrent genes increases. It is also noted that without positive feedback, little to no correlation is found among genes within a given system (figure A.18-C).

While the variance in the number of bound resources to individual genes increases and the anticorrelation falls as the number of concurrent genes is increased, measuring the variance in the total number of bound molecules among all genes in a system produces results which are not consistent with experimental data. Unfortunately, while the anticorrelation between two individual genes within a system drops as the number of genes increases, the total anticorrelation remains high, since the number of genes in the high state remains constant over the length of a trajectory. When considering the total number of bound molecules across all genes, any change in the variance trend is lost. This property can be found in figure A.19, which shows that the variance of the total bound molecules in a system across all genes as the number of genes increased. In the experimental chambers, variance in the measured fluorescence increased with the square of the chamber volume. This in turn caused CV^2 to remain relatively constant over multiple magnitude changes in abundance value. However, in the model, the variance in the number of bound molecules scales linearly with the number of genes, suggesting the model is not capturing the behavior responsible for the experimental noise measurements.

Different positive feedback binding curves were tested in an attempt to reduce the anticorrelation between genes and to determine whether the variance of the total bound resource in the system could be increased. The sigmoidal curve used in the previous model was taken to two extremes: a step function, which transitioned from a low binding to a high binding rate at a specific number of bound resources, and a linear function, which increased the binding rate linearly with bound resources. Characterization of the step function showed that while it was able to increase the variance in single genes of a system, the shift to the step function simply sharpened the transition times between high and low states for each gene without increasing the variance of the total bound resource population. The linear binding separated the transitions between the high and low states. However, the variance of the total population did not increase with the number of concurrent genes in the system.

3.4.2 Transient Ribosome Model

Because of the problems with the previous model, a new model was developed to help explain the behavior seen in the experimental chambers, based on the random telegraph model of gene bursting (depicted schematically in figure A.20-A). Importantly, the transient ribosome model would be simulated over a short time frame and not at steady state in order to capture the same transient behavior observed in the experimental chambers. In this model, genes within the system transition between an ON and OFF state independently of each other and produce mRNA molecules when in the ON state. However, the system has a limited number of ribosomes that can bind to mRNA molecules as they are produced. Ribosomes here represent any translational resource needed to produce protein. Under the time regime being tested, the mRNA molecules produced are assumed to not decay, and that ribosomes that bind to mRNA molecules are less likely to re-randomize and diffuse back into the global resource (unbinding is small). Bound ribosomes produce protein at a rate that decays exponentially over the length of the simulation, consistent with previously observed experiments[112].

Rate parameters for each of the rate constants were initially derived from literature sources[32, 39], and used as a starting point for a variety of different parameter sweeps. Concurrent genes were varied from 5 to 50 in 5 gene increments and simulated for 100 minutes.

The rate of translation decayed exponentially over time with a time constant set to match the transient behavior of the experimental data. Results from simulations are presented in figure A.20-B. Simulations increase size at constant concentration by increasing the number of concurrent genes with a proportional increase in the ribosome population. Measurements of the CV^2 and protein abundance values in the simulations show that with high transcriptional bursting, larger reaction chambers increases the steady state protein abundance of the system without changing the value of CV^2, consistent with the experimental data. Further examination of the simulations reveal that the shift to higher protein abundance values at similar noise magnitudes is due primarily to the timing and duration of the initial burst of activity for a given expression system.

In the transient ribosome model, at short times, the rate of any gene entering the ON state is the same. The first gene to transition to the ON state has access to the full ribosome resource pool and begins to capture resources. The gene produces several mRNA before returning to the OFF state, and the mRNA molecules sequester ribosomes from the pool. After a period of time, a second gene enters the ON state and produces mRNA molecules. However, a large proportion of the ribosomes have already been sequestered by the mRNA molecules generated from the first gene. At long times, genes that stochastically produce larger bursts of mRNA are left with few ribosomes in the pool, reducing their apparent burst size and limiting their influence on the total bound ribosome population. Additionally, the decrease in translational efficiency over time means that even if late mRNA transcripts are produced in a system with available ribosomes, those ribosomes will not be able to produce protein at a highly effective rate.

Stochastic simulations reveal that noise observed in the experimental system can be explained though the stochastic timing of the first transcriptional burst events. Interestingly, as the size of the chamber increases, a relatively small number of mRNA acquire a larger number of bound ribosomes. Instead of a larger number of mRNA being able to accumulate ribosomes when more mRNA and ribosomes are introduced into the pool, the system instead shifts a disproportionately large number of ribosomes to relatively few mRNA. Figure A.20-C shows this relationship graphically by ranking the timing of each mRNA produced in a given system with the number of proteins that were produced from that transcript. The figure clearly shows that even in systems with many genes producing mRNA, those mRNA that are produced earliest in time are the ones that produce the most protein. Additionally, the same number of mRNA transcripts produce the majority of the proteins in a given system, consistent with the idea that few genes dominate a chamber regardless of the size of the chamber. The model predicts that the larger protein populations found in larger reaction chambers resulted from the translational amplification of burst sizes, not the initiation of more bursts.

An additional experimental prediction was the decrease in noise at increased protein abundance faster than the canonical inverse relationship with abundance for systems containing the same number of concurrent genes (same chamber size). This hypersensitivity to abundance variations seen in the experimental chambers can be attributed to the random timing of the first transcriptional event in each simulation. Chambers that have a transcriptional event early in time are able to make full use of translation, and result in high steady state protein abundance values. The average number of ribosomes captured per mRNA molecule is also reduced, leading to smaller burst size and lower noise. Conversely, systems which have late initial transcriptional events both have less time to produce protein, and a reduced rate of translation.

3.5 *E. coli* Comparison and Steady State Model

The experimental chamber results suggest available translational resources are more likely to aggregate to regions of active transcription, causing expression bursts to self-reinforce. The idea can be extended to say that, in prokaryotic cells, the size of a translational burst (b) is directly correlated with the size of a transcriptional burst (B). Recent work measured transcriptional and translational burst sizes in *E. coli*, which revealed that large mRNA populations are strongly correlated with large translational burst sizes[89]. Additionally, large protein populations are strongly correlated with large translational burst sizes[116]. As seen in figure A.21, as transcriptional burst size varies over a half an order of magnitude, the translational burst size varies over a much larger range (three orders of magnitude). Fitting reveals that the translation and transcriptional burst sizes are related as $b = 0.25{*}B^{4.78}$.

It is critical to note that the mechanistic relationship between burst sizes is still unknown, as is whether a large transcriptional burst size drives a large translational burst size, or vice versa. However, the correlation does suggest strong cooperativity between translational and transcriptional bursting. Some form of positive feedback, in which a large transcriptional burst encourages the formation of a large translational burst, is a possible mechanism suggested by the data. Spatial effects are a likely source of feedback, as crowding due to RNAP or crowding-enhanced localization could increase the expression burst size[117, 118].

The previous model describing the behavior of the cell-free expression chambers was specific to the transient nature of the experimental systems. To create a model that describes the burst behavior measured in the *E. coli* data, the model should be valid in the steady state regime. New model parameters were adjusted to match literature values for *E .coli*[32]. An mRNA decay term was incorporated into the model, and protein decay was assumed to be equal to the doubling time of an *E. coli* cell. Ribosomes accumulated into a “local pool,” representing a region of spatial proximity to the gene, based on the mRNA population in the system instead of to individual mRNAs. The decay associated with translational efficiency was also removed, as the parameter was unique to the cell-free experimental chambers. For this model, it is assumed that the rate at which ribosomes leave the local pool is dependent on the number of mRNA in the system, as it is reasonable to assume that a high local population of mRNA retains ribosomes and continue to produce proteins due to localized crowding effects. Likewise, the rate of protein production is also said to be proportional to the number of mRNA, which argues that rapid rebinding of ribosomes to the mRNA facilitates more protein production. Simulations were run with a single gene under the assumption that the relationship between transcriptional and translational burst size in *E. coli* is specific to individual genes. The transcriptional and translational burst sizes of the system are measured as the number of mRNA produced per burst (transcriptional burst size), and the number of proteins produced from a single mRNA (translational burst size). It is assumed that bursts are well separated in time such that any mRNA produced in a burst decay before a second burst occurs. These assumptions are made to simplify measurements of burst size by avoiding cases where mRNA from a previous burst produce proteins that are then accounted for in the current burst.

Simulation results from the steady state model are shown in figure A.22. The comparison between the transcriptional and translational burst size reveal that a larger transcriptional burst results in a larger translational burst. For this model, an order of magnitude increase in transcriptional burst size is coupled with a similar order of magnitude increase in translational burst size. This relationship can be shown to be due to the influence on mRNA on both the rate of ribosomes entering the local pool and the rate of protein production by comparing the results to a model where either of these relationships is removed. In a system where the rate of protein production is influenced only on the number of ribosomes in the local pool, as shown in figure A.23, there is no translational burst size dependence on transcriptional burst size. However, the model does not predict the multiple magnitude increase in translational burst size over transcriptional burst sizes. Additional factors, including positive feedback pathways not accounted for in the model, may be responsible for the strong correlations between the burst sizes.

3.6 Conclusions

In this chapter, micro and nanofabricated experimental chambers that allowed for accurate control over confinement and resource sharing in gene expression were examined. Specifically, gene expression burst patterns were measured as the number of genes and resource molecules was increased: either through summing of individual chambers (discrete resources), or through an increase in chamber volume (shared resources). It was found that the total protein production scaled linearly with the amount of DNA and resources present, as expected. However, different resource sharing cases resulted in drastically different burst patterns. Composite sums of individual chambers (discrete resources) resulted in higher protein abundance through more frequent bursts, while chambers of increased volume (shared resources) increased protein production though an increase in burst size. The divergence of burst was present even though a constant ratio between resources and DNA was maintained, demonstrating the importance of resource sharing on expression bursting. As chamber size is increased with a shared resource pool, the expression system preferentially modulates burst size as a small number of genes increasingly use a larger proportion of the available resources. The results suggest that expression bursts display cooperativity (through self-reinforcement or positive feedback) that is controlled by the availability of global resources, and not intrinsic properties of the gene. Examination of the models found that genes in the system that produce mRNA early are those that most heavily pull on the resource pool, dominating the bursting behavior of the system. Subsequent bursts have less available resources to pull from and are unable to produce protein at a high rate due to the decay in translational efficiency due to resource limitations. This model suggests that burst size control may be the principle mechanism driving protein abundance changes observed between transcriptional and translational burst sizes in *E. coli*.

4. Crowding and Confinement Effects on Gene Expression

Previously, effects of spatial influences on gene expression were explored with cell-free expression chambers of varying sizes. The focus was on the idea that gene expression was influenced by resource sharing and confinement. This chapter explores the effects of macromolecular crowding and geometric confinement on gene expression and transcriptional bursting. Results from spatially resolved models are compared to two- and three-state models to explore the assumptions made about expression systems in cells and experiments. The importance of spatial considerations are detailed though the analysis of discrepancies between modeling methods.

4.1 Spatial Influences on Gene Expression

The two-state (random telegraph) model is widely used for bursty gene expression due to its ability to describe the behavior of gene expression independent of mechanistic causes. Different molecular mechanisms have both theoretically and experimentally been shown to cause bursty behavior, examples of which were described in detail in an earlier chapter. However, spatial considerations can influence assumptions inherent in the two-state model and significantly influence the behavior of bursty gene expression.

As described briefly in the introduction, previous work has started to reveal the relationship between spatial influences and gene expression. Rein ten Wolde and coworkers have shown that spatial diffusion and rapid rebinding of transcription factors can significantly increase the measured noise in a gene network, and that macromolecular crowding can enhance binding through changes to equilibrium constants[101, 102]. Previous work by Meyer et al. has also shown that highly crowded static environments significantly alter the behavior of gene expression by modifying the diffusive properties of molecules, and can drive what was a well-mixed constitutive process to behave in a bursty manner[59]. However, these works have either considered only (1) crowding using rescaled rate parameters in a well-mixed framework[101] or (2) static spatial crowding effects on gene expression, neglecting the influence of dynamic crowding molecules[59]. Stochastic effects associated with gene expression under the influence of dynamic crowding and confinement have not been fully explored (figure A.24). It has been shown that confinement and altered protein mobility can markedly influence the stochastic behavior of other biochemical reaction networks as well[103, 104].

In this chapter, it is shown that for a simple, spatially resolved model of gene expression, spatial considerations such as crowding by mobile molecules and geometric confinement can markedly influence the measured noise, and subsequently the inferred bursting parameters, of transcriptional bursting. Comparisons to two-state and three-state models reveal that significant aspects of the spatial model are not captured using the simpler models. It is shown that the relationship between burst frequency and burst size that is observed in the spatial data can be attributed to changes in the distribution of events due to spatial effects. In particular, the assumed distribution of events in a two-state model are ill-suited to describe the noise behavior at high crowding in the spatial model due to altered spatiotemporal correlations between molecules. This leads to changes in the noise magnitude of the mRNA population. These comparisons and subsequent analysis of the spatial model highlight the potential importance of spatial effects in the measurement and analysis of noise in gene expression.

4.2 Spatial Model of Gene Expression

Gene expression was modeled here using the Gillespie algorithm with spatial considerations incorporated. The system was first partitioned into a three-dimensional lattice of voxels. A voxel is a “volumetric pixel” and represents a three-dimensional subset of volume in space. Each voxel had the same characteristic length, width and height, all set equal (cubic voxels). Species were introduced into the space randomly and uniformly. Particles were allowed to diffuse, or “hop,” into an adjacent compartment (6 cardinal directions in three-dimensional space, no diagonal movements). Each particle diffused to an adjacent lattice site with a rate given by $\gamma = D/h^2$, where D is the diffusion coefficient for the particle in $\mu m^2/sec$, and h is the characteristic side length of a voxel.

Crowding by macromolecules was introduced by populating the space with crowder species that occupied single sites and excluded other molecules from occupying the same site at the same time. Crowding molecules were allowed to diffuse at a separate diffusion coefficient from the transcriptional particles. Attempts by particles to diffuse into a site occupied by a crowder were rejected. Crowding molecules were populated into the space randomly, where attempted insertions that conflicted with previously placed molecules were rejected. Crowding fraction was defined as the volume fraction occupied by crowding molecules over the total volume of the system. The boundaries of the space were modeled as hard-walls, such that any attempt to transition out of the voxel space was rejected.

A simple model of gene transcription was considered in which a single “gene” and a single “transcription factor,” each represented as a point particle, diffused in space. Transcription occurred at rate α when the two particles occupied the same voxel at the same time. Otherwise, no transcription occurred. The mRNA degraded at rate γ_m. The positions of mRNA molecules, which do not affect gene expression, were not tracked in order to save computational resources. The spatially resolved model was conceptually similar to the two-state model of gene bursting in which a gene can occupy an ON state where transcription occurs (co-localized particles) or an OFF state where no transcription occurs (separated particles).

The diffusion coefficient of the gene and transcription factor was taken to be 1 μm^2/sec, within the typical range of values measured in *E. coli*[119]. Crowding molecule diffusion coefficients ranging from 0.00001 to 0.001 μm^2/sec were considered to investigate the role of crowding molecules on mRNA production. The range of diffusion coefficients allowed for the systematic assessment of the influence of crowding molecules, which is typically most pronounced at slow diffusion coefficients. Additionally, the effect of confinement were explored by considering different system geometries at fixed volume: a bulk three-dimensional system (16x16x16), a confined slab-like system (32x32x4), and a two-dimensional system (64x64x1). Physically, the three regimes are representative of reaction systems in the cytoplasm, those in a confined region of the cell (e.g., the region between the nucleus and plasma membrane), and those confined to the plane of a cellular membrane, respectively. Using the same number of lattice sites for each case removed any ambiguity associated with changing volume. Multiple crowding fractions were considered, ranging from 0% to 50% by volume. For each crowding fraction and confinement case, 100 independent trajectories were generated. Each trajectory was simulated over 1000 minutes, with the first 300 minutes removed to ensure all data was analyzed at steady state.

4.3 Two-State Model Parameterization Methods

Multiple parameterizations of the two-state model were used to explore the relationship between the spatially resolved model of bursty gene expression and the simpler random telegraph model. In examining the spatial model, the co-localization and separation of the two gene expression particles share similarities to the ON and OFF states of the two-state model, respectively, as transcription only occurs during co-localization (the ON state). The “encounter method” assumes the behavior of the diffusing particles is directly correlated with the gene state of the random telegraph model. Two-state models were parameterized using the average particle encounter duration and time between particle encounters from the spatial model. “Encounter” refers to when the two particles occupy the same voxel at the same time. The average time spent apart between encounters and the average time spent together was used to parameterize k_{ON} and k_{OFF}:

$$\begin{gathered} k_{ON} = {}^{1}\!/\!\langle \tau_{Seperated} \rangle \\ k_{OFF} = {}^{1}\!/\!\langle \tau_{Co-localized} \rangle \end{gathered} \tag{4.1}$$

where $\langle \tau_{Seperated} \rangle$ is the mean time between particle encounters, and $\langle \tau_{Co-localized} \rangle$ is the mean time two particles occupy the same voxel. These equations are derived from an understanding of the two-state model under the assumption that the process is sufficiently bursty ($k_{OFF} >> k_{ON}$). Using the generated bursting rate parameters, two-state models were parameterized as a comparison point against which each spatial model at some crowding fraction and confinement was examined. This method is expected to give the correct average number of mRNA since the mean time between encounters and the mean encounter time give the correct fraction of time in the active state. However, it may lead to different noise characteristics due to spatial influences changing the distribution and variance of encounter times in the system. To summarize, the encounter duration and time between encounters for all trajectories of a given crowding and confinement case were averaged to generate a k_{ON} and k_{OFF} value for each comparable two-state model. Values were averaged to generate a single set of parameters for each case because the spatial distributions were also generated from a single set of parameters. Calculating separate two-state burst parameters from individual trajectories would over fit the data.

The theoretical noise framework described in chapter 2 provides a useful tool in relating the stochastic fluctuations of gene expression with the underlying characteristics of the system’s burst behavior. Previously, equations were described that calculate the burst size and burst frequency of bursty gene expression. The equations are leveraged to determine the burst parameters of the spatial simulations, under the assumptions that the noise framework applies fully to the spatial model and that the spatial model is accurately represented by a simple two-state model. Calculated burst size and burst frequency values were converted into two-state rate parameters by using the equations. To review, burst size and burst frequency of the two-state model were calculated from the noise magnitude and abundance of the mRNA population using the equations

$$B = CV^2\langle m_s\rangle \qquad (4.2)$$

$$b_f = \frac{\gamma_m}{CV^2}$$

From these values, a comparable two-state model was constructed by generating k_{ON} and k_{OFF} values using the equations

$$k_{ON} = b_f \qquad (4.3)$$

$$k_{off} = \frac{\alpha}{B}$$

which were again derived by examining the two-state model under the assumption that transcriptional bursts occur at well separated times (k_{OFF} >> k_{ON}). This method is referred to here as the “burst method,” due to its use of the burst equations from the noise framework. In summary, burst size and burst frequency values were calculated from the CV^2 and mRNA abundance of the spatial model for each trajectory. The geometric mean of all burst parameters for all trajectories in a crowding or confinement case was calculated and used to generate a single set of two-state parameters. The geometric mean was used to offset the influence of strong outliers that would erroneously shift the mean value of the burst parameters.

4.4 The Three-State Model

Motivated by the idea that molecule reencounters may play a significant role in the measured noise of the system, a three-state model was considered to determine if the properties of the spatial model are better captured. For this model, the gene is allowed to occupy one of three states: an ON state where transcriptional expression occurs, an intermediate state in which transitions to the ON state are highly likely, and finally a third state that infrequently transitions to the intermediate state. The intermediate and third states do not produce mRNA. The three-state model attempts to account for transcription factor behavior in a crowded environment, where the intermediate state accounts for times when the transcription factor is not in the active state but is still spatially correlated with the gene and therefore has a higher chance of returning to the active state. The third state represents the state when the transcription factor is well separated in space and is no longer spatially correlated. The model is described graphically in figure A.25.

The parameterization method described here was developed with the goal to create a three-state model where noise magnitude could be shifted through a single free parameter. Initial k_{OFF} and k_{ON} values were generated using the average encounter times of the spatial model (encounter method). It was assumed that the rate of transitioning from the active state to the intermediate state was unchanged compared to the two-state model ($k_{12} = k_{OFF}$), which assumed the duration particles remained encountered did not change between a three-state and a two-state model. The rate constant from the intermediate state to the ON state was assumed to be larger than the two-state rate parameter ($k_{21} > k_{ON}$) to account for the increased chance of reencounters. In order to reach the same number of mRNA at steady state as in the spatially resolved system, the combined time spent in the intermediate and third states was adjusted to account for the increase in the k_{21}, such that

$$k_{23} = k_{32}\left(\frac{k_{21}}{k_{ON}} - 1\right) \tag{4.4}$$

It was assumed that when $k_{21} = k_{ON}$, the three-state model generates the behavior of the two-state model by never entering the third state, instead transitioning between the ON and intermediate state with the same rates as the two state model. For this to be valid, the equation sets the transition rate to the third state (k_{23}) to be 0. Any increase in k_{21} over the calculated k_{ON} value was offset by an increase in k_{23}. It was assumed that the rate k_{32} was unchanged, and was set to a constant value for all crowding and confinement cases. It was also assumed that under the range of k_{21} values examined, the ratio of timing between the intermediate and third states was insensitive to the exact value of k_{32} and k_{23}. With these assumptions, the distribution of times between being in the ON state, and therefore the noise, was modulated by a single parameter (k_{21}). Increases in the value of k_{21} represented a change in the distribution of times spent in the active state, where short periods of rapid reencounters were punctuated by periods without encounters. For each spatially resolved crowding and confinement case, rate constants (k_{ON} and k_{OFF}) were first generated from the average encounter durations and time between encounters. k_{21} was then increased from the encounter method value (k_{ON}) until the distance between the mean value of the noise measurements from the spatially resolved simulations and the three-state model were minimized.

4.5 Spatially Resolved Model Results

Figure A.26 shows the results of the spatially resolved simulations. Figure A.26-A describes the time resolved mRNA population traces over a 100 minute portion of 10 simulation trajectories of the cubic 16x16x16 lattice space system at 0% crowding fraction. Each trajectory is assigned a random color. Figure A.26-B shows the mRNA population traces of the cubic 16x16x16 lattice space system at 50% crowding fraction. For each crowding and confinement case, the traces fluctuate around a common steady state mRNA population. However, in cases with high crowding fraction, several trajectories contain large fluctuations of populations, characterized by a strong burst of expression followed by decay over time. The large fluctuations are associated with enhanced mRNA production and are not present in systems that have low (below 30% crowding fraction by volume) macromolecular crowding. The likelihood of an expression “spike” occurring in the mRNA population increases as the crowding fraction of the system increases, while more confined spatial geometries also result in expression spikes occurring at lower crowding fractions. Figure A.26-C shows the mRNA population traces for the flat 64x64x1 lattice space system at 50% crowding. Interestingly, the most extreme crowding fraction and confinement geometry show spikes in expression that occur in close proximity. As will be discussed in detail later, large fluctuations in mRNA population occur when the two gene expression particles are locally trapped by crowding molecules and co-occupy a small effective volume for a short period of time.

The noise framework was used on the spatially resolved simulation results to measure the underlying burst parameters for each trajectory by examining the noise magnitude and abundance of each mRNA trace. Figure A.27-A shows the noise magnitude as measured by CV^2 as a function of mean mRNA abundance for the cubic 16x16x16 voxel space at a variety of different crowding fractions (from 0% to 50% by volume). Each small point represents the results of a single trajectory (colored by crowding fraction), and each large marker is the geometric mean of all the data points at a given crowding fraction. The geometric mean was used because the graph visualizes the data in log space, and the geometric mean minimizes the effects of outliers on the mean. As crowding fraction increased, the average steady state mRNA value increased as well, while the measured CV^2 decreased with an inverse relationship to mRNA abundance at low crowding fractions (below 30% by volume). However, behavior distinctly changed at higher crowding fractions, characterized by an increase in noise magnitude. Separately run, well-mixed models (figure A.27-B) were analyzed, where spatial diffusion considerations were removed. To account for the volume effects of the macromolecular crowders, the rate of transcription was modified based on the effective volume of the system. These simulations revealed that the driving force behind the increase in mean mRNA population was due primarily to the decrease in effective volume due to the crowding molecules, reducing the volume available for the gene expression molecules to interact. Notably, in these well mixed simulations, the noise magnitude continued to decrease at an inverse relationship with abundance for all crowding fractions considered, which was in contrast with the spatial simulations, where noise magnitude increased at higher crowding fractions.

To review, under the assumption that the system is adequately described by the random telegraph model, higher mRNA abundance values are achieved through an increase in burst size or an increase in burst frequency. Changes in burst dynamics are directly correlated with shifts in the distribution of noise magnitude and mRNA abundance. An increase in abundance due to increased burst size results in no change in CV^2 values, while an increase in abundance due to increased burst frequency results in a proportional decrease in CV^2. The results from the spatial model suggest the dynamics of bursting for gene expression in the spatial model are changing in two district regions of the crowding fractions tested. In the first region, which occurs over the range of crowding fractions from 0% to 30%, CV^2 drops inversely proportional to mRNA as the mRNA abundance value increases, which is consistent with an increase in burst frequency resulting from the excluded volume effects of the macromolecular crowding particles. However, at crowding fractions in excess of 30%, the system enters a second region where CV^2 begins to increase as abundance increases, which is indicative of an increase in burst size coupled with a decrease in burst frequency.

The geometry of the spatially resolved model, which was modified by changing the arrangement of voxels in space, also dramatically influenced the noise magnitude of mRNA production. Figure A.27-C shows noise magnitude and mRNA abundance for simulation trajectories at 50% crowding for three different voxel arrangements. In the least confined system (16x16x16 lattice), the distribution of points was relatively tight, with outliers at higher abundance and CV^2. These outliers corresponded to trajectories with large spikes in mRNA production, which drive the noise magnitude and the mRNA abundance values higher. In the moderately confined system (32x32x4 lattice), the distribution of the points was similar to the 16x16x16 lattice case, suggesting moderately confining the system space at high crowding was not significant enough to change the dynamics of bursty expression when compared to a more cubic system. In the most confined system (64x64x1 lattice), the distribution of points showed marked differences from the two other cases, with a broad distribution of steady state mRNA abundance values that extended over two orders of magnitude. The broad distribution was shown to be a consequence of long-lived spatial correlations that persisted due to confinement to two-dimensions. When two reacting particles are confined to a small volume by crowding particles, there are fewer degrees of freedom by which to escape from the confined subvolume. Similarly, when two particles are spatially segregated, crowding molecules are more likely to prevent the particles from encountering each other.

It was assumed in the spatial simulations that the macromolecular crowders diffused at a slower rate compared to the particles needed for gene expression in order to better observe the effects of crowding on transcriptional bursting. To determine how the rate of diffusion of the crowding molecules would affect the spatially resolved simulations, several mobility tests were conducted at different diffusion rates, shown in figure A.27-D. The most confined case at a high crowding fraction (64x64x1 lattice sites at 50% crowding fraction by volume) was used to emphasize the effects of any spatiotemporal correlations that may occur. The crowding molecules are characterized by diffusion coefficients 1000 to 100,000 times slower than the reacting particles, and their mobility both allows and constrains the local caging imposed by the crowders to eventually relax, although the timescale for this relaxation is influenced by the diffusion coefficient. This assertion is consistent with the figure, which shows that slower diffusion rates results in a broader distribution of steady state mRNA values and higher values of

CV^2. More mobile crowding molecules allow the reacting particles to explore the space more rapidly, and any localized confinement of the two particles into a smaller volume is shorter lived. As a consequence, large fluctuations in mRNA production are less pronounced, and the noise is reduced. Static crowding molecules were also considered, which at sufficiently high crowding fractions (above 30% crowding fraction by volume), resulted in trajectories that exhibited dramatically different dynamics: In some, the reacting particles were partitioned into separate subvolumes and were unable produce mRNA. In others, the particles were co-localized in a subvolume that resulted in high production of mRNA. Because these systems were static, there was no relaxation of the local spatial caging, thus resulting in divergent steady state behavior based on the random placement of the gene relevant particles. While these diffusion constants are too slow to represent realistic macromolecular crowding alone, these diffusion rates are useful in exploring and emphasizing the influence of mobile crowding on gene expression behavior.

4.6 Comparing Spatial Against Two- and Three-State Models

In this section, the degree to which the behavior of the spatially resolved model can be described by two- and three-state models of gene expression is explored. As described in detail previously, the behavior of individual simulation trajectories is characterized by the mean steady state value of mRNA and the noise magnitude measured by CV^2. The distribution of points in the CV^2 and mRNA abundance plane then provides a characterization of the gene expression burst behavior for a particular model compared to the spatially resolved model.

Figures A.28 through A.32 contain the accumulated results from all combinations of parameters tested for each model. Each figure shows the noise magnitude vs mRNA abundance values for a given model (spatial, encounter based two-state, burst equation based two-state, and three-state models) at multiple crowding fractions (10%, 30%, 50%), crowder diffusion constants (0.001 μm^2/sec to 0.00001 μm^2/sec), and confinement geometries (16x16x16, 32x32x4, 64x64x1). Each point represents a single mRNA population trajectory.

The spatial model results are described in figure A.28. Several trends are clearly shown through the comparison plots of CV^2 vs mRNA abundance. For all cases, as crowding fraction is increased, both the noise and the mRNA abundance increase as well. Additionally, shifts to higher crowding fraction result in a change in the distribution of CV^2 and mRNA abundance. As noted earlier, the shift to higher mRNA abundance is consistent with excluded volume effects associated with increased crowding. Additionally, the appearance of outliers is driven through short spatial correlations due to crowding molecules, which cause mRNA abundance values to deviate from steady state for a short amount of time.

Interestingly, the distribution of outliers change based on the diffusion rate of the crowding particles. When considering the cubic confinement geometry, a reduction in crowder diffusion rate results in a reduction in the maximum CV^2 and mRNA abundance values obtained. However, the absolute number of trajectories that deviate from the mean distribution increased, shifting from a distribution with a main cluster and few outliers to a single, long distribution. The behavior is attributed to the dynamics of molecular trapping: at slow crowder diffusion, relaxation time of spatial correlations is slow, such that deviations in mRNA abundance reach higher values due to extended trapping. Additionally, slow relaxation times result in fewer opportunities for trapping events to occur over each trajectory. At higher crowder diffusion rates, the spatial correlations do not last as long, such that spikes in mRNA abundance do not reach the same high values. Faster crowder diffusion allow for more opportunities for trapping events, increasing the number of trajectories that display deviations from the mean behavior.

In the spatial results figure, the flat 64x64x1 geometry at 50% crowding fraction displays significant changes in distribution associated with changes in crowder diffusion rate. At the slowest crowder diffusion rate, the distribution of mRNA abundance values ranges close to two orders of magnitude (some data points are missing due to axes constraints; full range is shown in figure A.26-C). As crowder diffusion rate decreases, the range of mRNA values decrease. The change in mRNA distribution reveals that when crowder diffusion rate is low, the spatial correlations last long enough to either drive mean mRNA abundance to extremely low or high values, depending on whether crowding molecules separate or trap the two transcriptional molecules.

The two-state models parameterized using the encounter method are shown in figure A.29. Clearly, drastically different behavior is displayed, as changes in crowding fraction influence CV^2 values differently than the comparable spatial model results. Changes in crowding fraction result in an increase in mRNA abundance values and a drop in CV^2, which follows an inverse relationship. The trend appears regardless of confinement geometry or crowder diffusion rate. Interestingly, the encounter method generates results which match the mRNA abundance and CV^2 values of the well-mixed, non spatial case generated previously. The results suggest mean encounter behavior is insensitive to changes in the rate at which crowders diffuse or the spatial geometry of the system.

Two-state models parameterized using the burst method are shown in figure A.30. When considering the 16x16x16 geometry, changes in crowding fraction result in similar trends in CV^2 and mRNA abundance compared to the spatial case. Initially, CV^2 decreases with increasing crowding fraction (10% to 30%). However, as crowding fraction continues to increase (30% to 50%), CV^2 increases, indicating a change in burst behavior. The results are also relatively insensitive to changes in crowder diffusion rate. The 32x32x4 spatial geometry reveal similar results, with CV^2 first decreasing, then increasing as crowding fraction increased. In the 64x64x1 spatial geometry, the 50% crowding fraction case shows significant dependence on the diffusion rate of the crowders. Slow crowder diffusion rate resulted in a larger distribution of mRNA and CV^2 values, while faster crowder diffusion rates resulted in significantly tighter mRNA values. While consistent with the change in distribution seen in the spatial results, the two-state model does not reach the same range of mRNA values, especially those at low mRNA abundance.

The three-state model results are shown in figure A.31. Like the two-state model parameterized using the burst equations, the three state model captures the trends of the spatial model when crowding fraction increases in both the 16x16x16 and the 32x32x4 spatial geometries, where noise magnitude initially decreases before increasing at higher mRNA abundance values. In the 64x64x1 spatial geometry at 50% crowding, the three-state model is better able to recreate the wide distribution of mRNA values similar to the spatial results. However, the three-state model is still unable to fully capture the distribution, as it does not reach the same low mRNA abundances at high CV^2 seen in the spatial results.

Finally, figure A.32 compiles the results from the previous four figures by showing the geometric means for all traces of a given crowding fraction for each model. When considering the cubic, 16x16x16 results at all crowder diffusion rates and at 10% and 30% crowding fraction, the centroids of the spatial model, encounter two-state model, and three-state model are consistent with each other, while the burst two-state model reaches higher CV^2 values. Interestingly, for the 50% crowding fraction case, the burst two-state model matches the centroids of both the three-state model and the spatial model, while the encounter two-state model reaches a significantly lower mean CV^2 value. These relationships are consistent in the 32x32x4 geometry as well. In the 64x64x1 confinement geometry, the 50% crowding case reveals the most deviation among the four different modeling methods. The encounter method mean CV^2 value is significantly lower than the spatial model CV^2 value, continuing to trend in an inverse relationship with mRNA abundance as crowding fraction increases. While both the burst two-state model and the three-state model show increases in mean CV^2 value, they are unable to match either the spatial CV^2 or mRNA abundance values.

The different parameterization methods for two- and three-state models of bursty gene expression were then compared directly to the spatial model at the slowest diffusion constant, 16x16x16 confinement geometry, and at a larger range of crowding fractions (0% to 50%) in order to better understand the discrepancies between models. The first comparison involved the encounter method, shown in figure A.33-A as colored points. They are contrasted with the results of the spatially resolved simulations (shown in grey). The spatially resolved results can be divided into two regions of behavior: the results between 0% and 30% crowding follow an inverse relationship between noise magnitude and mRNA abundance, where CV^2 decreasing inversely with mRNA abundance, while the results above 30% crowding fraction trend to higher CV^2 values as mRNA increases. In the first region, the distributions of the encounter method two-state model results are similar to the spatial results, following the same reduction in CV^2 as mRNA abundance increases. However, the encounter method deviates from the spatial model in the second region of CV^2 behavior. While the two-state model captures the same average steady state value of mRNA as in spatial model, the noise magnitude does not shift behavior, and instead continues to follow the same trend to lower CV^2 values at higher mRNA abundance. The behavior is in stark contrast to the sharp increase in CV^2 observed in the spatially resolved simulations. The difference clearly indicates that parameterizing the two-state model using the average encounter duration and average time between encounters is not sufficient to describe the noise characteristics of the spatially resolved system at high crowding fraction. Interestingly, the behavior of the encounter results is the same as the behavior shown in the well-mixed system at different effective volumes described earlier in the chapter.

In figure A.33-B, the results from the two-state model parameterized using the burst method are shown. The burst method two-state model produces qualitatively different results compared with the previous, encounter based two-state model. In particular, the burst method recreates the two region behavior characterized by a decrease in CV^2 at low crowding fraction and an increase in CV^2 at high crowding fraction, while also reaching the same mRNA abundance values. However, at all crowding fractions tested, an offset in noise magnitude is observed in the two-state results, with CV^2 values higher than those observed in the spatially resolved results. It is clear that assumptions in either the two-state model or in the burst equations used to generate the rate constants are not sufficient to accurately describe the noise behavior of the spatial model.

Finally, results from the three-state model with rate parameters generated through the method described previously are considered. Figure A.33-C reveals that the distribution of points in CV^2 and abundance space closely match those of the spatially resolved simulation. The CV^2 values of the three-state model are modulated without a change in average mRNA levels by optimizing a single burst parameter (k_{21}) such that rapid state changes between the ON and intermediate states were interspersed with state changes between the intermediate state and the OFF state. It was found that in order to drive the three-state model noise magnitude to levels similar to the spatial model at high crowding fraction, the free parameter (the rate of entering the ON state) needed to be increased by half an order of magnitude. Thus, high CV^2 at high crowding fraction was attributed to changes in the timing of encounter events. While this method most closely recreates the results from the spatially resolved model, it does not capture the appearance of rare events at high crowding fractions (large, short-lived deviations in mRNA abundance).

Comparisons in confinement changes are also examined to characterize effects on noise magnitude and steady state abundance of mRNA, as well as how well changes in confinement are captured using the various two- and three-state parameterization methods. Figure A.34 shows the three two- and three-state parameterization methods, which are compared against the spatial model as in Fig. A.33. All figure panels show the spatial model results over three different confinement cases (16x16x16, 32x32x4, 64x64x1) at 50% crowding fraction in gray with the two- and three-state model comparisons in color. The 50% crowding case was chosen to emphasis the differences in each parameterization model.

Figure A.34-A shows the CV^2 and mRNA abundance results for the two-state model parameterized using the encounter method. Similar to when crowding is varied, the increase in CV^2 due to increased geometric confinement is not captured by the encounter method. Very little change in noise magnitude or mRNA abundance occurs at drastically different confinement regimes, which is in stark contrast to the spatial results. The results reveal that the mean behavior of the encounters within the spatial system is insensitive to changes in geometric confinement under the same volume. The burst method, shown in figure A.34-B, better matches the 16x16x16 and the 32x32x4 lattice site cases when compared to the encounter method, but is unable to account for the large distribution in abundance and CV^2 values reached at the most confined test case (64x64x1). Finally, the three-state model is presented in figure A.34-C. While the model is best able to capture the CV^2 behavior of the spatial model at the highest confinement geometry tested, the three-state model is unable to capture the distribution of the spatial results, even at extremely high ON rates. The results imply the current method of parameterizing the three-state model is insufficient at describing the behavior at high geometric confinement and crowding.

4.7 Discussion

The different approaches used to parameterization simpler two- and three-state models are analogous to different methods of taking an experimental system, which may include parameters outside the scope of the random-telegraph model, and fitting it to a framework that can be more easily analyzed. The comparison of the spatially resolved model against each of the two- and three- state models have shown that the noise signatures found in the spatial model are difficult to recreate using a simpler model. But what about the spatial model is changing and causing these discrepancies?

The encounter method makes an assumption that is considered reasonable: the ON state in a two-state model is equivalent to when the two particles occupy the same voxel site in the spatial model, since this is the only period that transcription occurs. However, comparisons between the spatial results and the encounter method results show significant differences in measured CV^2 values at high crowding fractions and spatial geometries. While the encounter based two-state model produces steady state mRNA abundance values that are consistent the spatial model, the noise magnitude at high crowding fraction is significantly suppressed and qualitatively differs in trend.

The second method of generating two-state parameters relies on the noise analysis framework to generate two-state parameters using the noise magnitude and abundance values from the spatially resolved model. While better able to capture the increase in CV^2 observed in the spatially resolved simulations, an offset in the noise magnitude is introduced. The discrepancy between the two-state and spatial results could be due to two possibilities: the assumptions in the burst equations derived from the noise framework, or the assumptions associated with the two-state model (namely the assumption that the distribution between events in exponential).

In contrast to the two-state models, the three-state model is better able to capture the behavior of the spatially resolved simulations. Better fits are likely a product of parameter optimizations that minimize the distance between the geometric means of the three-state and spatial model results. The accuracy of the three-state model comes from the ability to tune the distribution of burst events to match those found in the spatial model at each different crowding fraction. However, long lived correlations in the spatial model that lead to spikes in mRNA production were still not captured. Additionally, the model failed to capture the noise behavior at the extreme geometric confinement and crowding cases considered.

The inadequacies of each simplified two-state model parameterization method are a direct result of the difference between the assumed exponential distribution of burst times for the two-state model and the actual distribution of encounter times in the spatial model. Comparisons of the encounter distributions are analyzed for the highest crowding fraction (50%) spatial model in the 16x16x16 confinement geometry, the comparable encounter based two-state model, the comparable burst based two-state model, and the comparable three-state model (figure A.35). The figure measures the time between events as the time between particle encounters in the spatial model and time between transitions to the ON state for the two- and three- state models. ECDF stands for the empirical cumulative distribution function, and is on a semi-log plot to better visualize the differences between the models.

The range of times between events for each model differs greatly. The spatial model spans the largest range, with encounter times occur at significantly lower values compared to the simpler models, due to increased rapid rebinding. Long periods without particle encounters are attributed to crowding molecules reducing the ability of spatially uncorrelated particles to become correlated. The encounter based two-state model spans the smallest range, as the model is incapable of recreating either the rapid reencounters required for short event times or the crowder induced long waiting times between encounters, due to the assumed exponential distribution of times inherent in the two-state model. Notably, the mean value of times for the spatial model and the encounter based two-state model are the same, as the mean encounter times from the spatial model were used to parameterize the two-state model.

The burst based two-state model displays an ECDF that is similarly shaped to the encounter based two-state model. The similarity of shape is due again to the assumption of an exponential distribution of times inherent in the two-state model. However, the curve is shifted to longer times between events, such that the model captures the long times between burst events at the expense of quick bursting events. The shift in the mean time between events is accounted for with an increase in the burst magnitude, such that the model retains the same steady state mRNA abundance values as the spatial model by producing less frequent, but larger, bursts. The three-state model distribution is better able to capture the noise magnitude of the spatial model by reaching similar rare, long times between transcriptional events while also reaching faster rebinding events than is found in the two-state models. However, the model still does not capture the extreme rapid rebinding values seen in the spatial simulations.

The differences among the ECDF curves for each model reveals the way each simpler model attempts to capture the behavior of the spatial model. Interestingly, the two-state model parameterized using the burst method is able to qualitatively capture the change in CV^2 behavior at high crowding fraction, even under the assumption that event timings were exponentially distributed. Further examination of the burst method two-state model revealed that the model accounts for the change in CV^2 by increasing burst size and decreasing burst frequency. In a spatially resolved system in which rapid rebinding is prevalent (at high crowding fractions), mRNA production events occur in a spatially and temporally correlated manner. However, because the two-state model assumes exponentially distributed times between bursts, rapid reencounters from the spatial model, which are extremely unlikely to occur in a two-state model, are instead grouped as single bursts. The grouping of encounters effectively increases burst size while decreasing burst frequency, due to the accumulation of many smaller encounters into single events.

The idea that rapid reencounters punctuated by long periods of time without encounters motivated the creation and analysis of the three-state model. The likelihood of rapid reencounters in the system is adjusted using a single parameter (k_{21}). Optimizing k_{21} such that the three-state model noise magnitude matched the spatial simulations revealed that the distribution of encounter events in the three-state model was broadening, consistent with crowding in the spatial model inducing rapid re-encounters, coupled with longer excursions once particles become uncorrelated. In the three-state model, this distribution is recreated by having the system rapidly transition between the ON and intermediate state (rapid reencounters) while also transitioning more readily from the intermediate to the OFF states (re-randomization). However, at high crowding and high geometric confinement, this method of parameterizing the three-state model breaks down. Because the parameterization method prioritizes the ability to adjust noise using a single parameter and focuses on changing the burst distribution by reducing the time spent in the intermediate state, extreme increases in k_{21} results in the saturation of the burst distribution. Under the current parameterization scheme, at high values of k_{21}, shifts in the value of k_{21} no longer change the distribution of events in time. In order to properly reach high noise values, a different method of parameterizing the three-state model is required.

As discussed, changes in the distribution of encounter times are responsible for the increase in noise magnitude at high crowding and confinement. In the spatial model, at high crowding fractions, strong, short lived excursions in mRNA expression away from steady state (spikes) provide for an extreme example of rapid reencountering. Over the crowding fractions and confinement regimes considered, expression spikes were most likely to occur at high crowding and at high confinement. In the least confined regime (16x16x16), large production spikes fail to appear at crowding fractions below 30%. Spikes are more likely to occur as the crowding fraction increases, due to "caging" caused by crowders confining the expression particles into a small volume for a short period of time.

In figures A.36 and A.37, the effects of spatial caging are shown by mapping the location of particle encounters over the course of a single trajectory. Figure A.36 shows the 16x16x16 space represented as a 3 dimensional set of points, where the size and color of each point represents the number of times particles encountered each other in that specific lattice site. Larger, redder points represent lattice sites with many particle encounters. Sites without particle encounters are blank. The trajectories used to generate the figures each contained a single expression spike due to crowder-driven caging. The caging is localized to the small area with large, red circles. The high number of encounters in a short period of time result in an expression spike. Analogous behavior can be seen in Figure A.37 for the most confined case (64x64x1). Caging occurs due to the diffusive nature of the crowding molecules, which can stochastically confine the expression particles in a much smaller volume, promoting interactions between the two particles and causing an increased rate of mRNA production. After a period of time, the stochastic motion of crowding molecules allows the two particles to separate and sample the larger volume. Large fluctuations in mRNA number are more likely at high crowding fractions and at high spatial confinement due to the increased likelihood of trapping. Additionally, spikes in mRNA abundance are influenced by the rate of diffusion of the crowders. In systems with slow crowder diffusion, mRNA spikes are less frequent but reach higher values, while systems with faster crowder diffusion result in mRNA spikes that are more frequent but lower in magnitude. This relationship is due to the change in cage timing, as slower crowders are slower to form, but cage particles for longer periods of time.

The relationship between burst size and burst frequency for increasing mRNA abundance values has been studied in a number of papers. So et al. measured an increase in transcriptional burst size with increasing expression level, which was attributed to modulation of the k_{OFF} parameter of the two-state model[120]. Additionally, Taniguchi et al. have shown through analysis of both protein and mRNA expression in *E. coli* that low expression levels are primarily dominated by intrinsic noise, while high expression levels are dominated by extrinsic noise[32]. These examples describe real-world experiments that have measured changes in noise behavior over a range of mRNA expression levels. The work comparing a spatial simulation model with two- and three-state models have shown that discrepancies arise when assumptions underlying the simpler models are not consistent with the behavior of the real system. This highlights the need to understand the relationship between experimental systems and the models used to describe them, as spatial or other factors that change the distribution of transcriptional events may be inconsistent with underlying model assumptions.

4.8 Conclusions

Transcriptional bursting is a commonly observed phenomenon in gene expression. While molecular mechanisms have been widely explored as the cause of gene bursting, spatial considerations have been neglected until recently. The assumptions regarding the two-state model of gene expression are considered adequate to describe transcriptional bursting in cells. However, extensive testing with two-state gene expression models shows that features of models that incorporate spatial details are not captured by simpler models in some cases. The development of the three-state model (motivated by observed changes in the distribution of times between transcriptional events in the spatial model) introduced an intermediate state to represent a spatial correlated state which was highly likely to reenter the active state. While the three-state model is best able to capture many of the behaviors of the spatial model, it does not capture the most extreme crowding properties seen, such as expression spikes due to molecular trapping and noise at high confinement. Further analysis of the bursting behavior of the spatial system show the discrepancies in the simpler models are due to differences in the assumed distribution of encounter times, as the spatial model does not burst with exponentially distributed wait times. Crowding and confinement both increase the likelihood of particle reencounters and the likelihood that spatially separated particles remained uncorrelated for longer times. The change in encounter distribution reduces the measured burst frequency and increases the measured burst size through the accumulation of rapidly occurring encounter events into single burst events. Comparisons among the models reveal the importance of considering spatial factors when examining the behavior of bursty gene expression, especially those under high crowding and confinement.

5. Conclusions

This work is focused on the observation and analysis of gene expression under the effects of spatial factors, such as macromolecular crowding and physical confinement. Experimental analysis of cell-free reaction chambers revealed that gene expression burst patterns were highly dependent on the method of resource allocation. While the composite of individual chambers increased protein abundance through more frequent bursts, chambers of increased volume increased protein abundance through larger bursts. Through the use of models, it was shown that the change in burst behavior at higher chamber volume was a product of the timing of initial bursts of activity in each chamber. In systems with a single resource pool and multiple genes, genes that produce mRNA early capture a disproportionate number of global resources, dominating the burst behavior. Additionally, changes in volume result in a small subset of genes utilizing an increasing amount of resources, reaching higher protein abundances through changes to burst size.

Subsequent simulations of spatial and two-state gene expression models revealed that noise and burst behavior were strongly influenced by macromolecular crowding and geometric confining effects. High crowding fraction was shown to increase the noise of systems above what would be expected in a non-spatial, well-mixed system, and it was shown that subsequent two-state models were unable to capture the same noise behavior. A three-state model was developed that was better able to capture the noise behavior of the spatial model, but was still unable to reproduce extreme crowding and confinement conditions. Discrepancies between the models were found to be directly related to the distribution of encounter events. While the two- and three-state models assume an exponential distribution between encounter events, the spatial model generates encounters that deviate significantly from that assumption.

Each of the models and experiments undertaken here share a significant point of interest: the resulting observations and conclusions could not have been explored without the consideration of time dependent measurements. While many experimental studies explore noise and bursting through imaging of individual mRNA molecules[89] or through flow cytometry[50], the results from the cell-free chambers could not have been obtained without tracking and imaging individual chambers over time. The cell-free chambers were subjected to significant time dependent resource allocation, where mRNA produced early in time were much more likely to produce protein compared to those produced late in time. Likewise, in the simulation of crowding and confinement, it was clear from the comparisons between the two-state models and the spatial model that steady state measurements of noise and abundance are not sufficient in describing the behavior of the system. The distribution of events, which was strongly dependent on the timing of encounter events, was shown to be important in the magnitude of noise in simulated systems. These experiments reveal both the importance of time dependent considerations when measuring the behavior of cellular systems, as well as the importance of spatial effects on noise in these systems.

Both the modeling of the experimental chambers and the simulation of crowded and confined spatial environments reveal the importance of considering spatial factors in the modeling and analysis of gene expression systems. The work here reveals additional avenues of inquiry, both in the analysis of spatial effects on the efficiency of resource utilization in transcriptional bursting, as well as in the measurement of noise and the characterization of bursting in more complex cell systems.

5.1 Transcriptional and Translational Burst Size in E. coli

At the end of chapter 3, the relationship between transcriptional and translational burst size in *E. coli* was examined, and it was noted that the translational burst size is strongly correlated with the transcriptional burst size. These experimental results present an intriguing avenue of inquiry regarding expression bursting. Organisms are tasked with optimizing the use of a limited reservoir of shared resources. While little is known about the benefits of gene bursting, the experimental results illustrate that bursting may be a method for organisms to share resources in a time delineated manner. Genes are constrained to limited periods of time where they draw heavily from the shared resource pool, while utilizing no resources for the remainder of time. This pattern shares many similarities to packet mode communication[121], where the capacity of a shared network is divided among a number of different messages. Several recent results have revealed that burst frequency saturates in many types of cells[35, 39, 116]. The preferential modulation of burst size instead of burst frequency for resource sharing could serve as an explanation for the saturation of burst frequency. Still other recent results have shown burst size increases in response to an increase in cell volume[122] or crowding in vesicles[123].

Further modeling and analysis of spatial considerations may reveal a clearer picture of how burst patterns lead to more efficient use of shared resources.

This line of inquiry again leads to the exploration of spatial factors, which have been shown to be important in the experimental expression chambers. The production of mRNA causes a recruitment of translational resources, including large macromolecules such as ribosomes. These ribosomes translate proteins by moving along the length of the mRNA transcript, and multiple ribosomes can bind to single transcripts. Additionally, the proteins are spatially correlated when they are produced, requiring some time to diffuse away from the mRNA transcripts. All of these factors could result in a locally crowded region, where ribosomes are more likely to rebind to nearby mRNA transcripts instead of diffusing away from active regions of translation. The production of mRNA transcripts in bursts could therefore cause the translational burst size to couple into the transcriptional burst size by increasing the crowding in the local area, thereby increasing the rate of ribosome rebinding.

Preliminary spatial models have been developed in order to determine whether ribosomes remaining spatially correlated with active transcriptional areas increases the overall usage efficiency of the resources available. The space is discretized into a cubic lattice of spaces in three dimensions. Ribosomes are randomly populated in the space and allowed to diffuse to nearest neighbor lattice sites according to a set diffusion rate. mRNA molecules are randomly added to the system over time, and at time 0, only one mRNA molecule exists. mRNA molecules do not diffuse and remain in a single lattice site. Here, efficiency is defined as the ratio of ribosomes localized in a site with active mRNA over the total number of ribosomes in the system. A highly efficient ribosome spends the majority of the time localized with an mRNA molecule, while a low efficiency ribosome spends much of its time searching the reaction space for an active mRNA region. Two scenarios are initially considered. In the “non-sticky” case, the appearance of mRNA does not change the diffusion rate of ribosomes into or out of a lattice site. In the “sticky” case, lattice sites with active mRNA molecules have a reduced diffusion rate out of the site, such that ribosomes that diffuse into the mRNA lattice site are less likely to leave.

Results of these simulations are presented in figure A.38. In the case where mRNA molecules do not interact with diffusion, the efficiency of the system scales linearly with the number of active mRNA molecules. The efficiency is consistent with the number of ribosomes in the space and the number of mRNA active mRNA molecules. In the case where mRNA molecules cause ribosomes to diffuse away more slowly, efficiency increases a significant amount compared to the non-sticky case. In these simulations, the appearance of an mRNA molecule causes nearby ribosomes to congregate around the local area. The efficiency of the sticky model is dependent on the value of the reduced diffusion rate in relation to the normal diffusion rate. Interestingly, when a new active mRNA molecule is produced in the sticky model, the efficiency of the total ribosome population does not increase linearly. Instead, subsequent mRNA produce a diminishing increase in efficiency. Examining the space reveals that when a new mRNA molecule arrives, active ribosomes leave previously activated mRNA sites and aggregate around the new mRNA. Over time, the relative number of ribosomes around any single mRNA molecules will be the same, since all mRNA molecules have the same reduced diffusion parameters.

Additional simulations are required to fully explore the spatial model of ribosome binding and resource efficiency. While these simulations show an increase in efficiency as the number of active mRNA molecules increase, further simulations at steady state (where mRNA molecules both arrive and decay) are required. Additionally, various spatial considerations such as geometric confinement and macromolecular crowding may influence the behavior of the resource molecules. Finally, the spatial distribution of active transcription sites, and whether or not resource intensive sites are spatially correlated, could have an effect on the measured resource efficiency of the system. The preliminary results do support the assertion that spatial effects can influence the sharing of resources, suggesting the correlations in transcriptional and translational burst size are a method of utilizing limited resources more efficiently.

5.2 Noise in the HIV-1 Negative Feedback Circuit

While the magnitude of noise in many gene expression systems has been shown to be influenced by parameters including crowding fraction, geometric confinement, and resource sharing, it can be difficult to understand the impact changes in bursting behavior have on cellular systems. To illustrate the impact of noise on real systems, the HIV-1 virus' behavior in transitioning from latency to active infection is described here.

Human immunodeficiency virus (HIV) actively replicates in CD4$^+$ T lymphocytes, weakening the patient's immune system and potentially leading to acquired immunodeficiency syndrome (AIDS), allowing for opportunistic infections to thrive[124]. However, an infected cell can enter a long-lived state where the virus does not replicate called proviral latency. This state fails to generate a substantial viral load, and is therefore unaffected by anti-retroviral therapies (ART), which typically target viral machinery. One difficulty in treating and curing HIV comes from the latent reservoirs of HIV virus, as any interruption in ART allows these reservoirs to reactivate, pushing viral loads to levels on the order of those before any treatment took place[125].

Research has shown that the ability for HIV to enter proviral latency is an evolutionary method of “bet hedging,” beneficial in surviving periods where environmental conditions are unfavorable by developing a long-lived viral reservoir[125]. The decision between active infection and proviral latency appears to be strongly tied to stochastic fluctuations in transcriptional activity and a combination of positive and negative feedback mechanisms[126].

In the current understanding of the HIV-1 gene expression network, HIV-1 mRNA is produced at a low basal rate. Once transcribed, mRNA molecules are serially spliced: the full mRNA molecule is fully transcribed before being spliced into smaller mRNA transcripts (unlike parallel splicing, where the mRNA is spliced as it is being transcribed)[127]. The full length mRNA transcripts are spliced into many transcripts, two of which encode for Tat and Rev. Tat is a transcription trans-activator protein that introduces a positive feedback loop by binding to the trans-activated response element (TAR)[128]. Tat binding alters the properties of the transcription complex, allowing transcription to occur at an accelerated rate. Rev, on the other hand, is a protein that binds as a tetramer to full-length mRNA transcripts, exporting them from the nucleus to be packaged in newly formed viruses[129]. Because Rev removes mRNA transcripts from the local pool, it acts as a negative feedback loop, driving the steady state value of proteins down[130]. The combination of positive and negative feedback reveals the importance of noise; latent HIV genes produce mRNA in a highly noisy manner, unable to accumulate sufficient Tat populations to fully activate the gene. However, stochastic fluctuations in the mRNA population will at some time produce a sufficient number of Tat molecules to drive transcription into a highly productive state. Once a sufficient population of rev molecules have been produced, negative feedback then drives the system to export the newly produced mRNA, thereby settling into a state of active infection.

Consider the simplified model presented in figure A.40. The model produces mRNA at some transcription rate, and accumulates it in the cell. The mRNA transcribes protein, which begins to reversibly bind to the mRNA population to create a new population of bound molecules. This complex of mRNA and protein is exported from the cell, which produces a negative feedback loop where the production of mRNA and protein facilitates the export of mRNA from the system. In the model, splicing is removed, and it is assumed mRNA is immediately available for translation. To test the influence of the negative feedback loop due to mRNA export, comparisons were made where the binding of the protein and the export of mRNA was removed (rates set to 0). The removal of protein binding is equivalent to increasing the rate at which mRNA is spliced, such that no full-length mRNA is present long enough for rev to bind and export from the system.

The results of the two cases are shown in figure A.41. Trajectories associated with negative feedback due to mRNA export have significantly different noise properties compared to those without mRNA export. All trajectories begin the simulation time in the ON state, resulting in a common rise in protein abundance as time moves forward. There are two main differences between the cases: the system without negative feedback has high noise, characterized by high steady state values and high steady state variability, while the system with negative feedback is characterized by lower noise and lower steady state values. Without negative feedback, the protein population is allowed to increase to high values before completely decaying when the gene transitions to the OFF state. In contrast, the system with negative feedback cannot reach similar steady state values because mRNA molecules are continuously removed from the system. However, negative feedback lowers the noise by reducing the rate at which the protein population decays when the gene transitions to the OFF state. The reduced rate of decay is due to the latent complex population (mRNA bound to protein) that slowly releases mRNA molecules back into the system, lengthening the time protein can be translated. The comparison between the two cases highlights the importance of noise and negative feedback in the activation of HIV: without negative feedback, the production of protein is noisy, readily crashing to a population of 0 when the gene is inactive. However, with negative feedback, when a gene activates and produces protein, the reduced noise “locks in” the decision to activate, retaining both a nonzero population of protein and an active export of mRNA transcripts. While this simplified model explores the effects of negative feedback on noise in HIV regulation, further modeling work is still needed to fully explore the complex regulatory structure of HIV. Additional work may consider spatial influences on the measured noise in HIV regulation, as factors including chromatin remodeling, nuclear transport, or spatial localization may influence the magnitude of noise. The model helps illustrate the importance of noise in the behavior of a gene expression system.

5.3 The importance of spatial considerations in gene expression

Spatial considerations have been shown throughout this work to be an important part of the analysis and characterization bursty gene expression. As experiments on cellular and synthetic systems continue to rely on noise measurements for the analysis of gene expression bursting, it becomes increasingly important to assure that spatial factors are properly accounted for. As was shown by the work here, failure to account for spatial interactions including crowding, confinement, or the spatial distribution of resources can lead to significant differences between inferred behaviors. Additionally, the use of simplified models that cannot adequately describe a particular gene expression system may lead to inferences about burst dynamics that are erroneous.

Avenues of future work have been detailed, including the examination of the efficiency of resource utilization. The consideration of spatial effects may play an important role in many other future studies. Measurements of both mRNA and protein populations produced from the same gene can give insight into the different ways spatial effects change transcription and translation separately. While the exploration of macromolecular crowding was done through simple diffusion of single lattice site particles, more complex crowding situations should be explored, including crowders that behave as long chain polymers, similar to Ficoll 70, a common experimental crowding agent. Additionally, several recent papers have shown that the interior of bacterial cells are spatially organized, such that transcription and translation occur in different regions of the cell[131, 132]. This spatial organization may impart some kind of efficiency in the use of limited cellular resources, such that transcription or translation occurs more readily compared to a system which is uniformly distributed with expression machinery.

List of References

1. Smalley, R.E., *Future global energy prosperity: the terawatt challenge.* Mrs Bulletin, 2005. **30**(06): p. 412-417.
2. BP, *BP Statistical Review of World Energy 2015*, in *BP Technical Report*. 2015.
3. IEA, *International Energy Agency Key World Energy Statistics*. 2014.
4. Lewis, N.S. and D.G. Nocera, *Powering the planet: Chemical challenges in solar energy utilization.* Proceedings of the National Academy of Sciences, 2006. **103**(43): p. 15729-15735.
5. Vesborg, P.C. and T.F. Jaramillo, *Addressing the terawatt challenge: scalability in the supply of chemical elements for renewable energy.* RSC Advances, 2012. **2**(21): p. 7933-7947.
6. Boden, T.A., G. Marland, and R.J. Andres, *Global, regional, and national fossil-fuel CO2 emissions.* Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tenn., USA doi, 2009. **10**.
7. Solomon, S., et al., *IPCC, 2007: summary for policymakers.* Climate change, 2007: p. 93-129.
8. Energy, U.D.o., *US Primary Energy Consumption by Source and Sector, 2008*. 2009, US Dept. of Energy.
9. Association, R.F., *Accelerating industry innovation: 2012 ethanol industry outlook*. 2012: Renewable Fuels Association.
10. Pimentel, D. and T.W. Patzek, *Ethanol production using corn, switchgrass, and wood; biodiesel production using soybean and sunflower.* Natural resources research, 2005. **14**(1): p. 65-76.
11. Yu, P., et al., *Conversion of food industrial wastes into bioplastics*, in *Biotechnology for Fuels and Chemicals*. 1998, Springer. p. 603-614.
12. Bhardwaj, R., et al., *Renewable resource-based green composites from recycled cellulose fiber and poly (3-hydroxybutyrate-co-3-hydroxyvalerate) bioplastic.* Biomacromolecules, 2006. **7**(6): p. 2044-2051.
13. Domenek, S., et al., *Biodegradability of wheat gluten based bioplastics.* Chemosphere, 2004. **54**(4): p. 551-559.
14. Sudo, K. and K. Shimizu, *A new carbon fiber from lignin.* Journal of applied polymer science, 1992. **44**(1): p. 127-134.
15. Kadla, J., et al., *Lignin-based carbon fibers for composite fiber applications.* Carbon, 2002. **40**(15): p. 2913-2920.
16. Roberts, T., *The Carbon Fiber Industry worldwide 2011-2020.* Materials Technology Publications, Watford, 2011. **6**: p. 29.
17. Benner, S.A. and A.M. Sismour, *Synthetic biology.* Nature Reviews Genetics, 2005. **6**(7): p. 533-543.
18. Andrianantoandro, E., et al., *Synthetic biology: new engineering rules for an emerging discipline.* Molecular systems biology, 2006. **2**(1).
19. Purnick, P.E. and R. Weiss, *The second wave of synthetic biology: from modules to systems.* Nature reviews Molecular cell biology, 2009. **10**(6): p. 410-422.
20. Khalil, A.S. and J.J. Collins, *Synthetic biology: applications come of age.* Nature Reviews Genetics, 2010. **11**(5): p. 367-379.

21. Weber, W., et al., *A synthetic time-delay circuit in mammalian cells and mice.* Proceedings of the National Academy of Sciences, 2007. **104**(8): p. 2643-2648.
22. Basu, S., et al., *Spatiotemporal control of gene expression with pulse-generating networks.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(17): p. 6355-6360.
23. Gardner, T.S., C.R. Cantor, and J.J. Collins, *Construction of a genetic toggle switch in Escherichia coli.* Nature, 2000. **403**(6767): p. 339-342.
24. Stricker, J., et al., *A fast, robust and tunable synthetic gene oscillator.* Nature, 2008. **456**(7221): p. 516-519.
25. Yokobayashi, Y., R. Weiss, and F.H. Arnold, *Directed evolution of a genetic circuit.* Proceedings of the National Academy of Sciences, 2002. **99**(26): p. 16587-16591.
26. Atsumi, S. and J.C. Liao, *Metabolic engineering for advanced biofuels production from Escherichia coli.* Current opinion in biotechnology, 2008. **19**(5): p. 414-419.
27. van Maris, A.J., et al., *Alcoholic fermentation of carbon sources in biomass hydrolysates by Saccharomyces cerevisiae: current status.* Antonie Van Leeuwenhoek, 2006. **90**(4): p. 391-418.
28. Mandal, D., et al., *The use of microorganisms for the formation of metal nanoparticles and their application.* Applied Microbiology and Biotechnology, 2006. **69**(5): p. 485-492.
29. Nozik, A., *Quantum dot solar cells.* Physica E: Low-dimensional Systems and Nanostructures, 2002. **14**(1): p. 115-120.
30. Jarboe, L., et al., *Development of ethanologenic bacteria*, in *Biofuels*. 2007, Springer. p. 237-261.
31. Struhl, K., *Fundamentally different logic of gene regulation in eukaryotes and prokaryotes.* Cell, 1999. **98**(1): p. 1-4.
32. Taniguchi, Y., et al., *Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells.* Science, 2010. **329**(5991): p. 533-538.
33. Golding, I., et al., *Real-time kinetics of gene activity in individual bacteria.* Cell, 2005. **123**(6): p. 1025-1036.
34. Peccoud, J. and B. Ycart, *Markovian modeling of gene-product synthesis.* Theoretical population biology, 1995. **48**(2): p. 222-234.
35. Sanchez, A. and I. Golding, *Genetic determinants and cellular constraints in noisy gene expression.* Science, 2013. **342**(6163): p. 1188-1193.
36. Raj, A., et al., *Stochastic mRNA synthesis in mammalian cells.* PLoS Biol, 2006. **4**(10): p. e309.
37. Suter, D.M., et al., *Origins and consequences of transcriptional discontinuity.* Current opinion in cell biology, 2011. **23**(6): p. 657-662.
38. Chubb, J.R., et al., *Transcriptional pulsing of a developmental gene.* Current biology, 2006. **16**(10): p. 1018-1025.
39. Dar, R.D., et al., *Transcriptional burst frequency and burst size are equally modulated across the human genome.* Proceedings of the National Academy of Sciences, 2012. **109**(43): p. 17454-17459.
40. Pedraza, J.M. and J. Paulsson, *Effects of molecular memory and bursting on fluctuations in gene expression.* Science, 2008. **319**(5861): p. 339-343.

41. Hebenstreit, D., *Are gene loops the cause of transcriptional noise?* Trends in Genetics, 2013. **29**(6): p. 333-338.
42. Kepler, T.B. and T.C. Elston, *Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations.* Biophysical journal, 2001. **81**(6): p. 3116-3136.
43. Simpson, M.L., C.D. Cox, and G.S. Sayler, *Frequency domain chemical Langevin analysis of stochasticity in gene transcriptional regulation.* Journal of theoretical biology, 2004. **229**(3): p. 383-394.
44. To, T.-L. and N. Maheshri, *Noise can induce bimodality in positive transcriptional feedback loops without bistability.* Science, 2010. **327**(5969): p. 1142-1145.
45. Suter, D.M., et al., *Mammalian genes are transcribed with widely different bursting kinetics.* Science, 2011. **332**(6028): p. 472-474.
46. Sanchez, A., et al., *Effect of promoter architecture on the cell-to-cell variability in gene expression.* PLoS Comput Biol, 2011. **7**(3): p. e1001100.
47. Levens, D. and D.R. Larson, *A new twist on transcriptional bursting.* Cell, 2014. **158**(2): p. 241-242.
48. Chong, S., et al., *Mechanism of transcriptional bursting in bacteria.* Cell, 2014. **158**(2): p. 314-326.
49. Zenklusen, D., D.R. Larson, and R.H. Singer, *Single-RNA counting reveals alternative modes of gene expression in yeast.* Nature structural & molecular biology, 2008. **15**(12): p. 1263-1271.
50. Blake, W.J., et al., *Noise in eukaryotic gene expression.* Nature, 2003. **422**(6932): p. 633-637.
51. Berg, O.G., *A model for the statistical fluctuations of protein numbers in a microbial population.* Journal of theoretical biology, 1978. **71**(4): p. 587-603.
52. Thattai, M. and A. Van Oudenaarden, *Intrinsic noise in gene regulatory networks.* Proceedings of the National Academy of Sciences, 2001. **98**(15): p. 8614-8619.
53. Marcello, T., et al., *Interferons α and λ inhibit hepatitis C virus replication with distinct signal transduction and gene regulation kinetics.* Gastroenterology, 2006. **131**(6): p. 1887-1898.
54. Beato, M., *Gene regulation by steroid hormones*, in *Gene Expression*. 1993, Springer. p. 43-75.
55. Paulsson, J., *Models of stochastic gene expression.* Physics of life reviews, 2005. **2**(2): p. 157-175.
56. Lee, W.-P. and W.-S. Tzou, *Computational methods for discovering gene networks from expression data.* Briefings in bioinformatics, 2009. **10**(4): p. 408-423.
57. Karlebach, G. and R. Shamir, *Modelling and analysis of gene regulatory networks.* Nature Reviews Molecular Cell Biology, 2008. **9**(10): p. 770-780.
58. Bolouri, H. and E.H. Davidson, *Modeling transcriptional regulatory networks.* BioEssays, 2002. **24**(12): p. 1118-1129.
59. Meyer, B., et al., *Geometry-induced bursting dynamics in gene expression.* Biophysical journal, 2012. **102**(9): p. 2186-2191.
60. Elcock, A.H., *Atomic-level observation of macromolecular crowding effects: escape of a protein from the GroEL cage.* Proceedings of the National Academy of Sciences, 2003. **100**(5): p. 2340-2344.

61. Ermak, D.L. and J. McCammon, *Brownian dynamics with hydrodynamic interactions.* The Journal of chemical physics, 1978. **69**(4): p. 1352-1360.

62. Gillespie, D.T., *Exact stochastic simulation of coupled chemical reactions.* The journal of physical chemistry, 1977. **81**(25): p. 2340-2361.

63. Gillespie, D.T., A. Hellander, and L.R. Petzold, *Perspective: Stochastic algorithms for chemical kinetics.* The Journal of chemical physics, 2013. **138**(17): p. 170901.

64. Klann, M., A. Ganguly, and H. Koeppl, *Hybrid spatial Gillespie and particle tracking simulation.* Bioinformatics, 2012. **28**(18): p. i549-i555.

65. Gillespie, D.T., *Stochastic simulation of chemical kinetics.* Annu. Rev. Phys. Chem., 2007. **58**: p. 35-55.

66. Li, H., et al., *Algorithms and software for stochastic simulation of biochemical reacting systems.* Biotechnology progress, 2008. **24**(1): p. 56-61.

67. 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, 2000. **104**(9): p. 1876-1889.

68. Rathinam, M., et al., *Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method.* The Journal of Chemical Physics, 2003. **119**(24): p. 12784-12794.

69. Leff, P., *The two-state model of receptor activation.* Trends in pharmacological sciences, 1995. **16**(3): p. 89-97.

70. Cox, C.D., et al., *Using noise to probe and characterize gene circuits.* Proceedings of the National Academy of Sciences, 2008. **105**(31): p. 10809-10814.

71. Weinberger, L.S., R.D. Dar, and M.L. Simpson, *Transient-mediated fate determination in a transcriptional circuit of HIV.* Nature genetics, 2008. **40**(4): p. 466-470.

72. Austin, D., et al., *Gene network shaping of inherent noise spectra.* Nature, 2006. **439**(7076): p. 608-611.

73. Ozbudak, E.M., et al., *Regulation of noise in the expression of a single gene.* Nature genetics, 2002. **31**(1): p. 69-73.

74. Arkin, A., J. Ross, and H.H. McAdams, *Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected Escherichia coli cells.* Genetics, 1998. **149**(4): p. 1633-1648.

75. Tranquillo, R.T., D.A. Lauffenburger, and S. Zigmond, *A stochastic model for leukocyte random motility and chemotaxis based on receptor binding fluctuations.* The Journal of cell biology, 1988. **106**(2): p. 303-309.

76. McAdams, H.H. and A. Arkin, *Stochastic mechanisms in gene expression.* Proceedings of the National Academy of Sciences, 1997. **94**(3): p. 814-819.

77. Swain, P.S., M.B. Elowitz, and E.D. Siggia, *Intrinsic and extrinsic contributions to stochasticity in gene expression.* Proceedings of the National Academy of Sciences, 2002. **99**(20): p. 12795-12800.

78. Raser, J.M. and E.K. O'Shea, *Noise in gene expression: origins, consequences, and control.* Science, 2005. **309**(5743): p. 2010-2013.

79. Elowitz, M.B., et al., *Stochastic gene expression in a single cell.* Science, 2002. **297**(5584): p. 1183-1186.

80. Cox, C.D., et al., *Frequency domain analysis of noise in simple gene circuits.* Chaos: An Interdisciplinary Journal of Nonlinear Science, 2006. **16**(2): p. 026102.

81. Cox, C.D., et al., *Analysis of noise in quorum sensing.* OMICS A Journal of Integrative Biology, 2003. **7**(3): p. 317-334.

82. Hiratani, I., et al., *Global reorganization of replication domains during embryonic stem cell differentiation.* PLoS Biol, 2008. **6**(10): p. e245.

83. Balázsi, G., A. van Oudenaarden, and J.J. Collins, *Cellular decision making and biological noise: from microbes to mammals.* Cell, 2011. **144**(6): p. 910-925.

84. Weinberger, A.D. and L.S. Weinberger, *Stochastic fate selection in HIV-infected patients.* Cell, 2013. **155**(3): p. 497-499.

85. Weinberger, L.S., et al., *Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity.* Cell, 2005. **122**(2): p. 169-182.

86. Weinberger, L.S. and T. Shenk, *An HIV feedback resistor: auto-regulatory circuit deactivator and noise buffer.* PLoS Biol, 2007. **5**(1): p. e9.

87. Paulsson, J., *Summing up the noise in gene networks.* Nature, 2004. **427**(6973): p. 415-418.

88. Simpson, M.L., C.D. Cox, and G.S. Sayler, *Frequency domain analysis of noise in autoregulated gene circuits.* Proceedings of the National Academy of Sciences, 2003. **100**(8): p. 4551-4556.

89. So, L.H., et al., *General properties of transcriptional time series in Escherichia coli.* Nat Genet, 2011. **43**(6): p. 554-60.

90. Harper, C.V., et al., *Dynamic analysis of stochastic transcription cycles.* PLoS Biol, 2011. **9**(4): p. e1000607.

91. Larson, D.R., *What do expression dynamics tell us about the mechanism of transcription?* Current opinion in genetics & development, 2011. **21**(5): p. 591-599.

92. Kaern, M., et al., *Stochasticity in gene expression: from theories to phenotypes.* Nature Reviews Genetics, 2005. **6**(6): p. 451-464.

93. Roberts, E., et al., *Noise contributions in an inducible genetic switch: a whole-cell simulation study.* PLoS Comput Biol, 2011. **7**(3): p. 1002010.

94. Zimmerman, S.B. and S.O. Trach, *Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli.* Journal of molecular biology, 1991. **222**(3): p. 599-620.

95. Ellis, R.J., *Macromolecular crowding: obvious but underappreciated.* Trends in biochemical sciences, 2001. **26**(10): p. 597-604.

96. Verkman, A.S., *Solute and macromolecule diffusion in cellular aqueous compartments.* Trends in biochemical sciences, 2002. **27**(1): p. 27-33.

97. Minton, A.P., *The effect of volume occupancy upon the thermodynamic activity of proteins: some biochemical consequences.* Molecular and cellular biochemistry, 1983. **55**(2): p. 119-140.

98. Eggers, D.K. and J.S. Valentine, *Molecular confinement influences protein structure and enhances thermal protein stability.* Protein Science, 2001. **10**(2): p. 250-261.

99. van den Berg, B., R.J. Ellis, and C.M. Dobson, *Effects of macromolecular crowding on protein folding and aggregation.* The EMBO journal, 1999. **18**(24): p. 6927-6933.

100. Fowlkes, J.D. and C.P. Collier, *Single-molecule mobility in confined and crowded femtolitre chambers.* Lab on a Chip, 2013. **13**(5): p. 877-885.

101. van Zon, J.S., et al., *Diffusion of transcription factors can drastically enhance the noise in gene expression.* Biophysical journal, 2006. **91**(12): p. 4350-4367.

102. Morelli, M.J., R.J. Allen, and P.R. Ten Wolde, *Effects of macromolecular crowding on genetic networks.* Biophysical journal, 2011. **101**(12): p. 2882-2891.

103. Abel, S.M., et al., *The membrane environment can promote or suppress bistability in cell signaling networks.* The Journal of Physical Chemistry B, 2012. **116**(11): p. 3630-3640.

104. Mlynarczyk, P.J., R.H. Pullen III, and S.M. Abel, *Confinement and diffusion modulate bistability and stochastic switching in a reaction network with positive feedback.* The Journal of chemical physics, 2016. **144**(1): p. 015102.

105. Kumar, N., A. Singh, and R.V. Kulkarni, *Transcriptional bursting in gene expression: analytical results for general stochastic models.* PLoS Comput Biol, 2015. **11**(10): p. e1004292.

106. McQuarrie, D.A., *Stochastic approach to chemical kinetics.* Journal of applied probability, 1967. **4**(3): p. 413-478.

107. Nitzan, A. and J. Ross, *A comment on fluctuations around nonequilibrium steady states.* Journal of Statistical Physics, 1974. **10**(5): p. 379-390.

108. Cai, L., N. Friedman, and X.S. Xie, *Stochastic protein expression in individual cells at the single molecule level.* Nature, 2006. **440**(7082): p. 358-362.

109. Bendat, J.S. and A.G. Piersol, *Random data: analysis and measurement procedures*. Vol. 729. 2011: John Wiley & Sons.

110. Retterer, S.T., et al., *Development and fabrication of nanoporous silicon-based bioreactors within a microfluidic chip.* Lab on a Chip, 2010. **10**(9): p. 1174-1181.

111. Siuti, P., S.T. Retterer, and M.J. Doktycz, *Continuous protein production in nanoporous, picolitre volume containers.* Lab on a Chip, 2011. **11**(20): p. 3523-3529.

112. Karig, D.K., et al., *Probing cell-free gene expression noise in femtoliter volumes.* ACS synthetic biology, 2013. **2**(9): p. 497-505.

113. Norred, S.E., et al., *Sealable Femtoliter Chamber Arrays for Cell-free Biology*. Journal of visualized experiments: JoVE, 2015(97).

114. Nomura, S.i.M., et al., *Gene expression within cell‐sized lipid vesicles.* ChemBioChem, 2003. **4**(11): p. 1172-1175.

115. Kato, A., et al., *Cell-sized confinement in microspheres accelerates the reaction of gene expression.* Scientific reports, 2012. **2**.

116. Dar, R.D., et al., *The Low Noise Limit in Gene Expression.* PloS one, 2015. **10**(10): p. e0140969.

117. Ge, X., D. Luo, and J. Xu, *Cell-free protein expression under macromolecular crowding conditions.* PLoS One, 2011. **6**(12): p. e28707.

118. Shimizu, Y., et al., *Cell-free translation reconstituted with purified components.* Nature biotechnology, 2001. **19**(8): p. 751-755.

119. Kumar, M., M.S. Mommer, and V. Sourjik, *Mobility of cytoplasmic, membrane, and DNA-binding proteins in Escherichia coli.* Biophysical journal, 2010. **98**(4): p. 552-559.

120. So, L.-h., et al., *General properties of transcriptional time series in Escherichia coli.* Nature genetics, 2011. **43**(6): p. 554-560.

121. Chandra, K., *Statistical multiplexing.* Encyclopedia of Telecommunications, 2003.

122. Padovan-Merhar, O., et al., *Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms.* Molecular cell, 2015. **58**(2): p. 339-352.

123. Hansen, M.M., et al., *Macromolecular crowding creates heterogeneous environments of gene expression in picolitre droplets.* Nature nanotechnology, 2016. **11**(2): p. 191-197.

124. Weiss, R.A., *E EWC.* Science, 1993. **260**: p. 1273.

125. Rouzine, I.M., A.D. Weinberger, and L.S. Weinberger, *An evolutionary role for HIV latency in enhancing viral transmission.* Cell, 2015. **160**(5): p. 1002-1012.

126. Rouzine, I.M., B.S. Razooky, and L.S. Weinberger, *Stochastic variability in HIV affects viral eradication.* Proceedings of the National Academy of Sciences, 2014. **111**(37): p. 13251-13252.

127. Kornblihtt, A.R., et al., *Alternative splicing: a pivotal step between eukaryotic transcription and translation.* Nature reviews Molecular cell biology, 2013. **14**(3): p. 153-165.

128. Bohan, C., et al., *Analysis of Tat transactivation of human immunodeficiency virus transcription in vitro.* Gene expression, 1991. **2**(4): p. 391-407.

129. Brice, P.C., A.C. Kelley, and P.J.G. Butler, *Sensitive in vitro analysis of HIV-1 Rev multimerization.* Nucleic acids research, 1999. **27**(10): p. 2080-2085.

130. Felber, B.K., C.M. Drysdale, and G.N. Pavlakis, *Feedback regulation of human immunodeficiency virus type 1 expression by the Rev protein.* Journal of virology, 1990. **64**(8): p. 3734-3741.

131. Bakshi, S., et al., *Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells.* Molecular microbiology, 2012. **85**(1): p. 21-38.

132. Llopis, P.M., et al., *Spatial organization of the flow of genetic information in bacteria.* Nature, 2010. **466**(7302): p. 77-81.

Appendix

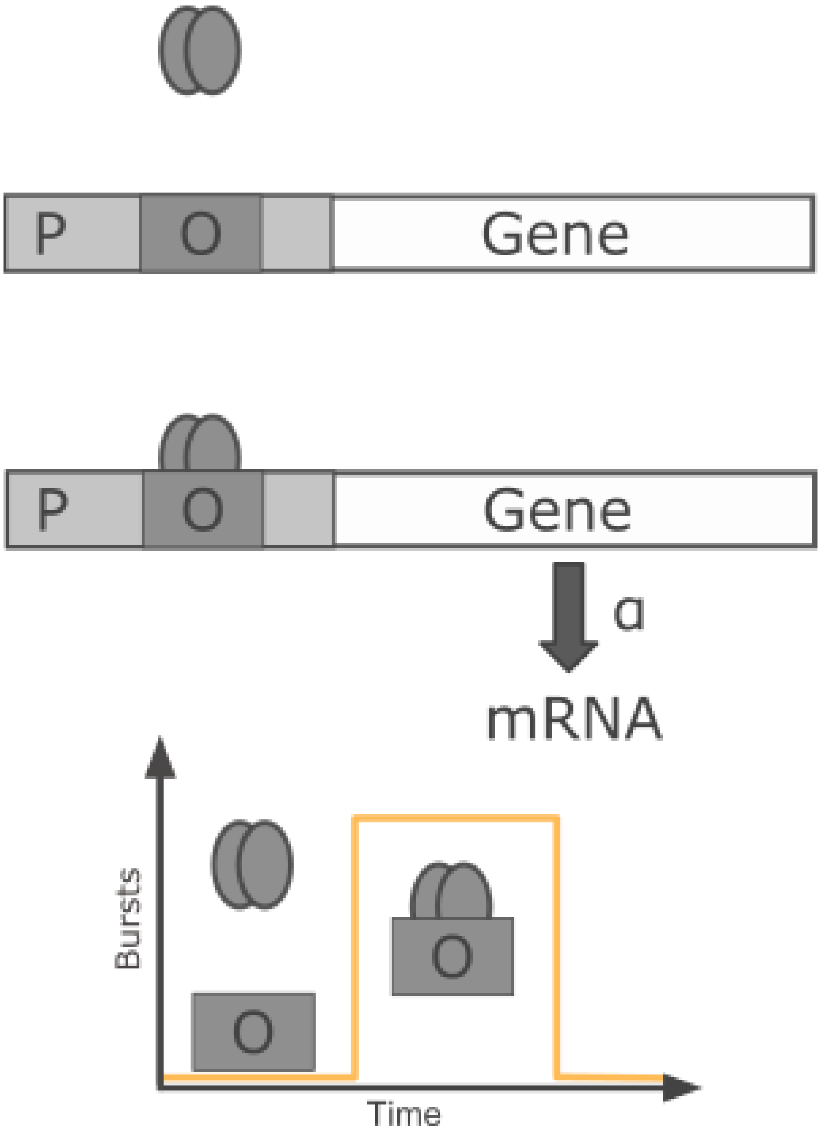

Figure A.1 – Example two state model of gene expression. The gene is activated through the binding of a molecular species to the operator site, which allows transcription to occur at a rate α. The activity of the gene is represented as a pulse function over time, where the binding of the molecular species causes the gene to change state instantaneously.

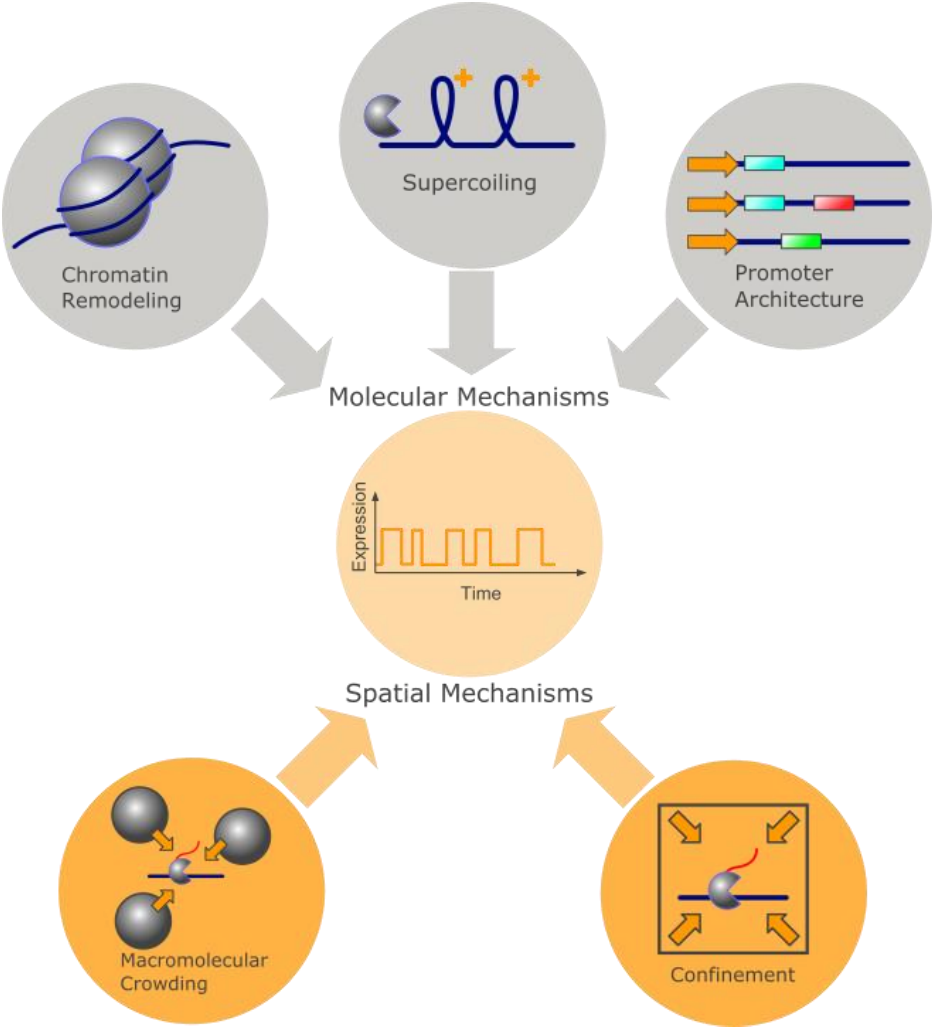

Figure A.2 – Cartoon graphic of the various suspected sources of transcriptional bursting. While there are a multitude of possible molecular mechanisms that can cause bursting in gene expression, spatial mechanisms, such as crowding and spatial confinement, are the focus of this thesis.

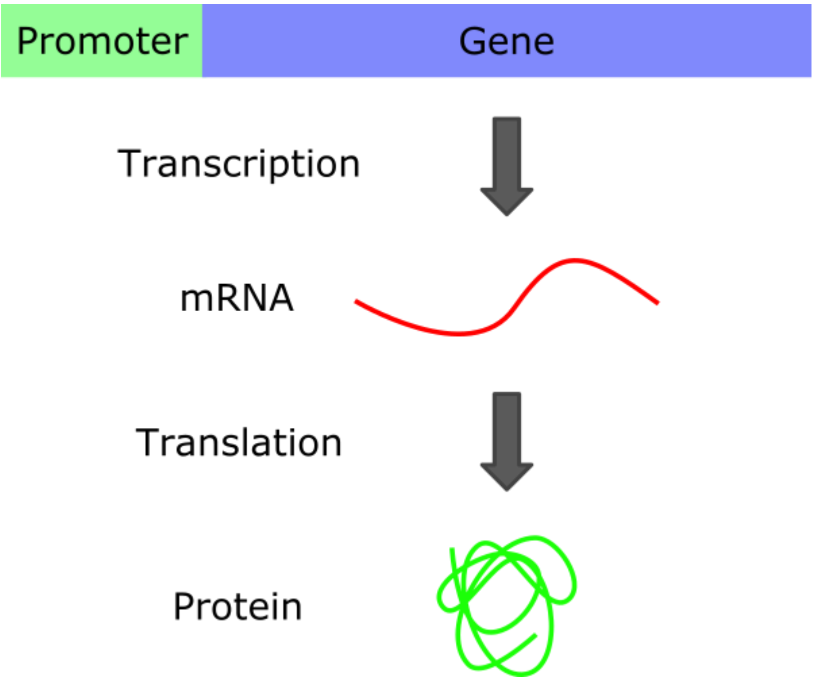

Figure A.3 – A simplified model of gene expression. In this model, a particular gene encodes for an mRNA molecule, which is transcribed when a polymerase attaches to the promoter region and moves along the length of a gene. The mRNA molecule is later translated by ribosomes to create proteins, which function in other areas of the cell.

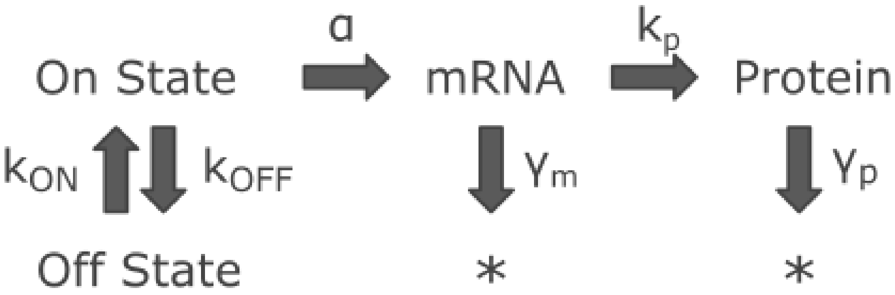

Figure A.4 – Schematic of the two state model. The two state model is a widely used model of bursty gene expression, where the low basal expression has been simplified to a state that produces no mRNA. * denotes a molecule decaying.

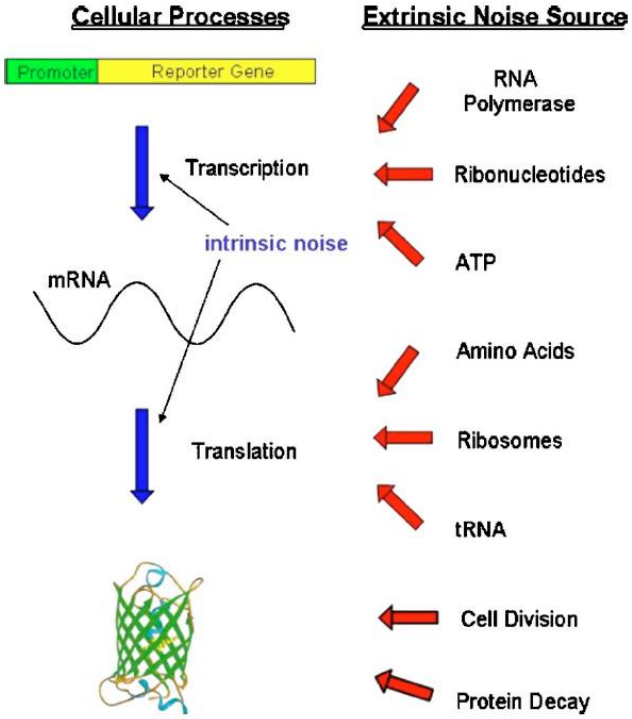

Figure A.5 – Intrinsic and extrinsic noise sources in gene expression. Intrinsic noise sources include both transcription and translation, and are inherent in the stochastic and discrete production of mRNA and proteins. Extrinsic noise, on the other hand, is a function of many global resources, things which are indirectly related to the main cellular process. Figure adapted from Cox et al, *Chaos* (2006)[80].

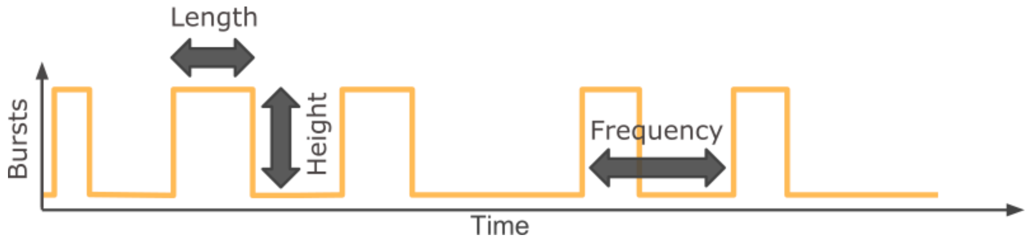

Figure A.6 – Example pulse train for a two-state bursty process. Burst size is a function of both the burst length (the duration the burst stays on) as well as the burst height (the amount of mRNA produced per unit time). Burst frequency is measured as the time between bursts, measured from where they start.

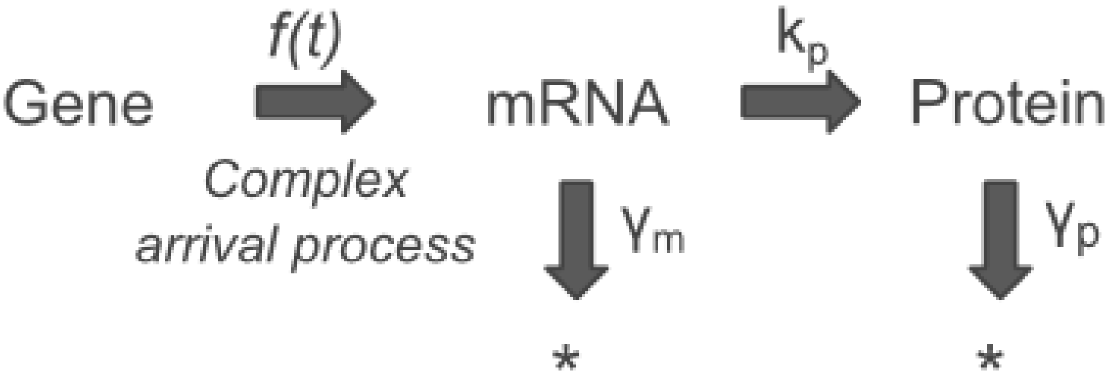

Figure A.7 – Model of complex arrivals of mRNA, based on Kulkarni et al.[105]. Instead of assuming an exponential distribution, mRNA arrive in bursts according to a function *f(t)* which describes the arrival rate of the mRNA bursts. mRNA subsequently produce protein and decay according to an exponential distribution.

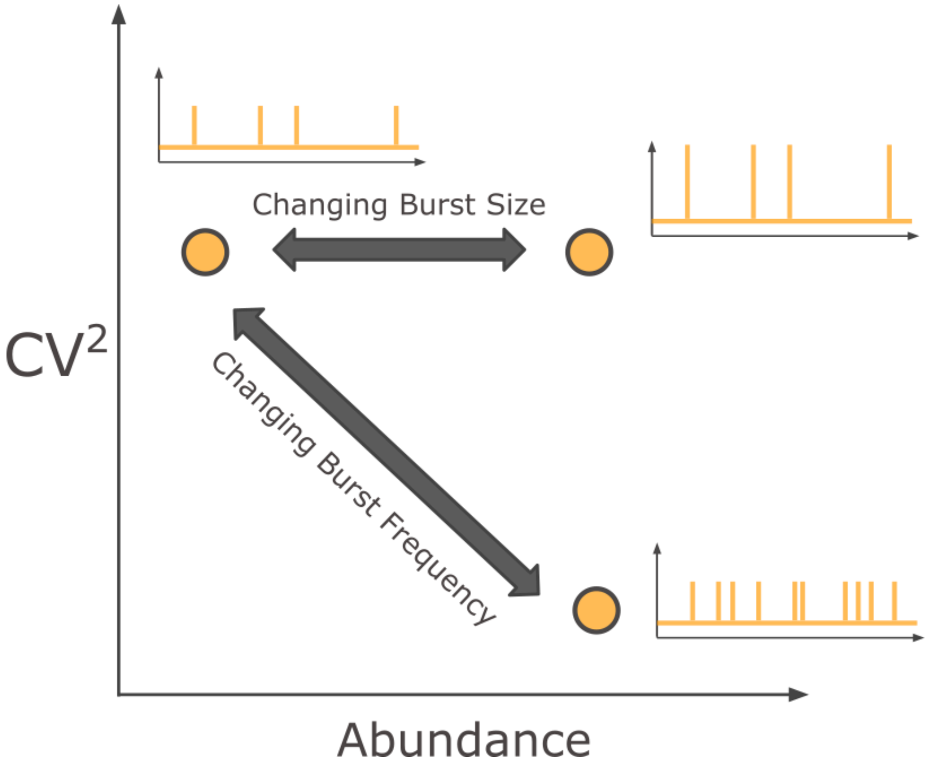

Figure A.8 – How changes in CV2 and abundance indicate shifts in burst frequency and burst size. A shift in burst size is consistent with a shift in abundance without a shift in CV^2, shown in the inset pulse trains as increase burst durations. A shift in burst frequency shows a decrease in CV^2 as abundance increases in an inverse relationship, demonstrated in the inset pulse trains by more closely spaced bursts in time.

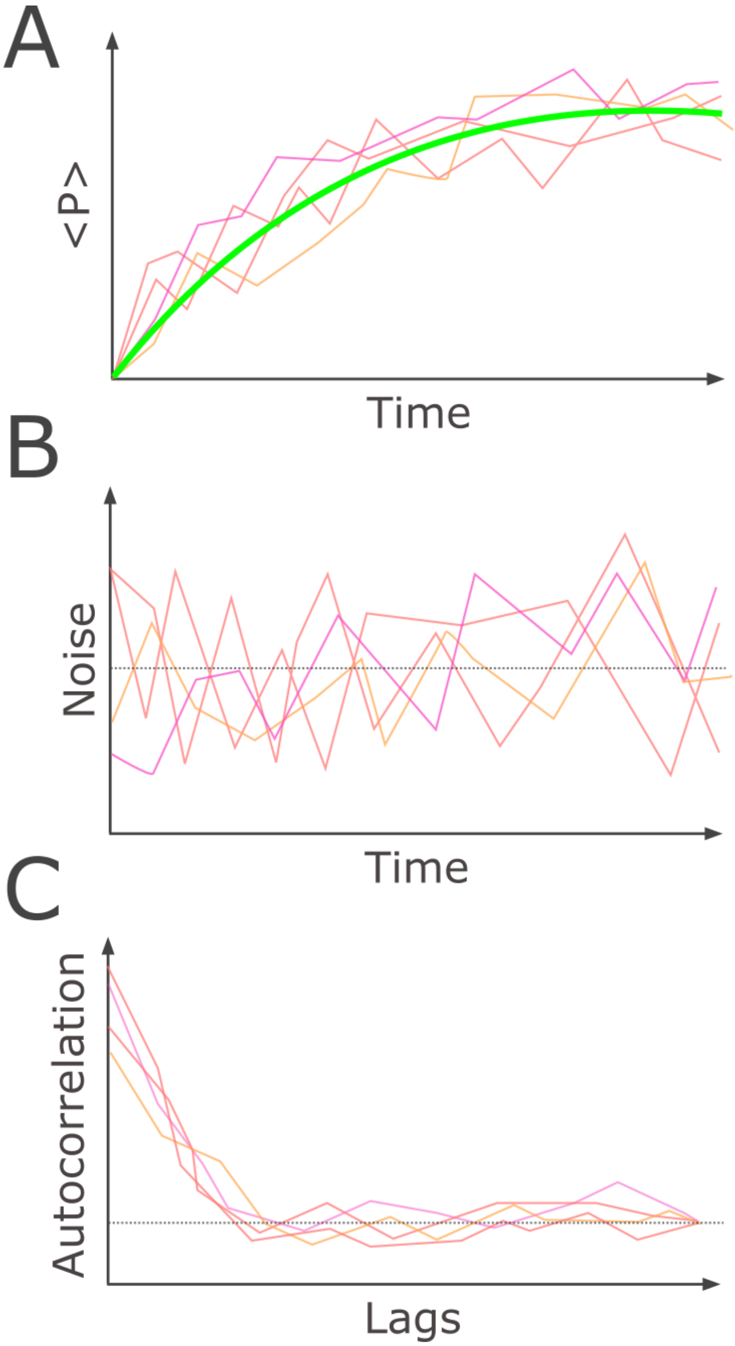

Figure A.9 – Example noise analysis process. In panel A, a group of traces is measured which has some general trend (green) associated with it. This general trend is removed using a gain factor to reveal panel B: the noise in the system. This noise is then autocorrelated, resulting in the traces in panel C. The 0 lag time value is equal to the variance of the trace, and is used to calculate CV^2. The dotted line in panels B and C represents a value of 0.

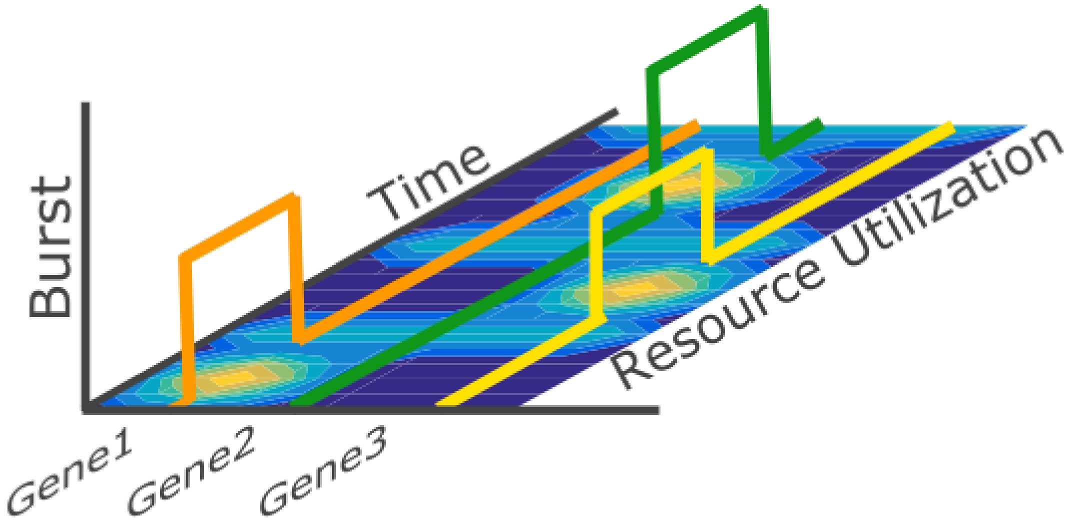

Figure A.10 – Time variant resource utilization. Bursty gene expression draws heavily from a shared pool of global resources, as shown by the color of the resource pool. However, resource utilization is done for limited duration and only when a gene is active (indicated by the step function in the “burst” axis). The bursts are separated in time, such that multiple genes share a single limited resource pool.

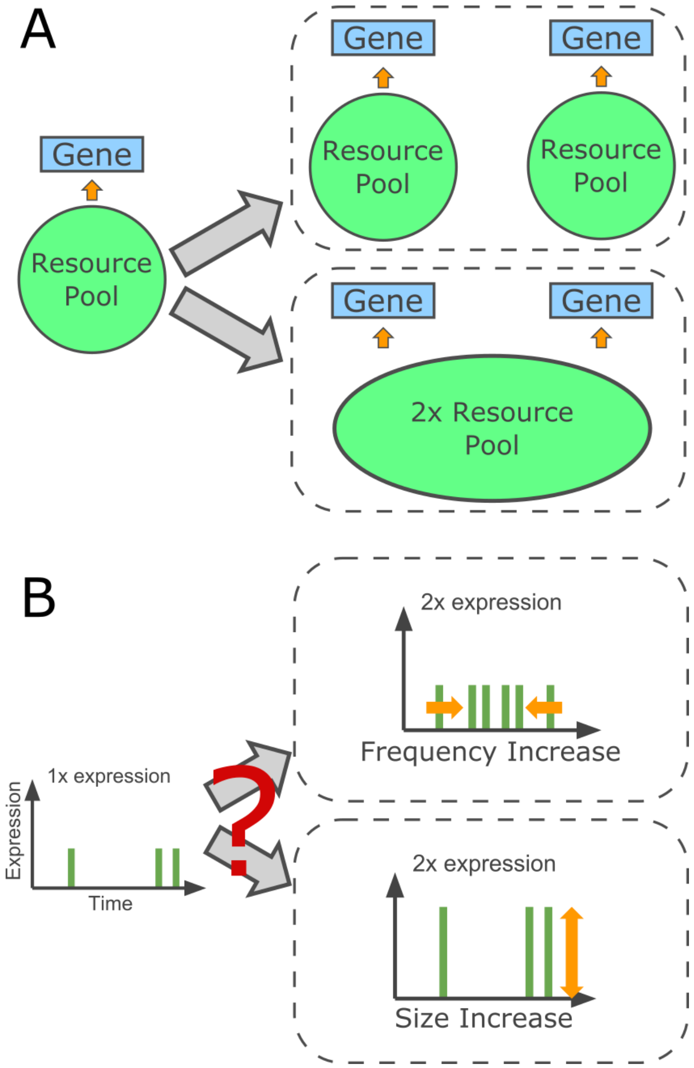

Figure A.11 - Resource use and bursty gene expression. (A) Protein abundance is increased through an increase in both the number of genes and the number of resources available. Resources can be shared either through enforced compartmentalization (top) or through a single shared resource pool (bottom). (B) Protein abundance change may be driven through an increase in burst frequency (top) or an increase in burst size (bottom). Different sharing scenarios were considered to determine whether it affects the expression bursting pattern.

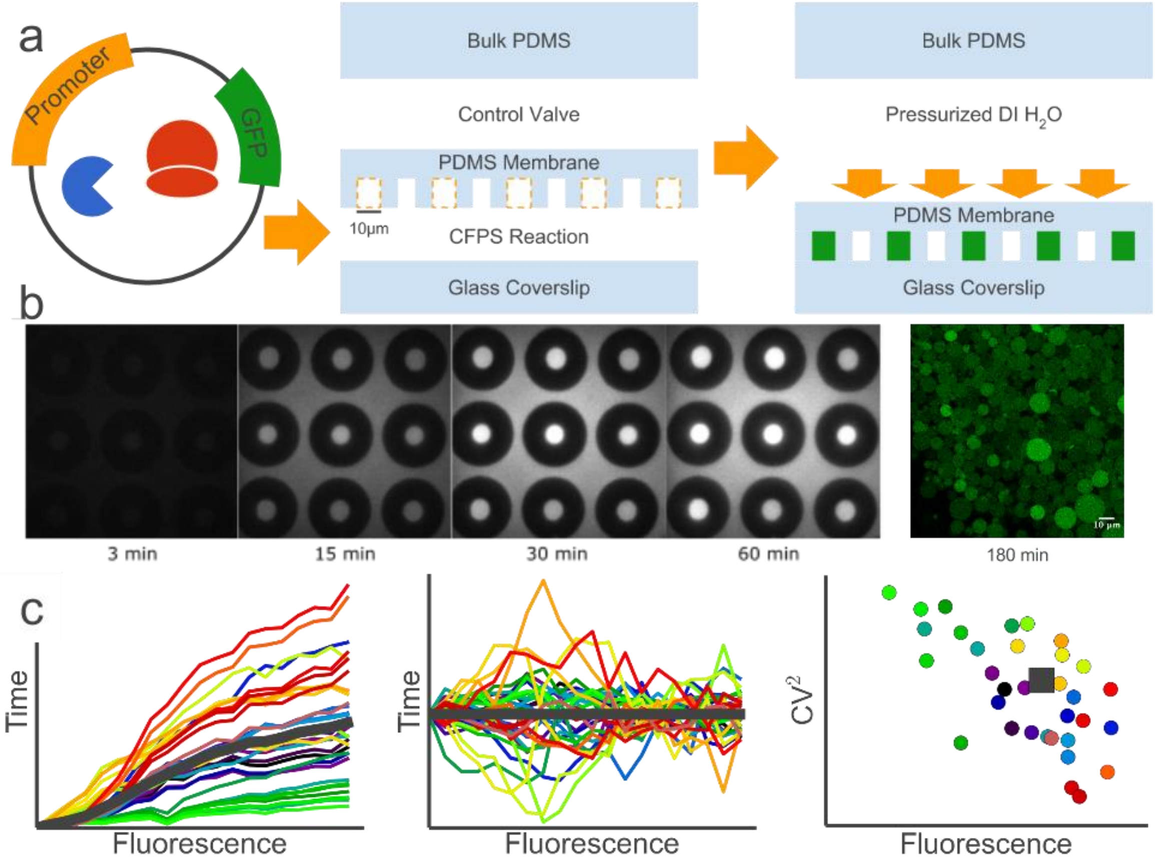

Figure A.12 - Confined cell-free gene expression and noise measurements. (a) Cell-free protein synthesis (CFPS) reactions were trapped within microfabricated chambers. (b) Time-lapse fluorescence microscopy was used to image the confined reactions every 3 minutes for 1 hour. Images are from an expression experiment performed in 10 μm-diameter reaction chambers show fluorescence intensity increasing over time. Scale bar, 20 mm. (Right) A representative z-slice of POPC vesicles expressing EGFP. (c) (left) The time history of the growth of the protein population was collected for each chamber. (middle) Gene expression noise was found by removing the deterministic general trend from each expression transient. (right) The CV2 and final fluorescence level (protein abundance) for individual chambers (colored circles) and for the average of all chambers (gray square) was determined. Adapted from submitted manuscript by Caveney et al.

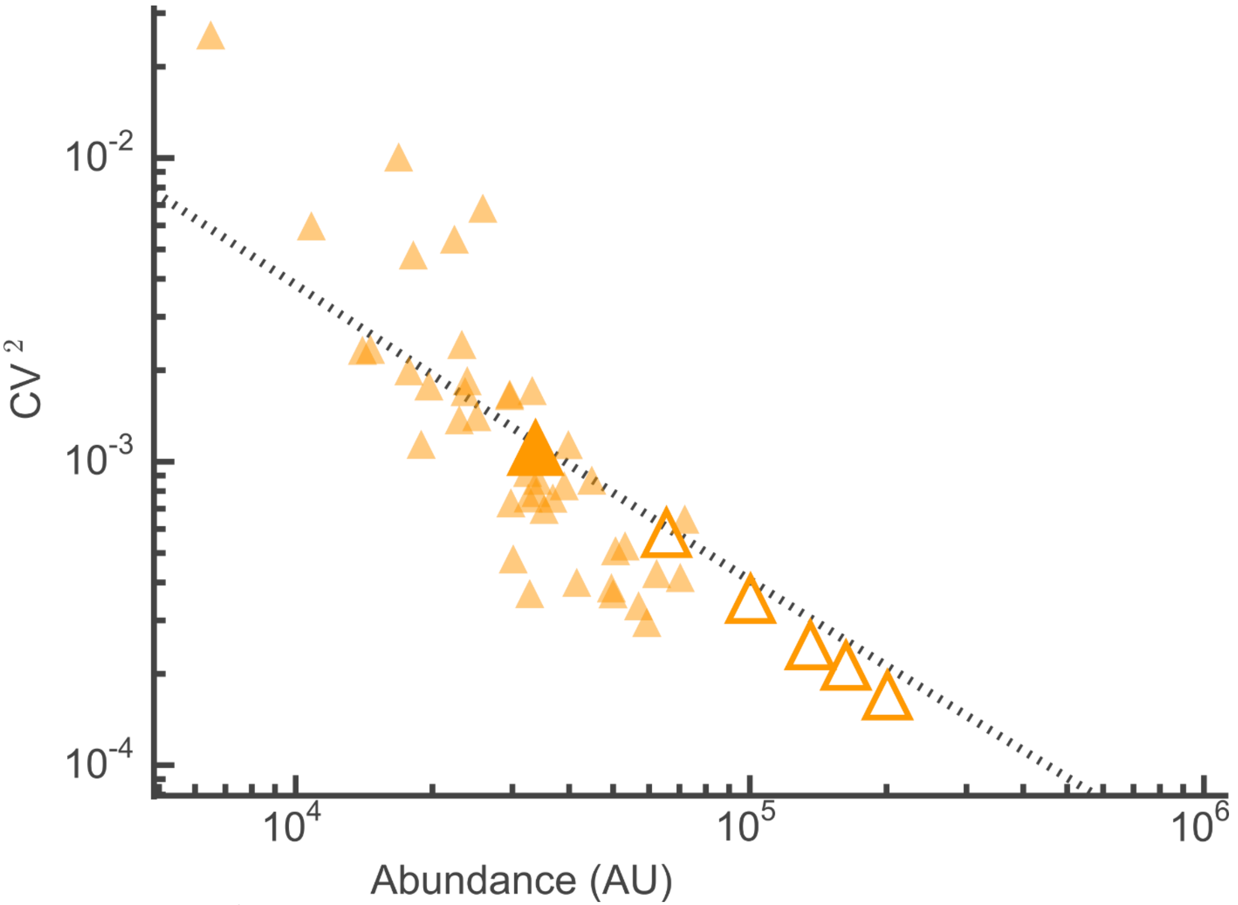

Figure A.13 – CV^2 vs Abundance for 2 µm individual and composite chambers. Individual chambers are denoted by filled triangles. The large filled triangle represents the mean value of all individual chambers. Empty triangles represent averages of composite chamber sums ranging from 2 summed chambers to 6 summed chambers.

Figure A.14 – Effects of resource pool size on gene expression noise in both microfluidic chambers and vesicles. (a) CV2 vs. abundance for 2, 5, and 10 µm diameter chambers. The small data points represent individual chambers while the large data points show the average behaviors for all chambers of a given size. Dashed gray line is a fit to the 2 µm chambers of the form a*(Abundance)^(-1), and highlights the Poissonian relationship between combinations of 2 µm chambers (open orange triangles). The inset shows volume vs. abundance is well approximated by a linear fit. (b) Histograms of abundance for 5 µm chambers and combinations of six 2 µm chambers (centroids in red box in (a)). Histograms are normalized and fit with normal distributions. (c) CV2 vs. abundance for vesicles ranging in diameter from 4 µm to 19 µm. Each data point is an individual vesicle. The orange points are vesicles with diameters 8-9 µm, and the blue points have diameters 18-19 µm. The solid gray line is a fit to all points of the form a*(Abundance)^b. Dashed lines are power fits to both size ranges with the exponent, b, equal to -2. The inset shows volume vs. abundance is well approximated by a linear fit. The orange region corresponds to the volume range of the chambers. (d) Same data in (a) without centroids. Dashed lines are power fits to each size chamber with the exponent, b, equal to -2.

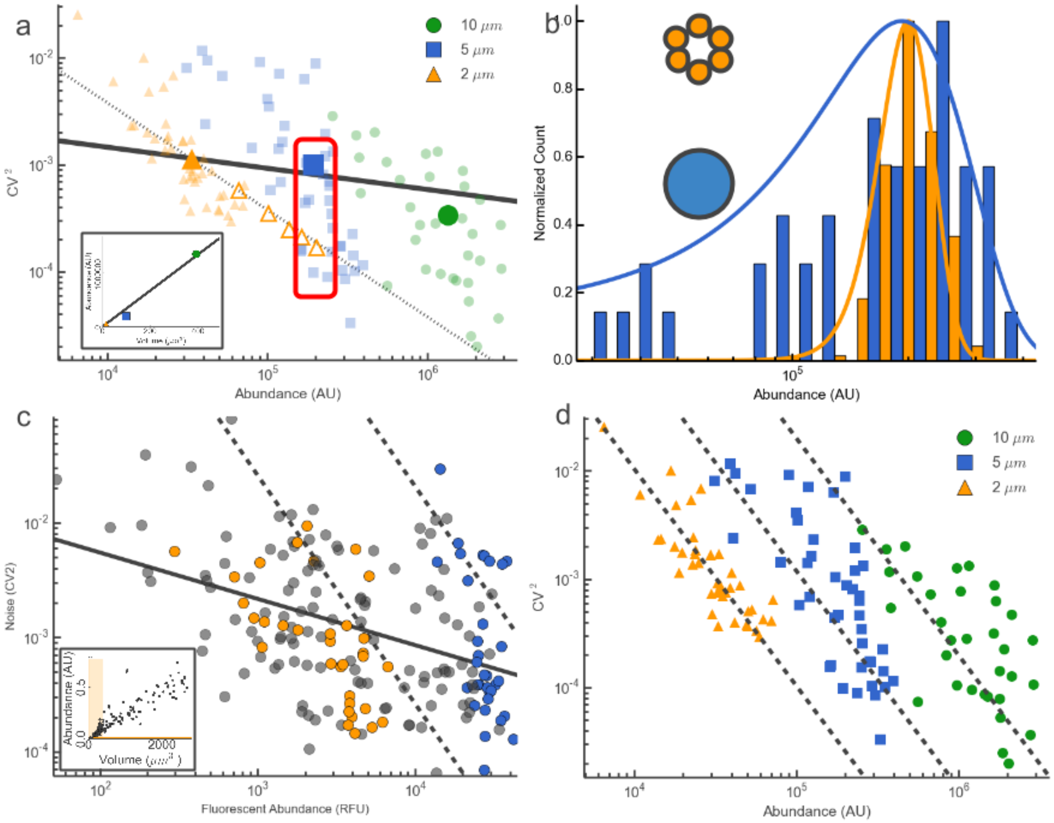

Figure A.14 continued

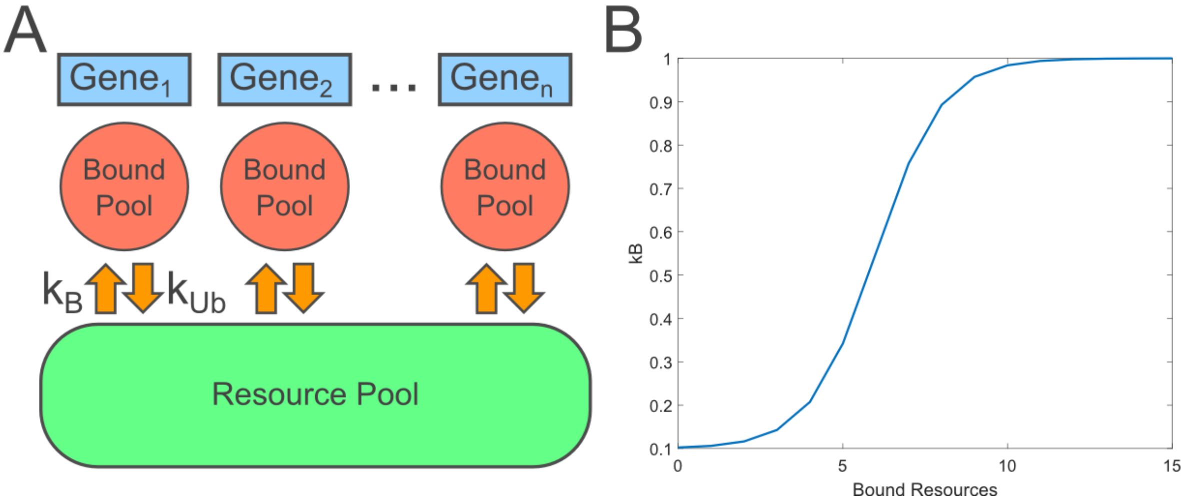

Figure A.15 – Ribosome binding model with positive feedback. A) A variable number of genes are placed in the system and resources enter a bound pool at a rate of k_B and leave at a rate k_{Ub}. The number of resources in the pool is proportional to the number of genes in the system. B) The binding curve of resources to genes was subjected to positive feedback, where bound resources increase the rate at which new resources enter the pool.

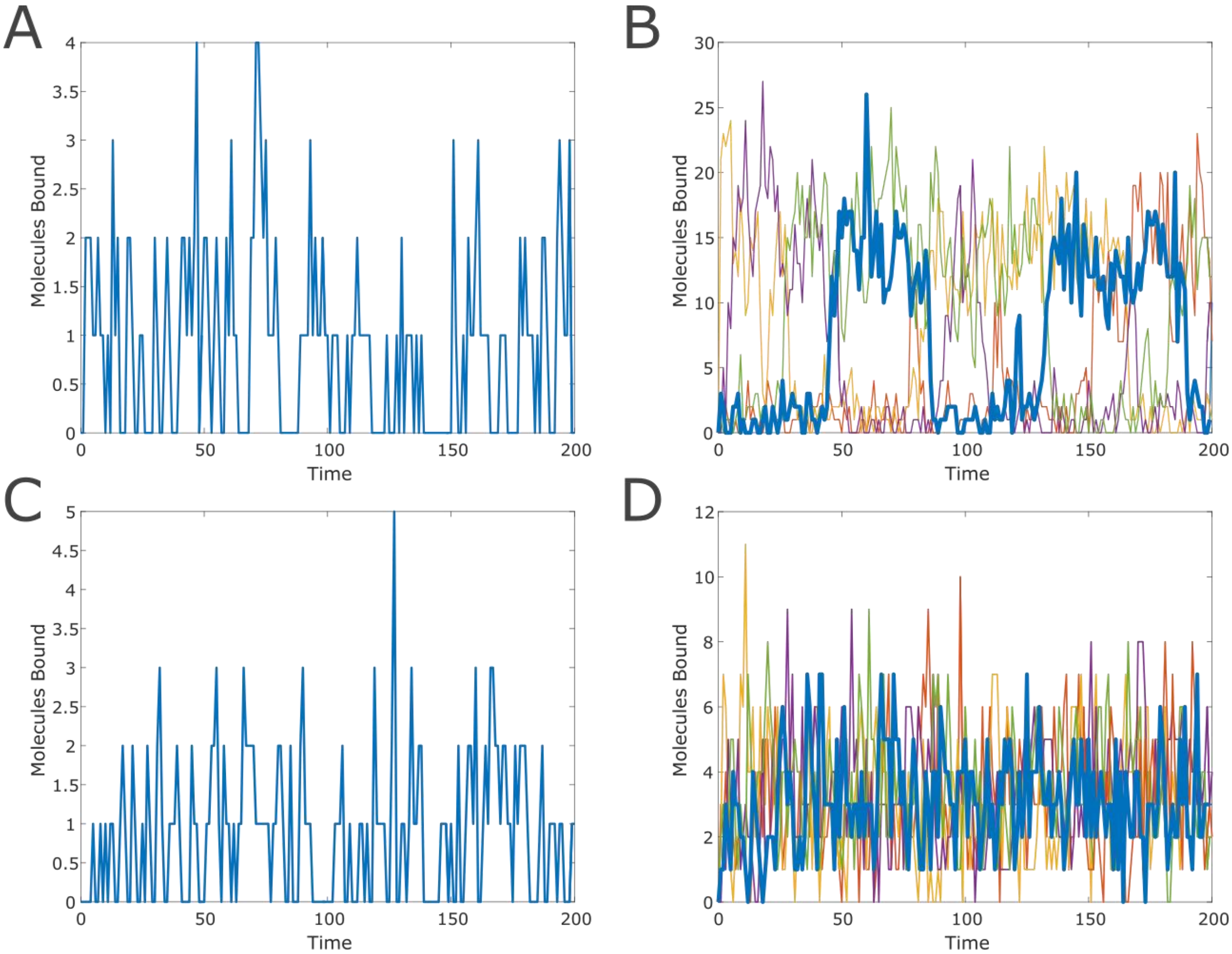

Figure A.16 – Bound molecule time traces for the shared resource pool model. A) and B) correspond to models with positive feedback, while C) and D) correspond to models without positive feedback. A) shows the number of bound resource molecules over time for a system with a single gene and a sigmoidal positive feedback curve. B) shows a system with 5 concurrent genes pulling from a single resource pool with positive feedback. Note how a single gene (outlined in blue) stochastically transitions from a low state to a high state. C) Shows a system with a single gene without positive feedback, while D) shows a system with 5 concurrent genes without positive feedback. Note that in the systems without positive feedback, there is no high or low state formation.

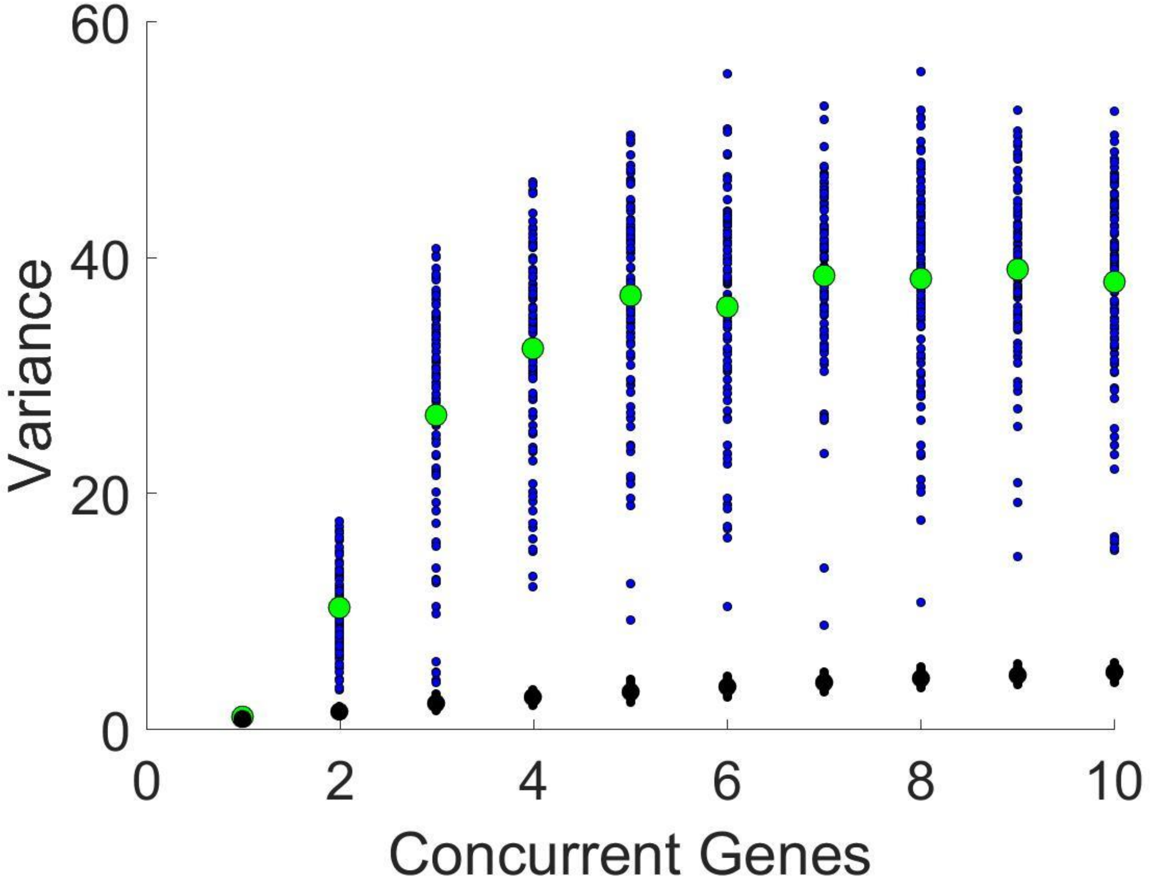

Figure A.17 - Variance points with and without positive feedback. Blue colored dots indicate one trajectory of some number of concurrent genes pulling from a pool of resources with positive feedback, with the larger green points denoting the mean. Black points show the variance and mean variance of the system without feedback. Note the extreme difference in the variance between the two cases at each number of genes.

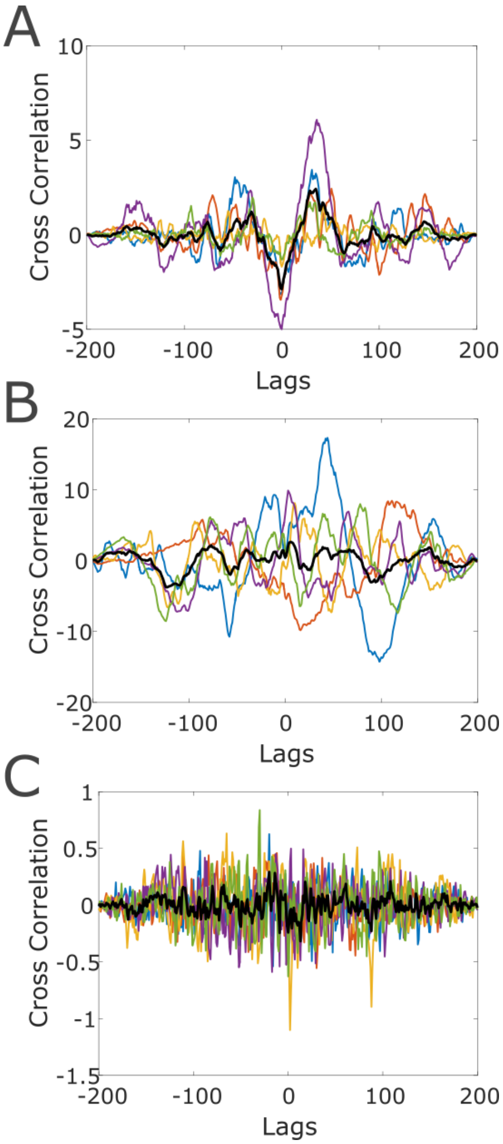

Figure A.18 – Cross correlation traces between genes. A) shows the strong anti-correlation at 0 lag between two genes with positive feedback in the same system. In systems with larger numbers of concurrent genes, as shown in B) where 5 concurrent genes are correlated amongst each other, the magnitude of the anti-correlation is reduced. For comparison, C) shows the correlation functions between genes in a 5 concurrent gene system without positive feedback.

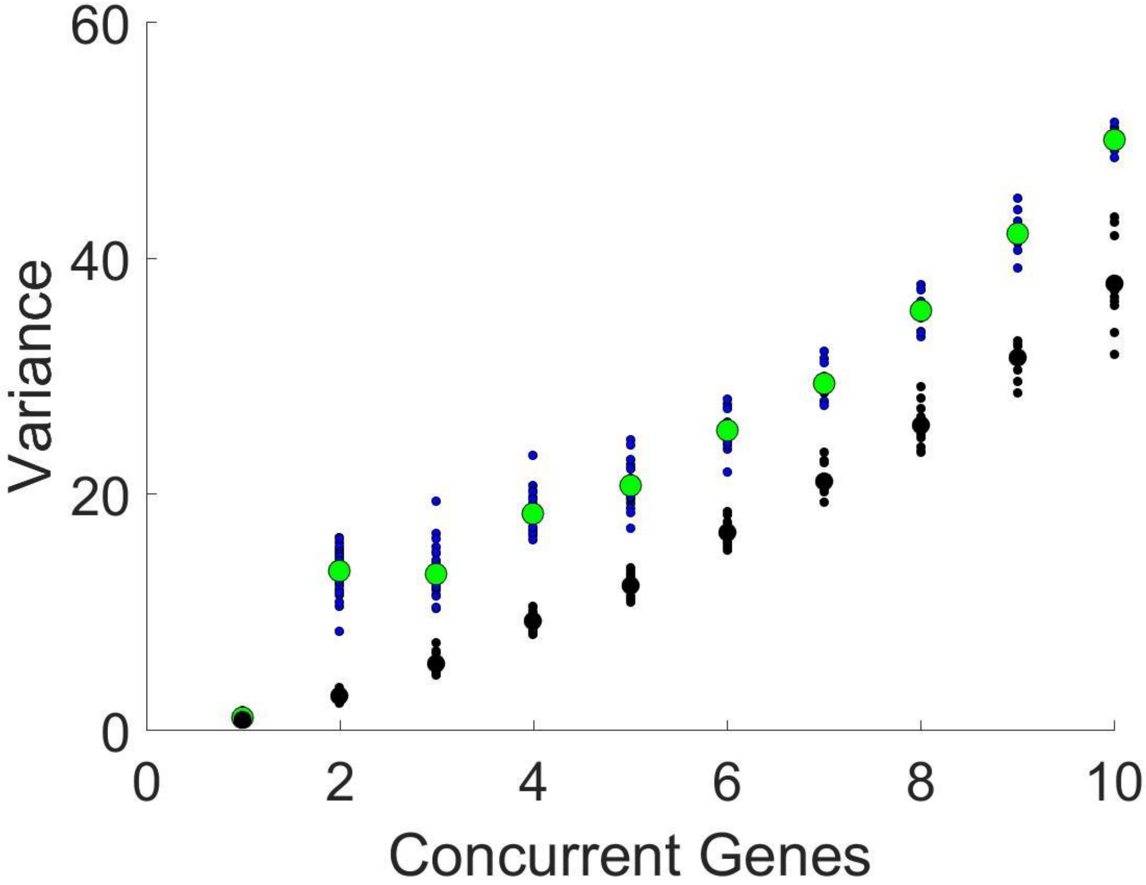

Figure A.19 – A comparison of variance in summed traces. Once all the genes in a given system are summed, the variance does not show a significant difference in behavior between the cases with positive feedback (blue points with green mean values) and those without positive feedback (black points).

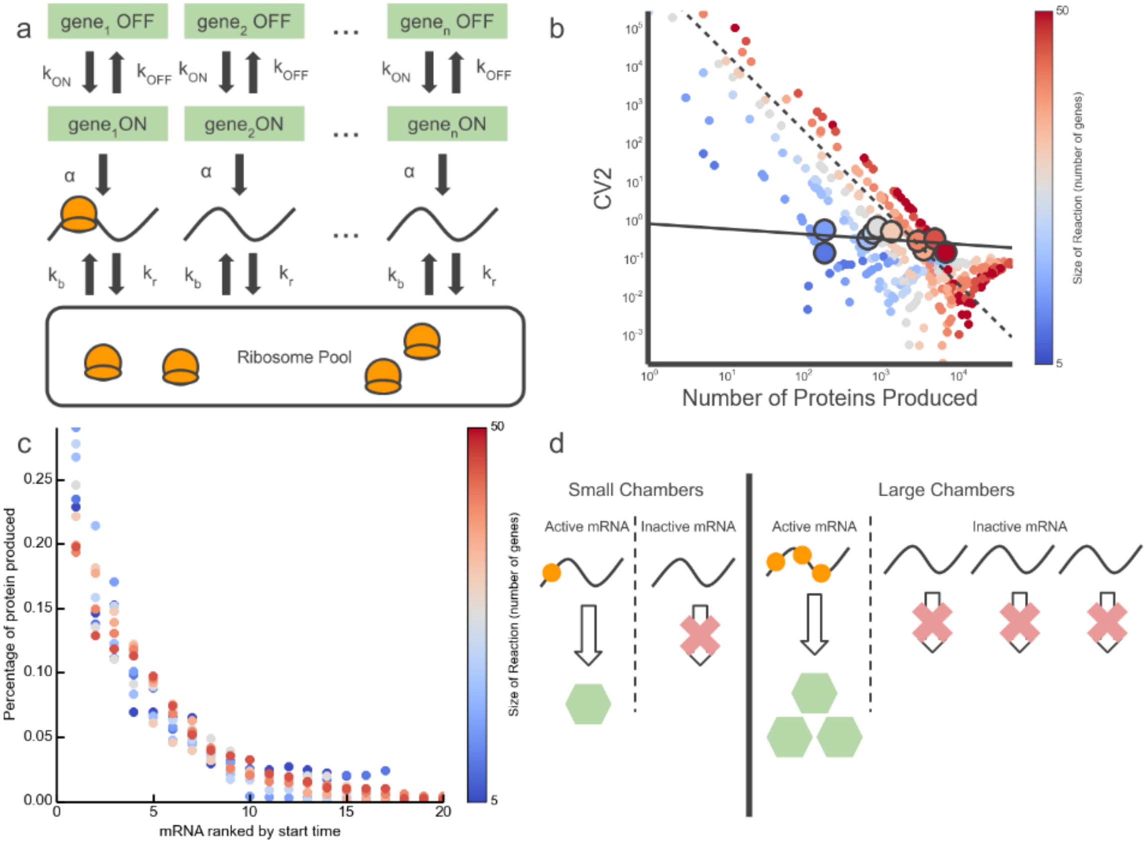

Figure A.20 – Model of the effects of resource pool size on expression bursting. (a) The model of resource sharing includes a resource pool of a limited number of reusable molecules, e.g. ribosomes, that associate with one of n genes at rate kn and return to the resource pool at rate gn. (b) CV2 vs Protein Abundance from the model described in (a). Colors represent the size of the reaction from 5 to 50 genes. Large points are geometric means. CV2 has a range of ~8 orders of magnitude while Protein Abundance spans ~5 orders of magnitude. The solid line is a power fit to all data points; the dashed line is a power fit with exponent -2 to one size chamber. (c) mRNA are ranked in the order they are produced. The amount of protein produced from each mRNA is normalized by the amount of protein the entire reaction produces. Points are colored by the reaction size. (d) Schematic of experimental results supported by the simulation in (a). Active mRNA in small chambers use a small resource pool (orange circles) and thus produce small amounts of protein (green hexagons). Conversely, active mRNA in large chambers use a large resource pool and thus produce large amounts of protein.

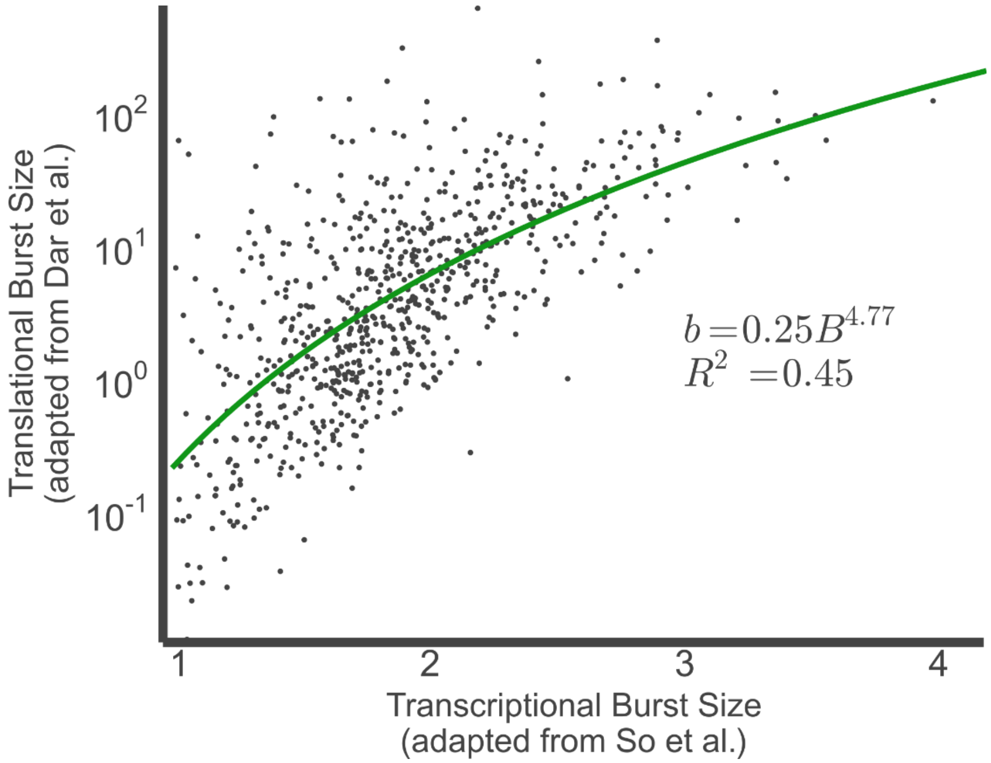

Figure A.21 – Transcriptional and translational burst size in *E. coli*. The comparison reveals strong correlations between the size of transcriptional bursting and the size of translational bursting. The solid line is a power law fit given by the equation in the graph. Each point represents data from an individual E. coli gene. Translational burst size adapted from Dar et al.[116], while transcriptional burst size adapted from So et al.[89].

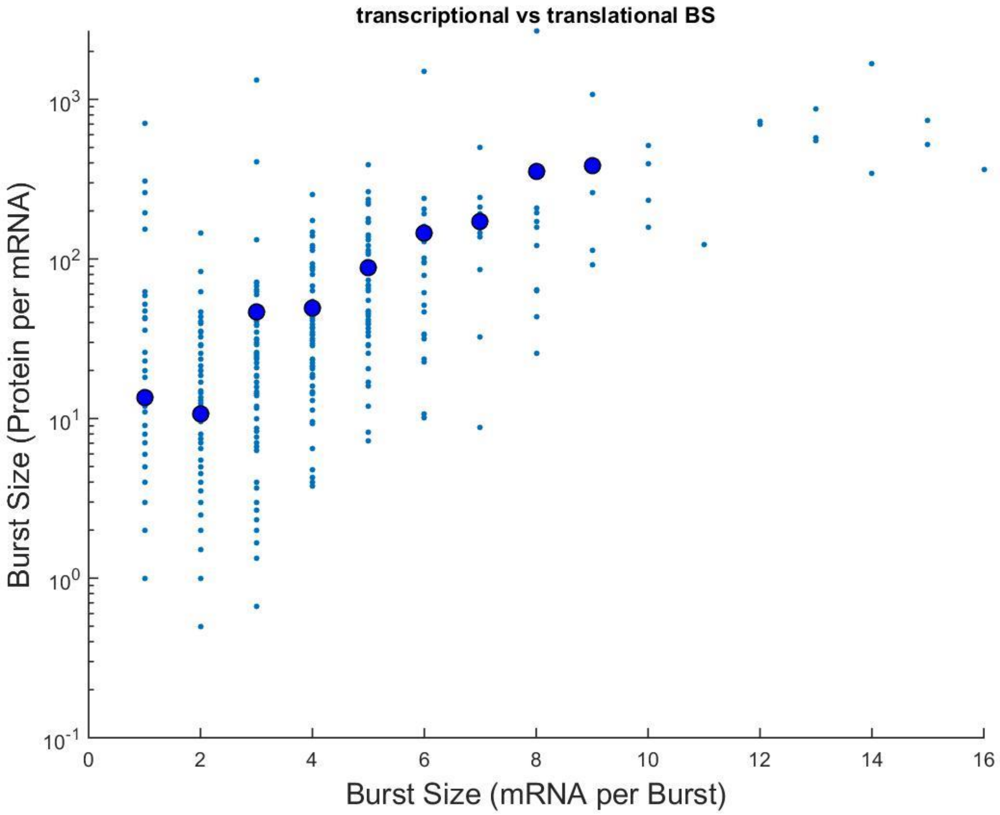

Figure A.22 – Transcriptional and translational burst size from model with two dependencies on mRNA population. Translational burst size increases two orders of magnitude as transcriptional burst size increases over one order of magnitude. Each small point represents the mean transcriptional and translational burst size of a single simulation trajectory. Large blue circles represent the average translational burst size of all simulations at a given transcriptional burst size.

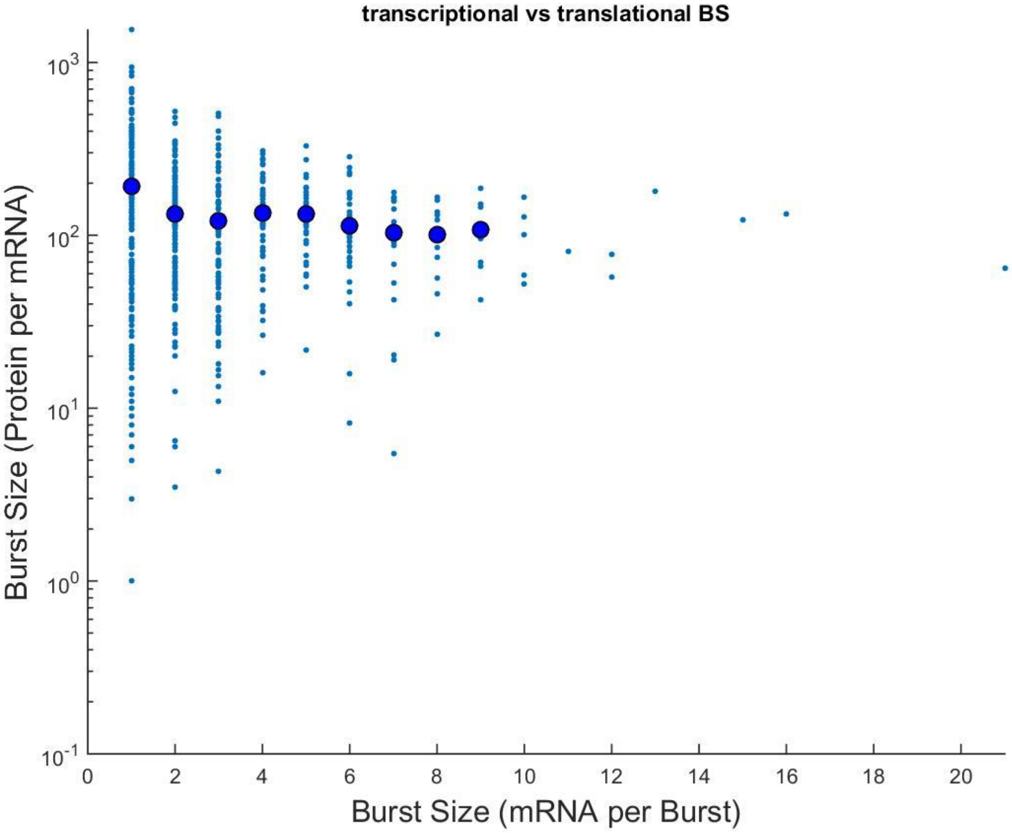

Figure A.23 – Transcriptional and translational burst size from model with one translational dependency on mRNA population. Translational burst size is independent of the transcriptional burst size in this model. Each small point represents the mean transcriptional and translational burst size of a single simulation trajectory. Large blue circles represent the average translational burst size of all simulations at a given transcriptional burst size.

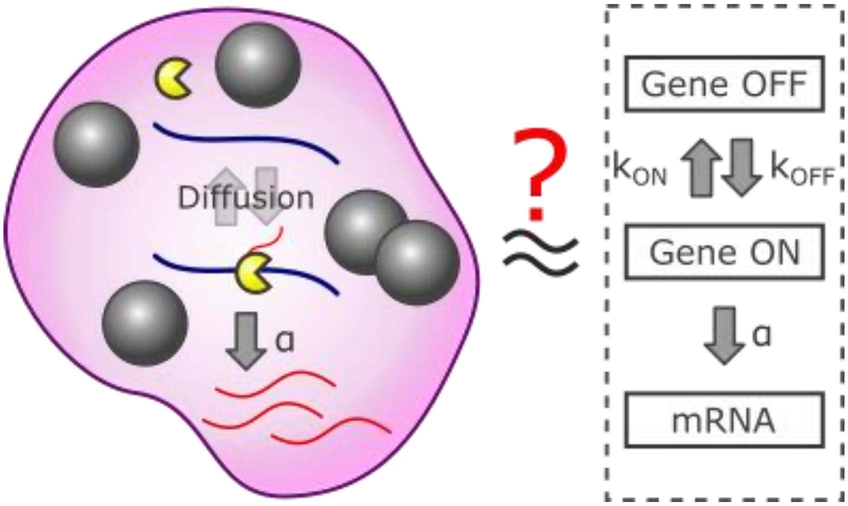

Figure A.24 – Can a cellular system be adequately described using a simple two-state model? In a cellular system, can various spatial factors, including crowding and confinement, be adequately captured using modified rate parameters in a two-state model?

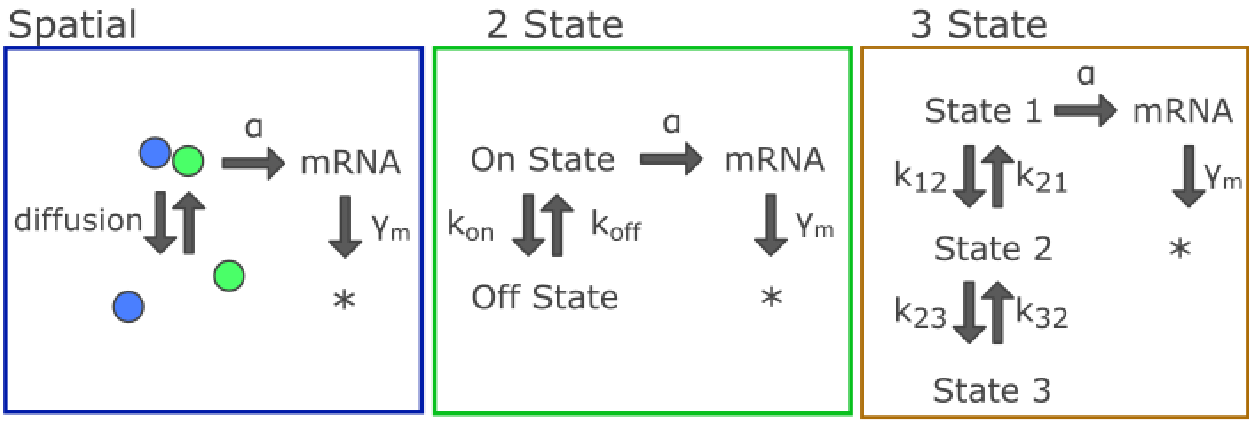

Figure A.25 – Spatial, two-state, and three-state models. All models transcribe mRNA through bursting by transitioning between a single ON state and one or more OFF states.

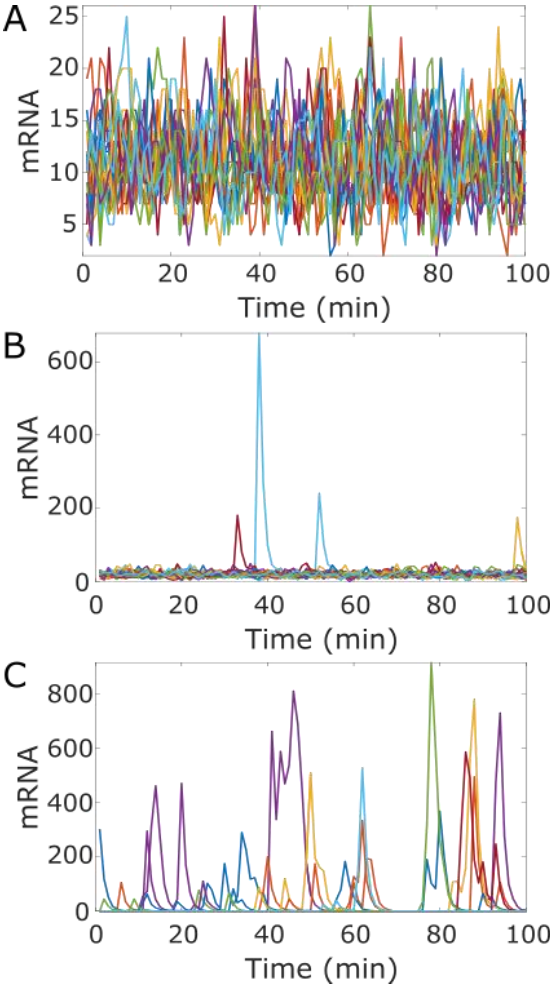

Figure A.26 – mRNA trajectories from spatially resolved simulations at various crowding and confinement regimes. (A) 20 sample mRNA traces over 100 minutes for a 16x16x16 cubic space with 0% crowding fraction. mRNA values vary around a steady state value set by the production and decay rate of the model. (B) 20 sample mRNA traces over 100 minutes for a 16x16x16 cubic space with 50% crowding fraction. In contrast to the 0% crowding case, short lived, but strong correlations driven by macromolecular crowding dynamics causes mRNA values to "spike," reaching high population values before decaying back to steady state. (C) 20 sample mRNA traces over 100 minutes for a 64x64x1 two dimensional space with 50% crowding fraction. Similar to the 50% case under the cubic geometry, strong, short lived correlations in macromolecular crowding cause "spikes" in mRNA populations. However, the addition of geometric confinement into two dimensions causes correlations that are longer lived, resulting in more frequent and longer spikes.

Figure A.27 – Noise analysis of the spatially resolved simulations at various crowding and confinement regimes.Noise magnitude as measured by CV2 is plotted against the average number of mRNA for the 16x16x16 lattice space with crowding fractions ranging from 0% to 50%. As the crowding fraction increases, the CV2 value initially decreases before increasing at high crowding fraction. The large fluctuations in mRNA as described in figure 3 result in outliers in CV2 and abundance space, and are more numerous at higher crowding fraction. (B) CV2 and average number of mRNA for the highest crowding fraction (50%) over the three different spatial geometries considered. Differences between the 16x16x16 space and the 32x32x4 space are minimal, while the highest geometric confinement at 64x64x1 results in a distribution of points which differs dramatically, revealing a wide range of abundance values over two orders of magnitude. (C) A plot of noise magnitude and mRNA population which reveals the difference in distributions when the diffusion coefficient of the crowding molecules is changed relative to the reacting particles. Slowly diffusing crowding molecules lead to the broadest resulting distribution of points, while faster crowder diffusion results in reduced noise magnitude. (D) A comparison of a well-mixed system at different crowding fractions where spatial effects due to crowding are incorporated into an effective volume term which modifies the transcription rate (colored points) against the spatial results (black points). This comparison reveals that the change in steady state mRNA population is due primarily to the excluded volume effects of the macromolecular crowders. The difference in noise magnitude is due to the spatial effects not captured by the well-mixed model.

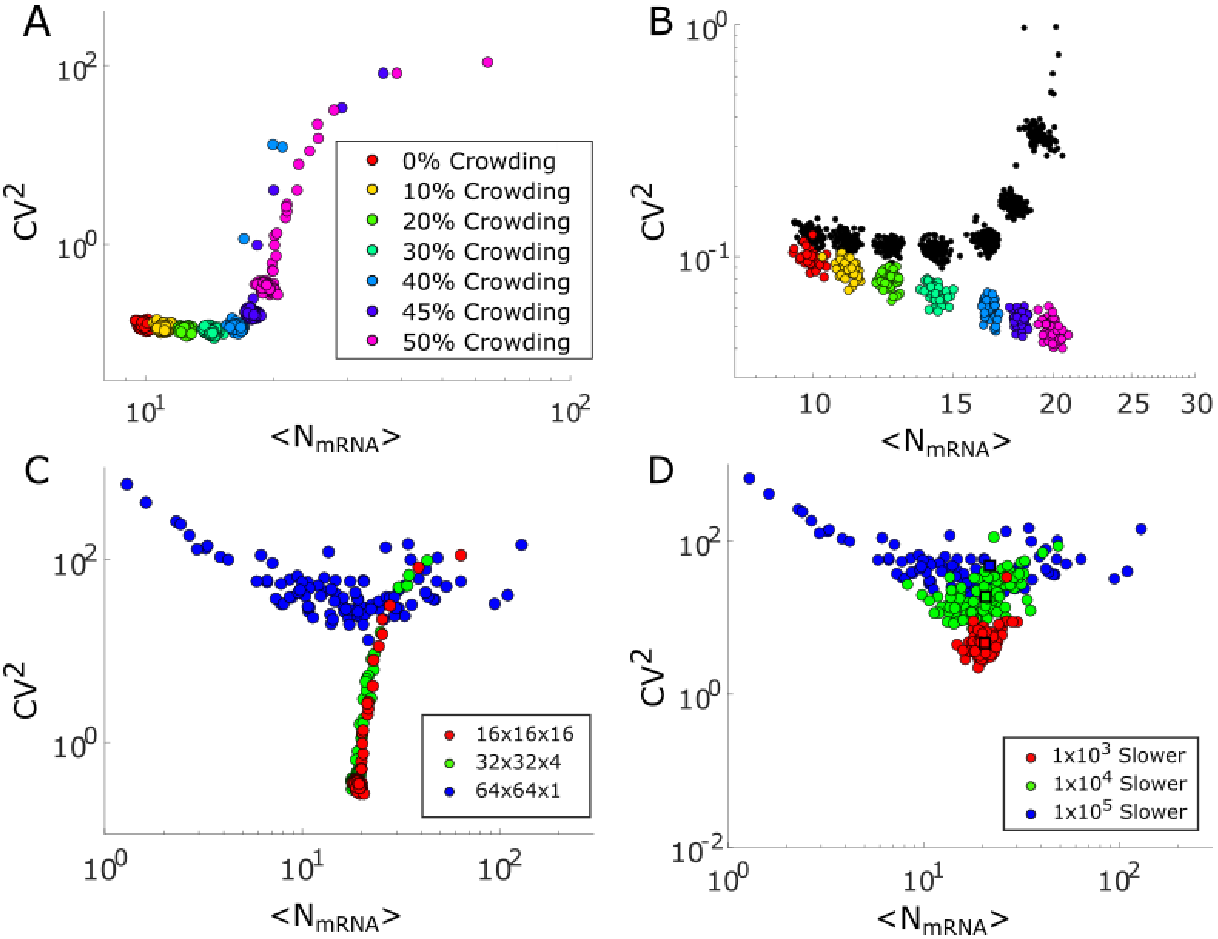

Figure A.27 continued

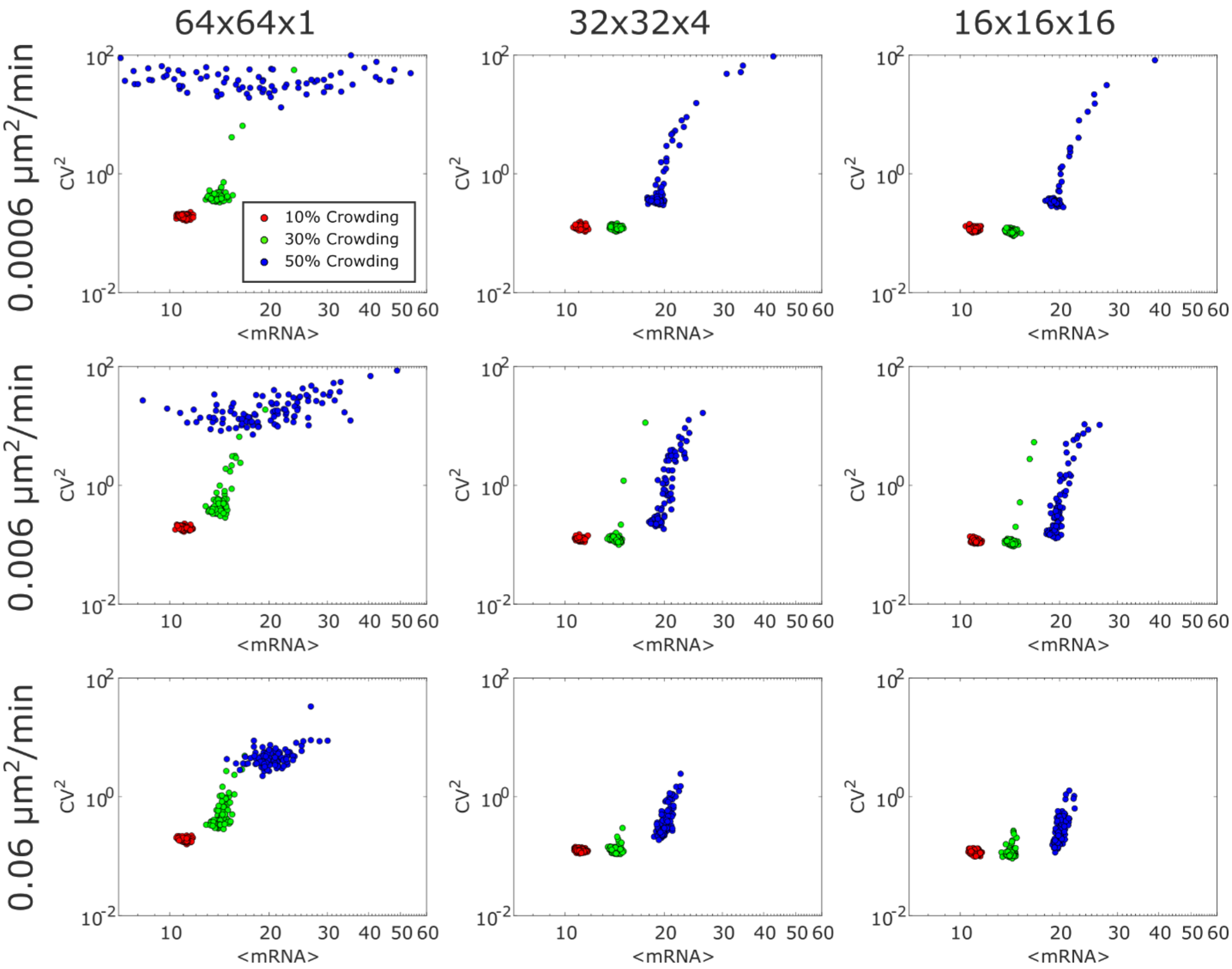

Figure A.28 – Compiled cases for the spatial model. Spatial model results are presented for all crowder diffusion rates (rows) and all confinement geometries (columns). Colors represent the various crowding fractions tested. Each data point represents the CV^2 and mRNA abundance value for a single simulation trajectory.

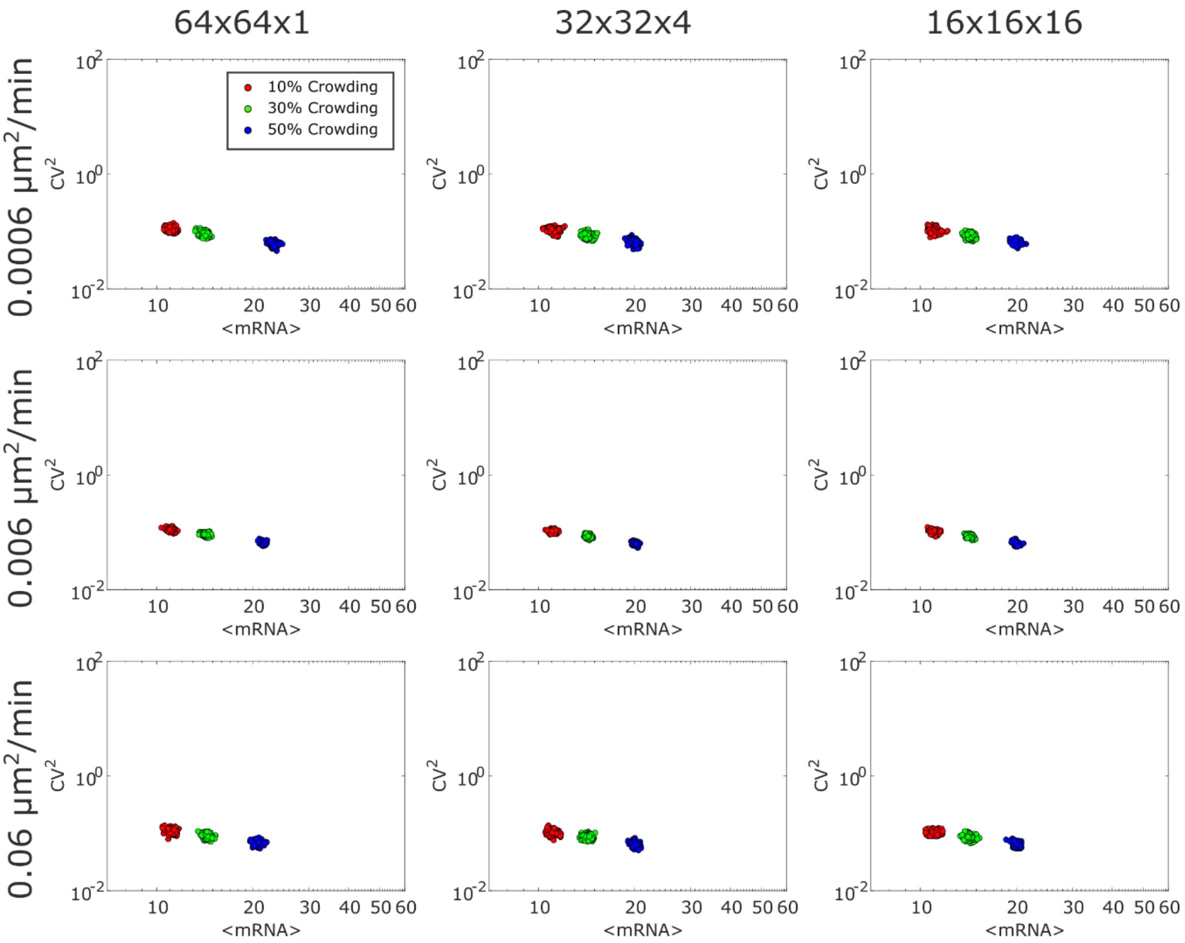

Figure A.29 – Compiled cases for the encounter method Encounter two-state model results are presented for all crowder diffusion rates (rows) and all confinement geometries (columns). Colors represent the various crowding fractions tested. Each data point represents the CV^2 and mRNA abundance value for a single simulation trajectory.

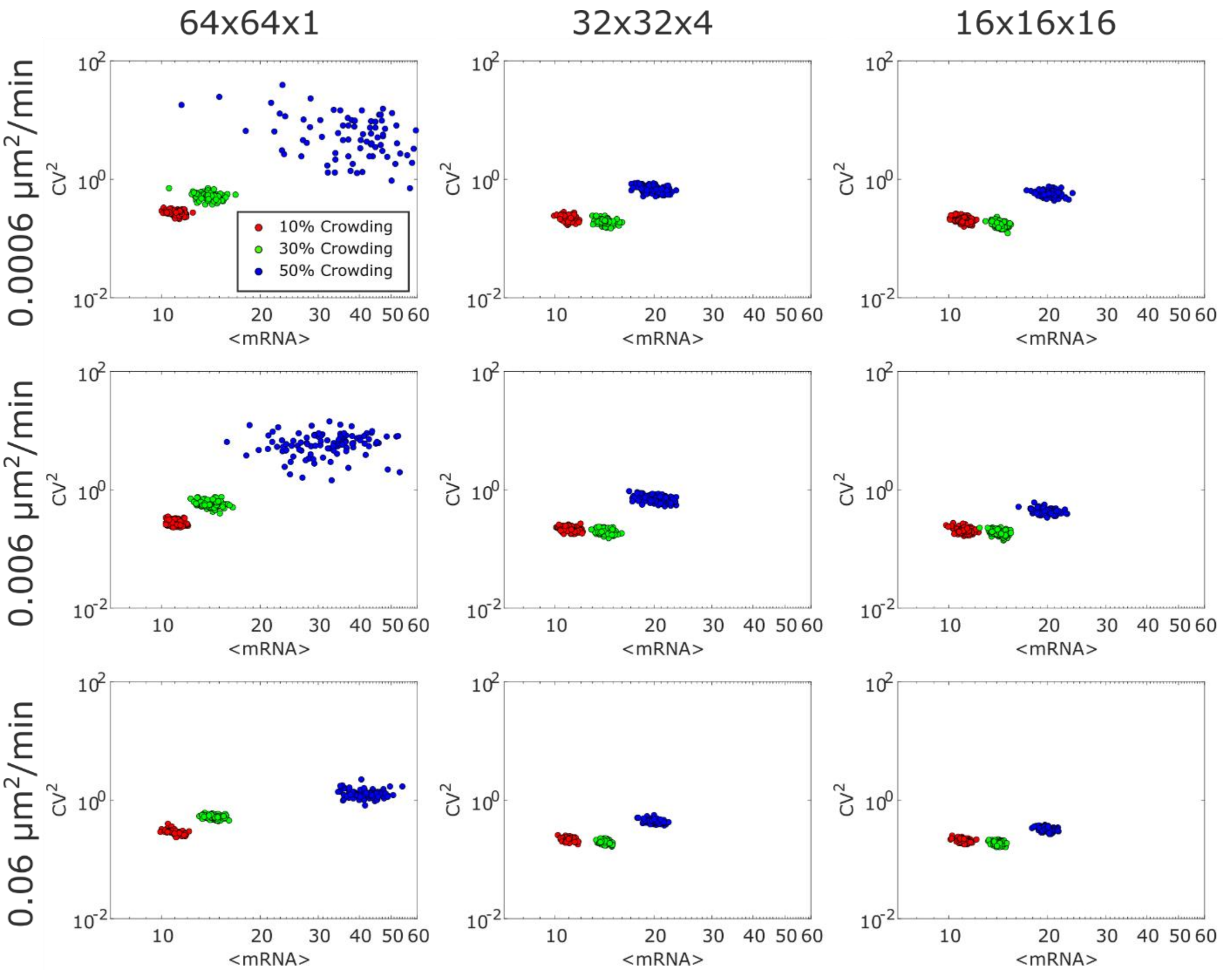

Figure A.30 – Compiled cases for burst method. Burst method two-state model results are presented for all crowder diffusion rates (rows) and all confinement geometries (columns). Colors represent the various crowding fractions tested. Each data point represents the CV^2 and mRNA abundance value for a single simulation trajectory.

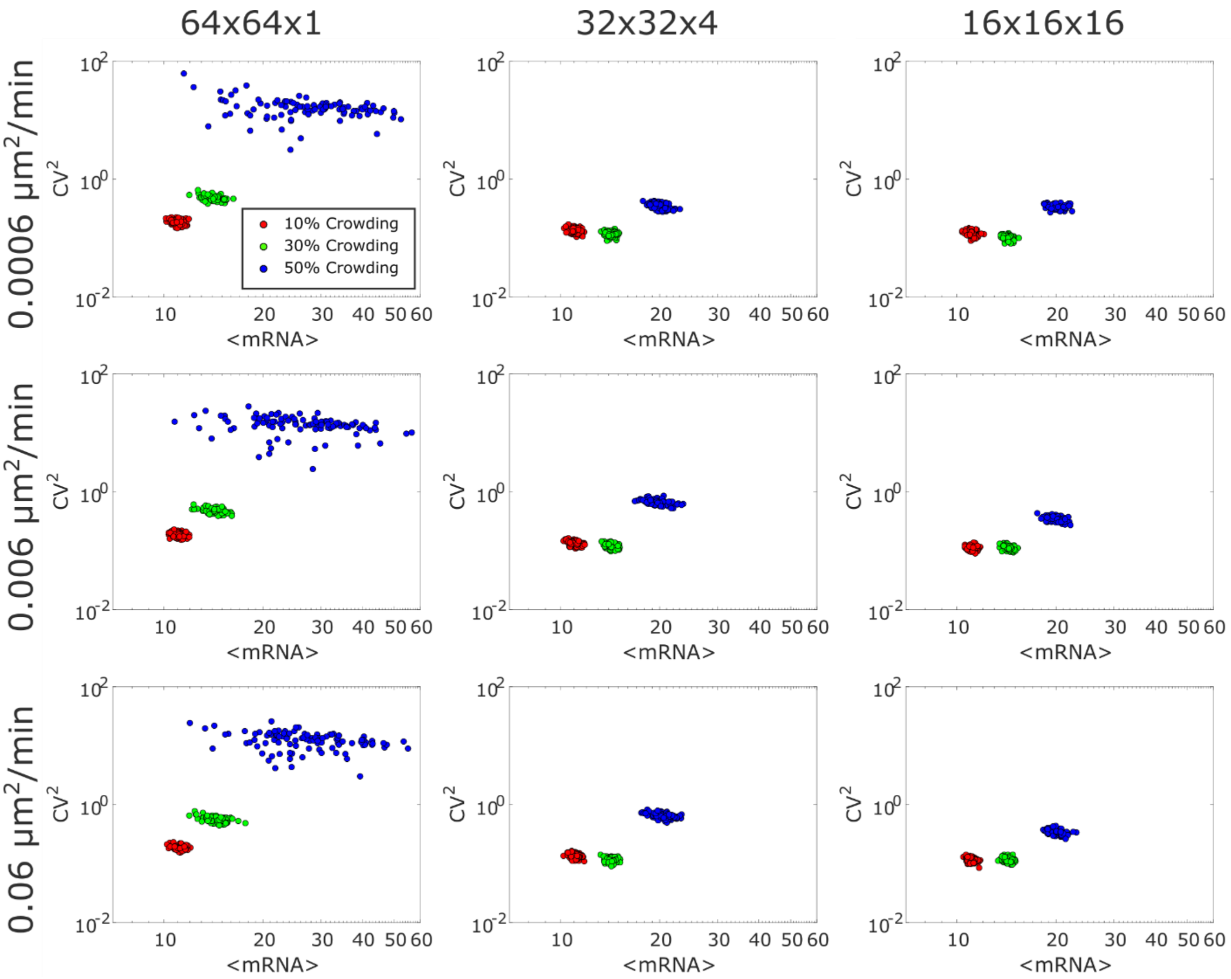

Figure A.31 – Compiled cases for 3-state model. Three-state model results are presented for all crowder diffusion rates (rows) and all confinement geometries (columns). Colors represent the various crowding fractions tested. Each data point represents the CV^2 and mRNA abundance value for a single simulation trajectory.

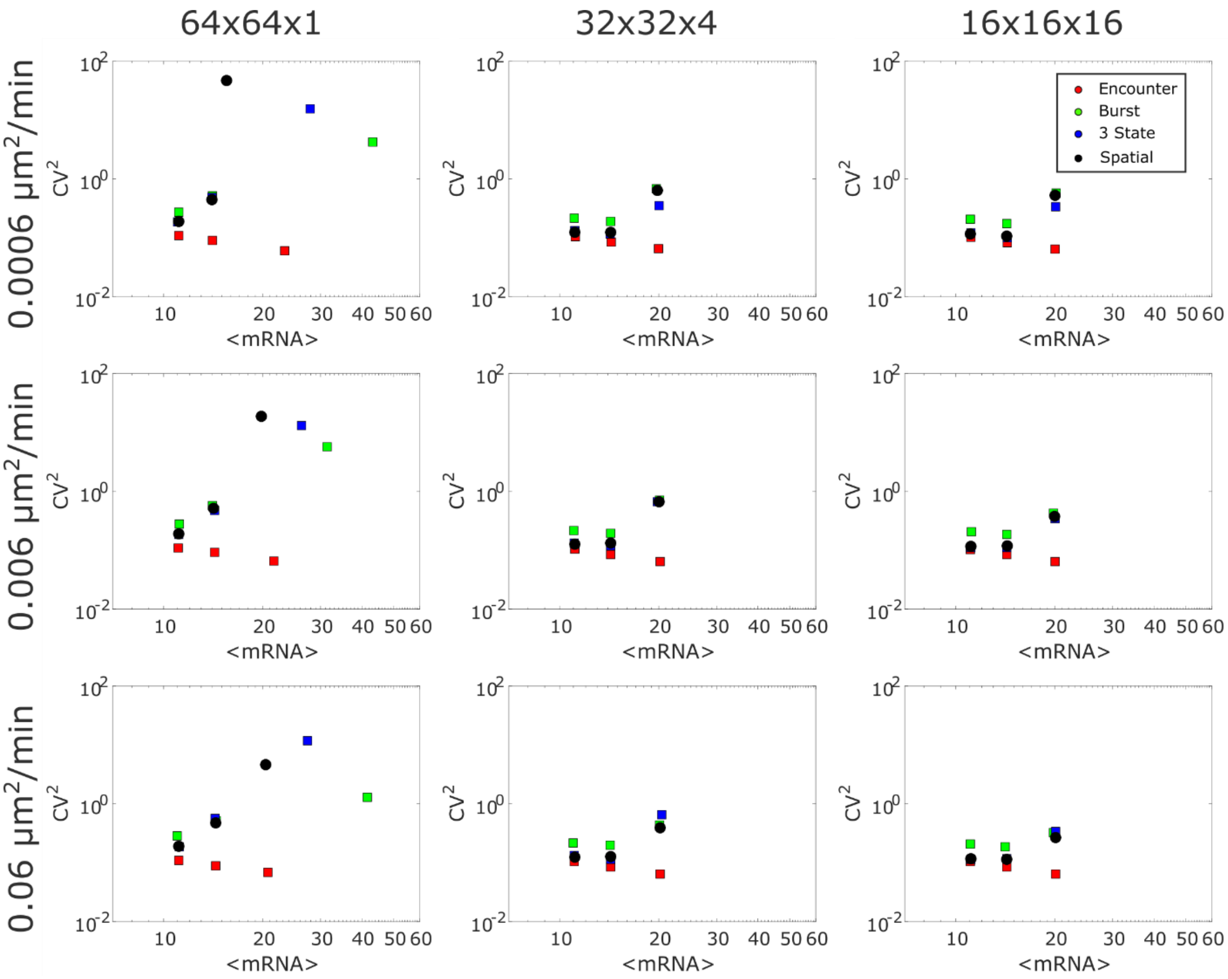

Figure A.32 – Complied means for all tested cases. All model results are presented for all crowder diffusion rates (rows) and all confinement geometries (columns). Colors represent the various modeling methods. Each data point represents the geometric mean of CV^2 and mRNA abundance values for a given model.

Figure A.33 – Two- and three- state model parameterizations compared against the cubic spatial model results. All comparisons were done at the 16x16x16 lattice geometric confinement over a range of crowding fractions (0% - 50%). Results from spatial simulations are shown in grey. (A) Results from the two-state model with rate constants generated from encounter times from each spatially resolved crowding case. This method does not capture the increase in CV2 at high crowding fraction seen in the spatial model. (B) Results from the two-state model with rate constants generated from burst equations calculated using CV2 and abundance from the spatial model. This method captures the increase in burst size, but introduces an error in the calculated CV2. (C) Results from a three-state model with rate constants generated from encounter data and systematic variation of the free parameter. The three-state model captures the bulk behavior of the spatial model, although it does not generate outliers like those seen in the distribution of rare events at high crowding fraction.

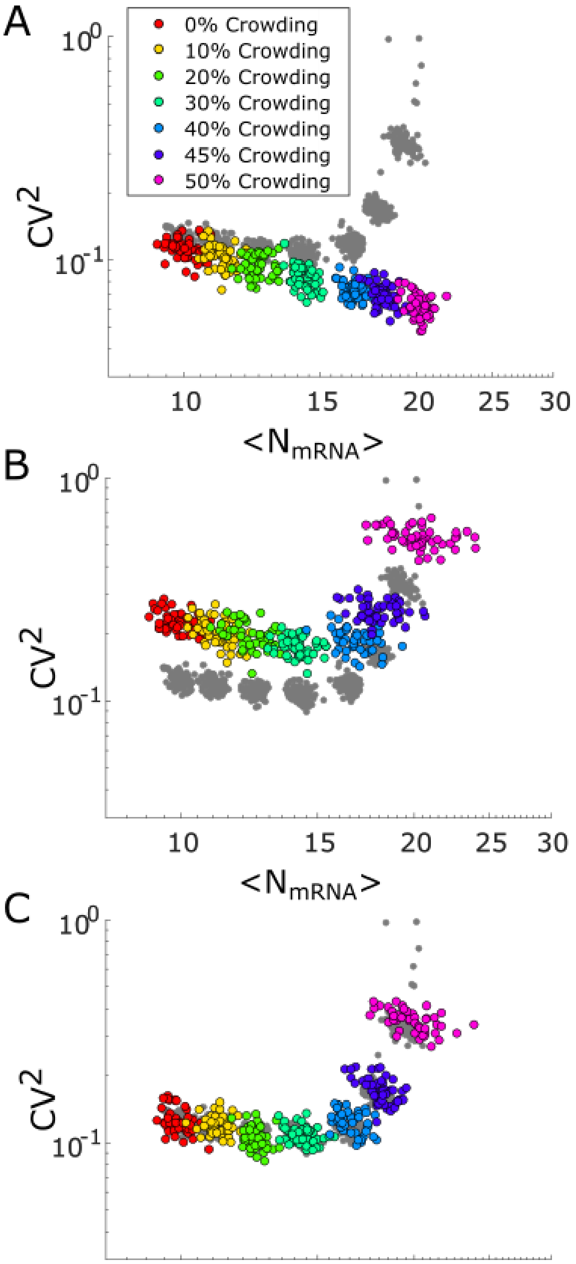

Figure A.33 continued

Figure A.34 – Two- and three- state model parameterizations compared against the confined spatial model results. All comparisons were done at 50% crowding fraction over the range of geometric confinement spaces tested (16x16x16, 32x32x4, 64x64x1). Results from spatial simulations are shown in grey. (A) Results from the two-state model with rate constants generated from encounter times from each spatially resolved crowding case. This method does not capture the increase in CV2 at high crowding fraction seen in the spatial model. (B) Results from the two-state model with rate constants generated from burst equations calculated using CV2 and abundance from the spatial model. This method captures the increase in burst size, but is unable to reach the noise magnitude values at the highest confinement case (64x64x1). (C) Results from a three-state model with rate constants generated from encounter data and systematic variation of the free parameter. The three-state model captures the bulk behavior of the spatial model, although it is unable to capture the full distribution of points at any of the confinement cases.

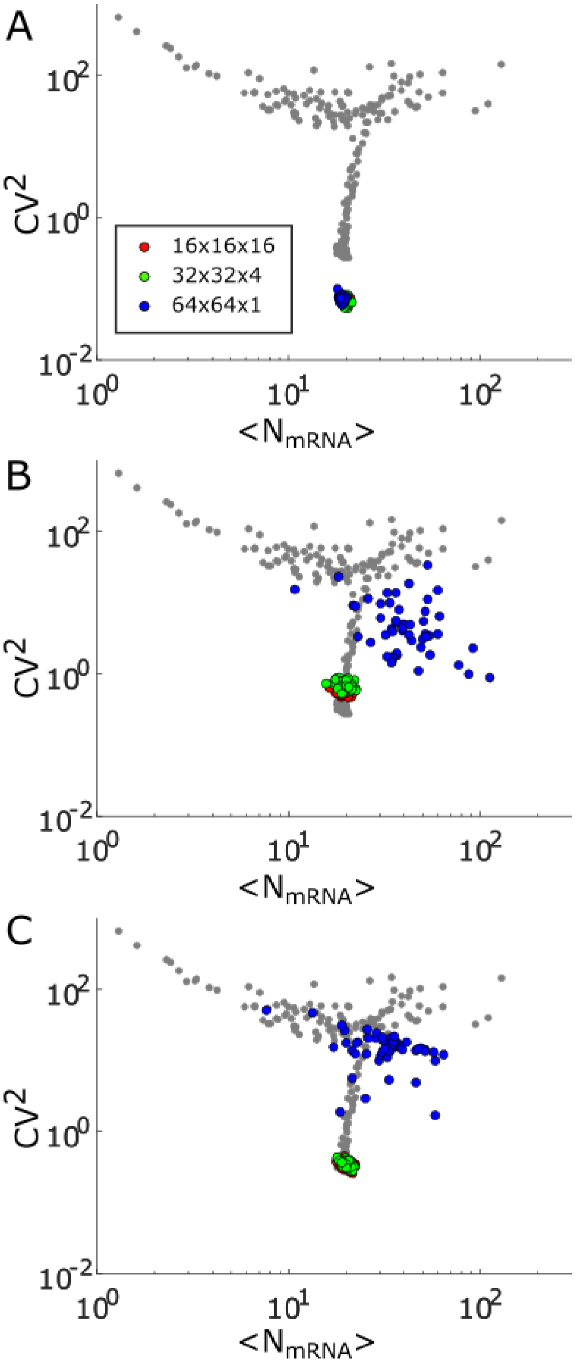

Figure A.34 continued

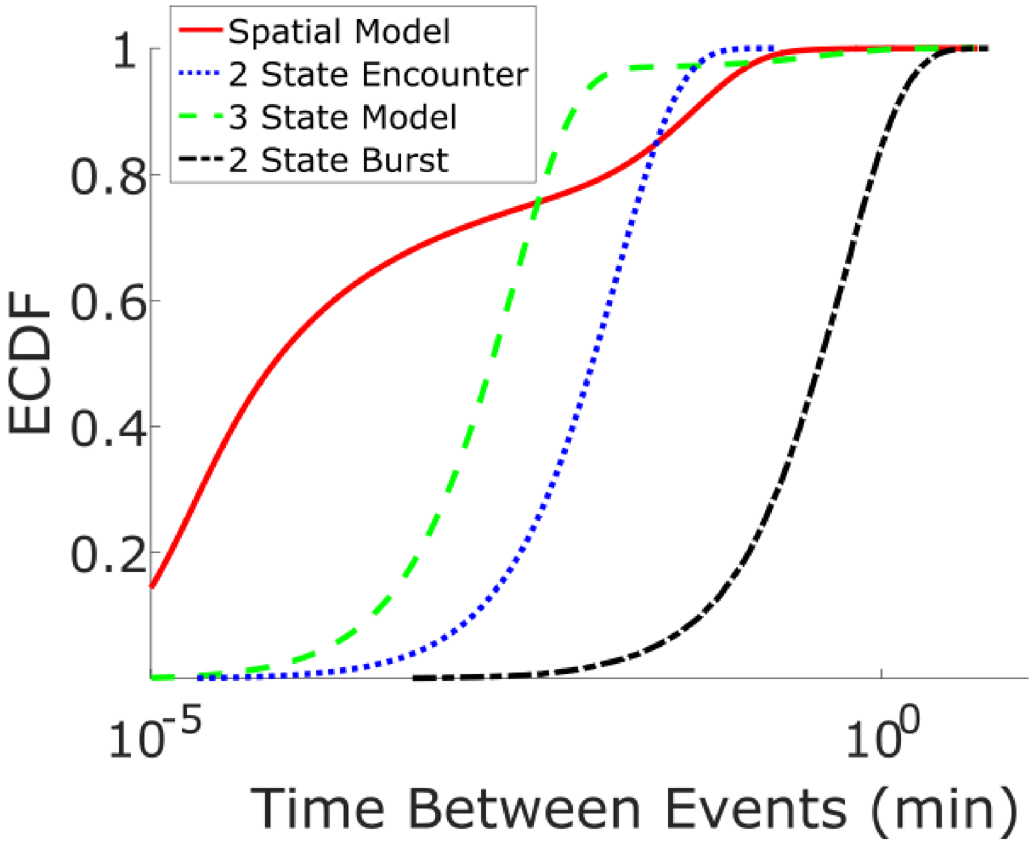

Figure A.35 – Empirical distribution function comparison among the distribution of time between encounters. Empirical distribution function comparison among the distribution of time between encounters for the spatial model at 50% crowding and 16x16x16 spatial confinement, the distribution in the two-state model based on the mean encounter values of the same spatial model, and the distribution in the three-state model based on the modified encounter values of the spatial model. It is clear from the comparison that the distributions are drastically different among the three models, with the distribution of the spatial model occupying much lower values as well as a long tail of high values not apparent in the two-state model distribution. The three-state model captures the behavior at long times, but still underestimates how short encounters occur in the spatial model.

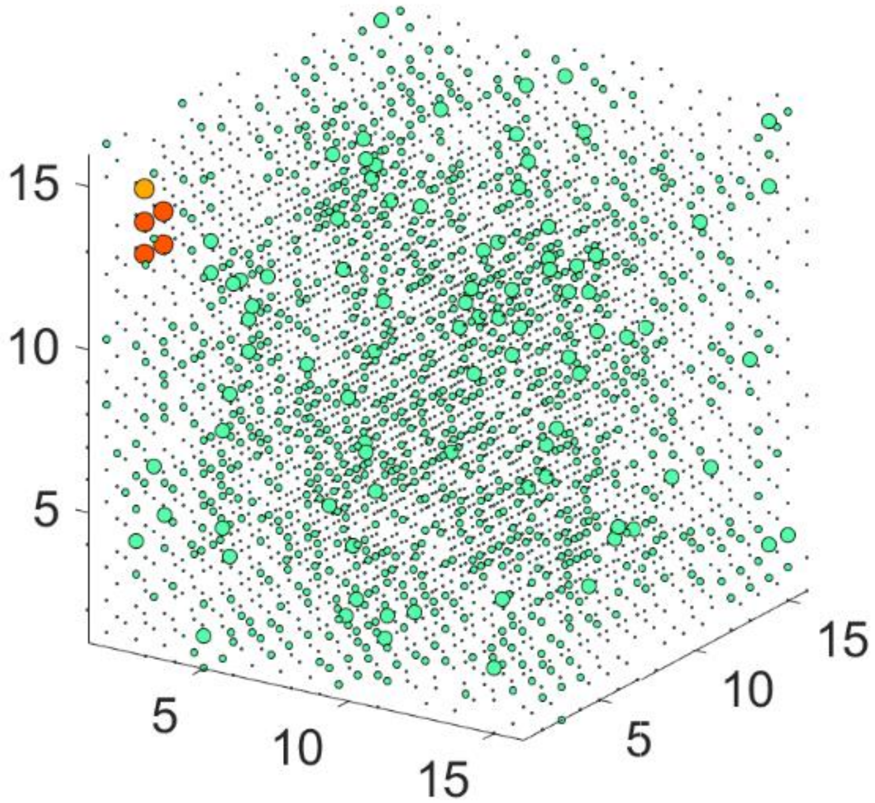

Figure A.36 – 16x16x16 encounter map. Each point represents the number of times the two particles encounter each other in a given lattice site. The size and color of the point indicates the number of times two particles encounter in a given lattice site (larger, redder points indicate more encounters).

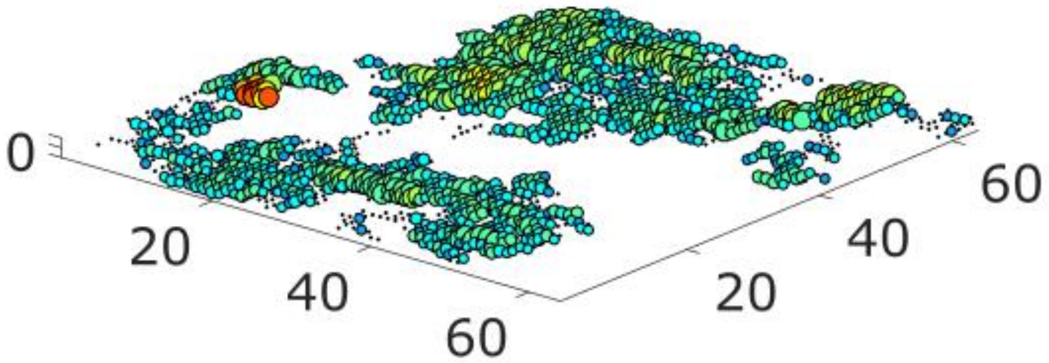

Figure A.37 – 64x64x1 encounter map . Each point represents the number of times the two particles encounter each other in a given lattice site. The size and color of the point indicates the number of times two particles encounter in a given lattice site (larger, redder points indicate more encounters).

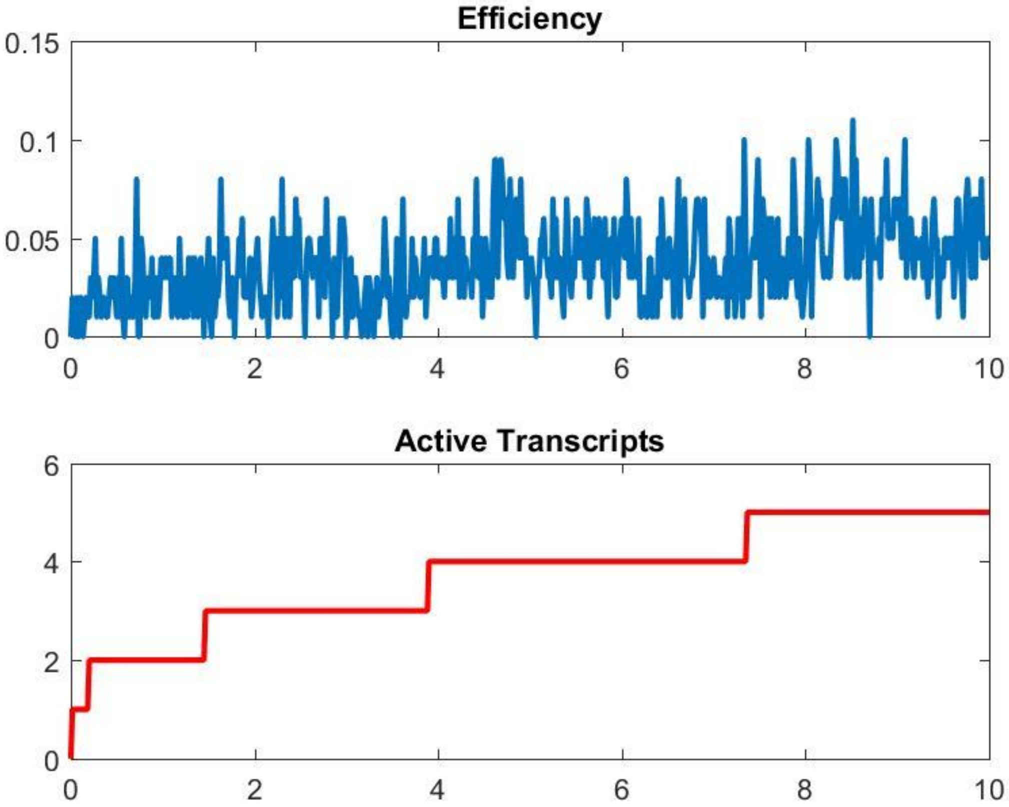

Figure A.38 – Spatial resource efficiency model results without sticky mRNA. The top figure shows the efficiency of ribosomes over time, where efficiency is defined as the number of ribosomes present in a lattice site with an active mRNA transcript, divided by the total number of ribosomes in the system. The bottom figure shows the number of active transcripts in the system over time.

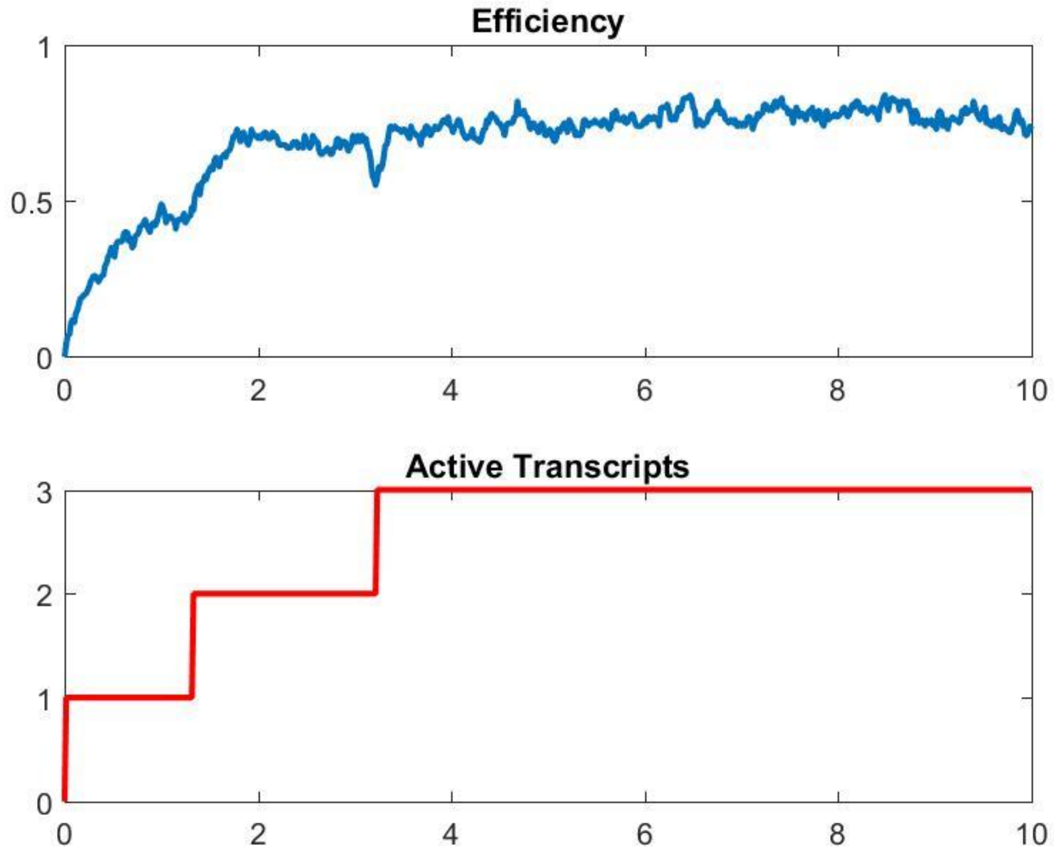

Figure A.39 – Spatial resource efficiency model results with sticky mRNA. The top figure again shows the efficiency of ribosomes over time. The bottom figure shows the number of active transcripts in the system over time. Notably, when mRNA transcripts become sticky, the efficiency of the ribosomes becomes much higher.

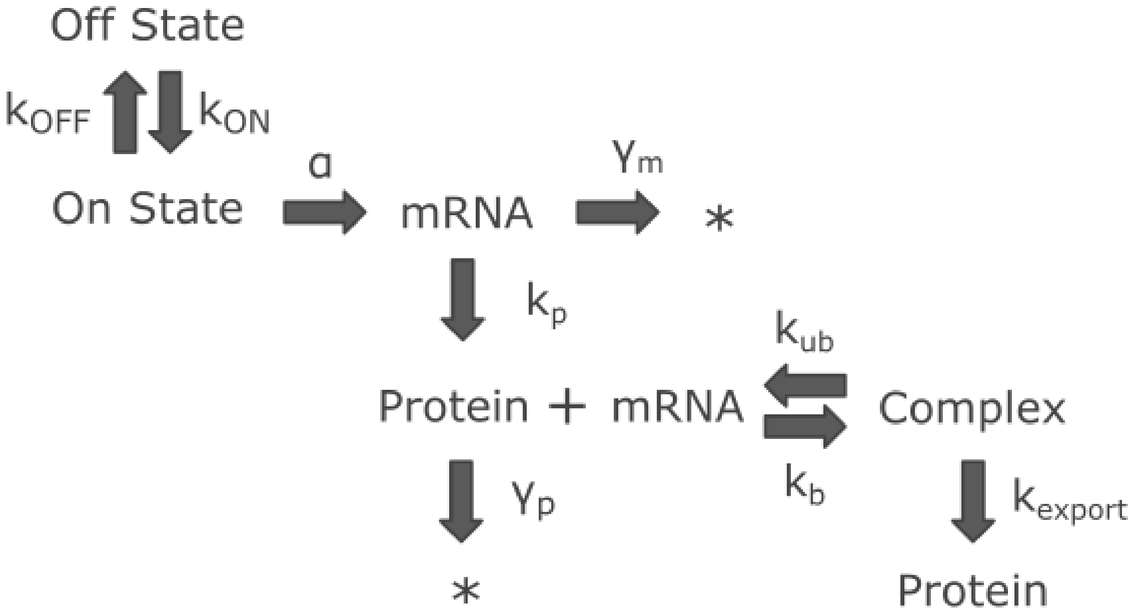

Figure A.40 – Schematic of the simplified HIV model. Here, a single gene is allowed to stochastically burst on and off, producing mRNA in the on state. Protein can bind to mRNA molecules to create a complex, which is exported from the system at rate k_{export}. The act of binding mRNA sequesters it from the system, introducing a negative feedback loop which drives steady state abundance and noise in protein population down.

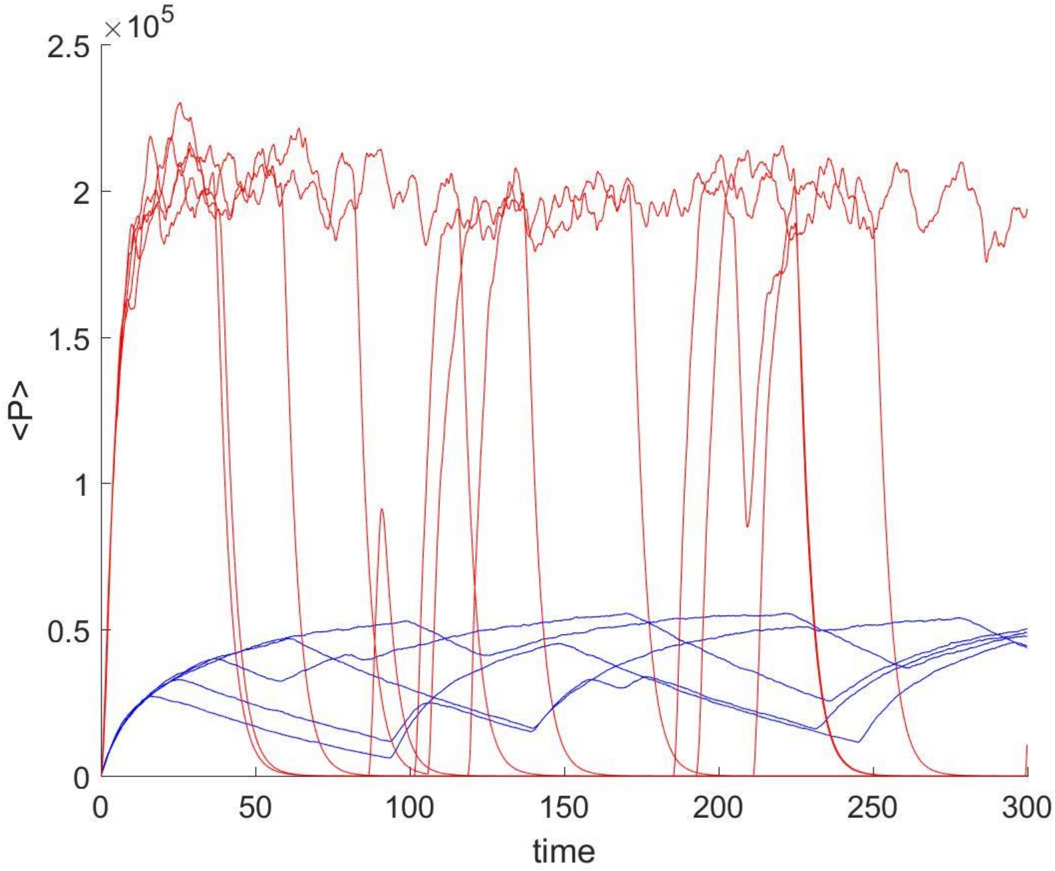

Figure A.41 – Results of the two HIV models, with and without mRNA export. Red traces represent the model without negative feedback or mRNA export. Steady state protein values reach 20,000 and rapidly decay to 0 when the gene stochastically turns off. Blue traces represent the model with negative feedback and mRNA export. Because of complex formation, steady state values do not reach the same high values as in the previous case. However, the reversible formation of the complex reduces the decay rate of the protein population, which maintains a non-zero protein population when the gene turns off.

Vita

Charles Chin was born in October of 1986 in Oak Ridge, Tennessee. He would remain in Oak Ridge his entire childhood, graduating from Oak Ridge High School in 2005. In high school, Charles was an avid archer, winning multiple state and regional championships throughout his career, and was also proficient in the violin, helping his class win regional and national orchestra competitions across the United States. After high school, Charles moved to Baltimore to study at Johns Hopkins University, studying in the competitive field of biomedical engineering. After graduating with a Bachelor's of Science in Biomedical Engineering, Charles returned to Oak Ridge to work at the world renowned Oak Ridge National Lab under Dr. Hsin Wang, which was on the imaging and analysis of induction heated spot welds for structural weaknesses and non-destructive testing. Soon after, Charles began work on a Master's degree in Biomedical Engineering, working under Dr. Zhili Zhang on the study and imaging of plasmon resonance of core-shell nanoparticles. Having graduated from that program in 2011, Charles was then inducted into the inaugural class of the Bredesen Center, an interdisciplinary program that collected graduate students from energy fields from climate studies to nuclear engineering, and placed them all into a program that fostered collaboration. This joint University of Tennessee (UTK) and Oak Ridge National Laboratory (ORNL) venture allowed Charles to continue to study at UTK while researching models of noise in gene expression systems at the lab. In his spare time, Charles enjoys music, art, and programming for his hobbies.