

Agent-Based Kinetics:
A Nonspatial Stochastic Method for Simulating
the Dynamics of Heterogeneous Populations

ALEX PLAKANTONAKIS

Correspondence should be addressed to alex.plaka@alumni.princeton.edu.

Version: 2024.1

Copyright © Alex Plakantonakis 2019-2024

This document is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivatives 4.0 International License
CC BY-NC-ND 4.0

Unless noted otherwise, all of the computer code cited
in this work is licensed under the terms of the
GNU General Public License (GPL), version 3.

Short Abstract

Molecular, cellular, or organismal populations are heterogeneous in that they are made up of distinct individuals, each endowed with a unique set of properties that determine how they will respond to various kinds of stimuli and conditions. While the mechanisms that give rise to population heterogeneity are in many cases known, its effects on the dynamics of coupled processes and population time trajectories are not well understood. We have addressed this issue by developing an algorithmic framework for taking into account different forms of heterogeneity, thus allowing for the systematic study of individual variation on the emergent properties of the collective. We show on theoretical grounds and through simulations that the stochastic method we have developed, termed *agent-based kinetics*, correctly predicts the time evolution of populations and can be flexibly implemented for use in a variety of fields, such as biochemistry, cell biology, ecology, and epidemiology. This monograph surveys examples from the aforementioned fields to showcase the use of the algorithm, compares its predictions to those of the deterministic and other stochastic approaches, and examines the roles of noise in low copy number populations and the presence of heterogeneity on a system's dynamics. Special emphasis is given to the simulation of modular collections of coupled processes, or motifs, that exhibit distinct qualitative behaviors, such as continuous and discontinuous (hysteretic) switching, and oscillatory activity. We also present how our method can be used to perform sensitivity analysis with respect to quantitative measures of population heterogeneity.

Abstract

The presence of heterogeneity is a salient feature of molecular, cellular, and organismal populations, and while significant progress has been made on understanding the mechanisms that give rise to it, its effects on a system’s dynamics are not well understood. The consequences of heterogeneity are especially important to consider when all members of a species’ population have intrinsically different propensities toward interacting with each other or are distinguishable with respect to one or more quantifiable traits, as can be the case for populations studied in a variety of fields (e.g., cell biology, ecology, social sciences). Accordingly, there is a need for flexible modeling approaches and simulation tools that can account for an arbitrary level of population heterogeneity, allow for the simulation of systems with dynamic parameter landscapes, and predict population time trajectories *and* the uncertainty associated with them.

We address these issues by developing a method we term *agent-based kinetics* (ABK), where we treat populations of species as consisting of a discrete number of agents/individuals to generate ensembles of stochastic time trajectories based on a predefined set of processes involving intra- and interspecies interactions. ABK is a kinetic Monte-Carlo algorithm, uses the Markov property of chemical reactions for simulating general processes (although it is easily adaptable to non-Markovian processes), and, in its simplest implementation, assumes a well-mixed system, thereby resulting in a nonspatial simulation. Given the computational cost of assessing if any of all possible transition events occur for each agent within a given time interval (rejection sampling), the algorithm is most efficient and best suited for simulating the dynamics of small (hundreds or thousands of agents) to low copy number populations.

We validate the algorithm’s effectiveness by simulating the time evolution of homogeneous populations (where all agents of a given species have the same kinetic rate constant associated with them) and comparing our results to predictions obtained through either solutions to ordinary differential equations (ODEs), the stochastic simulation algorithm (SSA, or Gillespie’s algorithm), or, in the simplest cases, the chemical master equation (CME). We begin by simulating simple irreversible processes (0th-3rd order, and then generalize to potential higher order processes in non-chemical systems) and use them as a basis for extending our analyses to reversible and birth-death processes, the Michaelis-Menten reaction scheme, and processes under regulatory control. We also discuss efficient uses of the algorithm in modeling concurrent processes featuring same- or mixed-order kinetics.

We continue the method’s validation by applying it to simulating signal-response (SR) motifs consisting of species with homogeneous populations. These motifs are modular collections of basic processes where the signal S represents an external forcing (which can be continuous or discrete in nature; the latter case can represent the population of a molecular messenger species, for instance) and interspecies connectivity can lead to nonlinear effects on the population size

of the response element R . Such nonlinearity is a hallmark of the emergence of complex behavior at the populations level. Specifically, we study linear, hyperbolic, and sigmoidal (or reversible switch) graded response motifs; discontinuous switches (of the reversible or irreversible type) exhibiting hysteresis; motifs producing an adaptive or homeostatic response; and oscillator systems (such as multi-component negative feedback loops). We use the 2-component negative feedback loop motif as an example to illustrate how the algorithm can be easily altered if a modeler is interested in assessing the effect of an explicit time delay in a specific (non-Markovian) process within a motif, and we compare the simulated time evolution to the numerical solution of a delay differential equation (DDE). Furthermore, we apply the ABK methodology to foundational models in population biology (Lotka-Volterra and Volterra models of predator-prey interactions), epidemiology (Kermack-McKendrick model of infectivity), and the study of oscillatory behavior in chemical systems (Brusselator). In the case of the Brusselator, our simulation correctly predicts relaxation oscillations and demonstrates that the algorithm can be used in systems exhibiting strong separation of timescales.

Throughout our investigations, our simulations yield average population trajectories that generally agree with deterministic expectations. However, they also predict various phenomena that the deterministic approach cannot and that have been theoretically predicted to occur and observed in real populations. Examples include the role of stochasticity in switching between a system's stable states (e.g., noise-induced switch state reversals in bistable switch motifs), and a behavior we term *persistent high-amplitude fluctuations* in systems where noise causes a stochastic trajectory to repeatedly take long excursions around a fixed point when damped oscillations are deterministically predicted.

We then proceed to show how the algorithm can be adjusted to simulate the time evolution of well-mixed populations participating in basic processes (1st order, 2nd order, and birth-death processes) that exhibit *compositional* and/or *temporal heterogeneity/diversity* with respect to the microscopic kinetic rate constants associated with specific agents. We theoretically and numerically compare the results of our analyses to those obtained from the simulation of homogeneous populations, thus highlighting the effects of heterogeneity on qualitative features and various statistical properties of the computed trajectories. We then generalize by showing how the method can be adapted to a graph-theoretic representation of agent interactions within populations.

We continue our investigations by revisiting several motifs and present effects of different forms of heterogeneity on their dynamics. Specifically, we show through different cases how population diversity can affect the sigmoidal response motif. We then present the hysteretic switch resulting from mutual activation with only one species exhibiting compositional heterogeneity in the Michaelis constant of an enzymatic reaction. Our simulations show that statistically significant deviations from the behavior of homogeneous populations can result from such a simple example of a heterogeneous population structure. We also revisit the case of the 2-component negative feedback loop motif, where we introduce heterogeneity in the agent-specific time delay value of the feedback

reaction and present its effects on the presence of oscillatory activity. Finally, we describe how sensitivity analysis can be performed using the ABK method to assess the effect of metrics quantifying the degree of heterogeneity in a population on a species' average time trajectory. We demonstrate how this analysis can be accomplished using the example of the mutual activation switch.

Taken together, the multitude of systems we have successfully simulated highlight the adaptability of the ABK methodology to different representations of a dynamical system, thus making it a suitable computational test bed for understanding the effects of the topology of interspecies connectivity, stochasticity, and heterogeneity on population dynamics.

Contents

1	Introduction	1
1.1	Layout of this manuscript	8
1.2	Methods: Computing environment and code availability	9
2	Agent-Based Kinetics for Homogeneous Populations: Basic Processes	11
2.1	Microscopic Rates and Rate Constants	12
2.2	1 st Order Processes	15
2.2.1	Transition Probability from the Integrated Rate Law	15
2.2.2	Transition Probability from the Differential Rate Law	17
2.2.3	Transition Probability of a Bernoulli Process	18
2.2.4	Relationship between P_{int} (or P_{ber}) and P_{dif}	20
2.2.5	From Individuals to Populations	21
2.2.6	The ABK Algorithm and its Implementation	24
2.3	2 nd Order Processes	34
2.3.1	$2\mathbf{A} \rightarrow \mathbf{X}$	34
2.3.2	$\mathbf{A} + \mathbf{B} \rightarrow \mathbf{X}$	47
2.4	3 rd Order Processes	55
2.4.1	$3\mathbf{A} \rightarrow \mathbf{X}$	55
2.4.2	$2\mathbf{A} + \mathbf{B} \rightarrow \mathbf{X}$	57
2.4.3	$\mathbf{A} + \mathbf{B} + \mathbf{C} \rightarrow \mathbf{X}$	58
2.5	0 th Order Processes	59
2.6	Generalization of the ABK Algorithm	62
2.6.1	Transition Probability Expressions	62
2.6.2	General ABK Algorithm Implementation	64
2.7	Concurrent Processes and Mixed-Order Kinetics	65
2.7.1	General Case of Concurrent Processes	67
2.8	Birth-Death Process	70
2.9	Processes under Regulatory Control	73
2.10	Reversible Processes	76
2.11	The Michaelis-Menten Reaction Scheme	78
2.11.1	Full Treatment	78
2.11.2	Abbreviated Treatment	79

3	Signal-Response Motifs of Homogeneous Populations, and Other Models	83
3.1	Simple Graded Response Systems	85
3.1.1	Linear Response Motif	85
3.1.2	Hyperbolic Response Motif	87
3.2	Continuous Switches	91
3.2.1	Multi-Step Ultrasensitivity	92
3.2.2	Ultrasensitivity from a Cyclic Enzymatic Process	96
3.3	Discontinuous Switches	102
3.3.1	The Mutual Inhibition Switch	103
3.3.2	The Irreversible Mutual Inhibition Switch	112
3.3.3	Hysteretic Switch from Mutual Activation	122
3.4	Adaptation and Homeostasis	127
3.4.1	Perfect Adaptive Response Motif	127
3.4.2	Imperfect Adaptive Response Motif	132
3.4.3	Homeostasis	136
3.5	Oscillator Systems	139
3.5.1	Negative Feedback 2-Component Motif	139
3.5.2	Negative Feedback 2-Component Motif with Explicit Time Delay	147
3.5.3	Negative Feedback 3-Component Motif	150
3.5.4	Activator-Inhibitor Motif	159
3.5.5	Substrate-Depletion Motif	163
3.6	Other Models	166
3.6.1	The Lotka-Volterra Predator-Prey Model	166
3.6.2	The Kermack-McKendrick Model	176
3.6.3	The Brusselator	180
4	Agent-Based Kinetics for Heterogeneous Populations of Well-Mixed Systems	188
4.1	Introduction: Types of Heterogeneity	189
4.2	Measures of Population Heterogeneity	189
4.3	0 th Order Processes	192
4.4	1 st Order Processes	193
4.4.1	Predicting the Time Trajectory of a Heterogeneous Population	194
4.4.2	ABK Implementation	196
4.5	2 nd Order Processes	201
4.5.1	ABK Implementation of a Heterologous 2 nd Order Process	204
4.5.2	Adjustments to the Algorithm for the 2 nd Order Homologous Process	208
4.6	Generalization to Graphs	209
4.7	Higher Order Processes	211
4.8	Heterogeneity in a Birth-Death Process	212
4.9	Heterogeneity in the Sigmoidal Switch Motif	216
4.10	Heterogeneity in the Hysteretic Switch from Mutual Activation	222

4.11	Negative Feedback 2-Component Motif with Heterogeneity in the Explicit Time Delay	228
4.12	Sensitivity Analysis with Respect to Measures of Population Het- erogeneity	235
4.12.1	Case Study: Sensitivity Analysis of the Mutual Activation Switch Motif	237
	Closing Remarks	243
A	The Chemical Master Equation for Selected Processes	245
A.1	1 st Order Process	246
A.1.1	Relationship to Deterministic Prediction	247
A.1.2	Determining the Variance from the CME	248
A.1.3	CME Solution	249
A.2	Birth-Death and 0 th Order Processes	250
A.2.1	Steady-State Solution	251
A.2.2	Relationship to Deterministic Prediction	252
A.2.3	Determining the Variance from the CME	253
A.3	Homologous 2 nd Order Process	257
A.3.1	Relationship to Deterministic Prediction	257
A.4	Heterologous 2 nd Order Process	260
B	The Stochastic Simulation Algorithm (SSA)	261
B.1	Algorithm Implementation	263
B.2	Relation to CME	265
C	A Primer on Stability Analysis	267
C.1	1-Dimensional Linear Systems	267
C.2	1-Dimensional Nonlinear Systems	269
C.3	2-Dimensional Linear Systems	270
C.4	2-Dimensional Nonlinear Systems	272
D	Selected Matlab Code	274
	Bibliography	288

List of Figures

2.1	Comparison of P_{int} (or P_{ber}) and P_{dif} as a function of Δt	21
2.2	Sample runs for $A \rightarrow X$ with populations of different sizes.	26
2.3	Time course of $A \rightarrow X$ with populations of different initial size.	27
2.4	Standard deviation of 1 st order process average trajectory.	29
2.5	Distribution of 1 st order transition events.	30
2.6	Effect of time interval Δt on accuracy of P_{int} (or P_{ber}) and P_{dif} expressions.	32
2.7	Comparison of transition probabilities for a 2 nd order process.	41
2.8	Choice of transition probability expression.	42
2.9	Time course of $2A \rightarrow X$ process for populations of different initial size.	45
2.10	Distribution of $2A \rightarrow X$ transition events.	46
2.11	Time course of $A + B \rightarrow X$ process for populations of different initial size.	52
2.12	Distribution of $A + B \rightarrow X$ transition events.	54
2.13	Time course of 0 th order process.	61
2.14	Examples of concurrent processes simulated using the ABK algorithm.	69
2.15	Time course of a birth-death process: $\emptyset \rightarrow A \rightarrow \emptyset$	72
2.16	Sample forms of the RSF.	75
2.17	Time course of 1 st order process with feedback regulation.	76
2.18	Time course of a reversible process.	77
2.19	Time course of enzymatic process using the full treatment of the Michaelis-Menten reaction scheme.	79
2.20	Time course of enzymatic process using the abbreviated treatment of the Michaelis-Menten reaction scheme.	81
2.21	Full vs. abbreviated treatments of Michaelis-Menten kinetics.	82
3.1	Linear response motif.	85
3.2	Linear response motif: SR curve.	87
3.3	Hyperbolic response motif.	87
3.4	Hyperbolic response motif: rate-balance plot and SR curve.	89
3.5	Abbreviated 2-step sigmoidal response motif.	92
3.6	Abbreviated 2-step sigmoidal response motif: SR curve.	94
3.7	2- and 3-step sigmoidal response motif.	94

3.8	Multi-step sigmoidal response motif: SR curve comparison. . . .	95
3.9	Sigmoidal response motif.	96
3.10	Sigmoidal response motif: rate-balance plot and SR curve.	99
3.11	Sigmoidal response motif: $N_{S_{50}}$ and maximum slope of SR curve.	101
3.12	Mutual inhibition switch motif.	104
3.13	Mutual inhibition switch motif: N_R^* bifurcation diagram with respect to n_H and phase plane.	106
3.14	Mutual inhibition switch motif: stochastic time trajectories in state space (case 1).	108
3.15	Mutual inhibition switch motif: stochastic time trajectories in state space (case 2).	109
3.16	Mutual inhibition switch motif: noise-induced switch state reversal.	111
3.17	Irreversible (one-way) mutual inhibition switch motif.	112
3.18	Irreversible mutual inhibition switch motif: phase plane and SR curve.	115
3.19	Irreversible mutual inhibition switch motif: stochastic SR curve.	117
3.20	Irreversible mutual inhibition switch motif: noise-induced switch state reversal.	118
3.21	Reversible mutual inhibition switch motif: stochastic SR curve.	121
3.22	Alternate mutual inhibition switch motif.	122
3.23	Mutual activation switch motif.	123
3.24	Mutual activation switch motif: phase plane.	124
3.25	Mutual activation switch motif: stochastic SR curve.	125
3.26	Mutual activation switch motif: noise-induced switch state reversal.	126
3.27	Perfect adaptive response motif.	127
3.28	Perfect adaptive response motif: rate-balance plot.	129
3.30	Perfect adaptive response: alternate motif.	129
3.29	Perfect adaptive response motif: time course.	130
3.31	Perfect adaptive response: alternate motif time course.	131
3.32	Imperfect adaptive response motif.	132
3.33	Imperfect adaptive response motif: predicted SR curve.	134
3.34	Imperfect adaptive response motif: time course.	135
3.35	Homeostatic response motif.	136
3.36	Homeostatic response motif: rate-balance plot and SR curve.	138
3.37	Negative feedback 2-component motif.	140
3.38	Negative feedback 2-component motif: time course (case 1).	142
3.39	Negative feedback 2-component motif: time course (case 2).	145
3.40	Negative feedback 2-component motif: time course (case 3).	146
3.41	Negative feedback 2-component motif with delay: time course.	149
3.42	Negative feedback 3-component motif.	150
3.43	Negative feedback 3-component motif: average time course, $N_S = 10$	153
3.44	Negative feedback 3-component motif: sample time course, $N_S = 10$	154
3.45	Negative feedback 3-component motif: average time course, $N_S = 30$	155

3.46	Negative feedback 3-component motif: sample time course, $N_S = 30$.	156
3.47	Negative feedback 3-component motif: average time course, $N_S = 50$.	157
3.48	Negative feedback 3-component motif: sample time course, $N_S = 50$.	158
3.49	Activator-inhibitor motif.	159
3.50	Activator-inhibitor motif: average and sample time trajectories.	161
3.51	Activator-inhibitor motif: coefficient of variation.	162
3.52	Substrate-depletion motif.	163
3.53	Substrate-depletion motif: sample and average time trajectories.	165
3.54	Lotka-Volterra (LV) model: phase plane and deterministic time trajectory.	169
3.55	Stochastic LV model: sample and average time trajectories.	171
3.56	Volterra model: phase plane.	174
3.57	Stochastic simulations of the Volterra model.	175
3.58	Simulation of the Kermack-McKendrick model.	178
3.59	Kermack-McKendrick $S-I$ state space.	179
3.60	The Brusselator reaction scheme.	180
3.61	Brusselator phase plane.	182
3.62	Sample oscillatory time trajectories of the Brusselator.	184
3.63	Sample time trajectory of the deterministically-damped Brusselator.	185
3.64	The modified Brusselator reaction scheme.	186
3.65	Sample oscillatory time trajectory of the modified Brusselator.	187
4.1	Population heterogeneity matrix for a 1 st order process.	193
4.2	Time trajectory and transition event distribution of a 1 st order process with two subspecies.	199
4.3	Standard deviation and coefficient of variation of a 1 st order process with two subspecies.	200
4.4	Population heterogeneity matrix for a 2 nd order process.	201
4.5	ABK simulation of a 2 nd order process with two subinteraction groups.	206
4.6	Standard deviation and coefficient of variation of a 2 nd order process with two subinteractions.	207
4.7	Graph representation of the 2 nd order process $A + B \rightarrow X$.	210
4.8	The effect of heterogeneity on a birth-death process.	214
4.9	Heterogeneity in the abbreviated 2-step sigmoidal response motif: SR curves.	219
4.10	Heterogeneity in the abbreviated 2-step sigmoidal response motif: time-dependence of population-based average rate constants.	220
4.11	Heterogeneity in the mutual activation switch motif: sample time trajectories.	224
4.12	Heterogeneity in the mutual activation switch motif: SR curve.	225
4.13	Heterogeneity in the mutual activation switch motif: differences in $\langle N_R^* \rangle$, $SDev\{\langle N_R^* \rangle\}$, and switch state.	226

4.14	Heterogeneity in the negative feedback 2-component motif with delay: sample time trajectories.	231
4.15	Heterogeneity in the negative feedback 2-component motif with delay: average time trajectories.	232
4.16	Heterogeneity in the negative feedback 2-component motif with delay: trajectory smoothing and peak detection.	233
4.17	Heterogeneity in the negative feedback 2-component motif with delay: distribution of peak heights and time intervals.	234
4.18	Heterogeneity in the mutual activation switch motif: sample time trajectories using CRN.	238
4.19	Heterogeneity in the mutual activation switch motif: mean time trajectories and sensitivity analysis.	240
4.20	Heterogeneity in the mutual activation switch motif: differential fraction of simulation runs in ON state.	241
4.21	Heterogeneity in the mutual activation switch motif: standard deviation of time trajectories and sensitivity analysis.	242

List of Tables

2.1	Transition probability expressions for process $A \rightarrow X$	20
2.2	Transition probability expressions for process $2A \rightarrow X$	39
2.3	Transition probability expressions for process $A + B \rightarrow X$	50
2.4	Transition probability expressions for process $3A \rightarrow X$	57
2.5	Transition probability expressions for process $2A + B \rightarrow X$	58
2.6	Transition probability expressions for process $A + B + C \rightarrow X$	58
3.1	Canonical vs. agent-based DEs in the Brusselator.	182
4.1	Heterogeneity in the mutual activation switch motif: p -values for statistically significant differences in $\langle N_R^* \rangle$	227

Chapter 1

Introduction

The simulation of natural phenomena requires careful consideration of the interplay between the level of detail that can serve as input and desired output for a model, and the efficient use of available computational resources. Generally, after the application of basic principles to an idealized system of interest yields a satisfactory understanding of the underlying processes, a more realistic model with explanatory power and that can make predictions commensurate with the level of complexity observed in the real world is ultimately desired.¹ To fulfill these goals, the challenge of progressing from a simple to a more realistic implementation of a model is to increase the level of detail while keeping the problem computationally tractable.

The most widely used approach for modeling the time evolution of populations in a variety of fields is differential calculus, and for good reason: calculus is an accessible and powerful mathematical language for expressing how quantities change with respect to each other and time, and various established techniques can be used to analyze the dynamics of systems exhibiting nonlinearity (as is most often the case in real systems). However, in the case of nonspatial simulations, this deterministic approach through the use of ordinary differential equations (ODEs) has several obvious limitations:

1. Continuity is assumed for the variable representing the population size, whereas real population sizes are *discrete* quantities. The validity of this assumption is arguably unimportant to consider for large populations. However, a mathematical representation of a system's processes must account for discretization in the regime of small population sizes, where

¹Of course, explanation and prediction are just two of the many reasons to build models. Others include (adapted from Epstein, 2008): models inform which experiments to conduct and guide the ensuing data collection, uncover a system's core dynamics, suggest dynamical analogies, discover new questions to ask, expose prevailing theories as incompatible with available observations or data, demonstrate the trade-offs between different approaches to solving a particular problem, highlight the connection between apparent simplicity and complexity of systems, train other modelers, and educate the general public. All of the above are relevant to the models we will present in this work.

chance events can have a dramatic effect on future population trajectories. This regime is present in many contexts within the biological and social sciences. Examples include a) there can be just a handful of intracellular copies of a particular molecule (e.g., chromosomal DNA is found in only one or two copies); b) the appearance of a heritable mutation in an individual and its initial intergenerational spread through a population; and c) the early stages of information dissemination (e.g., misinformation and rumor spreading) within social networks.

2. The deterministic approach gives insight into the expected population time trajectories but cannot provide a measure of the *uncertainty* in the future value of a population size. Alternatively stated, ODEs do not capture the variability that results from the randomness inherent to the occurrence of real processes. For instance, chemical reactions are fundamentally stochastic processes, therefore a probabilistic description is needed to model them. This is especially important in the case of low copy number populations (Halling, 1989), where fluctuations of any magnitude become more likely (first considered by Donnan, 1927) and may dominate the system's dynamics.
3. Even if one accepts the presence of a discrete population size, the use of ODEs implicitly assumes that populations are homogeneous; that is, there is *no variability* in the propensity of the members of a population to react or interact with each other as part of a process that affects the dynamics of the system. The assumption of homogeneity has the advantage of making the simulation of systems analytically and/or computationally tractable. However, real populations can be characterized by a statistical distribution of the value used to quantify a given trait, therefore they can be considered *heterogeneous* with respect to it. Here are some examples where consideration of the causes and/or effects of heterogeneity is important for understanding biological processes:
 - a) In the realm of biochemistry, many protein species in higher organisms are expressed as several closely related isoforms, each with its own characteristic catalytic rate or affinity toward various binding partners.
 - b) In cell biology, several research groups (see Taniguchi et al., 2010; Albayrak et al., 2016, among others) have documented a lack of correlation between the gene-specific number of mRNA and protein molecules in prokaryotic and eukaryotic cells, measured with single-molecule sensitivity in the same cell (for a review on this topic, see Liu et al., 2016). As a result, a *clonal* population of cells can exhibit a high degree of phenotypic heterogeneity at the level of individual cells, leading to both desirable and harmful effects depending on the biological context (Raj and van Oudenaarden, 2008).
 - c) Stochastically generated genetic variation leading to cellular heterogeneity occurs in multicellular organisms: mutations introduced randomly in the genome during the early stages of development and subsequently propagated through cell lineages lead to *somatic mosaicism*, the effects

of which are only beginning to be appreciated (Lim et al., 2017; Lodato et al., 2015; Priest et al., 2016; Ju et al., 2017). Mosaicism can also result from chromosomal nondisjunction events leading to, for example, *gynandromorphism* (or sex-chromosome mosaicism), where an organism exhibits both male and female characteristics.²

d) Numerous studies have reported on the diversity in cancer cell populations due to genetic or non-genetic factors and intratumor spatial heterogeneity (Vargas-Garcia et al. 2017; Nguyen et al. 2016; Swanton et al. 2015; Sharma et al. 2010; Gonzalez-Garcia et al. 2002; for a review, see Marusyk and Polyak, 2010; (for a review in the context of signal transduction networks, see Kolch et al., 2015)).

e) In epidemiology, real populations can have a skewed distribution of individual infectiousness (Lloyd-Smith et al., 2005) and “superspreading” events or individuals have been documented in disease outbreaks (for an example, see Lau et al., 2017). Accordingly, the presence or absence of such highly infectious individuals in the early stages of a disease outbreak can have a significant effect on the ensuing transmission dynamics and epidemic spread.

The first two of the above problems — namely, the discrete nature of populations and stochasticity — can be dealt with by the chemical master equation (CME) formalism developed more than 50 years ago (McQuarrie, 1967). The CME describes the time-dependent probability distribution of the discrete population size of the species involved in a set of processes. However, the CME can be solved analytically only in the simplest of cases (see Appendix A for a presentation of some of these cases). The need for a numerical algorithm that takes into account the inherent randomness of processes was addressed by Daniel Gillespie (1977), who proposed an elegant method for simulating such systems.³ The *stochastic simulation algorithm* (SSA) is a rejection-free algorithm that monitors the time between productive reaction events and updates the population sizes accordingly (see Appendix B for a description of the algorithm). The rationale behind ignoring unproductive or nonreactive collisions has to do with the assumption that the population is well-mixed or uniformly distributed within the reaction vessel’s volume. This can be achieved by continuous stirring (in the case of a chemical system), or by requiring that the number of molecules far exceeds the number of transition events. The latter assertion guarantees that the large number of nonreactive collisions between molecules randomizes their positions in space and thermal energies (and therefore velocities) according to the Maxwell-Boltzmann distribution. Thus, for a well-mixed system, the SSA ignores the exact positions and velocities of all molecules and simply focuses on the actual reaction events. This is a significant simplification

²As an interesting aside, rare gynandromorphs have provided the opportunity to perform elegant experiments on the interplay between genetic sex and hormonal influences on sexual differentiation (e.g., Agate et al., 2003).

³The algorithm was in fact first proposed by Kendall (1950) for the specific case of simulating birth-death processes.

that makes the simulation computationally efficient. In addition, Gillespie’s algorithm has been widely adopted in software packages (e.g., COPASI, Hoops et al., 2006; StochSS, Drawert et al., 2016). Finally, we note that the SSA is an exact numerical realization of the CME, since they are both derived (without approximation) from the basic premise of the stochastic formulation of chemical kinetics that any transition event has a finite probability of occurring (Gillespie, 1977; 2007).

Despite the SSA’s wide use, it cannot be easily adapted to account for nuanced situations involving intrapopulation heterogeneity. Consider for a moment the population of a *species* (molecular, organismal, etc.) that can be described as consisting of a small number of *subspecies*, each with its own *subpopulation* and set of parameters that affect the probability of an individual member of the subspecies undergoing a given process. In general, it may be the case that not all subspecies of a given species participate in the same set of processes (e.g., isoforms of a given protein species can have specific amino acid sequences resulting in different binding partners, subcellular localizations, etc.), thus necessitating a way of computing each subspecies’ time trajectory separately from the others. One can argue that the SSA can still be used by defining additional processes that treat subspecies as separate species. However, this approach is cumbersome for increasing levels of heterogeneity (especially at the limit of the number of subspecies approaching the species’ population size) and would substantially increase the computational time that the SSA requires for simulating the population time trajectories. It is therefore important to have an alternative simulation tool for probing population dynamics that can be easily configured to take into account an arbitrary level of population heterogeneity.

A reasonable approach is to explicitly model each individual’s fate and contribution to a population’s evolution. This approach is referred to as *agent-* or *individual-based modeling* (ABM or IBM).⁴ Several disciplines contributed to ABM’s development, most notably computer science and the study of complex systems. There is no universal consensus on how to precisely define the idea of this modular entity, the agent. The minimal requirements are that an agent is *autonomous* in that it can “operate” without external direction and can be represented as a software object (Macal and North, 2010). Typically, agents are treated as only being able to sense their *local* environment; this supports the aforementioned idea of agent autonomy because the system is thereby decentralized with respect to its behavior. It is common in applications of ABM for an agent to obey a set of rules that dictate its behavior based on its interactions with other agents and its environment. An early and famous example of how basic rules can give rise to complex behavior is Conway’s Game of Life (Gardner, 1970), which is essentially a spatial (2-dimensional) agent-based model implemented in the form of cellular automata. ABM has been used over the last two decades in a diverse array of fields, such as economics, sociology, archeology, epidemiology, and microbiology (see Macal and North, 2010, and references therein

⁴The term *particle-based modeling* is also sometimes found in the literature.

for examples of ABM applications; for a review of ABM in modeling microbial communities, see Hellweger and Bucci, 2009; Goroehowski, 2016, reviewed ABM models of cellular populations in synthetic biology; an example of modeling synthetic microbial biofilms can be found in Rudge et al., 2012). Interestingly, some studies have highlighted how different the predictions of the ABM approach can be from those of the deterministic one (e.g., Hellweger and Kianirad, 2007).

Despite the increased computational demands associated with tracking each agent in a population, this modeling paradigm offers the important advantage of modularity, thus allowing for explicit accounting of population heterogeneity. These properties also make agent-based models more comprehensible and intuitive to understand: *the trajectory of a population is a superposition of the fates of individual agents*. Furthermore, the agent-based approach allows for the study of how the elementary interactions between the components of a system give rise to the emergence of complex patterns and behaviors at the population level. This bottom-up approach is useful for testing our understanding of the basic processes that drive the dynamics of a system (as is of interest in the field of systems biology/ecology), as well as for the design of particular arrangements of species interactions that result in a desired behavior (i.e., synthetic biology).

Modeling heterogeneous populations has been explored by a handful of authors, especially in the context of ecology and evolutionary biology. Ecosystems are complex adaptive systems (Levin, 1998) and stochastic modeling approaches have been applied to ecological problems (Lecca et al., 2013a). It is clear that evolution cannot take place in the absence of heterogeneity, since selective pressures would have no trait variations in individuals to act on (Wagner and Altenberg, 1996). Given this perspective, a realistic representation of biological populations necessitates the inclusion of the prospect of heterogeneity. Van Nes and Scheffer (2005) have examined the effect of simple patterns of *spatial* heterogeneity due to varying environmental stresses as a system transitions to an alternate stable state as a result of a catastrophic bifurcation. Fredrick et al. (2013) performed a non-stochastic ABM simulation to show which mechanisms contribute to variability in the intracellular level of phosphorus in phytoplankton cells. Other researchers have modeled cell population dynamics by taking into account the heterogeneity that results from stochasticity in cell division and asymmetric partitioning of the intracellular volume and contents of the mother cell (Mantzaris, 2006; Charlebois et al., 2011; Roy and Klumpp, 2018).

In this work, we address the aforementioned limitations of the deterministic approach and account for population heterogeneity by establishing an algorithmic framework for implementing the nonspatial stochastic simulation of populations consisting of distinct agents. We treat agents in the simplest sense possible: they are individuals that probabilistically move in a system's state space based on a kinetic view of the involved processes. Specifically, we consider each agent participating in a process as being characterized by a single parameter: the kinetic rate constant.⁵ As we will discuss in chapter 2, a rate

⁵Processes subject to stimulatory or inhibitory regulation and enzymatic reactions require additional parameters, as will be discussed in Chapter 2.

constant value is a proxy for the propensity of a molecule/agent to react and can be used to calculate the probability of an agent-specific transition event. Furthermore, we will show over the course of the next few chapters that the ABK algorithm accurately predicts the time evolution of a species population where each of its members has a rate constant value associated with it.

It is important to comment on how such a value can be specified for each agent, given that kinetic constants are measured at the level of a population. Agents of the same species with intrinsically different propensities can have their rate constant values determined from measurements on a homogeneous population of this type of agent. A common example of this in biochemistry is the characterization of the kinetic properties of protein isoforms by performing *in vitro* experiments on homogeneous populations of the respective isoforms. However, obtaining such data can be laborious, the extent of heterogeneity within a population may not be known ahead of time, or such experiments may not be possible (as in the case of organismal populations). Despite these limitations, we believe that having a computational framework for assessing the effect of different statistical distributions of the agent-specific rate constant values (or other relevant parameters) on a system’s dynamics can be useful in understanding the potentially complex interplay between population heterogeneity and stochasticity.

We will concern ourselves with systems exhibiting two types of heterogeneity: *compositional* and/or *temporal*.⁶ The first type simply refers to species populations of kinetically distinct agents. On the other hand, temporal heterogeneity deals with populations with time-dependent kinetic rate constants, thus allowing the simulation of systems with dynamic parameter landscapes. We term our approach *agent-based kinetics* (ABK). In brief, the probability of each agent participating in a process is evaluated for a given interval of simulated time, and this numerical value is then compared to a random number drawn from the uniform distribution in the unit interval to assess the binary-valued state⁷ of this agent (a *two-state* agent) in the next time step. This repeated sampling of the occurrence of a process, performed over all agents and for all non-overlapping time intervals making up the total simulated time, constitutes a Monte Carlo simulation of the population time trajectory. ABK provides a *mesoscopic* representation of chemical systems by assuming they are well-mixed, thus allowing the consideration of every possible interaction or collision between molecules within a given time interval. That is, the method accounts for the kinetics of individual molecules/particles, but does not explicitly consider their physicochemical properties, diffusion, or other spatial considerations. In the specific case of biochemical reactions occurring in the intracellular milieu, the

⁶We will primarily deal with compositional heterogeneity in the case studies we will describe in the coming chapters. However, we do present how to account for time-dependent kinetic parameters when introducing a general framework for treating heterogeneous processes in Chapter 4.

⁷We use a binary value to keep track of the state each agent is in: 1 represents an agent who is ‘alive’ (or that has not been consumed), and 0 for a ‘dead’ agent (one that has been consumed, degraded, or transformed into an agent of a different species).

assumption of well-mixed molecular populations ignores the crowding effects or presence of structural barriers that may restrict the diffusion of molecules *in vivo* (Turner et al., 2004). In recognition of this limitation, we classify the simplest implementation of our method as *nonspatial*.

Finally, we briefly juxtapose our current work with related work done by other researchers. Vestergaard and Génois (2015) have adapted the SSA for simulating time-varying networks of individuals (temporal heterogeneity). Cai et al. (2013) applied the SSA to an epidemic-spreading model with heterogeneity in the infectivity of individuals (compositional heterogeneity). The ABK algorithm is simpler to set up and programmatically implement than the above adaptations of the SSA while also being capable of simultaneously accounting for both temporal and compositional heterogeneity. Furthermore, non-SSA modeling approaches that treat molecules as individual entities have been attempted before by Morton-Firth and Bray (1998)⁸ and Pogson et al. (2006). The work we present here differs from those efforts in several key ways:

1. The ABK transition probability expressions are algebraically different from those used in the aforementioned studies. Furthermore, we present no fewer than three different ways of deriving such suitable expressions for a given elementary process comprised of species with homogeneous populations.
2. The algorithmic structure of our method is different than that of the above studies. For instance, Morton-Firth and Bray consider in each time step of the simulation the possible interaction between a molecule and a “pseudo-molecule” for the case of a 1st order reaction. In contrast, our algorithm does not require consideration of any nonexistent or unintuitive pseudo-interactions.
3. We introduce a general framework for determining transition probabilities for a process of any nonnegative integral order (not just 1st or 2nd).
4. We demonstrate the algorithm’s effectiveness on basic processes and a large number of collections of such processes (termed *motifs*) that exhibit complex behavior such as adaptation, switching, oscillations, etc. To our knowledge, the systematic study of these motifs in the context of agent-based kinetic modeling and population heterogeneity has not been documented before.
5. We built the ABK algorithm specifically for the purpose of modeling populations that are heterogeneous with respect to the kinetic properties of each agent in a population. Heterogeneity was not addressed in either of the mentioned previous studies.
6. Our work allows us to probe the dynamics of populations that can be represented as graphs or networks. As such, our methodology has a wide-

⁸Further developed into the software package StochSim (Le Novère and Shimizu, 2001).

ranging scope and can be easily adapted for use in fields outside of the molecular sciences.

1.1 Layout of this manuscript

We first present the application of the ABK method to homogeneous populations before adapting it to handle cases of heterogeneity. This serves three primary purposes:

1. It allows us to establish the algorithm’s theoretical underpinnings and validate its predictions against those of the deterministic and stochastic formulations.
2. It highlights the advantages of the stochastic approach to modeling system dynamics over the deterministic one. It also sets the stage for showcasing the advantages of the ABK method over other stochastic methods that cannot be easily adapted to simulate heterogeneous populations.
3. It reflects the workflow we recommend for applying the ABK algorithm to specific problems. The code implementation of the algorithm leads to predictions that are compared to the results of other methods for homogeneous populations (ODE, SSA, CME) to ensure the code is bug-free. The code can then be adjusted to account for the presence of heterogeneity in the specific process(es) a researcher is interested in simulating.

In Chapter 2, we establish and validate the ABK methodology as applied to homogeneous populations. This allows us to compare our stochastic simulations to deterministic predictions (and, in some cases, to the solutions of the CME formalism) and explore the latter’s limitations. We begin by investigating irreversible processes of 1st and 2nd order, and then generalize to higher orders. We also address the simulation of 0th order processes. We then present how the algorithm can be used efficiently when simulating concurrent processes (i.e., when a species participates in two or more reaction paths/channels), and how to calculate transition probabilities in processes under regulatory control. The above few cases form a basis set for modeling more elaborate schemes, such as a birth-death process, reversible reactions, and the Michaelis-Menten reaction scheme.

Chapter 3 deals with the application of the ABK algorithm to collections of coupled processes, called *motifs*, whose study reveals that the arrangement of intra- and interspecies agent interactions can produce emergent properties or complex behavior at the level of a population. We again focus on the simulation of homogeneous populations to compare our results to those of the deterministic approach, which we obtain by employing methods from the analysis of nonlinear dynamical systems (see Appendix C for an introduction to Stability Analysis). In each motif, we assign a species as a “Response element” (R , the output) whose transient population level is dependent on the presence of a “Signal element” (S , the stimulus or input) and the precise connectivity defined by the motif’s

processes. As such, each motif can be thought of as a modular functional unit that acts on an input to produce an output. We characterize the sensitivity of the response in each motif with a Signal-Response (SR) curve showing the dependence of R on S . We present a series of such motifs in what can be considered as tutorial form and in each case we test that our algorithm works as expected and makes reasonable predictions.

All of the above simulations were performed on homogeneous populations. In Chapter 4, we apply ABK to study the *effect* of heterogeneity on the emergent properties and dynamics of the collective. We address this challenge by laying the foundations for how the algorithm can be adapted to simulate the time evolution of heterogeneous populations, where each agent has a specific kinetic constant value associated with each of the processes it participates in. We initially consider well-mixed systems exhibiting two types of heterogeneity: *compositional* and/or *temporal*. A species consists of *subspecies*, the population of each being made up of agents characterized by the same kinetic parameter value. We find it helpful to consider a variation of this concept for 2nd and higher order processes: intra- and interspecies interactions can be grouped into kinetically distinct *subinteractions*. On the other hand, temporal heterogeneity deals with populations where the time-dependence of the kinetic rate constants needs to be accounted for. We show how the ABK method can accommodate both types of heterogeneity and present simulations of small population sizes undergoing irreversible 1st, 2nd order, and birth-death processes. We then generalize this discussion to the application of the algorithm to graphs, where nodes represent agents and the rate constants act as the weights of the connecting edges. Therefore, the simulation of an arbitrary system (including non-biological ones) that can be abstracted as a network of interacting nodes is possible. We conclude Chapter 4 by revisiting several of the motifs presented in Chapter 3 to examine the effect of heterogeneity on their dynamics. We also present the algorithmic basis and an example of performing sensitivity analysis with respect to metrics of population heterogeneity.

1.2 Methods: Computing environment and code availability

We use the Matlab programming language (MathWorks, Natick, MA) for performing all of the computations and coding of the examples presented in this work, and its integrated development environment (IDE) for testing, debugging, and running the code. Matlab is a high-level programming language, meaning that it uses a level of abstraction to hide details of code execution at the machine level (e.g., it manages physical memory allocation and addressing), thus allowing the programmer to focus on algorithm development and testing of easily readable code. Moreover, Matlab provides a particularly wide range of built-in functions useful for both numerical and symbolic calculations, making it ideal for designing, testing, and deploying scientific and engineering applications. We

have written and tested all of our code in Matlab version R2017a.⁹ Simulations were run on a PC with an Intel® 5th generation dual-core processor operating at the frequency range of 1.6 – 2.7 GHz. We present selected examples of the code used in our simulations in Appendix D, and all of our code is publicly available at <https://github.com/alexplaka/ABK> (licensed under the terms of the GNU General Public License, version 3.0). We also point the reader to each section’s relevant code with hyperlinks to the appropriate repository folders in marginal notes throughout the text.

⁹The majority of the code can also be run using GNU Octave, a free alternative to Matlab (<https://www.gnu.org/software/octave>). However, we have not specifically tested our code in that environment and, as such, cannot provide any guarantees on its performance.

Chapter 2

Agent-Based Kinetics for Homogeneous Populations: Basic Processes

The method of *agent-based kinetics* (ABK) can be applied to study the time evolution of any system where there is information on the kinetic parameters describing the interactions between members of a population. We present the following arguments from the perspective of chemical kinetics, but they can be applied without loss of generality to any population composed of discrete individual entities (termed *agents*), be they members of a human population, cells, or protein molecules. Accordingly, the terms *reaction* and *process*, or *molecule* and *agent*, are used interchangeably in this chapter.

ABK depends on calculating the probability of a process happening within a specified time interval for *each* molecule in a population. Therefore, the *macroscopic* kinetic rate constants, which may depend on concentrations or population densities (and therefore the volume/area in which the population of molecules/agents is in), must be converted to their *microscopic* counterparts with units of number of molecules/agents undergoing a reaction/transition per second (sec^{-1}). We begin by formalizing this conversion of rates and rate constants, and then derive the probability of an agent's transition from one species (or class/type of agents) to another using different approaches. We will first consider elementary processes and then use them to build and characterize simple combinations of them, such as reversible and birth-death processes, as well as the Michaelis-Menten reaction scheme for enzymatic catalysis. Throughout this chapter we will examine the effectiveness of the ABK method by comparing it to deterministic (ODE) and stochastic (CME or SSA) predictions for the time evolution of homogeneous populations, which consist of agents with identical kinetic properties (we will extend these ideas to heterogeneous populations in Chapter 4).

2.1 Microscopic Rates and Rate Constants

Consider an elementary process of unspecified order in which species X is produced. We assume that the molecules are uniformly distributed within a volume V and present in a single phase. Let $[X]$ be the molar concentration of X . For such a process, the *macroscopic* rate r' is given by

$$r' = \frac{d[X]}{dt}. \quad (2.1)$$

Since the rate r' has units M/sec ,¹ multiplying both sides of equation 2.1 by the product of Avogadro's number and the volume ($\mathcal{N}_{Av}V$) results in the following expression, which we define to be the *microscopic* rate r ,

$$r \equiv \frac{dN_X}{dt}, \quad (2.2)$$

where N_X is the number of molecules or agents of X . Note that the units of r are the number of molecules of X produced per second (sec^{-1}). Even though N_X is a discrete quantity and can therefore only have integer values, we shall consider the process-specific expression for r and its consequences as the *deterministic approach* to studying the time evolution of populations.

The relationship between r and r' can also be expressed as follows:

$$r' = \frac{d[X]}{dt} = \frac{d\left(\frac{N_X}{\mathcal{N}_{Av}V}\right)}{dt} = \left(\frac{1}{\mathcal{N}_{Av}V}\right) \frac{dN_X}{dt} = \frac{1}{\mathcal{N}_{Av}V} r.$$

Therefore, the *macro-* and *microscopic* rates are generally related by

$$r' = \frac{1}{\mathcal{N}_{Av}V} r. \quad (2.3)$$

The result in equation 2.3 holds for a process of arbitrary overall order² since we have made no assumptions about how X is produced. It will be useful in that it allows for the easy conversion of *macroscopic* rate constants to their *microscopic* counterpart. We show below how this conversion can be done for processes of a specific order.

0th order rate constant. The rate law for the process $\emptyset \rightarrow X$, where \emptyset represents the (apparent) lack of dependence of the rate law on any reactants,³ is $r' = k'_0$, with k'_0 being the rate constant of the process. Note that in

¹Or, in general, units of concentration per unit of time.

²It is commonplace to refer to the *molecularity* of elementary processes. In this work, we use the term *order* instead, in recognition of the fact that the methods we will develop are general and extend beyond the realm of chemical reactions.

³We will use the symbol \emptyset throughout this work in reaction diagrams to indicate 0th order processes and to describe the result of a molecule's degradation (or, equivalently, an agent's "death").

this discussion we use the convention of the subscript indicating the order of the reaction. Substituting equation 2.3 for r' into the rate law, we get

$$\begin{aligned}\frac{1}{\mathcal{N}_{Av}V} r &= k'_0 \\ r &= k'_0 (\mathcal{N}_{Av}V).\end{aligned}\quad (2.4)$$

Let

$$k_0 = k'_0 (\mathcal{N}_{Av}V) \quad (2.5)$$

be the *microscopic* 0th order rate constant. The *microscopic* rate law then becomes

$$\frac{dN_X}{dt} = r = k_0, \quad (2.6)$$

as one would expect for a 0th order process by correspondence to the macroscopic (molar) rate law. Note that the units of both r and k_0 are sec^{-1} .

1st order rate constant. According to the law of mass action, the molar rate law for the process $A \rightarrow X$ is $r' = k'_1[A]$. Proceeding as before, we get

$$\begin{aligned}\frac{1}{\mathcal{N}_{Av}V} r &= k'_1 \left(\frac{N_A}{\mathcal{N}_{Av}V} \right) = \left(\frac{k'_1}{\mathcal{N}_{Av}V} \right) N_A, \\ r &= k'_1 N_A\end{aligned}\quad (2.7)$$

where N_A is the number of molecules or agents of A . Therefore, for a 1st order process,

$$k_1 = k'_1, \quad (2.8)$$

and the *microscopic* rate law is $r = k_1 N_A$. The above result is not surprising given that the macroscopic 1st order rate constant already has units of sec^{-1} .

2nd order rate constant. A second order process can have the form $2A \rightarrow X$ or $A + B \rightarrow X$. We distinguish these two cases by referring to them as *homologous* and *heterologous* 2nd order processes, respectively. We consider here the latter case, whose molar rate law is $r' = k'_2[A][B]$. Substituting equation 2.3 for r' into the rate law, we get

$$\frac{1}{\mathcal{N}_{Av}V} r = k'_2 \left(\frac{N_A}{\mathcal{N}_{Av}V} \right) \left(\frac{N_B}{\mathcal{N}_{Av}V} \right) = \left(\frac{k'_2}{\mathcal{N}_{Av}V} \right) N_A \left(\frac{N_B}{\mathcal{N}_{Av}V} \right).$$

After multiplying both sides by $\mathcal{N}_{Av}V$, we get

$$r = \left(\frac{k'_2}{\mathcal{N}_{Av}V} \right) N_A N_B. \quad (2.9)$$

Therefore, for 2nd order processes

$$k_2 = \frac{k'_2}{\mathcal{N}_{Av}V}, \quad (2.10)$$

and the *microscopic* rate law is $r = k_2 N_A N_B$. The reader can ascertain the fact that the result is identical for the case where $2A \rightarrow X$. Simple inspection of the units reveals that, indeed, k_2 has units of sec^{-1} (recall that k'_2 has units of $\text{M}^{-1} \text{sec}^{-1}$).

n^{th} order rate constant. An analysis similar to the one described above shows that for a reaction $nA \rightarrow X$, or, more generally, for any reaction of overall order n ,

$$k_n = \frac{k'_n}{(\mathcal{N}_A v V)^{n-1}}, \quad n \geq 0, \quad (2.11)$$

and k_n has units of sec^{-1} .

Equation 2.11 allows us to obtain the microscopic rate constant for any process, although chemical systems are practically limited to 3rd order processes due to the low probability of a simultaneous and effective collision between 3 molecules needed for a reaction occurrence.

The relationship between the microscopic rate constant k_n , obtained as described above, and the stochastic reaction constant c_n (as used by Gillespie in his formulation of the SSA) is that they are the same in all cases except for 2nd and higher order processes involving reactant molecules of the same species (e.g., $A + A \rightarrow X$). In those cases, c_n is greater than k_n by a statistical factor related to the number of distinct pairs of reactant molecules. For a 2nd (or higher) order process with two identical reactant molecules, this factor is $2! = 2$, and $c_2 = 2k_2$. Similarly, for three identical molecules the factor is $3! = 6$ (Gillespie, 1977, p. 2343), etc.

2.2 1st Order Processes

Consider a 1st order elementary process given by



We introduce three ways to evaluate the probability that a molecule of A transitions (is converted) to X within a finite time interval of arbitrary length. The first two extend the deterministic formulation of the rate law for this process to populations of discrete size, while the third treats transition events as Bernoulli trials. We will see that the different approaches yield identical or closely related expressions for the per-agent transition probability.

2.2.1 Transition Probability from the Integrated Rate Law

As shown in the preceding section, the microscopic rate law for such a process is $r = k_1 N_A$. Therefore,

$$-\frac{dN_A}{dt} = k_1 N_A. \quad (2.13)$$

Note that we derived this expression from the macroscopic rate law, which is valid in the thermodynamic limit of infinitely large population sizes. As we show in Appendix A.1, the above rate law is in fact true for the average molecular population size (i.e., $-d\langle N_A \rangle / dt = k_1 \langle N_A \rangle$; see discussion leading to equation A.12) and is a consequence of microphysical considerations that lead to the stochastic formulation of chemical kinetics (the chemical master equation, or CME). Given this relationship, we now take the continuous deterministic formulation (equation 2.13, without the designation for the average population size for notational simplicity) and explore ways to derive probability expressions for the transition $A \rightarrow X$, where we treat populations as consisting of distinct agents.

Separation of variables and integration over an arbitrary time interval $\Delta t = t - t_o$ yields the microscopic integrated rate law for this process

$$\frac{N_A(t)}{N_A(t_o)} = e^{-k_1 \Delta t}, \quad (2.14)$$

where $N_A(t_o)$ and $N_A(t)$ are the number of A agents at the beginning and end of time interval Δt , respectively.⁴

For a *single* agent of A , it is reasonable to consider that the larger the time interval Δt is the more likely its conversion to X becomes (or equivalently, the less likely it is *not* to be converted to X). Moreover, it is our expectation

⁴A familiar result can be obtained by considering the half-life, λ , the elapsed period of time for which the condition $N_A(t)/N_A(t_o) = 1/2$ is true. We then get $k_1 = \ln 2/\lambda$, whose substitution into the integrated rate law provides an alternate form of equation 2.14,

$$\frac{N_A(t)}{N_A(t_o)} = 2^{-\Delta t/\lambda}. \quad (2.15)$$

that the greater the fraction of the original population of A that has remained unreacted (or equivalently, the smaller the fraction that has reacted) the higher the probability of a reaction event *not* occurring in a given time interval Δt . In the limiting case where $N_A(t)/N_A(t_o) = 1$, it is clear that no reaction events have taken place within Δt . This suggests that we are dealing with a slow process and the probability of *no* reaction happening within a time interval of similar scale to Δt will be high. Conversely, we would expect the probability of *no* reaction to be low when the process occurs fast enough so that $N_A(t)/N_A(t_o) \rightarrow 0$ within Δt .

These observations serve as motivation for the following definition of the probability⁵ that a *single* molecule of A will *not* undergo a transition to X within time interval Δt :

$$P_{A \nrightarrow X} \equiv \frac{N_A(t)}{N_A(t_o)} = e^{-k_1 \Delta t}. \quad (2.16)$$

Since having no transition ($A \nrightarrow X$) and a transition ($A \rightarrow X$) are mutually exclusive and collectively exhaustive events,

$$P_{A \rightarrow X} = 1 - P_{A \nrightarrow X} \quad (2.17)$$

$$P_{A \rightarrow X} = 1 - e^{-k_1 \Delta t}. \quad (2.18)$$

We will refer to the transition probability in equation 2.18 as P_{int} to reflect the fact that it was derived using the integrated rate law for a 1st order microscopic process (equation 2.14) and the definition of the probability of *no* reaction occurring in Δt (equation 2.16). In the following section, we will also derive an expression for the transition probability using the differential form of the rate law.

$P_{A \rightarrow X}$ has the properties

$$\lim_{\Delta t \rightarrow 0} P_{A \rightarrow X} = 0 \quad (2.19)$$

$$\lim_{\Delta t \rightarrow \infty} P_{A \rightarrow X} = 1, \quad (2.20)$$

as is desired for a probabilistic description of a transition event with a finite, nonzero, positive rate constant. Also note that $P_{A \rightarrow X}$ does *not* depend on the population size of A , therefore it is constant for a given choice of Δt .

It is useful to remember that the above result holds for any 1st order process and not necessarily one involving a chemical reaction. For instance, consider the biologically common and important case of the dissociation of a ligand (L) bound to its receptor (R),



This process is typically reversible, although here we consider only the dissociation of the RL complex (we will discuss reversible processes in section 2.10).

⁵Sometimes referred to in the literature as the *survival probability*.

In accordance with our previous remarks, the likelihood of the ligand remaining bound to the receptor decreases as the considered time interval increases. Therefore, the per- RL complex probability of dissociation is

$$P_{RL \rightarrow R+L} = 1 - e^{-k_{diss}\Delta t}, \quad (2.22)$$

where k_{diss} is the 1st order dissociation rate constant.

2.2.2 Transition Probability from the Differential Rate Law

We start once again with separation of variables in the microscopic rate law of a 1st order process (equation 2.13)

$$-\frac{dN_A}{dt} = k_1 N_A \Rightarrow -\frac{dN_A}{N_A} = k_1 dt. \quad (2.23)$$

We then discretize the differential quantity dt so that the time interval over which the process is studied is sufficiently small but finite, $\Delta t = t - t_o$, such that the change in the number of agents of A , $\Delta N_A = N_A(t) - N_A(t_o)$, is also finite. Note that as we mentioned in the beginning of the preceding subsection, N_A in the differential rate law represents the average population size of A (see Appendix A.1), therefore ΔN_A need not be integer-valued. This discrete form of the differential rate law is

$$-\frac{\Delta N_A}{N_A(t_o)} \approx k_1 \Delta t. \quad (2.24)$$

Notice that the quantity on the left side of equation 2.24 is positive and has a range in the interval $[0, 1]$. Intuitively, the greater the measurable *fraction* of molecules of A undergoing a reaction in a finite time interval, the greater the likelihood of conversion of A to X . Hence, we define the probability of a *single* molecule of A reacting within time interval Δt as

$$P_{A \rightarrow X} \equiv -\frac{\Delta N_A}{N_A(t_o)} \approx k_1 \Delta t. \quad (2.25)$$

In addition to expression 2.18, this provides an alternate way to evaluate the probability of the occurrence of a reaction, and we will refer to it as P_{dif} to distinguish it from the one obtained using the integrated form of the rate law (P_{int} , equation 2.18). It is evident that the right side of equation 2.25 has no upper limit and, therefore, a judicious choice of Δt must be made. Equation 2.25 suggests an upper limit for the value of Δt is

$$\Delta t_{max} = \frac{1}{k_1}. \quad (2.26)$$

In the following subsections, we provide another proof for calculating per-agent transition probabilities and discuss the relationship between P_{int} and P_{dif} .

2.2.3 Transition Probability of a Bernoulli Process

Extracting transition probabilities from population-wide measurements

An alternate way to determine the probability of reaction for a single molecule of A is to treat the transition as a Bernoulli trial process. This is justified by the fact that within a time interval of arbitrary length there are only two possible outcomes: either a given molecule of A is converted to X , or it is not.

We are interested in determining the per-agent probability that the transition $A \rightarrow X$ occurs within a time interval Δt , which we divide into s subintervals of equal length. Let r be the observed average number of reaction events within a population of A molecules per unit of time⁶ (microscopic rate), $\hat{\pi}$ be the population-wide probability of a reaction occurring in an ensemble of N_A molecules of A within a sufficiently small interval $\Delta t/s$ such that $0 \leq \hat{\pi} < 1$, and p be the probability that *each* molecule of A has of reacting in $\Delta t/s$. Given the Bernoulli trial interpretation for this process, the expected value for the population-wide number of transition events is $s\hat{\pi}$, and the following condition is true

$$r\Delta t = s\hat{\pi}. \quad (2.27)$$

Since any molecule of A can react at any given time and we assume that a reaction event is independent of other such events within the population of A , the population-wide probability $\hat{\pi}$ must be the sum of the per-agent probabilities p . In other words,

$$\hat{\pi} = \sum_{j=1}^{N_A} p_j = pN_A. \quad (2.28)$$

Also, we know from our discussion of microscopic rates (section 2.1) that the microscopic rate law for a 1st order process is $r = k_1N_A$, where k_1 is the microscopic rate constant. Substituting for r and $\hat{\pi}$ into equation 2.27, we get

$$k_1N_A\Delta t = spN_A \quad (2.29)$$

$$p = \frac{k_1\Delta t}{s}. \quad (2.30)$$

This provides a way to determine the per-agent probability of a transition event from the kinetic parameter k_1 , which is *measured at the level of a population*. Notice that this approach contrasts to the method developed by Gillespie (described in Appendix B) where kinetic parameters are used to determine the *population-wide* probability of a reaction event.

Transition events as Bernoulli trials

We consider the number of time steps until a reaction/transition takes place for a particular agent of A , since at that point the agent is irreversibly converted

⁶The rate of a 1st order process can vary considerably depending on the fraction of molecules of A that remain unreacted. We assume here that Δt is small enough so that r can be approximated as constant.

to X . For instance, the probability that an A agent transitions in the first subinterval of time length $\Delta t/s$ is p , in the second subinterval qp , then q^2p , etc., where q is the probability that an agent of A does *not* transition during $\Delta t/s$. We also note that the probability of no reaction taking place in any of the s subintervals of Δt is $P_{A \rightarrow X} = q^s$. Summing the probabilities of all $s+1$ possible outcomes for a given agent, we get

$$P_{tot} = p + qp + q^2p + \dots + q^{s-2}p + q^{s-1}p + q^s. \quad (2.31)$$

It is apparent that the first s terms of this sum constitute a geometric series with common ratio q and first term p ,

$$P_{tot} = \sum_{i=1}^s q^{i-1}p + q^s.$$

Using the formula for the finite sum of a geometric series, we rewrite P_{tot} as

$$P_{tot} = \frac{p(1-q^s)}{1-q} + q^s = (1-q^s) + q^s = 1, \quad (2.32)$$

where we used the fact that $p+q=1$ to simplify the expression. Therefore, the number of time steps until a molecule of A reacts is geometrically distributed.

The per-agent probability that a transition takes place within Δt is

$$\begin{aligned} P_{A \rightarrow X} &= P_{tot} - P_{A \rightarrow X} \\ P_{A \rightarrow X} &= 1 - q^s = 1 - (1-p)^s. \end{aligned} \quad (2.33)$$

Substituting for p using equation 2.30 and taking the limit as the number of time subintervals becomes infinite gives

$$\begin{aligned} P_{A \rightarrow X} &= \lim_{s \rightarrow \infty} \left[1 - \left(1 - \frac{k_1 \Delta t}{s} \right)^s \right] \\ P_{A \rightarrow X} &= 1 - e^{-k_1 \Delta t}. \end{aligned} \quad (2.34)$$

We have arrived at the same result as in equation 2.18 (P_{int}), which was derived from the microscopic integrated rate law (subsection 2.2.1). We will refer to the above expression as P_{ber} to indicate the fact that it was derived from the Bernoulli trial interpretation of 1st order reaction events.

This result should come as no surprise since equation 2.33 is the cumulative distribution function (CDF) of the geometric distribution (where there is finite number, s , of time subintervals, and therefore a discrete number of reaction occurrence failures until a successful transition), while equation 2.34 is its continuous analogue (i.e., the CDF of the exponential distribution). The *memoryless* property of this distribution confirms what we have observed in the different forms of $P_{A \rightarrow X}$ we have derived: the probability of a reaction event depends on the duration of the time interval Δt and is independent of how much time had previously elapsed without an event occurrence. This agrees with known 1st order processes. For instance, in radioactive decay it is impossible to predict (according to quantum theory) when an unstable nucleus of an atom will decay, regardless of when the atom first came into existence (Loveland et al., 2006, p. 57).

2.2.4 Relationship between P_{int} (or P_{ber}) and P_{dif}

Given the above result (equation 2.34), we consider the Taylor expansion of $P_{A \rightarrow X}$ around $\Delta t = 0$,

$$\begin{aligned} P_{A \rightarrow X} &= k_1 \Delta t - \frac{(k_1 \Delta t)^2}{2!} + \dots + (-1)^{m+1} \frac{(k_1 \Delta t)^m}{m!} \\ P_{A \rightarrow X} &= k_1 \Delta t + o(\Delta t), \end{aligned} \tag{2.35}$$

where the little- o notation⁷ refers to quadratically small terms in Δt , and is used to indicate that what amounts to the error term of the above Taylor series expansion is small compared to Δt as $\Delta t \rightarrow 0$.

Therefore, for a sufficiently small Δt , the probability of the transition $A \rightarrow X$ can be approximated by

$$P_{A \rightarrow X} \approx k_1 \Delta t, \tag{2.36}$$

which is the same result as in equation 2.25 (P_{dif}), derived from the microscopic differential rate law (subsection 2.2.2). This shows that the P_{dif} way of calculating the probability of a reaction for a single molecule of A is a linear approximation to the result in equation 2.34 (P_{int} or P_{ber}), as shown in Figure 2.1. Table 2.1 summarizes our findings thus far.

$P_{A \rightarrow X}$	Derived from:
$P_{int} = 1 - e^{-k_1 \Delta t}$	Integration of microscopic rate law: $-\frac{dN_A}{dt} = k_1 N_A$.
$P_{ber} = 1 - e^{-k_1 \Delta t}$	Transition events treated as Bernoulli trials.
$P_{dif} = k_1 \Delta t$	1) Discretized microscopic rate law: $-\frac{\Delta N_A}{\Delta t} \approx k_1 N_A$. 2) Linear approximation to P_{int} or P_{ber} .

Table 2.1 – Transition probability expressions for a 1st order process $A \rightarrow X$.

⁷The little- o notation is defined such that when $h(x) = f(x) + o(g(x))$, which upon rearrangement is $h(x) - f(x) = o(g(x))$, indicates that

$$\lim_{x \rightarrow a} \frac{h(x) - f(x)}{g(x)} = 0.$$

This succinctly conveys the idea that $h(x) - f(x)$ is of smaller order than $g(x)$, or, alternatively, that $h(x) - f(x)$ is negligible compared to $g(x)$ for x in the neighborhood of some value a (Apostol, 1967, p. 286). Hence, if the above definition describes the Taylor expansion of $h(x)$, then $o(g(x))$ can be thought of as the error of the approximation to the function h .

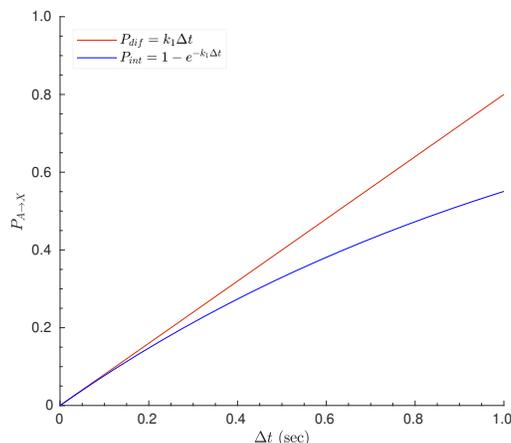


Figure 2.1 – Comparison of P_{int} (or P_{ber}) and P_{dif} as a function of Δt . P_{dif} is the linear approximation to P_{int} (or P_{ber}) in the neighborhood of $\Delta t = 0$. Note that P_{int} (or P_{ber}) asymptotically approaches 1 as $\Delta t \rightarrow \infty$, while P_{dif} increases without bound and reaches the value 1 when $\Delta t = 1/k_1$. The plotted data were calculated using the sample value of $k_1 = 0.8664 \text{ sec}^{-1}$ (which corresponds to a 1st order process with a half-life $\lambda = 0.8000 \text{ sec}$).

2.2.5 From Individuals to Populations

We hereby investigate what predictions the aforementioned per-agent probabilities make for the entire population. As we already showed, the probability that a molecule of A does *not* react is

$$P_{A \not\rightarrow X} = e^{-k_1 \Delta t},$$

and the probability of its transition to X within Δt is

$$P_{A \rightarrow X} = 1 - e^{-k_1 \Delta t}.$$

We consider the probability that μ transition events will occur in Δt within the population of N_A molecules. Thus, for $\mu = 0$ the population-wide probability is

$$\hat{P}_{\mu=0} = \binom{N_A}{0} (e^{-k_1 \Delta t})^{N_A} = e^{-k_1 N_A \Delta t}. \quad (2.37)$$

If we do a Taylor expansion of the above expression around $\Delta t = 0$, we get

$$\hat{P}_{\mu=0} = e^{-k_1 N_A \Delta t} = 1 - k_1 N_A \Delta t + o(\Delta t). \quad (2.38)$$

We continue by writing an expression for \hat{P}_{μ} when $\mu = 1$:

$$\hat{P}_{\mu=1} = \binom{N_A}{1} (1 - e^{-k_1 \Delta t}) (e^{-k_1 \Delta t})^{N_A - 1}. \quad (2.39)$$

Using the Taylor expansions $1 - e^{-k_1\Delta t} = k_1\Delta t + o(\Delta t)$, and $e^{-k_1\Delta t} = 1 - k_1\Delta t + o(\Delta t)$, we have

$$\begin{aligned}\hat{P}_{\mu=1} &= N_A(k_1\Delta t + o(\Delta t))(1 - k_1\Delta t + o(\Delta t))^{N_A-1} \\ &= k_1N_A\Delta t + o(\Delta t).\end{aligned}\quad (2.40)$$

Continuing this process with the per-agent probability expressions we have derived, we can write a general expression for \hat{P}_μ ,

$$\hat{P}_\mu = \binom{N_A}{\mu} (1 - e^{-k_1\Delta t})^\mu (e^{-k_1\Delta t})^{N_A-\mu}, \quad (2.41)$$

where this expression is true for an arbitrary integral number of transition events $0 \leq \mu \leq N_A$.

We want to obtain a measure of the population-wide probability, \hat{P}_μ , of more than one transition event occurring within Δt . If we make the same substitutions as above, then in the case of $\mu > 1$, we have

$$\hat{P}_{\mu>1} = \binom{N_A}{\mu} (k_1\Delta t + o(\Delta t))^\mu (1 - k_1\Delta t + o(\Delta t))^{N_A-\mu}. \quad (2.42)$$

It suffices to consider the above expression insofar as it shows that $\hat{P}_{\mu>1} \propto (\Delta t)^\mu$, and will therefore be $o(\Delta t)$ in the neighborhood of $\Delta t = 0$. Thus, $\hat{P}_{\mu>1}$ goes to zero faster than Δt for multiple reaction events.

This result along with the previous two (equations 2.38, 2.40) are identical with those obtained from the basic premise of stochastic chemical kinetics as $\Delta t \rightarrow 0$, and which lead to a derivation of the CME (Gillespie, 1992, pp. 420-423). We therefore conclude that our agent-based approach is consistent with the CME at the limit of $\Delta t \rightarrow 0$.

To show that 2.41 is a distribution, we consider the sum of probabilities for all possible numbers of transition events ($0 \leq \mu \leq N_A$),

$$\hat{P}_{tot} = \sum_{\mu=0}^{N_A} \binom{N_A}{\mu} (1 - e^{-k_1\Delta t})^\mu (e^{-k_1\Delta t})^{N_A-\mu}, \quad (2.43)$$

which is obviously a binomial expansion and is equal to 1.

$$\hat{P}_{tot} = [(1 - e^{-k_1\Delta t}) + (e^{-k_1\Delta t})]^{N_A} = 1 \quad (2.44)$$

Therefore, μ is a binomially-distributed variable with mean

$$\langle \mu \rangle = N_A (1 - e^{-k_1\Delta t}). \quad (2.45)$$

This is not a surprising result given that we assumed the 1st order process to consist of a sequence of Bernoulli trials.

Since Δt can be made arbitrarily small, we expect μ to be a small nonnegative integer in the general case. In accordance with this expectation, we consider the case where $N_A \gg \mu$, and $\binom{N_A}{\mu}$ becomes

$$\frac{N_A!}{(N_A - \mu)! \mu!} = \frac{N_A(N_A - 1) \cdots (N_A - \mu + 1)}{\mu!} \approx \frac{N_A^\mu}{\mu!}. \quad (2.46)$$

(We will have more to say about how reasonable this approximation is for small μ in the following paragraph.) We can substitute this result into equation 2.41, as well as the expansion $1 - e^{-k_1 \Delta t} = k_1 \Delta t + o(\Delta t)$, and $N_A - \mu \approx N_A$. The expression for \hat{P}_μ becomes

$$\begin{aligned} \hat{P}_\mu &= \frac{N_A^\mu}{\mu!} (k_1 \Delta t + o(\Delta t))^\mu (e^{-k_1 \Delta t})^{N_A} \\ \hat{P}_\mu &= \frac{(k_1 N_A \Delta t)^\mu}{\mu!} e^{-k_1 N_A \Delta t} + o(\Delta t). \end{aligned} \quad (2.47)$$

It is evident that as $\Delta t \rightarrow 0$, \hat{P}_μ is a homogeneous Poisson distribution⁸ with average number of transition events $\langle \mu \rangle = k_1 N_A \Delta t = r \Delta t$, where r is the microscopic rate of a 1st order process. Notice that this average value of μ is the first order approximation to the result from the binomial distribution (equation 2.45) as $\Delta t \rightarrow 0$.

The significance of this result that the number of transition events, μ , is Poisson-distributed is that the time between transition events (interoccurrence time) asymptotically approaches the exponential distribution as $\Delta t \rightarrow 0$. The same distribution is predicted starting from the basic premise of the stochastic formulation of chemical kinetics: namely, that any transition event has a finite probability of occurring (Gillespie, 1977, 2007). Indeed, the CME predicts that the interoccurrence time is exponentially distributed (see Appendix B), and this fact is precisely the starting point for the implementation of Gillespie's algorithm. We have thus shown that the per-agent transition probabilities we have derived lead to the expected distribution of the population-wide number of transition events in the limit of infinitesimally small time intervals.

This result depends heavily on the approximation to the binomial coefficient, which is reasonable only for small values of μ and $N_A \gtrsim 100$. For instance, for a modestly-sized population made up of 100 molecules in which two reactions happen within a given time interval, the binomial coefficient is $\binom{100}{2} = 4950$, while $\frac{100^2}{2!} = 5000$, for approximately a 1% difference between the two values. For smaller populations, the error increases rather fast (e.g., for $N_A = 10$ and $\mu = 2$, the error is 11%). However, in the special cases where $\mu = 0$ or $\mu = 1$, the result of equation 2.46 is exact. That is, $\binom{N_A}{0} = 1$ and $\binom{N_A}{1} = N_A$. Thus, if we assume that Δt is small enough so that *at most* one reaction event can occur in the population (which is supported by equation 2.42 and ensuing discussion),

⁸A homogeneous Poisson distribution assumes that the process occurs at a constant rate. In our case, as $\Delta t \rightarrow 0$ the rate of the process can be considered constant for the duration of the time interval.

then equation 2.47 is only an approximation insofar as $N_A - 1 \approx N_A$ (when $\mu = 1$).

We address the issue of how small Δt should be in order for $\mu \in \{0, 1\}$ at the end of the following subsection.

2.2.6 The ABK Algorithm and its Implementation

For the unidirectional or irreversible process $A \rightarrow X$, only molecules of A can transition to X and we need to keep track of the populations as a function of time. We represent the initial population of *each species* by an array (or row vector) of size equal to that of the initial population's, and each placeholder of the array can store the *state* of that particular member of the population. One way to represent the state of each molecule/agent is a binary system: 0 if the agent does not yet exist or has already been converted to another species, and 1 if it is a member of the population of that particular species (we can alternatively describe the state of each agent as 'dead' or 'alive'). This way, each member of the populations is treated as a distinct entity for which a transition probability can be calculated.

The ABK algorithm can be generally broken down into a small number of distinct steps. We consider the total time T over which we wish to simulate the process as consisting of non-overlapping subintervals Δt (whose duration can be fixed or variable depending on which method a modeler chooses to use). The n^{th} iteration of the algorithm determines the change in the population size for the n^{th} time interval Δt using the following sequence of steps:

1. Evaluate the probability of transition $P_{A \rightarrow X}$ for *each molecule of A* present at time t_n using either the P_{int} (or P_{ber}) or P_{dif} formalism described in the previous section.
2. Draw a random number r_u from the standard uniform distribution⁹ in the interval $(0, 1)$.
3. Check the condition that $r_u < P_{A \rightarrow X}$ for *each agent/molecule of A*. If the condition is true, then the state of that particular molecule of A is changed to 0, while a molecule of X is created and its state is set to 1. If the condition is false then the state of molecule A does not change.
4. Sum up the number of A and X agents/molecules in their respective populations after all probability conditions have been checked. This is the updated species population sizes at t_{n+1} , the time point corresponding to the end of time interval Δt .
5. Update the absolute time $t_{n+1} = t_n + \Delta t$.

⁹All pseudo-random number generators are based on deterministic algorithms, therefore the generated sequence of numbers has a characteristic period. We use here the Mersenne Twister generator (<http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/emt.html>) that has a particularly long period of $2^{19937} - 1$.

This procedure can be easily implemented in any programming language using the `for` loop structure (or its equivalent) for sampling through all members of the populations and for all time intervals until time T is reached.

We tested the ABK algorithm for the 1st order process $A \rightarrow X$ with a rate constant $k_1 = 0.8664 \text{ sec}^{-1}$, which corresponds to a half-life $\lambda = 0.8000 \text{ sec}$ (see representative Matlab code on page 274). Figure 2.2 shows two sample runs of the simulation for different initial populations of A , while the initial population of X is 0 in all cases. We used a fixed time step of $\Delta t = 0.10 \text{ sec}$ and the P_{int} method of determining transition probabilities (equations 2.18, 2.34). The increased granularity in the time evolution of the population of A becomes more evident as populations of smaller size are considered. The solid green curve is the solution to the microscopic differential rate law for the process (equation 2.14), and is shown for comparison to the results of the stochastic simulation.

The simulation of a process through the ABK methodology can be repeated a user-specified number of times such that the ensemble of simulated time trajectories can be used to obtain statistical measures of the system's evolution. As an example, we plot in Figure 2.3 the average and standard deviation for the population size of A , $N_A(t)$, after repeating the simulation 50 times. The time evolution for decreasing orders of magnitude of the initial population of A is shown, along with insets zooming in at the time interval from 2.0 – 2.4 sec to highlight the relative scale of the standard deviation in each case. The simulation curves match the theoretical curves obtained from the integrated rate law exceedingly well (coefficient of determination,¹⁰ $R^2 = 0.9999, 0.9999, 0.9998, 0.9987$ for $N_{A,i} = 10000, 1000, 100, 10$, respectively), especially considering the small sample trajectory size ($n = 50$). The most visibly granular time evolution of the population size is for $N_{A,i} = 10$, but it still agrees with the theoretical curve obtained from the integrated rate law (green dotted curve).

It is instructive to compare the standard deviation values obtained in each case and normalize them to the initial population size of A . Figure 2.4a shows that the relative standard deviation increases for progressively smaller populations and that they match those predicted by the CME. Specifically, the CME standard deviation for a 1st order process can be shown to be (see Appendix A.1 for a derivation of the expression for the variance, equation A.23),

$$SDev(\langle N_A(t) \rangle) = \sqrt{N_{A,i} e^{-k_1 t} (1 - e^{-k_1 t})}. \quad (2.49)$$

¹⁰The coefficient of determination, R^2 , is a statistic that can be interpreted as the proportion of the variance in the dependent variable that can be predicted from the independent variable. R^2 can be computed for a set of estimates y_i and the corresponding predicted values f_i using the definition

$$R^2 = 1 - \frac{SSR}{SST}, \quad (2.48)$$

where $SSR = \sum_i (y_i - f_i)^2$ is the sum of the squared residuals (a metric that is minimized in regression analyses), and $SST = \sum_i (y_i - \bar{y})^2$ is the total sum of squares. This generally works for assessing the fit of any data to either a linear or nonlinear predicted curve.

Follow this [link](#) to see all of the code used for the simulations in this section.

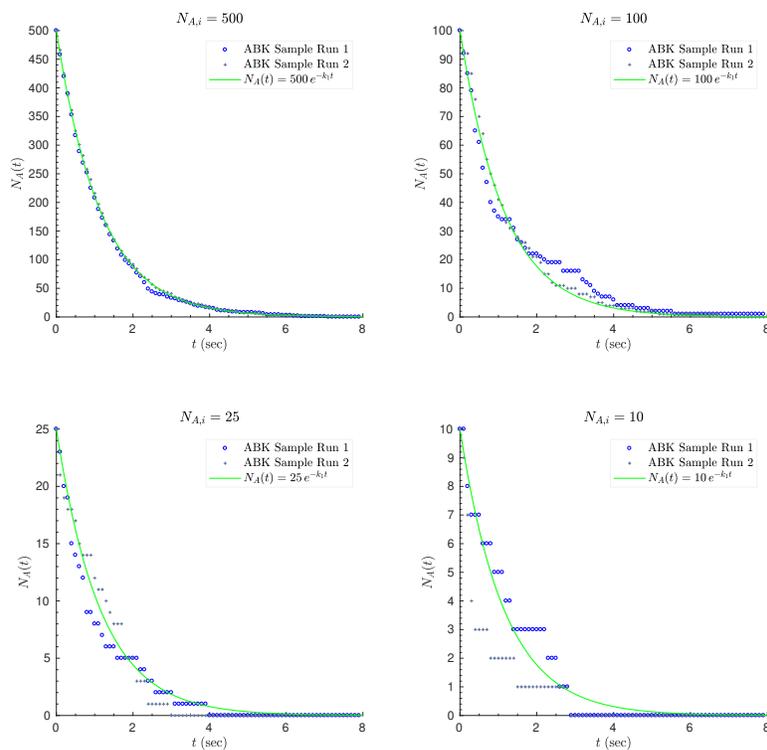


Figure 2.2 – Sample runs for $A \rightarrow X$ for populations of different sizes. Two sample simulation runs using the ABK algorithm are shown for the 1st order conversion $A \rightarrow X$. We consider initial populations of $N_{A,i} = 500, 100, 25, 10$ molecules/agents of A respectively. In all cases, the half-life of the process is $\lambda = 0.8000 \text{ sec}$, and therefore the microscopic rate constant is $k_1 = \ln 2 / \lambda = 0.8664 \text{ sec}^{-1}$. We calculated the transition probabilities using the P_{int} (equation 2.18) expression, and used fixed time step increments of $\Delta t = 0.10 \text{ sec}$. The green curve is the solution to the differential equation describing this process (i.e., the integrated rate law). Note that the plots differ in the scaling of the vertical axis.

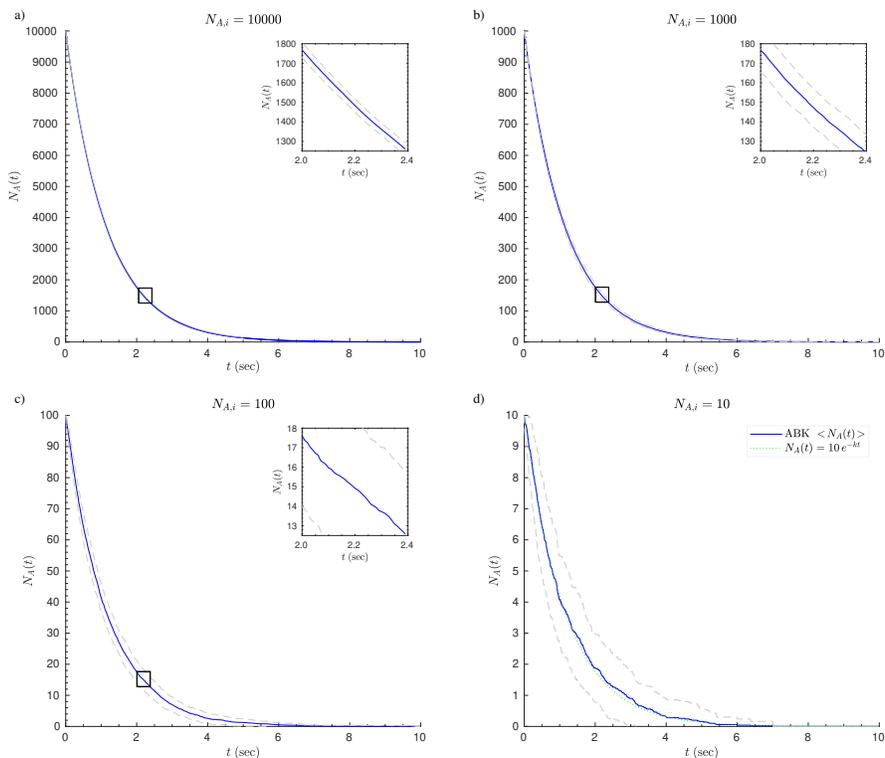


Figure 2.3 – Time course of $A \rightarrow X$ with populations of different initial size. ABK simulations of the 1st order process $A \rightarrow X$ with a fixed time step of $\Delta t = 0.01$ sec, rate constant $k_1 = 0.8664 \text{ sec}^{-1}$ ($\lambda = 0.8000 \text{ sec}$), and using the P_{int} transition probability expression were performed for homogeneous populations of initial size $N_{A,i} = 10000, 1000, 100, 10$ molecules/agents, respectively. In each case, we repeated the simulation 50 times and plotted the average population size of A and its standard deviation at each time step. Blue curve: average population size, $\langle N_A(t) \rangle$; gray dashed curves represent the one-standard-deviation envelope (average ± 1 standard deviation). For populations of initial size $N_{A,i} = 10000, 1000, 100$ (plots a-c) the inset graph is an enlarged window of the time interval $2.0 - 2.4$ sec to highlight the difference in the magnitude of the standard deviation. The dotted green curve in panel d ($N_{A,i} = 10$) corresponds to the solution to the differential equation (it is not shown in plots a-c for the sake of clarity since the simulated population average is virtually indistinguishable from the deterministic solution). Note that the vertical axis is scaled differently in the plots.

The maximum standard deviation occurs at $t = \ln 2/k_1 = \lambda$, and has a value of

$$SDev_{max} \equiv SDev \left(\langle N_A(\ln 2/k_1) \rangle \right) = \frac{\sqrt{N_{A,i}}}{2}. \quad (2.50)$$

Upon dividing both sides by $N_{A,i}$, we get

$$\frac{SDev_{max}}{N_{A,i}} = \frac{1}{2\sqrt{N_{A,i}}}. \quad (2.51)$$

Finally, we transform this equation into a linear relationship between logarithmic quantities,

$$\ln \left(\frac{SDev_{max}}{N_{A,i}} \right) = -\frac{1}{2} \ln(N_{A,i}) - \ln 2. \quad (2.52)$$

Figure 2.4b shows a graph of $\ln(SDev_{max}/N_{A,i})$ vs. $\ln(N_{A,i})$, where we observe a very close agreement between the CME (dotted line) and ABK data (correlation coefficient, $R = 0.9998$).

As we showed in a preceding section, the 1st order transition $A \rightarrow X$ is a Poissonian process for small Δt as simulated with the use of the ABK algorithm (equation 2.47). Therefore, the distribution of time between events, or interoccurrence time, should be described by the exponential probability density function (PDF)

$$f(t, k_1) = k_1 e^{-k_1 t}. \quad (2.53)$$

We now show that our simulations produce a distribution that agrees with the exponential probability density function. We simulated the time evolution of different initial populations of A and kept track of the time at which each agent transitioned to X . We repeated the simulation 1000 times for the same value of k as before, fixed time step increments of $\Delta t = 0.01$ sec, and the P_{ber} form of transition probabilities. We then constructed a histogram of the fraction of transition events within successive time intervals of the simulated time. Figure 2.5 shows such a histogram with 20 bins. We verified that the choice of initial population size, $N_{A,i}$, has no effect on the fractional values obtained, as would be expected for a 1st order process where the agents transition independently of each other. The histogram is in excellent agreement with the exponential PDF ($R^2 = 0.9997$), which is scaled appropriately to account for the chosen number of bins.¹¹ Therefore, the ABK algorithm produces a distribution of transition events that matches the theoretical expectation.

We now examine how the choice of using the P_{int} (or P_{ber}) or P_{dif} expressions of determining transition probabilities and the duration of time interval Δt affect the agreement between the ABK-simulated population time courses and the theoretically expected ones (i.e., based on the deterministic integrated rate law or CME treatment of this process). We use the following approach: 1) We determine the average population size and its standard deviation at each time

¹¹The scaled form of the exponential PDF is $f(t, k_1) = \frac{T}{\text{bins}} k_1 e^{-k_1 t}$, where T represents the total simulated time. In this case, for $T = 10$ sec and 20 bins, the scaling factor is $1/2$.

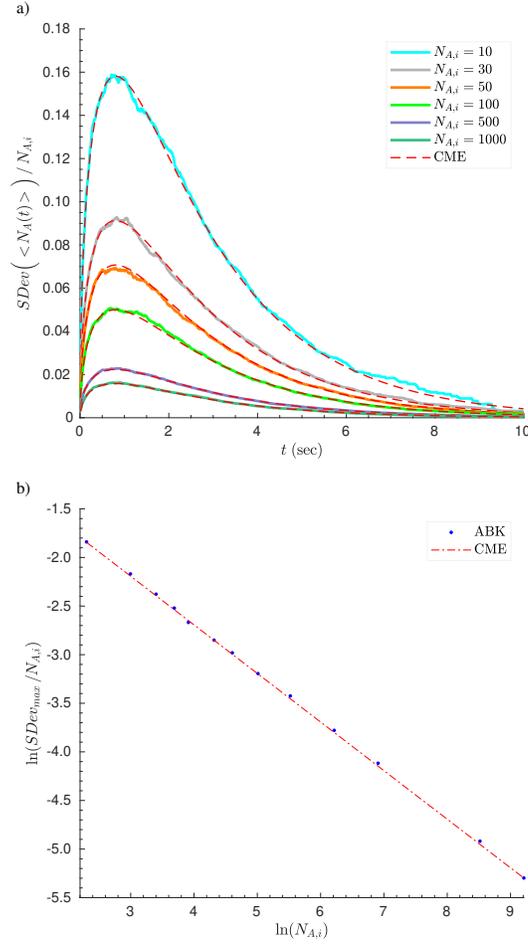


Figure 2.4 – Standard deviation of 1st order process average trajectory. A comparison between the relative standard deviation values obtained through repetitions of the ABK simulation ($n = 1000$) and the values predicted from the solution to the CME for different initial population sizes. **a)** Relative standard deviation for initial populations $N_{A,i} = 1000, 500, 100, 50, 30, 10$ as a function of time. **b)** We show the *maximum* relative standard deviation values (at $t = \ln 2/k_1 = \lambda$) for different initial population sizes (blue filled circles) and compare them to the theoretical prediction from the CME (dash-dotted red line; see equation 2.52; correlation coefficient, $R = 0.9998$).

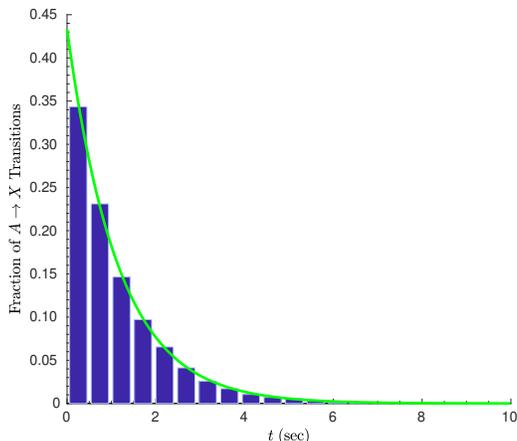


Figure 2.5 – Distribution of 1st order transition events. Histogram of the fraction of species A agent transitions within successive simulated-time intervals of duration 0.50 sec. The data was aggregated over 1000 repetitions of the ABK simulation with fixed time step increments of $\Delta t = 0.01$ sec, $N_{A,i} = 10$, and the P_{ber} form of transition probabilities. We have superimposed the PDF of the exponential distribution, $f(t, k_1) = k_1 e^{-k_1 t}$ (green curve), to highlight the agreement between the simulation results and theoretical expectation ($R^2 = 0.9997$).

point from an ensemble of simulated trajectories (in this case, $n = 500$), thus generating curves like those shown in Figure 2.3; 2) We then obtain the best-fit parameter k_{est} using $N_A(t) = N_{A,i} e^{-k_{est} t}$ for the curve corresponding to the average population size, and separately for the curves defining the one-standard-deviation envelope (see sample Matlab code for fitting a curve to a set of data points on page 288). We plot k_{est} vs. Δt for populations of different initial size (Figure 2.6), where the error bars represent the range of k_{est} values defined by the one-standard-deviation envelope of the average population time trajectories. The overall trend is that the simulations using the P_{int} implementation closely agree with the true value of $k_1 = 0.8664 \text{ sec}^{-1}$ (e.g., for $N_{A,i} = 10$, t -test $p = 0.5245$). Note that this result is typical for all of the differently sized populations presented here. On the other hand, the P_{dif} transition probabilities produce time trajectories overestimating the true value of k_1 as the fixed time interval Δt increases. This effect is much more pronounced for larger populations (clearly visible in Figure 2.6a,b), while for the small populations of $N_{A,i} = 25$ and 10 molecules/agents (Figure 2.6e,f) the P_{int} and P_{dif} results agree with each other and the true value of k_1 for smaller values of Δt . More specifically, for $N_{A,i} = 10$, the P_{dif} -obtained k_{est} agrees with the true value of k_1 for $\Delta t = 0.01 - 0.06$ sec (t -test $p = 0.0184$, signifying failure to reject the null hypothesis for a level of significance $\alpha = 0.01$), while at higher values of Δt , $k_{est} > k_1$. We observed the same effect for small populations of similar size: there is a Δt value below which the P_{dif} -derived k_{est} values are not statistically significantly

different from $k_1 = 0.8664 \text{ sec}^{-1}$. Moreover, the threshold Δt value for statistical significance increases as $N_{A,i}$ decreases. Finally, the error bars in the plots of Figure 2.6 verify that the variance of the average time trajectories increases with smaller population sizes, in accordance with our previous observations (Figures 2.3 and 2.4).

We conclude based on the above data that the P_{dif} implementation of the transition probability for a 1st order process can be used as an alternative to the P_{int} one for small population sizes and Δt values.

Computational speed using P_{int} vs. P_{dif}

Figure 2.6 clearly shows that using the ABK algorithm with the P_{int} transition probabilities gives consistently accurate results for a wide range of Δt values. However, even though P_{dif} is an approximation to P_{int} and its effectiveness is limited to smaller values of Δt , we determined if there is an advantage in computational speed given the algebraic simplicity of the P_{dif} expression. We used a simple Matlab script with a `for` loop to assess the speed of merely computing the transition probability and found that the P_{dif} implementation is faster than P_{int} by approximately up to an order of magnitude. This fact may be considered when seeking a balance between choosing a sufficiently small value of Δt (which leads to more iterations of the algorithm) and the savings in computational time afforded by computing the P_{dif} expression. The speed of computing P_{dif} may be especially useful to consider in cases where the transition probability needs to be computed at each iteration of the algorithm, such as when simulating systems of reactions that include 2nd order processes (as we will see in the next section), or when the algorithm uses variable time step increments (discussed below).

Fixed vs. variable Δt increments

The distribution of the length of Δt is also a consideration in the algorithm. We tested the following approaches:

1. Δt is of constant or fixed length throughout the simulation. All of the results presented in this section thus far have used fixed time step increments.
2. Δt is variable, depends inversely on the population size, and is reevaluated at each time-advancing iteration of the algorithm. One way to implement this is to consider the discretized differential rate law for this process (see subsection 2.2.2). We may choose the average change in the population size within Δt to be 1, such that

$$\begin{aligned}
 -\langle \Delta N_A \rangle &= 1 = k_1 N_A(t) \Delta t \\
 \Delta t &= \frac{1}{k_1 N_A(t)}. \tag{2.54}
 \end{aligned}$$

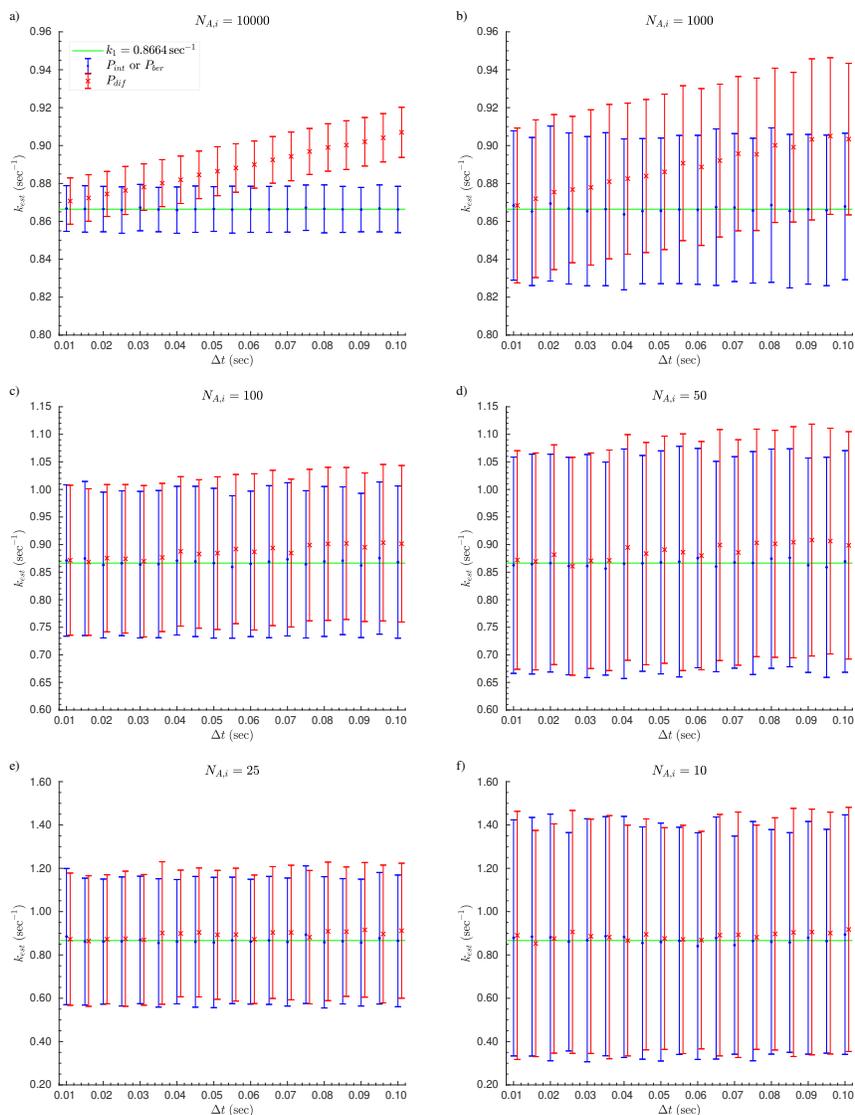


Figure 2.6 – Effect of time interval Δt on accuracy of P_{int} (or P_{ber}) and P_{dif} expressions. We ran the ABK simulation for the conversion $A \rightarrow X$ with half-life $\lambda = 0.8000 \text{ sec}$ ($k_1 = 0.8664 \text{ sec}^{-1}$), in populations of initial size $N_{A,i} = 10000, 1000, 100, 50, 25, 10$ molecules/agents, and different values of fixed time step increments (Δt) in the range of 0.01 to 0.10 sec. For each combination of $N_{A,i}$ and Δt values, we repeated the simulation 500 times and obtained the average population size for each time step in the reaction time course. We then fit the resulting time course curves to the equation $\langle N_A(t) \rangle = N_{A,i} e^{-k_{est} t}$ to obtain k_{est} , the rate constant which provides the best fit to the data. We performed the same procedure to the one-standard-deviation envelope of the time course data to obtain an estimate for the magnitude of the uncertainty in the k_{est} values (shown as error bars in the plots). The results for the different initial populations show how using P_{int} (or P_{ber} ; blue) or P_{dif} (red) affects k_{est} . The green line shows the true value of k_1 for comparison. Notice that the scale of the vertical axis values changes for each row in the tableau of plots.

This value of Δt will, on average, produce a change in the population size of just one molecule/agent. The choice of $\langle \Delta N_A \rangle = 1$ is arbitrary and can be made smaller¹² if a finer sampling of Δt is desired (at the cost of additional computational effort).

3. Given that the exponential distribution describes the time between events in a Poisson process (as a 1st order process is known to be), we can treat Δt as exponentially distributed with an average value obtained from the numerical result of equation 2.54. We therefore reevaluate Δt at every iteration (time step) of the algorithm by drawing a (pseudo-)random number from the exponential distribution with mean $1/k_1 N_A(t)$. Note that this choice of Δt will be used to evaluate the transition probability of each agent within a given time step.

We have indeed performed simulations using variable Δt values obtained through approaches 2 and 3, and observed just as good of an agreement of the stochastic average time trajectories (and their variances) with the deterministic (or CME) prediction (results not shown).

¹²A fractional value for $\langle \Delta N_A \rangle$ is acceptable since it represents the *average* change in population size during Δt .

2.3 2nd Order Processes

We consider the following 2nd order elementary processes



and



with microscopic kinetic constant k_2 (with units of sec^{-1}) and X is a generalized product species. We refer to 2.55 as a *homologous* 2nd order process because the molecules/agents participating in the reaction are of the same species, while 2.56 represents a *heterologous* process.

As we did for 1st order processes, we proceed to derive different forms of the transition probability and compare their accuracy.

2.3.1 $2A \rightarrow X$

Transition probability from the canonical rate law

The microscopic rate law for this process with respect to A , r_A ,¹³ is

$$r_A = -\frac{dN_A}{dt} = k_2 N_A^2. \quad (2.57)$$

This is the canonical form of a 2nd order reaction rate law as presented in the study of chemical kinetics. The underlying assumption that this statement applies to large populations of A will become apparent later in this subsection when we discuss the agent-based rate law.

We consider an arbitrary time interval $\Delta t = t_{n+1} - t_n$, where the variable n can represent the n^{th} time step of the ABK algorithm. Upon integration of the above rate law, we obtain

$$\frac{1}{N_A(t_{n+1})} - \frac{1}{N_A(t_n)} = k_2 \Delta t, \quad (2.58)$$

where $N_A(t_n)$ and $N_A(t_{n+1})$ are the number of molecules/agents of A at the beginning and end of Δt respectively. Upon rearrangement, we get

$$N_A(t_{n+1}) = \frac{N_A(t_n)}{k_2 N_A(t_n) \Delta t + 1}. \quad (2.59)$$

Note that taking the limit of the right-hand side of this equation as $\Delta t \rightarrow \infty$ results in $N_A(t_{n+1}) \rightarrow 0$. Given that we consider the members of the population A as discrete agents, it is easy to realize that the predicted end behavior is true if and only if the initial number of molecules of A , $N_{A,i}$, is an even number. In that case, all pairs of molecules of A will react with each other and none

¹³As opposed to the species-independent rate $r = -\frac{1}{2} \frac{dN_A}{dt} = \frac{dN_X}{dt}$.

will eventually be left. On the other hand, if $N_{A,i}$ is odd then we predict $N_A(t_{n+1}) \rightarrow 1$ after a sufficient amount of time has elapsed. This is the first example we have encountered where a differential equation fails to capture basic facts about the stoichiometry of a process, given the discrete nature of agents and the transition events between them.

As we did previously (equation (2.16)), we define the probability of *no reaction* occurring within Δt for a given molecule/agent of A as

$$P_{2A \rightarrow X} \equiv \frac{N_A(t_{n+1})}{N_A(t_n)} = \frac{1}{k_2 N_A(t_n) \Delta t + 1}, \quad (2.60)$$

and therefore,

$$P_{2A \rightarrow X} = 1 - \frac{1}{k_2 N_A(t_n) \Delta t + 1}. \quad (2.61)$$

This is the per-agent probability of a reaction happening in Δt . It explicitly depends on the population size of species A at the beginning of Δt (time t_n) and has a range of $[0, 1)$. We will refer to this expression from now on as P_{can} to reflect the fact that it was derived from the canonical form of the rate law. Notice that $P_{2A \rightarrow X}$ satisfies the Markov property in that it only depends on the last known population size of A , $N_A(t_n)$.

We will next show that P_{can} is in fact an approximation most suitable for simulating the time evolution of large populations of A .

Transition probability from the agent-based rate law

The rate law presented above (equation 2.57) is not entirely correct if one considers that two distinct molecules of A participate in the reaction. That is, if a particular member of the population of A does indeed react, then the size of the molecular population from which to choose a partner is $N_A - 1$. This is akin to considering the number of distinct pairs of agents of A in the entire population. The number of distinct pairwise interactions would normally include a statistical factor of $1/2$, given that all members of the population are identical in a homogeneous population. Thus, the number of possible pairs is $\frac{1}{2}N_A(N_A - 1)$. However, since the algorithm evaluates the transition probability (and tests it against a random number from the uniform distribution) for *each* agent of A , the $i - j^{th}$ agent interaction is considered separately from the $j - i^{th}$. Therefore, the number of distinct pairs is $N_A(N_A - 1)$ from the perspective of ABK, and the rate's dependence on the population size is more properly written as

$$r_A = -\frac{dN_A}{dt} = k_2 N_A (N_A - 1). \quad (2.62)$$

We refer to this as the *agent-based* form of the differential rate law, in recognition of the fact that it treats the population as a discrete quantity (i.e., the members of the population are distinguishable entities). This interpretation is supported by the fact that manipulation of the CME for this process results in an analogous form of the differential rate law of the average population size, $\langle N_A(t) \rangle$ (see derivation in Appendix A.3 leading to equation A.75).

We hereby summarize the steps required to get the agent-based integrated rate law. We begin by separating the variables,

$$-\frac{dN_A}{N_A(N_A - 1)} = k_2 dt,$$

and expanding the left-hand side of the equation by using partial fraction decomposition,

$$\frac{dN_A}{N_A} - \frac{dN_A}{N_A - 1} = k_2 dt.$$

Upon integration of both sides over the time interval $\Delta t = t_{n+1} - t_n$, we get

$$\begin{aligned} \ln\left(1 - \frac{1}{N_A(t_{n+1})}\right) - \ln\left(1 - \frac{1}{N_A(t_n)}\right) &= -k_2 \Delta t \\ 1 - \frac{1}{N_A(t_{n+1})} &= \left(1 - \frac{1}{N_A(t_n)}\right) e^{-k_2 \Delta t}. \end{aligned}$$

Finally, upon further rearrangement we get the *agent-based* integrated rate law

$$N_A(t_{n+1}) = \frac{N_A(t_n)}{N_A(t_n)(1 - e^{-k_2 \Delta t}) + e^{-k_2 \Delta t}}. \quad (2.63)$$

Note that in the limit of $\Delta t \rightarrow \infty$, the population size $N_A(t_{n+1}) \rightarrow 1$. Starting from the microscopic form of the agent-based differential rate law (equation 2.62) we have obtained an expression that predicts one molecule of A will remain as the pool of A molecules is depleted. This is a by-product of solving the differential equation analytically and makes intuitive sense only in the case of $N_{A,i}$ being an odd number.

We have thus derived two forms of the integrated rate law (canonical and agent-based, equations 2.59 and 2.63 respectively) that predict end behavior that does not take into account if the initial population size is even or odd. However, it is evident that the larger the population of A is, the more insignificant this difference in end behavior between the two forms is. The relationship between these forms can be easily seen by using the linear approximation $1 - e^{-k_2 \Delta t} \approx k_2 \Delta t$ (valid when $\Delta t \rightarrow 0$), which we substitute into both exponential terms in the denominator of expression 2.63 and rearrange to get

$$N_A(t_{n+1}) = \frac{N_A(t_n)}{k_2 [N_A(t_n) - 1] \Delta t + 1}. \quad (2.64)$$

This is identical to equation 2.59 for large populations of A where $N_A(t_n) - 1 \approx N_A(t_n)$. Note how the two forms of the integrated rate law are essentially the same for small Δt , while they deviate in terms of their end behavior. In any case, the ABK algorithm considers iterations of time intervals Δt that are as small as possible (with consideration given to the computational cost of choosing too small of a Δt value), therefore the choice of which integrated rate law to use in defining the transition probability is not particularly important for larger populations of A .

We now use equation 2.63 to define the probability of *no reaction* occurring in Δt as

$$P_{2A \rightarrow X} \equiv \frac{N_A(t_{n+1})}{N_A(t_n)} = [N_A(t_n) (1 - e^{-k_2 \Delta t}) + e^{-k_2 \Delta t}]^{-1}, \quad (2.65)$$

and the per-agent transition probability is then

$$P_{2A \rightarrow X} = P_{int} = 1 - [N_A(t_n) (1 - e^{-k_2 \Delta t}) + e^{-k_2 \Delta t}]^{-1}. \quad (2.66)$$

We will refer to this probability as P_{int} to reflect the fact that it was derived from the integrated rate law of the agent-based form of the differential equation describing the process. P_{int} is only defined in the interval $[0, 1)$ for $N_A(t_n) \geq 1$. It is important to notice that the transition probability for 1st order processes did *not* depend on the population size, while 2nd order processes require consideration of the population size, $N_A(t_n)$, at each iteration of the ABK algorithm (as equation 2.66 indicates).

We invite the reader to confirm that the Taylor expansion of the expression for P_{int} around $\Delta t = 0$ is

$$P_{int} = k_2 [N_A(t_n) - 1] \Delta t + o(\Delta t). \quad (2.67)$$

The usefulness of this result will become apparent in the discussion that follows on the derivation of the transition probability from the discretized differential rate law (P_{dif}).

P_{dif} from the differential rate law

We start with the agent-based 2nd order differential rate law $r_A = -\frac{dN_A}{dt} = k_2 N_A (N_A - 1)$ and discretize of the differential quantities dN_A and dt . That is, we assume that Δt is small but finite such that ΔN_A is also finite. We rewrite the discretized rate law as

$$-\frac{\Delta N_A}{\Delta t} \approx k_2 N_A (N_A - 1), \quad (2.68)$$

where $\Delta t = t_{n+1} - t_n$. We remind the reader that the population size N_A is time-dependent: $N_A(t_n)$. We then multiply both sides by $\Delta t / N_A(t_n)$ and use the same definition of P_{dif} as in equation 2.25,

$$P_{2A \rightarrow X} \equiv P_{dif} = -\frac{\Delta N_A}{N_A(t_n)}, \quad (2.69)$$

to obtain an expression for P_{dif}

$$P_{dif} = k_2 [N_A(t_n) - 1] \Delta t. \quad (2.70)$$

Notice that just like in the case of 1st order processes, the expression for P_{dif} is the linear approximation to P_{int} (equation 2.67) as $\Delta t \rightarrow 0$ and does not have

an upper bound. Thus, one must choose Δt with caution such that $P_{dif} < 1$ for all possible values of $N_A(t_n)$. Equation 2.70 suggests the upper limit for the value of Δt in each iteration of the ABK algorithm should be

$$\Delta t_{max} = \frac{1}{k_2 [N_A(t_n) - 1]}. \quad (2.71)$$

Transition probability of a Bernoulli process

We treat a 2nd order process in a similar way to that presented in subsection 2.2.3 for 1st order processes. We are interested in determining the probability that the transition $2A \rightarrow X$ occurs within a time interval $\Delta t = t_{n+1} - t_n$, which we divide into s subintervals of equal length. Let r_A be the observed average number of reaction events within a population of molecules of A per unit of time (microscopic rate),¹⁴ $\hat{\pi}$ be the population-wide probability of a reaction occurring in an ensemble of N_A molecules within a sufficiently small interval $\Delta t/s$ such that $0 \leq \hat{\pi} < 1$, and p be the probability that *each* molecule of A has of reacting in $\Delta t/s$.

We assume that a transition event is independent of other such events in the vicinity of a given molecule of A . Moreover, we ignore details inherent to the microscopic nature of 2nd order chemical processes, such as the proximity of two molecules of A , the energy and geometry of effective collisions, and the diffusion coefficients of the molecular species. Accordingly, we stress that treating this process as a Bernoulli trial will yield *an approximation of the transition probability* (while the same approach produced an exact result for 1st order systems, where transition events are truly independent of one another). We aim to show that this treatment will be phenomenologically useful in simulating 2nd order processes despite the fact it is an approximation of how real systems, chemical or otherwise, may behave.

Given these assumptions, we interpret this process as a sequence of Bernoulli trials, with the expected value for the population-wide number of transition events being $s\hat{\pi}$, and the following condition being true

$$r_A \Delta t = s\hat{\pi}. \quad (2.72)$$

The population-wide probability $\hat{\pi}$ must be the sum of the per-agent probabilities p as long as each agent's transition occurs independently of the others. In other words,

$$\hat{\pi} = \sum_{j=1}^{N_A} p_j = pN_A. \quad (2.73)$$

We then substitute the above expression and equation 2.62 for r_A into equation 2.72 to obtain

$$k_2 N_A (N_A - 1) \Delta t = spN_A, \quad (2.74)$$

¹⁴We assume that Δt is small enough so that r_A can be approximated as constant during this time interval.

and therefore the per-agent probability is

$$p = \frac{k_2(N_A - 1)\Delta t}{s}. \quad (2.75)$$

As previously, N_A stands for the population size at the beginning of time interval Δt (alternatively denoted as $N_A(t_n)$).

We then proceed as for 1st order processes (subsection 2.2.3) to arrive at

$$\begin{aligned} P_{2A \rightarrow X} &= \lim_{s \rightarrow \infty} \left[1 - \left(1 - \frac{k_2(N_A - 1)\Delta t}{s} \right)^s \right] \\ P_{2A \rightarrow X} &= 1 - e^{-k_2(N_A - 1)\Delta t}. \end{aligned} \quad (2.76)$$

This is recognizable as the mathematical form of the transition probability of a 1st order process, except the exponent has an additional term reflecting the fact that the probability depends on the population size of A at the beginning of Δt . We will heretofore refer to the result of equation 2.76 as P_{ber} , as an indication that it was derived from the Bernoulli trial interpretation of this process.

Given the above result for P_{ber} , the reader can verify that following the same arguments as the ones presented in subsection 2.2.5, the number of transition events, μ , is Poisson-distributed as $\Delta t \rightarrow 0$,

$$\hat{P}_\mu = \frac{(k_2 N_A (N_A - 1) \Delta t)^\mu}{\mu!} e^{-k_2 N_A (N_A - 1) \Delta t} + o(\Delta t). \quad (2.77)$$

(Compare the above expression with equation 2.47.) The average number of transition events is $\langle \mu \rangle = k_2 N_A (N_A - 1) \Delta t = r_A \Delta t$, where r_A is the microscopic rate of a homologous 2nd order process.

Comparison of transition probability expressions

We summarize the expressions for the probability of the transition $2A \rightarrow X$ that we have derived in Table 2.2. Figure 2.7 shows a comparison of the dif-

$P_{2A \rightarrow X}$	Derived from:
$P_{can} = 1 - \frac{1}{N_A(t_n) k_2 \Delta t + 1}$	Integration of canonical microscopic rate law: $-\frac{dN_A}{dt} = k_2 N_A^2$.
$P_{int} = 1 - \frac{1}{N_A(t_n)(1 - e^{-k_2 \Delta t}) + e^{-k_2 \Delta t}}$	Integration of agent-based microscopic rate law: $-\frac{dN_A}{dt} = k_2 N_A (N_A - 1)$.
$P_{ber} = 1 - e^{-k_2 [N_A(t_n) - 1] \Delta t}$	Reaction events treated as Bernoulli trials. Molecular details are ignored.
$P_{dif} = k_2 [N_A(t_n) - 1] \Delta t$	Discretized agent-based microscopic rate law: $-\frac{\Delta N_A}{\Delta t} = k_2 N_A (N_A - 1)$.

Table 2.2 – Transition probability expressions for a 2nd order homologous process $2A \rightarrow X$.

ferent expressions for the transition probability as a function of the time step interval in the range $0 \leq \Delta t \leq 0.10$ sec. We graphed the results for two separate 2nd order rate constants differing by an order of magnitude ($k_2 = 1.0$ and 0.10 sec^{-1}), and for two snapshots of the population size at an arbitrary time step of the simulation, $N_A(t_n) = 10$ and 50 agents respectively. As expected, the probability curves converge as $\Delta t \rightarrow 0$. However, for fast reactions¹⁵ (first row of plots in Figure 2.7) occurring in larger populations, the linear approximation P_{dif} grows too rapidly and the different forms of the $P_{2A \rightarrow X}$ curves significantly deviate from each other at low Δt values. We conclude that for fast reactions P_{int} should be used for evaluating transition probabilities at each time step of the ABK algorithm. For slower reactions (e.g., $k_2 = 0.10 \text{ sec}^{-1}$, second row of plots in Figure 2.7) and low copy number populations, the curves become virtually indistinguishable (for instance, at $\Delta t = 1/100$ sec) and P_{dif} is the preferred approach because of its simplicity and savings in computational time that it affords.

The plots in 2.7 suggest that the choice of which of the four different expressions for the transition probability of $2A \rightarrow X$ to use depends on the population size, $N_A(t_n)$, and Δt . When the population size is small and Δt is chosen to be small, P_{dif} and P_{ber} are excellent approximations to P_{int} . In fact, the smaller that rate constant k_2 is the greater the range of Δt values and populations for which the use of P_{dif} and P_{ber} is warranted. These observations are illustrated in qualitative form in Figure 2.8 and provide a guideline for the proper choice of a transition probability expression.

¹⁵Note that the microscopic rate constant $k_2 = 1.0 \text{ sec}^{-1}$ represents a very fast 2nd order process. For a volume of $1 \mu\text{m}^3 = 10^{-15} \text{ L}$ (approximate volume of an *E. coli* cell), we can obtain using equation 2.10 the macroscopic rate constant: $k'_2 \approx 6 \times 10^8 \text{ sec}^{-1} \text{ M}^{-1}$. This is indeed a very fast process with a rate constant that is within the range of measured catalytic rates for diffusion-limited enzymes (BioNumbers ID 103809: <http://bionumbers.hms.harvard.edu/bionumber.aspx?id=103809&ver=3>; Milo et al., 2010).

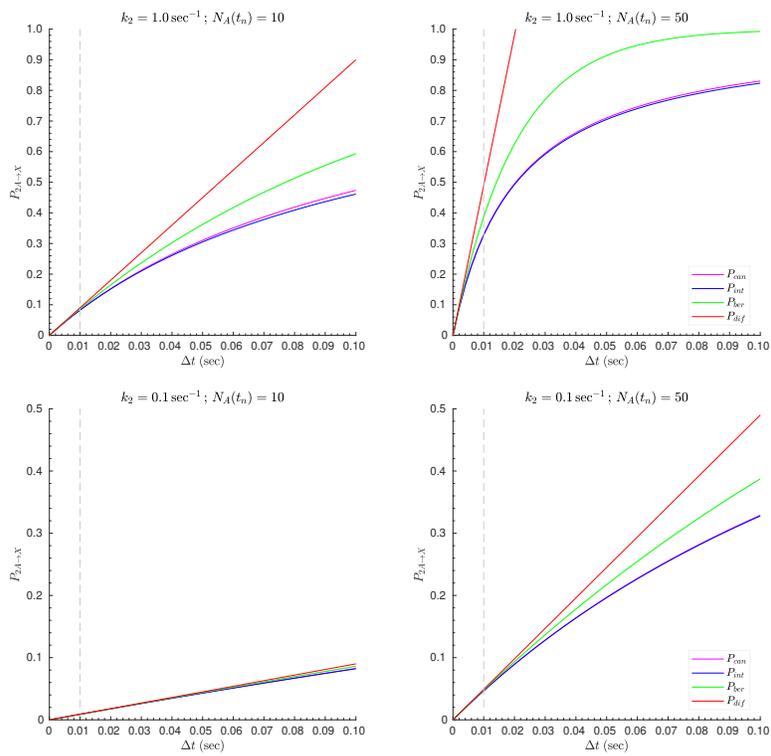


Figure 2.7 – Comparison of transition probabilities for a 2nd order process. The four derived expressions for the transition probability of $2A \rightarrow X$ are plotted as a function of the time interval Δt for two population sizes, $N_A(t_n) = 10$ and 50 agents of A . The vertical dashed gray line is drawn to highlight the difference between the curves at $\Delta t = 0.01$ sec. The first row of plots show the results for $k_2 = 1.0 \text{ sec}^{-1}$ (a very fast process), while the second row for $k_2 = 0.10 \text{ sec}^{-1}$. The legends apply to all four plots. Notice the difference in the scaling of the vertical axis in each row of the above tableau of plots.

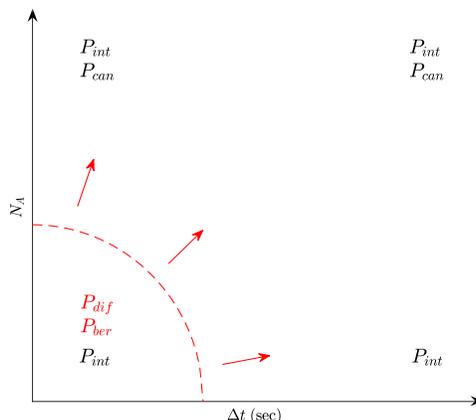


Figure 2.8 – Choice of transition probability expression. A qualitative summary of the proper choice for the transition probability expression in different regimes of population size, $N_A(t_n)$, and time step increment, Δt , used in the ABK algorithm. Given that P_{int} is the result of the agent-based treatment for a 2nd order process, it can be used throughout the diagram. P_{can} is best-suited for large populations, but P_{dif} and P_{ber} are good approximations to P_{int} for small population sizes and time step increments, and are preferred because of their simpler algebraic nature and resulting gains in computing speed. The red arrows indicate that the region where P_{dif} and P_{ber} can be used without significant loss of accuracy “expands” the slower the process is (i.e., as k_2 decreases).

ABK algorithm implementation

We hereby implement the ABK algorithm as described in subsection 2.2.6, except we modify step 3 to account for the fact that we are dealing with a 2nd order process. Essentially, the algorithm assesses if a given agent of A participates in a transition event and then *randomly chooses an interacting partner* (a second agent of A), justified by the fact that a selection of a reacting partner is required to reflect the stoichiometry of the process. We store the state of each member of the populations A and X in an array with a binary (0/1) value system. The algorithm then is:

1. Evaluate the transition probability $P_{2A \rightarrow X}$ for each agent of A present at time t_n using the appropriate probability expression (as shown in Figure 2.8 and the preceding discussion).
2. Draw a random number r_u from the standard uniform distribution in the interval $(0, 1)$.
3. Check the condition that $r_u < P_{2A \rightarrow X}$ for each agent of A . If the condition is true, then the state of that particular agent of A is changed to 0, a second agent of A is *randomly chosen* from the population and its state is changed to 0, while an agent of species X is created and its state is set to 1. If the condition is false then the agent states and population sizes do not

change.

Important note: One must be careful to take the second randomly chosen A agent out of the pool of available agents within a given time step. This ensures that an agent is not counted multiple times as part of transition events in the same time interval.

4. Sum up the number of A and X molecules in their respective populations after all probability conditions have been checked. This is the updated population size at the time point corresponding to the end of time interval Δt , or t_{n+1} .
5. Update the absolute time $t_{n+1} = t_n + \Delta t$.

We include a sample Matlab script to illustrate our programmatic implementation of the algorithm in Appendix D (p. 276).

We simulated the time course of a 2nd order process for homogeneous populations of different initial size, $N_{A,i}$, using the ABK algorithm. The results of simulations for populations consisting of thousands or hundreds of A agents agree exceedingly well with the time course predicted by the solutions to both the canonical and agent-based differential equations used to describe this process (equations 2.57 and 2.62). This is consistent with earlier arguments made in this section that for large populations the distinction between the two forms is not practically important. However, recall that the end-behavior of the solutions to the differential equations is different: in the canonical form (equation 2.59) the population size of A approaches 0, while in the agent-based form (equation 2.63) it approaches 1 regardless of the initial population size. We show in Figure 2.9 that our simulations using the P_{int} form¹⁶ of evaluating transition probabilities are consistent with this distinction becoming more pronounced for smaller population sizes. While for $N_{A,i} = 50$ the population size (averaged over 500 repetitions of the ABK algorithm; blue curve) closely follows the trajectory of the canonical form (red dotted line, $R^2 = 0.9995$), for even smaller populations with an odd initial number of agents the population size deviates from the prediction of the canonical form and approaches that of the agent-based form as t increases (Figure 2.9b). As noted previously, this is to be expected since each transition event consumes two agents of A , and therefore the final population will consist of a single individual if the initial number was odd. The effect is most pronounced for $N_{A,i} = 10$ ($R^2 = 0.8939$ for agreement with agent-based form; $R^2 = 0.9956$ for canonical form) where the difference between the agent-based deterministic and ABK time trajectories is clearly visible (Figure 2.9c).

It is more meaningful to compare $\langle N_A(t) \rangle$ obtained through ABK simulations with an average trajectory computed from the probabilistic formulation of chemical kinetics. Obtaining a solution for the CME describing a homologous 2nd order process is cumbersome due to the nonlinearity that the stoichiometry of the process imposes. Moreover, there is no closed-form solution for the moments of the time-dependent distribution of the molecular population size (see

¹⁶Similar results were obtained when using P_{can} , P_{ber} and P_{dif} to determine transition probabilities (data omitted in the interest of brevity).

Follow this [link](#) to see all of the code used for the simulations in this section.

section A.3). Given these difficulties in computing a solution to the CME, we instead use Gillespie’s algorithm (SSA; Appendix B) to obtain measures of the mean and standard deviation of the $N_A(t)$ trajectory.¹⁷ We remind the reader that the SSA provides a way to accurately sample the probability distribution that is predicted by the CME, therefore the two approaches can be considered to be equivalent. We show in Figure 2.9c that the ABK and SSA mean trajectories are virtually indistinguishable from each other ($R^2 = 0.9993$). Figure 2.9d shows a comparison of the relative standard deviation as observed through our simulations and the SSA. Once again, the agreement is excellent.

Finally, we plot the distribution of transition of events for two different initial population sizes $N_{A,i} = 50, 10$ (Figure 2.10) and compare them to the deterministic distribution calculated numerically using the solution to the agent-based microscopic rate law¹⁸ (equation 2.62). We numerically computed the deterministic distribution as follows: 1) the number of histogram bins defines a succession of time intervals; 2) the normalized difference between the population sizes of a species at the endpoints of each time interval provide the fraction of transition events within the specified time interval; 3) evaluation of these fractional values for all successive and non-overlapping time intervals yields the discretely-sampled PDF.

Comparison of the ABK data to the distribution obtained from the agent-based rate law is nearly perfect ($R^2 = 0.9999$) even for small population sizes of $N_{A,i} = 50$. In the case of low copy number populations, $N_{A,i} = 10$ is an example where the respective average time trajectories of the two approaches deviate from each other (see Figure 2.9c). Figure 2.10b shows that ABK predicts a greater fraction of transition events occurring in the first 10 sec of simulated time, while there is close agreement beyond this time point (overall $R^2 = 0.9963$). However, as we have discussed, it is more appropriate to compare the ABK method to the stochastic formulation of chemical kinetics. We have therefore computed the distribution of transition events obtained from the SSA average trajectory in the same way as described in the preceding paragraph. The agreement between the ABK and SSA distributions is excellent ($R^2 = 0.9989$ for $N_{A,i} = 10$).

¹⁷We wrote our own custom scripts to implement Gillespie’s algorithm. We also used the software StochSS to perform these simulations (Drawert et al., 2016). The two implementations gave virtually identical aggregate results.

¹⁸We have also checked our results against the PDF computed from the canonical form of the rate law and obtained a similarly good fit.

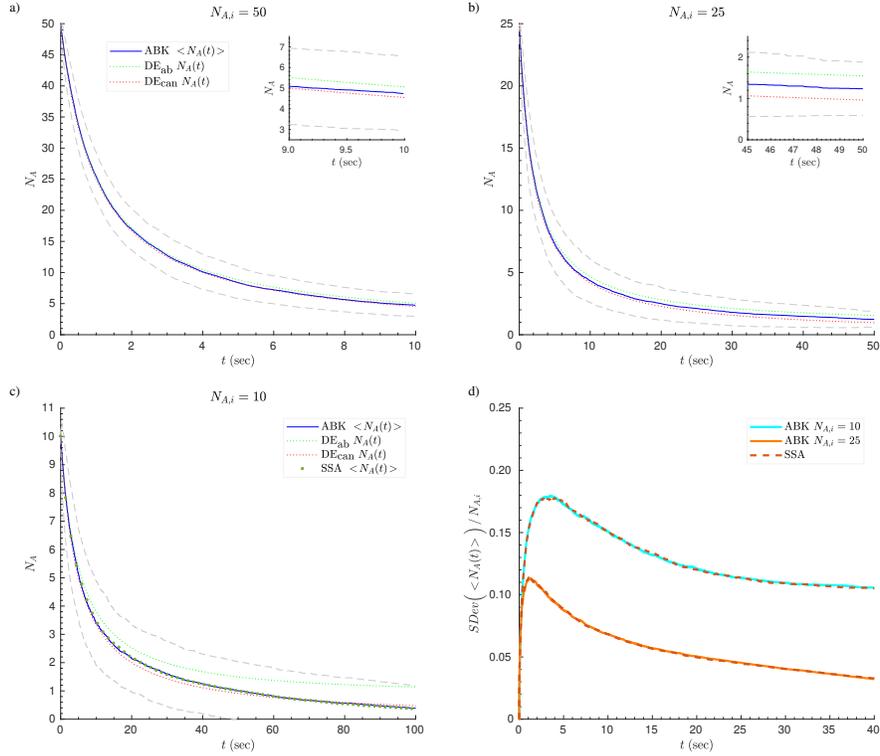


Figure 2.9 – Time course of $2A \rightarrow X$ process for populations of different initial size. a-c) For each of the initial populations of A , $N_{A,i} = 50, 25, 10$ agents, we simulated the time evolution for the process $2A \rightarrow X$ using the ABK algorithm. We used the following parameters and methods: fixed time step increments $\Delta t = 0.01$ sec, microscopic rate constant $k_2 = 0.01 \text{ sec}^{-1}$, and P_{int} for calculating transition probabilities (equation 2.66). We hereby plot the time dependence of the average population size of A ($n = 500$; blue curve) and its one-standard-deviation envelope (gray dashed curves). We also show the solutions to the agent-based and canonical differential equations (denoted by DE_{ab} and DE_{can} in the legend of plot a; green and red dotted curves, respectively) used to describe this process (equations 2.62 and 2.57). The inset plots in panels a, b focus on the end behavior of the simulation results compared to the solutions of the differential equations, and have the same aspect ratio between the t and $N_A(t)$ axes as their parent plots. Panel c ($N_{A,i} = 10$) also shows the average time trajectory obtained through Gillespie's algorithm (SSA; $n = 1000$). The agreement between the ABK and SSA mean trajectories is excellent ($R^2 = 0.9993$). Similar fits were obtained for $N_{A,i} = 50, 25$ (not shown). Note that the plots differ in the scaling of the axes. d) A comparison between the relative standard deviation values obtained through repetitions of the ABK simulation ($n = 5000$) and the SSA ($n = 5000$) for different initial population sizes, $N_{A,i} = 25, 10$, for the first 40 sec of simulated time ($R^2 = 0.9995, 0.9979$, respectively; $R^2 = 0.9923$ for the $N_{A,i} = 50$ curve, which is not shown).

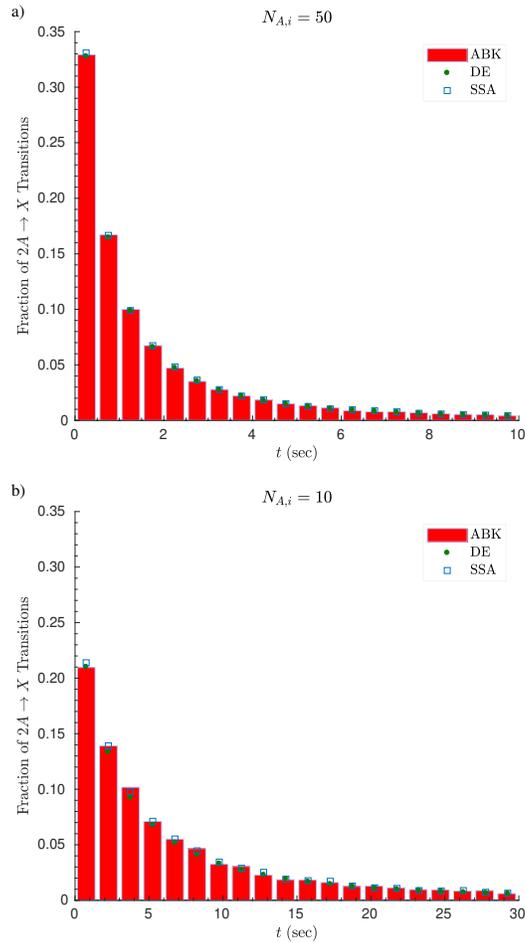


Figure 2.10 – Distribution of $2A \rightarrow X$ transition events. **a)** Distribution of transition events (shown as a histogram with 20 bins) for an ABK simulation (repeated 2000 times) with the indicated initial population size ($N_{A,i} = 50$) and the P_{ber} method of evaluating transition probabilities. The number of transitions in each time interval is normalized to $N_{A,i}$. We have superimposed the numerically-obtained distribution at selected time points (green circles) based on the agent-based microscopic rate law of this 2nd order process (equation 2.62; labeled “DE” in the legend) and the fraction of transitions obtained from an ensemble of SSA simulations ($n = 5000$; hollow blue squares). The agreement between these discretely-sampled PDFs with the ABK-obtained distribution is excellent ($R_{DE}^2 = 0.9999$, $R_{SSA}^2 = 0.9999$). **b)** Same as in panel a, except for the initial population size: $N_{A,i} = 10$ ($R_{DE}^2 = 0.9963$; $R_{SSA}^2 = 0.9989$). Note that the scale of the (horizontal) time axis is different in the two plots.

2.3.2 $A + B \rightarrow X$

Armed with the results of the analyses from the preceding sections on 1st and 2nd order homologous processes, we proceed more swiftly in our description of modeling heterologous processes using the ABK framework (although some modifications to the algorithm will be necessary to accommodate this scenario). We consider a system that consists of two species, A and B , capable of interacting with each other to produce X .

Transition probability from the integrated rate law

The microscopic rate law for this process with respect to A or B , r , is¹⁹

$$r = -\frac{dN_A}{dt} = -\frac{dN_B}{dt} = k_2 N_A N_B. \quad (2.78)$$

This is the canonical form of a 2nd order reaction rate law as presented in the study of chemical kinetics. The number of distinct pairs of interacting agents is $N_A N_B$, and consideration of the discrete nature of molecules/agents does not yield a distinct form of the rate law. Thus, there is only one expression for the differential rate law describing this process. We proceed to integrate this differential equation over the time interval $\Delta t = t_{n+1} - t_n$, obtaining

$$\frac{N_B(t_{n+1})/N_B(t_n)}{N_A(t_{n+1})/N_A(t_n)} = e^{k_2[N_B(t_n)-N_A(t_n)]\Delta t}. \quad (2.79)$$

Proceeding as before, we define the probability of no reaction happening in Δt as

$$P_{A+B \rightarrow X} \equiv \frac{N_A(t_{n+1})}{N_A(t_n)} = \frac{N_B(t_{n+1})}{N_B(t_n)}, \quad (2.80)$$

depending on which of the two reactant species we wish to evaluate an agent's transition probability of. The probability that a given agent of A or B , respectively, will react within Δt is

$$P_{\mathbf{A}+B \rightarrow X} = 1 - \frac{N_B(t_{n+1})}{N_B(t_n)} e^{k_2[N_A(t_n)-N_B(t_n)]\Delta t} \quad (2.81)$$

$$P_{A+\mathbf{B} \rightarrow X} = 1 - \frac{N_A(t_{n+1})}{N_A(t_n)} e^{k_2[N_B(t_n)-N_A(t_n)]\Delta t}, \quad (2.82)$$

where the bold designation of either A or B on the subscript of the left-hand side of the above equations indicates the reactant species that the expression applies to.

These transition probabilities are different from the ones we have previously encountered in that evaluation of the right side of equations 2.81 and 2.82 requires knowledge of either species population size at t_{n+1} . This makes use of

¹⁹The stoichiometric coefficients for this process being 1 ensures that this is the same as the species-independent rate $r = -\frac{dN_A}{dt} = -\frac{dN_B}{dt} = \frac{dN_X}{dt}$.

the ABK algorithm impossible since its purpose *is* to determine the population sizes at t_{n+1} .²⁰ We therefore abandon this approach and continue by using other methods to obtain suitable transition probability expressions.

P_{dif} from the differential rate law

We discretize the rate law (equation 2.78) to obtain

$$-\frac{\Delta N_A}{\Delta t} = -\frac{\Delta N_B}{\Delta t} \approx k_2 N_A N_B. \quad (2.83)$$

We then define the transition probabilities

$$P_{A+B \rightarrow X} \equiv -\frac{\Delta N_A}{N_A(t_n)} = P_{dif,A}, \quad (2.84)$$

and

$$P_{A+B \rightarrow X} \equiv -\frac{\Delta N_B}{N_B(t_n)} = P_{dif,B}. \quad (2.85)$$

This gives us the familiar form of P_{dif} evaluated with respect to a molecule/agent of A or B , respectively,

$$P_{dif,A} = k_2 N_B(t_n) \Delta t \quad (2.86)$$

$$P_{dif,B} = k_2 N_A(t_n) \Delta t. \quad (2.87)$$

Transition probability of a Bernoulli process

We choose to monitor the progress of the reaction by sampling the transition probability of the members of only one of the two species involved (A or B). From this perspective, if a molecule of A reacts then we randomly select a molecule of B from its population to satisfy the “rule” of the stated reaction

²⁰We attempted a modification of the algorithm to obtain an approximation for the time trajectory of the populations using the results of equations 2.81 and 2.82. In brief, given the initial population sizes at t_0 , we calculated the corresponding population sizes at t_1 (first iteration of the ABK algorithm) using the discretized differential form of transition probabilities (P_{dif} , described next in this section). Knowledge of the populations sizes at t_0 and t_1 served as a seed for using the probability expressions derived from the integrated rate law. Generally, we considered the time interval $\Delta t + \tau = t_{n+1} - t_n + \tau$, where the population sizes are known at times t_n and t_{n+1} . We then calculated the transition probabilities using the aforementioned equations *for the time interval* Δt , and reasoned that as $\tau \rightarrow 0$ it is an acceptable approximation to apply the calculated probabilities over the subsequent time interval τ to determine the result of the algorithm’s next iteration. The resulting time trajectories showed excellent agreement with the theoretically predicted curves (obtained by the analytic solution of the differential equation 2.78) for initial population sizes $N_{A,i}$ or $N_{B,i} > 50$. However, as we tried progressively smaller initial population sizes, the time trajectories (averaged over many repetitions of the algorithm) showed increased deviation from theory even when we considered smaller values of τ (such as $\tau = 1/5000$ sec; using even smaller values of τ surpassed the memory capacity of the computers we had available for running these simulations; results not shown). For this reason, and because there is no compelling theoretical basis for using this approximation in the first place, we abandon the use of the integrated rate law to model the time evolution of heterologous 2nd order processes.

($A + B \rightarrow X$). Alternatively, we can do the same from the perspective of the members of population B : we calculate the probability that each molecule of B has of reacting, and then randomly choose a molecule of A as its reacting partner if a reaction occurs within Δt .

We follow the same logic as in the previous section on homologous 2nd order processes, and consider the population-wide probability $\hat{\pi}_A$ of a reaction occurring to a member of the population of A within a time interval of length $\Delta t/s$, and the per-agent probability p_A in the same interval. Therefore, we have

$$r\Delta t = s\hat{\pi}_A. \quad (2.88)$$

Given the expression for r (equation 2.78), and assuming that a reaction event for an A molecule is independent of other such reactions in the population, then $\hat{\pi}_A = N_A p_A$, and equation 2.88 becomes

$$\begin{aligned} k_2 N_A N_B \Delta t &= s p_A N_A \\ p_A &= \frac{k_2 N_B \Delta t}{s}. \end{aligned} \quad (2.89)$$

As before, taking the limit as $s \rightarrow \infty$, we arrive at the expression for the probability that a molecule of A reacts within time interval $\Delta t = t_{n+1} - t_n$,

$$P_{A+B \rightarrow X} = 1 - e^{-k_2 N_B \Delta t}. \quad (2.90)$$

Similarly, we derive the corresponding expression for when an agent of B is sampled,

$$P_{A+B \rightarrow X} = 1 - e^{-k_2 N_A \Delta t}, \quad (2.91)$$

where N_A and N_B refer to the population sizes at time t_n (elsewhere denoted as $N_A(t_n)$ and $N_B(t_n)$, respectively). We will refer to these expressions as $P_{ber,A}$ and $P_{ber,B}$ to indicate the Bernoulli interpretation of reaction events and the species they specifically apply to. As in the case of a homologous 2nd order process, we stress that the probability expressions we have derived here are an approximation valid for small Δt .

Using the above result for $P_{ber,B}$ (or $P_{ber,A}$), the reader can verify that the same arguments presented in subsection 2.2.5 result in the number of transition events, μ , being Poisson-distributed as $\Delta t \rightarrow 0$,

$$\hat{P}_\mu = \frac{(k_2 N_A N_B \Delta t)^\mu}{\mu!} e^{-k_2 N_A N_B \Delta t} + o(\Delta t). \quad (2.92)$$

The average number of transition events is therefore $\langle \mu \rangle = k_2 N_A N_B \Delta t = r\Delta t$, where r is the microscopic rate of a heterologous 2nd order process (equation 2.78).

Summary of transition probability expressions

We summarize the expressions we have derived for the probability of the transition $A + B \rightarrow X$ in Table 2.3. All of our previous comments on the differences

$P_{A+B \rightarrow X}$	Derived from:
$P_{ber,A} = 1 - e^{-k_2 N_B(t_n) \Delta t}$	Reaction events treated as Bernoulli trials.
$P_{ber,B} = 1 - e^{-k_2 N_A(t_n) \Delta t}$	Molecular details are ignored.
$P_{dif,A} = k_2 N_B(t_n) \Delta t$	Discretized agent-based microscopic rate law: $-\frac{\Delta N_A}{\Delta t} = -\frac{\Delta N_B}{\Delta t} = k_2 N_A N_B.$
$P_{dif,B} = k_2 N_A(t_n) \Delta t$	

Table 2.3 – Transition probability expressions for a 2nd order heterologous process $A + B \rightarrow X$.

between P_{ber} and P_{dif} still apply. Most notably, P_{dif} increases without bound so care must be taken to ensure that it does not exceed 1. On any given time step of the simulation, the following condition must be satisfied,

$$\Delta t_{max} = \frac{1}{k_2 N(t_n)}, \quad (2.93)$$

where $N(t_n)$ represents either the population size of A or B at time t_n , depending on which P_{dif} expression is used.

ABK algorithm implementation

We will implement the ABK algorithm as described for homologous 2nd order processes, except that it is modified to account for the fact that we are dealing with a heterologous process. The algorithm can be run with respect to either population A or B ; that is, the probability of each member of the chosen population participating in a transition event with a member of the other species is evaluated. Our results indicate that identical results are obtained when using the algorithm for either choice of population, although our preference is to use the smaller, or limiting, species population because it reduces computational runtime.

We present an outline of the algorithm when choosing population B to monitor the time evolution of the system (with the understanding that the alternate choice requires the swapping of A and B in steps 1, 3):

1. Evaluate the transition probability $P_{A+B \rightarrow X}$ for each B agent present at time t_n using the appropriate probability expression (Table 2.3).
2. Draw a random number r_u from the standard uniform distribution in the interval $(0, 1)$.
3. Check the condition that $r_u < P_{A+B \rightarrow X}$ for each B agent. If the condition is true, then the state of that particular B agent is changed to 0, an A agent is randomly chosen from the population and its state is also changed to 0, while an X agent is created and its state is set to 1. If the condition is false then the populations do not change.

4. Sum up the number of A , B and X molecules in their respective populations after all probability conditions have been checked. This is the updated population size at the time point corresponding to the end of time interval Δt , or t_{n+1} .
5. Update the absolute time $t_{n+1} = t_n + \Delta t$.

We include a sample Matlab script to illustrate our programmatic implementation of the algorithm in Appendix D (p. 278).

We tested the algorithm for populations of different sizes, choice of transition probability expression (P_{ber} or P_{dif}), and choice of species population with respect to which transition probabilities are evaluated (i.e., A or B). For simulations with $k_2 = 0.01 \text{ sec}^{-1}$ and fixed time steps of $\Delta t = 0.01 \text{ sec}$, our results indicate that for larger populations ($N_{A,i}$ and $N_{B,i} > 100$) the agreement with the solution to the differential equation describing this system is virtually perfect (data not shown) using either P_{ber} or P_{dif} expressions, and regardless of which species we chose to run the algorithm with respect to. Figure 2.11a,b shows the results for two examples of smaller populations where $N_{B,i} = 0.7 N_{A,i}$ (i.e., species B is limiting). Notice that in the plot for $N_{A,i} = 10$ and $N_{B,i} = 7$, there is a small but noticeable deviation from the solution to the differential equation (approximately 1 – 2% difference in the neighborhood of $t = 30 \text{ sec}$).

The solution to the CME for a 2nd order heterologous process can be obtained analytically (Lecca et al., 2013b; see Appendix A.4, equation A.81), and is a time-dependent probability distribution of the population size that has no direct analogue to the ones typically encountered in science. We have calculated the first moment (mean) of the distribution and found that it also deviates from the deterministic prediction.²¹ In fact, the average trajectories from the ABK simulations agree nearly perfectly ($R^2 = 0.9997$) with those predicted from the solution to the CME (Figure 2.11b).²² We observed a similar deviation in the regime of small population sizes for homologous 2nd order processes, and attribute it to the discrete nature of the population’s makeup becoming more pronounced during the act of averaging over low copy number population sizes. We also calculated the central second moment (variance) and standard deviation of the distribution predicted by the CME and compared it to the standard deviation of the average time courses obtained through our simulations (Figure 2.11c,d). The agreement is once again excellent ($R^2 > 0.996$).

We also show the time-dependent distribution of transition events (normalized to $N_{B,i}$; shown as red bars in Figure 2.12) and compare it to the deterministic distribution for a 2nd order heterologous process (green circles). We also show the corresponding distribution obtained from the CME-predicted average

²¹The small difference between the deterministic time trajectories and those obtained from the CME solution has been reported before (Ishida, 1964). The explanation for this discrepancy is the presence of a covariance of N_A and N_B term in the differential equation for $d\langle N_B \rangle / dt$ (Lecca et al., 2013b, equation 2.114). Thankfully, this term is typically negligible.

²²A far less significant in magnitude difference in the average time trajectories was observed in the case of $N_{A,i} = 50$, $N_{B,i} = 35$, $N_{X,i} = 0$. Accordingly, we do not show the CME average trajectory in Figure 2.11a.

Follow this [link](#) to see all of the code used for the simulations in this section.

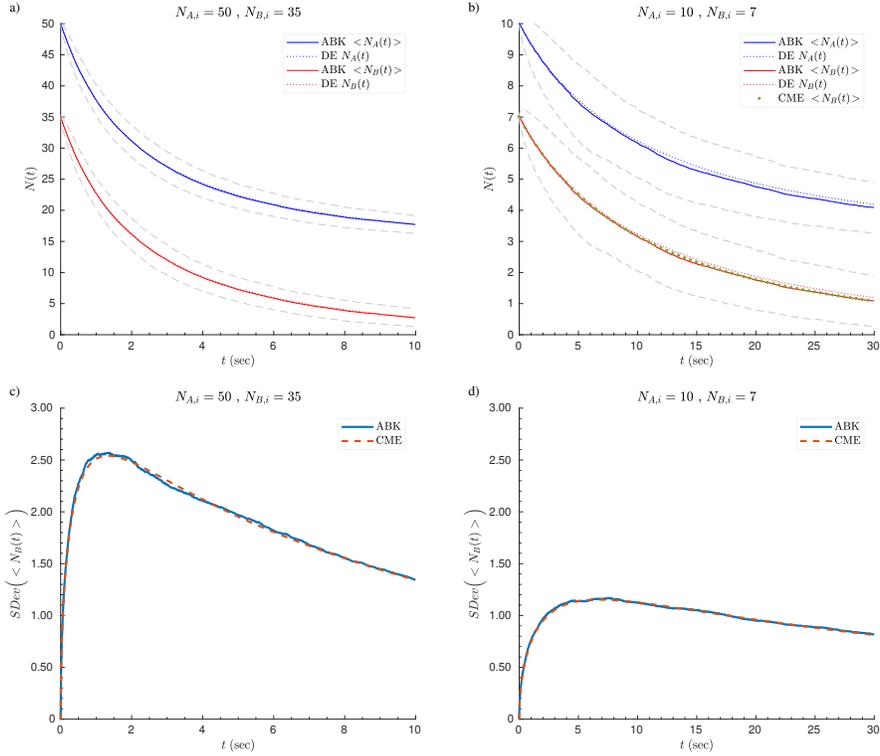


Figure 2.11 – Time course of $A + B \rightarrow X$ process for populations of different initial size. We simulated the time evolution of the heterologous 2^{nd} order process $A + B \rightarrow X$ using the ABK algorithm for two sets of initial number of A and B agents/molecules. In both cases, $\Delta t = 0.01 \text{ sec}$ (fixed time step increment), $k_2 = 0.01 \text{ sec}^{-1}$. **a)** $N_{A,i} = 50, N_{B,i} = 35, N_{X,i} = 0$. We performed simulations with respect to the members of population A using the P_{dif} form of transition probabilities. Average population sizes ($n = 500$) are depicted by solid curves: blue for species A and red for species B . The standard deviations were also calculated at each time step (dashed gray curves denote the one-standard-deviation envelope). The dotted curves correspond to the solution of the microscopic differential equation (DE) describing this process. The simulation agrees with the DE time trajectories ($R^2 = 0.9998$). The time evolution of species X is not shown for clarity. **b)** $N_{A,i} = 10, N_{B,i} = 7, N_{X,i} = 0$. In this case we performed simulations with respect to the members of population B using the P_{ber} form of transition probabilities. We also show the average time trajectory $\langle N_B(t) \rangle$ obtained through the solution to the CME (green filled circles). The CME trajectory is slightly different than the one predicted by the DE, and the average ABK trajectory ($n = 500$) more closely matches the CME's ($R^2 = 0.9997$). **c)** Comparison of the standard deviation obtained through repetitions of our simulations ($n = 5000$) and that predicted by the solution to the CME ($R^2 = 0.9970$) for the indicated initial population sizes. **d)** Same as in panel c, but with different initial population sizes ($n = 5000$; $R^2 = 0.9964$). Note that the horizontal time-axes are scaled differently in each column of panels.

population trajectories. We numerically computed the respective distributions in the same way as described in the preceding subsection (on homologous 2nd order processes). We repeat it here for the reader's convenience: 1) the number of histogram bins defines a succession of time intervals; 2) the normalized difference between the population levels of a species at the endpoints of each time interval provide the expected fraction of transition events within the specified time interval; 3) evaluation of these fractional values for all successive and non-overlapping time intervals yields the discretely-sampled PDF. The agreement between the results of the ABK simulation and the theoretically-predicted distributions is excellent ($R^2 > 0.996$).

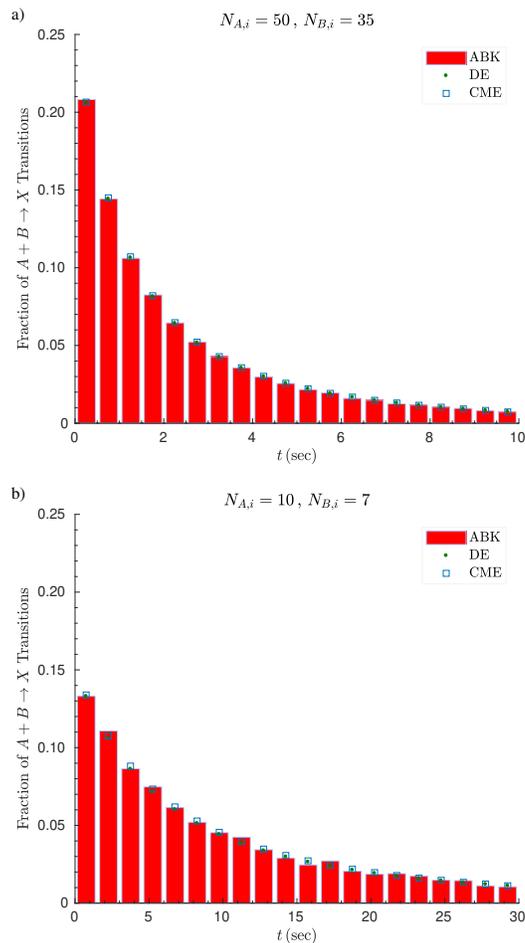


Figure 2.12 – Distribution of $A + B \rightarrow X$ transition events. We repeated the ABK simulation 2000 times (same parameters as in Figure 2.11) to obtain the histogram (20 bins) of the number of transitions in each time interval (normalized to $N_{B,i}$). **a)** Distribution of transition events for a simulation with the initial condition $N_{A,i} = 50, N_{B,i} = 35, N_{X,i} = 0$. We have superimposed the numerically-obtained distribution for the 2nd order deterministic process (labeled “DE” in the legend; green circles). We also show the distribution obtained from the CME (hollow blue squares). A comparison of these distributions with the one obtained through ABK simulations is excellent ($R_{DE}^2 = 0.9999, R_{CME}^2 = 0.9999$). **b)** Distribution of transition events for the set of initial population sizes $N_{A,i} = 10, N_{B,i} = 7, N_{X,i} = 0$. The agreement of the ABK and the theoretical distributions is once again excellent ($R_{DE}^2 = 0.9966, R_{CME}^2 = 0.9978$). Note that the horizontal time-axes are scaled differently in the two plots.

2.4 3rd Order Processes

Chemical reactions obeying 3rd order kinetics occur extremely rarely and it is thought that there are no 3rd order elementary processes. However, we briefly present the ABK implementation of such processes to demonstrate the generality of the hitherto derived methods and how they can be extended to potential higher order processes in non-chemical systems.

In keeping with the nomenclature we used when dealing with 2nd order processes, we will concern ourselves with the homologous elementary process



and the heterologous ones



with microscopic rate constant k_3 (units: sec^{-1}). X does not necessarily describe the formation of a single product or complex, but is meant to represent all possible products that can result from a considered reaction.

2.4.1 $3A \rightarrow X$

Transition probability from the canonical rate law

The microscopic rate law with respect to A , r_A , for the homologous process $3A \rightarrow X$ is²³

$$r_A = -\frac{dN_A}{dt} = k_3 N_A^3. \quad (2.96)$$

Upon integration over the time interval $\Delta t = t_{n+1} - t_n$ (where the variable n represents the n^{th} time step of the ABK algorithm) gives

$$\frac{N_A(t_{n+1})}{N_A(t_n)} = (2k_3 N_A(t_n)^2 \Delta t + 1)^{-1/2}. \quad (2.97)$$

Note that upon integration over a sufficiently large enough time interval Δt (not for the purpose of using the ABK algorithm, but for investigating the end-behavior of this solution) we find that $N_A \rightarrow 0$, as expected for large populations. The initial number of A agents will obviously affect the final number of untransformed/unreacted agents in a process where discrete members of a population interact (for instance, only when $N_{A,i}$ is a multiple of 3 will the population of A be completely depleted). This is another example of how differential calculus is not well-suited to describing low copy number populations.

²³As opposed to the species-independent rate law, $r = -\frac{1}{3} \frac{dN_A}{dt} = \frac{dN_x}{dt}$.

Given our standard definition of the probability that no reaction occurs in Δt , we obtain

$$P_{3A \rightarrow X} \equiv 1 - \frac{N_A(t_{n+1})}{N_A(t_n)} \quad (2.98)$$

$$P_{3A \rightarrow X} = 1 - (2k_3 N_A(t_n)^2 \Delta t + 1)^{-1/2}. \quad (2.99)$$

We will refer to this transition probability as P_{can} in reference to the fact that it was derived from the canonical rate law for this process.

The agent-based interpretation of the rate law

The canonical rate law presented above applies best to large populations. Considering the population as consisting of discrete agents suggests that the rate law should be restated as

$$r_A = -\frac{dN_A}{dt} = k_3 N_A (N_A - 1) (N_A - 2), \quad (2.100)$$

which can be integrated over time interval $\Delta t = t_{n+1} - t_n$ to yield the following expression,

$$\frac{\sqrt{N_A(t_{n+1}) [N_A(t_{n+1}) - 2]}}{N_A(t_{n+1}) - 1} = \frac{\sqrt{N_A(t_n) [N_A(t_n) - 2]}}{N_A(t_n) - 1} e^{-k_3 \Delta t}. \quad (2.101)$$

There is no way to rewrite this result through algebraic manipulation in the form $\frac{N_A(t_{n+1})}{N_A(t_n)} = f(N_A(t_n), k_3, \Delta t)$, where f is a function of the indicated variables. Therefore, we cannot use the integrated rate law to derive a suitable transition probability expression. Notice, however, the end behavior of this expression: as $\Delta t \rightarrow \infty$, $N_A \rightarrow 0$ or $N_A \rightarrow 2$ (this can be easily seen by examining the numerator of the left-hand side in equation 2.101). This observation again highlights the limitation in using differential equations to describe small populations of discrete agents, as the final number of remaining A agents depends on the size of the initial population.

Other transition probability expressions and summary

We proceed through the same derivation of treating reaction events as Bernoulli trials and derive the following familiar expression:

$$P_{ber} = 1 - e^{-k_3 [N_A(t_n) - 1] [N_A(t_n) - 2] \Delta t}. \quad (2.102)$$

Also, by discretizing the agent-based form of the rate law (equation 2.100), we determine

$$P_{dif} = k_3 [N_A(t_n) - 1] [N_A(t_n) - 2] \Delta t. \quad (2.103)$$

We summarize our findings for transition probability expressions that describe homologous 3rd order processes in Table 2.4. We have implemented the ABK algorithm as previously described and have obtained results similar to the ones shown for homologous 2nd order processes (data not shown).

Follow this [link](#) to see the Matlab code.

$P_{3A \rightarrow X}$	Derived from:
$P_{can} = 1 - (2k_3 N_A(t_n)^2 \Delta t + 1)^{-1/2}$	Integration of canonical microscopic rate law: $-\frac{dN_A}{dt} = k_3 N_A^3$.
$P_{ber} = 1 - e^{-k_3 [N_A(t_n) - 1] [N_A(t_n) - 2] \Delta t}$	Reaction events treated as Bernoulli trials.
$P_{dif} = k_3 [N_A(t_n) - 1] [N_A(t_n) - 2] \Delta t$	Discretized agent-based microscopic rate law: $-\frac{\Delta N_A}{\Delta t} = k_3 N_A (N_A - 1) (N_A - 2)$.

Table 2.4 – Transition probability expressions for a 3rd order homologous process $3A \rightarrow X$.

2.4.2 $2A + B \rightarrow X$

We now consider the heterologous 3rd order processes consisting of two interacting species, A and B , at a stoichiometry of $2 : 1$. We focus on this process with the understanding that our results will apply to the case $A + 2B \rightarrow X$ by a simple exchange of A and B in the expressions for the transition probability.

The canonical rate law for this process is

$$r_A = -\frac{dN_A}{dt} = k_3 N_A^2 N_B, \quad (2.104)$$

which applies to large populations of A and B . We are again specifically referring to the species-dependent rate law, r_A . The corresponding rate law r_B has the same algebraic form, even though r_A and r_B are not equal to each other and are related by a statistical factor based on the stoichiometry (in this case, $r_A = 2r_B$).

The agent-based interpretation of this process leads the following form of the rate law:

$$r_A = -\frac{dN_A}{dt} = k_3 N_A (N_A - 1) N_B. \quad (2.105)$$

As we saw in the preceding subsection, integrating the rate law of higher order processes can yield complicated expressions that cannot always be rearranged to fit the definition of the probability of no reaction $P_{no} \equiv N_A(t_{n+1})/N_A(t_n)$. We therefore focus on the Bernoulli trial interpretation of reaction events and the discretized version of the differential rate law to derive expressions for the transition probability (P_{ber} and P_{dif} , respectively), which, although they are approximations, are amenable to easy analysis and programmatic implementation.

As in the case of heterologous 2nd order processes, the implementation of the ABK algorithm requires a choice to be made with respect to which species population the per-agent transition probability is evaluated. We summarize the probability expressions in Table 2.5.

Follow this [link](#) to see the Matlab code.

$P_{2A+B \rightarrow X}$	Derived from:
$P_{ber,A} = 1 - e^{-k_3[N_A(t_n)-1]N_B(t_n)\Delta t}$ $P_{ber,B} = 1 - e^{-k_3N_A(t_n)[N_A(t_n)-1]\Delta t}$	Reaction events treated as Bernoulli trials.
$P_{dif,A} = k_3 [N_A(t_n) - 1] N_B(t_n) \Delta t$ $P_{dif,B} = k_3 N_A(t_n) [N_A(t_n) - 1] \Delta t$	Discretized agent-based microscopic rate law: $-\frac{\Delta N_A}{\Delta t} = k_2 N_A [N_A(t_n) - 1] N_B.$

Table 2.5 – Transition probability expressions for a 3rd order heterologous process $2A + B \rightarrow X$.

2.4.3 $A + B + C \rightarrow X$

The only other possible heterologous 3rd order process consists of three species of agents interacting with stoichiometry 1 : 1 : 1. We merely show in Table 2.6 the transition probability expressions using the Bernoulli trial interpretation and the discretized differential rate law. We do not show plots highlighting the agreement between the ABK algorithm and the solution to the differential equation describing this process in the interest of brevity. We note that we obtained identical results regardless of which of the three reactant species we chose to run the ABK algorithm with respect to.

Follow this [link](#) to see the Matlab code.

$P_{A+B+C \rightarrow X}$	Derived from:
$P_{ber,A} = 1 - e^{-k_3 N_B(t_n) N_C(t_n) \Delta t}$ $P_{ber,B} = 1 - e^{-k_3 N_A(t_n) N_C(t_n) \Delta t}$ $P_{ber,C} = 1 - e^{-k_3 N_A(t_n) N_B(t_n) \Delta t}$	Reaction events treated as Bernoulli trials.
$P_{dif,A} = k_3 N_B(t_n) N_C(t_n) \Delta t$ $P_{dif,B} = k_3 N_A(t_n) N_C(t_n) \Delta t$ $P_{dif,C} = k_3 N_A(t_n) N_B(t_n) \Delta t$	Discretized agent-based microscopic rate law: $-\frac{\Delta N_A}{\Delta t} = k_3 N_A N_B N_C.$

Table 2.6 – Transition probability expressions for a 3rd order heterologous process $A + B + C \rightarrow X$. We remind the reader of the equality $-\frac{\Delta N_A}{\Delta t} = -\frac{\Delta N_B}{\Delta t} = -\frac{\Delta N_C}{\Delta t}$ which follows from the equivalence between the corresponding differential rates.

2.5 0th Order Processes

We have not yet addressed 0th order processes because they are fundamentally different from the others we have analyzed: they do *not* depend on the presence of any agents such that a reaction event occurs (alternatively stated, they are not agent-based processes). This is obviously not true for real physical processes where there is always a 1st or higher order dependence on the participating species (reactants in the case of chemical reactions). However, real processes may *appear* to kinetically behave as 0th order under saturating conditions. Examples are a reaction that depends on the presence of an inorganic catalyst whose surface is saturated with reactant molecules, and biological catalysts obeying Michaelis-Menten kinetics operating near v_{max} . Transcription and translation are particularly important examples of biological processes that can proceed (and be modeled) through 0th order kinetics.

Here, we define a probability expression that allows us to stochastically model such processes. We use the notation



where the symbol \emptyset on the left-hand side represents the lack of dependence of the reaction rate on any reactants.²⁴ As such, it is equivalent to a “birth” process where X is produced at a constant rate,

$$r = \frac{dN_X}{dt} = k_0, \quad (2.107)$$

with k_0 being the 0th order microscopic rate constant (units: sec^{-1}).

It initially seems that since the rate does not depend on the population of any species, the ABK algorithm cannot be used to assess the time evolution of the process. That is, there is no probability condition to be checked *for each* agent since no reactant species exists in the model to begin with. However, we can still stochastically simulate this process by defining the probability that a “birth” event will take place within time interval Δt as

$$P_{\emptyset \rightarrow X} \equiv k_0 \Delta t. \quad (2.108)$$

As long as Δt is chosen so that $P_{\emptyset \rightarrow X} < 1$, the above expression accurately predicts the growth of population X . This definition is motivated by the observation that k_0 is the frequency of production of X (agents of X per unit of time) and can be scaled by the time interval we monitor the reaction progress for, Δt . This also agrees with the integrated rate law for this process, $N_X(t) = N_{X,i} + k_0 t$, for the time interval $[0, t]$. Rewriting this expression for the general time interval Δt , we get $\Delta N_X = k_0 \Delta t$. As long as we choose a small enough Δt such that, on

²⁴At first glance, the process 2.106 implies that an agent/molecule of X is produced out of nothing, thereby violating the Law of Conservation of Mass. What we mean here is that the synthesis of X *appears* to follow 0th order kinetics and the population sizes of the species involved in the reaction (the reactants) remain unchanged. We therefore omit them for simplicity and use \emptyset to denote them.

average, $\Delta N_X < 1$, then the above definition of the probability of a 0th order reaction is justified.

Equation 2.108 has the same algebraic form as the transition probability for a 1st order reaction (equation 2.25). However, there is an important difference in how the expression is used by the ABK algorithm for 0th order processes: the value of $P_{\emptyset \rightarrow X}$ is checked against a random number (drawn from the uniform distribution) just *once* in each time interval Δt (whereas $P_{A \rightarrow X}$ is checked for each A agent undergoing a 1st order process). Therefore, either the population size of X remains the same or it increases by *one* agent at the end of Δt .

ABK algorithm implementation

We hereby state how the algorithm is modified to account for the kinetics of a 0th order process: the following steps are executed just *once* in each time step of duration Δt . As we have done in previous sections, we define $\Delta t = t_{n+1} - t_n$, where n represents the n^{th} time step of the ABK algorithm.

Follow this [link](#) to see the Matlab code.

1. Evaluate the probability that a “birth” event occurs in Δt , $P_{\emptyset \rightarrow X}$ (as defined in equation 2.108).
2. Draw a random number r_u from the standard uniform distribution in the interval $(0, 1)$.
3. Check the condition that $r_u < P_{\emptyset \rightarrow X}$. If the condition is true, then a single agent X is created or “born” in the time interval Δt . If the condition is false then there is no change in the population size.
4. We update the population size of X at the end of time interval Δt , or t_{n+1} (the population has either increased by 1 or is the same as at time t_n).
5. Update the absolute time $t_{n+1} = t_n + \Delta t$.

Figure 2.13a shows the results we obtained when testing this algorithm for the simulation of a 0th order process. There is near-perfect agreement between the ensemble of ABK simulation runs and the deterministic or CME prediction for the average time trajectory of species X . We remind the reader that the deterministic and CME formalisms have the same form for the differential equation describing this process (equations 2.107 and A.48 in Appendix A.2). Finally, we compared the magnitude of the standard deviation from our simulations to that predicted by the CME (equation A.69) and again obtained excellent agreement (Figure 2.13b).

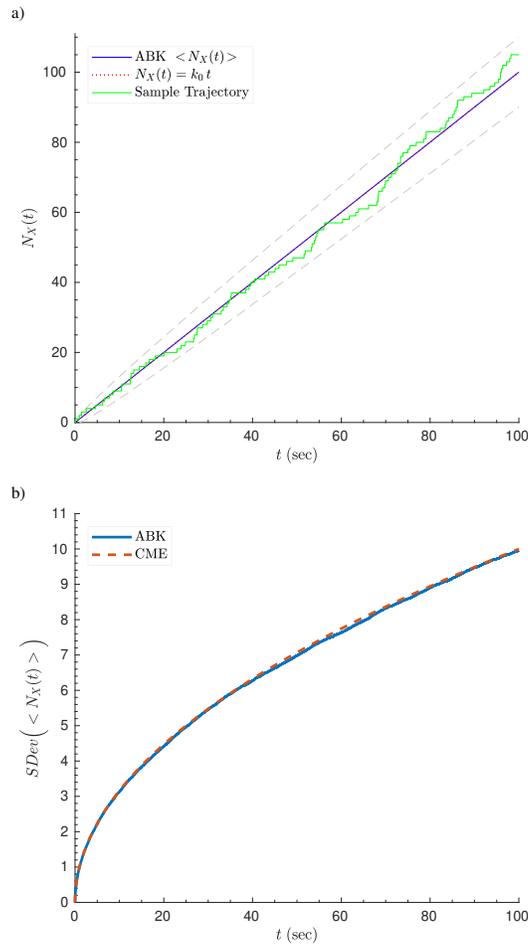


Figure 2.13 – Time course of 0th order process. Simulated time course for a 0th order process with $k_0 = 1 \text{ sec}^{-1}$, fixed time step increments $\Delta t = 0.01 \text{ sec}$, and $N_{X,i} = 0$. **a)** The green curve is a sample time trajectory of population X from a single simulation run. The blue curve is the average population size obtained from an ensemble of ABK simulations ($n = 10000$), and the gray dashed curves delineate the one-standard-deviation envelope. The solution to the differential equation describing this process is shown as a red dotted line (it agrees closely with the average trajectory obtained from the ABK simulation; $R^2 = 0.9999$). Note that the solution to the CME for $\langle N_X(t) \rangle$ is the same as the deterministic one. **b)** Comparison of the standard deviation in the population size of X obtained through an ensemble of ABK simulations ($n = 10000$; blue curve) to the CME prediction (equation A.69; red dashed curve). The agreement is excellent ($R^2 = 0.9994$).

2.6 Generalization of the ABK Algorithm

2.6.1 Transition Probability Expressions

We consider a process with the number of reactant species denoted by ξ . If a particular reactant species is R_a , where the subscript a is an integer in $[1, \xi]$, the microscopic agent-based differential rate law with respect to this species is

$$-\frac{dN_{R_a}}{dt} = k \prod_{i=1}^{\xi} f_i(N_{R_i}, m_i). \quad (2.109)$$

N_{R_i} represents the number of molecules/agents of R_i , m_i is a nonnegative integer representing the order (or multiplicity), of the process with respect to reactant R_i , and k is the microscopic rate constant for the process with overall order $M = \sum m_i$.²⁵ The function f_i depends on the number of agents that the population of each reacting species consists of, and we define it as

$$f_i = \begin{cases} \prod_{j=1}^{m_i} [N_{R_i} - (m_i - j)] & , \quad m_i \geq 1 \\ 1 & , \quad m_i = 0, \end{cases} \quad (2.110)$$

where the second part of this piecewise function deals with 0th order processes. For example, a process that depends on a species R_i with molecularity $m_i = 2$, would result in $f_i = N_{R_i}(N_{R_i} - 1)$ according to the above definition. This is consistent with the agent-based interpretation of processes that we have used in the preceding sections (e.g., equation 2.62).

We then define the per-agent *kinetic complementarity factor* (KCF), Ω , for a process of overall order $M \geq 1$ and with respect to reactant R_a as follows:

$$\Omega_{R_a} \equiv \frac{-dN_{R_a}/dt}{N_{R_a}}. \quad (2.111)$$

The precise form of the expression for Ω_{R_a} depends on which reactant we choose to monitor the reaction progress by. Ω_{R_a} represents the number of possible interactions that can lead to a reaction event *for a given agent* belonging to the population R_a , scaled by the value of the microscopic kinetic constant k . The intuition for the above definition lies in the fact that Ω ‘‘complements’’ the population size of R_a in the microscopic form of the differential equation,

$$-\frac{dN_{R_a}}{dt} = N_{R_a} \Omega_{R_a}. \quad (2.112)$$

For instance, in 1st order processes an agent undergoes a transition independently of others in the population, so $\Omega = k$. This is the only case where Ω

²⁵Note that the use of the subscripts a and i is meaningful here: R_a is a particular reactant species that equation 2.109 applies to, whereas R_i on the right-hand side of 2.109 is indexed over all reactant species.

does not include a time-dependent population size of a species. For the 2nd order process, $A + B \rightarrow X$, if we monitor the process with respect to A then $\Omega_A = k N_B(t)$; considering it with respect to B gives $\Omega_B = k N_A(t)$.

Therefore, given our treatment of 1st and higher order processes as Bernoulli trials, the general form for the transition probability of a process with respect to reactant R_a is

$$P_{ber,R_a} = 1 - e^{-\Omega_{R_a} \Delta t}. \quad (2.113)$$

Similarly, the discretized form of the differential rate law yields a probability expression of the form

$$P_{dif,R_a} = \Omega_{R_a} \Delta t. \quad (2.114)$$

We make two important comments on the limitations of obtaining numerical values for P_{ber} and P_{dif} , respectively:

1. The exponent in equation 2.113 must be small enough so that changes in its value give *distinct* values for the probability P_{ber,R_a} (which has an upper asymptotic limit of 1). With respect to numerical evaluation, we note the heuristic condition,

$$\Omega_{R_a} \Delta t \lesssim 35 \ni \text{num}(P_{ber,R_a}) < 1, \quad (2.115)$$

where $\text{num}(P_{ber,R_a})$ is the numerical value of the transition probability in double-precision format (i.e., with 15 decimal figures). That is, values of $\Omega_{R_a} \Delta t$ greater than 35 give probability values so close to 1 that they are beyond the double-precision level of accuracy and the computed P_{ber,R_a} values are stored in computer memory as 1.

This condition is useful in deciding how small Δt should/can be or if there is a maximum population size that can be effectively modeled using the ABK method. For instance, in a 2nd order process $A + B \rightarrow X$ with $k = 0.01 \text{ sec}^{-1}$ and $\Delta t = 0.01 \text{ sec}$, $P_{ber,A} = 1 - e^{-k N_B \Delta t}$ and the maximum population size of B (or A when using $P_{ber,B}$) is 3.5×10^5 agents in order to satisfy condition 2.115. Practically, this limitation is not of major concern because the ABK algorithm is most useful and computationally efficient when modeling populations 2-3 orders of magnitude smaller than this upper limit.

2. As we have previously mentioned, P_{dif} does not have an upper bound (unlike P_{ber}). Since we want $P_{dif} = \Omega_{R_a} \Delta t \leq 1$, the following condition must be true:

$$\Delta t \leq \frac{1}{\Omega_{R_a}}. \quad (2.116)$$

Note that ensuring Δt is small enough according to this condition makes the preceding note about the accuracy of the numerical value of P_{ber,R_a} a moot point.

The above conditions suggest that it is best to model large populations using the P_{ber} form of transition probabilities. Using condition 2.116 when Ω_{R_a} is large

results in Δt being unreasonably small from the perspective of the number of iterations required for the algorithm to run on. In that case, one can relax this restriction by making Δt bigger, however as long as condition 2.115 is satisfied then the use of P_{ber} is warranted.

2.6.2 General ABK Algorithm Implementation

An iteration of the ABK algorithm determines the change in the population sizes as a result of a given process in the time interval Δt . In the case of second and higher order processes, we first choose a single reactant species population (R_a) with respect to which the algorithm runs throughout the simulated time (i.e., sampling only members of the population R_a to assess if transition events occur in each time step). The limiting reactant species population is the best choice because it requires the smallest number of iterations (but the algorithm yields accurate results for any other choice as well). We then perform the following steps:

1. Evaluate the transition probability *for each agent of R_a* present at time t_n (where the variable n represents the n^{th} time step of the algorithm) using either the P_{ber} or P_{dif} formalism.
2. Draw a random number r_u from the standard uniform distribution in the interval $(0, 1)$.²⁶
3. Check the condition that $r_u < P$ *for each agent of R_a* . If the condition is true, then the state of that particular R_a agent (as well as any additional ones according to the stoichiometry of the process) changes to 0, while agents of the corresponding product species are created and their state changes to 1. If the condition is false the state of all agents does not change.
4. Sum up the number of agents in the populations of all species after all probability conditions have been checked. These are the updated species population sizes at the time point corresponding to the end of time interval Δt , or t_{n+1} .
5. Update the absolute time $t_{n+1} = t_n + \Delta t$, using either a fixed or variable time step increment.

²⁶Some implementations of pseudo-random number generators output a number in the semi-open interval $[0, 1)$. For instance, the Python function `random.random()` produces a number in this range. In this work, we use the Matlab function `rand` that returns a number in the open interval $(0, 1)$. The use of either language's library of functions is acceptable when implementing the algorithm.

2.7 Concurrent Processes and Mixed-Order Kinetics

We now consider the case where an agent of species A can participate in a number of distinct processes of varying overall order, each occurring independently of the others and with its own kinetic constant. Before we generalize, we present a treatment of the following example,



where we have serially numbered the kinetic constants that define this system of processes.²⁷ A member of species A can transition in any one of three different ways within a given time interval. The microscopic differential equation describing the rate of disappearance of A is

$$\begin{aligned} -\frac{dN_A}{dt} &= k_1 N_A + k_2 N_A N_B + k_3 N_A \\ -\frac{dN_A}{dt} &= N_A (k_1 + k_2 N_B + k_3). \end{aligned} \quad (2.118)$$

It is easy to see that the terms in parentheses are the KCFs for each of the three processes with respect to A . That is,

$$-\frac{dN_A}{dt} = N_A (\Omega_{A1} + \Omega_{A2} + \Omega_{A3}). \quad (2.119)$$

We will derive expressions for the total transition probability. First, we can take the sum of the KCFs,

$$\Omega_A = \Omega_{A1} + \Omega_{A2} + \Omega_{A3} = k_1 + k_2 N_B + k_3, \quad (2.120)$$

and use the P_{ber} formalism to determine the *total* probability, $P_{ber,A123}$, of an agent of species A being converted to either X or Y or Z ,

$$P_{ber,A123} = 1 - e^{-\Omega_A \Delta t} = 1 - e^{-(k_1 + k_2 N_B + k_3) \Delta t}. \quad (2.121)$$

An alternative approach is to discretize equation 2.118 and rearrange to get

$$-\frac{\Delta N_A}{N_{A,i}} = (k_1 + k_2 N_B + k_3) \Delta t = \Omega_A \Delta t. \quad (2.122)$$

Using definition 2.25, the *total* transition probability for an agent A within Δt is

$$P_{dif,A123} = \Omega_A \Delta t = (k_1 + k_2 N_B + k_3) \Delta t. \quad (2.123)$$

²⁷Here, the subscript of the kinetic constants does *not* refer to the overall order of each reaction (as was done earlier in this chapter).

As we have previously seen, this is the first order Taylor approximation around $\Delta t = 0$ to equation 2.121.

Let us compare these results to determining the probability of occurrence of each of the individual processes separately. Using the P_{ber} formalism, we have

$$P_{ber,A1} = 1 - e^{-k_1\Delta t} \quad (2.124)$$

$$P_{ber,A2} = 1 - e^{-k_2N_B\Delta t} \quad (2.125)$$

$$P_{ber,A3} = 1 - e^{-k_3\Delta t}. \quad (2.126)$$

Since a reaction event for a given agent A can only occur through one of the three processes and they occur independently of one another (by assumption), the total probability P_{A123} must be equal to

$$P_{A123} = P_{ber,A1} + P_{ber,A2} + P_{ber,A3}. \quad (2.127)$$

However, substituting the previously determined $P_{ber,A123}$ for P_{A123} and the expressions for the individual probabilities shows that the two sides of the above equation are only approximately equal:

$$1 - e^{-(k_1+k_2N_B+k_3)\Delta t} \approx (1 - e^{-k_1\Delta t}) + (1 - e^{-k_2N_B\Delta t}) + (1 - e^{-k_3\Delta t}). \quad (2.128)$$

Equality is approached only as $\Delta t \rightarrow 0$, underscoring the importance of using a small enough time interval. It is easier to see the relationship between the two sides of equation 2.128 by considering the linear approximation of the transition probability for each of the three processes,

$$P_{dif,A1} = k_1\Delta t \quad (2.129)$$

$$P_{dif,A2} = k_2N_B\Delta t \quad (2.130)$$

$$P_{dif,A3} = k_3\Delta t. \quad (2.131)$$

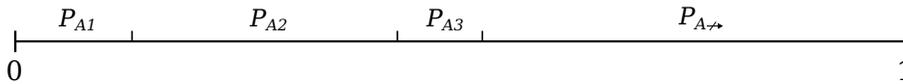
Similarly, substituting $P_{dif,A123}$ into the left side of equation 2.128, we get

$$P_{dif,A123} = P_{dif,A1} + P_{dif,A2} + P_{dif,A3} \quad (2.132)$$

$$(k_1 + k_2N_B + k_3)\Delta t = (k_1\Delta t) + (k_2N_B\Delta t) + (k_3\Delta t). \quad (2.133)$$

The two sides are obviously equal, however the requirement of small Δt still holds since all factors are the first order Taylor approximations to their respective terms in 2.128.

Therefore, for a sufficiently small time interval Δt , it is reasonable to divide the unit interval into non-overlapping contiguous subintervals whose length corresponds to the probability of each of the processes. Then, the remainder of the overall interval corresponds to the probability of no reaction, $P_{A\leftrightarrow}$.



It is easy to implement the ABK algorithm to determine which process occurs:

if a random number r_u drawn from the uniform distribution in the unit interval is less than the sum of individual event probabilities, then a process involving A occurs. In addition, the exact value of r determines *which* transition occurs based on the subinterval of P_{A123} it falls in.

Equation 2.133 suggests another way to view the relationship between the reaction probabilities: the fraction of P_{A123} that corresponds to the occurrence of the first reaction is Ω_{A1}/Ω_A , the second Ω_{A2}/Ω_A , and the third Ω_{A3}/Ω_A . That is, we can evaluate the numerical values of the transition probabilities,

$$P_{A1} = \frac{\Omega_{A1}}{\Omega_A} P_{A123} \quad (2.134)$$

$$P_{A2} = \frac{\Omega_{A2}}{\Omega_A} P_{A123} \quad (2.135)$$

$$P_{A3} = \frac{\Omega_{A3}}{\Omega_A} P_{A123}. \quad (2.136)$$

It is worth noting that considering all reactions an agent A participates in *all at once* contributes to the algorithm's computational efficiency. The procedure we have outlined above requires just a single random number to determine if a particular agent transitions and through which process.

We now extend these ideas to a general case where an agent has an arbitrary number of paths to transition to other species. We present our findings in abbreviated form.

2.7.1 General Case of Concurrent Processes

Consider a population of species R_a , where each agent can participate in up to h distinct processes, each of overall order 1 or greater. In agreement with our observations summarized in equation 2.119, the differential equation describing how the population of R_a changes with time is

$$-\frac{dN_{R_a}}{dt} = N_{R_a} \sum_{i=1}^h \Omega_{R_a,i}, \quad (2.137)$$

where $\Omega_{R_a,i}$ is the KCF of the i^{th} process with respect to agent R_a . Then, given the definition 2.16 of the probability of an agent *not* transitioning in any of the h available ways within time interval Δt ,

$$P_{R_a \rightarrow} = \exp\left(-\sum_{i=1}^h \Omega_{R_a,i} \Delta t\right), \quad (2.138)$$

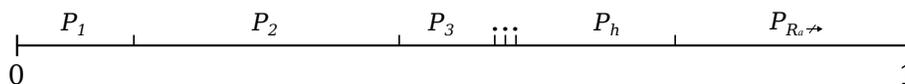
while the *total* probability of an agent R_a transitioning in any one of h ways is

$$P_{R_a,tot} = 1 - \exp\left(-\sum_{i=1}^h \Omega_{R_a,i} \Delta t\right). \quad (2.139)$$

$P_{R_a,tot}$ has an upper bound of 1 and consists of non-overlapping contiguous segments on the number line of the unit interval representing the probability of each of the possible transition paths. That is,

$$P_{R_a,tot} = \sum_{i=1}^h P_{R_a,i}, \quad (2.140)$$

where $P_{R_a,i} = 1 - e^{-\Omega_{R_a,i}\Delta t} \approx \Omega_{R_a,i}\Delta t$. For a sufficiently small Δt value, either the P_{ber} and P_{dif} formalisms can be used to determine the probability $P_{R_a,i}$. Thus, the transition path an agent R_a will take is determined by the subinterval a random number r_u drawn from the uniform distribution in the unit interval falls in. We illustrate this below (for the sake of simplicity and readability, we denote the probabilities $P_{R_a,i}$ as P_i).



This is an important result, as it suggests that if we want to efficiently model numerous processes, we must first identify all the ways in which an agent of a particular species can transition and then choose which process, if any, takes place in Δt using the procedure outlined above. We can then do this for all species that have multiple transition paths. The algorithm is the same as previously described (subsection 2.6.2) and only one random number is needed to decide between the multiple transition paths for a particular agent of a species. Also recall that the time evolution of each process can be modeled by choosing one and only one species with respect to which the transition probability is evaluated. Thus, we can efficiently simulate a series of processes that have a common reactant by simply determining which transition path each agent of R_a takes on each iteration of the algorithm. Programmatically, a simple for loop going through each of the R_a agents is enough to model the time evolution of all processes having R_a as a reactant.

We illustrate the implementation of the algorithm using this approach by presenting several examples in Figure 2.14. We have obtained excellent agreement with the deterministically predicted time trajectories for populations of all sizes, including for low copy number populations consisting of less than 10 agents.

Follow this [link](#) and this [link](#) to see the Matlab code.

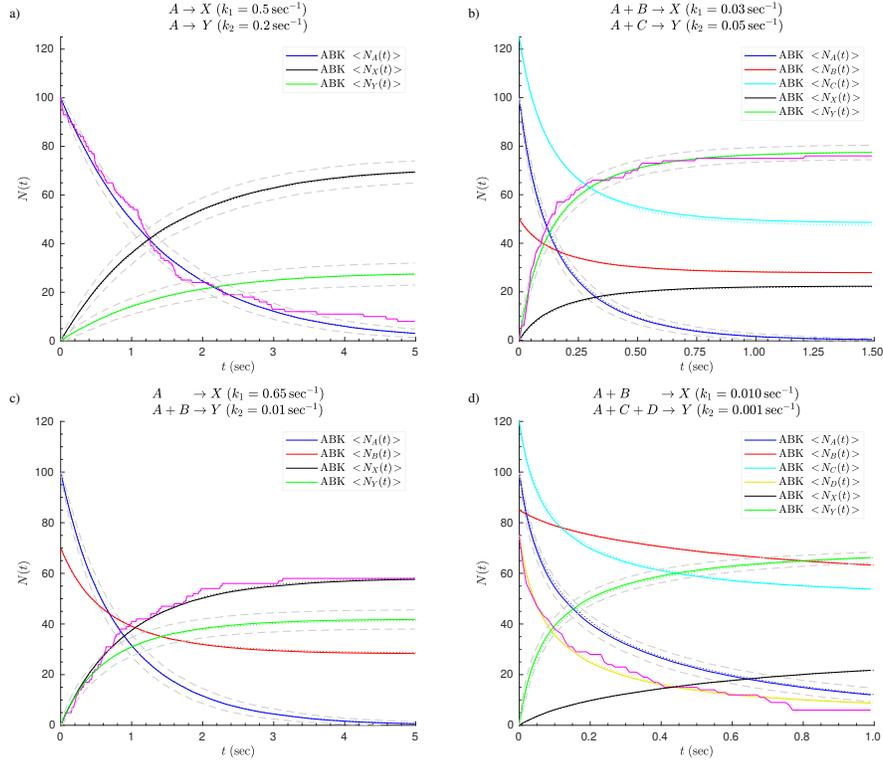


Figure 2.14 – Examples of concurrent processes simulated using the ABK algorithm. Panels *a* and *b* show concurrent processes of the same order (1st and 2nd respectively), while *c* and *d* show mixed-order processes. The rate constants for all processes are indicated in the respective plot titles. We used fixed time interval increments of $\Delta t = 0.01$ sec, the solid curves represent the average population size of the indicated species as simulated with the ABK algorithm ($n = 500$), the gray dashed curves are the one-standard-deviation envelopes of the average population size (not shown for all species for clarity), the dotted curves are the solution to the differential equation describing the time evolution of each species (not easily visible because of the close agreement with the simulation results), and the magenta curve is a sample time trajectory of a selected species population (not shown for all species to avoid overcrowding the plots). **a)** Initial population sizes: $N_{A,i} = 100$, $N_{X,i} = N_{Y,i} = 0$. Agreement between average ABK trajectories and deterministic predictions: $R_{N_A}^2 = 0.99998$, $R_{N_X}^2 = 0.99974$, $R_{N_Y}^2 = 0.99925$. **b)** $N_{A,i} = 100$, $N_{B,i} = 50$, $N_{C,i} = 125$, $N_{X,i} = N_{Y,i} = 0$. $R_{N_A}^2 = 0.99908$, $R_{N_B}^2 = 0.99571$, $R_{N_C}^2 = 0.99873$, $R_{N_X}^2 = 0.99872$, $R_{N_Y}^2 = 0.99852$. **c)** $N_{A,i} = 100$, $N_{B,i} = 70$, $N_{X,i} = N_{Y,i} = 0$. $R_{N_A}^2 = 0.99998$, $R_{N_B}^2 = 0.99892$, $R_{N_X}^2 = 0.99940$, $R_{N_Y}^2 = 0.99803$. **d)** $N_{A,i} = 100$, $N_{B,i} = 85$, $N_{C,i} = 120$, $N_{D,i} = 75$, $N_{X,i} = N_{Y,i} = 0$. $R_{N_A}^2 = 0.99830$, $R_{N_B}^2 = 0.99969 = R_{N_X}^2$, $R_{N_C}^2 = R_{N_D}^2 = 0.99735 = R_{N_Y}^2$. Note that the scale of the horizontal (time) axis varies between plots.

2.8 Birth-Death Process

We now present an important example of mixed-order processes with particular relevance to biological systems: the 0th order production (or “birth”) of molecules/agents belonging to a given species, and their 1st order degradation (or “death”). In summary,



where k_b is the microscopic 0th order rate constant for the production of A , and k_d the 1st order rate constant for its degradation. We can use the ABK algorithm to model this system by evaluating the effect of each of the two processes on the population size of A within a given time interval. The general way of dealing with concurrent processes we presented in the preceding section applies to reactions of order 1 or greater. When a 0th order process is one of the considered processes, it must be accounted for separately from the others because its progression does not depend on the population size of A . Therefore, we evaluate whether a single “birth” event occurs in each time step Δt (i.e., the population of A increases by 1) and whether *each* of the A agents present at the beginning of Δt undergoes a degradation event.

The differential equation describing this birth-death process is

$$\frac{dN_A}{dt} = k_b - k_d N_A. \quad (2.142)$$

The above equation is linear and separable, and therefore can be easily solved to obtain

$$N_A(t) = \frac{k_b}{k_d} (1 - e^{-k_d t}), \quad (2.143)$$

for an initial condition of $N_{A,i} = 0$. The steady-state value²⁸ can be easily computed to be

$$N_A^* = k_b/k_d. \quad (2.144)$$

We present here the results of a simulation with $N_{A,i} = 0$ and kinetic constants $k_b = 0.10 \text{ sec}^{-1}$ and $k_d = 0.01 \text{ sec}^{-1}$. Given these values for the kinetic constants, we get $N_A^* = 10$. The average trajectory of $N_A(t)$ obtained through an ensemble of ABK simulations ($n = 500$) closely matches the deterministic prediction (equation 2.143; $R^2 = 0.9981$; see Figure 2.15a). We also found excellent agreement between the standard deviation of the average ABK trajectory ($n = 5000$) and the theoretical prediction obtained through the CME for this birth-death process (Figure 2.15b; see Appendix A.2 for the derivation of the variance and standard deviation as predicted by the CME).

Finally, we plot a widely-used measure of noise, the coefficient of variation, η , defined as

$$\eta(t) = \frac{SDev(\langle N_A(t) \rangle)}{\langle N_A(t) \rangle}. \quad (2.145)$$

²⁸We will heretofore use the notation of a superscript asterisk to indicate a steady state value for a given quantity.

Follow this [link](#) to see the Matlab code.

Given that a birth-death process is Poissonian (see derivation leading to equation A.38 in Appendix A), it has the property that $SDev = \sqrt{\langle N_A \rangle}$. The coefficient of variation for such a process then is

$$\eta_P(t) = \frac{1}{\sqrt{\langle N_A(t) \rangle}}, \quad (2.146)$$

where η_P is the coefficient of variation for a Poissonian process. Figure 2.15 shows that $\eta(t)$, obtained from an ensemble of ABK simulations, matches $\eta_P(t)$. This indicates that the intrinsic noise in the random production and degradation of A fully accounts for the observed variance in the time trajectory.

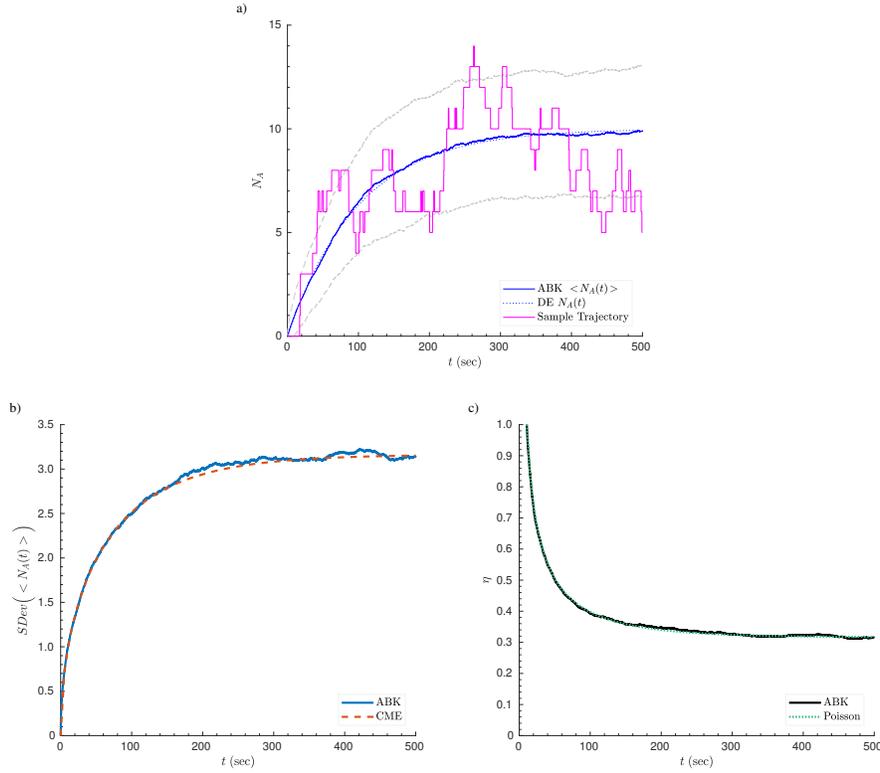


Figure 2.15 – Time course of a birth-death process: $\emptyset \rightarrow A \rightarrow \emptyset$. We used the ABK algorithm to simulate concurrent 0th and 1st order reactions (birth-death process) for the production and degradation of agents belonging to species A . Initial condition $N_{A,i} = 0$, kinetic constants $k_b = 0.10 \text{ sec}^{-1}$ and $k_d = 0.01 \text{ sec}^{-1}$. Algorithm parameters: fixed time step increments of $\Delta t = 0.01 \text{ sec}$, $P_{\emptyset \rightarrow X}$ for the 0th order process (equation 2.108), P_{ber} transition probabilities for the 1st order degradation of A (we obtained identical results using P_{dif}). **a)** The solid blue curve shows the average time trajectory of the population A , $\langle N_A(t) \rangle$, obtained from repeating the ABK simulation 500 times. The one-standard-deviation envelope is also shown (gray dashed curves), and the magenta curve is a sample time trajectory from a single simulation run. Finally, the blue dotted curve is the solution to the differential equation, $N_A(t) = \frac{k_b}{k_d} (1 - e^{-k_d t})$. The coefficient of determination when comparing $\langle N_A(t) \rangle$ and the deterministic trajectory is $R^2 = 0.9981$. **b)** Comparison of the magnitude of the standard deviation obtained from 5000 repetitions of the ABK simulation (blue curve) and as predicted by the CME (red dashed curve; see Appendix A.2 for the derivation). $R^2 = 0.9957$. **c)** Comparison of the coefficient of variation, η , obtained from the simulation data ($n = 5000$), and that expected for a Poissonian process (η_P).

2.9 Processes under Regulatory Control

We now consider processes whose rate constant is modulated by agents of the same or another species, usually in a nonlinear way, thus affecting the temporal evolution of the populations taking part in the process. Such regulatory feedback or feedforward mechanisms can be either positive or negative, leading to activation (up-regulation, or increase in the rate of the process) or repression (down-regulation, or decrease in the rate of the process), respectively.

We consider a 1st order process $A \rightarrow X$ that is regulated by X ,²⁹ and has a rate constant k_o in the absence of any regulation.³⁰ We introduce the function F , which we shall refer to as the microscopic *regulation scaling function* (henceforth abbreviated as RSF), such that the effective rate constant is

$$k = k_o F(N_X), \quad (2.147)$$

where F has the form

$$F(N_X) = \frac{K_{50}^{n_H} + \alpha N_X^{n_H}}{K_{50}^{n_H} + N_X^{n_H}} = \frac{1 + \alpha \left(\frac{N_X}{K_{50}}\right)^{n_H}}{1 + \left(\frac{N_X}{K_{50}}\right)^{n_H}}. \quad (2.148)$$

We assume that α , the Hill coefficient n_H , and K_{50} are time-independent parameters, while the population size N_X is time-dependent. The parameter α denotes the degree of activation or repression and is the asymptotic limit of F (i.e., the effective rate constant approaches αk_o as $N_X \rightarrow \infty$). In summary, the parameter α defines the mode or type of regulation:

$$\text{Type of Regulation : } \begin{cases} \alpha k_o < k \leq k_o & , 0 \leq \alpha < 1 \quad (\text{Repression}) \\ k = k_o & , \alpha = 1 \quad (\text{No Regulation}) \\ k_o < k \leq \alpha k_o & , \alpha > 1 \quad (\text{Activation}). \end{cases} \quad (2.149)$$

Note that αk_o is the minimum value that k can have in the case of repression, while it is the maximum value of k in the case of activation.

F also has the property that $F(0) = 1$, so $k = k_o$ in the absence of X agents. Complete repression corresponds to $\alpha = 0$, and $F \rightarrow 0 \Rightarrow k \rightarrow 0$ as $N_X \rightarrow \infty$ (Figure 2.16, blue curve). K_{50} is the number of agents of X producing half-maximal activation or repression³¹ with respect to the range of F , thus yielding

²⁹The same mathematical expressions and considerations apply to feedforward regulation; that is, if A up- or down-regulates its own conversion to X . In general, a process can be regulated by a number of different species. Here, we consider the simplest case where only species X affects the rate of the process.

³⁰For instance, many biological processes (e.g., transcription) are known to be “leaky,” thus maintaining a level of activity even in the absence of activation.

³¹ K_{50} is the microscopic form of EC_{50} , or half-maximal effective concentration, a commonly used term in the biochemical literature. We retain the subscript “50” to indicate 50% saturation or activity compared to the maximum/minimum level. The conversion can be accomplished by $K_{50} = EC_{50}(\mathcal{N}_{Av}V)$, where V is the system volume and \mathcal{N}_{Av} is Avogadro’s number.

the value,

$$F_{50} = 1 + \frac{\alpha - 1}{2}. \quad (2.150)$$

We also show the *microscopic* form of the well-known Hill function (Hill, 1910), which we shall refer to as H , and depends on N_X in the following way:

$$H(N_X) = \frac{N_X^{n_H}}{K_{50}^{n_H} + N_X^{n_H}} = \frac{\left(\frac{N_X}{K_{50}}\right)^{n_H}}{1 + \left(\frac{N_X}{K_{50}}\right)^{n_H}}. \quad (2.151)$$

The function H is a hyperbola for $0 < n_H \leq 1$ and sigmoidal for $n_H > 1$, has the property $H(0) = 0$, and monotonically increases to the asymptotic limit of 1. K_{50} is the number of agents of X which make $H = 1/2$ (i.e., half-maximal activation/repression). The dependence of the effective rate constant on H has the form

$$k = k_o + (\alpha k_o - k_o)H, \quad (2.152)$$

where the leading additive term, k_o , represents a vertical shift and the multiplicative factor $\alpha k_o - k_o$ stretches or compresses the curve of H depending on the value of α . These transformations ensure that H can be scaled to have the appropriate range and behavior for any regulatory process. Rearranging the above equation, we get

$$k = k_o + k_o(\alpha - 1)H \quad (2.153)$$

$$= k_o [1 + (\alpha - 1)H]. \quad (2.154)$$

Combining this observation with equation 2.147, the relationship between F and H becomes apparent.

$$F = 1 + (\alpha - 1)H \quad (2.155)$$

The reader can verify that the plots of F and $1 + (\alpha - 1)H$ are identical (Figure 2.16). Note that in the case of repression, $\alpha - 1 < 0$, thus k decreases as N_X increases.

Therefore, the dependence of k on either the scaled form of the Hill function (equation 2.154) or RSF (equation 2.147) is the same, and both implementations produce identical results when used to compute transition probabilities in the ABK algorithm. We verified this by simulating the time course of $A \rightarrow X$, with X regulating the kinetics of the process for initial populations $N_{A,i} = 20$ agents, $N_{X,i} = 0$ agents, and $K_{50} = 10$ agents of X (Figure 2.17). We used either the $P_{ber} = 1 - e^{-k\Delta t}$ or $P_{dif} = k\Delta t$ equations for evaluating the transition probability of each agent A within time interval Δt , with the rate constant k determined using either the Hill function or RSF. We have plotted time trajectories for processes without ($n_H = 1$) and with cooperativity ($n_H = 2$). In the case of complete repression ($\alpha = 0$) and $n_H = 2$, the change in population size is steeper than in the absence of cooperativity ($n_H = 1$) in the neighborhood of $N_X = 10 = K_{50}$. This is expected given that the steepness of

Follow this [link](#) to see the Matlab code.

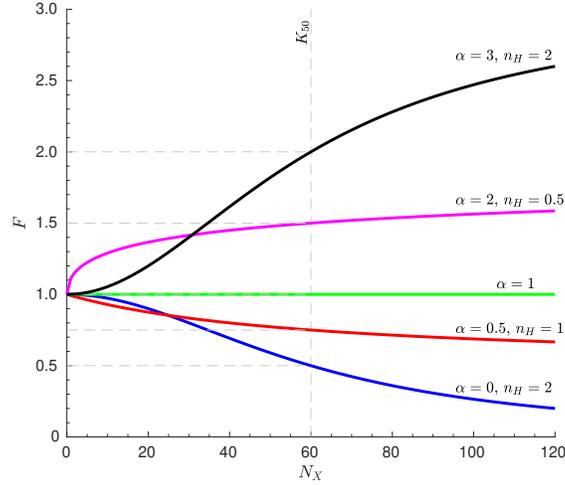


Figure 2.16 – Sample forms of the RSF. This figure shows the dependence of the regulation scaling function, F , on parameters α and n_H . In all cases the curves asymptotically approach the value of $F = \alpha$ as $N_X \rightarrow \infty$. In this example, $K_{50} = 60$ is marked by the vertical dashed gray line (agents of X producing half-maximal activation/repression), and its horizontal branches indicate the F_{50} values corresponding to the different curves (equation 2.150). Note that when $n_H = 1$ the resulting curve is hyperbolic, while for $n_H > 1$ the curve becomes increasingly sigmoidal, a sign of greater cooperativity. $\alpha = 1$ corresponds to no regulation ($F = 1$ for all population sizes of X , green line). The blue and red curves are examples of repression (down-regulation), while the magenta and black curves represent activation (up-regulation).

the RSF (or the Hill function) is greatest when the population size is K_{50} (see solid red and dot-dashed dark red curves in Figure 2.17).

Finally, we briefly present the case where the rate of a process is modulated by agents from multiple species. If the rate of the conversion $A \rightarrow X$ depends on *both* A and X agents,³² then the overall RSF is

$$F = F(N_A) F(N_X), \quad (2.156)$$

with each of the functions having its own set of parameters (α , K_{50} , n_H). Similarly, the overall Hill function is

$$H = H(N_A) H(N_X), \quad (2.157)$$

and all of our previous remarks on how to calculate the effective rate constant k still apply.

³²This can be thought of as a logical AND gate. Other logical relationships between the regulating species are also possible (OR, XOR, etc.).

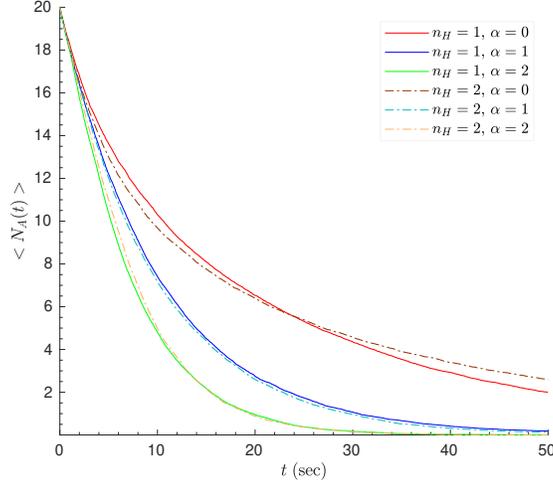


Figure 2.17 – Time course of 1st order process with feedback regulation. We simulated the conversion $A \rightarrow X$, where X regulates the kinetics of the reaction with rate constant k according to the scaling function F or $1 + (\alpha - 1)H$ (equations 2.147, 2.154). We repeated each simulation 500 times using the ABK algorithm (fixed time step increments $\Delta t = 0.01$ sec, P_{dif} transition probabilities) to obtain the average population size of A (along with its standard deviation; not shown). We used the following parameters for all simulations: $N_{A,i} = 20$ agents, $N_{X,i} = 0$, basal rate constant $k_o = 0.10 \text{ sec}^{-1}$, $K_{50} = 10$, with α and n_H values as indicated in the legend. All trajectories showed excellent agreement with the numerically-obtained solution to the differential equation used to describe these processes (not shown).

2.10 Reversible Processes

We continue our presentation of the ABK algorithm’s effectiveness in simulating the time course of basic reaction schemes by considering the reversible process



with 1st order rate constants k_f and k_r for the forward and reverse reactions respectively. Notice that synthesis and degradation of each of the species are not part of the reaction scheme. Therefore, the combined number of A and X agents, N_{TOT} , is a constant. The populations of A and X reach a steady state when the rates of the forward and reverse reactions are equal, leading to

$$\frac{N_A^*}{N_X^*} = \frac{k_r}{k_f}, \quad (2.159)$$

where the star in the superscript indicates the steady-state value of the respective populations. Equation 2.159 can be solved explicitly for N_A^* given the assumption,

$$N_{TOT} = N_A(t) + N_X(t) = N_A^* + N_X^*. \quad (2.160)$$

Solving for N_A^* , we get

$$N_A^* = \frac{k_r}{k_f + k_r} N_{TOT}. \quad (2.161)$$

We treat this reaction scheme as consisting of two separate 1st order processes and use the ABK algorithm to simulate each one separately from the other. Figure 2.18 shows the results of simulating $A \rightleftharpoons X$ for initial populations $N_{A,i} = 10$, $N_{X,i} = 10$ agents, and rate constants $k_f = 0.04 \text{ sec}^{-1}$ and $k_r = 0.20 \text{ sec}^{-1}$. Based on these parameters, we predict $N_A^* = 16.667$ (equation 2.161), while the average of 500 simulation runs using the ABK algorithm yielded a value of $\langle N_A(25.0 \text{ sec}) \rangle = 16.664$.

Follow this [link](#) to see the Matlab code.

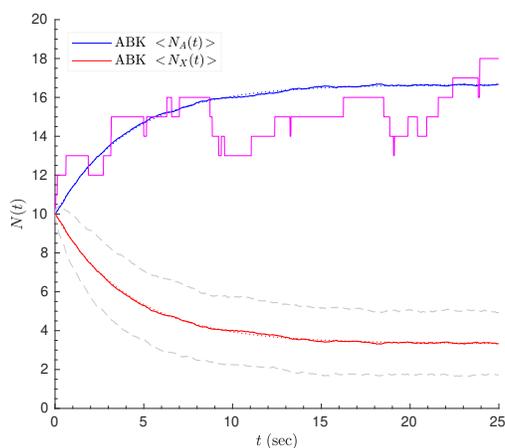


Figure 2.18 – Time course of a reversible process. We used the ABK algorithm to simulate the time evolution of the reversible process $A \rightleftharpoons X$, with forward and reverse 1st order rate constants $k_f = 0.04 \text{ sec}^{-1}$ and $k_r = 0.20 \text{ sec}^{-1}$. The initial population sizes are $N_{A,i} = 10$, $N_{X,i} = 10$ agents. Algorithm parameters: fixed time step increments $\Delta t = 0.01 \text{ sec}$, P_{dif} transition probabilities (P_{ber} produced identical results). The solid curves are the average time trajectory for species A and X over 500 repetitions of the simulation. The gray dashed curves show the one-standard-deviation envelope for the time course of species X (not shown for species A). The magenta curve is a sample trajectory for species A . Finally, the dotted curves show the solution to the differential equation describing this process (not easily visible because of the close agreement with the simulation results; $R^2 = 0.9987$).

2.11 The Michaelis-Menten Reaction Scheme

2.11.1 Full Treatment

We now analyze a very important process for biochemical systems where an enzyme, E , catalyzes the conversion $A \rightarrow X$. An accurate description of this process was first published by the German biochemist Leonor Michaelis and the Canadian physician Maud Menten (1913).³³

We consider here the most commonly used mechanistic scheme describing the overall conversion of the substrate A to X ,



which we shall refer to as the “full treatment” (as opposed to an abbreviated one presented in the next subsection). The forward and reverse reactions in $E + A \rightleftharpoons EA$ have rate constants k_f and k_r , respectively, and k_{cat} is the rate constant for the dissociation of the complex EA (commonly referred to as the enzyme-substrate complex) to the product X and the regenerated catalyst E . We assume that the rate of the uncatalyzed reaction is negligible in the absence of E .

In keeping with our earlier comments on the use of the ABK algorithm, we explicitly model each of the three reactions in this reaction scheme. We present the results of an ABK simulation with initial population sizes $N_{A,i} = 20$, $N_{E,i} = 5$, $N_{X,i} = 0$ in Figure 2.19. The average simulation trajectories are virtually indistinguishable from the numerical integration results of the differential equations describing the reaction scheme. Notice that the substrate population, N_A , is not significantly greater than the total number of enzyme molecules, therefore the assumption that N_{EA} is in a steady state (i.e., $dN_{EA}/dt = 0$) is not warranted. In fact, the black curve in Figure 2.19, representing the average population size of EA , computed from an ensemble of ABK simulations, reveals that N_A is not large enough to saturate the enzyme for any significant period of time.

We have also tested the ABK algorithm for many other choices of parameters and population sizes and have observed exceedingly good agreement with the deterministic predictions. We conclude that the full treatment of the Michaelis-Menten reaction scheme is ideal for modeling such processes regardless of how small the population sizes are, provided that the mechanistic details and kinetic parameters are known. More complicated schemes (e.g., when product formation is reversible, $EA \rightleftharpoons E + X$, or if an enzyme-product complex, EX , is also reversibly formed) can be easily accommodated by the algorithm as additional elementary processes.

We will next consider an abbreviated form of this reaction scheme that significantly reduces the computational time devoted to modeling enzymatic reactions obeying such kinetics. The proposed simplification hinges on whether the

³³The interested reader may consult Gunawardena (2012) for an absorbing perspective on the use of mathematical models in the context of Michaelis and Menten’s original studies.

Follow this [link](#) to see the Matlab code.

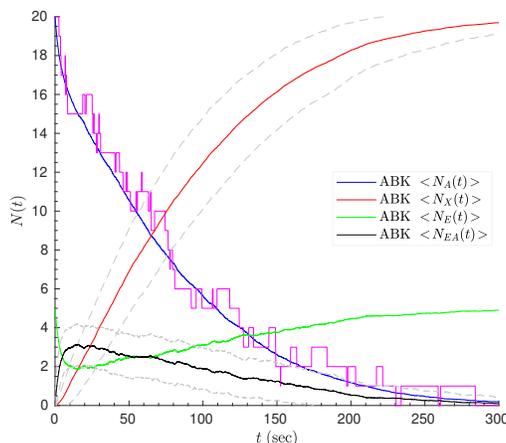


Figure 2.19 – Time course of enzymatic process using the full treatment of the Michaelis-Menten reaction scheme. We used the ABK algorithm to explicitly model each of the three reactions in the Michaelis-Menten scheme (2.162) with microscopic rate constants $k_f = 0.01 \text{ sec}^{-1}$, $k_r = 0.05 \text{ sec}^{-1}$, $k_{cat} = 0.05 \text{ sec}^{-1}$, and initial population sizes $N_{A,i} = 20$, $N_{E,i} = 5$, $N_{X,i} = 0$. Algorithm parameters: fixed time step increments $\Delta t = 0.01 \text{ sec}$, P_{dif} transition probabilities (use of P_{ber} yielded identical results). The solid curves are the average time trajectory ($n = 500$) for species A (substrate), EA (enzyme-substrate complex), E (free enzyme), X (product). The gray dashed curves show the one-standard-deviation envelope for the time course of species X and EA (not shown for species A and E for clarity), while the magenta curve is a sample trajectory for species A . Comparison of the average trajectories to the deterministic predictions: $R_{N_A}^2 = 0.9998$, $R_{N_E}^2 = R_{N_{EA}}^2 = 0.9970$, $R_{N_X}^2 = 0.9998$.

steady-state approximation is valid and a modeler’s desire to balance accuracy and speed of the algorithm when many processes are considered.

2.11.2 Abbreviated Treatment

It is more often true than not that the rate constants for the intermediate steps of an enzymatic reaction are not known. It is instead common for biochemists to experimentally determine the Michaelis constant, which “bundles” the kinetic parameters of the full treatment. Accordingly, it may be necessary to model an abbreviated treatment of the Michaelis-Menten scheme where the conversion $A \rightarrow X$ is treated as a single-step process.

It can easily be shown that assuming the scheme 2.162 and $\frac{dN_{EA}}{dt} \approx 0$, then for the time period during which this assumption is true,

$$-\frac{dN_A}{dt} = \frac{dN_X}{dt} = \frac{k_{cat}N_{E,i}N_A}{K_M + N_A}. \quad (2.163)$$

We define K_M as the *microscopic* Michaelis constant,

$$K_M = \frac{k_r + k_{cat}}{k_f}. \quad (2.164)$$

K_M represents the number of molecules of A that produce a half-maximal rate of product formation and is generally interpreted as a measure of the affinity between the enzyme and a substrate molecule (the smaller the value of K_M , the greater the affinity).³⁴ Equation 2.163 yields a hyperbolic curve when plotting the rate $-dN_A/dt$ as a function of N_A , and the maximal rate is $k_{cat}N_{E,i}$.³⁵

The assumption that the intermediate species EA is at a steady state on a given time scale is sometimes referred to as the *quasi-steady-state assumption*, and it is useful in reducing the dimensionality of the system at hand. In this case, it means that we will not have to explicitly keep track of the species EA in our ABK simulations.

We can use expression 2.163 and the methods developed in earlier sections to derive the transition probability for the process $A \rightarrow X$ during a time interval Δt of arbitrary length. Namely, we discretize equation 2.163, divide both sides by $N_A(t)$, and use our definition of P_{dif} to get

$$P_{dif} = k_{cat} \frac{N_{E,i}}{K_M + N_A(t_n)} \Delta t, \quad (2.165)$$

where $\Delta t = t_{n+1} - t_n$. We denote this transition probability as P_{dif} to indicate that it was derived from the discretized differential rate law describing the overall process. In accordance with the formalism we introduced earlier, the kinetic complementarity factor (KCF) with respect to the substrate species A , is

$$\Omega_A = k_{cat} \frac{N_{E,i}}{K_M + N_A(t_n)}. \quad (2.166)$$

This means that the transition probability can also be expressed as

$$P_{ber} = 1 - \exp\left(-k_{cat} \frac{N_{E,i}}{K_M + N_A(t_n)} \Delta t\right). \quad (2.167)$$

Notice that this expression fits the general form that we described for 1st order processes (after all, the overall reaction is $A \rightarrow X$), except the rate constant k_{cat} is multiplied by a nonlinear term in the KCF. We have already seen a similar case when we discussed processes under regulatory control. Since only one reaction and, therefore, one transition probability expression needs to be evaluated, the abbreviated method of modeling Michaelis-Menten kinetics is computationally

³⁴The Michaelis constant is typically expressed in units of molarity and can be easily converted to its microscopic form. In keeping with the notation we introduced in section 2.1, if K'_M is the molar (macroscopic) Michaelis constant, then for a process occurring in volume V , its microscopic counterpart is $K_M = K'_M(\mathcal{N}_{Av}V)$.

³⁵In practice, it is the *initial* rate vs. substrate population size that's hyperbolic, thus accounting for complicating factors on the reaction rate in real systems. Such factors include enzyme inhibition upon accumulation of product, progressive enzyme inactivation, and the effect of reversibility (in general, since an enzyme does not alter the equilibrium of an overall reversible process $A \rightleftharpoons X$, it can catalyze the reverse reaction as well). Since we have not explicitly modeled any processes in addition to those stated in reaction scheme 2.162 (i.e., we consider an idealized system without degradation, inhibition, or overall reversibility), the distinction of using initial rates is unimportant for our purposes here.

faster and easier to programmatically implement. Our approach has also been used by other researchers to simplify the modeling of enzymatic reactions using Gillespie’s algorithm (Rao and Arkin, 2003), leading to a propensity function that is consistent with the KCF we described above.

We have simulated the time evolution of the overall process $A \rightarrow X$ using this abbreviated treatment of Michaelis-Menten kinetics, with initial populations $N_{A,i} = 20$, $N_{E,i} = 5$, $N_{X,i} = 0$ (same initial conditions as in the full treatment example), and parameters $k_{cat} = 0.05 \text{ sec}^{-1}$, and $K_M = 10$. Note that this value of K_M is computed using equation 2.164 with $k_f = 0.01 \text{ sec}^{-1}$, and $k_r = 0.05 \text{ sec}^{-1}$ (again, same kinetic parameters as in the full treatment example). We summarize the results of the ABK simulation in Figure 2.20.

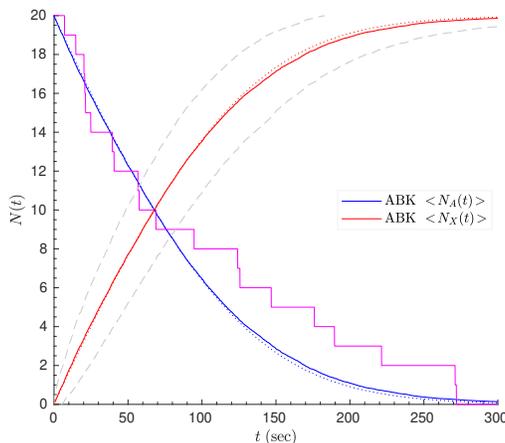


Figure 2.20 – Time course of enzymatic process using the abbreviated treatment of the Michaelis-Menten reaction scheme. We used the ABK algorithm to model the abbreviated form of the Michaelis-Menten scheme ($A \rightarrow X$, with rate equation 2.163). Model parameters: rate constant $k_{cat} = 0.05 \text{ sec}^{-1}$, microscopic Michaelis constant $K_M = 10$, and initial population sizes $N_{A,i} = 20$, $N_{E,i} = 5$, $N_{X,i} = 0$. Algorithm parameters: fixed time step increments $\Delta t = 0.01 \text{ sec}$, P_{dif} transition probabilities (equation 2.165; we obtained identical results when using P_{ber} , equation 2.167). The solid curves are the average time trajectories for species A and X over 500 repetitions of the simulation. The gray dashed curves show the one-standard-deviation envelope for the time course of species X (not shown for A), while the magenta curve is a sample trajectory for A . The dotted curves show the solution to the differential equation describing this process (equation 2.163; not easily visible because of the close agreement with the simulation results). Comparison of the average ABK and deterministic trajectories: $R^2 = 0.9995$.

Generally, the condition that satisfies the quasi-steady-state assumption is (Segel and Slemrod, 1989),

$$\frac{N_{E,i}}{K_M + N_{A,i}} \ll 1. \quad (2.168)$$

Clearly, the approximation is valid when the population of A is much greater than the number of available enzyme molecules E (i.e., excess substrate) and

for high K_M values. We have tested cases that satisfy these criteria and found that the full and abbreviated treatments give virtually identical results, thus justifying the modeling of Michaelis-Menten kinetics as a single-step reaction. The agreement was especially good during the time period when the enzyme is saturated and the quasi-steady-state assumption holds. This is clearly not the case in the example we have presented in this section, and the difference between the two approaches is visible in Figure 2.21.

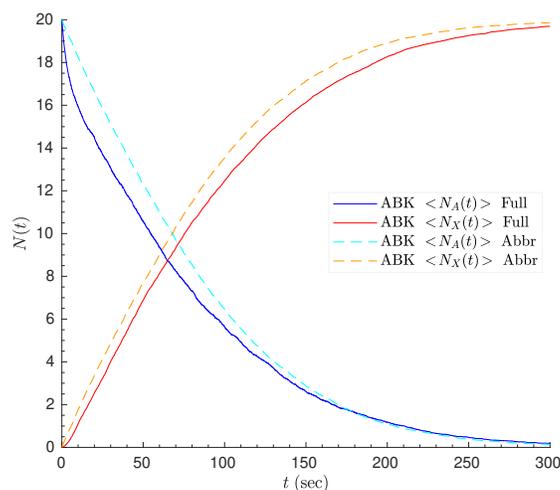


Figure 2.21 – Full vs. abbreviated treatments of Michaelis-Menten kinetics. The curves show the ABK simulation results for the full (reaction scheme 2.162) and abbreviated ($A \rightarrow X$, with rate equation 2.163) treatments of Michaelis-Menten kinetics. All kinetic parameters are the same as in Figures 2.19 and 2.20.

This example is instructive in helping us decide when a chosen treatment of this reaction scheme is appropriate for use with the ABK algorithm. When the mechanistic details and kinetic parameters of an enzymatic reaction are known with confidence, then the full treatment of the reaction scheme will give an accurate temporal evolution of all species. The same method is preferable when the ratio of substrate to enzyme molecules is not large (as in our example). On the other hand, use of the abbreviated treatment is justified when condition 2.168 is met.

Chapter 3

Signal-Response Motifs of Homogeneous Populations, and Other Models

In this chapter we examine the effectiveness of the ABK method in describing what we shall refer to as *motifs*, modules or sets of interspecies interactions that produce qualitatively distinct types of behavior. We will refer to species whose quantifiable levels we interpret as input to the system as the *signal* or *stimulus* (or *external forcing*) S , while the output is the population size of species R , the *response*. We find this is a useful way to idealize the interspecies connectivity patterns in these motifs to describe a diverse array of phenomena. We will mainly present our findings from the perspective of molecular signaling in biological systems, but these motifs can also be considered to occur at higher levels of hierarchical organization (e.g., physiological, organismal, societal). We invite the reader to apply these concepts to as many different contexts as possible.

The motifs we will present serve as units of “structure” from which complex systems are constructed, and their properties are important to understand when using a bottom-up approach to designing systems where the emergence of complex behavior is desired (as in the field of synthetic biology). We take a similar approach to Tyson et al. (2003) and start by presenting simple motifs that, upon modification, exhibit increasingly complex behavior and nuanced signal-response relationships that could not have been easily predicted by mere inspection of the interspecies connectivity or knowledge of the system’s kinetic parameters. We will also find that simulations performed using the ABK methodology can be used to probe the effect of stochasticity on the time trajectories of species with small population sizes.

We stress here that the investigations we present in this chapter are on systems comprising homogeneous species populations, as was the case in the preceding chapter. Our goal is to further validate the ABK method and establish it as suitable for simulating complex systems whose time evolution can

still be compared to deterministic predictions. We will turn our attention to heterogeneous populations in the next chapter.

Finally, we will consider well-known models that are loosely based on physical or biological processes, but are nonetheless useful because of their mathematical simplicity. Specifically, we consider the Lotka-Volterra predator-prey competition model, the Kermack-McKendrick model of infectivity, and the Brusselator. We will show that application of the ABK methodology is straightforward even for those models and provides insight into the role of stochasticity in system dynamics.

3.1 Simple Graded Response Systems

We begin with simple motifs that are straightforward extensions of the processes we presented in Chapter 2 and that produce *graded* (or continuous) *responses*. First, we consider a birth-death process of species R and add a signal species, S , that promotes the birth of the response element. This results in the steady state of R increasing linearly with the population level of S . On the other hand, when an agent/molecule of S promotes one direction of a reversible process, the steady-state level of R is hyperbolic with increasing S . This type of response is sometimes referred to “Michaelian” because it features the same characteristics of initial increase followed by saturation as in the Michaelis-Menten description of enzyme kinetics (although they are different sets of quantities that display this hyperbolic relationship: reaction velocity vs. level of substrate in Michaelis-Menten kinetics, while in our case the steady-state population size of R vs. S).

3.1.1 Linear Response Motif

We consider an extension of the one-species birth-death process we presented in section 2.8 by adding species S that promotes the synthesis of R as shown below.

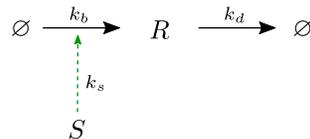


Figure 3.1 – Linear response motif. A birth-death process where S promotes the synthesis of R . We assume that the population size of S does not change.

The processes governed by kinetic constants k_b (0th order synthesis of R) and k_d (1st order degradation of R) are straightforward based on the reactions we have examined in the previous chapter. The process marked by a green dashed arrow in the above figure signifies that S promotes the synthesis of R via a process that can be viewed as either 0th order with a rate constant of bundled terms $k_s N_S$, or 1st order with respect to N_S and rate constant k_s . The reason for this dual interpretation is that we assume that the population of S agents, N_S , does *not* change.

This motif can be considered a simplified description of the following biological process: a repressed gene is transcribed at a basal level (“leaky” activity with rate constant k_b), thereby producing mRNA transcripts (R) which degrade with 1st order rate constant k_d . A surge in transcription occurs when inducer molecules (S) bind to the repressor, thus allowing the RNA polymerase complex to bind to the DNA and synthesize transcripts with rate constant k_s (note that the repressor, DNA molecule, and RNA polymerase are not explicitly shown as distinct species in the motif). The more inducer molecules are present, the more the repressor-gene complex binding equilibrium is pushed toward the dis-

sociated state, thus allowing transcription to occur. A particularly well-studied example of this type of regulation is when repression of the *lac* operon by the tetrameric *lac* repressor is relieved by the inducer allolactose. In our case, since we assume that the population of species S (the inducer) does not change over time, S plays the role of isopropyl β -D-1-thiogalactopyranoside (IPTG), the non-hydrolyzable synthetic analog of allolactose.

This motif is described deterministically by the ODE

$$\frac{dN_R}{dt} = (k_b + k_s N_S) - k_d N_R, \quad (3.1)$$

where we have grouped the terms related to the synthesis of R in parentheses. By assumption, N_S is constant, therefore $k_b + k_s N_S = \text{constant}$ (notice the resemblance to equation 2.142). Setting $dN_R/dt = 0$, we find the steady-state value for the population of R ,

$$N_R^* = \frac{k_b + k_s N_S}{k_d}. \quad (3.2)$$

Based on this formula, it is clear that if enough time passes to allow the population of R to reach its steady-state value, then performing separate experiments with a greater number of S agents should lead to an increase in N_R^* . This equation then captures exactly what we mean by a linear response motif: increasing the signal/stimulus (or input) increases the response (the output) in a *linear* way.

We verified this proposal by using the ABK algorithm to model this set of processes. In our example, we assumed kinetic constants of $k_b = 0.02 \text{ sec}^{-1}$, $k_s = 1.00 \text{ sec}^{-1}$, $k_d = 0.02 \text{ sec}^{-1}$, for an initial population $N_{R,i} = 0$, and a total simulated time of 700 sec (this is significantly longer than the amount of time required to reach the steady state) using fixed time step increments of $\Delta t = 0.02 \text{ sec}$. We ran separate experiments/simulations for N_S values ranging from 0 to 20 agents/molecules (with N_S staying constant in each experiment), and estimated N_R^* by averaging the population size of R over the course of the last 300 sec of simulated time. This allowed us to account for the variation in population size due to stochasticity once the steady state had been reached (we do not show the time course curves here). Figure 3.2 shows the deterministic signal-response (SR) curve (blue line; equation 3.2) with the corresponding values obtained through the ABK simulation. The agreement is excellent (correlation coefficient, $R = 0.9993$).

Follow this [link](#) to see the Matlab code.

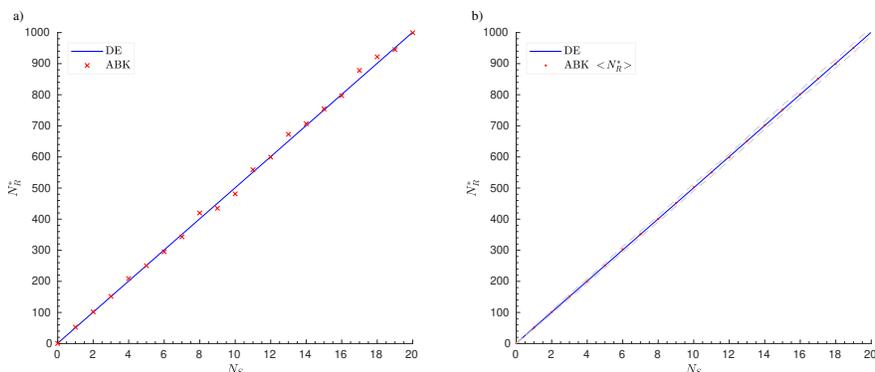


Figure 3.2 – Linear response motif: SR curve. We simulated the time evolution of the reaction scheme outlined in Figure 3.1 using the ABK algorithm. Simulation parameters: kinetic constants of $k_b = 0.02 \text{ sec}^{-1}$, $k_s = 1.00 \text{ sec}^{-1}$, $k_d = 0.02 \text{ sec}^{-1}$, for an initial population $N_{R,i} = 0$, and a total simulated time of 700 sec using fixed time step increments of $\Delta t = 0.02 \text{ sec}$. We performed separate simulations for distinct integer values of S agents ($N_S = 0 - 20$) promoting the production of R , where we assumed that the population of S remains constant. **a)** We plot here the estimated steady-state population size of R , N_R^* (averaged over the course of the last 300 sec of simulated time) vs. N_S for each of the performed simulations (red crosses). The blue curve (marked “DE” in the legend) is the deterministic SR curve (equation 3.2). The agreement between the ABK simulations and the deterministic SR curve is excellent ($R = 0.9993$). **b)** We repeated the ABK simulation ($n = 100$) for each value of N_S to obtain an average value and standard deviation of N_R^* . The resulting $\langle N_R^* \rangle$ vs. N_S curve had an exceedingly good fit to the expected straight line ($R = 0.9999$). The gray dashed lines outline the one-standard-deviation envelope.

3.1.2 Hyperbolic Response Motif

We now examine a hyperbolic relationship between the signal and response elements. Consider a reversible process between species A and R , where the signal S promotes the forward conversion of A to R , as shown below.

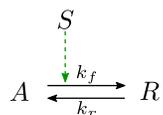


Figure 3.3 – Hyperbolic response motif. k_f and k_r are the microscopic rate constants for the forward and reverse reactions respectively. We assume that S promotes, or favors, the forward transition $A \rightarrow R$, and that the number of S agents does not change.

We further assume that no additional synthesis or degradation processes are taking place, therefore the number of S agents and the combined population of A and R stay constant. In other words,

$$N_{TOT} = \text{constant} = N_A(t) + N_R(t) \quad (3.3)$$

$$N_S = \text{constant}. \quad (3.4)$$

We note that we can alternatively think of S as a catalyst being regenerated during the conversion of A to R . In that case, we can simply write the motif in chemical reaction notation as $A + S \rightleftharpoons R + S$.

Once we have set the size of the combined populations of A and R (N_{TOT}), equation 3.3 relates N_A and N_R in an exact way, thus the system can be described by only one variable. Accordingly, we choose to monitor the time evolution of species R to understand the dynamics of the system. Given these assumptions, we use the law of mass action to deterministically describe $N_R(t)$,

$$\frac{dN_R}{dt} = k_f N_S (N_{TOT} - N_R) - k_r N_R. \quad (3.5)$$

To find the steady-state population of R , N_R^* , we set $dN_R/dt = 0$ and solve for N_R^* .

$$N_R^* = \frac{N_{TOT} N_S}{\frac{k_r}{k_f} + N_S} \quad (3.6)$$

Thus, we expect that as long as we simulate the time evolution of this reaction scheme for enough time so that the steady state is reached, the SR curve should be hyperbolic and in the limit of large N_S , $N_R^* \rightarrow N_{TOT}$ (i.e., response saturation).

Figure 3.4a shows an easy graphical way to understand this relationship. The rate-balance plot shows how the rates of synthesis and degradation of N_R vary as a function of N_R . These rates correspond to the first and second terms in the linear differential equation 3.5, respectively. Since the rates of synthesis and degradation of R are exactly equal at the intersections of these lines, the plot shows the progression of N_R^* values as N_S increases. It is apparent that increasing N_S causes the spacing between the intersections to increase more slowly, as would be predicted by a hyperbolic relationship between N_R^* and N_S (equation 3.6).

We tested this prediction by simulating this reaction scheme using the ABK algorithm. We considered a combined population size of A and R , $N_{TOT} = 100$, initial populations $N_{A,i} = 25$ and $N_{R,i} = 75$, and microscopic kinetic constants $k_f = 0.01 \text{ sec}^{-1}$ and $k_r = 0.10 \text{ sec}^{-1}$. We treated the forward reaction either as a 2nd order process (with rate constant k_f) or as a 1st order process with an effective rate constant of $k_f N_S$ (because N_S is constant by assumption), and obtained identical results using both approaches. We then performed distinct simulations for integer values of N_S ranging from 0 to 40, for a total simulated time of 50 sec with fixed time step increments $\Delta t = 0.01 \text{ sec}$. In each case we determined N_R^* by averaging the population size N_R over the last 25 sec of the simulation to account for stochastic fluctuations once the steady state had been reached. Figure 3.3b shows the SR curve we obtained through simulation (red crosses) and the deterministic prediction (blue curve; equation 3.6). The agreement between simulation and the predicted hyperbolic curve is excellent (coefficient of determination, $R^2 = 0.9844$). Repeating the simulation for each N_S value ($n = 100$) to obtain an average value for N_R^* makes the fit to the hyperbolic curve nearly perfect (Figure 3.3c).

Follow this [link](#) to see the Matlab code.

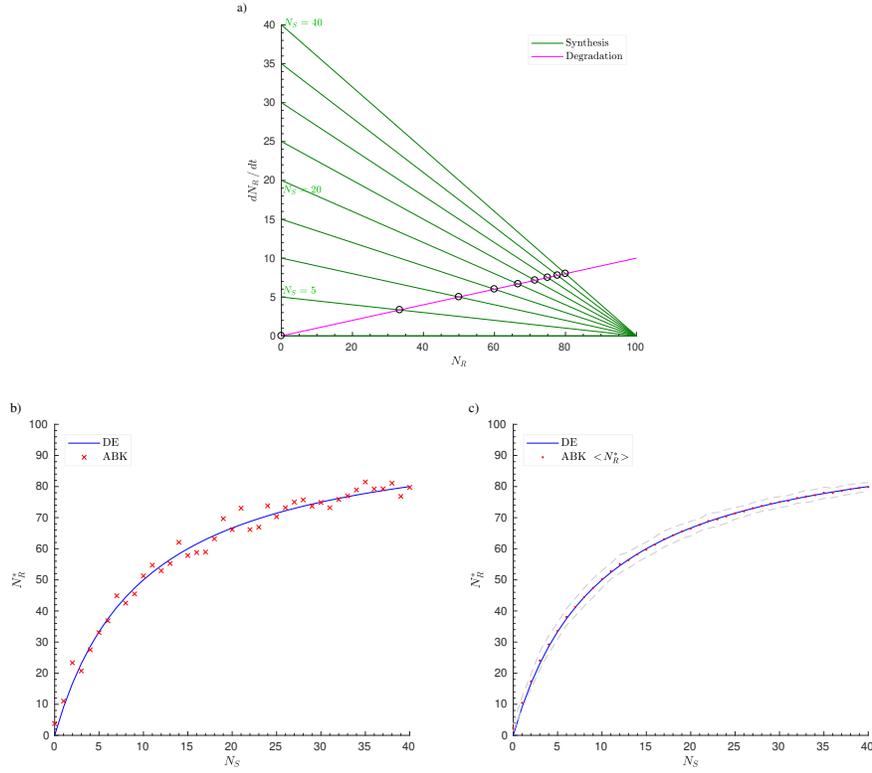


Figure 3.4 – Hyperbolic response motif: rate-balance plot and SR curve. We simulated the time evolution of the motif outlined in Figure 3.3. Simulation parameters: kinetic constants $k_f = 0.01 \text{ sec}^{-1}$, $k_r = 0.10 \text{ sec}^{-1}$, initial populations $N_{A,i} = 25$ and $N_{R,i} = 75$. **a)** Rate-balance plot showing the rates of synthesis (green) and degradation (magenta) of N_R as a function of N_R . Synthesis also depends on N_S , so we have plotted the synthesis curves for values of $N_S = 0 - 40$ (shown here in multiples of 5; selected N_S values are printed on the graph). The intersections (marked by the hollow black circles) represent the steady-state values of population R , N_R^* . **b)** We used the ABK algorithm to model this motif for a total simulated time of 50 sec using fixed time step increments of $\Delta t = 0.01 \text{ sec}$. We performed separate simulations for distinct integer values of S agents ($N_S = 0 - 40$), and assumed that N_S remains constant. We estimated N_R^* in each simulation by averaging over the course of the last 25 sec of simulated time. We plot here N_R^* vs. N_S for each of the performed simulations (red crosses). The blue curve (marked “DE” in the legend) is the deterministic SR curve (equation 3.6). The agreement between the ABK simulations and the deterministic prediction is excellent ($R^2 = 0.9844$). **c)** We repeated the simulation ($n = 100$) for each value of N_S to obtain the average and standard deviation of N_R^* . The resulting curve for $\langle N_R^* \rangle$ gave an exceedingly good fit to the expected curve ($R^2 = 0.9994$). The one-standard-deviation envelope is also shown (gray dashed curves).

The motif we have examined in this section is commonly used within the context of biological signaling. For example, a ligand (the signal) binding to a monomeric protein receptor leads to the receptor response (the output)¹ and dissociation of the ligand-receptor complex is a reversible reaction. Another common example is the reversible enzymatic phosphorylation of a chemical species, catalyzed by a kinase and phosphatase that are *far from saturation*. The latter condition assures that the enzymes respond linearly to the level of substrate (approximated by the rate constants k_f and k_r in equation 3.5) and the analysis we have presented above still holds.

¹Typically, the response is a conformational change that relays the binding event to the intracellular domain of the receptor.

3.2 Continuous Switches

We continue our survey of motifs producing graded, or continuous, responses by presenting two variations of the hyperbolic response motif that result in *ultrasensitivity* (or sensitivity amplification): the response has a heightened sensitivity to a narrow range of signal level. The characteristic shape of an SR curve in these cases is sigmoidal.² We refer to this response type as a *sigmoidal switch* because, given a steep enough sigmoidal curve, a small change in the signal significantly changes the level of the response element. Therefore, for the majority of the possible signal levels, the response element is in either a low or high copy number state that correspond to the switch being in the OFF and ON state respectively. Moreover, the signal must remain constant to maintain a certain level of response.

It is important to consider the mechanisms responsible for converting a hyperbolic response to a sigmoidal one. An easy way to accomplish this conversion is to take the hyperbolic response motif and render it dependent on multiple agents of S . That is, the rate of one direction in a reversible process is dependent on two (or more) agents of S . Indeed, we show that this dependence results in SR curve *sigmoidicity*; however, this way of implementing a dependence on multiple agents of S is arguably unrealistic. For instance, if S is a kinase that phosphorylates a suitable substrate at two distinct sites, then a more realistic representation would be to model the phosphorylation events as separate processes that can occur serially. We thus model the motif as a multi-step process (essentially, a series of reversible reactions) where one S agent acts on each step. Our simulations correctly predict that as the number of considered steps increases, the SR curve becomes increasingly sigmoidal. This way of decreasing the range of signal required to produce a large response differential is well established in the study of molecular signal transduction systems (Ferrell, 1996).

We also explore another way of producing a sigmoidal response: *0th order ultrasensitivity*. In the motifs described in the preceding paragraph, we assumed that the kinetics of the reversible processes follow the law of mass action. Even if one considered these processes to be enzymatic in nature, it was implicit that they operated far from saturation so that the approximation of 1st or 2nd order kinetics was reasonable. If, however, we explicitly treat these processes as obeying Michaelis-Menten kinetics, then in the regime of activity saturation (0th order kinetics) a sigmoidal system response is produced (Goldbeter and Koshland, 1981). We analyze this case to find that the nonlinearity associated with Michaelis-Menten kinetics results in a sigmoidal SR curve. Furthermore, we show that the ABK simulation results are in sharp agreement with the deterministic predictions.

In summary, we consider the same reversible process as in Figure 3.3, but the following changes represent two distinct mechanisms (Ferrell, 1996) for achieving ultrasensitivity: a) two or more S agents (instead of one) are needed for the

²The mechanisms we discuss here are distinct from cooperativity, which is an additional way of producing a sigmoidal response.

forward process $A \rightarrow R$, and b) if the interconversion between A and R is catalyzed by enzymes that are near saturation.

3.2.1 Multi-Step Ultrasensitivity

As we discussed in the hyperbolic response motif, the signal S promotes the conversion of A to R . For instance, if we consider R to be the phosphorylated version of A , then S is a kinase that catalyzes this reaction and we assume that its activity is far from saturation. Only one S agent was needed for this conversion in the hyperbolic case, which is akin to A being phosphorylated at a single site. However, what if A is phosphorylated at two distinct sites? What effect would this dual-site phosphorylation requirement have on the SR curve? An initial attempt at answering these questions entails incorporating the need for two agents of S in the conversion $A \rightarrow R$ in the deterministic treatment of this system. This is equivalent to considering that two molecules of the kinase S perform separate phosphorylation events at the two sites in A and R is essentially A -bisphosphate.

First, we present a simple version of this motif to illustrate how this small modification can produce a sigmoidal response. We refer to it as the “abbreviated 2-step sigmoidal response motif,” as shown below.

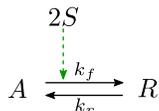


Figure 3.5 – Abbreviated 2-step sigmoidal response motif. k_f and k_r are the microscopic rate constants for the forward and reverse reactions respectively. We assume that two agents of S are needed to promote the forward transition $A \rightarrow R$, and that the number of S agents does not change.

The deterministic rate law of the forward process has a dependence on the square of the signal population level and implicitly assumes that two S agents act simultaneously (or, within the same time interval of size Δt) on A .

As we have done previously, we assume that N_S and $N_{TOT} = N_A(t) + N_R(t)$ are constant. Given the dual requirement for S agents, the differential equation for this motif is identical to 3.5, except for N_S being replaced by N_S^2 ,

$$\frac{dN_R}{dt} = k_f N_S^2 (N_{TOT} - N_R) - k_r N_R. \quad (3.7)$$

Notice that the forward process is now theoretically 3rd order.³ Given that such higher order reactions do not occur in chemical systems, we do not explicitly

³Equation 3.7 is the canonical form of the differential equation. Since the population of S is a discrete quantity, the agent-based interpretation of this process yields the differential equation

$$\frac{dN_R}{dt} = k_f N_S (N_S - 1) (N_{A+R} - N_R) - k_r N_R, \quad (3.8)$$

model the interactions between species A and S , and the forward process is essentially treated as 1st order with a N_S^2 algebraic term in the transition probability expression (i.e., the KCF with respect to each A agent is $\Omega_A = k_f N_S^2$). We will present an arguably more realistic interpretation of the processes in this motif in the next subsection (the full multi-step sigmoidal response motif).

Solving for the steady-state value N_R^* , we get

$$N_R^* = \frac{N_{TOT} N_S^2}{\frac{k_r}{k_f} + N_S^2}. \quad (3.9)$$

This is the same result as in equation 3.6 except for the N_S term being squared, hence a sigmoidal SR curve is expected. Also note that the value of N_S that produces a half-maximal N_R^* is $N_{S,50} = \sqrt{k_r/k_f}$.

We show our results from a simulation of this motif with parameters $k_f = 0.0001 \text{ sec}^{-1}$, $k_r = 0.10 \text{ sec}^{-1}$, for an initial population $N_{A,i} = 25$ and $N_{R,i} = 75$. We observed a sigmoidal SR curve⁴ in excellent agreement with the deterministic prediction (Figure 3.6).

Follow this [link](#) to see the Matlab code.

Full multi-step sigmoidal response motif

A more realistic description of this motif would consider the steps occurring serially. We illustrate below (Figure 3.7a) how the motif can be modified to examine the effect of explicitly including an intermediate species. For instance, using our analogy of protein phosphorylation, the 2-step process can be represented with an additional intermediate, the mono-phosphorylated species B , while R is the dual-site phosphorylated form of A . One agent of S (a kinase) promotes each of the two forward processes. The reverse processes represent the action of a phosphatase on either R or B . As we have previously mentioned, the discussion that follows assumes that both the kinase and phosphatase activities are far from saturation.

where we substituted $N_S(N_S - 1)$ for N_S^2 in recognition of the fact that for the simultaneous effect of two distinct S agents, there are $N_S - 1$ available S agents after the first has been chosen. We have simulated this motif using this agent-based interpretation and obtained an excellent fit to the deterministic prediction based on equation 3.8 ($R^2 > 0.999$). We do not show these results here in the interest of brevity and will focus instead on the canonical interpretation.

⁴We used the same parameters as in the hyperbolic response motif, except here $k_f = 0.0001 \text{ sec}^{-1}$ instead of 0.01 sec^{-1} . The reason for this difference is that in this case the forward process is formally 3rd order (not 2nd) and one would expect the rate constant to be of smaller value for a termolecular reaction event/process. Therefore, we deem it not entirely meaningful to compare the hyperbolic vs. 2-step motifs with the same numerical value for k_f . We have, however, compared the two motifs (with $k_f = 0.01 \text{ sec}^{-1}$) to satisfy our curiosity on the mathematical consequence of replacing N_S with N_S^2 in equation 3.5. Indeed, we observed the conversion of hyperbolicity to sigmoidicity (SR curve not shown) with the simulation results again agreeing with deterministic predictions.

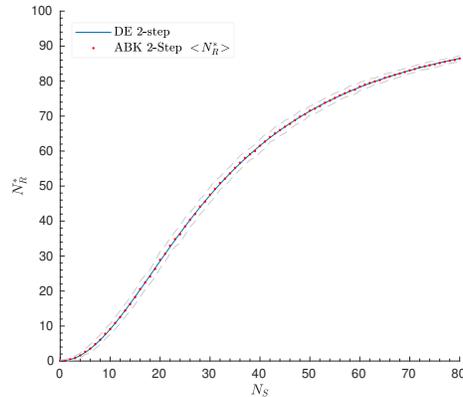


Figure 3.6 – Abbreviated 2-step sigmoidal response motif: SR curve. We simulated the time evolution of the reaction scheme outlined in Figure 3.5. Simulation parameters: kinetic constants $k_f = 0.0001 \text{ sec}^{-1}$, $k_r = 0.10 \text{ sec}^{-1}$, for initial populations $N_{A,i} = 25$ and $N_{R,i} = 75$. We used the ABK algorithm to model this motif for a total simulated time of 100 sec using fixed time step increments of $\Delta t = 0.01 \text{ sec}$, and performed separate simulations for integer values of S agents ($N_S = 0 - 80$). We assumed that the population of S remains constant in each simulation. We estimated N_R^* by averaging N_R over the course of the last 50 sec of simulated time to account for the variation in population size due to stochasticity once the steady state had been reached. We then repeated the simulation ($n = 100$) for each value of N_S to obtain the average and standard deviation of N_R^* . We plot here the average steady-state population size of R , $\langle N_R^* \rangle$ vs. N_S for each of the performed simulations (red filled circles). The blue curve is the deterministic SR curve (equation 3.6). The one-standard-deviation envelope is also shown (gray dashed curves). The fit between the ABK and deterministic curves is excellent ($R^2 = 0.9999$).

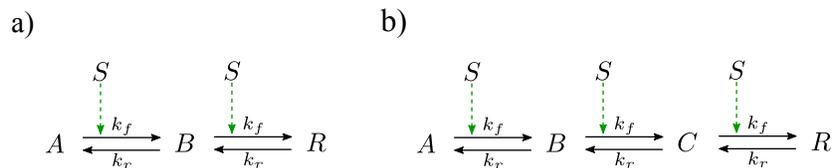


Figure 3.7 – 2- and 3-step sigmoidal response motif. **a)** R is formed in a 2-step process, with B being an intermediate. **b)** R is formed in a 3-step process, with species B and C being intermediates. We assume that k_f and k_r are the microscopic rate constants for the forward and reverse reactions in each of the steps, and that the number of S agents does not change.

The discussion can be easily extended to 3- or multi-step processes (Figure 3.7b). In all cases, we make the assumption that all species cannot be synthesized or degraded beyond the processes that are explicitly stated. Therefore, the following condition is true for the 2-step motif: $N_{TOT} = N_A(t) + N_B(t) + N_R(t) = \text{constant}$.

We present here the results of simulating the 2- and 3-step motifs and use the same values of k_f and k_r for each of the forward and reverse processes as we

Follow this [link](#) to see the Matlab code.

did in the hyperbolic response motif (or 1-step process, Figure 3.3). This allows us to directly compare these motifs and observe the effect of having additional steps on the SR curve. The interested reader can state the differential equations and derive the analytic expressions for the deterministic steady-state values of the different species (we do not present this treatment here in the interest of brevity). We show a summary of the deterministic and ABK simulation results in Figure 3.8, where the transition from hyperbolicity to sigmoidicity is clearly seen as the number of steps increases.

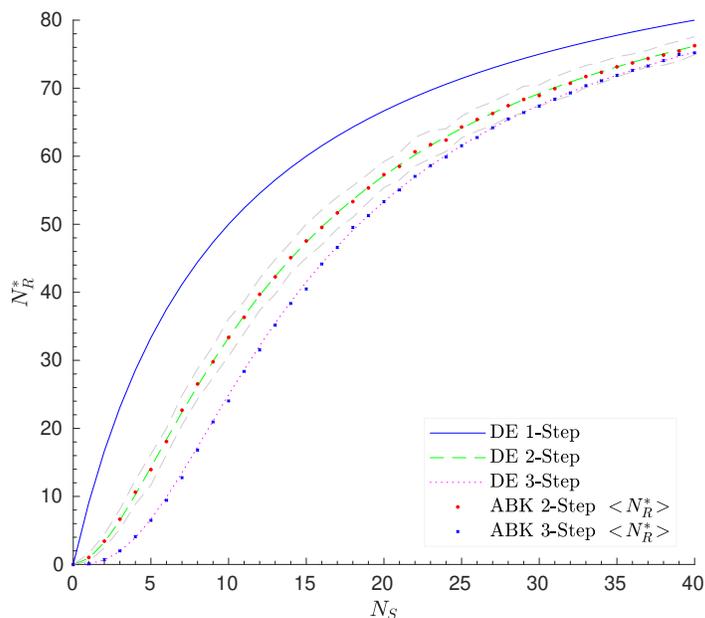


Figure 3.8 – Multi-step sigmoidal response motif: SR curve comparison. We simulated the time evolution of the reaction schemes outlined in Figure 3.7*a,b*. Simulation parameters: kinetic constants $k_f = 0.01 \text{ sec}^{-1}$ and $k_r = 0.10 \text{ sec}^{-1}$ for each of the forward and reverse processes, total simulated time of 100 sec with fixed time step increments of $\Delta t = 0.01 \text{ sec}$. Initial populations: 2-step motif, $N_{A,i} = 25$, $N_{B,i} = 75$, $N_{R,i} = 0$; 3-step motif, $N_{A,i} = 25$, $N_{B,i} = 75$, $N_{C,i} = 0$, $N_{R,i} = 0$. We performed separate simulations for integer values of S agents ($N_S = 0 - 40$) and assumed that N_S remains constant in each. N_R^* was estimated by averaging over the course of the last 50 sec of simulated time. We repeated the simulation ($n = 100$) for each value of N_S to obtain the average and standard deviation of N_R^* . We plot here $\langle N_R^* \rangle$ vs. N_S (red filled circles for the 2-step process; blue crosses for the 3-step process). There was excellent agreement between our simulation results and the deterministic SR curves ($R^2 > 0.9998$ for both the 2- and 3-step motifs). The one-standard-deviation envelope (gray dashed curves) is only shown for the 2-step motif. The green and magenta curves are the deterministic SR curves for the 2- and 3-step motifs respectively (marked “DE” in the legend; curves obtained either analytically or numerically). The deterministic hyperbolic response curve (or 1-step process; blue curve) is shown for reference. Note that the ABK simulation results for the 1-step process are plotted in Figure 3.4*c*.

The above presentation shows that a sigmoidal response can result from considering multiple intermediate steps of an overall reversible process, where each of the reactions can be kinetically described according to the law of mass action. However, these results also hold for enzymatic processes as long as the enzymes operate far from saturating conditions. We will now examine reversible enzymatic processes where the full form of Michaelis-Menten kinetics is used and no restriction on the degree of enzyme saturation is required.

3.2.2 Ultrasensitivity from a Cyclic Enzymatic Process

The reversible phosphorylation of protein molecules, acted on by kinase and phosphatase enzymes and due to a second messenger molecule acting as the signal (e.g., cAMP), is an example of a common biological process that behaves as a reversible switch (Bray, 1995). An analogous example of this motif's use in signal transduction is the reversible interconversion of guanine nucleotide-binding proteins (G proteins). In this case, G proteins are not modified covalently, but their activities are instead modulated by their binding to either GDP or GTP that leads to allosteric conformational transitions. GTPase-activating proteins (GAPs) promote the hydrolysis of GTP to GDP, and guanine nucleotide-releasing proteins (GNRPs) induce the release of GDP. The G protein then binds to a new molecule of GTP and the cycle of conversions continues.

We consider a *reversible* enzymatic process where S promotes the production of R from A (Figure 3.9a). Notice that this motif is schematically identical to the motif that produced a hyperbolic response (Figure 3.3), except in this case we assume that the interconversions between A and R are enzymatic and follow Michaelis-Menten kinetics⁵ (we will use curved arrows to indicate processes obeying such kinetics in this work).

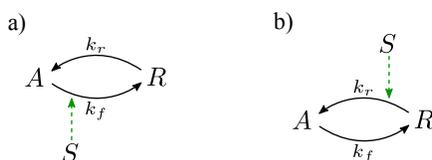


Figure 3.9 – Sigmoidal response motif. k_f and k_r are the microscopic rate constants for the forward and reverse reactions respectively. Both reactions follow Michaelis-Menten kinetics and therefore require the specification of the additional parameters $K_{M,f}$ and $K_{M,r}$ (microscopic Michaelis constants for the forward and reverse reactions). **a)** ON switch: S promotes (or catalyzes) the forward transition $A \rightarrow R$, therefore the steady-state level of R is directly proportional to S . **b)** OFF switch: S promotes (or catalyzes) the reverse transition $R \rightarrow A$, and the steady-state level of R is inversely proportional to S . In both cases, we assume that the number of S agents does not change.

In addition to the rate constants k_f and k_r , we need to use two additional parameters, the microscopic Michaelis constants $K_{M,f}$ and $K_{M,r}$ that represent the number of agents of A and R that produce a half-maximal rate of the

⁵This motif is sometimes referred to in the literature as an *enzymatic futile cycle*.

forward and reverse reactions, respectively. Figure 3.9 presents two motifs: a) R increases when S levels go up (ON switch), and b) increasing S levels lead to a decrease in R (OFF switch, a decreasing sigmoidal curve). Here, we will present our findings only for the ON switch.⁶

We will further assume that species A and R are not produced or degraded through any unspecified processes, therefore $N_A(t) + N_R(t) = N_{TOT}$. Since $N_A(t) = N_{TOT} - N_R(t)$, the deterministic description of this motif is 1-dimensional, and the relevant differential equation is

$$\frac{dN_R}{dt} = \frac{k_f N_S (N_{TOT} - N_R)}{K_{M,f} + (N_{TOT} - N_R)} - \frac{k_r N_R}{K_{M,r} + N_R}. \quad (3.10)$$

It is evident from this equation and our knowledge of the Michaelis-Menten equation that the maximal rate of the forward reaction is $v_{max,f} = k_f N_S$.

It is worthwhile to clarify how our presentation of this motif fits within the context of biological processes. Based on our assumptions (namely, S is regenerated or not consumed during the conversion $A \rightarrow R$), S is an enzyme that catalyzes the forward reaction and k_f is the rate constant of the rate-determining step in the enzyme's mechanism (typically denoted as k_{cat} ; see section 2.11). A very common form of biological signaling involves phosphorylation of substrates or signaling (macro-)molecules, therefore S can be thought of as a kinase that catalyzes the phosphorylation of A to produce species R (which can be the biologically activated form of A , for instance). If the level of kinase can be tuned according to an upstream signal (say, the level of a hormone or second messenger), then in our simplified model S is a proxy for the signal whose levels affect the overall rate of the forward reaction. Similarly for the reverse reaction, $v_{max,r} = k_r$, which suggests that either $k_r = k_{cat,r} N_P$, where N_P is the number of enzyme (e.g., phosphatase) molecules catalyzing the reverse reaction or that k_r is in fact the rate-determining-step rate constant for the reverse reaction and $N_P = 1$. The latter case then assumes that there is always one phosphatase molecule present and in both cases the level of phosphatase is not regulated by any additional processes. Either case is consistent with the motif we present in this subsection, which we try to keep as simple as possible and with the smallest number of processes and molecular species.

To find the steady-state level of the population of R , N_R^* , we set equation 3.10 equal to 0 to obtain

$$k_f N_S (N_{TOT} - N_R^*) (K_{M,r} + N_R^*) = k_r N_R^* [K_{M,f} + (N_{TOT} - N_R^*)], \quad (3.11)$$

which can be rearranged to reveal a quadratic function of N_R^* . We will not present the analytic solution to this equation here (the interested reader can apply the quadratic formula and sort through all of the constants). The solution for N_R^* is known as the Goldbeter-Koshland function (Goldbeter and Koshland, 1981). In this work, we computed the deterministic N_R^* values numerically.

⁶Although not shown here, we have confirmed that our simulations match the deterministic predictions for the case of the OFF switch.

Follow this [link](#) to see the Matlab code.

We present an example of this motif with kinetic constants $k_f = 0.05 \text{ sec}^{-1}$, $k_r = 1 \text{ sec}^{-1}$, and Michaelis constants $K_{M,f} = 30$, $K_{M,r} = 5$. Figure 3.10a shows the rate-balance plot indicating how the rates of synthesis and degradation of N_R vary as a function of N_R . These rates are given by the first and second terms in the differential equation 3.10 respectively. Since both reactions follow Michaelis-Menten kinetics, their rate curves are hyperbolic and their intersections are the steady-state values N_R^* . It is easy to see that in the cases where the synthesis curve (green) intersects the degradation curve (magenta) when it plateaus, a small change in N_S produces an disproportionately larger increase in N_R^* (for instance, when $N_S = 20 \rightarrow 35$, $N_R^* \approx 15 \rightarrow 65$, a more than 3-fold increase in the response compared to that of the signal). For this reason, this type of signal-response mechanism has been termed *0th order ultrasensitivity* (LaPorte and Koshland, 1983) to reflect the fact that when opposing processes are near saturation (the hyperbolic curve’s plateau is the 0th order regime) the response level is particularly sensitive to an increase (or decrease) in the signal.

We simulated the time evolution of the A and R populations using the ABK algorithm for the abbreviated form of Michaelis-Menten transition probabilities. We remind the reader that using the abbreviated treatment of Michaelis-Menten reactions (subsection 2.11.2) does *not* take into account the presence of the enzyme-substrate complex as a distinct species, and is therefore a good approximation only for an enzyme population of small size relative to the substrate’s.⁷ In the context of our presentation of this motif, this approximation is valid for small N_S values. A more careful approach would use the full treatment of Michaelis-Menten kinetics for higher values of N_S . Nevertheless, we proceed with the use of the abbreviated form of transition probabilities for this idealized motif.

We considered initial population sizes $N_{A,i} = 50$ and $N_{R,i} = 50$, and simulate this system for a total time of 1000 sec using fixed time step increments of $\Delta t = 0.02 \text{ sec}$. We performed separate simulations for values of $N_S = 1 - 60$, each of which had approached the steady-state values by at most $t = 500 \text{ sec}$. We averaged $N_R(t)$ during the last 400 sec of each simulation to obtain an estimate for N_R^* . The resulting SR curve, shown in Figure 3.10b, shows excellent agreement with the deterministic prediction of a sigmoidal response (obtained by numerical solution for N_R^* in equation 3.11). We also repeated the simulation ($n = 100$) for each value of N_S and obtained $\langle N_R^* \rangle$ that again agrees with the theoretical SR curve ($R^2 = 0.9991$). We do not show this plot because the agreement to theory is nearly perfect and the one-standard-deviation envelope is so small that it is hardly visually discernible ($\sigma \sim o(1/10)$).

Some notes on the steepness of the sigmoidal SR curve

We will now make some remarks that are useful in assessing systems that show reversible switching behavior, as well as for designing switches with desired properties. We begin by showing the deterministic maximum slope of the sigmoidal

⁷For a brief discussion on how comparable levels of enzyme and substrate affect the motif’s ultrasensitivity and maximal response, see Ferrell and Ha (2014).

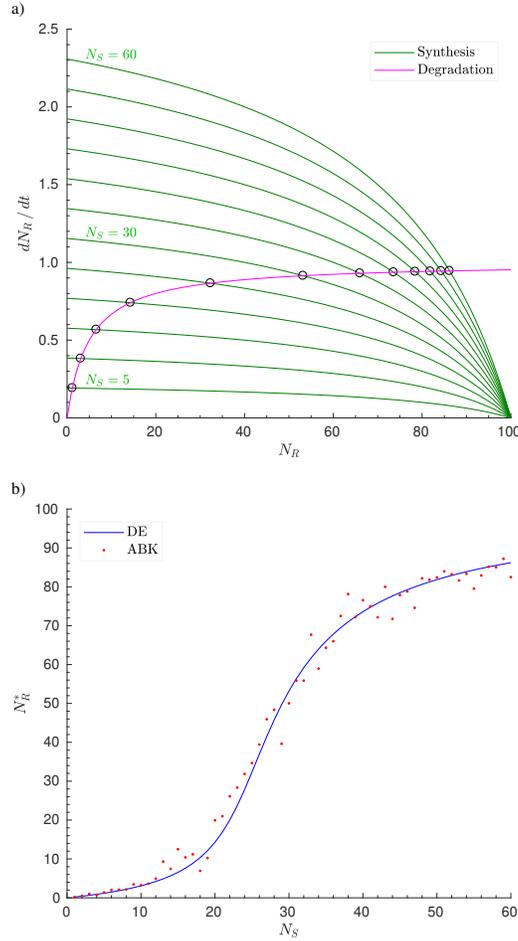


Figure 3.10 – Sigmoidal response motif: rate-balance plot and SR curve. We simulated the time evolution of the motif outlined in Figure 3.9. Simulation parameters: kinetic constants $k_f = 0.05 \text{ sec}^{-1}$ and $k_r = 1 \text{ sec}^{-1}$; Michaelis constants $K_{M,f} = 30$ and $K_{M,r} = 5$; initial population sizes $N_{A,i} = 50$ and $N_{R,i} = 50$. **a)** Rate-balance plot showing the rates of synthesis (green) and degradation (magenta) of N_R as a function of N_R . Since synthesis also depends on N_S , we have plotted the synthesis curves for values of $N_S = 1 - 60$ (shown here in multiples of 5; selected N_S values are highlighted in green text). The intersections (marked by the hollow black circles) represent the steady-state values N_R^* . **b)** We used the ABK algorithm to model this motif for a total simulated time of 1000 sec using fixed time step increments of $\Delta t = 0.02 \text{ sec}$. We performed separate simulations for distinct integer values of S agents ($N_S = 1 - 60$) promoting the production of R , and assumed that N_S remains constant. We plot here N_R^* (averaged over the course of the last 400 sec of simulated time) vs. N_S for each of the performed simulations (red filled circles). The blue curve (marked “DE” in the legend) is the deterministic signal-response curve (equation 3.6). The agreement between the simulations and the theoretical curve is excellent ($R^2 = 0.9899$).

SR curve occurring at the point $(N_{S_{50}}, N_{TOT}/2)$, where $N_{S_{50}}$ corresponds to the population size of S that produces a half-maximal steady-state response. The steeper the curve is at this point, the more switch-like the response is.⁸

Substituting the coordinates of this point into equation 3.11 allows us to solve for $N_{S_{50}}$, obtaining

$$N_{S_{50}} = \frac{k_r (N_{TOT} + 2K_{M,f})}{k_f (N_{TOT} + 2K_{M,r})}. \quad (3.12)$$

We then implicitly differentiate the steady-state condition in equation 3.11 to obtain $\frac{dN_R^*}{dN_S}$. We subsequently substitute our result for $N_{S_{50}}$ (equation 3.12) and $N_R^* = N_{TOT}/2$ into $\frac{dN_R^*}{dN_S}$ to derive an expression for the maximal slope of the sigmoidal curve as a function of the system's parameters.

$$\max \frac{dN_R^*}{dN_S} = \frac{1}{4} \left(\frac{k_f}{k_r} \right) \frac{N_{TOT} (N_{TOT} + 2K_{M,r})^2}{K_{M,f} (N_{TOT} + 2K_{M,r}) + K_{M,r} (N_{TOT} + 2K_{M,f})} \quad (3.13)$$

Increasing k_f and/or decreasing k_r increases the curve's steepness, as does the size of the combined population, N_{TOT} .⁹ This observation suggests that, for a given set of kinetic parameters, smaller population sizes result in less-pronounced switching and require higher levels of $N_{S_{50}}$ when $K_{M,f} > K_{M,r}$. That is, a bigger S population is needed for switching to occur. Figure 3.11 summarizes these results. We are unaware of any experimental data that address this deterministic prediction. In future work, we plan on performing stochastic simulations to further explore the implications of the above observations on the switching behavior of systems with small population sizes.

⁸There are other measures of a sigmoidal curve's steepness. Goldbeter and Koshland (1981) used the macroscopic concentration ratio $EC_{90} : EC_{10}$, where EC stands for effector (or signal) concentration needed to produce a 90% and 10% response respectively. In this work, the microscopic equivalent is $N_{S_{90}} : N_{S_{10}}$. This ratio is 81 for a hyperbolic curve and less than that value for a sigmoidal curve. It is related to the Hill coefficient by $N_{S_{90}}/N_{S_{10}} = 81^{1/n_H}$.

⁹Notice that the expression is overall a quadratic function of N_{TOT} .

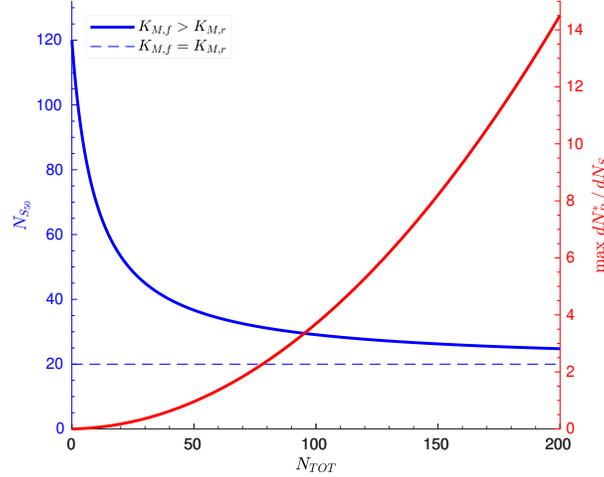


Figure 3.11 – Sigmoidal response motif: $N_{S_{50}}$ and maximum slope of SR curve. We used the parameters $k_f = 0.05 \text{ sec}^{-1}$, $k_r = 1 \text{ sec}^{-1}$, $K_{M,f} = 30$, $K_{M,r} = 5$, and equations 3.12 and 3.13 to show the dependence of $N_{S_{50}}$ (blue curve, left vertical axis) and $\max dN_R^*/dN_S$ (red curve, right vertical axis) on the total population N_{TOT} . The blue curve is typical of the case where $K_{M,f} > K_{M,r}$: it monotonically decreases and asymptotically approaches $\frac{k_r}{k_f}$ (which has a value of 20 for the parameters we used) for large N_{TOT} . $N_{S_{50}}$ can only stay constant if $K_{M,f} = K_{M,r}$ (blue dashed line) and has a value of $\frac{k_r}{k_f}$. For the case when $K_{M,f} < K_{M,r}$, $N_{S_{50}}$ starts lower than $\frac{k_r}{k_f}$ and approaches the same horizontal asymptote (curve not shown). Finally, the red curve has a parabolic form for all parameter values. Therefore, the maximal slope of the SR curve increases monotonically as N_{TOT} increases.

A note on nomenclature. We have so far examined the linear, hyperbolic, and sigmoidal response motifs which give rise to simple input-output relationships. In most cases, the motifs were 1-dimensional (i.e., that can be fully characterized by the time evolution of one species) and with only one unique steady-state value for a given set of parameters. As we consider motifs that result in more complex behavior, we change our terminology to reflect the fact that not all steady states act as “attractors,” as has been the case thus far. We will refer to the population sizes that make the differential equations describing the system equal to zero as *fixed points*, whose stability can be determined using methods in nonlinear dynamics (see Appendix C).

3.3 Discontinuous Switches

The sigmoidal switches we described in the preceding section produce a graded response. We continue with a discussion of *discontinuous switches*, where a small change in the signal can result in large and abrupt changes in the steady-state level of the response element. First, we establish the basis for this behavior by analyzing the dynamics of a motif with two mutually inhibitory species in the absence of a signal. We show that this leads to *bistability* (switch-like behavior) and that the phase plane or state space is split into two basins of attraction. In a deterministic sense, one of two species dominates the other depending on which basin of attraction the initial condition lies in. Since there is no signal to potentially reset the state of the deterministic system, one species dominates indefinitely. A typical behavior of these motifs is that the state of the switch depends on the history of previous states. Namely, in the case of an *irreversible* (or *one-way*) *switch*, as the level of signal surpasses a threshold value, the system switches to its other attractor and is *not* capable of switching back even when the signal decreases to values below the threshold. This phenomenon of a system maintaining memory of past states after any transient signals have decayed is termed *hysteresis*, and its usefulness to biological systems¹⁰ is that it favors the robust maintenance of a switch’s state.

However, the stochastic approach provides insight into the role of noise when considering such systems. In the following sections, we will describe the presence of spontaneous *stochastic switching* events in ABK simulations, where a system can “flip” from one attractor state to another as a result of fluctuations inherent to the system’s processes (i.e., noise). We specifically explore how stochastic fluctuations in the species’ population sizes of the mutual inhibition switch motif probabilistically affect the state of the switch. We show how an ensemble of ABK stochastic simulation runs can be used to calculate the probability of the switch being in one state or the other at a given time point. Therefore, the deterministic prediction of an outcome based on the initial population sizes can be replaced with a probabilistic one.

We then continue our treatment of bistable switches by extending the case of mutual inhibition through adding the effect of a signal species, S , that can relieve one of the inhibitory interactions. We see that this scenario also leads to bistability and again observe the phenomenon of hysteresis. We present through our application of the ABK method to this system how stochasticity affects the transition from one switch state to the other. Furthermore, we show how a small change in the parameters of the same motif can transform it into a hysteretic *reversible* (or two-way) switch: the system transitions to a new attractor state at a critical level of S , whereas it switches to the original attractor at a lower S value than the one required for the initial transition.

We complete our investigations on switches by presenting the case of mutual activation (or positive feedback) where two species promote each other’s synthesis. This motif is also capable of exhibiting switch-like behavior.

¹⁰Hysteresis has also been characterized in many physical systems (e.g., magnets).

Biological significance

In this section, we apply the ABK method to probe the role of stochasticity in bistable systems and illustrate its potential in contributing to the study of a multitude of phenomena spanning many scientific fields and disciplines, some of which we briefly review below.

The phenomenon of *stochastic switching* has been implicated in different biological contexts, a fact that underscores its importance in population dynamics (Tsimring, 2014; Eldar and Elowitz, 2010). An early investigation on this phenomenon was performed by Arkin et al. (1998) who showed how the lytic/lysogenic fate of an *E. coli* cell infected by phage λ depends on a stochastic switch. In addition, some bacterial gene networks that control metabolism, the stress response, and pathogenicity have been observed to use stochastic switching to generate phenotypically heterogeneous cellular populations, a strategy known as *bet hedging* (for reviews of these observations, see Raj and van Oudenaarden, 2008; Eldar and Elowitz, 2010). In other fields (such as ecology and paleoclimatology), the study of a system's transition to an alternate attractor state and early warning signals that could predict such transitions is an active area of research (Scheffer et al., 2012; 2009).

Of particular medical importance is the observation of subpopulations of persister cells that are drug-tolerant in bacterial (Balaban et al., 2004) and cancer cell (Sharma et al., 2010) populations. These are all instructive examples of how organisms have evolved to take advantage of noise, a seemingly counterintuitive proposal given that noise is considered an impediment to reliable control in a system's decision-making process. However, the existence of distinct subpopulations of a species generated through spontaneous stochastic switching events has been theoretically predicted to be the preferred strategy under some conditions for adapting to fluctuating environmental stimuli and stresses (Kussell and Leibler, 2005), thus conferring a fitness advantage to the population as a whole.

3.3.1 The Mutual Inhibition Switch

We begin by analyzing the dynamics of the mutual inhibition between two species, A and R , and will, for the moment, ignore the presence of an extrinsic signal S . Figure 3.12 shows the motif that we will refer to as the *mutual inhibition switch*. The 0th order basal rates of synthesis of species A and R are parameterized by the rate constants k_a and k_r respectively, and the 1st order degradation rate constants are κ_a and κ_r . Moreover, the two species inhibit each other's synthesis, with half-maximal inhibitory activity of each effector defined by the microscopic parameters $K_{50,A}$ and $K_{50,R}$. Finally, we use the Hill coefficients $n_{H,A}$ and $n_{H,R}$ to denote the cooperativity of inhibition *due to* species A and R respectively.

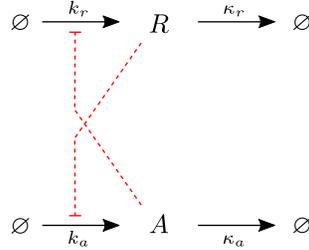


Figure 3.12 – Mutual inhibition switch motif. k_a and k_r are the microscopic 0th order rate constants for the synthesis of A and R , and κ_a and κ_r are the microscopic 1st order rate constants for the degradation of A and R respectively. A inhibits the synthesis of R with half-maximal inhibition occurring with $K_{50,A}$ agents/molecules of A , and Hill coefficient $n_{H,A}$. Similarly, the inhibition of A by R is parameterized by $K_{50,R}$ and $n_{H,R}$.

We can describe this motif deterministically by the following set of differential equations,

$$\frac{dN_A}{dt} = k_a \frac{K_{50,R}^{n_{H,R}} + \alpha_r N_R^{n_{H,R}}}{K_{50,R}^{n_{H,R}} + N_R^{n_{H,R}}} - \kappa_a N_A \quad (3.14)$$

$$\frac{dN_R}{dt} = k_r \frac{K_{50,A}^{n_{H,A}} + \alpha_a N_A^{n_{H,A}}}{K_{50,A}^{n_{H,A}} + N_A^{n_{H,A}}} - \kappa_r N_R, \quad (3.15)$$

where we have used the general form of the RSF (equation 2.148). These equations can be used to analyze the dynamics of any regulatory effect between species A and R because the parameter α denotes the degree of activation/repression. In this example, we will assume mutual inhibition and set the parameters $\alpha_a = \alpha_r = 0$, signifying *complete* mutual repression of the two species. To further simplify our analysis, we will assume that the cooperativity and strength of mutual inhibition between A and R are the same. In summary, our assumptions are:

$$\begin{aligned} \alpha_a &= \alpha_r = 0 \\ K_{50,A} &= K_{50,R} \equiv K_{50} \\ n_{H,A} &= n_{H,R} \equiv n_H. \end{aligned} \quad (3.16)$$

Accordingly, the differential equations can be rewritten in simplified form.

$$\frac{dN_A}{dt} = k_a \frac{K_{50}^{n_H}}{K_{50}^{n_H} + N_R^{n_H}} - \kappa_a N_A = k_a \frac{1}{1 + \left(\frac{N_R}{K_{50}}\right)^{n_H}} - \kappa_a N_A \quad (3.17)$$

$$\frac{dN_R}{dt} = k_r \frac{K_{50}^{n_H}}{K_{50}^{n_H} + N_A^{n_H}} - \kappa_r N_R = k_r \frac{1}{1 + \left(\frac{N_A}{K_{50}}\right)^{n_H}} - \kappa_r N_R \quad (3.18)$$

Despite these simplifications, finding the fixed points is algebraically challenging in the general case,¹¹ but can be determined numerically.

¹¹For an even more simplified version of this system of differential equations amenable to easy algebraic analysis, see Kaplan and Glass, 1995, pp. 235-236.

We consider the following values for the parameters of the mutual inhibition switch motif: $k_a = k_r = 1.0 \text{ sec}^{-1}$, $\kappa_a = \kappa_r = 0.01 \text{ sec}^{-1}$, and $K_{50} = 50$ agents/molecules of A or R . We used the Symbolic Math toolbox in Matlab to numerically solve for the fixed points and assess their stability (see Appendix C). We show in Figure 3.13a the bifurcation diagram for the fixed points N_R^* as a function of the Hill coefficient n_H . The blue curve corresponds to the stable fixed points (we will also refer to them as *attractors* or *sinks*), while the red dashed curve signifies unstable fixed points. The system is monostable for $n_H \leq 2$ with $N_R^* = 50$, and becomes bistable for $n_H > 2$, where $N_R^* = 50$ now becomes unstable. Thus, a transcritical pitchfork bifurcation occurs at $n_H = 2$.

To further understand the dynamics when $n_H > 2$, we show the phase plane plot of this system for a particular value of the Hill coefficient, $n_H = 3$ (Figure 3.13b), where we plot the nullclines for N_A (red curve for which $dN_A/dt = 0$) and N_R (blue curve for which $dN_R/dt = 0$). The intersections between the nullclines are the fixed points whose stability can be easily assessed by visual inspection of the direction field (green arrows). The plot reveals two stable fixed points on the upper left and lower right corners of the phase plane (blue-filled circles), while a third fixed point (red-filled circle) exists at (50, 50). Notice that trajectories are attracted to the latter fixed point along the diagonal, but are repelled from it along the orthogonal direction. Such fixed points are known as *saddles*. Finally, we show the boundary between the two basins of attraction, known as the *separatrix* (dashed orange line along the diagonal). That is, initial population sizes above the separatrix tend towards the stable fixed point on the upper left corner of the phase plane, while trajectories move toward the other stable fixed point when below the separatrix. We shall heretofore refer to the two stable fixed points on the upper left and lower right corners of the state space as “sink 1” and “sink 2” respectively.

A note on nomenclature. The term *phase plane* is reserved for use in systems described by continuous variables. The solutions to the deterministic equations assume continuity and therefore the direction field can be theoretically calculated at every point in the plane. On the other hand, our agent-based simulations can only produce trajectories that “jump” from one discrete state to another. Therefore, we shall refer to this plot as the *state space* of the discrete dynamical system when overlaying trajectories obtained through ABK simulations.

The role of stochasticity in determining the switch state

We now examine the results of implementing the ABK algorithm to simulate the time evolution of the populations $N_A(t)$ and $N_R(t)$ given the kinetic and inhibition parameter values listed in Figure 3.13 and Hill coefficient $n_H = 3$. We simulated the system for a total time of 2000 sec with fixed time step increments of $\Delta t = 0.02 \text{ sec}$. In addition, we considered the initial population sizes to be $N_{A,i} = N_{R,i} = 10$. We show the results of one such simulation plotted parametrically with respect to time on the state space in Figure 3.14a. Notice

Follow this [link](#) to see the Matlab code.

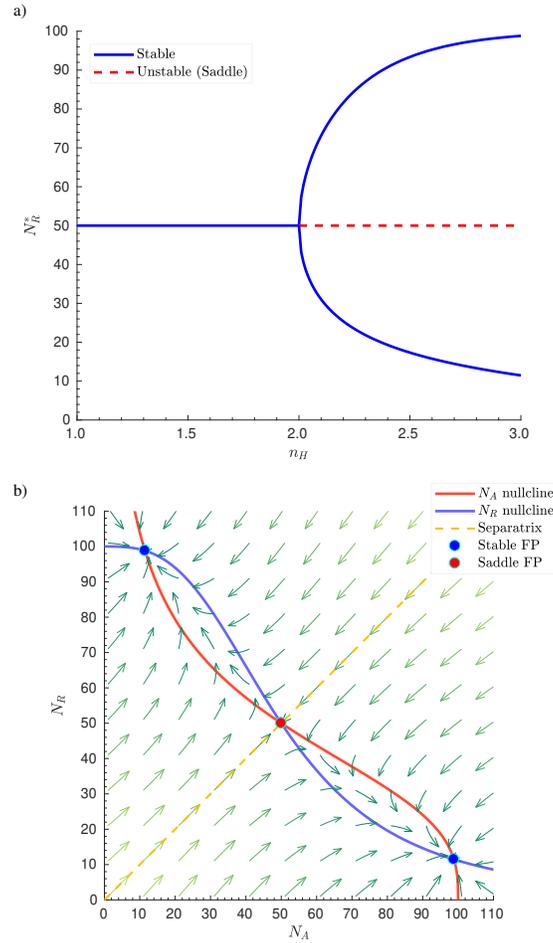


Figure 3.13 – Mutual inhibition switch motif: N_R^* bifurcation diagram with respect to n_H and phase plane. We consider the mutual inhibition switch motif with parameters $k_a = k_r = 1.0 \text{ sec}^{-1}$, $\kappa_a = \kappa_r = 0.01 \text{ sec}^{-1}$, and $K_{50} = 50$ agents/molecules of A or R . **a)** We varied the parameter n_H and used Matlab to numerically solve for the fixed points in the system of differential equations 3.17 and 3.18. We then assessed their stability using Linear Stability Analysis (see Appendix C). A transcritical pitchfork bifurcation occurs at $n_H = 2$ and the system becomes bistable for $n_H > 2$. Blue solid curve: stable fixed points; red dashed curve: unstable (or saddle) fixed points. **b)** Phase plane of this system calculated for $n_H = 3$. We show the nullclines (solid curves) and their intersections (blue-filled circles for stable fixed points and red-filled circle for the saddle point; legend key: FP, fixed point). At each ordered pair (N_A, N_R) we evaluated the velocity vector $(dN_A/dt, dN_R/dt)$ to produce the direction field (green arrows; lighter shades of green correspond to greater amplitude of the velocity vectors). The separatrix is also shown (orange dashed line).

that this initial condition lies on the separatrix, and therefore the deterministic trajectory (black crosses) strictly follows a linear path to the saddle (50, 50), as would be expected based on the direction field shown in Figure 3.13b. Since the deterministic trajectory stays along the separatrix, it has no chance to “curve” one way or the other toward the stable fixed points, and settles instead on the saddle (i.e., the separatrix *is* the stable manifold of the saddle). However, in accordance with our expectation of how real systems behave, the ABK simulation shows that stochastic fluctuations will cause the trajectory to settle in one of the two basins of attraction. The sample trajectory shown in Figure 3.14a starts at (10, 10) at $t = 0$ sec, and ends at (96, 15) at $t = 2000$ sec, which is in the neighborhood of the numerically-evaluated stable fixed point at (98.80, 11.47), or sink 2.

We repeated the simulation ($n = 500$) to compute and plot the average population trajectories (Figure 3.14b). We observed that the populations converged on the neighborhood of either sink with equal likelihood, as would be expected when stochasticity determines the switch state given that the initial condition lies on neither basin of attraction.

We show the results of one additional case that highlights the dependence of this motif’s behavior on stochasticity: the values of the kinetic and inhibition parameters are the same as above except the initial population sizes are $N_{A,i} = 10$, $N_{R,i} = 11$. Since this initial condition lies above the separatrix, the deterministic solution converges on sink 1, the stable fixed point ($N_A = 11.47$, $N_R = 98.80$). Figure 3.15 shows a sample ABK trajectory where the populations are observed to converge on sink 2 (instead of sink 1) with a frequency of 50.2% ($n = 1000$). We do not show the average trajectories in this case because they are nearly identical to those in the preceding figure (3.14b).

It is evident that for initial population sizes close to the nullcline, the probability of going to one sink versus the other is approximately 50%, as expected for a stochastic process with no initial bias toward either sink. We also performed simulations for other initial conditions and observed the probability of convergence on either sink change accordingly. For instance, an initial condition in the lower basin of attraction, ($N_{A,i} = 75$, $N_{R,i} = 50$), converges on sink 2 with a frequency of 94.4% ($n = 500$; total simulated time of 2000 sec for each run), while random fluctuations (noise) of sufficient amplitude divert the remaining 5.6% of the trajectories to sink 1’s state-space neighborhood. In another example, starting from an initial condition very close to sink 1 ($N_{A,i} = 10$, $N_{R,i} = 90$), the system converged to sink 2 ($N_A^* = 98.8$, $N_R^* = 11.47$) only once in 4000 simulation runs (total simulated time of 2000 sec for each run). It is noteworthy that even when starting so close to a sink, a trajectory converging on the other sink is still possible.¹²

¹²If we had greater computational resources, we would perform simulations for a range of integer values of population sizes representing the entirety of the state space shown in the preceding figures. We would then be able to obtain estimates of the scalar-valued function $P(N_{A,i}, N_{R,i})$, denoting the probability of convergence on either sink within a specified amount of simulated time.

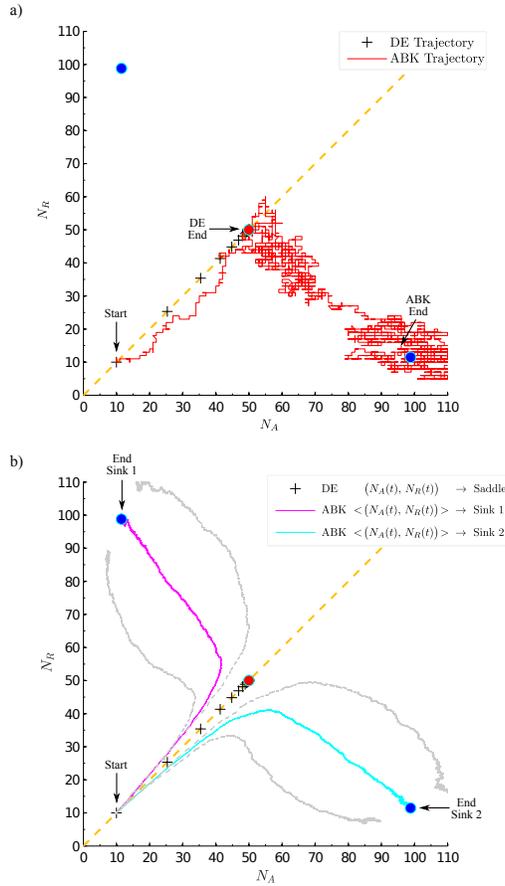


Figure 3.14 – Mutual inhibition switch motif: stochastic time trajectories in state space (case 1). We used the ABK algorithm to simulate the time evolution of the mutual inhibition switch motif. Model parameters: $k_a = k_r = 1.0 \text{ sec}^{-1}$, $\kappa_a = \kappa_r = 0.01 \text{ sec}^{-1}$, $K_{50} = 50$, $n_H = 3$, with initial populations sizes $N_{A,i} = N_{R,i} = 10$. We simulated this system for 2000 sec with fixed time step increments of $\Delta t = 0.02 \text{ sec}$. We refer to the two stable fixed points as sink 1 and sink 2 in the figure. **a)** Sample run of the ABK simulation where the time trajectory of the populations N_A and N_R is plotted parametrically on the state space (red path). We label the beginning and end of the ABK time trajectory on the plot to aid its visualization (although it is impossible to fully convey the time dependence of the trajectory on a static image such as this). The solution to the differential equation given these initial population sizes is also shown (black crosses); the deterministic time trajectory ends at the saddle (50, 50) because it “flows” along its stable manifold. In this simulation run, the populations converge to sink 2, the stable fixed point with coordinates ($N_A = 98.80$, $N_R = 11.47$), and fluctuate stochastically in its vicinity. **b)** We repeated the simulation of this motif 500 times to observe the frequency with which the populations approach the two stable fixed points. We then classified the time trajectories according to which sink they approached and then averaged them (magenta and cyan curves for sinks 1 and 2, respectively). We have also plotted the one-standard-deviation envelopes for each average trajectory (gray dashed curves). The system goes to sink 1 ($N_A = 11.47$, $N_R = 98.80$) 49.2% of the time and to sink 2 ($N_A = 98.80$, $N_R = 11.47$) 50.8%.

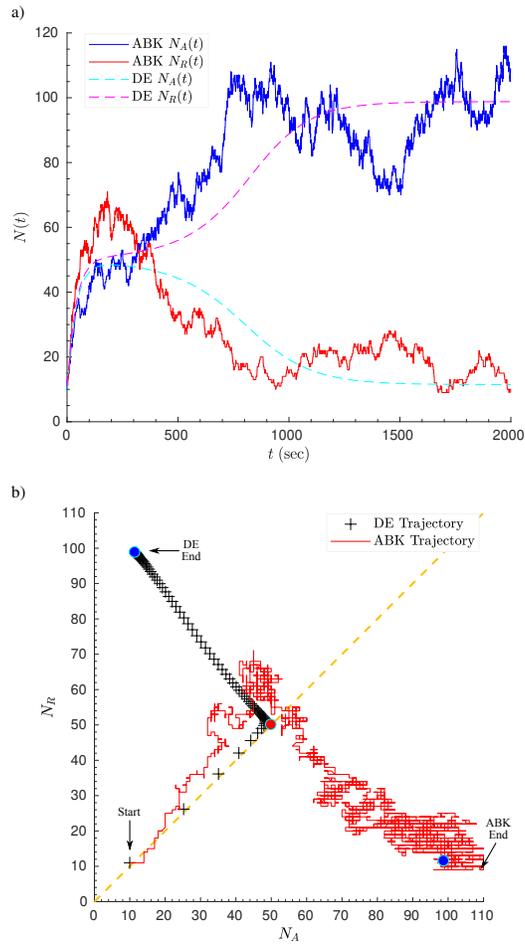


Figure 3.15 – Mutual inhibition switch motif: stochastic time trajectories in state space (case 2). We used the same parameter values as indicated in Figure 3.14, except the initial population sizes are $N_{A,i} = 10$, $N_{R,i} = 11$. **a)** Time course of a sample run where the ABK trajectories are shown along with the deterministic solutions. **b)** Same sample simulation as in **a)**, but plotted parametrically on the state space. Notice that the deterministic solution converges on sink 1 (upper left corner), while the ABK trajectory takes a brief excursion in the upper basin of attraction before ultimately ending up in the neighborhood of sink 2 (lower right corner). The frequency of convergence on sink 2 was 50.2% ($n = 1000$).

We tested whether simulating this motif for a sufficiently long period of time leads to a *spontaneous* reversal in the switch state (i.e., a switch “flip”). By spontaneous, we mean that a reversal occurs as a result of the *intrinsic noise* that is characteristic of the motif’s processes, rather than an external signal influencing the interspecies inhibition or some other aspect of the system (we will investigate the latter case in the next subsection). We ran a simulation of the mutual inhibition motif with the same parameters as above for a total simulated time of 5×10^5 sec. We used the initial condition of $(N_{A,i} = 30, N_{R,i} = 100)$, which is a point in state space within sink 1’s basin of attraction. The system is therefore expected to settle in the vicinity of sink 1. Indeed, the ABK-simulated sample trajectory quickly converged on the vicinity of sink 1, however we subsequently observed two reversals of the switch state in the time window of $2.0 - 3.5 \times 10^5$ sec (shown in Figure 3.16). If we take sink 1 and 2 to represent the switch in the ON (high N_R) and OFF (low N_R) state respectively, then the first “flip” turns the system from the ON to the OFF state, while the second reversal occurrence restores the ON state. We refer to each instance of this behavior as a *noise-induced switch state reversal*.

We have thus shown that in the absence of a signal, convergence on either sink occurs with a finite probability that depends on the initial population sizes. Reaching either sink should not be considered an irreversible process, as our simulations have highlighted the importance of stochastic fluctuations in the spontaneous reversal from one switch state to the other. These random fluctuations, or noise, allow the system to sample a variety of population sizes in state space and can lead the trajectory to cross the separatrix into either basin of attraction. Even if the system is in the immediate neighborhood of a sink, there is a small but finite probability of a noise-induced excursion of sufficient magnitude through state space to the adjacent basin of attraction, possibly leading to a switch state reversal.

The state space can alternatively be characterized by a potential energy function whose minima are at the sink locations, while the separatrix is a potential energy barrier. In this sense, noise can provide the requisite amount of “activation energy” in order for the system to cross that barrier.

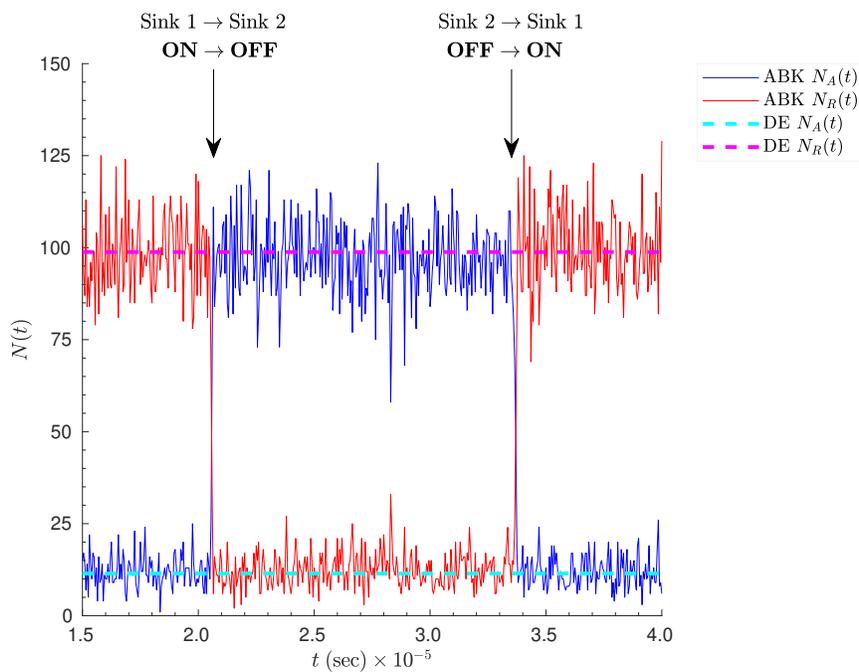


Figure 3.16 – Mutual inhibition switch motif: noise-induced switch state reversal. We used the ABK method to simulate the mutual inhibition switch motif with the same kinetic and inhibition parameters as in Figures 3.14 and 3.15. In this case, the initial population sizes are in sink 1's basin of attraction: $N_{A,i} = 30$ and $N_{R,i} = 100$. We recorded the population time trajectories of species A and R for a total of 5×10^5 sec of simulated time. The system quickly settles in the state-space neighborhood of sink 1, but two spontaneous switch state reversals occur within the time window shown in the figure. We refer to sink 1 and 2 as the ON (high N_R) and OFF (low N_R) switch state respectively. We also show the deterministic trajectories (dashed lines) that converge on sink 1 (ON state) given the initial condition.

3.3.2 The Irreversible Mutual Inhibition Switch

We now consider an extension of the mutual inhibition switch motif by including the presence of an extrinsic signal S that inhibits the inhibition of the synthesis of R by A (Figure 3.17). Therefore, S effectively turns the switch ON, which we take to mean a state with a large population of R . This motif has been constructed as a synthetic switch in *E. coli* (Gardner et al., 2000), where A and R (in our nomenclature) are repressors which inhibit each other's promoter and species S is the inducer IPTG.

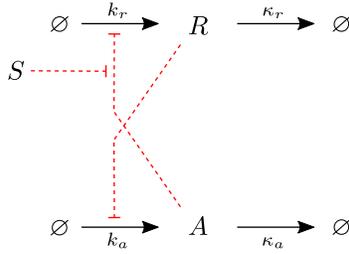


Figure 3.17 – Irreversible (one-way) mutual inhibition switch motif. All reactions and parameters are the same as listed for the mutual inhibition switch motif (Figure 3.12), except for the additional presence of the signal S that inhibits the inhibition of the synthesis of R by A . The parameter $K_{50,S}$ represents the number of agents/molecules of S needed for half-maximal inhibition, and the Hill coefficient $n_{H,S}$ captures the cooperativity of this additional regulatory interaction. We assume that the population size of S remains constant.

Since the signal species S mediates a regulatory interaction, specification of the following additional parameters is dictated by the RSF: the degree of activation/repression α_s , $K_{50,S}$ is the number of S molecules that produce a half-maximal regulatory effect, and $n_{H,S}$ is the Hill coefficient. Given these parameters, we first state the general form of the system of differential equations for this motif.

$$\begin{aligned} \frac{dN_A}{dt} &= k_a \frac{K_{50,R}^{n_{H,R}} + \alpha_r N_R^{n_{H,R}}}{K_{50,R}^{n_{H,R}} + N_R^{n_{H,R}}} - \kappa_a N_A \\ \frac{dN_R}{dt} &= k_r \frac{K_{50,A}^{n_{H,A}} + \alpha_a \left(N_A \frac{K_{50,S}^{n_{H,S}} + \alpha_s N_S^{n_{H,S}}}{K_{50,S}^{n_{H,S}} + N_S^{n_{H,S}}} \right)^{n_{H,A}}}{K_{50,A}^{n_{H,A}} + \left(N_A \frac{K_{50,S}^{n_{H,S}} + \alpha_s N_S^{n_{H,S}}}{K_{50,S}^{n_{H,S}} + N_S^{n_{H,S}}} \right)^{n_{H,A}}} - \kappa_r N_R \quad (3.19) \end{aligned}$$

The only difference between this system of equations and the one presented in the last subsection is the term corresponding to N_A in the RSF for the regulation of R . Here, N_A has been multiplied by the RSF describing the additional regulatory effect by S (compare equation 3.19 with 3.15). Tuning the parameters α_s , α_r , and α_a thus allows us to study any combination of activation and inhibition that may exist between the species in this motif.

In this example, we assume the same kinetic parameters as in the last subsection for the mutual inhibition between A and R . In addition, S completely represses the inhibition of the synthesis of R by A ($\alpha_s = 0$). In summary, our assumptions for the parameter values are:

$$\begin{aligned}\alpha_a &= \alpha_r = \alpha_s = 0 \\ K_{50,A} &= K_{50,R} \equiv K_{50} \\ n_{H,A} &= n_{H,R} \equiv n_H.\end{aligned}\tag{3.20}$$

The differential equations can be rewritten in simplified form as

$$\begin{aligned}\frac{dN_A}{dt} &= k_a \frac{K_{50}^{n_H}}{K_{50}^{n_H} + N_R^{n_H}} - \kappa_a N_A \\ \frac{dN_R}{dt} &= k_r \frac{K_{50}^{n_H}}{K_{50}^{n_H} + \left(N_A \frac{K_{50,S}^{n_{H,S}}}{K_{50,S}^{n_{H,S}} + N_S^{n_{H,S}}}\right)^{n_H}} - \kappa_r N_R,\end{aligned}\tag{3.21}$$

or, equivalently

$$\begin{aligned}\frac{dN_A}{dt} &= k_a \frac{1}{1 + \left(\frac{N_R}{K_{50}}\right)^{n_H}} - \kappa_a N_A \\ \frac{dN_R}{dt} &= k_r \frac{1}{1 + \left(\frac{N_A}{K_{50}} \frac{K_{50,S}^{n_{H,S}}}{K_{50,S}^{n_{H,S}} + N_S^{n_{H,S}}}\right)^{n_H}} - \kappa_r N_R.\end{aligned}\tag{3.22}$$

Finally, for the sake of simplicity we consider the inhibition by S to be non-cooperative ($n_{H,S} = 1$) and take $K_{50,S}$ to have the same value as the corresponding parameter for the other species (i.e., $K_{50,S} = K_{50}$; the strength of inhibition is the same). All other parameters remain the same: $k_a = k_r = 1 \text{ sec}^{-1}$, $\kappa_a = \kappa_r = 0.01 \text{ sec}^{-1}$, $K_{50} = 50$, $n_H = 3$.

Given these parameters, we analyze the deterministic behavior of this motif by numerically solving the above differential equations. First, recall that the phase plane for $N_S = 0$ is equivalent to that shown in Figure 3.13b, with the unstable fixed point, or saddle, at $(50, 50)$ and the system is bistable. We then plot the phase plane in Figure 3.18a,b for two different nonzero values of N_S and observe that as N_S increases the saddle and stable fixed point on the lower right (N_A is large and N_R is small; sink 2) approach one another, eventually collide, and finally annihilate each other when $N_S = 12$.¹³ The disappearance of those fixed points makes the system monostable, with sink 1 featuring a large R population remaining as the only attractor. In summary then, for $N_S > 12$ the populations converge on sink 1 and the switch is irreversibly in the ON state (high N_R^*).

These considerations lead to the theoretical SR curve shown in Figure 3.18c. The upper stable branch corresponding to high N_R^* is disconnected from the

¹³This is a classic case of a saddle-node bifurcation where two fixed points collide and disappear.

lower part in the N_S range of bistability ($0 \leq N_S \leq 12$). This means that a system in the region of high N_R^* (ON state; sink 1) has no way of switching to high N_A^* (OFF state; sink 2). Even if N_S decreased, the system would remain in the ON state. This SR curve and resulting system behavior is the hallmark of an irreversible (or one-way) switch.

Results of ABK simulations

We used the ABK algorithm to simulate the time evolution of this motif with two sets of initial population sizes: $N_{A,i} = 100$, $N_{R,i} = 0$ (which represent the switch effectively being in the OFF state), and $N_{A,i} = 10$, $N_{R,i} = 10$. We performed a separate simulation for each integer value of N_S in the range $[0, 20]$ for a total simulated time of 5000 sec.¹⁴ We estimated the steady-state levels N_A^* and N_R^* by averaging the population sizes over the last 1000 sec of the simulation (see relevant Matlab code in Appendix D on page 279). We show the resulting SR curve from the ABK simulations superimposed on the deterministic curve in Figure 3.19a. The open red circles or cyan cross markers and their error bars represent $\langle N_R^* \rangle \pm 1 \text{ } SD_{ev}$ ($n = 100$) for each of the two initial conditions. It is important to clarify that the $\langle N_R^* \rangle$ values in between the two stable branches (high and low values of N_R^* ; blue curves of the deterministic SR curve) do *not* mean that the populations reached some intermediate size, as the dynamics of the system in the bistable region dictate that the population trajectories reach either of the two sinks. The average is taken over an ensemble of simulation runs with individual switch outcomes being either in the ON or OFF state.¹⁵ Rather, these values reflect the fact that due to the stochastic nature of the processes, sink 1 (ON state; high N_R^*) can be reached with increasing probability as N_S increases. It is evident in the plot that for $3 \leq N_S \lesssim 14$, N_R^* can be found in either sink with a finite probability determined by the magnitude of N_S . For instance, when $N_S = 9$, $\langle N_R^* \rangle \approx 60 \pm 40$, which is at the halfway point between the N_R^* values corresponding to the two sinks and the error bars span the full range between them. A reasonable interpretation of this is that simulations of the system end up in the ON state approximately 50% of the time. Indeed, in

Follow this [link](#) to see the Matlab code.

¹⁴We increased the total simulation simulation time from 2000 sec (see preceding subsection) to 5000 sec because when fixed points collide and annihilate, they create a “bottleneck” where the trajectory flows very slowly through that region (some authors call this a “ghost node” or critical slowing down). The increased amount of simulated time ensures that the system has enough time to reach a stable fixed point.

¹⁵Imagine for a moment that a molecular realization of this motif occurs inside of a biological cell. Then, an experimenter taking a single measurement of N_R^* from an ensemble of cells (for instance, by coupling the motif’s response element to a fluorescent marker) would observe a smooth SR curve transition from one switch state to the other. In actuality, however, each measurement is the *average of distinct subsets of cells* that are in the ON or OFF state. This agrees with the observation of a bimodal distribution by Gardner et al. (2000, see Figure 5) when they experimentally constructed and tested this motif in bacterial cells. Other researchers have documented this effect when studying natural systems (e.g., see Ferrell and Machleder, 1998, for a characterization of the all-or-none cell fate switching in *Xenopus leavis* oocytes).

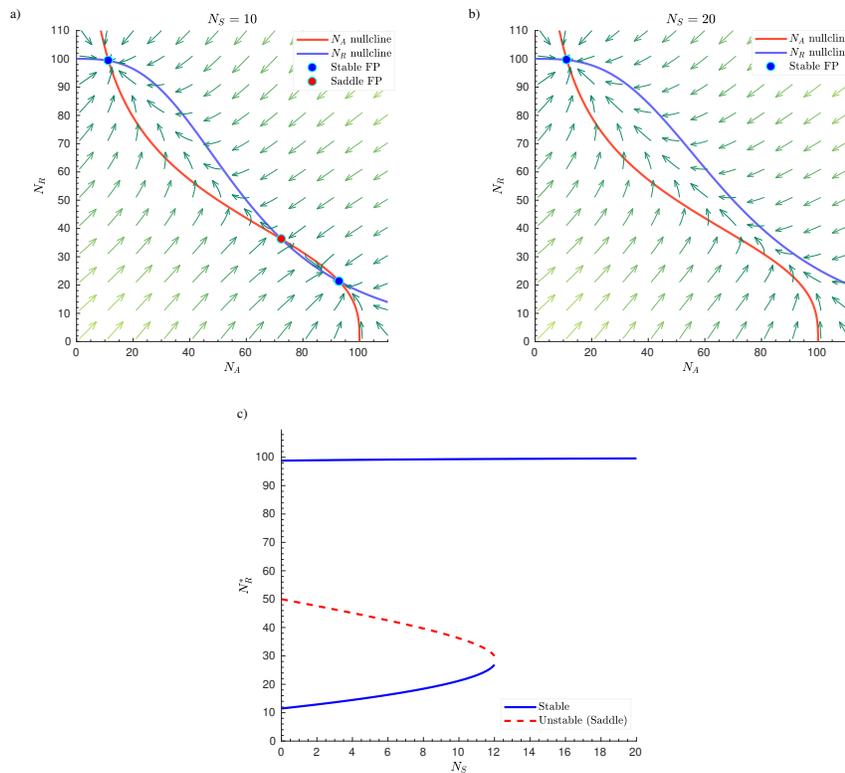


Figure 3.18 – Irreversible mutual inhibition switch motif: phase plane and SR curve. All parameters are the same as indicated in Figure 3.13b (which is the phase plane for this motif with $N_S = 0$), with the additional parameters $\alpha_s = 0$, $K_{50,S} = 50$ and $n_{H,S} = 1$. **a)** Phase plane for $N_S = 10$. Note that the saddle and stable fixed point on the lower right “move” toward each other as N_S increases. **b)** Phase plane for $N_S = 20$. The saddle and stable fixed point on the lower right of plot a have “collided” and disappeared. The system has now become monostable. **c)** SR curve (or, equivalently stated, bifurcation diagram with respect to N_S) showing the range of N_S values, $0 \leq N_S \leq 12$, for which the system is bistable. The red dashed curve corresponds to the saddle. Notice that the upper stable branch exists for all values of N_S and is disconnected from the lower unstable/stable region, therefore a system “caught” in the high N_R^* regime is unable to return to the lower one. This is the hallmark of an irreversible switch.

our simulations the system reached the ON state with a frequency of 52% when $N_S = 9$.

We plot the percentage of simulation runs that the system reached the ON state for each sampled value of N_S in Figure 3.19b. It is noteworthy that starting from the switch in the OFF state, the transition to ON starts earlier than predicted by the deterministic analysis, which predicts a sudden jump at $N_S = 12$ from OFF to ON. In the stochastic case, we see that the switch turns to the ON state in a substantial fraction of the simulation runs for $6 \leq N_S \leq 12$ when $N_{A,i} = 100$, $N_{R,i} = 0$, and for $0 \leq N_S \leq 8$ when $N_{A,i} = 10$, $N_{R,i} = 10$. This suggests that when stochastic fluctuations make the convergence on either sink possible, the probability that a given system switches to the ON state is nonzero for *smaller* N_S than the deterministic critical value at which the system becomes monostable (in this case, $N_S = 12$). It is apparent then that the initial population sizes markedly affect the transition to the ON state. In the example with initial populations $N_{A,i} = 10$, $N_{R,i} = 10$, the system is much more likely to end up in the ON state even when $N_S = 0$ (58% of the time in our simulations). Therefore, our work suggests that a system's transition from one switch state to another is strongly dependent on the initial state of the system (in accordance with our earlier observations on the signal-free mutual inhibition switch motif). The dependence can be qualitatively assessed by observing where on the state space the initial condition lies and how the direction field and proximity to the stable fixed points affects the possible stochastic trajectories of the system.

As previously shown in Figure 3.18, the presence of S leads to the saddle and fixed point corresponding to sink 2 getting closer together in the phase plane. Trajectories that are in the neighborhood of sink 2 can transition to sink 1 when stochastic fluctuations of sufficient magnitude allow the system to flow along the direction dictated by the saddle's unstable manifold, thus entering sink 1's basin of attraction. It is then reasonable to propose that such reversals will be more common than in the signal-free case (Figure 3.16) because smaller fluctuations will be needed given the saddle's increased proximity to sink 2. Indeed, we show an example of a noise-induced switch state reversal (sink 2 to 1) in Figure 3.20. Notice that the scale of simulated time needed to observe such a reversal is two orders of magnitude smaller than in the signal-free case (10^3 vs. 10^5 sec), suggesting a higher likelihood of a trajectory switching to sink 1's basin of attraction. The above hypothesis also suggests that sink 1 to 2 reversals will be correspondingly less frequent since the saddle is moving further away from sink 1 as N_S increases. The system would therefore have a bias toward settling in sink 1 as N_S increases within the range of deterministic bistability. This agrees with what we have shown to be the case in our simulations (Figure 3.19).

Becskei et al. (2001) have observed such a switch-state reversal in *Saccharomyces cerevisiae* cells while studying the behavior of a synthetic eukaryotic gene switch. The authors investigated the switching behavior of an artificial positive-feedback loop, unlike the mutual inhibition motif we have simulated here. Nonetheless, we remark here the experimental observation of a spontaneous transition to an alternate stable state of a real system. Interest-

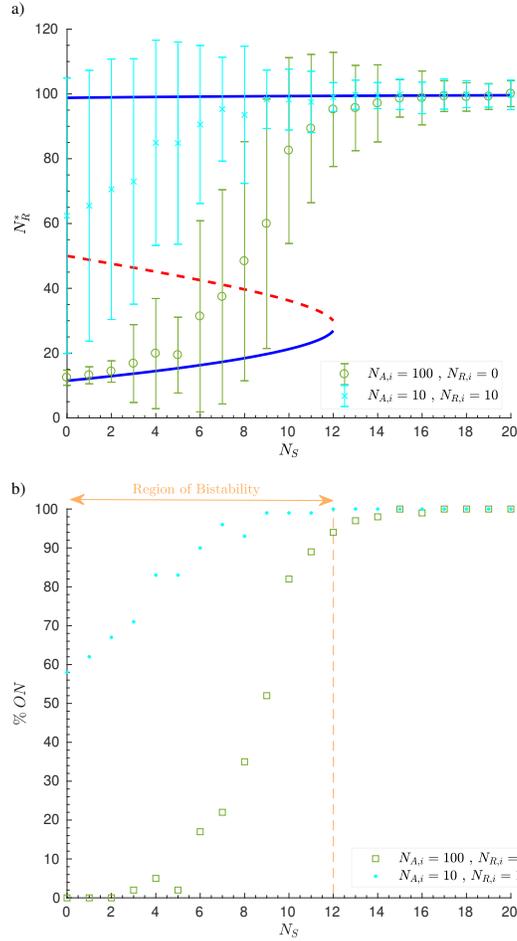


Figure 3.19 – Irreversible mutual inhibition switch motif: stochastic SR curve. We used the ABK algorithm to simulate the motif presented in Figure 3.17 with complete repression for all three inhibitory interactions and the following parameters: $k_a = k_r = 1 \text{ sec}^{-1}$, $\kappa_a = \kappa_r = 0.01 \text{ sec}^{-1}$, $K_{50} = K_{50,S} = 50$, $n_H = 3$, $n_{H,S} = 1$. We simulated the motif for 5000 sec with fixed time step increments of $\Delta t = 0.02 \text{ sec}$ and for integer values of the interval $0 \leq N_S \leq 20$. We then averaged the population sizes over the last 1000 sec to estimate N_R^* for each simulation. We repeated this procedure 100 times. We show the results of the simulation for two sets of initial population sizes: $N_{A,i} = 100, N_{R,i} = 0$ (i.e., the switch effectively being in the OFF state) and $N_{A,i} = 10, N_{R,i} = 10$. **a)** SR curve for the results of the ABK simulation superimposed on the deterministic bifurcation diagram (3.19c). The cyan crosses and green circles represent the average steady-state values for the population of R, $\langle N_R^* \rangle$, for the two different initial conditions (each half of the error bars is one standard deviation from the mean). **b)** We show the frequency of the system being in the ON state for each N_S value. The vertical orange dashed line shows the critical N_S value at which the deterministically-predicted region of bistability ends.

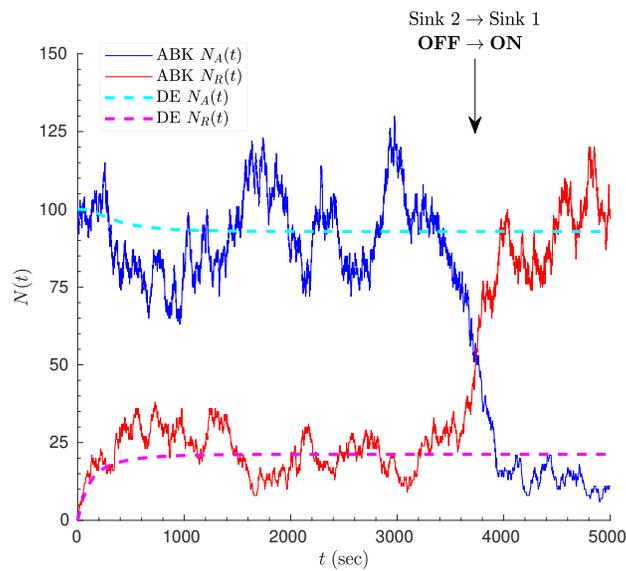


Figure 3.20 – Irreversible mutual inhibition switch motif: noise-induced switch state reversal. We used the ABK method to simulate the mutual inhibition switch motif with the same kinetic and inhibition parameters as in Figure 3.19 and $N_S = 10$. The initial population sizes, $N_{A,i} = 100$ and $N_{R,i} = 0$, are in sink 2's basin of attraction. We recorded the population time trajectories of species A and R for a total of 5000 sec of simulated time. The system quickly settles around sink 2, but a noise-induced switch state reversal occurs in the time window of 3400 – 3800 sec. Once again, sinks 1 and 2 represent the ON (high N_R) and OFF (low N_R) switch state respectively. We also show the deterministic time trajectories (dashed curves) that converge on sink 2 (OFF state) given the initial condition.

ingly, Becskei et al. only observed such transitions *to* the ON state, as we predict would be the case for our simulated system as N_S increases within the region of bistability ($0 \leq N_S \leq 12$).

We will not further elaborate here on the dynamics of this motif, other than to summarize that the ABK algorithm provides an easy and convenient way of understanding how real populations (made of a discrete number of agents/-molecules) behave in ways that deviate from deterministic predictions. Thus, an ensemble of ABK simulations can be used to obtain statistical measures of the system's possible time trajectory, thereby making probabilistic predictions about its behavior. This capability will be especially important when considering the presence of heterogeneity in species populations.

Conversion to a reversible switch

The example switch we presented above is not exceptionally robust with respect to its irreversibility. We briefly present here the result of modifying just one of the system’s parameters so that it is converted from an irreversible to a reversible switch.

We consider the same motif (Figure 3.17) with the following parameters: $k_a = k_r = 1 \text{ sec}^{-1}$, $\kappa_a = 0.005 \text{ sec}^{-1}$, $\kappa_r = 0.01 \text{ sec}^{-1}$, $K_{50} = K_{50,S} = 50$, $n_H = 3$, $n_{H,S} = 1$. This is the exact same set of parameters as used in the experiments described in the preceding subsection, except κ_a is half of its previous value and the degradation rates of the two species are no longer equal. Species A is degraded more slowly than R and has a chance to accumulate, thus opposing (or at least slowing down) the transition to the ON state (high N_R^*). The consequence of the parameter change is that the switch remains discontinuous, but big jumps in population sizes are expected for *two* critical values of N_S and the switch becomes reversible. That is, in addition to turning the system OFF→ON as N_S increases, a system in the ON state *can* switch back to OFF when N_S decreases sufficiently. This behavior is summarized by the deterministic bifurcation diagram presented in Figure 3.21a (colored in the same way as in Figure 3.19a). The region of bistability is now $11 \leq N_S \leq 74$ and the switch can be turned ON as N_S surpasses the critical population size of 74, but can also be turned OFF when N_S decreases beyond 11 and the system “falls off” the left edge of the upper set of stable fixed points (upper part of blue curve).

We have simulated this system using the ABK algorithm for the same two sets of initial population sizes as before: $N_{A,i} = 100$, $N_{R,i} = 0$, and $N_{A,i} = 10$, $N_{R,i} = 10$. We show the results of the simulations in Figure 3.21. Our observations and comments are similar to the ones for the irreversible switch case.

The above case study suggests that altering the rate of just one of the processes can have a profound effect on the dynamics of a hysteretic switch. A potential application of this is the use of a pharmacological intervention (such a competitive inhibitor to a suitably chosen enzyme) that renders a switch reversible, thus allowing for the reversal of a pathogenic state. This might be useful, for instance, in controlling the reversibility of cell cycle checkpoint switches whose aberrant regulation can lead to cancerous phenotypes (Kolch et al., 2015). Thus, the selective manipulation of a system’s dynamical properties can be a general strategy for system control with important applications in many fields.

Alternate mutual inhibition motif

The results we have presented in subsections 3.3.1 and 3.3.2 are based on motifs of mutual inhibition where the regulatory interactions are directly effected from one species to another. Thus, we used the RSF to mathematically describe the mode of regulation in those cases. Here, we present an alternate version of this motif that also leads to behavior characteristic of a hysteretic switch. In this version, inhibition is brought about by promoting the degradation or conver-

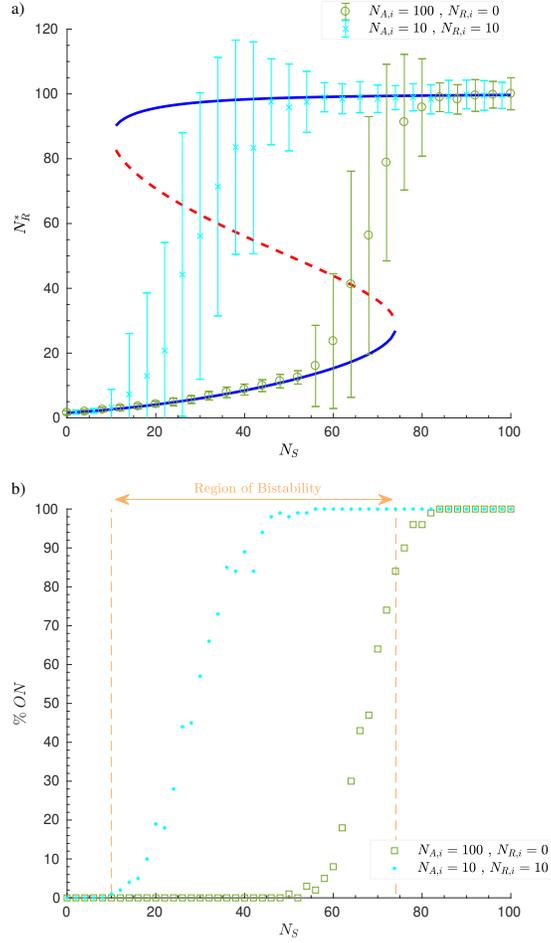


Figure 3.21 – Reversible mutual inhibition switch motif: stochastic SR curve. We used the ABK algorithm to simulate the motif presented in Figure 3.17 and assumed complete repression for all three regulatory interactions. Additional parameters: $k_a = k_r = 1 \text{ sec}^{-1}$, $\kappa_a = 0.005 \text{ sec}^{-1}$, $\kappa_r = 0.01 \text{ sec}^{-1}$, $K_{50} = K_{50,S} = 50$, $n_H = 3$, $n_{H,S} = 1$. We simulated the motif for 2500 sec with fixed time step increments of $\Delta t = 0.02 \text{ sec}$ and for integer values of the interval $0 \leq N_S \leq 100$. We then averaged the population sizes over the last 1000 sec to obtain an estimate for N_R^* in each simulation. We repeated this procedure 100 times. We show the results of the simulations for two sets of initial population sizes: $N_{A,i} = 100$, $N_{R,i} = 0$ (i.e., the switch effectively being in the OFF state) and $N_{A,i} = 10$, $N_{R,i} = 10$. **a)** SR curve for the results of the ABK simulation superimposed on the deterministic bifurcation diagram. The cyan crosses and green circles are the average steady-state values for the population of R , $\langle N_R^* \rangle$, for the two different initial conditions (each half of the error bars is one standard deviation from the mean). **b)** We plot the frequency of the system being in the ON state for each sampled N_S value. The vertical orange dashed lines show the N_S values that define the deterministically-predicted range of bistability.

sion of a species to another. Specifically, R promotes/catalyzes the conversion of species E to Ep (equivalent to a phosphorylation reaction in biological processes), thereby effectively inhibiting E by keeping its population low. In return, E promotes the degradation of R , thus completing the inhibitory loop.

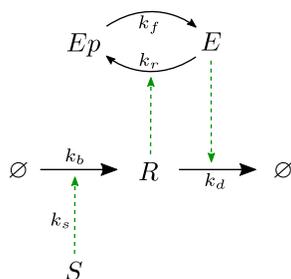


Figure 3.22 – Alternate mutual inhibition switch motif. Species E is reversibly converted to Ep in reactions following Michaelis-Menten kinetics (with kinetic rate constants k_f and k_r , and Michaelis constants $K_{M,f}$ and $K_{M,r}$). R promotes the conversion of E to Ep , but is not consumed in the reaction. Therefore, R effectively inhibits E . In turn, E enhances/catalyzes the degradation rate of R , thus forming a mutual inhibition loop.

We do not show here a mathematical analysis and ABK simulation of this motif, as we have already discussed the range of possible system behaviors in our previous examples.

3.3.3 Hysteretic Switch from Mutual Activation

Here, we present an additional example of a hysteretic, reversible switch based on mutual activation (instead of inhibition). Consider the motif in Figure 3.23. The signal S promotes the synthesis of R and species E is reversibly phosphorylated to form Ep in a Michaelis-Menten fashion,¹⁶ a reaction that is promoted by the response element R . In turn, Ep enhances the synthesis of R , thus forming a mutual activation loop. Essentially, R stimulates its own production, thus forming an *autocatalytic* or positive feedback loop.

¹⁶The interconversion between species E and Ep needs to follow Michaelis-Menten kinetics to observe the dynamics we will describe here. A reversible conversion between E and Ep that follows the law of mass action (as in section 2.10) yields a linear response motif, a case we have already investigated and will not present here.

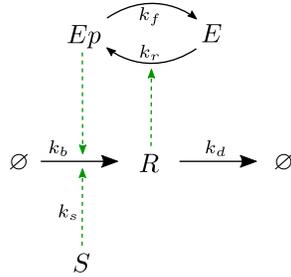


Figure 3.23 – Mutual activation switch motif. Species E is reversibly converted to Ep in reactions following Michaelis-Menten kinetics (with kinetic rate constants k_f and k_r , and Michaelis constants $K_{M,f}$ and $K_{M,r}$). R promotes the conversion of E to Ep , but is not consumed in the reaction. Moreover, Ep enhances the rate of formation of R , thus forming a mutual activation loop. Kinetic parameters for species R : k_b and k_s are the 0th order basal and maximal (in the presence of the signal S) rate constants for the formation of R , respectively, and k_d is the 1st order degradation rate constant. We assume no further synthesis and degradation reactions for E and Ep (therefore, $N_E + N_{Ep} = \text{constant} = N_{E,TOT}$) and that the population size of S remains constant.

The differential equations used to describe this motif are

$$\frac{dN_R}{dt} = (k_b N_{Ep} + k_s N_S) - k_d N_R \quad (3.23)$$

$$\frac{dN_{Ep}}{dt} = k_r N_R \frac{N_{E,TOT} - N_{Ep}}{K_{M,r} + (N_{E,TOT} - N_{Ep})} - k_f \frac{N_{Ep}}{K_{M,f} + N_{Ep}}, \quad (3.24)$$

where in the first equation we have grouped the terms corresponding to the synthesis of R in parentheses, and in the second we used $N_E + N_{Ep} = \text{constant} = N_{E,TOT}$. As we have previously done with hysteretic switches, we illustrate the deterministic dynamics that this motif can exhibit by plotting the phase plane for a particular value of N_S that falls within the region of bistability (Figure 3.24). The saddle and stable fixed point on the lower left corner of the phase plane (switch state: OFF) get closer as N_S increases and eventually collide with each other at $N_S \approx 19$. The system becomes monostable beyond this threshold and the only remaining attractor is the fixed point on the right side of the figure (switch state: ON). Notice that a transition from OFF to ON results in higher population sizes for *both* species R and Ep .

We succinctly present the results obtained from applying the ABK algorithm to this motif by plotting the deterministic bifurcation diagram (with respect to N_S) and superimposing the stochastic simulation results for two different initial population sizes (Figure 3.25). We obtained similar results to the case of a reversible switch obtained through mutual inhibition (Figure 3.21). We conclude our presentation of this motif by showing an example of noise-induced switch state reversals in Figure 3.26. Our previous comments and observations on such transitions (Figures 3.16 and 3.20) apply here as well.

We further note that just as in the previous examples we have considered, changes in the kinetic parameters of this system can result in the switch becoming irreversible. Mutual inhibition and activation motifs are therefore capable

Follow this [link](#) to see the Matlab code.

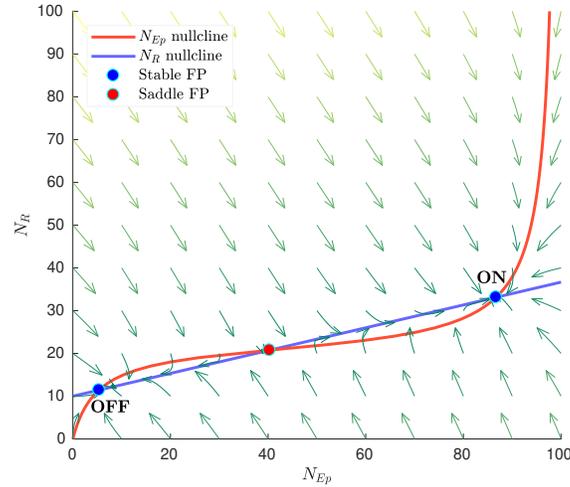


Figure 3.24 – Mutual activation switch motif: phase plane. We consider the mutual activation motif with the following parameters: $k_b = 0.02 \text{ sec}^{-1}$, $k_s = 0.05 \text{ sec}^{-1}$, $k_d = 0.075 \text{ sec}^{-1}$, $k_f = 1 \text{ sec}^{-1}$, $k_r = 0.05 \text{ sec}^{-1}$, $K_{M,f} = 5$, $K_{M,r} = 10$, $N_{E,TOT} = 100$. Here we show the phase plane for $N_S = 15$. Note that the saddle and stable fixed point on the lower left “move” toward one another (and eventually annihilate each other) as N_S increases. The switch states are: ON (high N_R and N_{Ep}) and OFF (low N_R and N_{Ep}), as indicated in the plot.

of producing a hysteretic switch response, whose (ir-)reversibility, viewed as a property of the system, depends on the specific values of the kinetic parameters characterizing the relevant processes. Importantly, we will revisit the mutual activation switch motif in Chapter 4 (section 4.10), where we will examine the effect of population heterogeneity on hysteretic switches.

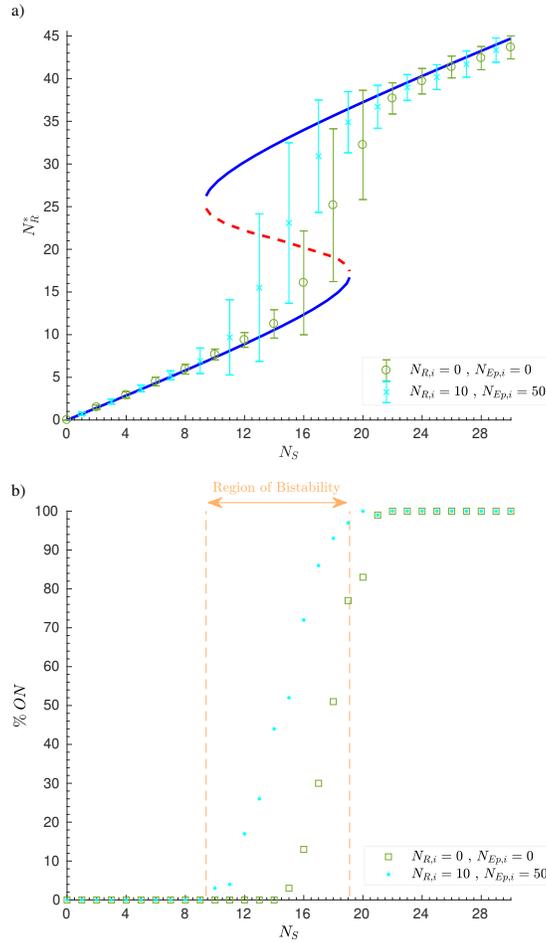


Figure 3.25 – Mutual activation switch motif: stochastic SR curve. We used the ABK algorithm to simulate the motif presented in Figure 3.23 with the same parameters as listed in Figure 3.24, and integer values of the signal in the interval $0 \leq N_S \leq 30$. The total simulated time was 1500 sec with fixed time step increments of $\Delta t = 0.02$ sec. We averaged the population sizes over the last 750 sec to estimate N_R^* for each simulation. We repeated this procedure 100 times. We show the results of the simulation for two sets of initial population sizes: $N_{R,i} = 0, N_{Ep,i} = 0$ (i.e., the switch starts out being in the OFF state), and $N_{R,i} = 10, N_{Ep,i} = 50$. **a)** SR curve for the results of the ABK simulation superimposed on the deterministic bifurcation diagram. The cyan crosses and green circles are the average steady-state values for the population of R , $\langle N_R^* \rangle$, for the two different initial conditions (each half of the error bars is one standard deviation from the mean). **b)** We plot the frequency of the system being in the ON state as a function of N_S . The vertical orange dashed lines show the range of N_S values for which the system is predicted to be bistable.

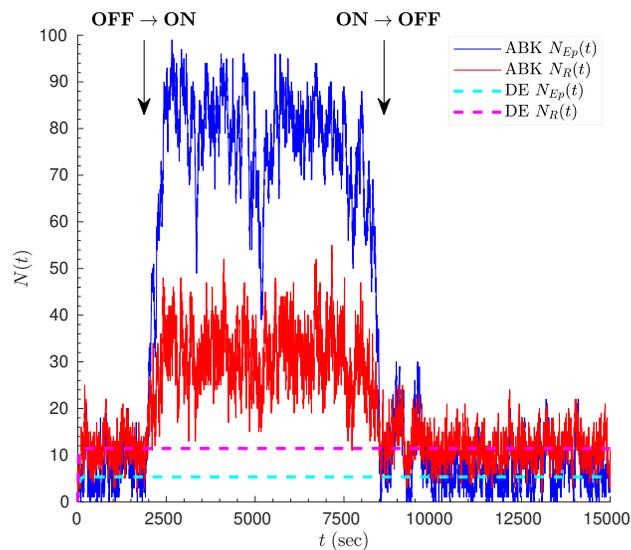


Figure 3.26 – Mutual activation switch motif: noise-induced switch state reversal. We used the ABK method to simulate the mutual inhibition switch motif with the same kinetic and inhibition parameters as in Figure 3.24. The switch states are: ON (high N_R and N_{Ep}) and OFF (low N_R and N_{Ep}). We recorded the population time trajectories of species Ep and R for a total of 1.5×10^4 sec of simulated time. The initial population sizes are in the basin of attraction of state OFF: $N_{Ep,i} = 0$ and $N_{R,i} = 0$. The system initially settles in the neighborhood of state OFF, but two noise-induced switch state reversals occur at the indicated time points, thus eventually returning the system to the OFF state. We also show the deterministic trajectories (dashed curves) which converge on the OFF state given the initial condition.

3.4 Adaptation and Homeostasis

Systems sometimes need to be able to resist changes due to varying levels of a signal. We explore how this is made possible through prototypical motifs that exhibit the phenomena of *adaptation* and *homeostasis*. In both cases, the steady-state level of response is largely — if not completely — independent of the presence of signal, while the difference between them is that there is a transient response to the signal level in the adaptive case.

3.4.1 Perfect Adaptive Response Motif

We build on the linear response motif (presented in subsection 3.1.1) to examine the case where changes in the level of stimulus produce a transient response that invariably returns to a signal-independent steady state. We refer to this response mechanism as *perfect adaptive response*. Chemotactic systems are the best-studied biological examples of this type of response.

In the linear response motif, a signal S activates the production of R . In this case, we maintain this relationship between S and R , but add a second pathway: S also promotes the production of species X , which in turn promotes the degradation of R . The following figure shows the reaction scheme we will consider here.

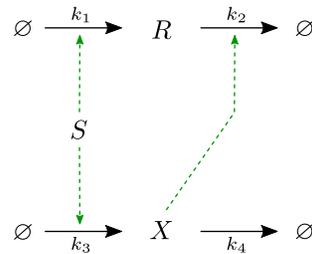


Figure 3.27 – Perfect adaptive response motif. k_1 and k_3 are the microscopic rate constants for the synthesis of R and X , while k_2 and k_4 are the rate constants for the degradation of R and X , respectively. We assume that S promotes, or favors, both syntheses reactions and that the number of S agents does not change. X promotes the degradation of R .

Our intuitive expectation is that increasing levels of S increase the number of R and X agents (at a rate proportional to k_1 and k_3 , respectively), but X increases R 's degradation rate thereby not allowing the population of R to grow over long time scales. Notice that S regulates R in two different ways: a) directly, by stimulating its production, and b) indirectly, by stimulating the production of X which inhibits R by promoting its degradation. Therefore, S can be thought of as regulating R via two distinct pathways with opposite regulatory effects. This topology of interactions is sometimes referred to in the literature as an *incoherent feedforward loop* (iFFL; Ferrell, 2016).

k_1 and k_3 are the 1st order microscopic rate constants for the production of R and X . We have seen before that such “birth” processes can also be modeled

as 0th order in the ABK algorithm because we assume that S is not consumed while promoting these syntheses reactions (alternatively, S can be thought of as being regenerated during each reaction event). k_2 is the 2nd order microscopic rate constant for the degradation of R , but we assume X is *not* consumed during this process (or is regenerated). Finally, k_4 is the 1st order rate constant for the degradation of X .

We now state the differential equations describing this reaction scheme and then examine the deterministic predictions for the steady-state level of R .

$$\frac{dN_R}{dt} = k_1 N_S - k_2 N_X N_R \quad (3.25)$$

$$\frac{dN_X}{dt} = k_3 N_S - k_4 N_X \quad (3.26)$$

We find the steady-state value of the population of R , N_R^* , by setting both of the above equations equal to zero and combining them to obtain

$$N_R^* = \frac{k_1 k_4}{k_2 k_3} \quad (3.27)$$

$$N_X^* = \frac{k_3 N_S}{k_4} \quad (3.28)$$

This reveals the crucial property of this motif: the steady-state value for the population of R does *not* depend on the signal population, N_S .

We present the results of using the ABK algorithm to model the time evolution of the populations of R and X in a system with kinetic constants $k_1 = 0.500 \text{ sec}^{-1}$, $k_2 = 0.005 \text{ sec}^{-1}$, $k_3 = 0.100 \text{ sec}^{-1}$, $k_4 = 0.050 \text{ sec}^{-1}$, and initial populations $N_{R,i} = N_{X,i} = 0$. First, we graph the rate-balance plot showing how the rates of synthesis and degradation of R (first and second terms of equation 3.25, respectively) depend linearly on the population size of R (Figure 3.28). The intersections of the synthesis and degradation curves represent the steady state of the system. It is clear in the figure that the steady states produced for the four different values of N_S are predicted to be the same ($N_R^* = 50$), in agreement with the expectation that N_R^* does not depend on N_S .

We then used the ABK algorithm to model this system for a total simulated time of 300 sec with fixed time step increments $\Delta t = 0.01 \text{ sec}$, and with N_S increased in a step-wise manner by 10 agents every 60 sec (sampled N_S values: 15, 25, 35, 45). In the last step, we maintained $N_S = 45$ for the final 120 sec of the simulation. Figure 3.29a shows that the average time trajectories of N_R and N_X after 500 repetitions of the algorithm are in excellent agreement with the numerical solution to the differential equations 3.25 and 3.26.

Follow this [link](#) to see the Matlab code.

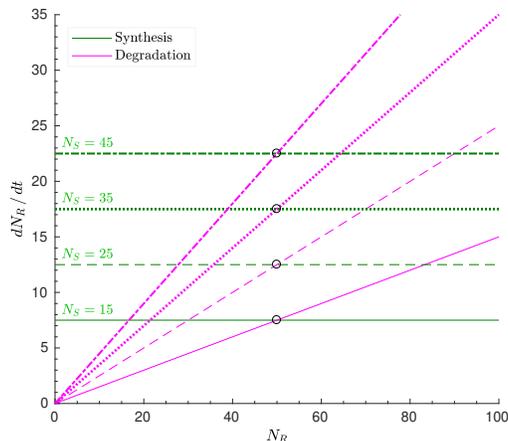


Figure 3.28 – Perfect adaptive response motif: rate-balance plot. We graph the rates of N_R synthesis ($dN_R/dt = k_1 N_S$; green curves) and N_R degradation ($dN_R/dt = k_2 N_X N_R$; magenta curves) as a function of its population size, for four different signal values: $N_S = 15$ (solid curves), 25 (dashed curves), 35 (dotted curves), 45 (dot-dashed curves). We generated the degradation curves by using the steady-state value of N_X , $N_X^* = \frac{k_3 N_S}{k_4}$. The intersections (marked by black open circles) represent the steady state of the population of R , N_R^* . Kinetic parameters: $k_1 = 0.500 \text{ sec}^{-1}$, $k_2 = 0.005 \text{ sec}^{-1}$, $k_3 = 0.100 \text{ sec}^{-1}$, $k_4 = 0.050 \text{ sec}^{-1}$. Given these parameter values, the predicted steady-state response is $N_R^* = 50$.

Perfect adaptive response from hyperbolic response motif

We briefly show another way to obtain a perfect adaptive response that utilizes the motif producing a hyperbolic SR curve (Figure 3.3). This new motif has S promoting the production of R and species X , the latter promoting the conversion of R back to A (see Figure 3.30 below), thus leading to adaptation. The reader will notice that the topology of interspecies interactions is the same as in the case we presented above, thus this motif is an iFFL.

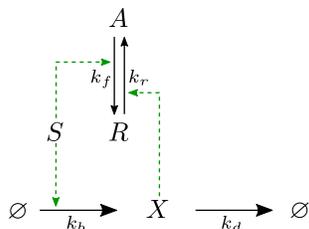


Figure 3.30 – Perfect adaptive response: alternate motif. We supplement the hyperbolic response motif (Figure 3.3) with a pathway involving species X . S promotes the synthesis of both R and X , but X promotes the conversion of R back to A . We assume that the population of S does not change. In addition, X is not consumed during the conversion $R \rightarrow A$.

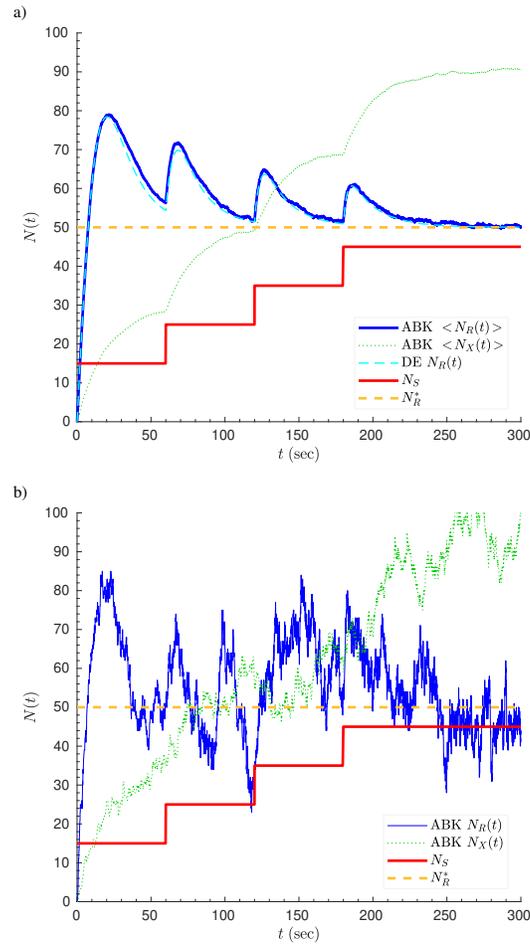


Figure 3.29 – Perfect adaptive response motif: time course. We used the ABK algorithm to model the time evolution of the motif shown in Figure 3.27. All kinetic constants are as indicated in Figure 3.28 with initial populations $N_{R,i} = N_{X,i} = 0$. For all simulations, we used fixed time step increments of $\Delta t = 0.01$ sec and the P_{ber} or P_{dif} transition probability expressions (we obtained identical results using both approaches). **a)** Average population sizes from 500 repetitions of the simulation. The deterministic time course for N_R (cyan dashed curve) is also shown. The orange dashed line is the predicted steady state $N_R^* = 50$. **b)** We show a sample run of the ABK-simulated population time trajectories.

This motif is a 2-dimensional system with variables N_R and N_X because the populations of A and R are related by $N_A(t) + N_R(t) = N_{TOT} = N_{A,i} + N_{R,i}$ (we have assumed the absence of any additional synthesis and degradation reactions). Therefore, the differential equations for this system are

$$\begin{aligned}\frac{dN_R}{dt} &= k_f N_S N_A - k_r N_X N_R \\ \frac{dN_X}{dt} &= k_f N_S (N_{TOT} - N_R) - k_r N_X N_R,\end{aligned}\quad (3.29)$$

and

$$\frac{dN_X}{dt} = k_b N_S - k_d N_X. \quad (3.30)$$

The reader can confirm that the deterministic prediction for the steady-state value of the population size of R is

$$N_R^* = \frac{\rho}{1 + \rho} N_{TOT}, \quad (3.31)$$

where

$$\rho = \frac{k_f k_d}{k_r k_b}. \quad (3.32)$$

The important thing to note about N_R^* is that it once again does *not* depend on N_S . Therefore, despite an initial transient change, N_R approaches its steady-state level. We show an ABK simulation sample run of this motif in Figure 3.31.

Follow this link to see the Matlab code.

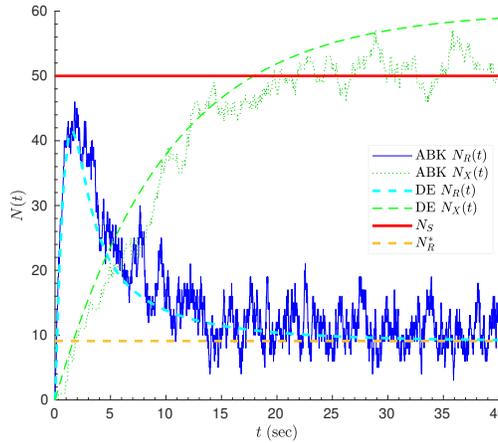


Figure 3.31 – Perfect adaptive response: alternative motif time course. We simulated the motif presented in Figure 3.30 using the ABK algorithm with the following parameters: kinetic constants $k_b = 0.10 \text{ sec}^{-1}$, $k_d = 0.10 \text{ sec}^{-1}$, $k_f = 0.01 \text{ sec}^{-1}$, $k_r = 0.10 \text{ sec}^{-1}$; initial populations $N_{A,i} = 100$, $N_{R,i} = 0$ (i.e., $N_{TOT} = 100$), $N_{X,i} = 0$; $N_S = 50$ (held constant). The motif was simulated for a total time of 40sec with fixed time step increments of 0.01 sec. Given these parameters, $N_R^* \approx 9$ agents. The dashed curves show the deterministic predictions for each of the two species. Also note that the average trajectories from an ensemble of ABK simulations agreed with deterministic predictions (not shown).

We have also attempted simulating the same motif but with additional processes accounting for the synthesis and degradation of A and R . We again observed perfect adaptation (results not shown).

3.4.2 Imperfect Adaptive Response Motif

We considered the case of perfect adaptive response where N_R always returns to the same level regardless of N_S . Here we examine a similar motif exhibiting adaptation, but with N_R^* having a dependence on the magnitude of N_S . We therefore expect a transient change in the response (caused by a change in the level of signal) to be followed by a return to a new and different basal steady-state level. We refer to this adaptive response as *imperfect*.

In the reaction scheme we presented in Figure 3.27, S promotes the synthesis of both R and X , with X promoting the degradation of R . Therefore, X is indirectly inhibiting the growth of the population of R . This suggests that we can modify the motif and model the *explicit* inhibition, or repression, of the formation of R by X , as shown in Figure 3.32 below.

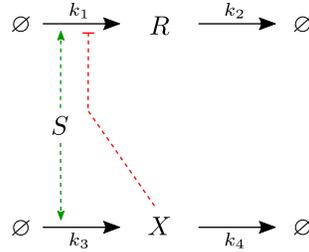


Figure 3.32 – Imperfect adaptive response motif. S promotes, or favors, both synthesis reactions and the number of S agents does not change. X inhibits the synthesis of R (red arrow) and we assume that the number of X agents does *not* change as a result of this regulatory activity. All kinetic rate constants are as described in Figure 3.27, except k_2 is now a 1st order constant for the degradation of R .

We use an appropriate regulation scaling Function (RSF, as described in section 2.9) to model the X -mediated inhibition of the 0th order synthesis of R . The RSF has the general form (equation 2.148)

$$F(N_X) = \frac{K_{50}^{n_H} + \alpha N_X^{n_H}}{K_{50}^{n_H} + N_X^{n_H}} = \frac{1 + \alpha \left(\frac{N_X}{K_{50}}\right)^{n_H}}{1 + \left(\frac{N_X}{K_{50}}\right)^{n_H}}.$$

We will assume for the purposes of this discussion that $\alpha = 0$ (complete repression). K_{50} is the number of agents of X that produce half-maximal repression and n_H is the Hill coefficient indicating the degree of cooperativity of the process.

We begin by stating the differential equations describing this motif.

$$\frac{dN_R}{dt} = k_1 N_S \left[1 + \left(\frac{N_X}{K_{50}} \right)^{n_H} \right]^{-1} - k_2 N_R \quad (3.33)$$

$$\frac{dN_X}{dt} = k_3 N_S - k_4 N_X \quad (3.34)$$

Setting these equations equal to zero to evaluate the fixed point of the system and combining them to obtain expressions only in terms of N_S , we get

$$N_R^* = \left(\frac{k_1}{k_2} \right) \frac{N_S}{1 + \left(\frac{k_3}{k_4 K_{50}} N_S \right)^{n_H}} \quad (3.35)$$

$$N_X^* = \frac{k_3 N_S}{k_4}. \quad (3.36)$$

Notice that N_X^* has the same form as in the case of perfect adaptation. However, equation 3.35 shows a clear dependence of N_R^* on N_S , and the exact form of the signal-response (SR) curve depends on the Hill coefficient n_H (see Figure 3.33). In the limiting case of $n_H = 0$, the SR curve is linear, as one would expect since this choice of n_H is indicative of no regulatory cross-talk between the two pathways (in essence, the motif becomes the separate birth-death processes for R and X , respectively). For $n_H = 1$, the SR curve is hyperbolic and approaches a maximum value of $\frac{k_1 k_4}{k_2 k_3} K_{50}$ as $N_S \rightarrow \infty$. In the cases where $n_H > 1$, $N_R^* \rightarrow 0$ as $N_S \rightarrow \infty$, as we would expect since the increased cooperativity of inhibition eventually shuts down the formation of R completely. This end behavior is preceded, however, by an initial increase in N_R^* , with a maximum, denoted by $N_{R,max}^*$, occurring at

$$N_S = (n_H - 1)^{-1/n_H} \left(\frac{k_4 K_{50}}{k_3} \right), \quad \{n_H > 1\}, \quad (3.37)$$

which upon substitution into equation 3.35 yields the $N_{R,max}^*$ value of

$$N_{R,max}^* = \frac{k_1 k_4}{k_2 k_3} K_{50} \left[\frac{(n_H - 1)^{\frac{n_H - 1}{n_H}}}{n_H} \right]. \quad (3.38)$$

We have highlighted the maximum N_R^* values for $n_H = 2, 3, 4$ (filled orange circles in Figure 3.33). We have also plotted $N_{R,max}^*$ for $n_H = 1.1 \rightarrow 100$ ($n_H = 100$ effectively represents infinite cooperativity), shown as an orange dashed line in the same figure. Notice that this curve consists of ordered pairs $(N_S, N_{R,max}^*)$, where each quantity is a function of n_H . This curve is therefore a parametric representation of the indicated variables with respect to n_H , where the “flow” through the curve is initially fast and then slows down as $n_H \rightarrow \infty$. We invite the interested reader to continue with the mathematical analysis of this curve’s behavior, which we do not present here in the interest of brevity. We

Follow this [link](#) to see the Matlab code.

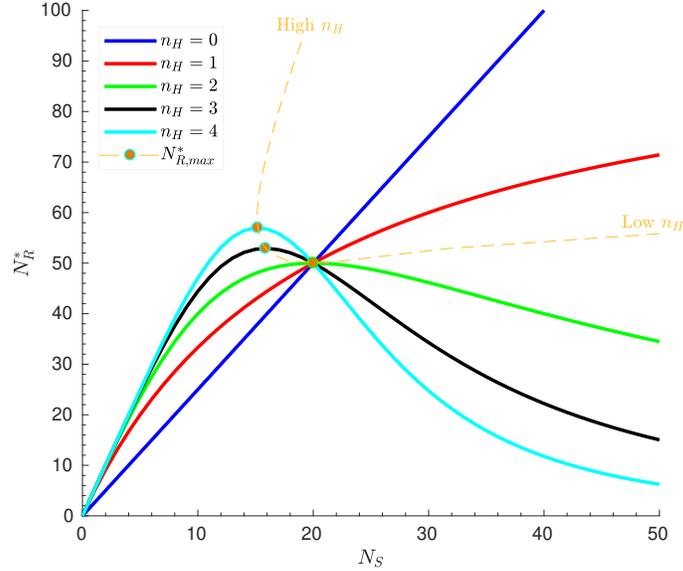


Figure 3.33 – Imperfect adaptive response motif: predicted SR curve. We show the deterministically predicted SR curves for different values of n_H (equation 3.35). The curves were generated using the motif presented in this section (Figure 3.32) with kinetic constants $k_1 = 0.50 \text{ sec}^{-1}$, $k_2 = 0.10 \text{ sec}^{-1}$, $k_3 = 0.10 \text{ sec}^{-1}$, $k_4 = 0.05 \text{ sec}^{-1}$, $K_{50} = 40$, and the indicated n_H values. The maximal steady-state response, $N_{R,max}^*$, is also plotted for a range of Hill coefficient values, $n_H = 1.1 \rightarrow 100$ (orange dashed curve), with the lower values of n_H starting on the right side of the plot and passing by the N_R^* maxima corresponding to $n_H = 2, 3, 4$ (filled orange circles).

merely mention the interesting case where the smallest value of $N_{R,max}^*$ occurs at $n_H = 2$.

As we did in the case of perfect adaptation, we simulated the time evolution of this motif using the ABK algorithm. We show the results of these simulations for $n_H = 1$ and $n_H = 2$ in Figure 3.34. As in the case of perfect adaptation, the average species trajectories agreed with the deterministic expectations.

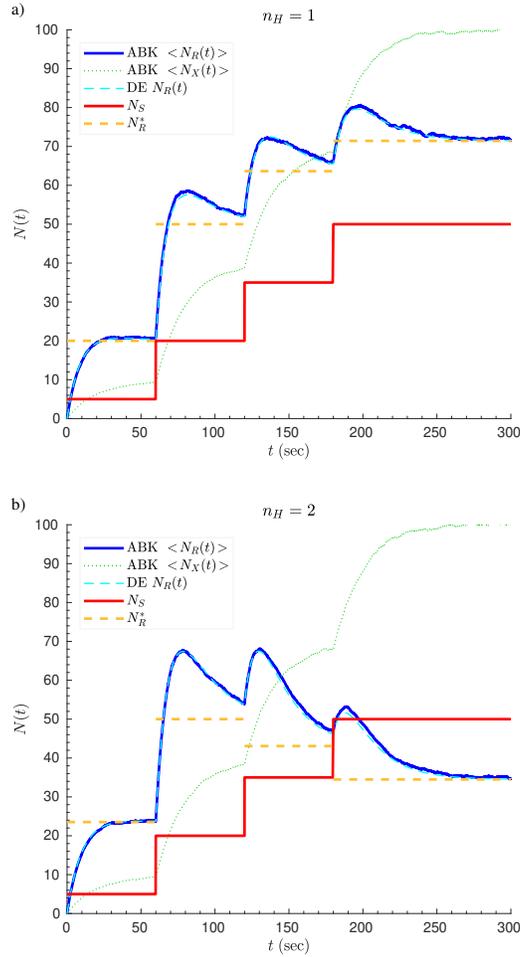


Figure 3.34 – Imperfect adaptive response motif: time course. We used the ABK algorithm to model the time evolution of the motif shown in Figure 3.32 with kinetic rate constants $k_1 = 0.50 \text{ sec}^{-1}$, $k_2 = 0.10 \text{ sec}^{-1}$, $k_3 = 0.10 \text{ sec}^{-1}$, $k_4 = 0.05 \text{ sec}^{-1}$, $K_{50} = 40$, and initial populations $N_{R,i} = N_{X,i} = 0$. We ramped up the signal population size (red line) at 60 sec intervals (N_S values: 5, 20, 35, 50). We used fixed time step increments of $\Delta t = 0.01 \text{ sec}$ for all simulations, and the P_{ber} or P_{dif} transition probability expressions (we obtained identical results using both approaches). **a)** Average population sizes from 500 repetitions of the simulation with Hill coefficient $n_H = 1$. The deterministic time course for the population of R (cyan dashed curve) is also shown. The orange dashed lines are the predicted steady-state N_R^* values for the time period corresponding to a given N_S value. **b)** Same as in **a**, except we ran the simulations with $n_H = 2$.

3.4.3 Homeostasis

We present here a motif exhibiting *homeostasis*, where the steady-state response level is confined to a small range of values for a wide range of signal levels (i.e., a largely signal-independent steady state). Moreover, this motif's behavior is different from adaptation because it shows no transient response to a change in signal levels.

We show the homeostasis motif in Figure 3.35. Its key feature is a negative feedback loop where species E promotes the synthesis of R , while R promotes the conversion of E into Ep (e.g., a phosphorylation reaction). In this case, the signal S promotes (or catalyzes) the degradation of R . Therefore, one would expect that as S increases the population of R would decrease in size, which in turn allows E to accumulate and, therefore, R to rebound. Thus, the negative feedback loop ensures the level of R stays approximately the same.

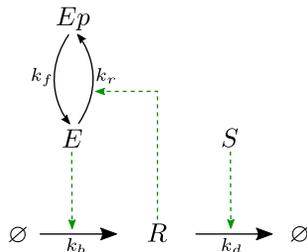


Figure 3.35 – Homeostatic response motif. R promotes the conversion of E to Ep , while E promotes the synthesis of R . This antagonistic relationship between R and E is a (two-component) negative feedback loop. S favors the degradation of R . We assume that a) no further synthesis and degradation processes occur for species E and Ep , thus $N_E(t) + N_{Ep}(t) = \text{constant} = N_{E,TOT}$, and b) the population size of S stays constant.

The homeostatic response motif is described by the 2-dimensional set of differential equations,

$$\frac{dN_R}{dt} = k_b N_E - k_d N_S N_R \quad (3.39)$$

$$\frac{dN_E}{dt} = k_f \frac{N_{E,TOT} - N_E}{K_{M,f} + (N_{E,TOT} - N_E)} - k_r N_R \frac{N_E}{K_{M,r} + N_E}, \quad (3.40)$$

where we have assumed that the conversion between E and Ep follows Michaelis-Menten kinetics and used the restriction $N_E + N_{Ep} = N_{E,TOT}$ to substitute for N_{Ep} in equation 3.40. Setting both equations equal to zero allows us to obtain expressions for the steady-state values N_R^* and N_E^* (although combining them to derive an expression for N_R^* in terms of N_S is algebraically cumbersome). We numerically determine N_R^* for different values of N_S to graph the SR curve in Figure 3.36b (shown in blue). A qualitative way of assessing what the SR curve looks like is by constructing a rate-balance plot, as we have done before. Figure 3.36a shows the rate-balance plot, where the intersections between the rates of N_R synthesis and degradation denote the values of N_R^* . The decreasing

sigmoidal curve for the rate of N_R synthesis is due to the negative feedback loop and does not depend on N_S . Therefore, when the degradation rate curve (which depends on N_S) intersects the synthesis curve at its region of steepest decline, N_R^* falls within a narrow range of values.

We compare the results of applying the ABK method to simulating this motif with the deterministic SR curve in Figure 3.36b. The agreement is excellent ($R^2 = 0.9717$).

Follow this [link](#) to see the Matlab code.

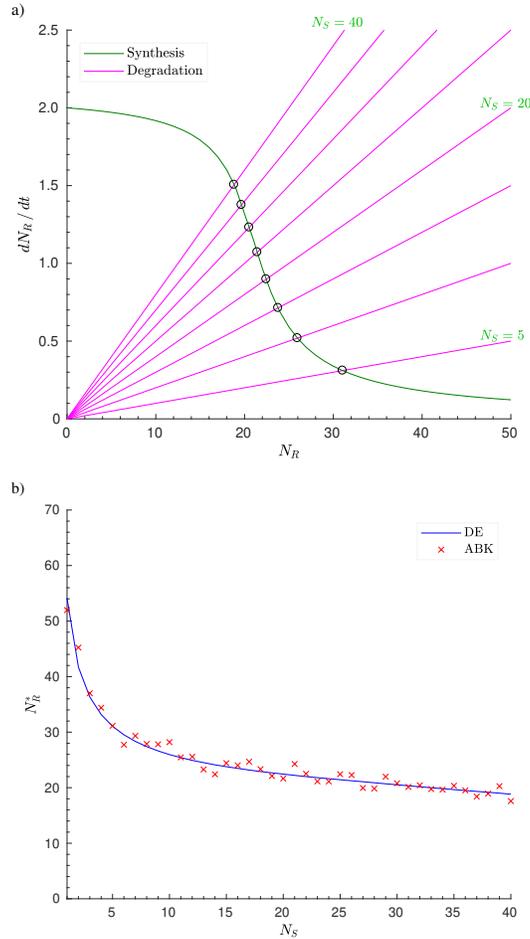


Figure 3.36 – Homeostatic response motif: rate-balance plot and SR curve. We used the ABK algorithm to model the time evolution of the motif shown in Figure 3.35, with kinetic rate constants $k_b = 0.02 \text{ sec}^{-1}$, $k_d = 0.002 \text{ sec}^{-1}$, $k_f = 1.00 \text{ sec}^{-1}$, $k_r = 0.05 \text{ sec}^{-1}$, $K_{M,f} = 5$, $K_{M,r} = 10$, with initial populations $N_{R,i} = 20$, $N_{E,i} = 70$, and $N_{E,TOT} = 100$. **a)** Rate-balance plot showing the rates of synthesis (green) and degradation (magenta) of R as a function of N_R . Since degradation depends on N_S , we have plotted the synthesis curves for values of $N_S = 1 - 40$ (shown here in multiples of 5; selected N_S values are highlighted in green). The intersections (marked by the hollow black circles) represent the steady-state values of population R , N_R^* . **b)** We used the ABK algorithm to model this motif for a total simulated time of 1500sec using fixed time step increments of $\Delta t = 0.02 \text{ sec}$. We performed separate simulations for distinct integer values of S agents ($N_S = 1 - 40$) promoting the degradation of R , where we assumed that the population of S remains constant. We plot here N_R^* (averaged over the course of the last 750sec of simulated time to account for the variation in population size due to stochasticity once the steady state had been reached) vs. N_S for each of the performed simulations (red crosses). The blue curve is the deterministic signal-response curve (obtained for each value of N_S by numerically determining N_R^* from the differential equations describing this motif). The agreement between the simulations and the theoretical curve is excellent ($R^2 = 0.9717$).

3.5 Oscillator Systems

We now turn our attention to oscillators. We begin by presenting the 2-component negative feedback motif, which is not theoretically capable of oscillations. However, we observe in our simulations that the level of response has quasi-regular large fluctuations around its predicted steady-state level. We term this phenomenon *persistent high-amplitude fluctuations* (or PHAF). If this behavior is experimentally verified to be true, it would mean that systems in the regime of small population sizes can generate oscillatory activity more easily than expected based on deterministic arguments. We also use the case of the 2-component negative feedback motif to illustrate how the ABK algorithm can be adapted to explicitly model time delays in one or more processes. Time delays are well-known to result in oscillations and, indeed, our simulations support this expectation and are in agreement with the solution of the system's delay differential equation, or DDE. We note that the algorithm allows for specified delay values to be associated with specific agents. As a result, a delay becomes another parameter that can be taken into account when modeling heterogeneous populations (we present this case in the next chapter, section 4.11).

Increasing the number of components in a negative feedback loop is known to result in oscillatory activity. For instance, adding just one additional species to the aforementioned 2-component loop can be thought of as an implicit form of time delay, thus affecting the temporal dynamics of an effector species regulating its downstream target. We verify this prediction in the 3-component negative feedback motif. We also study alternative ways to produce oscillations: the activator-inhibitor and substrate-depletion motifs. In all of these cases, we observe persistent high-amplitude fluctuations for sets of parameters that are deterministically predicted to act as damped oscillators and we speculate on how stochasticity can be the origin of this phenomenon.

3.5.1 Negative Feedback 2-Component Motif

We first consider motifs that feature a negative feedback loop. We show in Figure 3.37 the simplest such motif characterized by the regulatory interaction between species R and X/Xp . Specifically, R enhances the production of Xp , which in turn promotes the degradation of R , thus completing the feedback loop.¹⁷ This kind of regulatory loop can produce an adaptive response for the right set of parameters (Ferrell, 2016), a behavior we showcased in section 3.4 by simulating iFFLs. We have verified that this motif results in adaptation, but will focus instead on its oscillatory dynamics in the ensuing presentation. Specifically with regard to oscillatory activity, mathematical analysis of a simplified 2-component negative feedback loop has shown that it is *not* capable of producing sustained oscillations for any set of parameters (Griffith, 1968). However, as we shall see, our presentation of applying the ABK algorithm to simulating this motif shows that stochasticity can produce sustained high-amplitude fluctuations with irregular periodicity.

¹⁷The negative feedback can also take the form of Xp directly inhibiting the synthesis of R .

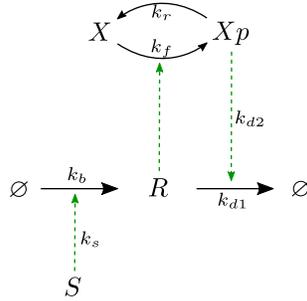


Figure 3.37 – Negative feedback 2-component motif. R promotes the conversion of X to Xp , which in turn has an inhibitory effect on R by promoting its degradation (or, alternatively, by inhibiting the synthesis of R ; not shown). Thus, an overall negative feedback loop is formed where R ultimately promotes its own degradation by regulating an intermediate species, X . The signal, S , enhances the synthesis of R , and we assume that the population size of S , N_S , remains constant. k_b and k_s are the basal 0th and 1st order (with respect to S) rate constants for the synthesis of R , respectively. Similarly, k_{d1} and k_{d2} are the basal 1st order (with respect to R) and overall 2nd order (1st order with respect to each, R and Xp) rate constants for the degradation of R , respectively. We assume that the interconversions between X and Xp follow Michaelis-Menten kinetics and that no further synthesis or degradation processes occur for these species (i.e., $N_{X,TOT} = N_X(t) + N_{Xp}(t) = \text{constant}$).

We assume that the combined population size of X and Xp remains constant, which we denote by $N_{X,TOT}$. Given this restriction, the deterministic description of this motif requires the use of only two variables, $N_R(t)$ and $N_{Xp}(t)$. Their respective differential equations are

$$\frac{dN_R}{dt} = (k_b + k_s N_S) - (k_{d1} N_R + k_{d2} N_R N_{Xp}) \quad (3.41)$$

$$\frac{dN_{Xp}}{dt} = k_f N_R \frac{N_{X,TOT} - N_{Xp}}{K_{M,f} + (N_{X,TOT} - N_{Xp})} - k_r \frac{N_{Xp}}{K_{M,r} + N_{Xp}}. \quad (3.42)$$

Note that we have grouped in parentheses the terms that correspond to the synthesis and degradation of R in equation 3.41. k_b represents the rate constant for the basal level of the synthesis of R , while $k_s N_S$ is a measure of how much S increases the rate of the process (as in previous cases in this chapter, we assume $N_S = \text{constant}$). Similarly, we use two degradation terms: one with 1st order rate constant k_{d1} that reflects the basal rate of R 's degradation in the absence of Xp , and a 2nd order process which depends on both R and Xp (rate constant k_{d2}).

ABK simulation

We begin by presenting the results of simulating this motif with kinetic constants $k_b = 0 \text{ sec}^{-1}$, $k_s = 0.100 \text{ sec}^{-1}$, $k_{d1} = 0 \text{ sec}^{-1}$, $k_{d2} = 0.005 \text{ sec}^{-1}$, $k_f = 0.010 \text{ sec}^{-1}$, $k_r = 1.000 \text{ sec}^{-1}$, Michaelis constants $K_{M,f} = K_{M,r} = 10$, a constant signal population size $N_S = 30$, initial population sizes $N_{R,i} = 0$,

Follow this [link](#) to see the Matlab code.

$N_{Xp,i} = 0$, and $N_{X,TOT} = N_X(t) + N_{Xp}(t) = 100$. We refer to this set of parameters as case 1. The 2-dimensional system of differential equations describing this motif has a single stable fixed point which we have evaluated using numerical methods: $N_R^* = 56.99$ and $N_{Xp}^* = 10.52$. Linearization in the neighborhood of the fixed point yields the eigenvalues $\lambda_{1,2} = -0.0385 \pm 0.0486i$, whose negative real part and presence of an imaginary component predict decaying (or damped) oscillations in the deterministic time trajectory. Figure 3.38a shows the excellent agreement between the average time trajectories obtained from the ABK simulation ($n = 500$) and the deterministic ones. However, inspection of individual simulation trajectories shows the presence of high-amplitude and irregular fluctuations for both species R and Xp that occur well beyond the expected time window of damped oscillatory activity ($0 < t < 150$ sec). We observed this behavior in all simulation runs we attempted (see Figure 3.38b for a representative example) and for arbitrarily long simulated time periods. We further analyzed these fluctuations by computing a measure of the noise in the simulation trajectories, the coefficient of variation,¹⁸

$$\eta(t) = \frac{\sigma(t)}{\langle N(t) \rangle}. \quad (3.43)$$

This quantity can be computed for each of the species in the system whose average time trajectory and standard deviation have been computed. Assuming that the synthesis and degradation/elimination processes that each species participates in are independent of each other, then a species can be thought of as being part of a birth-death process (section 2.8). Such a process obeys Poisson statistics¹⁹ and has the property $\sigma_P = \sqrt{\langle N \rangle}$. Therefore, the coefficient of variation is $\eta_P(t) = 1/\sqrt{\langle N(t) \rangle}$, which we consider to be a measure of the level of intrinsic noise expected for a system obeying Poisson statistics. We show a comparison of η and η_P for both species in Figure 3.38c and observe that the fluctuations for species R are higher than expected ($\eta_{N_R} \approx 2\eta_P$). We propose that the high variability in the level of R results from the (damped) oscillatory dynamics of the system leading to a long excursion around the fixed point when a perturbation of sufficient magnitude occurs. Such perturbations in N_R can occur through the regulatory effect of species Xp , whose small population size is subject to large relative changes due to stochasticity.

We show the phase plane for this motif in Figure 3.38d to illustrate how the observed large fluctuations in N_R can occur. The deterministic trajectory follows a spiraling path toward the fixed point, as would be expected for decaying oscillations given the aforementioned eigenvalues. However, a perturbation in N_R can result in a stochastic trajectory taking a long path around the fixed point (as shown by the light brown dashed curve in the figure).

We now discuss the effect of noise in the small population size of Xp (recall that $N_{Xp}^* \approx 10$). In this regime of a low copy number population, noise-induced

¹⁸Introduced in equation 2.145 and reproduced here in more succinct notation.

¹⁹It is reasonable to question this assumption. Some authors have argued that the negative binomial distribution is more appropriate for describing chemical systems (Paulsson et al., 2000).

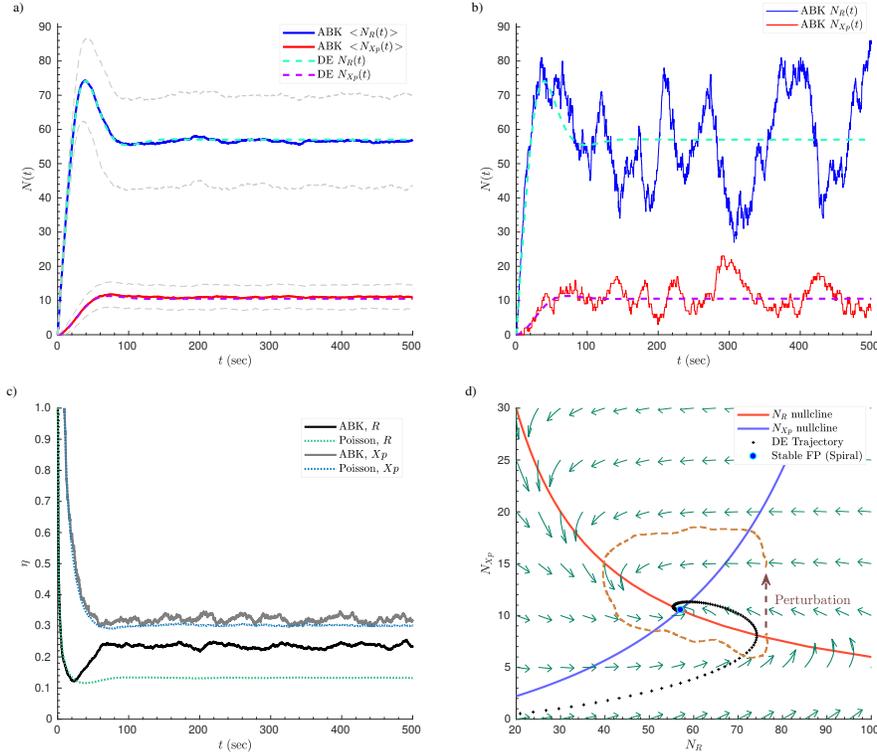


Figure 3.38 – Negative feedback 2-component motif: time course (case 1). We used the ABK algorithm to model the time evolution of the motif shown in Figure 3.37 with kinetic constants $k_b = 0 \text{ sec}^{-1}$, $k_s = 0.100 \text{ sec}^{-1}$, $k_{d1} = 0 \text{ sec}^{-1}$, $k_{d2} = 0.005 \text{ sec}^{-1}$, $k_f = 0.010 \text{ sec}^{-1}$, $k_r = 1.000 \text{ sec}^{-1}$, and Michaelis constants $K_{M,f} = K_{M,r} = 10$. We used a constant signal population size $N_S = 30$, initial population sizes $N_{R,i} = 0$, $N_{Xp,i} = 0$, and $N_{X,TOT} = N_X(t) + N_{Xp}(t) = 100$. We simulated this motif for a total time of 500 sec with fixed time step increments of $\Delta t = 0.01 \text{ sec}$. **a)** We plot the average time trajectory ($n = 500$) of the populations N_R and N_{Xp} (blue and red curves respectively) along with their one-standard-deviation envelopes (gray dashed curves). The deterministic trajectories are also shown (labeled “DE” in the legend). **b)** We show a sample simulation run that is representative of the persistent high-amplitude fluctuations observed in our simulations. **c)** We compare the coefficient of variation, $\eta(t)$, for each of the species with $\eta_P(t)$ for a Poisson process. The black curve indicates that the trajectory of species R shows a significantly larger deviation from its mean than expected. **d)** Phase plane for this motif. The deterministic trajectory spirals into the fixed point (i.e., decaying oscillations) at the intersection of the two nullclines. An intrinsic perturbation (i.e., noise; brown dashed arrow) sends the system along a stochastic counterclockwise excursion around the fixed point. Note that the direction and size of the perturbation can vary and is only shown here as an example. We show a smoothed-out cartoon version of a sample stochastic trajectory (light brown dashed curve) to facilitate visualization. The deterministic direction field is depicted as green arrows.

changes in N_{Xp} propagate to its regulatory target by producing large changes in the transition probability for the degradation of R . For example, if $N_{Xp}(t) = 6$ at some time t and over a subsequent time interval τ the population doubles (i.e., $N_{Xp}(t + \tau) = 12$), then the per- R -agent probability of a degradation event doubles as well. This is most easily seen using the P_{dif} formulation of transition probabilities. That is, for our example,

$$\frac{P_{dif,R}(t + \tau)}{P_{dif,R}(t)} = \frac{k_{d2}N_{Xp}(t + \tau)}{k_{d2}N_{Xp}(t)} = 2, \quad (3.44)$$

and the population of R would decrease accordingly since each agent has a higher probability of degradation. Thus, the relative change in the population size of Xp results in an equivalent change in the probability of R 's disappearance, leading to a transient decrease in N_R . A similar argument can be made for changes in N_{Xp} that lead to a large increase in N_R .

The above hypothesis implies that if N_{Xp}^* increases (as a consequence of changes in the motif's parameters), the high-amplitude fluctuation in R would decrease since the noise-induced relative changes in N_{Xp} will be smaller. We have simulated this motif for several other parameter combinations and have indeed found this to be the case. In the interest of brevity, we present here just one of those additional cases that can be thought of as demonstrating the diametrically opposite effect: a reduction in noise level of Xp . Figure 3.39 shows our results for this example (case 2). The chosen parameters (same as in case 1, except $k_s = 0.300 \text{ sec}^{-1}$ and $k_f = 0.10 \text{ sec}^{-1}$) result in the deterministic steady-state population levels $N_R^* = 19.7$, $N_{Xp}^* = 91.5$ (these values effectively represent the reverse of case 1 in terms of steady-state population sizes; here, $N_R^* < N_{Xp}^*$). Note that the stable fixed point in this case is *not* reached via a spiral trajectory within state space (i.e., there are no decaying oscillations), as revealed by both eigenvalues being real and negative ($\lambda_1 = -0.4460$, $\lambda_2 = -0.0703$). We show the phase plane for this system in panel d of the same figure, where the intersection of the nullclines corresponds to the fixed point. It is clear that when there is no spiraling tendency in the system, a perturbation in the population sizes will *not* result in a long excursion around the fixed point. Instead, the trajectory has small-amplitude fluctuations in the immediate neighborhood of the fixed point.

We also plot the noise levels in panel c and observe that the time trajectory of species Xp is significantly *less* noisy than expected for a Poisson process. We propose the reason for this observation has to do with the fact that we have modeled species X and Xp as not being synthesized or degraded through additional processes and can only be converted from one to the other (see Figure 3.37). Keep in mind that in both cases we have examined so far, we have used $N_{X,TOT} = 100$. Therefore, when the population of R is small, its relative increase results in an equivalent increase in the probability of the transition $X \rightarrow Xp$ (using the same argument about transition probabilities as shown in the example of equation 3.44). However, a commensurate increase in the population size of Xp cannot occur because 1) the reverse reaction ($Xp \rightarrow X$) is much faster than the forward one ($k_r = 10k_f$), especially as $N_{Xp}(t)$ is close to $N_{X,TOT}$; and 2) $N_{X,TOT}$ is a ceiling for the values $N_{Xp}(t)$ can have when $N_R(t)$

is low. Thus, the population size of Xp cannot fluctuate to the extent that R 's regulatory effect would indicate and the time trajectory of Xp has a reduced noise profile.

Finally, we show an additional case (Figure 3.40; case 3) for the purpose of better visualizing how dramatic these persistent fluctuations can be. We have used a set of parameters (listed in the figure) that results in a steady-state value of the response element being an order of magnitude higher than in case 1 ($N_R^* = 519.7$). We only show a sample time course for this case where the amplitude and irregularity of the fluctuations in N_R are clearly visible. Note that this case is closely related to case 1, in that the fixed point is a stable spiral. Thus, perturbations routinely cause a stochastic trajectory to take a long excursion around the fixed point (as shown in Figure 3.38d).

We also point out that we have qualitatively reproduced all of our results in cases 1-3 using Gillespie's algorithm (using the software package StochSS; Drawert et al., 2016). Therefore, our observations are not specific to the implementation of the ABK algorithm as compared to the SSA, and the algorithmic framework we have presented in this work is consistent with the predictions the SSA makes for systems composed of homogeneous populations.

We conclude that noise-induced excursions around the fixed point can result when a stable spiral is deterministically predicted in the dynamics of the system. We term this phenomenon *persistent high-amplitude fluctuations* (PHAF).²⁰ By "high-amplitude," we mean relative to the intrinsic noise that a population's steady-state size would exhibit if considered as a consequence of a Poisson process (as shown in Figure 3.38c). We deliberately do not use the term "oscillations" to describe this effect because of the irregularity in the observed amplitude and time interval between successive population size crests or troughs. This behavior is particularly sensitive to intrinsic noise in the low copy number population size of a species involved in regulatory interactions.

The phenomenon we have observed is reminiscent of *coherence resonance*, where an optimal level of noise produces coherent oscillations (El-Samad and Khammash, 2006). However, our present observations only depend on the *intrinsic noise* of the involved processes since we have not added any tunable sources of extrinsic noise in the above computational studies. We also do not consider our results to be an example of *stochastic resonance*, where noise can be beneficial in the detection of an output signal (for a review that discusses the more recent and broader definition of this phenomenon, see McDonnell and Abbott, 2009). Furthermore, stochastic resonance does *not* occur because the system resonates at a particular frequency, but rather there is a noise-induced signal enhancement. Instead, we believe that our observations are more in line with those made by McKane et al. (2007), who argue that this is a true resonance phenomenon: assuming the intrinsic noise is white (i.e., is Gaussian-distributed and with constant power for all frequencies), it includes the system's natural resonant frequency, which results in the amplification of fluctuations in

²⁰Some authors refer to this phenomenon as *noise-induced oscillations* (e.g., Thomas et al., 2013).

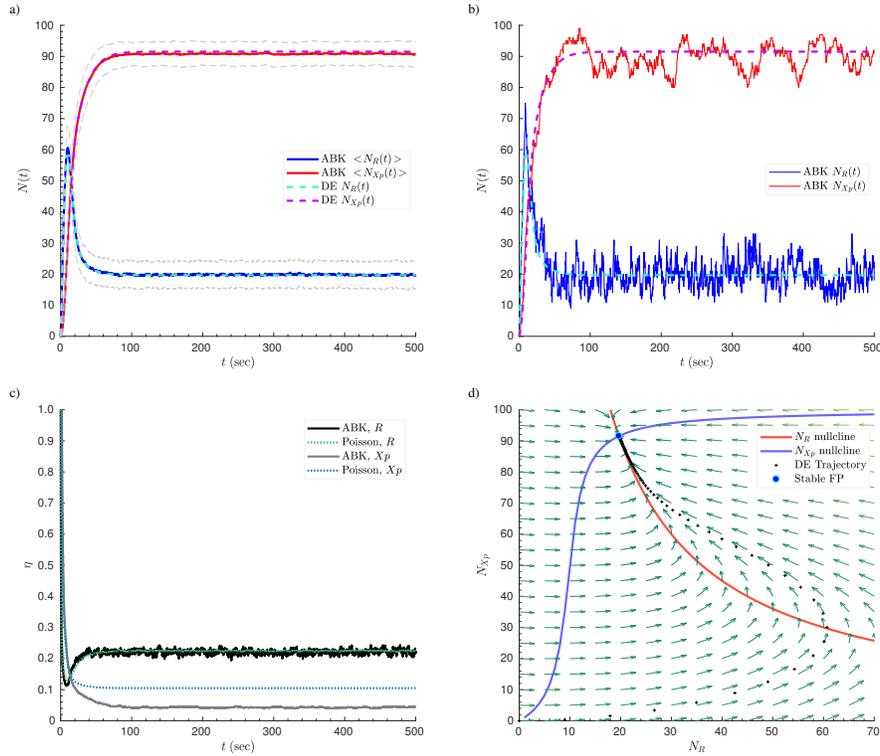


Figure 3.39 – Negative feedback 2-component motif: time course (case 2). We used the ABK algorithm to model the time evolution of the motif shown in Figure 3.37 with kinetic constants $k_b = 0 \text{ sec}^{-1}$, $k_s = 0.300 \text{ sec}^{-1}$, $k_{d1} = 0 \text{ sec}^{-1}$, $k_{d2} = 0.005 \text{ sec}^{-1}$, $k_f = 0.10 \text{ sec}^{-1}$, $k_r = 1.000 \text{ sec}^{-1}$, and Michaelis constants $K_{M,f} = K_{M,r} = 10$. We used a constant signal population size $N_S = 30$, initial population sizes $N_{R,i} = 0$, $N_{Xp,i} = 0$, and $N_{X,TOT} = N_X(t) + N_{Xp}(t) = 100$. We simulated this motif for a total time of 500 sec with fixed time step increments of $\Delta t = 0.01 \text{ sec}$. Stability analysis: fixed point $N_R^* = 19.7$, $N_{Xp}^* = 91.5$; eigenvalues: $\lambda_1 = -0.4460$, $\lambda_2 = -0.0703$. **a)** Average time trajectories ($n = 500$) of the populations N_R and N_{Xp} (blue and red curves respectively) along with their one-standard-deviation envelopes (gray dashed curves). The deterministic trajectories are also shown (labeled “DE” in the legend). **b)** Sample ABK simulation run. **c)** Comparison of the coefficient of variation, $\eta(t)$, for each of the species with $\eta_P(t)$ for a Poisson process. The trajectory of species Xp shows a significantly smaller deviation from its mean than expected. **d)** Phase plane for this motif. The deterministic trajectory does *not* spiral into the fixed point at the intersection of the two nullclines. Therefore, a stochastic trajectory cannot take an excursion around the fixed point. The deterministic direction field is depicted as green arrows.

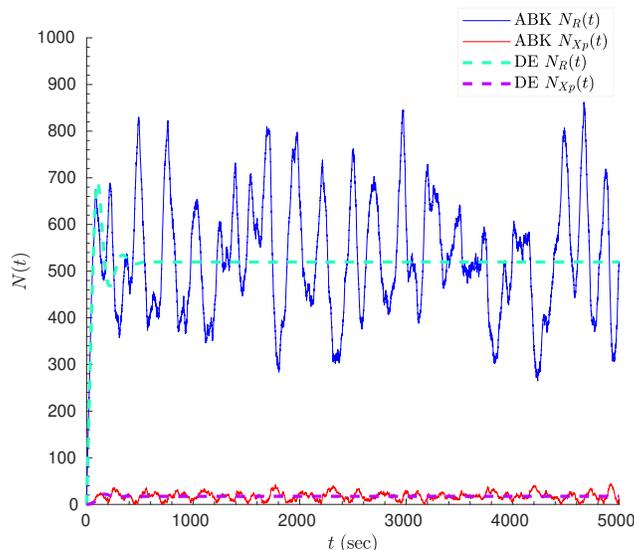


Figure 3.40 – Negative feedback 2-component motif: time course (case 3). We show a single ABK simulation run of the same motif as in Figure 3.38, with kinetic constants $k_b = 0 \text{ sec}^{-1}$, $k_s = 0.300 \text{ sec}^{-1}$, $k_{d1} = 0 \text{ sec}^{-1}$, $k_{d2} = 0.001 \text{ sec}^{-1}$, $k_f = 0.0015 \text{ sec}^{-1}$, $k_r = 0.800 \text{ sec}^{-1}$, and Michaelis constants $K_{M,f} = 30$, $K_{M,r} = 1$. We used a constant signal population $N_S = 30$, initial population sizes $N_{R,i} = 0$, $N_{Xp,i} = 0$, and $N_{X,TOT} = N_X(t) + N_{Xp}(t) = 1000$. We simulated this motif for a total time of 5000 sec with fixed time step increments of $\Delta t = 0.01 \text{ sec}$. Deterministic analysis: fixed point $N_R^* = 519.7$, $N_{Xp}^* = 17.3$; eigenvalues: $\lambda_{1,2} = -0.00985 \pm 0.02648i$. The stochastic trajectory of species R shows persistent high-amplitude fluctuations from its mean. We note that the average trajectory agrees with the deterministic trajectory (plot of average trajectories not shown).

the population sizes. The same authors go on to perform simulations and theoretically show that these amplified fluctuations occur in systems with finite population sizes (i.e., away from the thermodynamic limit) when damped oscillations are deterministically predicted. It is also noteworthy that Vilar et al. (2002) made similar *in silico* observations of (intrinsic-)noise-induced persistent fluctuations/oscillations. Importantly, this phenomenon has been observed in an experimental setting (e.g., Geva-Zatorsky et al., 2010).

In this work, we will see additional instances of PHAF in the case of the 3-component negative feedback loop (subsection 3.5.3) and a variant of the predator-prey competition model (subsection 3.6.1). We hope that this investigation will inspire and inform further efforts toward the experimental verification of the presence of such effects.

3.5.2 Negative Feedback 2-Component Motif with Explicit Time Delay

A *delay*, or *time lag*, within a feedback loop has been observed in real populations of living organisms to cause stable oscillations or irregular fluctuations about an equilibrium population size (for a summary of the latter behavior in an ecological context, see Krebs 1972, pp. 201-202). A time delay means that a species population changes at a rate that is dependent on population sizes at some time in the past, therefore the involved processes are *non-Markovian*. It is well-known that such time delays can cause a (molecular) system to repeatedly overshoot and undershoot a steady-state value, thereby producing oscillations (Novák and Tyson, 2008; Bratsun et al., 2005). Time delays can be specified *explicitly* for a given set of processes, but in reality are *implicit* because of some limitation that a natural system may be subjected to (for instance, diffusion of molecules across cellular compartments, or the presence of many intermediate species in a long sequential or cyclical set of processes). Given the role that delays may play in producing oscillatory behavior, we first investigate the effect of introducing an *explicit time delay* into the simulation of the negative feedback 2-component motif.

We now show how to account for the non-Markovian nature of a process with delay while using the ABK algorithm. We specifically use a time delay δ in the negative feedback reaction where X_p promotes the degradation of R (see Figure 3.37).²¹ To clarify what we mean, the (partial) differential equation for the population $N_R(t)$ can be rewritten (see equation 3.41 for comparison) as

$$\frac{\partial N_R(t, \delta)}{\partial t} = (k_b + k_s N_S) - (k_{d1} N_R(t) + k_{d2} N_R(t) \mathbf{N}_{X_p}(t - \delta)), \quad (3.45)$$

where we have highlighted in bold the term with the time delay. The expression for dN_{X_p}/dt is the same as in equation 3.42 since there are no time delays in it (by assumption). We have also noted all time-dependent terms in this differential equation to further emphasize that *only* the rate of change in N_R depends on the size of the X_p population at a previous time point $t - \delta$.

Given this explicit time delay, the probability (P_{dif}) that *each* agent of species R is degraded in a 2nd order process with rate constant k_{d2} is

$$P_{\mathbf{R}+X_p \rightarrow \emptyset}(t, \delta) = k_{d2} N_{X_p}(t - \delta) \Delta t, \quad (3.46)$$

for $t > \delta$.²²

We show the results of simulating this motif for a total of 2000 sec with a time delay of fixed duration $\delta = 50$ sec in Figure 3.41 (all other parameters are identical to those in Figure 3.38). Panel *a* shows a representative sample run of the simulation that reveals oscillations of higher amplitude and regularity than those in Figure 3.38b. Therefore, incorporating an explicit time delay in the

²¹The choice of *only* this reaction having a time delay is purely arbitrary and made for the purpose of demonstrating the versatility and extensibility of the ABK algorithm in simulating a diverse array of modeling studies or *in silico* experiments.

²²As noted in subsection 2.3.2, the P_{ber} probability expression can also be used. The kinetic complementarity factor (KCF; see section 2.6) is $\Omega_R(t) = k_{d2} N_{X_p}(t - \delta)$.

Follow this [link](#) to see the Matlab code.

agent-based stochastic simulation results in robust oscillations for a motif that cannot have stable oscillations in the absence of a delay. Panel b shows that the average time trajectories of the populations are decaying oscillations. The apparent decay is *not* representative of individual simulation runs, since they show unattenuated oscillatory activity (we verified this by running simulations of longer time periods). Instead, the apparent decay of the average trajectories is the result of stochastic fluctuations leading to variance in the time between successive peaks, essentially producing small phase differences in the trajectories (dephasing) that smooth out and diminish the value of the average.

We briefly note the ease with which the ABK algorithm can be altered to account for variability in the make-up of a population. In our example above, we treated all R agents identically in that the transition probability of each agent has the same dependence on δ . However, having only a subset of R agents be sensitive to a delay can be easily implemented using our algorithm. Furthermore, it is possible to specify a different value of δ *for each* R agent, or to have δ sampled from a user-specified statistical distribution (we will present these cases in the next chapter, section 4.11). Thus, one can easily model a population that is heterogeneous with respect to a delay parameter (we will extend this idea to any parameter the system depends on in Chapter 4).

The example we have presented in this subsection shows that a modeler can easily implement explicit time delays using the ABK algorithm. As we will see in the next subsection, introducing additional intermediate species in the feedback loop is an *implicit* form of time delay that also leads to stable oscillations.

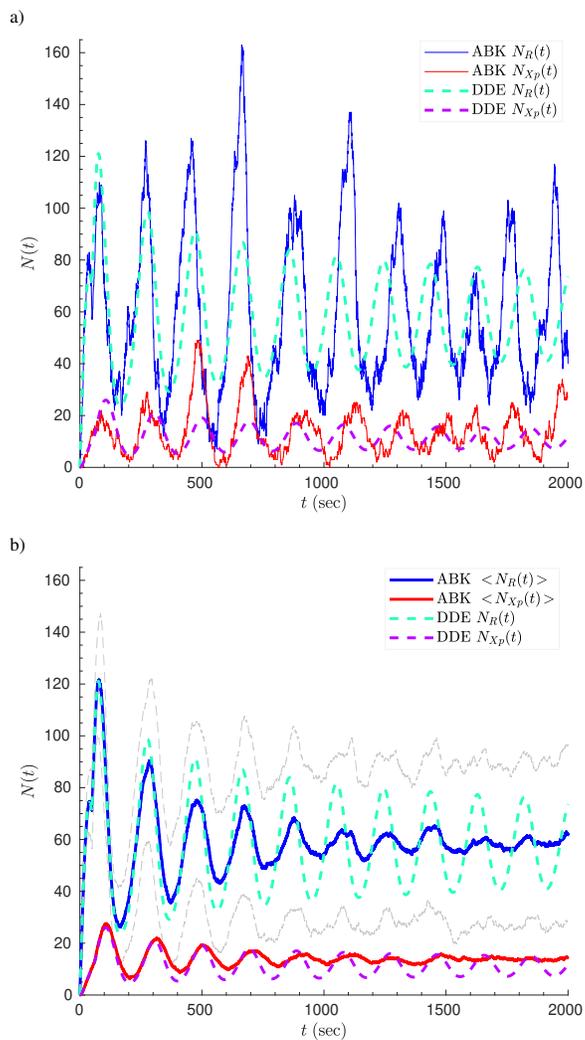


Figure 3.41 – Negative feedback 2-component motif with delay: time course. We used the ABK algorithm to model the time evolution of the negative feedback 2-component motif with an explicit time delay of $\delta = 50$ sec in the feedback reaction ($R + Xp \rightarrow \emptyset$). All other parameters are the same as those in Figure 3.38, except we simulated the system for 2000 sec. The first 50 sec of the trajectories were simulated without the time delay in the ABK (solid curves) and deterministic trajectories (dashed curves). The latter were evaluated by solving the ODE for the initial 50 sec, then obtaining the numerical solution to the set of delay differential equations, or DDEs, using the initial 50 sec period as history. **a)** Sample time trajectory. **b)** Average population sizes after 100 repetitions of the simulation. The gray dashed curves denote the one-standard-deviation envelope for $\langle N_R(t) \rangle$.

3.5.3 Negative Feedback 3-Component Motif

We now explore the same negative feedback loop as the one presented in the preceding subsection, but with an additional component, species W . Such a 3-component motif is theoretically capable of stable oscillations (i.e., a limit cycle) and has been used in the proof-of-principle design of a synthetic oscillatory network of three transcriptional repressors, termed “the repressilator” (Elowitz and Leibler, 2000). The presence of an additional component is a way of introducing an *implicit* time delay to the feedback loop such that the response element’s change in population size effectively depends on the negative effector’s population level at a previous time. This is a well-known criterion for generating an oscillatory response in chemical and biological systems (Novák and Tyson, 2008). We show the processes and regulatory interactions in this 3-component motif below (Figure 3.42).

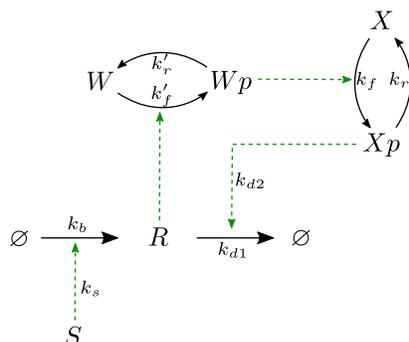


Figure 3.42 – Negative feedback 3-component motif. R promotes the conversion of W to Wp , which in turn promotes the production of Xp that has an inhibitory effect on R by promoting its degradation (or, alternatively, by inhibiting R 's synthesis; not shown). Thus, an overall negative feedback loop is formed where R ultimately promotes its own degradation by regulating two intermediate species, W and X . The signal, S , enhances the synthesis of R and we assume that the population size of S , N_S , remains constant. The rate constants are indicated in the figure. k_b and k_s are the basal 0th and 1st order (with respect to S) rate constants for the synthesis of R , respectively. Similarly, k_{d1} and k_{d2} are the basal 1st order (with respect to R) and overall 2nd order (1st order with respect to each, R and Xp) rate constants for the degradation of R , respectively. We assume that the interconversions between W and Wp as well as X and Xp follow Michaelis-Menten kinetics and that no further synthesis and degradation processes occur for species W/Wp and X/Xp . That is, $N_{W,TOT} = N_W(t) + N_{Wp}(t) = \text{constant}$ and $N_{X,TOT} = N_X(t) + N_{Xp}(t) = \text{constant}$.

The system can be deterministically analyzed with the following set of differential equations,

$$\frac{dN_R}{dt} = (k_b + k_s N_S) - (k_{d1} N_R + k_{d2} N_R N_{Xp}) \quad (3.47)$$

$$\frac{dN_{Wp}}{dt} = k'_f N_R \frac{N_{W,TOT} - N_{Wp}}{K_{M,f} + (N_{W,TOT} - N_{Wp})} - k'_r \frac{N_{Wp}}{K_{M,r} + N_{Wp}} \quad (3.48)$$

$$\frac{dN_{Xp}}{dt} = k_f N_{Wp} \frac{N_{X,TOT} - N_{Xp}}{K'_{M,f} + (N_{X,TOT} - N_{Xp})} - k_r \frac{N_{Xp}}{K'_{M,r} + N_{Xp}}, \quad (3.49)$$

where we have made the same assumptions as in the 2-component case. We initially consider this motif with kinetic constants $k_b = 0 \text{ sec}^{-1}$, $k_s = 0.400 \text{ sec}^{-1}$, $k_{d1} = 0 \text{ sec}^{-1}$, $k_{d2} = 0.001 \text{ sec}^{-1}$, $k_f = k'_f = 0.010 \text{ sec}^{-1}$, $k_r = k'_r = 1.000 \text{ sec}^{-1}$, Michaelis constants $K_{M,f} = K'_{M,f} = 10 = K_{M,r} = K'_{M,r}$, a constant signal population $N_S = 10$, initial population sizes $N_{R,i} = 0$, $N_{Wp,i} = 100$, $N_{Xp,i} = 100$, total population of species X/Xp , $N_{X,TOT} = N_X(t) + N_{Xp}(t) = 250$, and for species W/Wp , $N_{W,TOT} = N_W(t) + N_{Wp}(t) = 250$. Given these parameters, a single stable fixed point is deterministically predicted with population values $N_R^* = 93.65$, $N_{Wp}^* = 86.81$, $N_{Xp}^* = 43.14$. Linearization around this point gives the eigenvalues $\lambda_1 = -0.047342$, $\lambda_{2,3} = -0.00045305 \pm 0.013501i$. Thus, we qualitatively predict damped oscillations as the trajectory to the fixed point.

We simulated the time evolution of this motif using the ABK algorithm and we show the average time trajectories ($n = 200$) of populations R , Wp , and Xp in Figure 3.43a. For clarity, panels b, c, and d separately show each of the average population trajectories with their one-standard-deviation envelope, along with the deterministic prediction obtained from the numerical integration of the system of differential equations presented above. The oscillations in the average trajectories appear to decay faster than in the deterministic solution. We attribute this to the effect of noise in dephasing the oscillatory activity in individual simulation runs (especially after the first full oscillation in the system's time evolution; see Figure 3.44). We note that the role of stochasticity in dephasing oscillator systems controlling the cell cycle of individual cells has been experimentally observed (Di Talia et al., 2007).

Figure 3.44 is identical to 3.43, except it shows the results of a sample simulation run. We again observe high-amplitude, irregular, and persistent fluctuations (PHAF: see subsection 3.5.1) even as the oscillatory behavior is deterministically predicted to decay and for arbitrarily long periods of time (only shown here for $t \leq 5000 \text{ sec}$). The fact that we have made similar observations in the 2- and 3-component feedback loops supports our proposal that noise affects the dynamics of the system such that it continuously over/under-shoots the fixed point, while the intrinsic delay due to the presence of the additional intermediate W/Wp amplifies the effect. Note that sustained oscillations have been observed in the stochastic simulation of a genetic circadian oscillator model (using Gillespie's algorithm; Vilar et al., 2002), where the deterministic prediction is fast decay to a stable fixed point (see Figure 5 in Vilar and coworkers' study). Combined with our studies, these results highlight the potential effect of noise on intracellular regulation. In this case, noise may increase the *robustness* of

Follow this [link](#) to see the Matlab code.

the oscillatory regime by expanding the region in parameter space where stable oscillations are possible for a given level of noise.

Finally, there is a discrepancy between the simulation results and the deterministic trajectory shown in Figure 3.43d. The midpoint of the oscillation range in the average population size of species Xp , $\langle N_{Xp}(t) \rangle$, is consistently higher than the deterministic one. We have thoroughly inspected and verified that the code used in the implementation of the algorithm is correct. One possible explanation emerges from Figure 3.44d. The high-amplitude oscillations/fluctuations in $N_{Xp}(t)$ cannot extend symmetrically from the true equilibrium population size because it is rather close to zero (i.e., $N_{Xp}^* = 43.14$, which is less than the apparent oscillation amplitude). Accordingly, $\langle N_{Xp}(t) \rangle$ has a greater value than expected. For comparison, we also show the average time course for the species in this motif, simulated with the same set of parameters except for $N_S = 30$ and $N_S = 50$ (Figures 3.45 and 3.47). The difference between the stochastic and deterministic trajectories of Xp decreases as N_S increases (the fit is nearly perfect for $N_S = 50$). However, further work will be required to determine if this discrepancy is due to the inherent dynamics of the system. Others have attributed the observation of similar effects in computational experiments to the presence of noise (Paulsson et al., 2000).

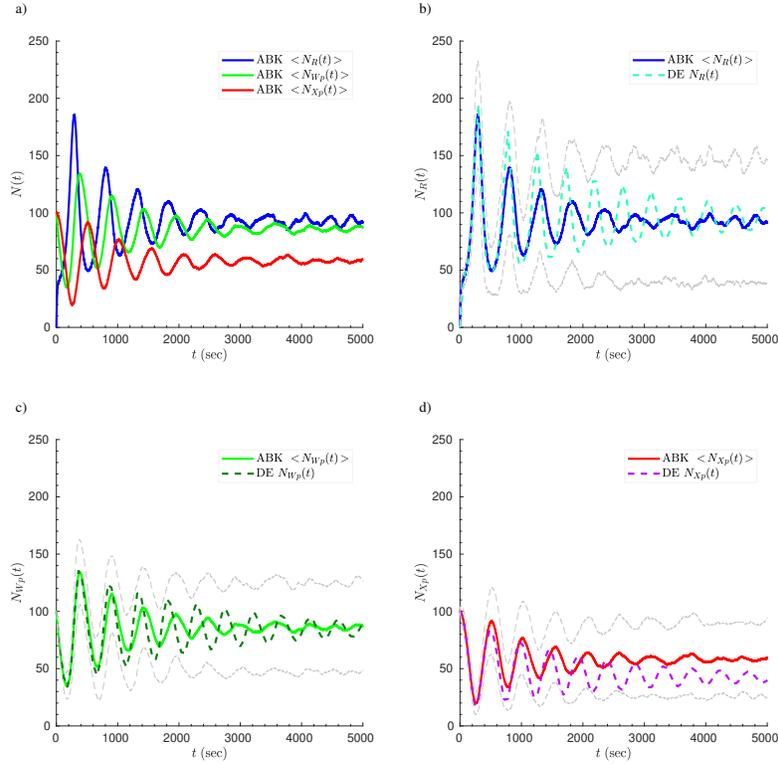


Figure 3.43 – Negative feedback 3-component motif: average time course, $N_S = 10$. We used the ABK algorithm to model the time evolution of the motif shown in Figure 3.42 with kinetic constants $k_b = 0 \text{ sec}^{-1}$, $k_s = 0.400 \text{ sec}^{-1}$, $k_{d1} = 0 \text{ sec}^{-1}$, $k_{d2} = 0.001 \text{ sec}^{-1}$, $k_f = k'_f = 0.010 \text{ sec}^{-1}$, $k_r = k'_r = 1.000 \text{ sec}^{-1}$, and Michaelis constants $K_{M,f} = K'_{M,f} = 10 = K_{M,r} = K'_{M,r}$. We used a constant signal population $N_S = 10$, initial population sizes $N_{R,i} = 0$, $N_{Wp,i} = 100$, $N_{Xp,i} = 100$, $N_{X,TOT} = N_X(t) + N_{Xp}(t) = 250$, and $N_{W,TOT} = N_W(t) + N_{Wp}(t) = 250$. We simulated this motif for a total time of 5000 sec with fixed time step increments of $\Delta t = 0.02 \text{ sec}$. **a)** Plot of the average time trajectory ($n = 200$) of the populations N_R , N_{Wp} , N_{Xp} (blue, green and red curves, respectively). **b-d)** Separate plots of the average time trajectories for each of the species R , Wp , Xp , along with their one-standard-deviation envelopes (gray dashed curves). The deterministic trajectories are also shown (labeled “DE” in the legend).

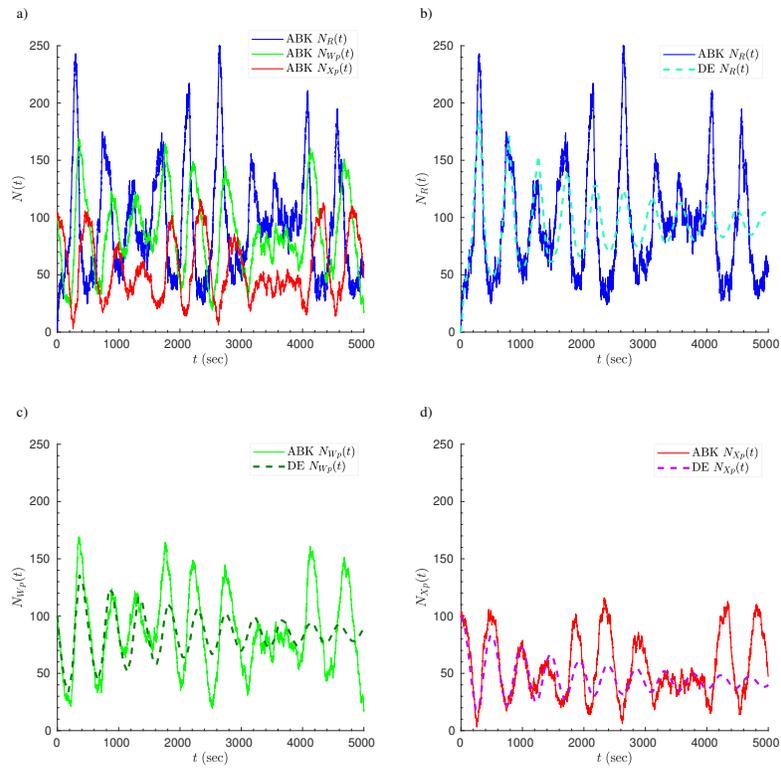


Figure 3.44 – Negative feedback 3-component motif: sample time course, $N_S = 10$. We show a sample simulation run with the same parameters as in Figure 3.43. The population time trajectories are representative of the high-amplitude irregular fluctuations that persist even as the deterministic solution decays to an equilibrium value.

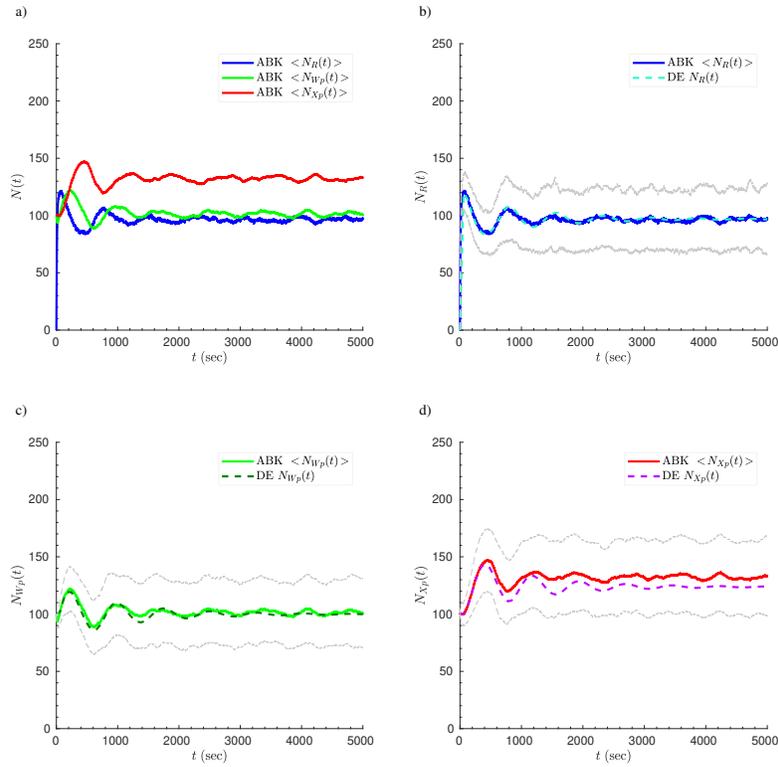


Figure 3.45 – Negative feedback 3-component motif: average time course, $N_S = 30$. We used the same parameters as indicated in Figure 3.43, except $N_S = 30$. Stable fixed point: $N_R^* = 97.58$, $N_{Wp}^* = 100.95$, $N_{Xp}^* = 123.06$. Linearization around this point gives the eigenvalues $\lambda_1 = -0.12363$, $\lambda_{2,3} = -0.00086808 \pm 0.0083141i$. The eigenvalues indicate that the decay to the stable fixed point is faster than in the case of $N_S = 10$.

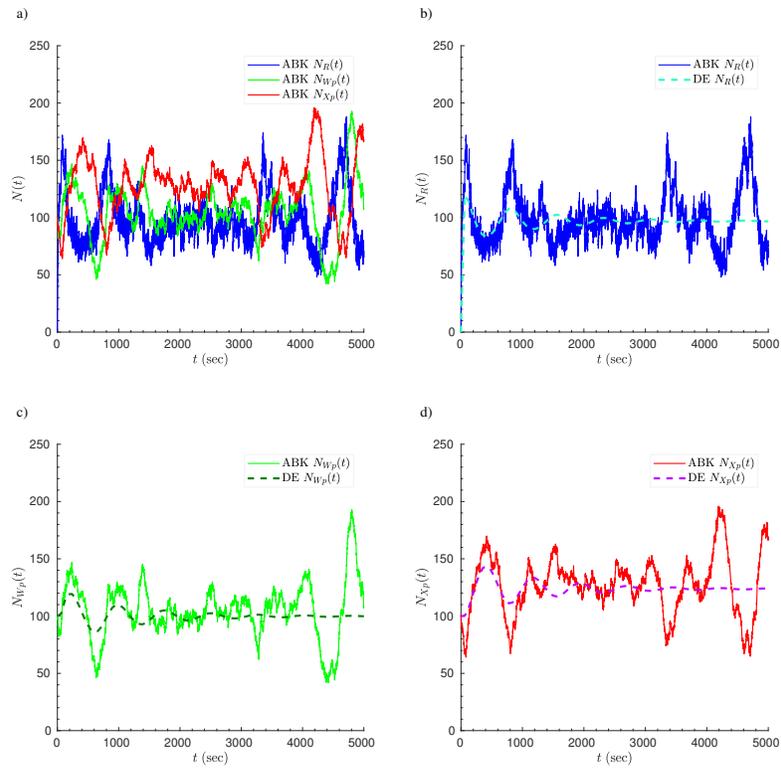


Figure 3.46 – Negative feedback 3-component motif: sample time course, $N_S = 30$. We show a sample simulation run with the same parameters as in Figure 3.45. The population time trajectories are representative of the high-amplitude irregular fluctuations that persist even as the deterministic solution decays to an equilibrium value.

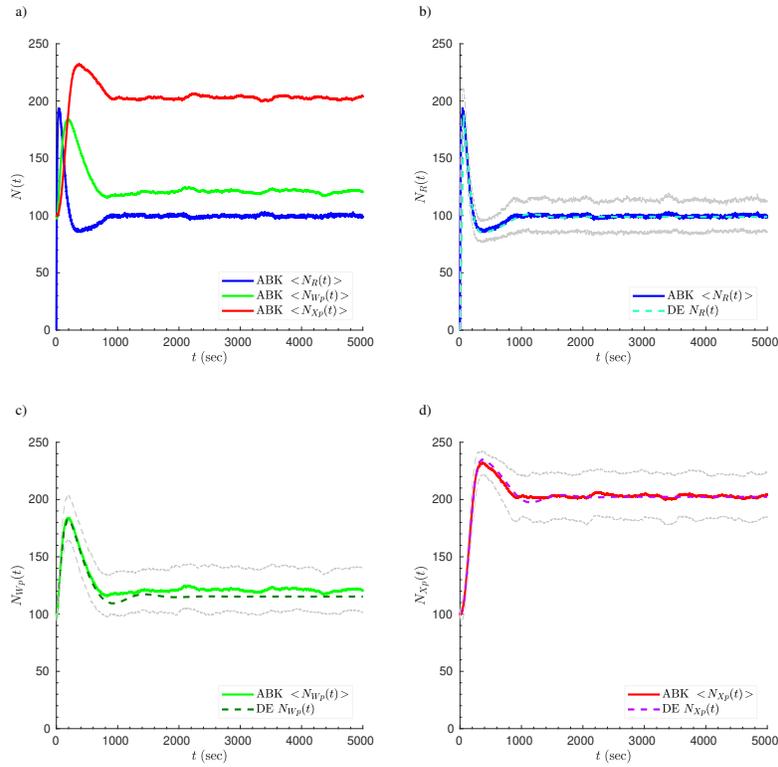


Figure 3.47 – Negative feedback 3-component motif: average time course, $N_S = 50$. We used the same parameters as indicated in Figure 3.43, except $N_S = 50$. Stable fixed point: $N_R^* = 98.85$, $N_{Wp}^* = 115.28$, $N_{Xp}^* = 202.33$. Linearization around this point gives the eigenvalues $\lambda_1 = -0.20252$, $\lambda_{2,3} = -0.0023036 \pm 0.0060281i$. The eigenvalues indicate that the decay to the stable fixed point is faster than in the case of $N_S = 30$.

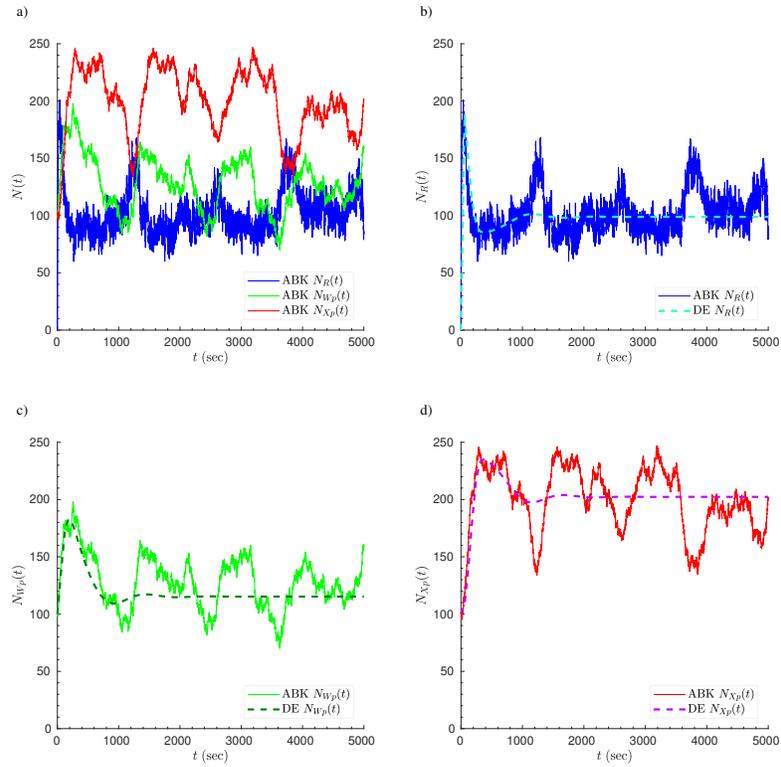


Figure 3.48 – Negative feedback 3-component motif: sample time course, $N_S = 50$. We show a sample simulation run with the same parameters as in Figure 3.47. The population time trajectories are representative of the high-amplitude irregular fluctuations that persist even as the deterministic solution quickly decays to an equilibrium value.

3.5.4 Activator-Inhibitor Motif

We now present another motif that is capable of producing oscillatory activity or *excitability*. The latter term refers to a system's ability to take a large-amplitude excursion into its phase space and subsequently return to its initial state. In the context of transient cellular differentiation, this motif has been shown to be responsible for the onset and exit of *Bacillus subtilis* from a competent state that allows the bacterium to take up exogenous DNA from its surroundings (Süel et al., 2006; Cagatay et al., 2009).

The motif we consider here consists of two modules: a) a mutual activation module where R and Ep stimulate each other's synthesis (this is the exact same motif as presented in subsection 3.3.3 that results in a bistable hysteretic switch), and b) R also stimulates the synthesis of species X that acts as an inhibitor to R (negative feedback loop; see Figure 3.49 below).

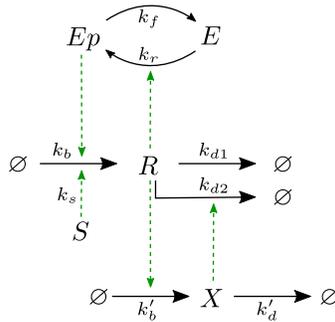


Figure 3.49 – Activator-inhibitor motif. The indicated processes involving species S , R , and E/Ep are identical as in the mutual activation hysteretic switch motif presented in Figure 3.23. The inhibitory part of this motif involves species X , whose synthesis is stimulated by R , while X promotes the degradation of R . This combination of the mutual activation module and the inhibition of R by X can produce oscillatory behavior for the right set of the system's parameters.

We consider here the case where X inhibits R by promoting its degradation (although explicit inhibition of R 's synthesis is also possible). Moreover, the diagram indicates two transitions for the degradation of R : a 1st order process with respect to R and rate constant k_{d1} that can be thought of as the basal rate of R 's degradation (i.e., it reflects the “leakiness” of the overall degradation process given an intrinsic average finite lifetime of an agent of species R), and a 2nd order process (1st with respect to R and X respectively) with rate constant k_{d2} . We further assume that the species from which the green arrows originate (the effectors) are not consumed in the processes they regulate.

We can reason why this motif can produce oscillatory activity as follows: the bistable switch established by the mutual activation between R and Ep (with S providing constant forcing) leads to either a high or low population level of R . As N_R increases so does the population of X (since the former promotes the latter's synthesis), which in turn increases R 's degradation rate thereby decreasing its

population. But as N_R gets low, so does N_X (X gets degraded faster than it is synthesized), thus allowing R to rebound and this cycle of interactions continues. However, one must tune the system's parameters to observe these oscillations. Here, we do not discuss the mathematics of determining the set of parameters for which a Hopf bifurcation occurs (the point in parameter space where a stable oscillation, or limit cycle, appears). Instead, we use a parameter set that leads to decaying oscillations and employ the ABK algorithm to model this motif for small populations of the involved species.

First, we show the system of differential equations that can be inferred from the motif's elementary processes shown in Figure 3.49.

$$\frac{dN_R}{dt} = (k_b N_{Ep} + k_s N_S) - k_{d1} N_R - k_{d2} N_R N_X \quad (3.50)$$

$$\frac{dN_{Ep}}{dt} = k_r N_R \frac{N_{E,TOT} - N_{Ep}}{K_{M,r} + (N_{E,TOT} - N_{Ep})} - k_f \frac{N_{Ep}}{K_{M,f} + N_{Ep}} \quad (3.51)$$

$$\frac{dN_X}{dt} = k'_b N_R - k'_d N_X \quad (3.52)$$

We assume that the conversion between E and Ep follows Michaelis-Menten kinetics and that $N_{E,TOT} = N_E(t) + N_{Ep}(t) = \text{constant}$. We show in Figure 3.50 the results of numerical integration of this system of equations (the deterministic time trajectory) and using the ABK methodology for this motif with the parameters $k_b = 0.01 \text{ sec}^{-1}$, $k_s = 0.05 \text{ sec}^{-1}$, $N_S = 15$, $k_{d1} = 0.0 \text{ sec}^{-1}$, $k_{d2} = 0.0075 \text{ sec}^{-1}$, $N_{E,TOT} = 100$, $k_f = 1 \text{ sec}^{-1}$, $k_r = 0.05 \text{ sec}^{-1}$, $K_{M,f} = 5$, $K_{M,r} = 10$, $k'_b = 0.001 \text{ sec}^{-1}$, $k'_d = 0.01 \text{ sec}^{-1}$, and initial population sizes $N_{R,i} = 0$, $N_{Ep,i} = 100$, $N_{X,i} = 0$. Numerical integration of the differential equation results in an attracting spiral fixed point with $N_R^* = 47.37$, $N_{Ep}^* = 93.31$, $N_X^* = 4.74$ (eigenvalues $\lambda_{1,2} = -0.021038 \pm 0.014763i$, $\lambda_3 = -0.08904$; the decaying oscillations are hardly visible in the deterministic trajectories shown in the figure, but are indeed present). The average time trajectories from the ABK simulation ($n = 500$; panel a) show excellent agreement with the deterministic curves. Panel b shows a sample trajectory revealing persistent high-amplitude fluctuations for all three population sizes, and are especially pronounced for $N_R(t)$. Note that $N_{E,TOT} = 100$ agents, which means that these fluctuations can only extend into the lower population sizes for N_{Ep} since its steady state is already close to $N_{E,TOT}$. This results in the ABK-simulated $\langle N_{Ep}(t) \rangle$ appearing slightly smaller than the deterministic solution to N_{Ep} (this is similar to what we saw in the 3-component negative feedback motif in the preceding section; see Figure 3.43). Notice that the population N_X stays in the single-digit range (near its fixed point) and is susceptible to changes in N_R (R stimulates the production of X). In addition, the low copy number of N_X makes it especially sensitive to perturbations due to noise, resulting in commensurately large changes in N_R since X stimulates the degradation of R . For instance, if $N_X(t) = 2$ for some time t and over a small time interval τ the population increases by only two agents (i.e., $N_X(t + \tau) = 4$), then the per- R -agent probability of a degradation event (through the process with rate constant k_{d2})

Follow this [link](#) to see the Matlab code.

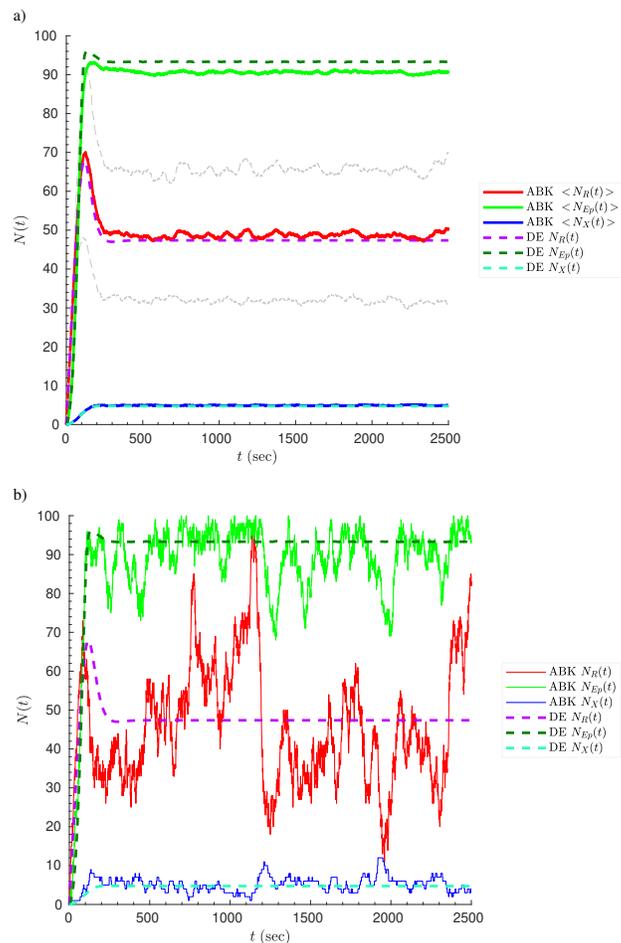


Figure 3.50 – Activator-inhibitor motif: average and sample time trajectories. We simulated the motif (Figure 3.49) for 2500 sec with fixed time step increments of $\Delta t = 0.02$ sec. Parameters: $k_b = 0.01 \text{ sec}^{-1}$, $k_s = 0.05 \text{ sec}^{-1}$, $N_S = 15$, $k_{d1} = 0 \text{ sec}^{-1}$, $k_{d2} = 0.0075 \text{ sec}^{-1}$, $N_{E,TOI} = 100$, $k_f = 1 \text{ sec}^{-1}$, $k_r = 0.05 \text{ sec}^{-1}$, $K_{M,f} = 5$, $K_{M,r} = 10$, $k'_b = 0.001 \text{ sec}^{-1}$, $k'_d = 0.01 \text{ sec}^{-1}$. Initial population sizes: $N_{R,i} = 0$, $N_{Ep,i} = 100$, $N_{X,i} = 0$. **a)** The average time trajectory of the species populations N_R , N_{Ep} , and N_X after 500 repetitions of the ABK simulation are shown in thick solid lines. The dashed curves correspond to the solution of the differential equations describing this motif. The one-standard-deviation envelope for the population of R , N_R , is also shown (gray dashed curves). **b)** A sample time trajectory of the population sizes are superimposed on the solution to the differential equations (obtained through numerical integration). We again observed high-amplitude irregular fluctuations for species R .

doubles. That is, for our example,

$$\frac{P_{dif,R}(t + \tau)}{P_{dif,R}(t)} = \frac{k_{d2}N_X(t + \tau)}{k_{d2}N_X(t)} = 2, \quad (3.53)$$

and the population of R would decrease accordingly (this is the same argument as presented in subsection 3.5.1 leading to equation 3.44).

We also plot the coefficient of variation, η , for each of the species in Figure 3.51. The noise we observe in our simulations is greater for species R than expected for a Poisson process, while the opposite is true for species X (we made similar observations when studying the 2-component negative feedback motif; see Figures 3.38c and 3.39c).

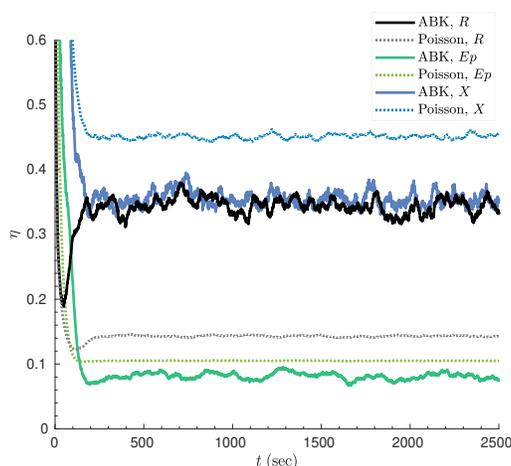


Figure 3.51 – Activator-inhibitor motif: coefficient of variation. Comparison of the coefficient of variation, $\eta(t)$, for each of the species with its predicted value, $\eta_P(t)$, assuming a Poisson process. The trajectory of species R shows a significantly larger deviation from its mean than deterministically expected, while the average trajectory of X is less noisy than Poissonian statistics would predict.

As we saw in the preceding subsection on negative feedback loops, regulatory interactions mediated by species with small population sizes can lead to large fluctuations in the populations sizes of other species, even when the system's parameters do not explicitly allow for the presence of a deterministic limit cycle. The present case appears to be yet another example of an excitable system exhibiting persistent high-amplitude fluctuations.

3.5.5 Substrate-Depletion Motif

The substrate-depletion motif, shown in the figure below, can also produce oscillations. The substrate is species X in the sense that it can be converted to R , while R and Ep form a mutual activation loop. When the population of X is large, a surge of R production ensues and the pool of X agents gets depleted. As the production of R slows down and its population decreases through degradation, X rebounds again and the cycle continues.

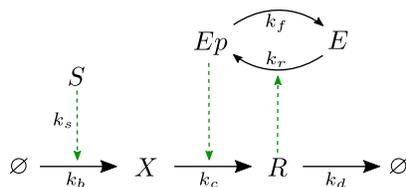


Figure 3.52 – Substrate-depletion motif. Species S promotes the synthesis of the substrate X that is converted to R . Species R and Ep form a positive feedback (mutual activation) loop. The right combination of parameter values can produce oscillatory behavior in this motif.

The set of differential equations for this motif is

$$\frac{dN_R}{dt} = (k_{c1}N_X + k_{c2}N_XN_{Ep}) - k_dN_R \quad (3.54)$$

$$\frac{dN_{Ep}}{dt} = k_rN_R \frac{N_{E,TOT} - N_{Ep}}{K_{M,r} + (N_{E,TOT} - N_{Ep})} - k_f \frac{N_{Ep}}{K_{M,f} + N_{Ep}} \quad (3.55)$$

$$\frac{dN_X}{dt} = (k_b + k_sN_S) - (k_{c1}N_X + k_{c2}N_XN_{Ep}) . \quad (3.56)$$

Notice that the conversion of X to R is achieved through two processes: the 1st order conversion $X \rightarrow R$ with rate constant k_{c1} (the effective basal conversion rate) and the 2nd order transition $X + Ep \rightarrow R$ with rate constant k_{c2} (we are assuming that Ep is *not* consumed in this process). As in previous cases, we also assume that E and Ep are not synthesized nor degraded via processes beyond the indicated Michaelis-Menten conversion between the two. Therefore, $N_{E,TOT} = N_E(t) + N_{Ep}(t) = \text{constant}$.

We simulated this motif using the ABK algorithm and compared the results we obtained to the numerical solution of the system of differential equations presented above. We used the parameters $k_b = 0.00 \text{ sec}^{-1}$, $k_s = 0.5 \text{ sec}^{-1}$, $N_S = 10$, $k_{c1} = 0.001 \text{ sec}^{-1}$, $k_{c2} = 0.00903225 \text{ sec}^{-1}$, $k_d = 0.1 \text{ sec}^{-1}$, $N_{E,TOT} = 100$, $k_f = 1 \text{ sec}^{-1}$, $k_r = 0.005 \text{ sec}^{-1}$, $K_{M,f} = 5$, $K_{M,r} = 10$, and initial population sizes $N_{R,i} = 0$, $N_{Ep,i} = 100$, $N_{X,i} = 0$. Given these values, the motif is deterministically predicted to have a single fixed point at $N_R^* = 50$, $N_{Ep}^* = 1.468$, $N_X^* = 350.644$, with eigenvalues $\lambda_{1,2} = \pm 0.0270i$, $\lambda_3 = -0.234$ (the real component of the complex pair of eigenvalues is approximately zero: $Re(\lambda_{1,2}) < 1 \times 10^{-9}$). We chose k_{c2} to have this particular value (shown to 6

Follow this [link](#) to see the Matlab code.

significant figures) because it makes the real component of $\lambda_{1,2}$ effectively zero given the other aforementioned kinetic parameters, thus producing a limit cycle whose amplitude decays only according to the exponential $e^{-\lambda_3 t}$. The result is a deterministic time trajectory with sizable oscillations in the order of 10^5 seconds of simulated time.

We employed the ABK algorithm and simulated this motif for 2500 sec with fixed time step increments of $\Delta t = 0.02$ sec. We show a sample and average ($n = 200$) time trajectory of species R in Figure 3.53. Note that we do not show all three species in the same plot because they oscillate around population sizes of different orders of magnitude: $N_{Ep}(t)$ oscillates in the single-digit population size regime, while $N_X(t)$ wildly oscillates from as little as 20 to 1200 agents in our simulations.

Our observations for this motif are similar to those presented in the preceding subsections on other oscillator systems. The population of R is not just dependent on the oscillatory trajectory inherent to the dynamics of the system, but also on random fluctuations that can profoundly affect the probability of transitions involving the low copy number population N_{Ep} in the 2nd order conversion of X to R . This results in larger than predicted oscillations in N_R (in fact, they would best be characterized as “spikes”; see Figure 3.53a), with the time interval between oscillations showing some variance (we performed spectral analysis on the data confirming this variation in the frequency domain; see Figure 3.53c). A simulated time trajectory thus appears at times out-of-phase compared to the deterministic solution. This is in agreement with our previous observations that stochasticity can lead to dephasing of individual simulation runs. The irregularity in the frequency and amplitude of the oscillations/spikes results in an average trajectory (Figure 3.53b) that appears aperiodic and noisy. Finally, we also plot the coefficient of variation, $\eta(t)$, which confirms that the high-amplitude oscillations make the average trajectory of R have a far greater variance than expected for a Poisson process.

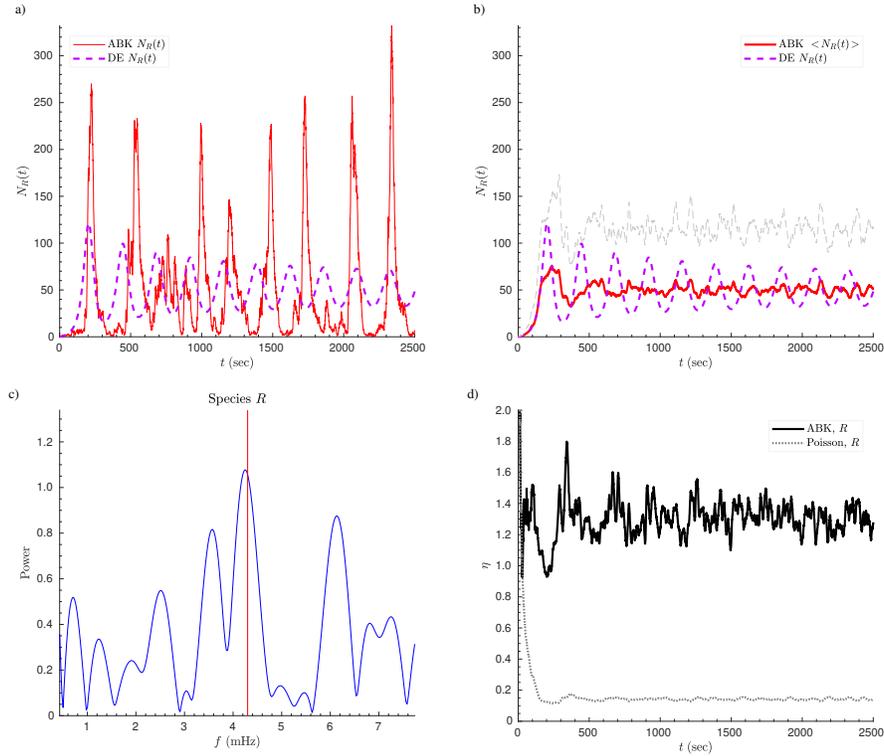


Figure 3.53 – Substrate-depletion motif: sample and average time trajectories. We analyzed the substrate-depletion motif (Figure 3.52) using the ABK algorithm. Parameters: $k_b = 0.00 \text{ sec}^{-1}$, $k_s = 0.5 \text{ sec}^{-1}$, $N_S = 10$, $k_{c1} = 0.001 \text{ sec}^{-1}$, $k_{c2} = 0.00903225 \text{ sec}^{-1}$, $k_d = 0.1 \text{ sec}^{-1}$, $N_{E,TOT} = 100$, $k_f = 1 \text{ sec}^{-1}$, $k_r = 0.005 \text{ sec}^{-1}$, $K_{M,f} = 5$, $K_{M,r} = 10$. ABK simulation parameters: 2500 sec of simulated time with fixed time step increments of $\Delta t = 0.02 \text{ sec}$. Initial population sizes: $N_{R,i} = 0$, $N_{E_p,i} = 100$, $N_{X,i} = 0$. In the interest of clarity, we show the results only for species R (X and E_p show similar behavior). **a)** Sample time trajectory, $N_R(t)$, using the ABK algorithm superimposed on the numerical solution to the system of differential equations (DE). **b)** Average time trajectory $\langle N_R(t) \rangle$ ($n = 200$). The gray dashed curve is the mean plus one standard deviation. **c)** Power spectrum of a time trajectory ($t = 500 - 2500 \text{ sec}$) in a sample ABK simulation run. The horizontal axis represents frequency in units of mHz. A periodic population trajectory corresponds to a peak at a finite frequency in the spectrum. The vertical red line shows the theoretically predicted oscillation frequency, obtained through the imaginary component of the eigenvalues, $\text{Im}(\lambda_{1,2})/2\pi = 0.0270/2\pi = 4.30 \text{ mHz}$, corresponding to a period of approximately 233 sec. **d)** Comparison of the coefficient of variation, $\eta(t)$, for species R with that of a Poisson process, $\eta_P(t)$. The trajectory of species R shows a significantly larger deviation from its mean than expected.

3.6 Other Models

We end this chapter with implementing the ABK method to a handful of models whose utility lies in their mathematical simplicity. We first discuss two species competition models: the Lotka-Volterra model of predator-prey interactions and the Kermack-McKendrick model of infectivity spread. In the former case, our simulations correctly predict oscillatory activity, while, in the latter, we discuss how the occurrence of a species extinction event, an *absorbing state* of the discrete dynamics (a system state that does not change once it is reached), affects the average population time trajectories obtained from an ensemble of ABK simulations. Finally, we present a much-studied model of oscillations originally proposed as an analogue of oscillating chemical systems: the Brusselator. This is an interesting case study because *relaxation oscillations* are predicted for the right set of kinetic parameters. Our simulations indeed reproduce this general behavior and show that our methodology can effectively model processes operating at different time scales. As a side note, the original form of the Brusselator includes an arguably unrealistic termolecular process. We address this by simulating a more realistic version of the Brusselator, which comprises only up to 2nd order processes.

3.6.1 The Lotka-Volterra Predator-Prey Model

We now discuss a well-known model of competition between a predator and a prey species that leads to oscillations in their respective population sizes. The usefulness in considering this model is its generality: the idea of competition can be used to describe a variety of phenomena beyond organismal ecology. For instance, the interplay between immune system cells and virus particles in an infected patient, or the sales of competing retail formats (Hung et al., 2017), are both examples of applying the model to a diverse array of situations.

The model, proposed independently by the mathematicians Alfred Lotka and Vito Volterra (Lotka, 1925; Volterra, 1926) almost a century ago, is simple and not meant to strictly represent elementary interactions between agents of the two species. It is therefore a *phenomenological model* that correctly predicts oscillatory dynamics but also has other weaknesses making it unsuitable for use in describing the dynamics of real populations (we will expose some of these shortcomings in the discussion that follows). Even so, we interpret the model in terms of agent interactions using the ABK methodology to show that it agrees with the deterministic predictions while also allowing for more realistic scenarios such as extinction events due to stochastic fluctuations in small population sizes.

We consider agents of two species, R (the prey; e.g., rabbits) and F (the predators; e.g., foxes). We will use the kinetic rate constants k_b , k_d , k'_b , k'_d , where the subscripts b and d refer to the birth and death processes of a species. We define the following processes to describe the intra- and interspecies agent

interactions (equivalently stated, the model’s mechanism):



It is our aim to define these elementary processes in a way that makes sense for the system at hand. The first two processes describe the birth and death of R , respectively. k_b is 1st order kinetic constant representing the intrinsic growth rate of species R , where agent R gives “birth” to another agent of the same species. On the other hand, k_d is a 2nd order constant reflecting the stoichiometric relationship between R and F leading to the consumption of R (alternatively, one can imagine a fox killing a rabbit).²³ Similarly, the latter two processes describe the birth and death of F , with k'_b and k'_d being a 2nd and 1st order constant respectively. The process 3.59 describing the birth of F depends on R but does *not* directly involve its consumption. Considering this assumption at the level of populations, it means that the growth of species F depends on the population size of R . Intuitively, the predator’s growth rate correlates with the availability of prey. Finally, the disappearance of F is a 1st order degradation process.

This set of processes leads to the 2-dimensional set of differential equations,

$$\frac{dN_R}{dt} = k_b N_R - k_d N_R N_F \quad (3.61)$$

$$\frac{dN_F}{dt} = k'_b N_R N_F - k'_d N_F, \quad (3.62)$$

where N_R and N_F denote the discrete and time-dependent size of the species populations. Setting both equations equal to zero, we can easily obtain two fixed points: $(0, 0)$ and $\left(\frac{k'_d}{k'_b}, \frac{k_b}{k_d}\right)$, where the ordered pairs represent (N_R^*, N_F^*) in phase/state space. We then linearize by evaluating the Jacobian matrix at each of the two fixed points, revealing that the origin is a *saddle* and the point $\left(\frac{k'_d}{k'_b}, \frac{k_b}{k_d}\right)$ a *center*. We are primarily interested in the latter, since closed trajectories are predicted to occur in its vicinity. First, we must confirm that the prediction of a center from linearization indeed results in a *nonlinear* center. To show this, it suffices to find a quantity, $E(N_R, N_F)$, that is invariant through a given trajectory²⁴ and that the conserved quantity has an extremum at the putative center’s coordinates in phase space (Strogatz, 2015, Theorem 6.5.1, p. 163). To find $E(N_R, N_F)$, we obtain an expression for $\frac{dN_R}{dN_F}$ that upon separation of variables and integration reveals the conserved quantity

$$E(N_R, N_F) = k'_d \ln N_R + k_b \ln N_F - k'_b N_R - k_d N_F. \quad (3.63)$$

²³From the perspective of population biology, notice that the model ignores the “death” of R agents *not* being the result of predation (i.e., due to natural causes).

²⁴That is, to show that this is a conservative system.

It is then easy to show that the point $\left(\frac{k'_d}{k'_b}, \frac{k_b}{k_d}\right)$ is an extremum of E , thereby satisfying the theorem's requirements. We conclude that $\left(\frac{k'_d}{k'_b}, \frac{k_b}{k_d}\right)$ is a *nonlinear center*. It is important to note this point in phase space is *not* enclosed in an attracting limit cycle, but is instead a center. This means that the deterministic solution to the set of differential equations has a family of closed trajectories around the center (i.e., the fixed point is *neutrally stable*) leading to oscillations of differing amplitudes depending on what the initial condition is. This is perhaps the model's greatest weakness: *it is unstable to the smallest of perturbations*. A “nudge” of arbitrary strength pushes the system to a new closed trajectory. This has generally not been observed to be true in real predator-prey populations where the oscillations tend to have a characteristic amplitude. In terms of performing a stochastic simulation on this system, we expect stochastic fluctuations to severely affect the oscillation amplitude.

The eigenvalues of the nonlinear center provide information about the frequency of the deterministically predicted oscillations. These eigenvalues are purely imaginary and depend only on the prey's birth and the predator's death rate constants: $\lambda = \pm\sqrt{k_b k'_d}i$. Thus, the predicted oscillation frequency is

$$f = \frac{\sqrt{k_b k'_d}}{2\pi}. \quad (3.64)$$

We now present an example of this model with the following parameters: $k_b = 1 \text{ month}^{-1}$, $k_d = 0.01 \text{ month}^{-1}$, $k'_b = 0.01 \text{ month}^{-1}$, $k'_d = 1 \text{ month}^{-1}$.²⁵ We show the phase/state space in Figure 3.54a, where we plot the deterministic trajectory (black curve) for a total time of 20 months, superimposed on the direction field predicted by the differential equations 3.61 and 3.62. The trajectory moves counterclockwise around the center, which occurs at $(N_R^*, N_F^*) = (100, 100)$, with the initial population sizes of $N_{R,i} = 70$ and $N_{F,i} = 70$ (red star in the figure) being part of the closed loop. In fact, any initial condition that is not on the axes will be part of a unique closed trajectory representing unattenuated oscillations. This clarifies our earlier comment that the model is unstable: any initial condition is part of a different loop in phase/state space. We therefore expect the amplitude of a trajectory obtained through simulation to be particularly sensitive to the fluctuations inherent to a stochastic system.

We simulated the Lotka-Volterra (LV) model using the ABK methodology using the same parameters as noted in the preceding paragraph and for a total simulated time of 50 months with fixed time increments of $\Delta t = 0.01$ months. We noticed two limiting cases where stochastic fluctuations led to extinction events: a) if the population of R dwindles to zero (i.e., R becomes extinct),²⁶ then species F also eventually vanishes since it has no R agents to consume (co-extinction); b) if F goes extinct, then the population of R follows unhindered

Follow this [link](#) to see the Matlab code.

²⁵The choice of these numerical values and time units is arbitrary and made for the purpose of demonstrating how to model this system using ABK.

²⁶The model dictates that the birth of an R agent depends on the presence of a “parent” R agent (process 3.57). Therefore, if at any time during the course of the simulation $N_R(t) = 0$, then the species is extinct.

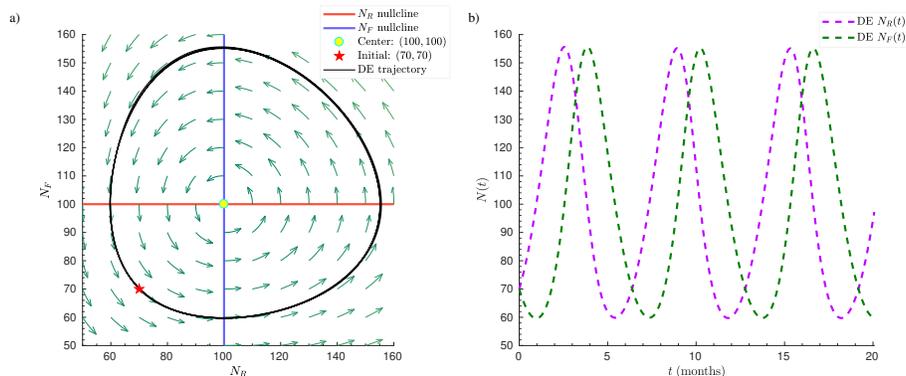


Figure 3.54 – Lotka-Volterra (LV) model: phase plane and deterministic time trajectory. **a)** We evaluated the velocity vector $(dN_R/dt, dN_F/dt)$ at each ordered pair (N_R, N_F) according to the differential equations (DEs, equations 3.61 and 3.62) to construct the direction field (green arrows). We also show the two nullclines obtained by setting the DEs equal to zero. Kinetic parameters: $k_b = 1 \text{ month}^{-1}$, $k_d = 0.01 \text{ month}^{-1}$, $k'_b = 0.01 \text{ month}^{-1}$, $k'_d = 1 \text{ month}^{-1}$. The center occurs at $(100, 100)$ and the initial condition is shown as a red star at $(70, 70)$. The deterministic trajectory (black curve) starts at the initial condition coordinate and traverses a closed path in the counterclockwise direction with a frequency of $f \approx 159 \times 10^{-3} \text{ month}^{-1}$ and period $T \approx 6.28 \text{ months}$ (calculated using equation 3.64). **b)** Time-series representation of the solution to the DEs.

exponential growth. We repeated the simulation 1000 times and removed trajectories that followed either of these limiting cases to assess the average behavior of an ensemble of extinction-free trajectories. In all, we removed 343 out of 1000 repetitions of the simulation, which suggests that there is approximately a probability of 0.35 for one or both species becoming extinct given these kinetic parameters and initial population sizes.²⁷ Moreover, the number of extinction events is approximately split in half between the two cases mentioned in the beginning of this paragraph. Being able to obtain an estimate for the probability of extinction is a distinct advantage of stochastic simulations compared to their deterministic analogue.

We show a typical extinction-free trajectory in Figure 3.55a, where a large variation in the amplitude of the oscillations is evident. Panel b of the same figure shows the average trajectory of the 657 extinction-free simulation trials. The observed damping of the oscillations in the average trajectory is not representative of the individual simulation runs. Instead, it is the result of stochastic fluctuations, which, in addition to affecting the amplitude, lead to phase shifts in individual trajectories that produce the observed attenuation in the aggregate.

It is instructive to perform a simulation of this model for the same parameters but a different set of initial conditions: $N_{R,i} = 100$ and $N_{F,i} = 100$. The deterministic solution shows no change in the population sizes since they start out right at the predicted center. However, the stochastic simulation reveals

²⁷Pooling these results with additional simulations gave 35.25% ($n = 2000$) as the estimate for the frequency of an extinction event.

that the population sizes oscillate (a sample run is shown in Figure 3.55c) with enough variation in the amplitude, phase, and frequency of the oscillations such that the average trajectory agrees with the deterministic prediction (panel d of the same figure).

It is also noteworthy that the average trajectories for the two initial conditions we simulated are different with regard to the oscillatory activity in the ensemble of simulations (Figure 3.55b,d). When the initial populations are at the center and stochastically fluctuate in its state-space neighborhood, there is little tendency for the system to complete a cycle in the first few months of simulated time. Even though the individual trajectories eventually settle into a robust high-amplitude oscillatory pattern (see Figure 3.55c), they show no coherence, thus resulting in an average trajectory without oscillations. In contrast, we see that in the case of $N_{R,i} = 70$ and $N_{F,i} = 70$, the individual trajectories start out oscillating coherently (as evidenced by the first 10 – 20 months of the average trajectories in Figure 3.55b) before fluctuations result in reduced phase synchronization and attenuation of the ensemble’s observed oscillatory pattern.

The above presentation of the predator-prey competition model showcases the ability of stochastic simulation methods to provide deeper insights in the dynamics of phenomenological models than the deterministic approach. The advantage of using the ABK algorithm, however, lies in allowing the study of the effects of population heterogeneity on the system’s dynamics. We will explore these effects in future investigations.

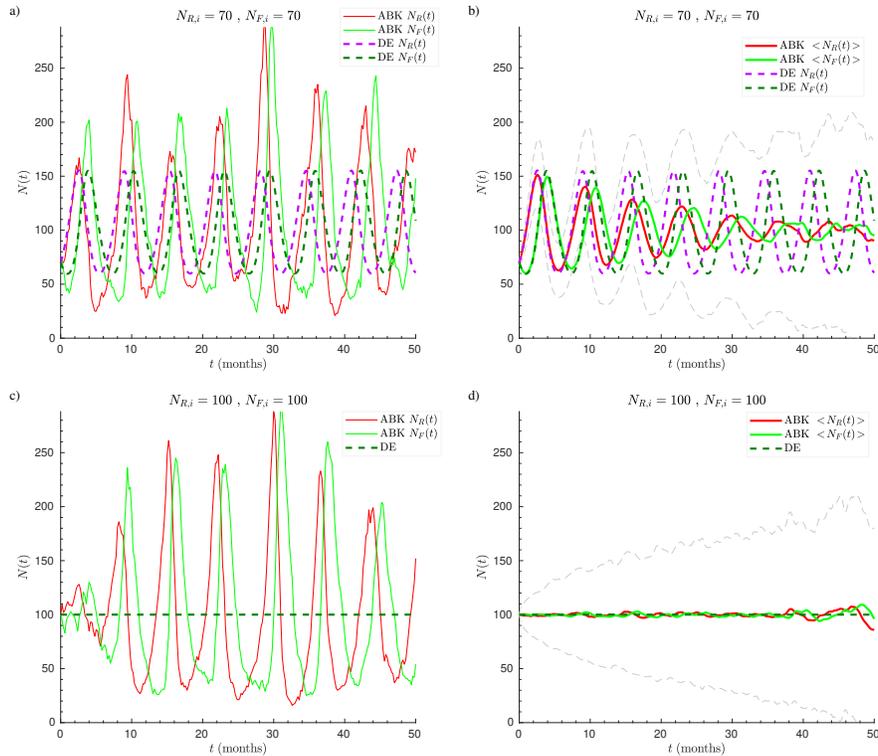


Figure 3.55 – Stochastic LV model: sample and average time trajectories. We used the ABK algorithm to simulate the Lotka-Volterra model with parameters $k_b = 1 \text{ month}^{-1}$, $k_d = 0.01 \text{ month}^{-1}$, $k'_b = 0.01 \text{ month}^{-1}$, $k'_d = 1 \text{ month}^{-1}$. Total simulated time: 50 months; fixed time step increment $\Delta t = 0.01$ months. We show here a sample and the average time trajectory of the populations for two initial conditions: $N_{R,i} = 70$ and $N_{F,i} = 70$ (plots *a*, *b*), and $N_{R,i} = 100$ and $N_{F,i} = 100$ (plots *c*, *d*). The average population sizes were obtained by repeating the simulation 1000 times and averaging over the extinction-free trajectories ($n = 657$ in plot *b*, $n = 701$ in plot *d*). All simulation results are superimposed on the deterministic solution (labeled “DE” in the legends). We also show the one-standard-deviation envelope for $\langle N_R(t) \rangle$ (gray dashed curve in panels *b*, *d*).

The Volterra predator-prey model

We now briefly present a variation of the aforescribed predator-prey model, where we consider the growth of species R as density-dependent. The so-called Volterra equations are the same as 3.61 and 3.62, except the term corresponding to the birth of R agents includes a factor featuring the carrying capacity, K :

$$\begin{aligned}\frac{dN_R}{dt} &= k_b N_R \left(1 - \frac{N_R}{K}\right) - k_d N_R N_F & (3.65) \\ \frac{dN_F}{dt} &= k'_b N_R N_F - k'_d N_F.\end{aligned}$$

Notice that the dN_F/dt equation is the same as 3.62. It is easily seen that as $K \rightarrow \infty$, equation 3.65 reduces to 3.61. Importantly, the two models have different dynamical behavior: the Lotka-Volterra model features neutrally stable closed orbits in phase space because of the presence of a conserved quantity (equation 3.63), while, as we will demonstrate, the introduction of a finite carrying capacity rules out the possibility of limit cycles in the Volterra system.

We are interested in the nontrivial fixed point for this 2-dimensional system, given by

$$N_R^* = \frac{k'_d}{k'_b} \quad (3.66)$$

$$N_F^* = \frac{k_b}{k_d} \left(1 - \frac{k'_d}{k'_b K}\right). \quad (3.67)$$

Note that the factor in parentheses is always positive for $0 < N_{R,i} < K$ since $N_R(t) \rightarrow N_R^*$ as $t \rightarrow \infty$ and $N_R^* < K$.²⁸ Another way to state this is to substitute equation 3.66 into the term in parentheses of equation 3.67 to get

$$\left(1 - \frac{k'_d}{k'_b K}\right) = \left(1 - \frac{N_R^*}{K}\right) > 0. \quad (3.68)$$

We then perform stability analysis to assess the system's behavior in the neighborhood of the above fixed point. It can be shown using methods described in Appendix C that the trace and determinant of the system's Jacobian matrix are

$$\tau = -k_b \frac{k'_d}{k'_b K} \quad (3.69)$$

$$\Delta = k_b k'_d \left(1 - \frac{k'_d}{k'_b K}\right). \quad (3.70)$$

Since all parameters in our system are positive (by definition), it follows that $\tau < 0$ and $\Delta > 0$, which ensures that the fixed point is either a stable node

²⁸The reader can verify this by inspecting equation 3.65.

or spiral. Distinguishing between these two possibilities can be determined by evaluating the discriminant of the characteristic equation (see equation C.32 and its solution, C.33), such that when $\tau^2 - 4\Delta < 0$ the eigenvalues will be complex, resulting in damped oscillations. Obtaining an algebraic expression for the discriminant, we get

$$\tau^2 - 4\Delta = k_b k'_d \left(k_b \frac{k'_d}{(k'_b K)^2} + 4 \frac{k'_d}{k'_b K} - 4 \right). \quad (3.71)$$

The sign of the discriminant depends on the sign of the term in parentheses, which can be analyzed as a quadratic function of $(k'_b K)^{-1}$. Subsequently, we obtain the condition for a stable spiral,

$$K > \left[\frac{2k'_b}{k_b} \left(\sqrt{1 + \frac{k_b}{k'_d}} - 1 \right) \right]^{-1}. \quad (3.72)$$

For the same kinetic parameters as the ones we used in the Lotka-Volterra model ($k_b = 1 \text{ month}^{-1}$, $k_d = 0.01 \text{ month}^{-1}$, $k'_b = 0.01 \text{ month}^{-1}$, $k'_d = 1 \text{ month}^{-1}$), the right side of the above inequality evaluates to 120.71. Therefore, any value for the carrying capacity above this threshold will yield damped oscillations, as can be clearly seen in the phase plane of the Volterra system with $K = 500$ (Figure 3.56). It is evident by comparing this figure to 3.54a that the N_R nullcline now has a nonzero slope and $N_F^* = 80$ (as opposed to 100 in the Lotka-Volterra case). Examining equation 3.67, it is clear that $N_F^* \rightarrow 100$ as $K \rightarrow \infty$, which agrees with our earlier assertion that the Volterra system converges to the Lotka-Volterra one as the carrying capacity becomes infinite.

We continue by performing an ABK simulation of this model, as shown in Figure 3.57. Panel a shows the species trajectories for a single simulation run, revealing that persistent high-amplitude fluctuations (PHAF) occur for this system. This is in agreement with our previous observations on oscillatory systems with stable spiral fixed points (see section 3.5). Meanwhile, the average of the trajectories from an ensemble of simulations ($n = 1000$) agrees with the deterministic expectation (panel b). As we have done before, we can verify the presence of high-amplitude fluctuations by plotting the coefficient of variation for both species trajectories (panel c). Indeed, η is more than three times higher than would be expected for a Poisson process ($\eta_P = 1/\sqrt{\langle N(t) \rangle}$). Finally, we computed the discrete Fourier transform to obtain the power spectrum of the species trajectories for each of the simulation runs. Figure 3.57d shows that the average of all the power spectra for species R trajectories has a peak at a frequency of approximately $134 \times 10^{-3} \text{ month}^{-1}$, which is within 5% of the theoretically predicted oscillation frequency $f = \text{Im}(\lambda_{1,2})/2\pi = 141.46 \times 10^{-3} \text{ month}^{-1}$ (corresponding to a period of approximately 7.069 months). We will further investigate the reasons for the discrepancy between the observed and deterministically predicted frequency value in future work.

We have also performed simulations for values of K that result in the fixed point being a stable node, and verified that high-amplitude fluctuations do not

Follow this [link](#) to see the Matlab code.

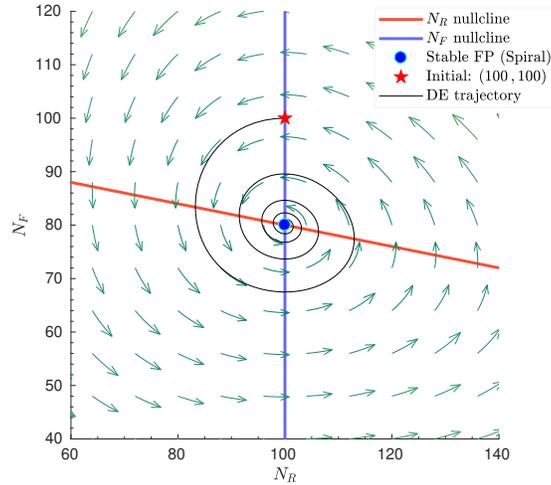


Figure 3.56 – Volterra model: phase plane. a) We evaluated the velocity vector (dN_R/dt , dN_F/dt) at each ordered pair (N_R, N_F) according to the differential equations (equations 3.65 and 3.62) to construct the direction field (green arrows). We also show the two nullclines. The kinetic parameters are the same as listed in Figure 3.54 and $K = 500$. The stable fixed point occurs at $(100, 80)$ and the initial population sizes are shown on the plane as a red star at $(100, 100)$. The deterministic trajectory (black curve) spirals in a counterclockwise direction into the stable fixed point (the eigenvalues are $\lambda_{1,2} = -0.10000 \pm 0.88882i$).

occur in those cases (not shown here). This is in keeping with our results from studying oscillatory systems that these fluctuations only occur when damped oscillations are predicted using the deterministic formalism. McKane and Newman (2005) have made similar *in silico* observations when studying the Volterra system.

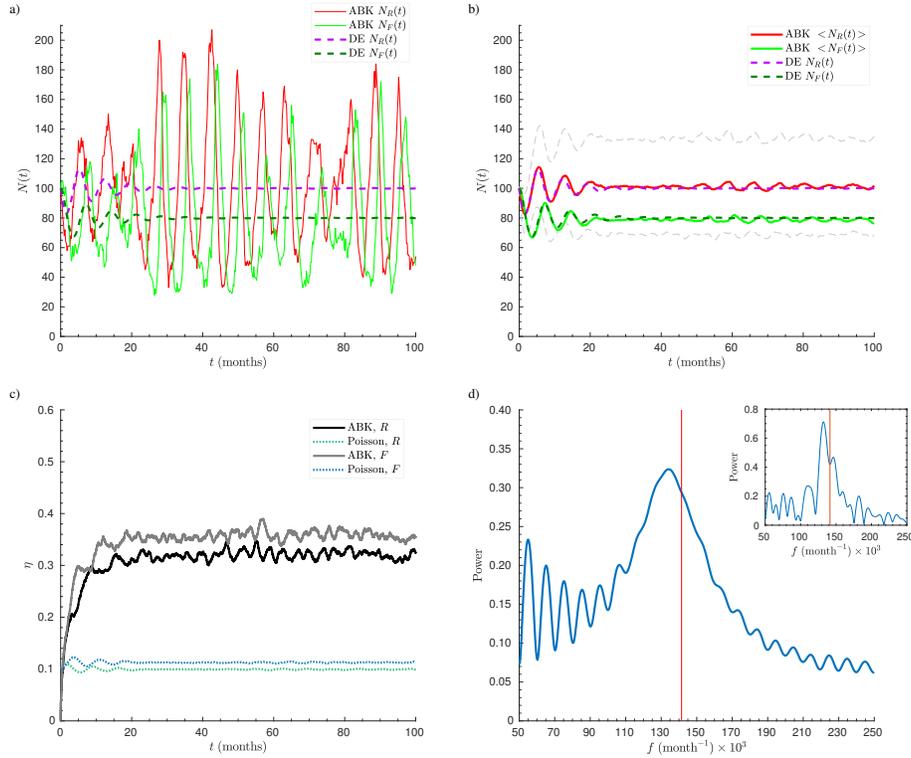


Figure 3.57 – Stochastic simulations of the Volterra model. We used the ABK algorithm to simulate the Volterra model with parameters $k_b = 1 \text{ month}^{-1}$, $k_d = 0.01 \text{ month}^{-1}$, $k'_b = 0.01 \text{ month}^{-1}$, $k'_d = 1 \text{ month}^{-1}$, $K = 500$. Total simulated time: 100 months; fixed time step increment $\Delta t = 0.01$ months. **a,b)** Sample and the average time trajectory ($n = 1000$) of the populations for the initial condition $N_{R,i} = 100$ and $N_{F,i} = 100$. All simulation results are superimposed on the deterministic solution (labeled “DE” in the legends). We also show the one-standard-deviation envelope for $\langle N_R(t) \rangle$ (gray dashed curve in panel b). **c)** Comparison of the coefficient of variation, $\eta(t)$, for species R and F with that of a Poisson process. Both species trajectories show a significantly larger deviation from their mean than expected. **d)** Average power spectrum of the N_R trajectories obtained in the ensemble ($n = 1000$) of ABK simulations. A periodic population trajectory corresponds to a peak at a finite frequency in the spectrum. The vertical red line shows the theoretically predicted oscillation frequency, obtained through the imaginary component of the eigenvalues, $f = \text{Im}(\lambda_{1,2})/2\pi = 141.46 \times 10^{-3} \text{ month}^{-1}$, corresponding to a period of approximately $T \approx 7.07$ months. Inset: power spectrum for the sample N_R trajectory shown in panel a.

3.6.2 The Kermack-McKendrick Model

We present the application of the ABK method to the model proposed by Kermack and McKendrick (1927) on the evolution of an epidemic, a foundational model in the field of epidemiology. We are especially interested in demonstrating the effect that an absorbing state of the discrete dynamics has on observing deviations from the deterministic predictions when simulating the time evolution of small population sizes.

We consider the spread of a disease through a population made up of three types (species, or compartments) of agents: Susceptible (S) agents who can become infected (I) upon contact with other infected agents. I individuals can in turn recover from the infection (R). We denote the number of discrete agents in the population as N_S , N_I , and N_R , respectively. The system of differential equations that Kermack and McKendrick proposed is

$$\frac{dN_S}{dt} = -k_c N_S N_I \quad (3.73)$$

$$\frac{dN_I}{dt} = k_c N_S N_I - k_d N_I \quad (3.74)$$

$$\frac{dN_R}{dt} = k_d N_I, \quad (3.75)$$

where the constant k_c is a measure of the average rate of contagion or infectivity, and k_d is a measure of the average recovery rate for infected (I) agents. The reader can ascertain that the above system of differential equations implies that $N_S + N_I + N_R = \text{constant}$, which suggests that the time scale over which these processes occur is small compared to the lifespan of the population's members. Therefore, we assume there are no deaths in the population and births or emigration/immigration are absent (i.e., no demographic processes). Also notice that the differential equations for N_S and N_I have the same form as the ones in the Lotka-Volterra model except that there is no term for the “birth” of S agents. This model can accordingly be viewed as a competition between species S and I .

The equations show that healthy or susceptible agents become infected at a rate proportional to $N_S N_I$, the number of distinct pairwise interactions between members of these two populations. Moreover, recoveries from infection occur at a constant rate in proportion to the number of infected people present at any given time. We thus transform the system of equations into an equivalent set of distinct processes that we can explicitly model using the ABK method:



k_c is the microscopic 2nd order rate constant for the conversion of S to I , and k_d is the microscopic 1st order rate constant for the conversion of I to R .

We simulated this reaction scheme by implementing the ABK algorithm for concurrent processes (I participates in both processes; see section 2.7). We

evaluated the transition probability forms P_{ber} or P_{dif} with respect to I with KCFs $\Omega_I = k_c N_S(t_n) \Delta t$ and $\Omega_I = k_d \Delta t$ for the two respective processes. For this particular example, we used $k_c = 0.03 \text{ days}^{-1}$ and $k_d = 0.05 \text{ days}^{-1}$ with fixed time step increments of $\Delta t = 0.01$ days. We observed that the initial number of I agents serves as a predictor of the agreement between the simulation average time trajectories and the numerical solution to the system of differential equations (Figure 3.58). When the initial number of infected agents, $N_{I,i}$, is very small, say 1 or 2, then there was significant deviation from the deterministic trajectories. This is not a failure of the ABK algorithm, but rather a consequence of the fact that $N_I = 0$ represents the extinction of I agents, an *absorbing state* of the discrete dynamics (i.e., once it is reached, the system state does not change). For instance, if we start monitoring the population when there is only one I agent, then we would expect there is a sizable probability that it will recover before converting healthy individuals to infected ones. The apparent discrepancy in the figure is then simply the result of averaging over trajectories in which the populations of I and R do not grow and no epidemic occurs. For the parameters we used in this simulation (Figure 3.58b) we observed that in approximately 10% of the simulations, N_I decreases to zero within the first few simulated days and no epidemic occurs (also see Figure 3.59, ABK sample trajectory #1 drawn in S - I state space).²⁹ This approach suggests that an ensemble of agent-based simulated trajectories can be used to approximate the probability of a species extinction. Removing the trajectories in which the disease did not spread from the ensemble of simulations produced average trajectories that match the deterministic predictions (not shown).

We finally add that our simulations correctly predicted other expected properties of this model. For instance, it can be shown (Strogatz, 2015) that the following condition has to be true to have an epidemic:³⁰

$$0 < \frac{k_d}{k_c N_{S,i}} < 1. \quad (3.78)$$

Note that the term “epidemic” formally means that the number of infected people increases before it goes down. Alternatively stated, $N_I(t)$ and dN_R/dt have a maximum at a time t_{peak} , the peak of the epidemic. An ensemble of our simulations correctly predicted this t_{peak} value (e.g., see solid red curve in Figure 3.58a).

This simple example highlights the importance of stochasticity in modeling processes with small populations. Since every epidemic can be thought of as

²⁹There are three variables in this model and therefore one should visualize the full state space in three dimensions. However, we ignore the number of recovered individuals, N_R , for the purposes of this figure to keep these trajectories easily traceable and visualized on a 2-dimensional plane.

³⁰This condition can be restated in terms of the epidemiological metric R_0 , the *basic reproductive number* (the expected number of secondary infections from an infected individual given a fully susceptible population of healthy individuals). In the model presented in this section, we define $R_0 = \frac{k_c N_{S,i}}{k_d}$. Thus, the condition 3.78 can be restated as $0 < 1/R_0 < 1$. Equivalently, an epidemic occurs when $R_0 > 1$.

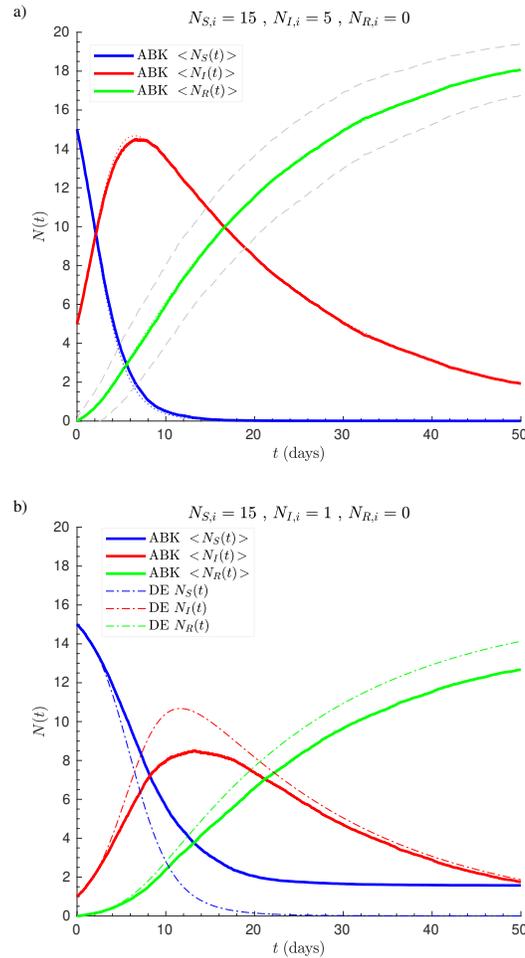


Figure 3.58 – Simulation of the Kermack-McKendrick model. We show here the results from two sets of simulations of the Kermack-McKendrick model with kinetic constants $k_c = 0.03 \text{ days}^{-1}$ and $k_d = 0.05 \text{ days}^{-1}$. ABK simulation parameters: we used the P_{ber} or P_{dif} transition probability expressions evaluated with respect to I for both processes, with fixed time step increments of $\Delta t = 0.01$ days. The two plots are different in the initial number of infected agents, $N_{I,i}$, as indicated in the plot headers, but all other parameters are identical. In both plots, the average trajectories ($n = 500$) of the populations are shown as solid curves (the one-standard-deviation envelope for $N_R(t)$ is also shown on plot a). The deterministic trajectories are shown as dotted curves in plot a and as dash-dot curves in plot b. The apparent significant deviation of the average simulated trajectories from the deterministic predictions in plot b is due to the lone I agent recovering before the epidemic can spread (i.e., the system enters an absorbing state).

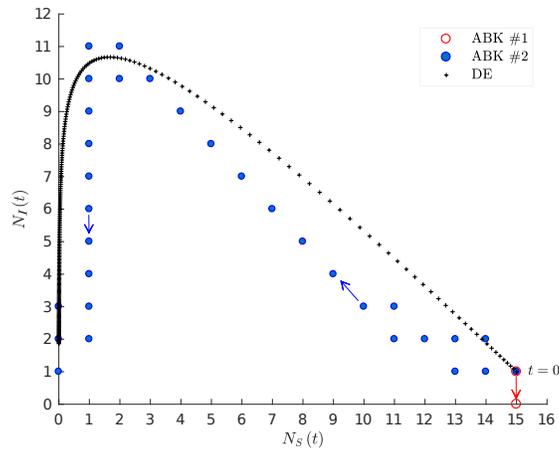


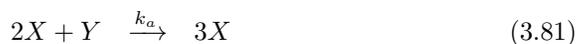
Figure 3.59 – Kermack-McKendrick S - I state space. We present here a parametric representation of the trajectories of species S and I . Here, we ignore species R to keep this plot 2-dimensional. The initial condition $N_{S,i} = 15$, $N_{I,i} = 1$ is designated by the label $t = 0$ on the lower right-hand corner of the state space. We show the deterministic trajectory/path through S - I space (labeled “DE” in the legend), evaluated by numerical integration of the differential equations describing the model until time $t = 50$ days, as well as two sample trajectories obtained through the ABK simulation (labeled as ABK #1 and ABK #2). The density of points along the deterministic path is inversely proportional to the speed of the trajectory through state space. The arrows indicate the flow of the agent-based simulation through state space as time progresses. Notice that for sample ABK trajectory #1 (red open circles), N_I becomes and stays zero (an absorbing state of the system), so no epidemic occurs. On the other hand, ABK trajectory #2 mirrors the deterministic path and the disease spreads through the population.

starting with a single infected individual (“patient zero”), these results agree with the logical premise that containing the spread of the disease at the earliest possible time is the best way to protect a susceptible population from an epidemic. Thus, either through chance events (e.g., patient zero recovers before infecting others) or through intervention according to public health policy/strategy (e.g., infected individual is quarantined), the epidemic does not occur.

3.6.3 The Brusselator

We now apply the ABK method to a much-studied mathematical model called the *Brusselator*, originally devised to understand the origin of far-from-equilibrium instabilities in chemical systems (Prigogine and Lefever, 1968) and the dynamics of spatiotemporal chemical oscillations.³¹

The model we present here has only two species, X and Y , and is defined by the following processes,



where k_b is the 0th order rate constant for the production (or “birth”) of X , k_c is the 1st order rate constant for the conversion of X to Y , k_a is the 3rd order rate constant for the autocatalytic production of X from Y , and k_d is the 1st order rate constant for the degradation of X .³² The following figure summarizes this reaction scheme in a succinct way.

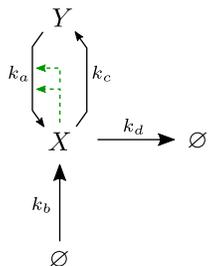


Figure 3.60 – The Brusselator reaction scheme. A summary of the set of reactions defining the Brusselator. The two dashed green arrows indicate that two molecules of X catalyze the synthesis of X from Y .

The two green arrows indicate that two molecules of X catalyze the conversion of a molecule of Y to X . This trimolecular (or termolecular) reaction is what makes the Brusselator not likely to be physically realistic: very few 3rd order reactions are known to chemistry since they require the simultaneous

³¹Such oscillations were first observed by Belousov in the 1950s and later elaborated on by Zhabotinsky and others. See Winfree, 1984, for a brief historical account of that early work.

³²The original version of this reaction set included additional species whose population sizes (or concentrations, in a chemical sense) were assumed to remain constant (as Prigogine and Lefever also assumed). For instance, one of those species makes the conversion of X to Y (with rate constant k_c) a 2nd order process. We omit these additional species in our model description for simplicity and without loss of generality (their concentrations can be thought of as being absorbed into the kinetic rate constants).

effective collision between three atoms/molecules.³³ The nonlinearity of this autocatalytic reaction in the differential equations describing the Brusselator is necessary for its oscillatory dynamics (Tyson, 1973). That is, if reaction (3.81) used only one molecule of X as a reactant and the reaction was bimolecular, then a limit cycle is not theoretically possible.

Given this set of reactions, the Brusselator can be described by the following set of differential equations (DEs) in terms of the kinetic constants and the populations sizes N_X and N_Y :

$$\frac{dN_X}{dt} = k_b + k_a N_X^2 N_Y - (k_c + k_d) N_X \quad (3.83)$$

$$\frac{dN_Y}{dt} = k_c N_X - k_a N_X^2 N_Y. \quad (3.84)$$

We will refer to this as the *canonical* set of DEs. For small populations sizes, it is reasonable to consider the agent-based interpretation of the model, so that the (slightly) modified set of *agent-based* DEs is

$$\frac{dN_X}{dt} = k_b + k_a N_X (N_X - 1) N_Y - (k_c + k_d) N_X \quad (3.85)$$

$$\frac{dN_Y}{dt} = k_c N_X - k_a N_X (N_X - 1) N_Y. \quad (3.86)$$

Essentially, only the term describing the trimolecular step is different from the canonical system of DEs, since when one X agent is recruited the pool of remaining and available X agents is $N_X - 1$. Given our previous investigations on 2nd and higher order processes (see Chapter 2), we expect the agent-based DEs to be a better fit to the results of our simulations for small population sizes.

Table 3.1 summarizes a comparison of the main results regarding the fixed point and Hopf bifurcation given the two DE interpretations we presented above. Linear stability analysis of this 2-dimensional system shows that a (supercritical) Hopf bifurcation occurs when the relationship between the kinetic constant parameters indicated in the 2nd row of Table 3.1 is satisfied. A limit cycle occurs when k_c is greater than the expression on the right side of the bifurcation condition. This allowed us to set the parameters so that we could model this system under predicted stable oscillatory activity. We show the Brusselator phase plane in Figure 3.61 for one such set of parameters: $k_a = 1.0 \times 10^{-4} \text{sec}^{-1}$, $k_b = 0.30 \text{sec}^{-1}$, $k_c = 0.075 \text{sec}^{-1}$, and $k_d = 0.020 \text{sec}^{-1}$. The figure shows the difference in the limit cycle predicted from the canonical and agent-based DEs (gray vs. black trajectories), with the latter producing higher amplitude oscillations.

We modeled the Brusselator for three sets of kinetic parameters, as shown in Figure 3.62 (panel b corresponds to the same parameters as presented in

Follow this [link](#) to see the Matlab code.

³³Catalyzed bimolecular reactions are made out of three components, however in this case a reaction does not occur with 3rd order kinetics because a stable intermediate (an enzyme-substrate complex in enzymatic reactions, for example) occurs first before the rate-determining step of catalytic conversion occurs.

	Canonical DEs	Agent-based DEs
Fixed Point:	$N_X^* = k_b/k_d$ $N_Y^* = \frac{k_c k_d}{k_a k_b}$	$N_X^* = k_b/k_d$ $N_Y^* = \frac{k_c k_d}{k_a(k_b - k_d)}$
Hopf bifurcation condition:	$k_c = k_d + \frac{k_a k_b^2}{k_d^2}$	$k_c = \left[k_d + \frac{k_a k_b (k_b - k_d)}{k_d^2} \right] \left(1 - \frac{k_d}{k_b} \right)$
Oscillation freq.:	$f = \frac{1}{2\pi} \sqrt{\frac{k_a k_b^2}{k_d}}$	$f = \frac{1}{2\pi} \sqrt{\frac{k_a k_b (k_b - k_d)}{k_d}}$

Table 3.1 – Canonical vs. agent-based DEs in the Brusselator. Results from the canonical and agent-based DE interpretations of the Brusselator model. The first row shows the state space coordinates (N_X^*, N_Y^*) for the theoretically predicted fixed point. Evaluating the Jacobian at the fixed point and setting the trace equal to zero yields the condition for a (supercritical) Hopf bifurcation. The fixed point turns from stable to unstable as k_c surpasses the value given by the indicated expression. The bifurcation condition allows the calculation of the oscillation frequency as the limit cycle is born.

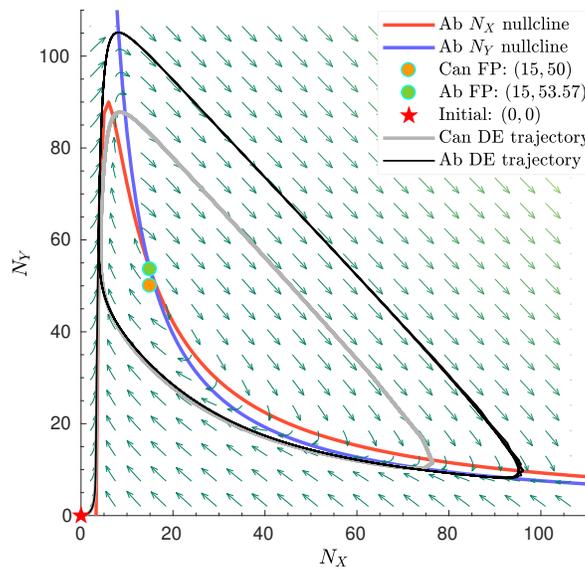


Figure 3.61 – Brusselator phase plane. Phase plane for the Brusselator (canonical DEs 3.83 and 3.84; agent-based DEs 3.85 and 3.86) with kinetic parameters $k_a = 1.0 \times 10^{-4} \text{ sec}^{-1}$, $k_b = 0.30 \text{ sec}^{-1}$, $k_c = 0.075 \text{ sec}^{-1}$, $k_d = 0.020 \text{ sec}^{-1}$, and initial population sizes $N_{X,i} = 0$, $N_{Y,i} = 0$ (a sample time trajectory for this parameter set is shown in Figure 3.62b). The direction field is shown as green arrows (lighter shades of green correspond to velocity vectors with greater amplitude). The nullclines, fixed point, and the predicted deterministic trajectory (gray curve for the canonical limit cycle, black curve for the agent-based one) are also shown. Abbreviations in legend: “Ab”, agent-based; “Can”, canonical; “DE”, differential equation; “FP”, fixed point.

the phase plane analysis in Figure 3.61). In all three cases we observed stable oscillations (as expected, since we made sure the condition for a Hopf bifurcation was satisfied), but with increasingly irregular amplitude and frequency when the considered populations grew smaller (see this progression from panel c to b to a). We found that the population amplitudes in our simulations more closely matched the numerical solution to the agent-based DEs for smaller populations (panels a and b) and the canonical DEs for larger populations (panel c). These simulations show that the ABK algorithm can effectively predict the system's trajectory even when the involved species are synthesized and degraded at widely separated time scales. For instance, N_Y builds up slowly before collapsing within a very short time period in each cycle.

We have also calculated the average population trajectories for species X and Y , and they follow a pattern of decaying oscillations (in the phase plane, the average trajectory spirals inward to the fixed point). This has been observed before in computational efforts to stochastically model the Brusselator reaction scheme (Lumsden and Rowe, 1984). We do not show plots of these average trajectories here because we have encountered the same phenomenon in other oscillatory motifs presented in this chapter (e.g., see subsection 3.5.3 on the 3-component negative feedback motif and subsection 3.6.1 on the Lotka-Volterra model; see Figure 3.55 in particular).

We also show the result of a simulation with values of kinetic constants such that the deterministic criterion for stable oscillations is not met (while still being close to a Hopf bifurcation in parameter space). Instead, the deterministic prediction (in either the canonical or agent-based formulations of the set of DEs) is that populations oscillate with a decaying amplitude to their steady-state values. We refer to this system as the deterministically-damped Brusselator. However, we observe oscillations with no apparent decay in our ABK simulations. Figure 3.63 shows a sample time trajectory for a total simulated time of 6000sec. We have run these simulations for up to 5×10^4 sec and there is no indication of any decay in the oscillatory activity. We see therefore that the oscillations in this form of the Brusselator are only damped in the deterministic sense, while the stochastic simulations reveal sustained noisy oscillations. We have already described several examples of this behavior in section 3.5.

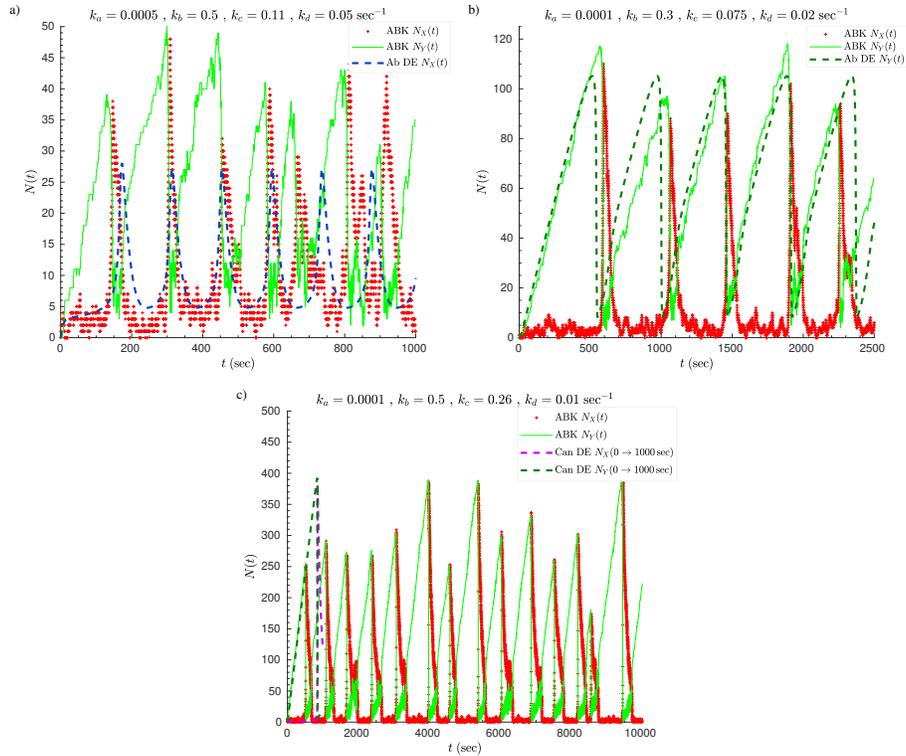


Figure 3.62 – Sample oscillatory time trajectories of the Brusselator. We simulated the Brusselator for three distinct sets of microscopic kinetic parameters, (as indicated in the headers of panels a-c; all kinetic constants have units of sec^{-1}), for which stable relaxation oscillations are predicted. We used the ABK algorithm with a fixed time step increment of $\Delta t = 0.01 \text{ sec}$ in panel a, and $\Delta t = 0.02 \text{ sec}$ in panels b, c. Note that the aspect ratio between the t and $N(t)$ axes is the same in all three panels, although the scale is different. We also show the agent-based deterministic time trajectory for $N_X(t)$ in plot a, the agent-based deterministic time trajectory for $N_Y(t)$ in panel b, and the canonical DE solution for $N_X(t)$ and $N_Y(t)$ only during the first 1000 sec in c. Some or part of the DE curves were omitted for ease of visualizing the stochastic trajectories. Abbreviations in legend: “Ab DE”, agent-based differential equation; “Can DE”, canonical differential equation.

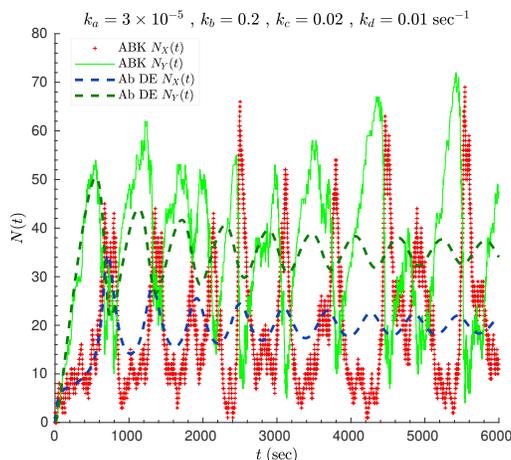
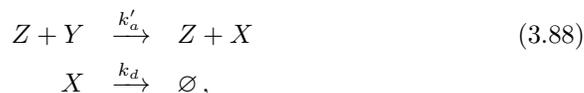


Figure 3.63 – Sample time trajectory of the deterministically-damped Brusselator. We simulated the Brusselator for a set of kinetic constants (all constants are listed in the header and have units of sec^{-1}) leading to damped oscillations using either the canonical or agent-based DEs. We used the ABK algorithm for a total simulated time of 50000 sec (showing here only up to 6000 sec) with a fixed time step increment of $\Delta t = 0.02$ sec. The stochastic simulation yielded persistent relaxation oscillations for the total duration of the simulated time. Abbreviation in legend: “Ab DE”, agent-based differential equation.

Modified Brusselator

We conclude our brief presentation of the Brusselator by showing a nominally modified version of it. We introduce a new intermediate species, Z , to avoid the chemically unrealistic trimolecular step. The full set of processes defining this modified Brusselator are,



where the numbered processes have replaced the trimolecular step (process 3.81). Notice that the synthesis and degradation processes of X are the same as before, as is the 1st order conversion of X to Y . Two molecules of X combine (or dimerize) to form Z in a reversible process (with rate constants k_f and k_r for the forward and reverse reactions respectively). Z in turn catalyzes the conversion of Y to X (with rate constant k'_a), however this process is 2nd order now instead of 3rd. Thus, this is an example of indirect autocatalysis where X catalyzes its own formation through the intermediate species Z . The following figure summarizes this modified form of the Brusselator reaction scheme.

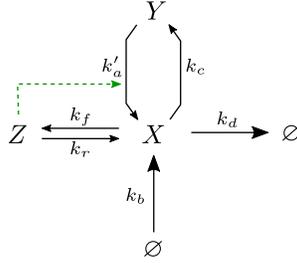


Figure 3.64 – The modified Brusselator reaction scheme. A summary of the modified set of reactions defining the Brusselator. The dashed green arrow indicates that a molecule of species Z catalyzes the synthesis of X from Y . Since Z can be thought of as a dimer of X , this scheme features an indirect autocatalytic loop.

This set of processes can be summarized in an updated system of canonical DEs in terms of the discrete population sizes N_X , N_Y , and N_Z .

$$\frac{dN_X}{dt} = k_b - 2k_f N_X^2 + 2k_r N_Z + k'_a N_Z N_Y - (k_c + k_d) N_X \quad (3.89)$$

$$\frac{dN_Y}{dt} = k_c N_X - k'_a N_Z N_Y \quad (3.90)$$

$$\frac{dN_Z}{dt} = k_f N_X^2 - k_r N_Z \quad (3.91)$$

The corresponding system of agent-based DEs is obtained by replacing N_X^2 with $N_X(N_X - 1)$ in the above equations.

It is clear that the steady-state value for the population of species Z is

$$N_Z^* = \frac{k_f}{k_r} (N_X^*)^2. \quad (3.92)$$

Substitution of this result into $\frac{dN_X}{dt}$ leads to a cancellation of the terms due to the reversible reaction 3.87, leaving us with expressions for $\frac{dN_X}{dt}$ and $\frac{dN_Y}{dt}$ that are identical to the original Brusselator's, except $k'_a \frac{k_f}{k_r}$ takes the place of the 3rd order constant k_a (we note this simplification to underscore the relationship between this modified Brusselator and the original). However, we will make no assumptions about the reversible process being fast compared to the other processes in the scheme when performing simulations. In fact, as we will see in the example below, the population of Z is *not* in a steady state for our choice of parameters (see blue time trajectory in Figure 3.65).

For notational simplicity, we let $K = \frac{k_f}{k_r}$, where K is the equilibrium constant for the reversible reaction. The coordinates of the canonical fixed point can therefore be easily shown to be

$$N_X^* = \frac{k_b}{k_d}, \quad N_Y^* = \frac{k_c k_d}{(k'_a K) k_b}, \quad N_Z^* = K \left(\frac{k_b}{k_d} \right)^2. \quad (3.93)$$

The corresponding quantities for the agent-based DEs are

$$N_X^* = \frac{k_b}{k_d} \quad , \quad N_Y^* = \frac{k_c k_d}{(k'_a K)(k_b - k_d)} \quad , \quad N_Z^* = K \frac{k_b (k_b - k_d)}{k_d^2} . \quad (3.94)$$

We simulated the modified Brusselator with parameters $k'_a = 0.01 \text{ sec}^{-1}$, $k_b = 0.30 \text{ sec}^{-1}$, $k_c = 0.075 \text{ sec}^{-1}$, $k_d = 0.02 \text{ sec}^{-1}$, $k_f = 0.01 \text{ sec}^{-1}$ and $k_r = 1 \text{ sec}^{-1}$, such that $K = 0.01$ and the product $k'_a K = 0.0001 \text{ sec}^{-1}$. This set of parameters is the same as the one presented for the original form of the Brusselator in Figure 3.62b (with $k_a = k'_a K$). Our simulation results again showed excellent agreement with the agent-based differential equations used to describe this modified form of the Brusselator. We end this subsection by showing in Figure 3.65 a sample time series obtained from the ABK simulation of this reaction scheme. Our general observations and comments on the original Brusselator apply here as well.

Follow this [link](#) to see the Matlab code.

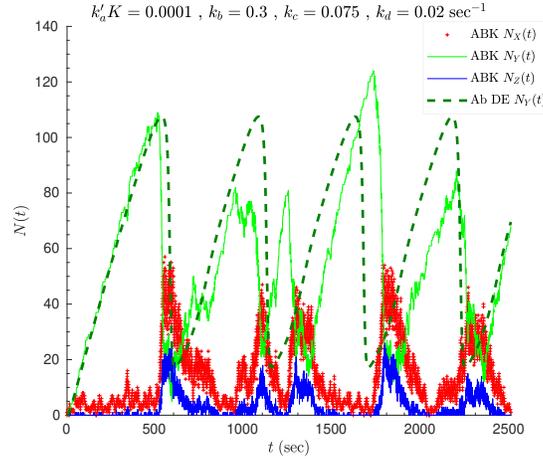


Figure 3.65 – Sample oscillatory time trajectory of the modified Brusselator. We simulated the modified Brusselator for the indicated set of kinetic constants (all constants have units of sec^{-1}) for which stable relaxation oscillations are predicted. We used $k'_a = 0.01 \text{ sec}^{-1}$, $k_b = 0.30 \text{ sec}^{-1}$, $k_c = 0.075 \text{ sec}^{-1}$, $k_d = 0.02 \text{ sec}^{-1}$, $k_f = 0.01 \text{ sec}^{-1}$ and $k_r = 1 \text{ sec}^{-1}$, such that $K = 0.01$ and the product $k'_a K = 0.0001 \text{ sec}^{-1}$. We implemented the ABK algorithm with a fixed time step increment of $\Delta t = 0.01 \text{ sec}$ and for a total simulated time of 2500 sec. We also show the agent-based deterministic time trajectory for $N_Y(t)$. Abbreviation in legend: “Ab DE”, agent-based differential equation.

Chapter 4

Agent-Based Kinetics for Heterogeneous Populations of Well-Mixed Systems

In the previous two chapters we established that the ABK algorithm can be used to describe systems made out of agents with identical kinetic properties (i.e., homogeneous populations). Some of the motifs we have presented can give rise to population heterogeneity in higher-level structures (e.g., stochastic switching or out-of-phase intracellular oscillators consisting of molecular components result in cellular phenotypic variation). In this chapter, we lay the foundations for the ABK algorithm to be used for studying the effect of heterogeneity on population dynamics. Specifically, we describe how the algorithm can be adapted to modeling heterogeneous populations whose agents can undergo 1st or 2nd order transitions and birth-death processes. We then generalize for higher order processes and develop a graph-theoretic view of intra- and interspecies agent interactions. We also revisit several of the motifs presented in Chapter 3 to examine case studies highlighting how the nominal presence of population heterogeneity can affect the emergent properties of the system, such as noise-induced switch state reversals and oscillatory behavior. Finally, we discuss how sensitivity analysis can be performed to assess the effect of various metrics of heterogeneity on a population's average time trajectory.

4.1 Introduction: Types of Heterogeneity

We define heterogeneity with respect to the *intrinsic kinetic rate constant* that an agent of a species population has. That is, we associate each agent with a kinetic parameter (with units of reciprocal time) that reflects its intrinsic propensity with respect to a transition that a member of the species can undergo. This agent-specific constant is the same as the microscopic constant, k , and can be obtained from measurements on the kinetics of homogeneous populations.

We define a group of agents belonging to a given species that have an intrinsic rate constant that is different from the one of other members in the population as a *subspecies* (with a corresponding *subpopulation* size). An example of a commonly encountered case in biochemistry is a eukaryotic protein's isoforms, which typically have different rate constants with respect to their activity (e.g., catalysis, binding to other molecules, etc.). In the limiting case where all agents have different rate constants, then the number of subspecies is the same as the population size.

The above definition of subspecies is most appropriate for species undergoing a 1st order process. For 2nd and higher order processes, we find it more useful to consider the *interactions* between agents. In those cases, we refer to kinetically distinct types of intra- or interspecies interactions as *subinteractions*. We will further elaborate on this choice of nomenclature in the following sections.

There are two general types of heterogeneity we will primarily concern ourselves with:

1. *Compositional heterogeneity*: A species consists of several subspecies and agents in different subpopulations have different intrinsic rate constant values. Thus, the probability of transition for a given agent depends on which subpopulation it belongs.
2. *Temporal heterogeneity*: The rate constants for the agents of a given population or subpopulation are time-dependent.

We will first present how to measure the degree of compositional heterogeneity in a population, and then discuss how we can adapt the ABK algorithm to simulate the time evolution of species undergoing basic processes where one or both of the above types of heterogeneity are present.

4.2 Measures of Population Heterogeneity

We begin by defining a basic measure of population heterogeneity: the *population richness*, C , signifying the number of distinct types of agents within a population. We note that this measure can be generally applied to different hierarchical levels of a population: 1) it can describe the number of species in a system (e.g., a community, in an ecological context), or 2) the number of kinetically distinct subspecies in a species population. We are more interested in the latter interpretation, but for the sake of generality we will distinguish between

these two contexts by including a subscript of the species name when describing its number of subspecies. As an example, the notation $C = 3$ denotes a system or community with 3 species, while $C_A = 3$ means that species A consists of 3 subspecies. It is evident that the richness does not take into account the abundance of the different species/subspecies and is therefore a rather crude measure of a population’s heterogeneity.

We present a more useful quantitative measure of the degree of compositional heterogeneity, the *index of heterogeneity*, defined as

$$\psi = 1 - \frac{1}{N(N-1)} \sum_{h=1}^C n_h(n_h - 1), \quad (4.1)$$

where N is the total number of entities (e.g., agents) within the population, and n_h is the number of entities within the h^{th} type or group.¹ The summation term runs over all types, quantified by C or C_A depending on the context. The above relation can be interpreted as the probability that two randomly chosen (without replacement)² entities belong to different types. As such, ψ has the desirable property $0 \leq \psi \leq 1$. Note that ψ is a simple transformation of a measure of diversity in a population first proposed by Simpson (1949), such that it increases with increasing heterogeneity.³ Therefore, the interval’s endpoints correspond to the boundary cases of a homogeneous ($\psi = 0$) and fully heterogeneous population ($\psi = 1$).

The index of heterogeneity is a measure of the relative number of distinct types of subpopulations, but does not capture information about the numerical variability of the kinetic rate constants. To have a full appreciation of population heterogeneity, one needs to supplement ψ with measures of the species- or interaction-specific rate constants’ statistical distribution properties, such as their means and variances. We establish the definition of these properties below and highlight our notational convention for them.

¹We use the terms “entities”, and “groups” or “types” for the sake of generality. We will use more specific terms when discussing 1st and 2nd order processes in the following sections.

²If we had chosen to represent this probability in terms of sampling the first entity and then having it replaced in its subpopulation before selecting a second entity, the index of heterogeneity would then be

$$\psi = 1 - \frac{1}{N^2} \sum_{h=1}^R n_h^2. \quad (4.2)$$

The difference between the two forms of ψ is small for large population sizes. However, for low copy number populations, the issue of choosing with or without replacement can give significantly different results. We therefore prefer the formulation given in equation 4.1 because it is in keeping with the agent-based interpretation we have used throughout this work and is applicable to small population sizes.

Note that this version of ψ (equation 4.2) is sometimes used in other fields. For instance, a machine learning method used for classification or regression is constructing a *decision tree* (or a *random forest*, which considers an ensemble of decision trees), where ψ is referred to as the Gini impurity index (the notation often found for this index in this context is G) and is used as a cost function to assess how the training examples assigned to a particular node are mixed among classes. Optimally, $\psi = 0$ and all examples fall under one class.

³The index of heterogeneity is also sometimes referred to as the Gini-Simpson index in the ecology literature, and Gibbs-Martin index in sociology and psychology.

We define the *fractional abundance* of the h^{th} subspecies or subinteraction as

$$\chi_h = \frac{N_h}{N}, \quad (4.3)$$

where h is an integer in the interval $[1, C]$ with C being the total number of subspecies or types of subinteractions that pertain to a given process, N_h represents the population size of the h^{th} type of subspecies/subinteractions, and N is the total number agents of a given species or possible subinteractions.

We now illustrate how this notation can be specifically used in the context of 1st and 2nd order processes. Given that a species A with C_A subspecies undergoes a 1st order process, the fractional h^{th} subspecies abundance is

$$\chi_{Ah} = \frac{N_{Ah}}{N_A}, \quad (4.4)$$

where N_{Ah} is the number of the h^{th} subspecies agents and N_A is the total number of A agents. Similarly, for a 2nd order process with two interacting species, A and B , the fractional h^{th} subinteraction abundance is

$$\chi_{ABh} = \frac{N_{ABh}}{N_{AB}}, \quad (4.5)$$

where N_{ABh} represents the number of the h^{th} type of subinteractions, and N_{AB} is the total number of possible A - B interspecies interactions. We establish the convention of the sequence of species names in the subscript to denote the species with respect to which the ABK algorithm is run for such a heterologous 2nd order process. The notation N_{BA} vs. N_{AB} then conveys information about how the process was simulated, in addition to the total number of possible B - A (or A - B) interspecies agent interactions (we will revisit this convention in section 4.5). Also note that this general notation can be easily adjusted and extended for homologous 2nd order (e.g., χ_{AA}) and higher order processes.

These fractional abundances allow for the calculation of the mean of the process-specific rate constant, \bar{k} ,

$$\bar{k} = \sum_{h=1}^C \chi_h k_h. \quad (4.6)$$

For simplicity in the presentation of the above equation, we have written the subscripts in simplified form by omitting the species names participating in the process. The variance can then be easily calculated,

$$Var\{k\} = \overline{k^2} - \bar{k}^2. \quad (4.7)$$

Finally, we note that information-theoretic measures of heterogeneity (e.g., Shannon's diversity index) can also be found in the literature of different fields of study. However, we will not use them here because we find that ψ and the above statistical properties of k adequately quantify the different aspects of population heterogeneity that are relevant to this work.

4.3 0th Order Processes

Since 0th order processes result in the creation (or “birth”) of agents of a given species, the probability of such an event is population-based and evaluated the same way as described in Chapter 2. In this case, we need only concern ourselves with assigning intrinsic rate constant values to the newly created agents, thereby generating heterogeneity within a population. The rate constant values can be assigned according to a known distribution obtained from experimental evidence, or conjecturally based on a user-specified rule or statistical distribution that the modeler wants to test. For instance, one can assign k values to agents that come into existence by randomly sampling a normal distribution with a specified mean \bar{k} and standard deviation σ_k . In this case, if r_n is a random number drawn from the standard normal distribution (i.e., a normal distribution with mean 0 and standard deviation of 1) then the k value for the i^{th} agent can be calculated by

$$k_i = \bar{k} + \sigma_k r_n. \quad (4.8)$$

Note that a new random number needs to be drawn every time a new agent is created. As a result, the population of a given species will consist of agents whose intrinsic rate constants are normally distributed around a specified mean value.

This approach to creating heterogeneous populations through a 0th order birth process assumes knowledge of all other processes that agents of a given species can participate in. Accordingly, a separate k value would need to be generated for each of the relevant processes. For instance, if a species A can transition through either a 1st or 2nd order process (i.e., concurrent processes; see section 2.7) then a newly synthesized A agent needs to have two rate constant values computed as described above.

4.4 1st Order Processes

To simulate 1st order processes where, in the general case, compositional and temporal heterogeneity are present, we construct at the beginning of the simulation a *population heterogeneity matrix* (henceforth abbreviated as PHM) of size $s \times p$, where s represents the number of time steps the algorithm will run through and p is the total number of agents belonging to a particular species. Therefore, each species has its own PHM. Note that if the agent-specific rate constants are time-independent, then the PHM reduces to a row vector with dimensions $1 \times p$.

$$\begin{array}{c}
 \begin{array}{c} s \\ \text{time steps} \end{array} \left\{ \begin{array}{c} \overbrace{\left[\begin{array}{cccc} k_1(t_0) & k_2(t_0) & \dots & k_p(t_0) \\ k_1(t_1) & k_2(t_1) & \dots & k_p(t_1) \\ \vdots & \vdots & \ddots & \vdots \\ k_1(t_s) & k_2(t_s) & \dots & k_p(t_s) \end{array} \right]}^{p \text{ agents}}
 \end{array}
 \right.
 \end{array}$$

Figure 4.1 – Population heterogeneity matrix for a 1st order process. We define the *population heterogeneity matrix* (PHM) to be the $s \times p$ matrix of the kinetic rate constants for a population exhibiting both compositional and temporal heterogeneity. The number of time steps considered in an ABK simulation is s , and p is the total number of agents of a given species.

The number of rows in the PHM, s , can be set at the beginning of the simulation if fixed time step increments are used for a preset total simulated time. Alternatively, the number of rows of the PHM can change if the time step durations are variable and/or if the simulation runs until some condition is met (such as a population size reaching a maximum value or a species going extinct). However, we note that changing the size of the PHM during the course of a simulation can have a negative effect on the algorithm’s performance as the number of species increases (recall that each species has its own PHM) because Matlab searches for a contiguous block of computer memory to store the matrix in. Repeated resizing of the matrix leads to additional computational overhead for managing memory allocation. We recommend that a predetermined number of time steps (e.g., for fixed values of Δt) is used so that memory is preallocated for each species PHM. This ensures the most efficient use of a computer system’s memory resources during the simulation.

The same concern exists for the number of columns in the PHM: agents of a given species are being synthesized or degraded during the course of a simulation, so the choice of how to represent the size of the population is important. One option is for p to represent the number of agents that are “alive” or exist at any given time point in the simulation. The obvious disadvantage of this approach is the frequent changing of the PHM size leading to inefficient use of computer memory. In the general case where many processes may be involved, it is our recommendation to set p to a user-defined maximum value (which can

be updated only if a population grows too much), so that the number of columns in the PHM represents the intrinsic rate constants for existing agents as well as those that may have already transitioned at a previous time step or will be synthesized at a future time. A modeler can specify the values of the rate constants in a number of ways, one of which we presented in our discussion of 0th order processes. For instance, all possible rate constant values can be set prior to the beginning of the simulation based on a discrete or discretely-sampled continuous distribution (such as the normal distribution). Alternatively, the rate constants can be dynamically assigned as agents come in and out of existence based on a specified distribution. Rules can also be devised that change the distribution of rate constant values across a species population in a time-dependent manner.

In summary, the PHM captures both the compositional and temporal heterogeneity of a species population, and can be flexibly adapted to a modeler's preferences or needs.

4.4.1 Predicting the Time Trajectory of a Heterogeneous Population

We now show how to calculate the statistical properties for a 1st order process occurring in a heterogeneous population. We aim to determine the time-dependent probability distribution of the species population size while taking into account an arbitrary level of heterogeneity within it. This will allow us to compute the distribution's moments, thus obtaining information about the uncertainty in the expected trajectory. We note that the CME is only appropriate for describing the time evolution of homogeneous populations. We propose here an alternate method which only uses basic concepts in combinatorics and probability theory.

It is easiest to understand our approach by first assuming a homogeneous population. Consider the process $A \rightarrow X$, with species A having a nonzero initial population size $N_{A,i}$, while $N_{X,i} = 0$. At any given time t , let the probability that an agent of species A has already transitioned to X at some time in the interval $(0, t)$ be $P_{A \rightarrow X}(t)$. Its complement, $P_{A \leftrightarrow X}(t) = 1 - P_{A \rightarrow X}(t)$, represents the probability that an agent remains a member of species A at time t . Since an agent can exist in one of two mutually exclusive and collectively exhaustive states (a binary agent), whether or not an agent of species A has transitioned can be considered a Bernoulli trial. It is then reasonable to examine the answer to this question: what is the population-wide probability that the population of A consists of n agents at time t ? Clearly,

$$\hat{P}(n, t) = \binom{N_{A,i}}{n} [P_{A \rightarrow X}(t)]^{N_{A,i}-n} [P_{A \leftrightarrow X}(t)]^n. \quad (4.9)$$

We have previously shown (see section 2.2) that the per-agent probabilities have the form

$$P_{A \rightarrow X} = 1 - e^{-kt} \quad (4.10)$$

$$P_{A \leftrightarrow X} = e^{-kt}, \quad (4.11)$$

and \hat{P} can be rewritten as

$$\hat{P}(n, t) = \binom{N_{A,i}}{n} (1 - e^{-kt})^{N_{A,i}-n} (e^{-kt})^n. \quad (4.12)$$

We remind the reader that this treatment is practically the same as presented in subsection 2.2.5 and the above equation is the same result as equation 2.41, except we previously considered the number of transition events, μ , which can be easily related to n by $n = N_{A,i} - \mu$. We also refer the reader to that subsection for showing that n (or μ) is indeed binomially distributed.

Population homogeneity is manifested in all agents being characterized by the same numerical value of the rate constant k . Therefore, the per-agent probabilities $P_{A \rightarrow X}$ and $P_{A \leftrightarrow X}$ always have the same value for a specified t . There will be a number of distinct ways, or combinations, that the population can exist with n agents, each with the same probability value, $(1 - e^{-kt})^{N_{A,i}-n} (e^{-kt})^n$. Since the system can be in the same state in a number of ways, the total probability $\hat{P}(n, t)$ will accordingly be the sum of the probabilities for each of the possible combinations. This allows the grouping of a number of additive terms exactly given by the binomial coefficient (which enumerates the possible combinations of n agents of species A). This grouping is not possible, however, when a population is heterogeneous, and a full listing of the agent-specific probabilities is necessary.

We take the symmetry-breaking effect of heterogeneity into account by using a subscript to denote the agent-specific rate constant k (following the convention used in Figure 4.1). $\hat{P}(n, t)$ has the form,

$$\hat{P}(n, t) = \sum_{s=1}^{\binom{N_{A,i}}{n}} \left[\prod_{h(s) \neq j(s)}^{N_{A,i}-n} (1 - e^{-k_h t}) \prod_{j(s) \neq h(s)}^n (e^{-k_j t}) \right], \quad (4.13)$$

where the indices h and j are uniquely determined by each particular combination of agents that the sum runs through (sampled by the index s). The reader will observe the correspondence of the above equation to 4.12: the products in the bracketed part of the expression are the probabilities of agents who have and have not transitioned in the time interval $(0, t)$ respectively. The sum takes into account all possible combinations of n existing agents within the population at time t .

We can calculate $\hat{P}(n, t)$ for all integral values $n \in [0, N_{A,i}]$ and for each value of t we are interested in, thereby obtaining the time-dependent probability distribution of the population size. We can then determine the first and second raw moments,

$$\langle N_A(t) \rangle = \sum_{n=0}^{N_{A,i}} n \hat{P}(n, t) \quad (4.14)$$

$$\langle [N_A(t)]^2 \rangle = \sum_{n=0}^{N_{A,i}} n^2 \hat{P}(n, t), \quad (4.15)$$

and subsequently determine the variance,

$$\sigma^2 \equiv \text{Var}(\langle N_A(t) \rangle) = \langle [N_A(t)]^2 \rangle - \langle N_A(t) \rangle^2. \quad (4.16)$$

We note that predicting the time trajectory of a heterogeneous population in this way is prohibitively computationally intensive for initial population sizes of $N_{A,i} > 100$. This is due to the large task of generating a list of all possible combinations of agents in the population, calculating the respective probabilities, and performing algebraic operations on them. However, the procedure we have outlined is exact and can, in principle, be used to predict the statistical properties of the trajectory of a heterogeneous population of any size.

4.4.2 ABK Implementation

Specifically with respect to a 1st order process, the propensity for an agent of a given species to transition is independent of other agents, therefore the algorithm works the same way as is described in Chapter 2 (subsection 2.2.6). The only difference is that the transition probability for the i^{th} agent is calculated using k_i , obtained from the i^{th} column of the species PHM (and the row which corresponds to the current time step in the simulation if k_i varies with respect to time).

We show a simple example of using the algorithm for the process $A \rightarrow X$, where species A consists of two subspecies, denoted by $A1$ and $A2$ (see Matlab code on page 283). Since we will consider a 1st order process by itself, we set the maximum number of agents to the initial population size of A . This way, the number of columns in the PHM, p , is set and needs no adjustment at any time during the simulation as long as the agent's binary state (0: 'dead', 1: 'alive') is stored in memory.

We first simulate the case where agents of each subspecies transition to X with time-invariant microscopic rate constants $k_{A1} = 0.3 \text{ sec}^{-1}$ and $k_{A2} = 0.7 \text{ sec}^{-1}$, respectively. We consider the situation where the initial population sizes⁴ are $N_{A,i} = 10$ and $N_{X,i} = 0$, and with initial subpopulation sizes $N_{A1,i} = N_{A2,i} = 5$ agents (richness $C_A = 2$; index of heterogeneity $\psi_A = 5/9 \approx 0.5556$, evaluated using equation 4.1 in section 4.2). Figure 4.2a shows the time trajectory for the population of species A , averaged over 2000 repetitions of the simulation (solid blue curve), and compares it to the deterministic trajectory of a homogeneous population with microscopic rate constant $k = \bar{k} = 0.5 \text{ sec}^{-1}$ (dotted green curve). The discrepancy between the two curves is due to the different kinetics of the two subpopulations: $A2$ transitions to X faster than $A1$, so that after a sufficient period of time has elapsed the population of A consists of mostly $A1$ agents. We show that this is indeed the case in Figure 4.2b, where we plot a histogram of transition events for each subspecies within 1 sec intervals. We have superimposed the probability density function (PDF) of the

Follow this [link](#) to see all of the relevant code.

⁴In keeping with the notation we have used throughout this work, when the subscript i follows a comma it denotes the initial population size. This is not to be confused with the i^{th} agent whose rate constant k_i can be found in the i^{th} column of the PHM.

exponential distribution, $f(t, k) = k e^{-kt}$, for each of the two subspecies (with rate constants k_{A1} and k_{A2}) on the histogram. The excellent fits ($R_{A1}^2 = 0.9986$, $R_{A2}^2 = 0.9997$) are in agreement with the expectation that a 1st order process follows Poissonian statistics, therefore the event interoccurrence times are exponentially distributed (also see subsection 2.2.5 and Figure 2.5 for a presentation of these arguments as they apply to homogeneous populations). Note that the plot shown in 4.2b is independent of the initial population size, $N_{A,i}$ (or the initial subpopulation sizes, $N_{A1,i}$ and $N_{A2,i}$, for that matter), as would be expected for a 1st order process where agent transitions occur independently of one another.

The above observations further imply that if we were to monitor $\langle k(t) \rangle$, the population-wide average of the rate constant at a given time t , it would approach the value of the slowest subpopulation (the one with agent-specific rate constant k_{A1} , in this case). We have confirmed that this is indeed the case (not shown). We will refer to this phenomenon of the kinetically slower subspecies making up the entire species population at later time stages as *enrichment*.⁵ We will see more examples of this behavior in this chapter as we explore more case studies of the effect of heterogeneity on population trajectories.

We also show in Figure 4.3 the relative standard deviation and coefficient of variation, η (a measure of the amount of noise inherent to the process), for the trajectories of heterogeneous populations with $\bar{k} = 0.5 \text{ sec}^{-1}$, but for several different pairs of values for k_{A1} and k_{A2} (alternatively stated, for different values of $\text{Range}\{k_{A1}, k_{A2}\}$ or $\text{Var}\{k\}$, the latter as evaluated using equation 4.7). We then compare the results of our simulations with the predicted standard deviation obtained through the method we presented in the preceding subsection (see equation 4.13 and ensuing discussion). It is evident in Figure 4.3a that in all cases the fits to the predicted relative standard deviation are excellent ($R^2 > 0.9980$). It is important to note how the values of the subpopulation rate constants affect the standard deviation and coefficient of variation. The larger the $\text{Range}\{k_{A1}, k_{A2}\}$ is, the smaller $\eta(t)$ is, indicating that the level of noise decreases as the subspecies are increasingly kinetically distinguishable. This observation supports the hypothesis that heterogeneity can have significant effects on the observed time trajectories of populations as well as in the uncertainty in them. It further suggests that, conversely, the level of population heterogeneity

⁵We can state this in the general case for the slowest subspecies, $A1$, as

$$\lim_{t \rightarrow \infty} \frac{N_{A1}(t)}{N_A(t)} = 1, \quad (4.17)$$

which reflects the fact that it dominates the population's composition and, therefore, its kinetics at larger times scales. That is, $N_{A1}(t)$ asymptotically approaches $N_A(t)$, the former considered as the time trajectory of a distinct subpopulation. We have confirmed this asymptotic relationship in our simulated time trajectories.

We can also write a similar expression for the species with the fastest kinetics:

$$\lim_{t \rightarrow \infty} \frac{N_{A2}(t)}{N_A(t)} = 0. \quad (4.18)$$

That is, $N_{A2}(t)$ is negligible compared to $N_A(t)$ at $t \rightarrow \infty$.

can potentially be inferred from experiments or field measurements that allow determination of the statistical properties of a population's time trajectory.

We can therefore see that a 1st order process involving a heterogeneous population is essentially a collection of 1st order processes, one for each of the subspecies. We have also investigated the kinetics of fully heterogeneous populations (where all agents have different k values) of different sizes and have obtained similar results to the ones we presented above (not shown). Importantly, we have demonstrated through the simple case of only two subspecies how heterogeneity affects the statistical properties of simulated time trajectories and that the ABK algorithm gives results that agree with theoretical expectations.

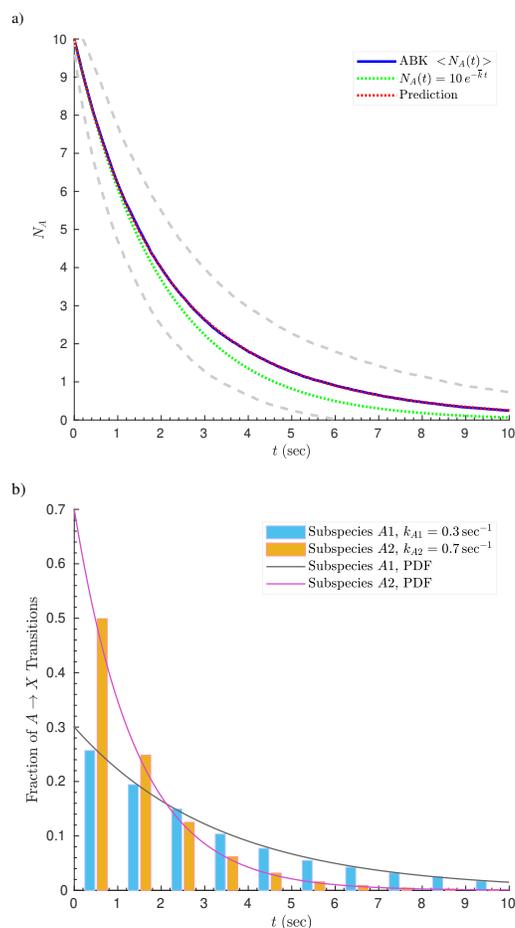


Figure 4.2 – Time trajectory and transition event distribution of a 1st order process with two subspecies. We consider the process $A \rightarrow X$, where the initial population of species A , $N_{A,i} = 10$ agents, consists of two subspecies with equal initial subpopulations sizes. The microscopic rate constants for the subspecies in this simulation are $k_{A1} = 0.3 \text{ sec}^{-1}$ and $k_{A2} = 0.7 \text{ sec}^{-1}$ (average rate constant $\bar{k} = 0.5 \text{ sec}^{-1}$). **a)** The average time trajectory for the population of species A is shown, obtained from an ensemble of 2000 repetitions of the ABK simulation (solid blue curve, with its one-standard-deviation envelope delineated by the gray dashed curves). We also show the deterministic trajectory for a homogeneous population where all agents have the same rate constant $k = \bar{k} = 0.5 \text{ sec}^{-1}$ (dotted green curve). The predicted time trajectory for the specified heterogeneous population (dotted red curve, and labeled as “Prediction” in the legend; see equation 4.13 and surrounding text) is in near perfect agreement with the average obtained through our simulations ($R^2 = 0.9999$). **b)** Histogram of the fractional amount of agent transitions, grouped by subspecies, within successive simulated-time intervals of 1 sec duration (aggregated over all 2000 repetitions of the simulation). Subspecies $A2$ (orange bars) follows faster kinetics so that in the later time stages the population of A consists almost entirely of subspecies $A1$ agents (light blue bars). This explains the discrepancy between the simulation and deterministic prediction in panel a. We have superimposed the probability density function (PDF) of the exponential distribution, $f(t, k) = k e^{-kt}$, for the subspecies with rate constants k_{A1} and k_{A2} (dark gray and purple solid curves respectively). The agreement of the PDFs with the respective histograms is excellent ($R_{A1}^2 = 0.9986$, $R_{A2}^2 = 0.9997$).

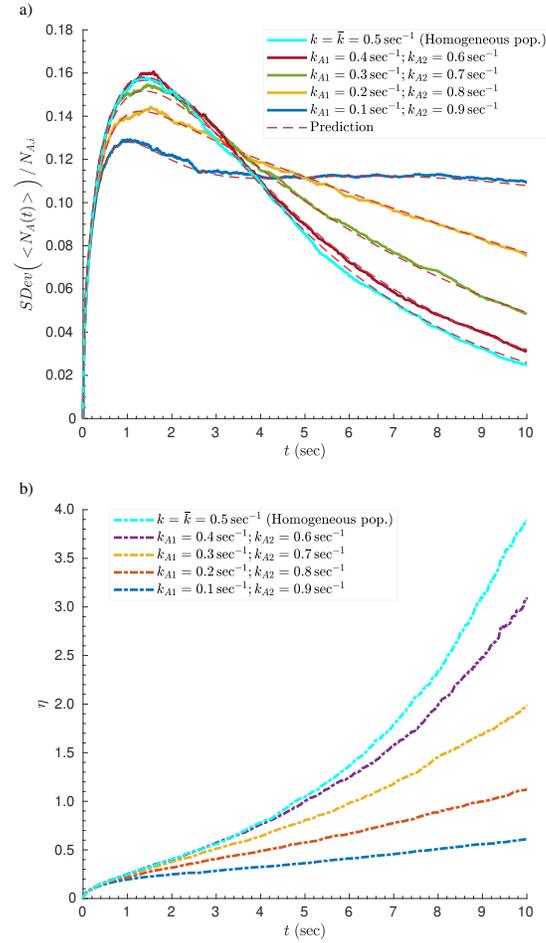


Figure 4.3 – Standard deviation and coefficient of variation of a 1st order process with two subspecies. We consider the process $A \rightarrow X$, where the initial population of species A , $N_{A,i} = 10$ agents, consists of two subspecies with equal initial subpopulations sizes, and average rate constant $\bar{k} = 0.5 \text{ sec}^{-1}$. **a)** Comparison of the relative standard deviation for the entire population's average time trajectory and its predicted value (calculated using equation 4.13 to obtain the second central moment). We show the results for the simulation of several examples of heterogeneous populations with the same structure (i.e., $N_{A,i} = 10$, and equal initial subpopulations sizes) and $\bar{k} = 0.5 \text{ sec}^{-1}$, but with increasingly different values of k_{A1} and k_{A2} . The standard deviation of a homogeneous population's trajectory with $k = \bar{k} = 0.5 \text{ sec}^{-1}$ is also shown (cyan curve). In all cases, the agreement with the predicted curves (obtained using equation 4.13 to calculate the moments) was excellent ($R^2 \geq 0.9980$). **b)** Plot of the coefficient of variation, η , obtained from the ABK simulation of the same cases as in panel a).

4.5 2nd Order Processes

To simulate 2nd order processes, we need to modify the dimensions of a PHM so that both types of heterogeneity (compositional and temporal) can be dealt with in the general case. We begin by considering the heterologous process $A+B \rightarrow X$, with p and q being the initial number of A and B agents respectively. We can then construct a $p \times q \times s$ matrix of the kinetic rate constants, as illustrated in 4.4 below.

$$\begin{array}{c}
 \begin{matrix}
 \left[\begin{array}{cccc}
 k_{1,1}(t_s) & k_{1,2}(t_s) & \dots & k_{1,q}(t_s) \\
 k_{2,1}(t_s) & k_{2,2}(t_s) & \dots & k_{2,q}(t_s) \\
 \vdots & \vdots & \ddots & \vdots \\
 k_{p,1}(t_s) & k_{p,2}(t_s) & \dots & k_{p,q}(t_s)
 \end{array} \right] \\
 \dots \\
 \left[\begin{array}{cccc}
 k_{1,1}(t_1) & k_{1,2}(t_1) & \dots & k_{1,q}(t_1) \\
 k_{2,1}(t_1) & k_{2,2}(t_1) & \dots & k_{2,q}(t_1) \\
 \vdots & \vdots & \ddots & \vdots \\
 k_{p,1}(t_1) & k_{p,2}(t_1) & \dots & k_{p,q}(t_1)
 \end{array} \right] \\
 \dots \\
 \left[\begin{array}{cccc}
 k_{1,1}(t_0) & k_{1,2}(t_0) & \dots & k_{1,q}(t_0) \\
 k_{2,1}(t_0) & k_{2,2}(t_0) & \dots & k_{2,q}(t_0) \\
 \vdots & \vdots & \ddots & \vdots \\
 k_{p,1}(t_0) & k_{p,2}(t_0) & \dots & k_{p,q}(t_0)
 \end{array} \right]
 \end{matrix}
 \end{array}
 \begin{array}{c}
 \underbrace{\hspace{10em}}_{q \text{ agents } B} \\
 \underbrace{\hspace{10em}}_{s \text{ time steps}}
 \end{array}
 \end{array}$$

Figure 4.4 – Population heterogeneity matrix for a 2nd order process. We define the population heterogeneity matrix (PHM) to be the $p \times q \times s$ matrix of the kinetic rate constants for a population exhibiting both compositional and temporal heterogeneity. p and q are the initial number of agents of species A and B respectively, and the number of time steps considered in an ABK simulation is s .

Recall that one can use the ABK algorithm to simulate the process $A+B \rightarrow X$ with respect to either species A or B . That is, one can evaluate the transition probability for each of the agents belonging to the population of one of the two species (see section 2.3 for a treatment of homogeneous populations). Here, we use the convention of assigning the row dimension of the PHM to the species with respect to which the transition probabilities are computed. Therefore, the PHM has dimensions $p \times q \times s$ if the 2nd order process is considered with respect to A , and $q \times p \times s$ if with respect to B .

We note that all kinetic constant values in the PHM apply to unique pairs of agents, and each pair deals with a distinct interspecies agent interaction. For instance, running the algorithm with respect to A and assuming that agents labeled as 1 and 2 of either species have not been consumed yet, then $k_{1,2}$ does *not* have the same meaning as $k_{2,1}$: the former applies to the interaction between agent 1 of species A and agent 2 of species B ; $k_{2,1}$, on the other hand, applies to the interaction between agent 2 of A and agent 1 of B .

Therefore, we can calculate the probability of all possible interactions within a given time step by considering all entries in the PHM. Given a PHM having the same form as the one shown in Figure 4.4, then for the i^{th} A agent interacting

with the j^{th} B agent, we have

$$P_{ber, \mathbf{A}_i+B_j \rightarrow X}(t_n) = 1 - e^{k_{i,j}(t_n) \Delta t} \quad (4.19)$$

$$P_{dif, \mathbf{A}_i+B_j \rightarrow X}(t_n) = k_{i,j}(t_n) \Delta t, \quad (4.20)$$

where boldface A in the subscripts indicates that the simulation is run with respect to species A . The algorithm can then proceed as it does when simulating homogeneous populations (that is, sequentially working on each agent of A , or row of the PHM, sampled by the counter variable i), but considering each possible interaction with the j^{th} agent of B by drawing a random number from the uniform distribution and comparing it to the value obtained from one of the above probability expressions. The algorithm takes into account each of these probability values only when both agents forming an interacting pair have not been consumed in a previous time step. Notice that this method allows for simulating the time trajectory of a *fully* heterogeneous population by evaluating all possible interactions for a transition event.

Moreover, it can be easily shown that this method reduces to the one we have previously seen for homogeneous populations. To see this, we consider the total probability that the i^{th} agent of A will transition in the n^{th} time step of the simulation,

$$P_{tot, A_i}(t_n) = \sum_{j=1}^q b_j k_{i,j}(t_n) \Delta t, \quad (4.21)$$

where we used the P_{dif} form of transition probabilities and b_j is a boolean factor with values 0 or 1 reflecting if agent j of species B has been consumed in a previous time step or not, respectively. Alternatively, we can think of b_j as the binary state value of agent j .

Essentially, we want the summation to extend only over agents of B which are available at time t_n by not having been consumed at a previous time step. We can then evaluate the sum by first recognizing that for a compositionally homogeneous population all entries in the PHM are identical; that is, $k_{i,j}(t_n) = k(t_n)$, and we can take the $k(t_n)$ term outside of the sum along with Δt . Moreover, summation over all terms gives $\sum_{j=1}^q b_j = N_B(t_n)$, since only agents of B present at time t_n are counted. Thus, we obtain,

$$P_{tot, A_i}(t_n) = k(t_n) N_B(t_n) \Delta t. \quad (4.22)$$

The expression on the right side of equation 4.22 is the same as 2.86 (derived for the treatment of homogeneous populations). The above equality also extends to the case of using the P_{ber} form of transition probabilities in the limit of $\Delta t \rightarrow 0$ (also see discussion in section 2.7 on this argument).

We present an example of this approach by considering initial population sizes $N_{A,i} = 10$ and $N_{B,i} = 7$, $N_{X,i} = 0$, where, in the simplest case, there are only two numerically distinct and time-invariant rate constant values.⁶ We

⁶We present this simple example to illustrate how the ABK algorithm works. We remind the reader that, in general, all $k_{i,j}$ entries can be numerically distinct.

denote these values by k_{BA1} and k_{BA2} , and we interpret them as defining two kinetically distinct types of *subinteractions* between agents of A and B . Notationally, we assign meaning to the order of the species names in the subscript: the first species is indicative of the ABK algorithm running with respect to it (as explained in section 2.3; this notation was introduced in section 4.2 in the context of subinteraction abundances). In this case, our notation shows that we will simulate the time evolution of this process with respect to species B . We therefore construct the PHM with dimensions 7×10 ($q \times p$) so that exactly half of the possible pairwise interactions are characterized by k_{BA1} and the rest by k_{BA2} . As a simple example, we provide below a choice of PHM that fulfills this criterion with values of $k_{BA1} = 0.004 \text{ sec}^{-1}$ and $k_{BA2} = 0.006 \text{ sec}^{-1}$ (we show a simulated time trajectory using these subinteraction kinetic parameters later in this section)

$$\begin{bmatrix} 4 & 4 & 4 & 4 & 4 & 6 & 6 & 6 & 6 & 6 \\ 4 & 4 & 4 & 4 & 4 & 6 & 6 & 6 & 6 & 6 \\ 4 & 4 & 4 & 4 & 4 & 6 & 6 & 6 & 6 & 6 \\ 4 & 4 & 4 & 4 & 4 & 6 & 6 & 6 & 6 & 6 \\ 4 & 4 & 4 & 4 & 4 & 6 & 6 & 6 & 6 & 6 \\ 4 & 4 & 4 & 4 & 4 & 6 & 6 & 6 & 6 & 6 \\ 4 & 4 & 4 & 4 & 4 & 6 & 6 & 6 & 6 & 6 \end{bmatrix} \times 10^{-3} \text{ sec}^{-1} \quad (4.23)$$

Notice that for a specific agent of species B (represented by a row in the PHM) the 2nd order rate constant has one value when interacting with agents A_{1-5} and a different value with A_{6-10} . This means that heterogeneity is dependent on the identity of an agent belonging to species A . Therefore, we can say that A consists of two subspecies, A_1 and A_2 , while the population of species B is homogeneous.⁷ Although this assessment of the origin or nature of heterogeneity is in general useful to consider, we find that examining the populations at the level of subinteractions is appropriate for the general case of no symmetry in the PHM, and is consistent with the way the ABK algorithm works (i.e., by evaluating the probability of an interaction between two agents).

The computation of ψ in this case requires consideration of the number of subinteractions (not agents; see section 4.2). In this sense, ψ measures the degree of heterogeneity in subinteractions. The richness, C_{BA} , is the number of kinetically distinct subinteractions between species B and A . The index of heterogeneity, ψ_{BA} , can then be calculated by considering the number of each type of subinteraction (n_{BAh} in equation 4.1) and the total number of possible subinteractions ($N_{BA} = pq$). For the case we described above, evaluation of these quantities yields $\psi_{BA} \approx 0.507$. The subinteraction abundances can also be easily determined, $\chi_{BA1} = \chi_{BA2} = 1/2$.

Using combinatorics to obtain a theoretical prediction of the average trajectories (and their standard deviations) of heterogeneous populations undergoing a 2nd order process is a tedious and extremely computationally intensive pro-

⁷Examining this with respect to A yields the same conclusion: an agent of A has the same rate constant regardless of the B agent it interacts with (see any column in the above PHM).

cess even for small population sizes. The calculation goes along the same lines as outlined for 1st order processes (subsection 4.4.1), however one needs to be especially careful to correctly account for all possible agent combinations that result in a successful transition event or failure. Given how computationally expensive and cumbersome this calculation is, we abandon its use as a theoretical predictor of statistical measures of heterogeneous population trajectories. We continue by presenting the results of our ABK simulations with confidence that they represent the true effects of heterogeneity on the trajectories of well-mixed populations. Our confidence stems from the following observations: a) the algorithm, as presented above, correctly reproduces the species trajectories in the limiting case of homogeneous populations (which we have thoroughly checked against theoretical predictions; see section 2.3); b) some of the trends that are present in 1st order processes as the degree of heterogeneity increases (especially the ones regarding the uncertainty in the population trajectories) are also present here.

4.5.1 ABK Implementation of a Heterologous 2nd Order Process

The algorithm parses the PHM left-to-right and row-by-row to exhaustively evaluate whether a transition event will occur for each possible B - A interaction within a given time step. Because we (arbitrarily) ordered the rate constant values in the PHM in a symmetrical way for the purposes of this example, care must be taken to avoid any possible biases. Here, for a given agent of B (or row of the PHM shown in expression 4.23), reading the matrix from left to right will consider all of the slower subinteractions before the faster ones (since $k_{BA1} < k_{BA2}$). We assume here a well-mixed system so that there should be no predetermined order in the pairwise interactions the algorithm considers. We avoid this unnecessary bias by *randomizing the order in which the column entries of the PHM are sampled when the algorithm parses a row*.

We simulated the time evolution of this process for several sets of subinteraction rate constant values (k_{BA1} and k_{BA2}) with the PHM having the same symmetry as in expression 4.23. All cases that we examined have the same initial average rate constant value, $\bar{k} = 0.005 \text{ sec}^{-1}$, and measures of heterogeneity $\chi_{BA1} = \chi_{BA2} = 0.5$, and $\psi_{BA} \approx 0.507$. We used the ABK methodology⁸ for heterogeneous populations discussed in the preceding paragraphs with fixed time step increments of duration $\Delta t = 0.01 \text{ sec}$, for a total simulated time of 200 sec, and obtained the average time trajectories of the populations after 1000 repetitions of the simulation (Figure 4.5a,b). The greater the difference between k_{BA1} and k_{BA2} , the more the average trajectory differs from the deterministic one with rate constant \bar{k} . This is evident in Figure 4.5b where the average simulated trajectory shows slower kinetics after the first 20 sec. It is reasonable to propose that the faster subinteraction dominates the kinetics initially, but in doing so it depletes the populations of *both* species so

Follow this [link](#) to see all of the relevant code.

⁸The Matlab code used for this simulation is included in Appendix D (on page 285).

that the slower subinteraction becomes underrepresented in the list of possible interactions. This is not surprising since a 2nd order process depends on interspecies agent interactions (unlike 1st order processes). We show that this is indeed the case in Figure 4.5c,d, where we plot a histogram of the fraction of species B agents undergoing a transition within successive time intervals of 10 sec duration. We have superimposed the predicted distribution of transition events for homogeneous populations with rate constant k_{BA1} or k_{BA2} , respectively (numerically calculated from the solution to the ODE or CME for this process; see Figure 2.12 and accompanying discussion for calculating these values). The plots show that the faster subinteraction consistently transitions faster than in the corresponding homogeneous population, while the trend is reversed for the slower subinteraction. This can be seen more clearly in the insets of 4.5c,d, where we have plotted the deviation of the fraction of transition events in the heterogeneous case from its homogeneous population counterpart (deviation = simulated heterogeneous – theoretical homogeneous). k_{BA1} always shows negative deviations, while k_{BA2} shows positive deviations. The initial magnitude of the deviations (i.e., within the first 100 sec) becomes larger as the difference between the subinteraction rate constants increase. In the later time stage ($t \gtrsim 100$ sec) when the populations are already mostly depleted, the deviations approach zero.

We remind the reader that for 1st order processes we observed that the temporal distribution of the subspecies transition events followed Poissonian statistics, as expected for a process where agents transition independently of each other. We have shown through our example that in the case of a 2nd order process the distribution of subinteraction transition events does *not* follow the statistics obtained for the same process occurring in a homogeneous population. We have not yet resolved on theoretical grounds what the precise distribution is.

We have verified that using this algorithm for a homogeneous population (where all PHM entries have the same value) yields the same time trajectory and transition event distribution as the one we presented in subsection 2.3.2. We have also performed experiments on heterogeneous populations for a variety of initial population sizes, values of the subinteraction rate constants, and methods of computing transition probabilities (P_{ber} or P_{dif}), and obtained similar results to the ones we presented above.

We show the standard deviation and coefficient of variation for some of these cases in Figure 4.6. Although we do not have a theoretical prediction to compare these curves to, the overall trend is the same as what we observed for 1st order processes (see Figure 4.3): the larger the $Range\{k_{BA1}, k_{BA2}\}$ is, the smaller $\eta(t)$ becomes, indicating that the level of noise decreases as the subinteractions are increasingly kinetically distinguishable. Our simulations therefore suggest that heterogeneity appears to have significant effects on the statistical properties of an ensemble of population time trajectories.

We are unaware of any measurements on real populations whose heterogeneity with respect to its agent-specific kinetic parameters is precisely known.

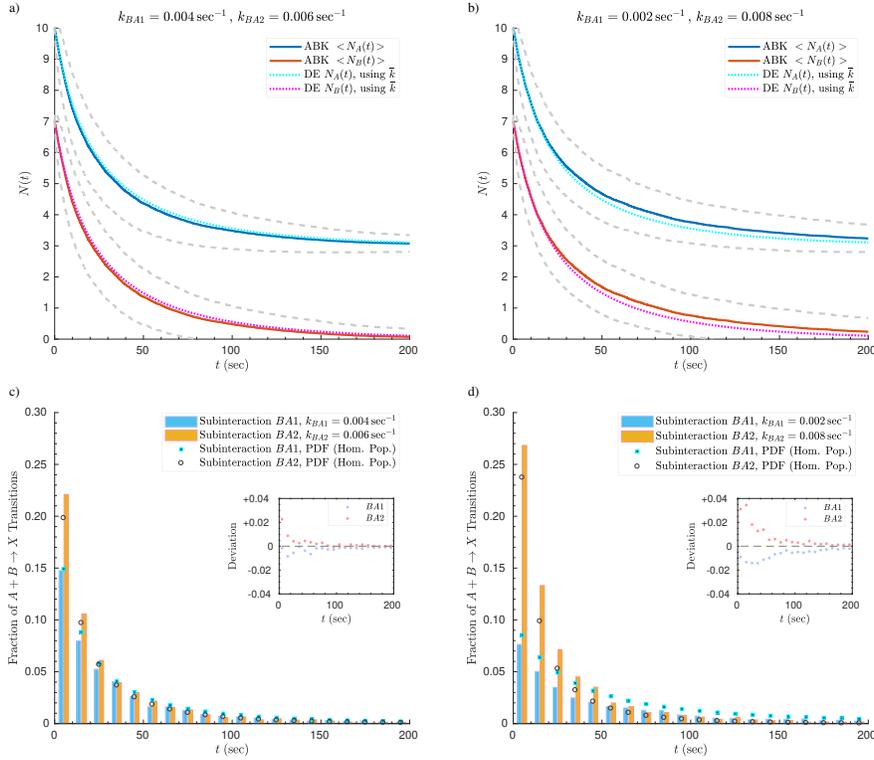


Figure 4.5 – ABK simulation of a 2nd order process with two subinteraction groups. We consider the process $A + B \rightarrow X$, with initial population sizes of $N_{A,i} = 10$ and $N_{B,i} = 7$ agents, and two subinteraction groups, each with an abundance of $\chi_{BA} = 0.5$ (thereby each occupying half of the PHM entries, as exemplified in expression 4.23). The microscopic rate constants for the subinteractions are indicated in each panel. We ran the ABK simulation with respect to species B using fixed time step increments of duration $\Delta t = 0.01 \text{ sec}$ for a total simulated time of 200 sec. **a)** $k_{BA1} = 0.004 \text{ sec}^{-1}$ and $k_{BA2} = 0.006 \text{ sec}^{-1}$. The average time trajectory for the population of species A and B are shown, obtained from 1000 repetitions of an ABK simulation (one-standard-deviation envelopes are delineated by the gray dashed curves). We also show the deterministic (numerical solution to ODEs) prediction based on the average rate constant $\bar{k} = 0.005 \text{ sec}^{-1}$ (dotted curves). **b)** Same as in **a**, except $k_{BA1} = 0.002 \text{ sec}^{-1}$ and $k_{BA2} = 0.008 \text{ sec}^{-1}$. **c,d)** Histogram of the fraction of agent transitions, grouped by subinteraction, within successive simulated-time intervals of 10 sec duration (aggregated over all 1000 repetitions of the simulation). The theoretically predicted fraction of transitions within a pure (homogeneous) population consisting of only one type of subinteraction (numerically calculated from the solution to the ODE for this process; use of the CME yields approximately the same results; see Figure 2.12) is superimposed on the histogram (light blue squares for subinteraction $BA1$ and open black circles for $BA2$). The inset plot shows the difference between the observed (i.e., simulated) fraction of subinteraction transitions and the numerically-computed theoretical fraction for each subinteraction homogeneous population counterpart (deviation = simulated heterogeneous – theoretical homogeneous). Subinteraction $BA2$ (orange bars) follows faster kinetics, thereby quickly depleting the pool of available agents that can be part of a transition. This results in the slower subinteraction, $BA1$ (light blue bars), exhibiting even slower kinetics than its homogeneous population counterpart.

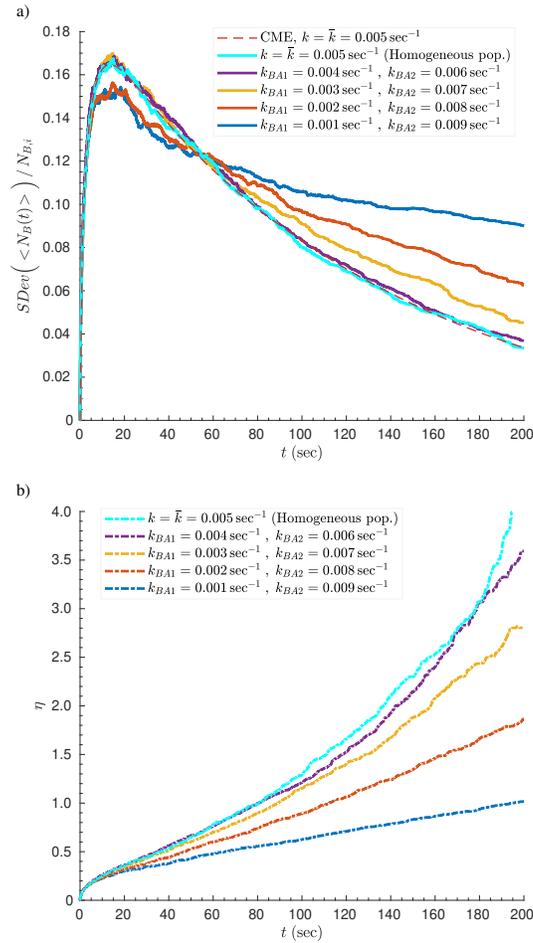


Figure 4.6 – Standard deviation and coefficient of variation of a 2nd order process with two subinteractions. We consider the process $A + B \rightarrow X$ with the same simulation parameters as in Figure 4.5. We show the results for the ABK simulation of several examples with increasingly different values of k_{BA1} and k_{BA2} . **a)** Comparison of the relative standard deviation for species B's average time trajectory, $\langle N_B(t) \rangle$. The standard deviation of a homogeneous population's ABK trajectory with $k = \bar{k} = 0.005 \text{ sec}^{-1}$ and the corresponding CME prediction are also shown (solid cyan and red dashed curves, respectively). **b)** Plot of the coefficient of variation, η , obtained from the ABK simulation of the same cases as in panel a.

Therefore, we cannot at the moment compare our algorithm's predictions with real datasets.

4.5.2 Adjustments to the Algorithm for the 2nd Order Homologous Process

Our treatment of the 2nd order homologous process $A + A \rightarrow X$ is the same as for the heterologous one, except for the following:

- the time-invariant PHM is a square matrix of size $p \times p$, since both the rows and columns represent the number of agents belonging to the same species, A . If temporal heterogeneity also needs to be considered, the PHM has dimensions $p \times p \times s$, where s is the number of time steps.
- the number of distinct pairwise interactions is $p^2 - p = p(p - 1)$, because an agent cannot interact with itself in a second order process. As such, the diagonal entries $k_{i,i}$, with the index i taking the integer values from 1 to p , are *not* included as the algorithm parses through the possible interactions within a given time step.

Other than the omission of the PHM's diagonal entries the algorithm proceeds in the same way as described for the heterologous case by advancing row by row until all possible pairwise interactions are considered. Note that the algorithm considers both instances of a pairwise interaction as possible transitions. For instance, the interaction between agents A_1 with A_2 is considered separately from that between A_2 and A_1 . This is consistent with the version of the algorithm we used for simulating this process in a homogeneous population (subsection 2.3.1).

It is reasonable for the case of the homologous process $A + A \rightarrow X$ to assume that the order in which a particular agent-agent interaction is considered does not affect the value of its transition probability. This is manifested in the PHM with the condition that the rate constant values $k_{i,j} = k_{j,i}$ (provided that $i \neq j$) resulting in a diagonally symmetric PHM matrix in which there are at most $\frac{p(p-1)}{2}$ unique values of k .

We have confirmed that when using a PHM reflecting a homogeneous population, the algorithm produces indistinguishable time trajectories from the ones obtained in subsection 2.3.1. Moreover, heterogeneous populations consisting of two or more subinteraction groups produces results similar to the ones presented above for the heterologous case. We do not show these results here in the interest of brevity.

Follow this [link](#) to see the Matlab code.

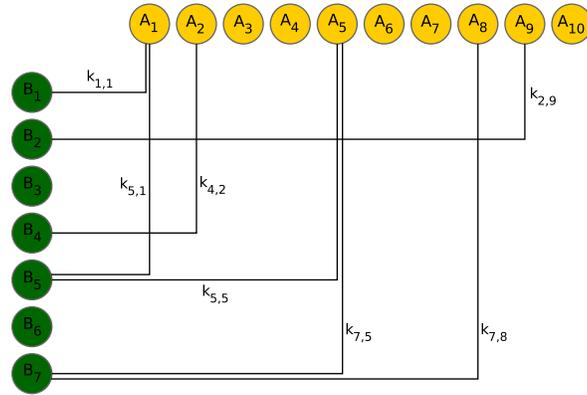
4.6 Generalization to Graphs

We have up to now implicitly assumed that the agents of different species populations can interact in 2nd and higher order processes without restrictions. For instance, in the process $A + B \rightarrow X$, any A agent has the potential to react with any B agent. This makes sense in a well-mixed system, but restrictions may arise due to spatial constraints or other reasons specific to the system at hand. We therefore create a more abstract representation of agent interactions: intra- or interspecies interactions can be represented by a graph whose *nodes* are the agents and *edges* are the possible interactions between them. This way of representing populations allows one to potentially take into account *spatial heterogeneity*, where agents may be distributed across separate “compartments” or sequestered in such a way so that only specific agent interactions are possible. We consider a more general version of this to be *topological heterogeneity*, in recognition of the fact that reasons other than spatial separation may result in different instantiations of node connectivity.

We show an example of such a graph in Figure 4.7, along with its PHM, a sparse matrix with only a handful of nonzero entries. The figure depicts a time-independent PHM, however the dependence on time of the node connectivity can be easily implemented: the PHM can be extended to have a third dimension that allows for time-dependent variability in the graph topology and the values of the rate constants (similarly to Figure 4.4).

The ABK algorithm can be easily adapted to accommodate cases such as these. We have tested the time-independent model described in Figure 4.7 and performed analyses similar to the ones presented in section 4.5. We do not show these results here because they mirror our observations shown in Figure 4.5 (except the overall kinetics are slower given the dearth of possible agent interactions). We will further explore the capabilities of generalizing the ABK algorithm to graphs in future studies.

Follow this [link](#) to see the Matlab code.



$$PHM = \begin{bmatrix} k_{1,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & k_{2,9} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & k_{4,2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ k_{5,1} & 0 & 0 & 0 & k_{5,5} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & k_{7,5} & 0 & 0 & k_{7,8} & 0 & 0 \end{bmatrix} \quad (4.24)$$

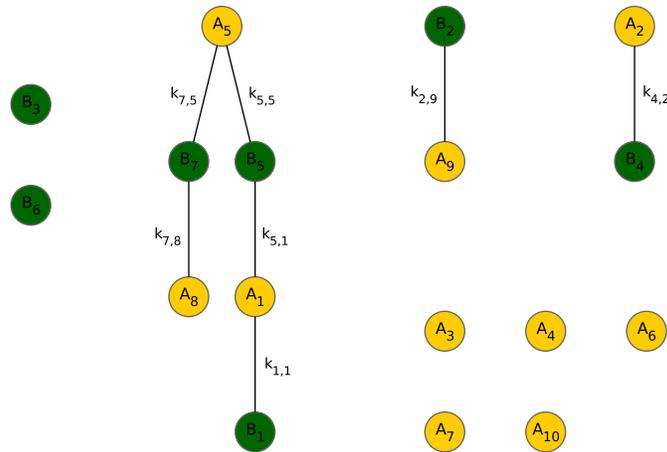


Figure 4.7 – Graph representation of the 2nd order process $A + B \rightarrow X$. An example graph representation of a heterologous 2nd order process with $N_{A,i} = 10$ and $N_{B,i} = 7$, in which only seven B – A interacting pairs are possible. Each agent is a node (color-coded by species: A , yellow; B , green). Only nodes connected by (undirected) edges represent possible interacting interspecies pairs. The top panel shows an arrangement of the nodes that parallels the rectangular structure of the PHM (shown in the middle panel; equation 4.24). The bottom panel shows an alternative, and arguably clearer, view of the graph that could be indicative of the constraints (e.g., spatial structure) that give rise to the observed node connectivity.

4.7 Higher Order Processes

The ABK algorithm can be easily extended to model higher order processes by adjusting the size of the PHM: an n^{th} order process would require the PHM to have $n+1$ dimensions (where the additional dimension is needed for cases involving time-dependent rate constants). The algorithm would proceed by evaluating the transition probability for each n -tuple agent interaction, whose rate constant is represented at a unique position in the PHM. All other considerations with respect to how the algorithm works remain the same.

For instance, all entries in the PHM represent distinct interspecies agent interactions in fully heterologous cases with n species. On the other hand, in homologous cases where there is only one species (e.g., $3A \rightarrow X$ for a 3rd order process), all dimensions of the PHM are of equal size but entries that represent interactions of an agent with itself must be skipped. For a 3rd order process with p agents belonging to a species, the size of the time-invariant PHM is $p \times p \times p$, but p^2 elements along the diagonals correspond to agent self-interactions. Thus, the total number of possible interactions that the algorithm calculates a probability for at each time step is $p^3 - p^2 = p^2(p-1)$. In general, for a n^{th} order homologous process, the number of distinct possible interactions is $p^n - p^{n-1} = p^{n-1}(p-1)$.

In cases where the species interactions are neither fully heterologous nor homologous, the number of PHM entries to be skipped must be carefully considered since it depends on the precise interspecies connectivity as defined by the system's processes. For instance, in the process $2A + B \rightarrow X$ where species A and B consist of p and q agents respectively, the time-invariant PHM has size $p \times p \times q$, and the number of possible interactions is $p^2q - pq = pq(p-1)$.

4.8 Heterogeneity in a Birth-Death Process

We now return to a simple reaction scheme we previously examined in section 2.8. We consider the 0th order production (birth) and 1st order degradation (death) of agents belonging to species A ,



Our previous implementation of simulating the time evolution of this scheme focused on a homogeneous population of A ; that is, the birth process produced identical A agents with respect to the k_d constant characterizing their degradation. Here, we explore the effect of heterogeneity on the time evolution of this system.

We take the birth process, whose rate constant is k_b , to yield a population of A agents with a normal distribution in k_d . We assign the k_d value of the i^{th} newly created agent by drawing a random number r_n from the standard normal distribution, and then using equation 4.8, we have

$$k_{d,i} = \bar{k}_d + \sigma_{k_d} r_n, \quad (4.26)$$

where \bar{k}_d is the mean and σ_{k_d} the standard deviation of the degradation rate constant distribution. The ABK algorithm stores each agent's k_d value in the one-dimensional PHM for the 1st order degradation of A (a matrix of size $1 \times p$, as shown in Figure 4.1 and discussed in section 4.4).

We simulate this birth-death process with parameters $k_b = 0.10 \text{ sec}^{-1}$, $\bar{k}_d = 0.010 \text{ sec}^{-1}$, $\sigma_{k_d} = 0.0025 \text{ sec}^{-1}$. Notice that k_b and \bar{k}_d have the same numerical values that we used in the case of a homogeneous population of A (section 2.8). We can therefore assess the effect of heterogeneity by directly comparing our simulations to those shown in Figure 2.15 and the corresponding deterministic results.

We show the average and one-standard-deviation envelope of $N_A(t)$ obtained using the ABK algorithm in Figure 4.8a. The average time trajectory differs from the deterministic expectation (equation 2.143), which is evaluated using the average degradation rate constant \bar{k}_d ,

$$N_A(t) = \frac{k_b}{\bar{k}_d} \left(1 - e^{-\bar{k}_d t} \right). \quad (4.27)$$

The reason for the observed discrepancy is that the distribution of k_d is time-dependent even though the 0th order birth process produces agents according to a preset normal distribution. This is clearly shown in the inset plot of Figure 4.8a: The average value of the rate constant given the population structure at time t and obtained through an ensemble of ABK trajectories, $\langle k_d(t) \rangle$, decreases and approaches a steady state value of approximately $\langle k_d \rangle^* \approx 0.0093 \text{ sec}^{-1}$ when $t > 500 \text{ sec}$, which is also when the average population size approaches its steady state value of $\langle N_A \rangle^* \approx 10.5$ (deterministic $N_A^* = k_b/\bar{k}_d = 10$). Meanwhile, the standard deviation of the k_d distribution remains the same ($\sigma_{k_d} = 0.0025 \text{ sec}^{-1}$; green dashed lines in the inset plot define the one-standard-deviation envelope).

Follow this [link](#) to see the Matlab code.

As we have seen when examining the time evolution of populations undergoing 1st and 2nd order processes, agents with higher kinetic rate constants transition with higher probabilities, thus leaving the population with a higher abundance of more slowly transitioning agents. We observe the same phenomenon here with the faster degradation of some agents resulting in an *enrichment* of “slower” A agents, and therefore a lower observed average value of k_d than the birth process produces.

If the birth-death process is considered as a simple example of population dynamics from the general perspective of evolution,⁹ then the agents are individual organisms and their capacity for survival is determined by a trait quantified through the agent-specific value of k_d . Therefore, the observed effect is an example of *directional selection*, with the trait’s distribution shifting with the passage of time while retaining its shape (in this case, a normal distribution with a specified standard deviation). The population shifts toward being made up of agents who are more likely to have lower k_d values and therefore survive longer.

We continue our analysis by showing the time-dependence of the standard deviation, obtained from the ensemble of ABK simulations, in Figure 4.8b. We then compare it to two sets of predictions for the same quantity: the CME-derived standard deviation (equation A.62) that uses the value of the constant \bar{k}_d ,

$$SDev(\langle N_A(t) \rangle; \bar{k}_d)_{CME} = \sqrt{\frac{k_b}{\bar{k}_d} (1 - e^{-\bar{k}_d t})}, \quad (4.28)$$

and the same expression that uses the observed time-dependent $\langle k_d(t) \rangle$ instead (i.e., obtained from the ensemble of ABK simulations, as shown in the inset of Figure 4.8a):

$$SDev(\langle N_A(t) \rangle; \langle k_d(t) \rangle)_{CME} = \sqrt{\frac{k_b}{\langle k_d(t) \rangle} (1 - e^{-\langle k_d(t) \rangle t})}. \quad (4.29)$$

We note that the above expressions of the standard deviation were derived from the CME while assuming that k_d is time-invariant. The justification for using $\langle k_d(t) \rangle$ in equation 4.29 is that we can imagine solving the CME for small enough non-overlapping time intervals such that k_d is approximated as constant in each of them. We used $\Delta t = 0.01$ sec in our simulations, an interval duration that is sufficiently small so that equation 4.29 is a good fit to the standard deviation obtained from our simulation data ($R^2 = 0.9867$). Also notice that this CME-derived prediction of the standard deviation depends on having a *posteriori* knowledge of $\langle k_d(t) \rangle$, since it can only be obtained by performing the ABK simulation before equation 4.29 can be evaluated. We are not aware of another method for theoretically predicting the statistical properties of hetero-

⁹A 0th order birth process is an unrealistic model of population growth through asexual or sexual reproduction, but it is reasonable to consider this simple example as an approximation to the case where the population size increases at a constant rate.

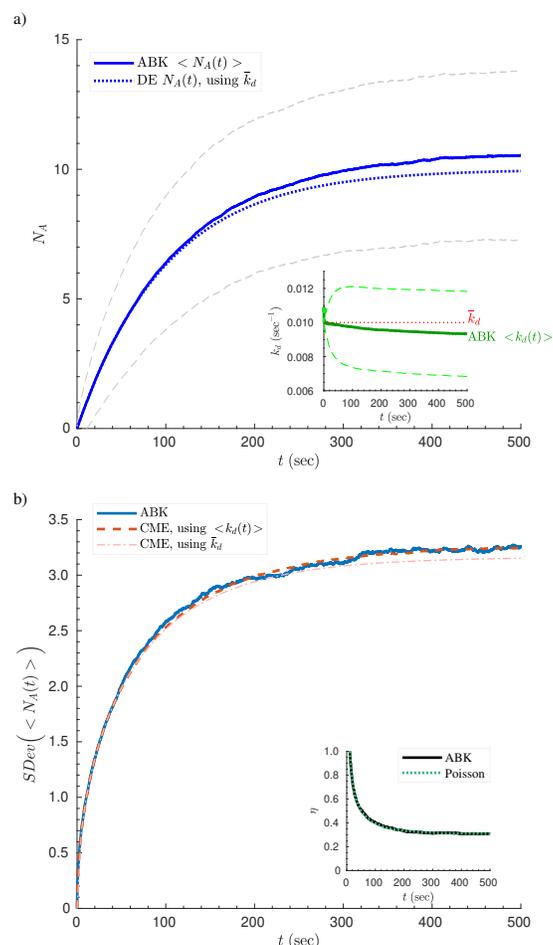


Figure 4.8 – The effect of heterogeneity on a birth-death process. We simulated the reaction scheme $\emptyset \xrightarrow{k_b} A \xrightarrow{k_d} \emptyset$ using the ABK algorithm for a total simulated time of 500 sec with fixed time step increments of $\Delta t = 0.01$ sec. Kinetic parameters: $k_b = 0.10 \text{ sec}^{-1}$, $k_d = 0.010 \text{ sec}^{-1}$, $\sigma_{k_d} = 0.0025 \text{ sec}^{-1}$. We show here the results of an ensemble of 5000 repetitions of the simulation. **a)** Comparison of the time trajectory $N_A(t)$ obtained through an ensemble of ABK simulations (solid blue curve) and the deterministic expectation for a homogeneous population with rate constant $k_d = \bar{k}_d = 0.010 \text{ sec}^{-1}$ (dotted blue curve). The one-standard-deviation envelope for our simulation results is also shown (dashed gray curves). Inset plot: time trajectory of the time-dependent population-wide average degradation rate constant $\langle k_d(t) \rangle$ (solid green curve) and its one-standard-deviation envelope (bounded by the dashed light green curves). The dotted red line corresponds to the mean, \bar{k}_d , of the normal distribution of newly created A agents through the 0th order birth process. **b)** Plot of the standard deviation obtained through the ensemble of ABK simulations (solid blue curve) and the predicted standard deviation using the CME with either \bar{k}_d (dash-dotted faint red curve) or the observed $\langle k_d(t) \rangle$ (dashed red curve). The agreement between our simulation and the latter curve is excellent ($R^2 = 0.9867$). Inset plot: comparison of the coefficient of variation, η , from our simulations and that expected for a Poisson process ($R^2 = 0.9999$).

geneous population trajectories.¹⁰ We propose that performing a computational experiment through the use of ABK to obtain an ensemble of trajectories is an effective method for estimating these statistical properties.

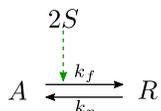
We proved in Appendix A.2 that a birth-death process is Poissonian and with standard deviation given by $SDev(\langle N_A(t) \rangle) = \sqrt{\langle N_A(t) \rangle}$. Evaluating this expression using the average ABK trajectory, $\langle N_A(t) \rangle_{ABK}$, gave values virtually indistinguishable from $SDev(\langle N_A(t) \rangle; \langle k_d(t) \rangle)_{CME}$ (not shown). Finally, the inset plot in Figure 4.8b shows the coefficient of variation for our simulation results, η , and that predicted for a Poisson process, $\eta_P = 1/\sqrt{\langle N_A(t) \rangle}$. The agreement is excellent ($R^2 = 0.9999$). We therefore conclude that the birth-death process of a heterogeneous population is consistent with Poissonian statistics.

Given our observations in simulations of 1st order systems (section 4.4) and those presented above, we predict that the presence of heterogeneity in the 1st order degradation rate constant of any species will produce a shift in the population-wide distribution of the rate constant value resulting in an enrichment of more slowly transitioning agents.

¹⁰The combinatorial method we introduced in section 4.4 is of dubious value in this case because agents are continuously synthesized and degraded. Any enumeration of agent combinations within the population could only be performed synchronously with the stochastic simulation, as this would be the only way full knowledge of the population's makeup at a particular time point can be obtained. Moreover, the method would only make predictions for the duration of the current time step Δt , which would make this approach prohibitively computationally more expensive than the ABK algorithm.

4.9 Heterogeneity in the Sigmoidal Switch Motif

We now examine the effect of compositional heterogeneity on continuous switches (see section 3.2). Specifically, we revisit the abbreviated 2-step sigmoidal response motif, shown below (this is the same as Figure 3.5):



As a reminder, k_f and k_r are the microscopic rate constants for the forward and reverse reactions respectively, and we make the following assumptions: a) two agents of S are needed to promote the forward transition $A \rightarrow R$; b) the population size of S agents, N_S , is held constant; c) $N_{TOT} = N_A(t) + N_R(t)$ is constant.

We consider several cases of compositional heterogeneity in the population of either species A or R (or both) with respect to the process they participate in. Since there are no additional birth or death processes beyond the interconversion between A and R , we use a PHM array for either species that holds the specified values for the kinetic rate constants. That is, the i^{th} value in the PHM is used to calculate the transition probability of the i^{th} agent according to the species it belongs to. To further speed up the simulation, we use a 1:1 mapping to keep track of A - R interconversions: the i^{th} A agent is converted to the i^{th} R agent and vice-versa, so that an agent of the respective species takes on the value of the rate constant assigned to it in its i^{th} PHM entry (we simulated this motif with homogeneous populations using the same mapping). This way of modeling the motif is consistent with the interpretation that an agent of a subspecies is merely modified according to the forward and reverse processes, as would be the case when a protein molecule is phosphorylated/de-phosphorylated and heterogeneity can result from a change in the protein's isoform-dependent kinetic properties upon the modification. As we did when simulating this motif with homogeneous populations, we note that the forward reaction is essentially modeled as a 1st order process with N_S only affecting the numerical value of the transition probability (i.e., we do not explicitly model A - S agent interactions; the KCF with respect to each A agent is $\Omega_A = k_f N_S^2$; see subsection 3.2.1 for a discussion on this). Therefore, a 1-dimensional PHM array is appropriate for simulating the forward reaction. We remind the reader that our treatment here is general and not meant to be limited to biochemical systems. As such, the presence of heterogeneity that we will examine is represented through idealized scenarios that can be used to assess the effect of diversity on population time trajectories and the system's emergent properties.

We present below a summary of the cases we investigated with the simplest possible heterogeneous population structure: the indicated species consists of two subspecies, each with an equal number of entries with a distinct value in

the PHM. For instance, heterogeneity in species A involves two subspecies ($A1$, $A2$), each with $N_{TOT}/2$ placeholders in the PHM for the respective k_f values. In our simulations, we used $N_{TOT} = 100$, therefore each subspecies has 50 entries in the PHM for k_f (and similarly for species R and k_r values).

We list the cases here according according to the shorthand notation we will use to refer to them in the ensuing figures.

HET1 k_f : Species A consists of two subspecies with k_f values of 0.50×10^{-4} and $1.50 \times 10^{-4} \text{ sec}^{-1}$.

HET1 k_r : Species R consists of two subspecies with k_r values of 0.05 and 0.15 sec^{-1} .

HET2 k_f : Species A consists of two subspecies with k_f values of 0.25×10^{-4} and $1.75 \times 10^{-4} \text{ sec}^{-1}$.

HET2 k_r : Species R consists of two subspecies with k_r values of 0.025 and 0.175 sec^{-1} .

HET1 k_f and k_r : Species A consists of two subspecies with k_f values of 0.50×10^{-4} and $1.50 \times 10^{-4} \text{ sec}^{-1}$, and species R of two subspecies with k_r values of 0.05 and 0.15 sec^{-1} .

HET2 k_f and k_r : Species A consists of two subspecies with k_f values of 0.25×10^{-4} and $1.75 \times 10^{-4} \text{ sec}^{-1}$, and species R of two subspecies with k_r values of 0.025 and 0.175 sec^{-1} .

Notice that all of the above six cases result in the full population's average values of $\langle k_f \rangle = 1.00 \times 10^{-4} \text{ sec}^{-1}$, $\langle k_r \rangle = 0.10 \text{ sec}^{-1}$, as was the case when we simulated homogeneous populations in subsection 3.2.1. In all of the simulations we will present, the initial population sizes are $N_{A,i} = 25$ and $N_{R,i} = 75$. The agents were initially randomly assigned to their PHM placeholders so that the initial average rate constant values are approximately those given above.

We first show the SR curves for the above cases in Figure 4.9a. It is clear that the presence of heterogeneity results in the SR curve being different from the one computed through the simulation of homogeneous populations. The figure reveals the following trends: a) Heterogeneity in the k_r values of species R shifts the SR curve to the left of the homogeneous one. Accordingly, the $N_{S,50}$ value (producing a half-maximal response in N_R^*) is lower than in the homogeneous case. Also, the maximum slope of the sigmoidal curve and maximum value of N_R^* remain the same.¹¹ We have also performed simulations with normally-distributed rate constant values where the number of subspecies is the species

Follow this [link](#) to see the Matlab code.

¹¹The reader will also notice that N_R^* is nonzero for low N_S values (e.g., for $N_S = 0$) in the HET2 k_r case (curve consisting of green triangles in Figure 4.9a). This is because the slower subspecies (with the smaller agent-specific k_r value of 0.025 sec^{-1}) dominates the population of R , thus slowing the progression of population trajectories to the steady state within the simulated 100 sec time period. Consequently, the N_R^* obtained from our simulations is nonzero. This misrepresents the system's long-term behavior because, indeed, $N_R \rightarrow 0$ as $t \rightarrow \infty$, albeit with slower kinetics than in the homogeneous case.

population size and all R agents have a unique k_r value. Figure 4.9b (cyan and magenta curves) shows that such heterogeneity in k_r again results in a shift of the SR curve to the left of the homogeneous case. b) Heterogeneity in the k_f values of species A shifts the SR curve to the right of the homogeneous case, resulting in the computed $N_{S,50}$ value being larger (Figure 4.9a). Moreover, the maximum slope of the sigmoidal curve and maximum value of N_R^* decrease. c) The HET1/2 k_f and k_r cases (Figure 4.9a) show that heterogeneity in both species balances the above two effects producing an SR curve that is virtually indistinguishable from the homogeneous case ($R^2 = 0.9998$). Although one may be tempted to think that this balancing effect from the dual presence of heterogeneity may be general, we show through a counterexample in Figure 4.9b (green dotted curve) that it is not. The observed SR curve is obviously dependent on the species population structures and degree of heterogeneity (as quantified by the various metrics of population diversity; see section 4.2).

We also plot the time-dependent population-based average rate constant values¹² for three of the above cases in Figure 4.10. In all cases, the same effect as previously illustrated when studying heterogeneity in a birth-death process (Figure 4.8a) is present: the subpopulation with the higher rate constant gets depleted by a surge of transition events in the early time stages of the simulation, while the slower subpopulation dominates the species population at later time stages. The extent of this enrichment in agents with slower rate constants is modulated by the value of N_S and which species has a heterogeneous population. When species R consists of two subspecies (Figure 4.10b), $\langle k_r \rangle$ settles to a higher steady state (although still lower than the initial $\langle k_r \rangle$ value) as N_S increases. This is because the forward process becomes quadratically faster as N_S increases, thus having a greater tendency to replenish the pool of fast R agents at any given time point. On the other hand, when heterogeneity in species A is present (Figure 4.10a) $\langle k_f \rangle$ settles to a lower steady state as N_S increases. This can be explained by the fact that the forward process depends on N_S^2 and the faster A agents are more likely to transition to R as N_S increases, thereby leaving the slower A agents to dominate the population of A . Finally, case HET1 k_f and k_r (Figure 4.10c,d) is interesting because the presence of heterogeneity in both A and R balance each other out in this case, and $\langle k_f \rangle$ reaches the same steady state for all values of N_S . The initial bump in $\langle k_f \rangle$ for low N_S values is due to a surge of fast R agents being converted to A . Recall the initial population sizes, $N_{A,i} = 25$ and $N_{R,i} = 75$, therefore, the high initial population of fast R agents leads to a surge of transitions to A while the probability of a forward process transition is small for low N_S values.

The above results suggest that caution should be exercised when interpreting experimental results obtained from population-based or batch measurements. For instance, a researcher trying to measure the $N_{S,50}$ value or cooperativity (e.g., Hill coefficient) of a sigmoidal response can get significantly different estimates depending on the degree of heterogeneity in the population of the in-

¹²The average rate constant given the structure of the corresponding species population at time t . This can show which subspecies population becomes enriched as simulated time elapses.

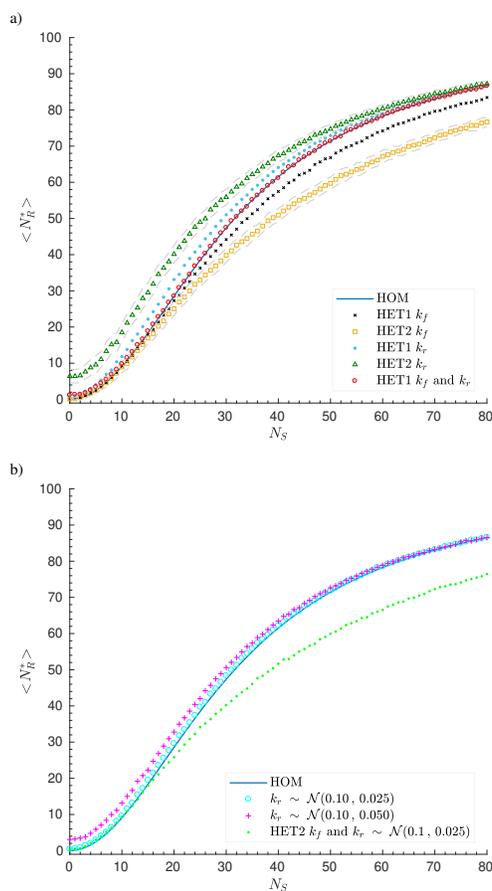


Figure 4.9 – Heterogeneity in the abbreviated 2-step sigmoidal response motif: SR curves.

For both plots, the simulation parameters are: kinetic constants $\langle k_f \rangle = 0.0001 \text{ sec}^{-1}$, $\langle k_r \rangle = 0.10 \text{ sec}^{-1}$, initial populations $N_{A,i} = 25$ and $N_{R,i} = 75$, and total simulated time of 100 sec with fixed time step increments of $\Delta t = 0.01 \text{ sec}$. We performed separate simulations for integer values of S agents, $N_S = 0 - 80$. We estimated N_R^* by averaging N_R over the course of the last 50 sec of simulated time to account for the variation in population size due to stochasticity once the steady state had been reached. We then repeated the simulation ($n = 100$) for each value of N_S to obtain the average and standard deviation of N_R^* . We plot here the average steady-state population size of R , $\langle N_R^* \rangle$ vs. N_S for each of the simulated cases. We show the resulting SR curves in two separate panels for clarity. **a)** We used the ABK algorithm to model this motif for the two-subspecies cases described in the text (we do not show case HET2 k_f and k_r , which produced virtually identical results to case HET1 k_f and k_r , to avoid overcrowding the plot). The blue curve is the reference SR curve obtained by simulating homogeneous populations (Figure 3.6). The one-standard-deviation envelope is also shown for the HET2 k_f and HET2 k_r SR curves (gray dashed curves). **b)** We performed additional simulations using the same methods as above. The notation we use for the normal distribution is $\mathcal{N}(\mu, \sigma)$. We show here two examples of heterogeneity in which all R agents have a unique k_r value, with k_r sampled from the normal distribution $\mathcal{N}(0.10, 0.025)$ and $\mathcal{N}(0.10, 0.050)$. We also plot the SR curve where both species A and R feature heterogeneous populations: A consists of two subspecies with k_f values of 0.25×10^{-4} and $1.75 \times 10^{-4} \text{ sec}^{-1}$ (same as case HET2 k_f) and species R with $k_r \sim \mathcal{N}(0.10, 0.025)$.

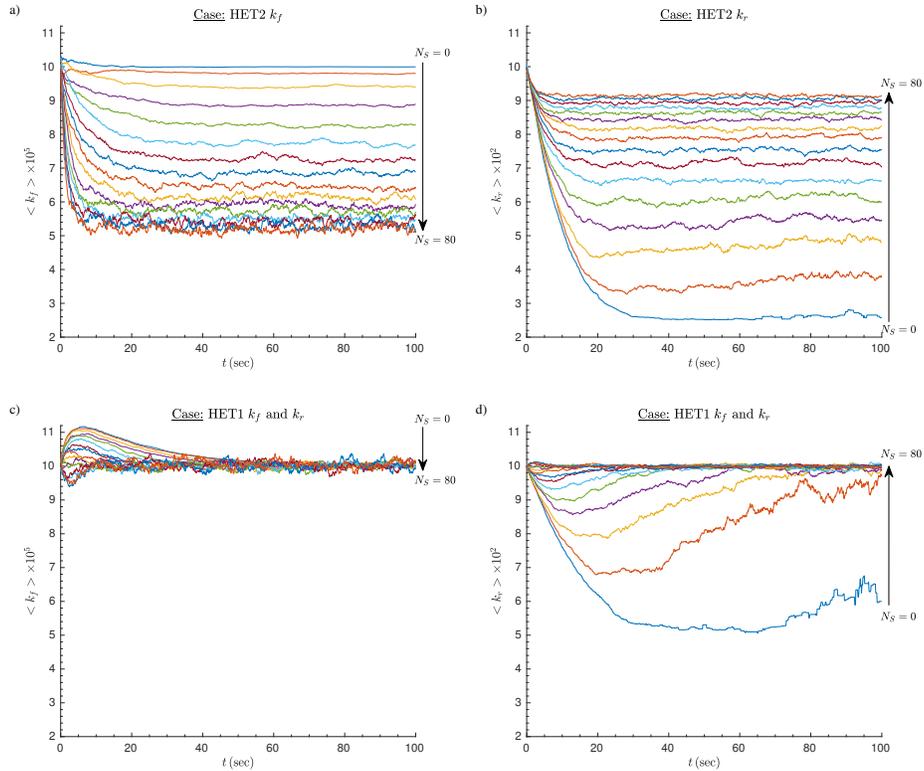
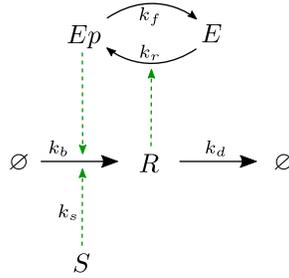


Figure 4.10 – Heterogeneity in the abbreviated 2-step sigmoidal response motif: time-dependence of population-based average rate constants. We plot the time-dependence of the average rate constant values, $\langle k_f(t) \rangle$ and $\langle k_r(t) \rangle$, with units of sec^{-1} , determined from the transient population structure of species A and R respectively (i.e., depending on which agents belonging to each subspecies are ‘alive’ at time t). We show the following cases: **a)** HET2 k_f , **b)** HET2 k_r , **c,d)** HET1 k_f and k_r . Each plot shows a time series of the indicated average rate constant for selected values of N_S in the simulated range $N_S = 0 - 80$ (only N_S values that are multiples of 5 are shown for ease of visualization). A black vertical arrow on the side of each panel indicates the general progression of the series as N_S increases. Notice that the scale of the vertical axes differs depending on which rate constant is being plotted.

volved species. Therefore, it is clear that knowledge of the presence of underlying population heterogeneity is essential for assessing the accuracy of estimated parameters and interpreting population-based metrics of a system's dynamical behavior.

4.10 Heterogeneity in the Hysteretic Switch from Mutual Activation

We now revisit the hysteretic switch motif presented in subsection 3.3.3 and consider the effect of heterogeneity in just one of the involved processes. For convenience, we show below the motif's interspecies connectivity (this is identical to Figure 3.23), where the mutual activation loop formed by species R and Ep can lead to bistability for a range of parameter values.



We remind the reader that we assume no further synthesis and degradation reactions for E and Ep , therefore $N_E(t) + N_{Ep}(t) = \text{constant} = N_{E,TOT}$. In the following discussion, we will use the same set of parameters as we did when simulating this motif with homogeneous populations, except for the parameter $K_{M,r}$. Specifically, instead of using a fixed value for the Michaelis constant of the reverse reaction for all agents belonging to species E , we examine here the case of two E subspecies, $E1$ and $E2$, with equal subpopulation sizes. For instance, setting an initial population size $N_{E,i} = N_{E,TOT} = 100$, then 50 $E1$ agents have $K_{M,r} = 1$ and 50 $E2$ agents have $K_{M,r} = 19$, such that $\langle K_{M,r}(t=0) \rangle = 10$ and $SDev\{K_{M,r}(t=0)\} \approx 9.05$. Notice that this average value of $K_{M,r}$ is the same as the fixed one we used when simulating a homogeneous E population. We encode this heterogeneity using a 1-dimensional PHM vector,¹³ the i^{th} entry of which serves as a placeholder for $K_{M,r}^{(i)}$ of the i^{th} E agent *when it is in existence*. That is, upon a conversion of an Ep agent to the i^{th} E agent, the latter takes on the value of $K_{M,r}^{(i)}$ in the PHM. Accordingly, the time-dependent population-wide average $\langle K_{M,r}(t) \rangle$ fluctuates around the average value of 10 and the size of the fluctuations varies inversely with the population size $N_E(t)$.¹⁴ We show

Follow this [link](#) to see the Matlab code.

¹³The reader will note that the conversion of E to Ep is a 2nd order process based on our definition of this motif, since R promotes this reaction without being consumed. Consequently, a 2-dimensional PHM would be needed in the general case where heterogeneity exists in the E - R agent interactions. However, we assume for the sake of simplicity in the present case that the population of R is homogeneous with respect to the E - R interactions, and the PHM “collapses” to a 1-dimensional vector describing only the variation of parameter $K_{M,r}$ in species E .

¹⁴The assigned $K_{M,r}^{(i)}$ in the PHM is time-independent, however the population-wide $\langle K_{M,r}(t) \rangle$ varies depending on the structure of the E population (i.e., which E agents are ‘alive’ or ‘dead’ at a particular time point). This is a convenient way of implementing the

that this is indeed the case in Figure 4.11, where we compare the species time trajectories when the system is in the OFF or ON switch states. In the latter case, the population $N_E(t)$ is small and the fluctuations of $\langle K_{M,r}(t) \rangle$ become more pronounced.

We now investigate if the presence of heterogeneity has a statistically significant effect on the SR curve of this system. We simulated this motif with the same parameters as listed in Figure 4.11, taking into account the heterogeneity in the parameter $K_{M,r}$ to obtain the SR curve for two sets of initial population sizes: $N_{R,i} = 0$, $N_{Ep,i} = 0$, $N_{E,i} = 100$ and $N_{R,i} = 10$, $N_{Ep,i} = 50$, $N_{E,i} = 50$. As we described in subsection 3.3.3, these initial conditions yield significantly different curves. We show a comparison of these results to the deterministic and ABK-simulated SR curves for the respective homogeneous populations in Figure 4.12. It is evident that the heterogeneous population SR curves, sampled at integer values of N_S , appear to lie above the respective homogeneous ones in part of or beyond the deterministic region of bistability. We address if this trend is statistically significant by applying the two-sample (heterogeneous vs. homogeneous populations) right-tailed t -test. Therefore, for each value of N_S , we evaluate the test statistic

$$t_{N_S} = \frac{\langle N_{R,HET}^* \rangle - \langle N_{R,HOM}^* \rangle}{\sqrt{\frac{1}{n} (s_{HET}^2 + s_{HOM}^2)}}, \quad (4.30)$$

where the designations HET and HOM denote quantities obtained from simulations of heterogeneous and homogeneous populations respectively. The quantity s is the observed standard deviation, and n is the number of observations in each sample (i.e., the number of times we repeated the simulation). In this case, $n = 1000$, which is large enough to ensure that there is a high number of degrees of freedom, so the t -distribution is essentially the normal distribution. We can then conclude that $\Delta \langle N_R^* \rangle \equiv \langle N_{R,HET}^* \rangle - \langle N_{R,HOM}^* \rangle$ is greater than 0 at the 99% confidence interval when $t_{N_S} > 2.326$ (this is the normal distribution's z -value for $\alpha = 0.01$). We plot $\Delta \langle N_R^* \rangle$ for each value of N_S in Figure 4.13a, and emphasize the statistically significant points by enlarging them (Table 4.1 lists the relevant p -values). We also show the difference in the standard deviations $SDev\{\langle N_{R,HET}^* \rangle\} - SDev\{\langle N_{R,HOM}^* \rangle\}$ as a function of N_S in Figure 4.13b. Finally, we plot the difference in the percentage of simulations where the system ended up in the ON state in Figure 4.13c, where we define $\Delta \%ON \equiv \%ON_{HET} - \%ON_{HOM}$. It is noteworthy that $\Delta \%ON = 7 - 9\%$ for some values of N_S in the region of bistability, indicating that this nominal presence of heterogeneity (involving only one parameter in one process of the motif) can have a significant impact on the observable properties of the system. For instance, an experimenter harvesting a population of cells at a particular time point after the beginning of an experiment can record significantly different measurements of the aggregate behavior, as long as the measured quantity

presence of heterogeneity in a species with a fixed range in population size, as in this case where $0 \leq N_E(t) \leq N_{E,TOT}$.

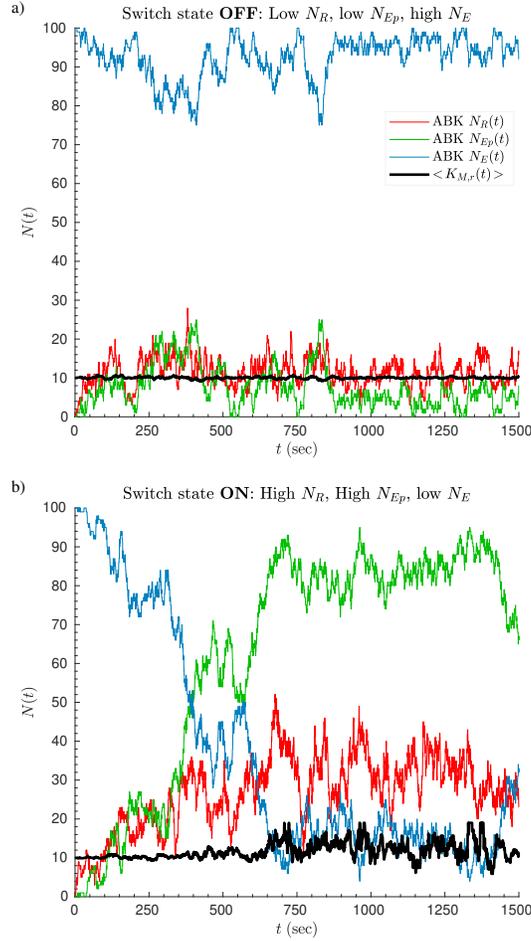


Figure 4.11 – Heterogeneity in the mutual activation switch motif: sample time trajectories. We used the ABK algorithm to model the time evolution of species R , Ep and E in the mutual activation hysteretic switch motif. We consider this motif with the following parameters: $k_b = 0.02 \text{ sec}^{-1}$, $k_s = 0.05 \text{ sec}^{-1}$, $k_d = 0.075 \text{ sec}^{-1}$, $k_f = 1 \text{ sec}^{-1}$, $k_r = 0.05 \text{ sec}^{-1}$, $K_{M,f} = 5$, $N_{E,TOT} = 100$, $N_S = 15$, and initial population sizes $N_{R,i} = 0$, $N_{Ep,i} = 0$, $N_{E,i} = 100$. The population of species E is heterogeneous with respect to the Michaelis constant for the reverse reaction in the conversion $Ep \rightleftharpoons E$: $\langle K_{M,r}(t=0) \rangle = 10$, $SDev\{K_{M,r}(t=0)\} \approx 9.05$, with two subspecies, $E1$ and $E2$, of equal initial subpopulation sizes (fractional abundance $\chi_{E1} = 0.5 = \chi_{E2}$; index of heterogeneity $\psi = 0.50$). In each case, the simulation was performed with fixed time step increments of $\Delta t = 0.01 \text{ sec}$ for a total simulated time of 1500 sec. Given the initial condition, this system composed of homogeneous populations is deterministically predicted to stay in the OFF state (low N_R ; deterministic $N_R^* = 11.43$). In panel *b*, a noise-induced transition to the ON state occurs in the time window of 300 – 600 sec (high N_R ; deterministic $N_R^* = 33.12$). We also plot $\langle K_{M,r}(t) \rangle$, whose fluctuations become noticeably larger when N_E is low (switch state ON). Note that the legend in the top panel applies to both plots.

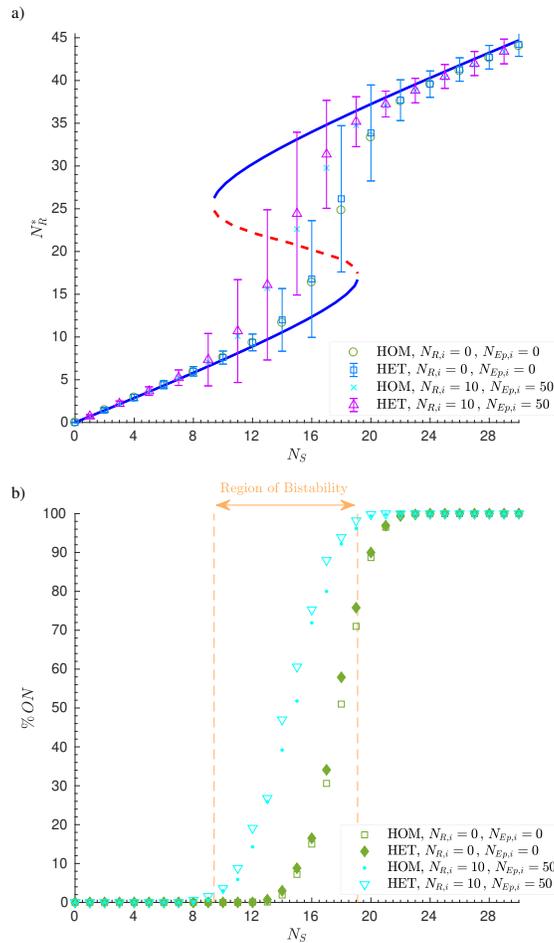


Figure 4.12 – Heterogeneity in the mutual activation switch motif: SR curve. We used the ABK algorithm to simulate the mutual activation switch motif with the same parameters as listed in Figure 4.11 that account for heterogeneity in $K_{M,r}$ values in the population of species E , and integer values of the signal species in the interval $0 \leq N_S \leq 30$. The total simulated time was 1500 sec with fixed time step increments of $\Delta t = 0.02$ sec. We averaged the population sizes over the last 750 sec to estimate N_R^* for each simulation. We then repeated this procedure 1000 times to obtain statistical measures of the N_R^* distribution. We show the results of the simulation for two sets of initial population sizes: $N_{R,i} = 0, N_{Ep,i} = 0, N_{E,i} = 100$ (i.e., the switch starts out in the OFF state's basin of attraction), and $N_{R,i} = 10, N_{Ep,i} = 50, N_{E,i} = 50$. We then compare them to identical simulations of the respective homogeneous populations ($n = 1000$). The legends in both panels indicate whether the curves were obtained from simulating homogeneous or heterogeneous populations with the descriptors HOM and HET respectively. **a)** $\langle N_R^* \rangle$ from the ABK simulation of the system with a heterogeneous E population superimposed on the deterministic SR curve and average N_R^* obtained from simulating homogeneous populations. We only show the error bars in the heterogeneous case for clarity (each half of the error bars is one standard deviation from the mean). **b)** Plot of the frequency of the system being in the ON state for each N_S value. The vertical orange dashed lines show the range of N_S values for which the system is deterministically predicted to be bistable.

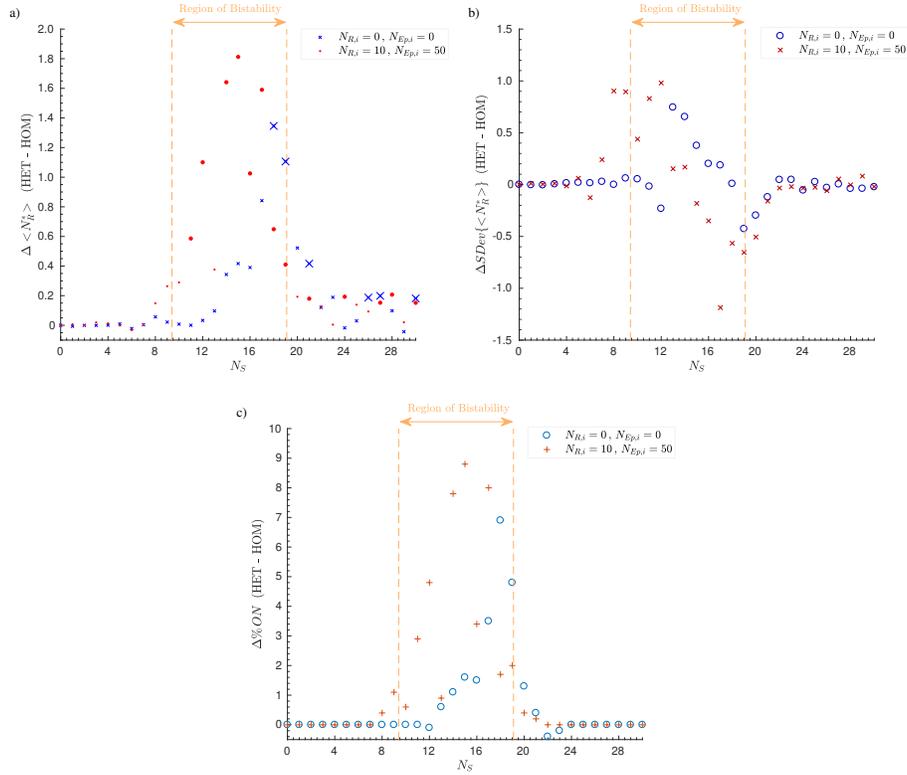


Figure 4.13 – Heterogeneity in the mutual activation switch motif: differences in $\langle N_R^* \rangle$, $SDev\{\langle N_R^* \rangle\}$, and switch state. We further analyze the results presented in Figure 4.12 by showing the difference in $\langle N_R^* \rangle$, $SDev\{\langle N_R^* \rangle\}$, and the percentage of simulations ending up in the ON switch state. The deterministically predicted region of bistability is shown as vertical dashed orange lines in all panels. **a)** We plot $\Delta \langle N_R^* \rangle \equiv \langle N_{R,HET}^* \rangle - \langle N_{R,HOM}^* \rangle$ as a function of N_S . The enlarged markers correspond to statistically significant differences in $\langle N_R^* \rangle$ at the 99% confidence interval (see Table 4.1 for a list of the p -values). **b)** Difference in the standard deviation, $\Delta SDev\{\langle N_R^* \rangle\} \equiv SDev\{\langle N_{R,HET}^* \rangle\} - SDev\{\langle N_{R,HOM}^* \rangle\}$, as a function of N_S . **c)** Difference in the percentage of simulations where the system ended up in the ON state at the end of simulated time (1500 sec), defined as $\Delta \%ON \equiv \%ON_{HET} - \%ON_{HOM}$, as a function of N_S .

$N_{R,i} = 0, N_{Ep,i} = 0, N_{E,i} = 100$		$N_{R,i} = 10, N_{Ep,i} = 50, N_{E,i} = 50$	
N_S	p	N_S	p
18	2.15×10^{-4}	11	9.84×10^{-3}
19	5.88×10^{-4}	12	5.75×10^{-4}
21	8.08×10^{-3}	14	6.64×10^{-5}
26	1.25×10^{-3}	15	1.25×10^{-5}
27	6.41×10^{-4}	16	3.45×10^{-3}
30	1.88×10^{-3}	17	1.44×10^{-7}
		18	1.49×10^{-3}
		19	2.48×10^{-3}
		21	5.82×10^{-3}
		24	1.29×10^{-3}
		27	6.71×10^{-3}
		28	3.70×10^{-4}
		30	8.10×10^{-3}

Table 4.1 – Heterogeneity in the mutual activation switch motif: p -values for statistically significant differences in $\langle N_R^* \rangle$. We list here the p -values for the cases where $\Delta \langle N_R^* \rangle \equiv \langle N_{R,HET}^* \rangle - \langle N_{R,HOM}^* \rangle > 0$ at the 99% confidence interval.

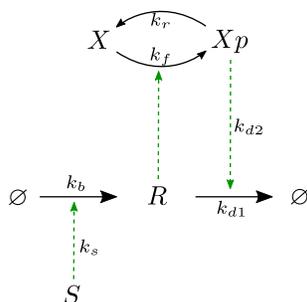
correlates with the sum of each individual cell’s switch state (e.g., fluorescence intensity of a reporter protein such as GFP).

This case study highlights the importance of having a simulation tool for modeling the effects of heterogeneity on system dynamics because it is difficult or impossible to have an *a priori* expectation for how heterogeneity will influence the system’s emergent properties. Here, we observed that the mutual activation motif has a small but statistically significant tendency toward higher N_R^* values and being in the ON switch state when there is diversity in the $K_{M,r}$ values of E agents.

4.11 Negative Feedback 2-Component Motif with Heterogeneity in the Explicit Time Delay

We now return to a motif we examined for the case of homogeneous populations in subsection 3.5.2, where we saw the effect of specifying an explicit time delay in one of the processes on oscillatory activity. Here, we will extend this discussion to different ways that heterogeneity could be manifested and their effect on the system's oscillatory dynamics.

We remind the reader the processes involved and their connectivity in the negative feedback motif below (same as in Figure 3.37):



The computational experiments we will present will again focus on the regulation of the 2nd order degradation of R (with rate constant k_{d2}) by species Xp , and introduce heterogeneity in the following two ways:

1. The delay value, δ , is not constant, but is instead sampled from a user-specified distribution. Specifically, we use here a normally-distributed delay value. The delay for the i^{th} agent in the population of R , δ_i , in each time step of the simulation is determined by drawing a random number, r_n , from the standard normal distribution, and evaluating the expression

$$\delta_i = \bar{\delta} + \sigma_\delta r_n, \quad (4.31)$$

where the mean and standard deviation of the distribution are given by the parameters $\bar{\delta}$ and σ_δ respectively. We will use the values $\bar{\delta} = 50$ sec and $\sigma_\delta = 10$ sec for the studies we will present later in this section.

The probability that the i^{th} R agent is degraded depends on N_{Xp} and is computed in each time step of duration Δt using¹⁵

$$P_{\mathbf{R}_i + Xp \rightarrow \emptyset}(t, \delta_i) = k_{d2} N_{Xp}(t - \delta_i) \Delta t, \quad (4.32)$$

for $t > \delta_i$. We assume that all agents of R have a nonzero delay value associated with their probability of being degraded. Notice that the population of Xp remains homogeneous in this scenario, hence it is the total

¹⁵We show here the form of transition probability derived from the discretized differential kinetic rate law for this process, or P_{dif} . We have also performed these simulations using the P_{ber} form and obtained results indistinguishable (in the aggregate) from the ones presented below.

population size, N_{X_p} , that is used in the above expression. It is only the variation in the delay parameter associated with individual R agents that constitutes the presence of compositional and temporal heterogeneity in the motif.

2. Only a fraction of the population of R consists of agents with a nonzero delay value (we will refer to this subspecies as $R1$), and the remainder have $\delta_i = 0$ (i.e., no delay; subspecies $R0$). The average fractional abundance of the delay-sensitive subspecies we use here is $\langle \chi_{R1} \rangle = 1/2$, so that when an R agent is created through a 0th order process (with rate constant k_b and k_s) we randomly choose if it has a nonzero delay with probability $1/2$. This way, at any given time of the simulation approximately half of the population of R consists of agents with a nonzero δ value and whose transition probability is given by equation 4.32.

The corresponding transition probability for an agent of subspecies $R0$ is, of course,

$$P_{R0+X_p \rightarrow \emptyset}(t) = k_{d2} N_{X_p}(t) \Delta t, \quad (4.33)$$

reflecting the fact that no delay is taken into account. Therefore, heterogeneity is encoded in the presence of two distinct subspecies of R , an agent being distinguished as a member of either subspecies by the boolean attribute of dependence on a delay with respect to the degradation process. Beyond this, the $R1$ delay could be a constant value or sampled from a distribution. Here, we will explore both of these possible cases.

We simulated four distinct cases of this motif to explore the effects of the different forms of heterogeneity outlined above: **a)** *all* agents of species R are sensitive to a constant delay value of $\delta = 50$ sec, therefore this case is identical to the one presented in subsection 3.5.2 on homogeneous populations;¹⁶ **b)** *all* agents of species R are sensitive to a normal distribution of delay values with $\bar{\delta} = 50$ sec and $\sigma_\delta = 10$ sec, and the details of the implementation are described in 1, above; **c)** only subspecies $R1$ (with $\langle \chi_{R1} \rangle = 1/2$) is sensitive to a constant delay value of $\delta = 50$ sec; **d)** only subspecies $R1$ (with $\langle \chi_{R1} \rangle = 1/2$) is sensitive to a normal distribution of delay values with $\bar{\delta} = 50$ sec and $\sigma_\delta = 10$ sec. We show the results of these simulations in Figures 4.14-4.17 (the plot labels in figures 4.14 and 4.15 preserve the case identification a-d). We will not dwell on the features of the simulated trial and average trajectories that we have encountered before (see section 3.5). Instead, we highlight the effects of heterogeneity on the system's behavior.

The sample time trajectories in Figure 4.14 show that splitting the population of R into two distinct subspecies results in the loss of consistent oscillatory activity (panels c, d). Moreover, where a peak is unmistakably present, its

¹⁶The only parameter that is different is the initial period when the system was simulated without time delays. Here we use the time period 0 – 100 sec (instead of 0 – 50 sec in Figure 3.41); this accounts for δ being normally-distributed in some of the ensuing cases. This parameter change causes the DDE solution to feature a smaller amplitude of the predicted oscillations.

Follow this [link](#) to see the Matlab code.

height or value of N_R at the top of the peak is smaller than when all R agents are delay-sensitive (plots a, b). The difference in peak heights between the different cases also results in a proportional change in the standard deviation of the average trajectories (e.g., shown for $\langle N_R(t) \rangle$ in Figure 4.15).

We quantified this difference in peak heights by applying a filtering method to smooth out the stochastic trajectories and then detecting peaks that correspond to instances of oscillatory activity (see details in the caption of Figure 4.16). We used a set of peak detection parameters that reliably picked out the desired feature in the simulated trajectories of cases **a**, **b**, and then applied the same parameters for the remaining cases. This resulted in progressively fewer peaks being detected when distinct R subspecies are present and a distribution of delay values is used (985, 950, 706, 676 detected peaks in cases **a-d**, respectively; this data was aggregated by repeating the simulation of each case 100 times). The average of the detected peak heights in the respective cases are 104.6 ± 24.8 , 100.0 ± 21.2 , 80.5 ± 8.4 , 77.6 ± 8.2 (mean \pm standard deviation), indicating that the peak heights are significantly reduced in the presence of distinct subspecies. We also plot the observed distribution of peak heights in Figure 4.17a, which verifies the much narrower range of peak heights in cases **c**, **d**.

We also collected data on the time intervals between successive detected peaks. This allowed us to plot the observed distributions of peak intervals for each of the four cases (Figure 4.17b). It is clear that the distributions become wider when the delay value is not constant and the population of R consists of distinct subspecies. This is at least in part due to the fact that fewer peaks were detected when subspecies were present, thus the time intervals between them tend to be longer (for an example of this, compare panels a and b in Figure 4.16).

Our computational experiments using the ABK method have once again shown that population heterogeneity with respect to only one process can significantly affect the behavior of a relatively simple system. In this case, we saw that a distribution of the delay parameter of a degradation process leads to less well-defined oscillatory activity produced by the 2-component negative feedback loop.

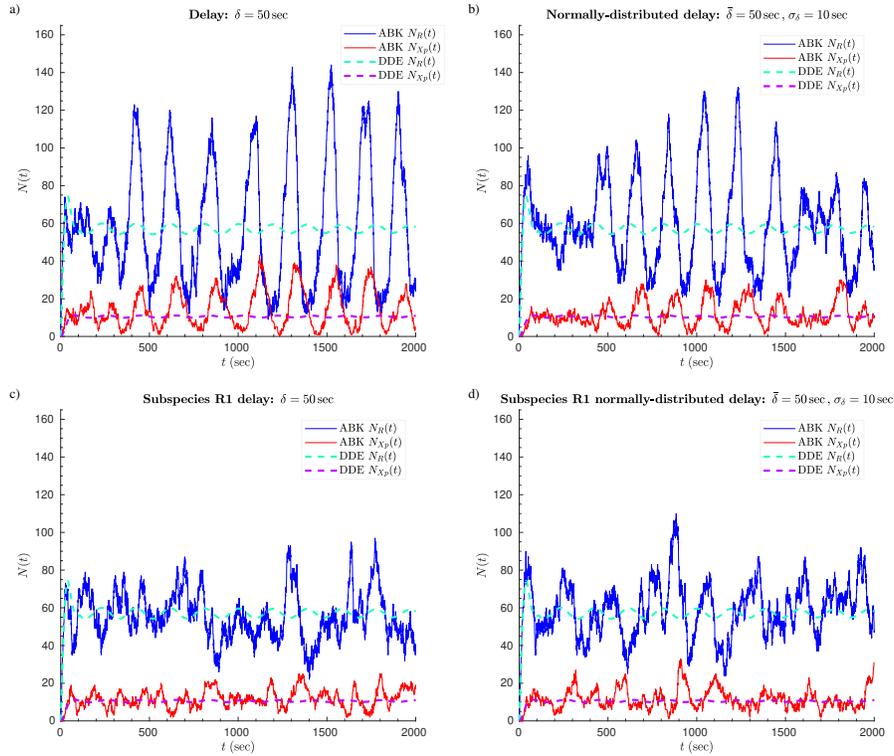


Figure 4.14 – Heterogeneity in the negative feedback 2-component motif with delay: sample time trajectories. We used the ABK algorithm to model the time evolution of the negative feedback 2-component motif with a delay in the feedback reaction ($R + Xp \rightarrow \emptyset$). We use the same kinetic parameters as we did in subsection 3.5.2 (see Figure 3.41), except the initial period when the system (ABK and deterministic trajectories) was simulated without time delays is 100 sec (instead of 50 sec) to account for the cases when $\delta_i > 50$ sec. The numerical solution to the set of DDEs use the initial 100 sec period as history. For convenience, we list the kinetic and simulation parameters again: $k_b = 0 \text{ sec}^{-1}$, $k_s = 0.100 \text{ sec}^{-1}$, $k_{d1} = 0 \text{ sec}^{-1}$, $k_{d2} = 0.005 \text{ sec}^{-1}$, $k_f = 0.010 \text{ sec}^{-1}$, $k_r = 1.000 \text{ sec}^{-1}$, Michaelis constants $K_{M,f} = K_{M,r} = 10$, constant signal population size $N_S = 30$, initial population sizes $N_{R,i} = 0$, $N_{Xp,i} = 0$, and total population of species X and Xp , $N_{X,TOT} = N_X(t) + N_{Xp}(t) = 100$. We simulated this motif for a total time of 2000 sec, with fixed time step increments of $\Delta t = 0.01$ sec. The four cases we show here (a-d) are described in the text. We have omitted the time trajectory of subspecies $R1$ in panels c, d for clarity.

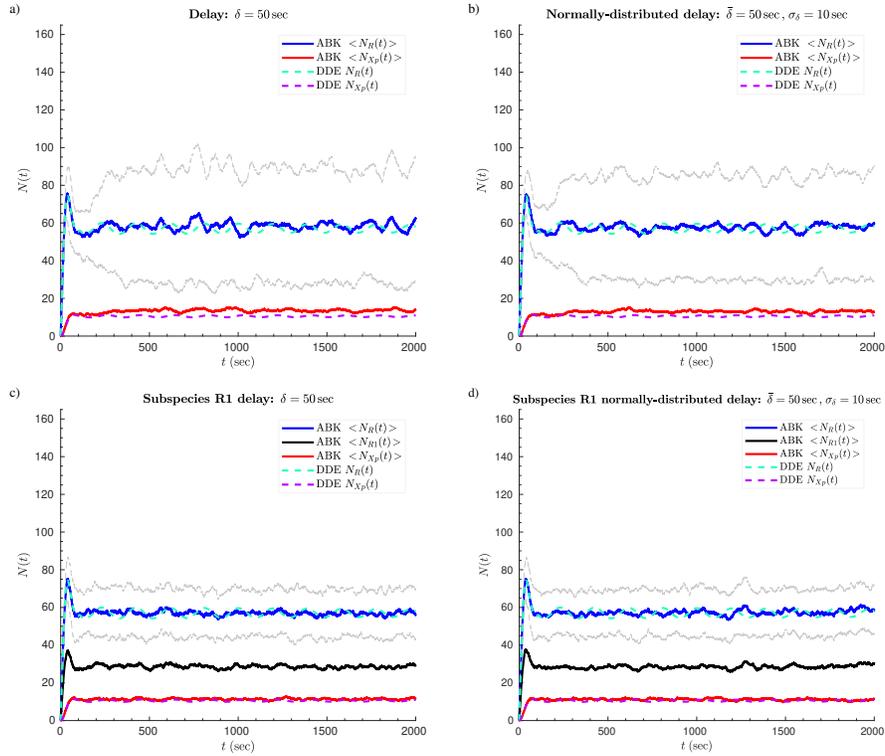


Figure 4.15 – Heterogeneity in the negative feedback 2-component motif with delay: average time trajectories. We show here the average population sizes after 100 repetitions of the simulation. The four cases we show here (a-d) are described in the text. The gray dashed curves denote the one-standard-deviation envelope for $\langle N_R(t) \rangle$ (not shown for species X_p). The average trajectory of subspecies $R1$, $\langle N_{R1}(t) \rangle$, is also shown in plots c and d (solid black curve), and is approximately half of $\langle N_R(t) \rangle$, as expected given that the 0th order birth process produces R agents such that $\langle \chi_{R1} \rangle = 1/2$. Notice that the standard deviation of $\langle N_R(t) \rangle$ when the population of R consists of distinct subspecies (plots c and d) is smaller than when all R agents are delay-sensitive (plots a and b). This is because the fluctuating population sizes are confined to a narrower range of values when there are distinct subspecies (also supported by the observed distribution of peak heights in Figure 4.17a).

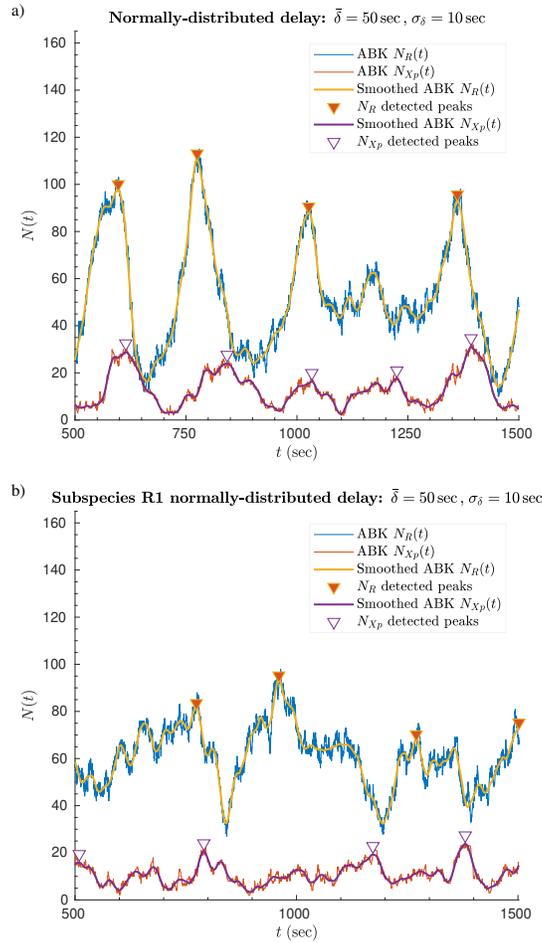


Figure 4.16 – Heterogeneity in the negative feedback 2-component motif with delay: trajectory smoothing and peak detection. We used the Savitzky-Golay smoothing filter to remove the noise and obtain smooth time trajectories (Matlab function: `sgolayfilt`, run with a frame length of 60sec and a polynomial order of 5; see documentation at <http://www.mathworks.com/help/signal/ref/sgolayfilt.html>). We then detected the peaks using the Matlab function `findpeaks` with parameters: species R : minimum peak prominence = 26 (consider peaks that drop off on both sides by at least this number), minimum peak height = 65, minimum peak distance = 45 sec (avoid detecting peaks that are too close in time to each other); species Xp : minimum peak prominence = 7, minimum peak height = 14, minimum peak distance = 45 sec. We show here representative simulated trajectories, along with their smoothed counterparts and detected peaks. We performed the smoothing and peak detection procedures for all simulated trajectories (100 repetitions for each of the four cases described in the text).

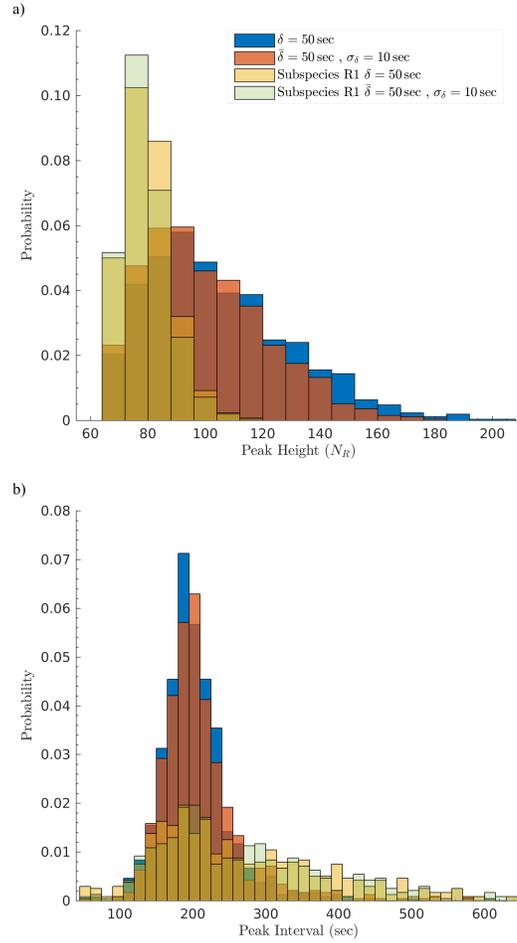


Figure 4.17 – Heterogeneity in the negative feedback 2-component motif with delay: distribution of peak heights and time intervals. Since the number of N_R detected peaks was different in each of the four cases we examined (985, 950, 706, 676 peaks in cases *a-d*, respectively; this data was aggregated by repeating the simulation of each case 100 times), we compare them by normalizing their respective histograms and using the same bin width. We then superimpose them by plotting each with a different level of transparency so that the distributions remain visible to the reader. **a)** Distributions of detected peak heights. The average of the peak heights in the respective cases, as ordered from top to bottom in the legend, are 104.6 ± 24.8 , 100.0 ± 21.2 , 80.5 ± 8.4 , 77.6 ± 8.2 (mean \pm standard deviation, in units of the population size N_R). **b)** Distribution of the time intervals between successive peaks. The respective means are 193.9 ± 50.6 , 207.1 ± 58.4 , 244.8 ± 139.9 , 241.4 ± 151.3 (mean \pm standard deviation, in units of sec). Note that the legend, shown only in *a*, applies to both plots.

4.12 Sensitivity Analysis with Respect to Measures of Population Heterogeneity

Sensitivity analysis is used to assess how the dynamical behavior of a system's output variables depends on variations in the parameters of the involved processes. The method we present in this section allows us to determine the effect of heterogeneity on simulated population time trajectories. For instance, we can vary the initial species- or interaction-specific index of heterogeneity, ψ_i ,¹⁷ by using an identical sequence of (pseudo-)random numbers (accomplished by specifying the same seed *value* for the random number generator) for each pair of simulations with parameter values ψ_i and $\psi_i + \epsilon$, where ϵ represents the magnitude of a small but finite perturbation to ψ_i . This method is sometimes referred to as the *common random numbers* procedure (CRN; Glasserman and Yao, 1992). After repeating this procedure many times, each with a different seed value for the random number generator, we can then compute the finite difference between each pair of mean trajectories and estimate their sensitivity (Gillespie et al., 2013; Rathinam et al., 2010) to changes in ψ ,

$$\text{sens}\{N_j(t; \psi_i), \epsilon\} = \frac{\langle N_j(t; \psi_i + \epsilon) \rangle - \langle N_j(t; \psi_i) \rangle}{\epsilon}, \quad (4.34)$$

where we have used the notation $N_j(t; \psi_i)$ to denote the dependence of the j^{th} species' time trajectory to ψ_i . Note that ψ_i can represent the degree of heterogeneity in any of the species in the system participating in a 1st order process or any intra/interspecies interaction in 2nd and higher order processes. Furthermore, we can also compute the sensitivity of the variance or higher order moments with respect to ψ_i by appropriately adjusting the numerator of equation 4.34.

Although we have singled out ψ for assessing the effect of heterogeneity on population dynamics, sensitivity analysis can be performed with respect to any parameter of interest associated with a *particular* agent or agent-agent interaction, or other population-wide metrics such as the average, \bar{k} , and variance, $\text{Var}\{k\}$, of the kinetic constant values.

Other software tools and packages exist that use deterministic derivative-based methods for performing sensitivity analysis with respect to parameters of ODE models (e.g., Domijan et al., 2016), which is computationally much faster than using ensembles of stochastic simulations (with either ABK or SSA) to accomplish the same analysis. However, as we have seen in several examples in this work, ODEs do not give insight into the uncertainty in the population trajectories and small population sizes can give surprising results that the deterministic approach may inadequately describe or altogether miss. Importantly, only the ABK method can address the effect of heterogeneity on population dynamics. We therefore recommend that an ABK-based sensitivity analysis be

¹⁷We emphasize that this is the initial value of $\psi_i = \psi(t=0)$, because the index of heterogeneity will be time-dependent as the population structure and size of a species or subspecies change during their time trajectories.

performed when simulating systems that include species with small population sizes and where compositional/temporal heterogeneity are known or suspected to exist.

4.12.1 Case Study: Sensitivity Analysis of the Mutual Activation Switch Motif

We now present an example of applying the CRN procedure and sensitivity analysis described in the beginning of this section. We return to the hysteretic switch resulting from the mutual activation motif (section 4.10), and use equation 4.34 to compute the sensitivity of both species R and Ep to heterogeneity in the Michaelis constant $K_{M,r}$. We use the same set of parameters as listed in Figure 4.11, except we use the following two marginally different initial populations of E with respect to $K_{M,r}$:

1. 50 agents of subspecies $E1$ have $K_{M,r} = 1$ and 50 $E2$ agents have $K_{M,r} = 19$. We will refer to simulations of this initial population structure as HET1 (initial index of heterogeneity, $\psi_{i,HET1} = 0.505\overline{0}$; computed using equation 4.1). Note that species E has the same population structure as the one we investigated in section 4.10.
2. 51 agents of subspecies $E1$ have $K_{M,r} = 1$ and 49 $E2$ agents have $K_{M,r} = 19$. We will refer to this as HET2 ($\psi_{i,HET2} = 0.5048$).

Recall that the initial population size $N_{E,i} = N_{E,TOT} = 100$, and that $N_E(t) + N_{Ep}(t) = N_{E,TOT} = 100$. Accordingly, we need only show the time trajectory of either species Ep or E .

In summary, the only difference between these two population structures is that a single $E2$ agent now belongs to subpopulation $E1$ in HET2, thus having a low $K_{M,r}$ value. Considering that the KCF of the i^{th} E agent is

$$\Omega_E^{(i)} = k_r \frac{N_R(t)}{K_{M,r}^{(i)} + N_E(t_n)}, \quad (4.35)$$

a lower $K_{M,r}$ value will result in a higher transition probability. Therefore, all of the ensuing results from this computational experiment are dependent on this differential propensity of an E agent to transition to Ep .

We first present a sample time trajectory of species R and Ep in Figure 4.18. We have used the same seed value for the random number generator to simulate the trajectories of the two cases, which start in the neighborhood of the OFF switch state (initial condition: $N_{R,i} = 0$, $N_{Ep,i} = 0$, $N_{E,i} = 100$) and begin deviating from each other at $t \approx 53$ sec. From that point forward, the trajectories differ significantly with respect to the switch state the system is in (OFF: low N_R , low N_{Ep} , high N_E ; ON: high N_R , high N_{Ep} , low N_E). This sample computation shows that a change in a parameter value of a single E agent can have a dramatic effect on the computed population time trajectories, especially in a system that is prone to noise-induced switch state reversals.

We simulated this motif with the HET1 and HET2 population structures 1000 times, as described above. This allowed us to compute the mean trajectories and their standard deviations for all species. Figure 4.19a shows that the average population sizes for both species R and Ep increase with time. This

Follow this [link](#) to see the Matlab code.

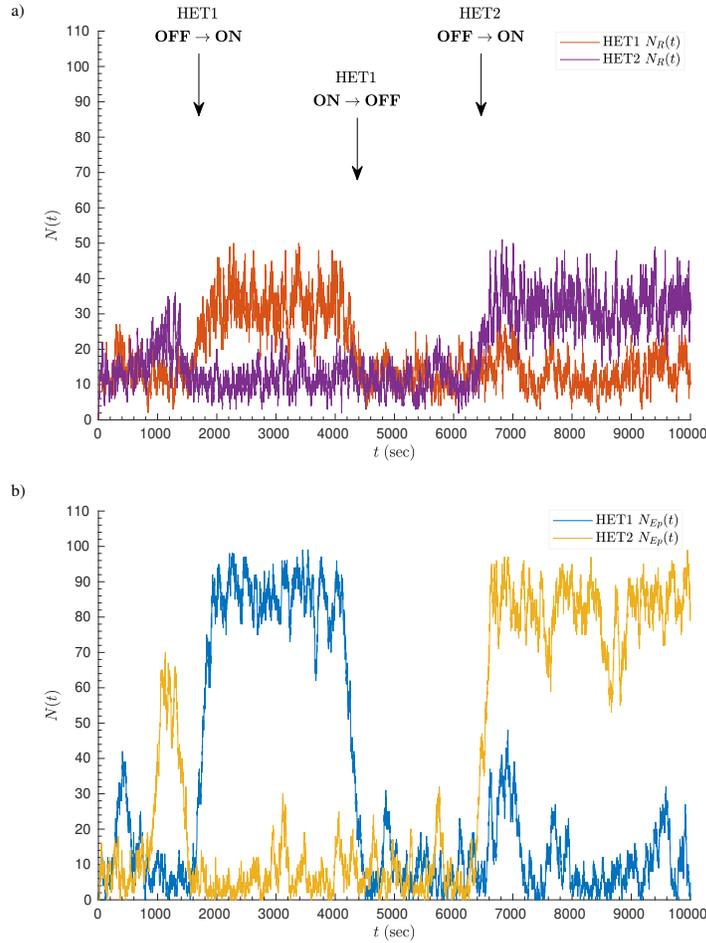


Figure 4.18 – Heterogeneity in the mutual activation switch motif: sample time trajectories using CRN. We simulated the hysteretic switch motif resulting from the R - E_p activation loop using the same parameters as indicated in section 4.10 (initial condition: $N_{R,i} = 0$, $N_{E_p,i} = 0$, $N_{E,i} = 100$, and $N_S = 15$), except here the simulated time is 0 – 10000 sec and we compare two different initial population structures of species E : HET1, 50 agents of subspecies $E1$ have $K_{M,r} = 1$ and 50 $E2$ agents have $K_{M,r} = 19$; and HET2, 51 agents of subspecies $E1$ have $K_{M,r} = 1$ and 49 $E2$ agents have $K_{M,r} = 19$. We used the CRN method by initializing the random number generator in both simulations with the same seed value (in this case, the seed value is 8; Matlab command: `rng(8)`). We observed that the trajectories of the same species from the two simulations begin deviating from each other at $t \approx 53$ sec. For clarity, we show the trajectories in two separate panels: **a)** species R , **b)** species E_p . We indicate in panel a) three noise-induced switch state reversals that occurred in the simulated time window. The simulations end up at different switch states at $t = 10000$ sec. Switch states: OFF, low N_R , low N_{E_p} (high N_E); ON, high N_R , high N_{E_p} (low N_E).

is an unexpected trend because at any given time point the system is found in one of the two switch states. This increase is indicative of the fact that the cumulative probability of a noise-induced switch state reversal (from initial state OFF to ON) increases with time, therefore an average over all simulation runs at a particular time point reflects the proportion of runs where the system is in the ON state. Moreover, the HET2 average trajectories are higher in value than the corresponding HET1 ones, suggesting a higher propensity for the system to be in the ON state in the HET2 case. This effect is more pronounced when comparing the $\langle N_{Ep}(t) \rangle$ trajectories because the species-specific difference in population sizes between the ON and OFF states is greater: $N_{Ep,OFF}^* = 5.35$, $N_{Ep,ON}^* = 86.69$ vs. $N_{R,OFF}^* = 11.43$, $N_{R,ON}^* = 33.12$ (these are the deterministically-predicted stable fixed points assuming homogeneous populations). That is, $N_{Ep,ON}^* - N_{Ep,OFF}^* > N_{R,ON}^* - N_{R,OFF}^*$. Accordingly, trajectories that have transitioned to the ON state will have a greater numerical affect on $\langle N_{Ep}(t) \rangle$ than $\langle N_R(t) \rangle$. Thus, the magnitude of the average trajectories is a reflection of the aggregate behavior and can be thought of as a proxy for the propensity of the system to be in the ON state.

We demonstrate that the average trajectories reveal the differential tendency to be in one of the switch states by showing the percentage of simulation runs that are in the ON state in Figure 4.20. It is clear that the HET2 case produces a greater proportion of ON states than HET1, with the difference being as high as 4% (see inset plot in Figure 4.20). We see therefore that the presence of a single E agent in case HET2 that can transition to Ep more easily results in significant differences with respect to the switch state the system is in.

We now proceed with performing sensitivity analysis on this motif's behavior with respect to heterogeneity. The difference between the values for the respective indices of heterogeneity is

$$\epsilon = \psi_{i,HET1} - \psi_{i,HET2} = 0.50\overline{50} - 0.50\overline{48} = 2.\overline{02} \times 10^{-4}. \quad (4.36)$$

We can then use equation 4.34 to compute the sensitivity of the average trajectories and their standard deviations (shown in Figures 4.19b and 4.21b, respectively). The sign of the sensitivity values for either species R or Ep is predominantly negative because HET2 produces higher-valued mean trajectories and standard deviations than HET1. Moreover, the magnitude of the sensitivity for species Ep is greater than for R , as explained in the preceding paragraphs.

The above results, obtained by performing computational experiments using the ABK algorithm, suggest that a single change in the structure of a population can have measurable effects on the system's simulated time evolution. Furthermore, our presentation of this motif as a case study on the use of sensitivity analysis with an ensemble of ABK simulations demonstrates that this method can provide useful insights on the role of heterogeneity on a system's dynamical properties.

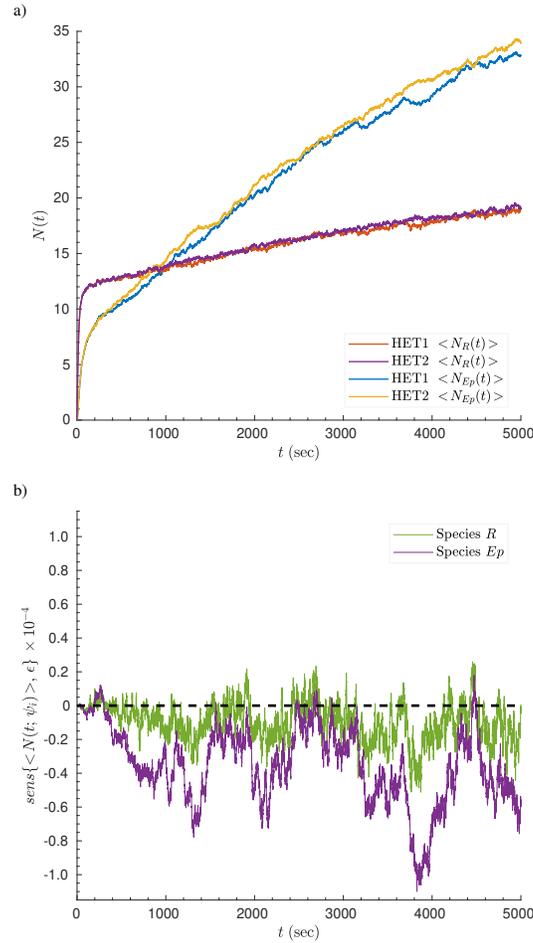


Figure 4.19 – Heterogeneity in the mutual activation switch motif: mean time trajectories and sensitivity analysis. We repeated the simulation of the mutual activation motif 1000 times (simulated time: 0 – 5000 sec). Each repetition was initialized with a different seed value for the random number generator (see Figure 4.18 for a sample simulation using the CRN method). This ensemble of simulation runs allowed us to calculate the average trajectories of all species given the two different initial population structures of E (HET1, HET2, as defined in the text and Figure 4.18). **a)** Mean trajectories for species R and Ep given HET1 and HET2. **b)** Sensitivity of species R and Ep given their respective average trajectories and the difference in the index of heterogeneity, $\epsilon = \psi_{i,HET1} - \psi_{i,HET2} = 2.02 \times 10^{-4}$ (computed using equation 4.34). The HET2 average trajectories are consistently higher than the HET1 ones, therefore the sensitivity values are predominantly negative.

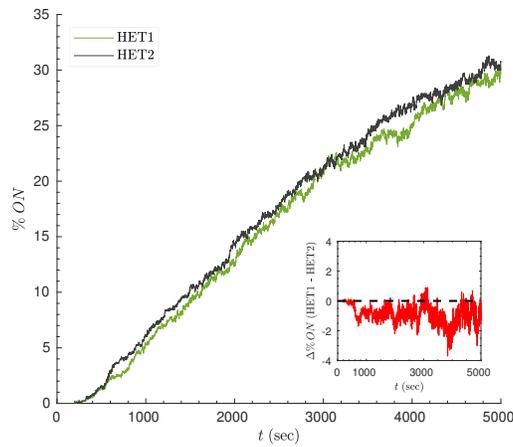


Figure 4.20 – Heterogeneity in the mutual activation switch motif: differential fraction of simulation runs in ON state. We plot the percentage of simulation runs ($n = 1000$) that were classified as being in the ON state at each time point in the simulations. All simulation parameters are the same as in Figure 4.19. We used the trajectories of species E_p for the classification because the difference between the deterministically-predicted fixed points is larger than for species R ($N_{E_p,OFF}^* = 5.35$, $N_{E_p,ON}^* = 86.69$). Specifically, we used the following heuristic for the classification: $|N_{E_p}(t) - N_{E_p,ON}^*| < \frac{1}{4} (N_{E_p,ON}^* - N_{E_p,OFF}^*)$. The difference in % ON between the HET1 and HET2 cases is statistically significant ($p = 2 \times 10^{-225}$ for the full time series; computed using the two-sample t -test). The inset plot shows more clearly the time dependence of the difference between the proportion of trajectories in the ON state: $\Delta\%ON \equiv \%ON_{HET1} - \%ON_{HET2}$.

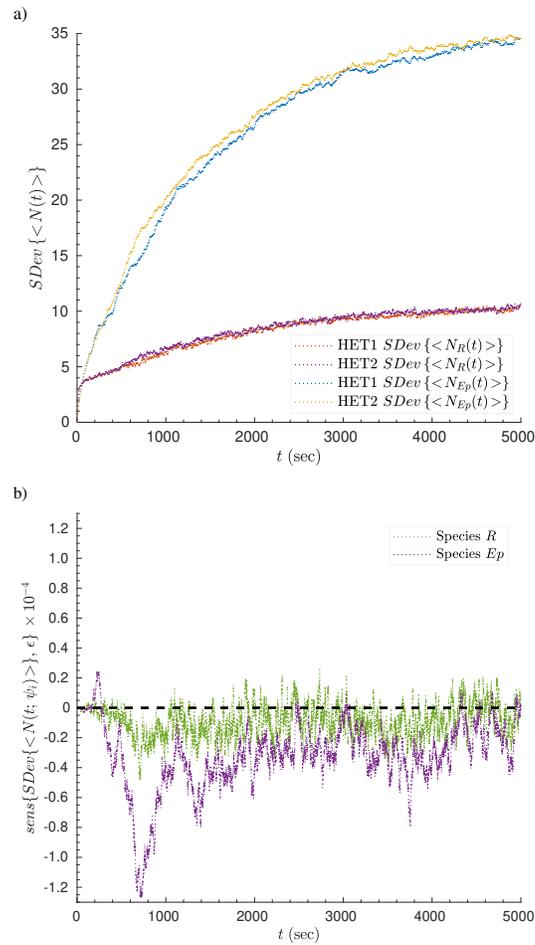


Figure 4.21 – Heterogeneity in the mutual activation switch motif: standard deviation of time trajectories and sensitivity analysis. All parameters and methods are the same as in Figure 4.19, except we plot the time dependence of the standard deviation from the ensemble of simulated trajectories and its sensitivity to ψ_i . The sensitivity of the standard deviation, $sens\{SDev\{\langle N_j(t; \psi_i) \rangle\}, \epsilon\}$, was computed using the general form in equation 4.34, where the mean was replaced by the standard deviation. Note that the species-specific color scheme is the same as in Figure 4.19.

Closing Remarks

In this work, we have considered populations as consisting of distinct individuals, or agents, whose kinetic properties can be explicitly specified and taken into account when simulating their trajectory in a system's state space. Using the algorithmic framework we developed in the preceding chapters, a modeler can perform computational experiments to investigate the effects of heterogeneity on a system's dynamics. Moreover, the assumption of a well-mixed system allows for an investigation of the mechanisms that give rise to or the consequences of heterogeneity in idealized scenarios where any potentially complicating factors (such as the presence of spatial barriers) are absent. Given the generality of our approach, we believe that the results we have presented in this work can be relevant to all fields where the notion of a (heterogeneous) population exists.

In summary, we have shown that ABK is a flexible and easily extensible method for simulating the time evolution of homogeneous or heterogeneous populations. Some of its advantages over other methods are: a) it can be easily implemented to account for both compositional and temporal heterogeneity in populations; b) it is useful in modeling a diverse array of phenomena spanning many fields within the biological and physical sciences; c) it is not limited to the simulation of up to bimolecular reactions/processes (higher order processes can be simulated, if necessary); d) it can be used to model agent interactions that can be represented as a directed or undirected graph. Moreover, although we have focused here on the binary representation of an agent's state, we will show in future work how the method can be easily extended to handle multi-state agents (Stefan et al., 2014) and the specification of rules governing state-dependent agent interactions.

The ABK method is undoubtedly computationally intensive, and the SSA (Gillespie's algorithm) is the preferred approach when homogeneity is assumed. Indeed, one should simulate all the processes and motifs we presented in Chapters 2 and 3 using the SSA if time and computational speed are of importance. Our aim in those chapters was to validate the ABK method for a large number of cases where we could compare its output to predictions obtained either through the deterministic formalism, CME, or SSA (this also allowed us to consider the limitations of the different approaches). Furthermore, we have used the method to simulate the basic processes and collections thereof that form a large part of a basis set for the functionality that higher order systems depend on. In the context of cell biology, for instance, we envision that researchers aiming to

simulate the behavior of a whole cell (Karr et al., 2015) could use ABK for cellular subsystems or modules where there is prior knowledge of and data on the presence of heterogeneity (e.g., subpopulations of a given protein’s isoforms in eukaryotic cells). Alternatively, an exploratory approach can be used to examine the effects of a distribution of parameter values across the agents of a species population on an observable property or properties of the system.

The Matlab code we have authored to perform the simulations presented in this work should be considered a prototype, and we are keenly aware of the fact that it can be improved to achieve better performance. For instance, some parts of the code can be vectorized to take advantage of libraries that are optimized for matrix manipulation and computation. Importantly, since our method relies on creating an ensemble of simulations to obtain statistical measures of a system’s dynamical behavior, parallelization of the executed code in multi-core processors and clusters will yield a significant boost in performance. We will work toward optimizing the ABK codebase and plan on coding a programming interface that will automate part of the simulation workflow, thus facilitating the method’s adoption and use by researchers. Part of this effort will require porting the code to a more widely used open-source language (such as Python) so that it can be freely and more easily accessed by interested users around the world.

Given the generality and wide applicability of the ABK methodology, we anticipate that it will be a useful modeling tool for probing the complex interplay between interspecies and inter-agent connectivity, stochasticity in low copy number populations, and compositional or temporal heterogeneity. It is our hope that the cases we have investigated in this work will inspire further computational studies and inform the design of experiments aimed at testing their predictions in real populations.

Appendix A

The Chemical Master Equation for Selected Processes

The *chemical master equation* (CME) describes the time-dependent probability distribution of the discrete population size of the species involved in a set of processes. As we will see by briefly studying a handful of case studies, the general form of a CME can be succinctly written as

$$\frac{\partial P_n(t)}{\partial t} = \sum_{n'} T(n|n')P_{n'}(t) - \sum_{n''} T(n''|n)P_n(t), \quad (\text{A.1})$$

where $P_n(t)$ is the probability¹ of the population size being n at time t , n' and n'' are the allowed alternate states of the system that transition to and from n respectively. $T(n|n')$ and $T(n''|n)$ are the conditional transition probabilities of going from a given state to another. The sum runs over all possible population sizes n' and n'' that are related to n according to the stoichiometry of the defined processes. Essentially, the first term in the above sum represents the processes through which the population can end up having size n (i.e., flux entering state n), while the second term shows how a population of size n can transition to other states (flux exiting state n).

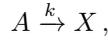
It is evident that the CME is really an infinite set of coupled equations, since it can be written for each of the possible combinations of n and n' or n'' . Hence, it cannot be solved analytically in all but the simplest of cases. Here, we present how to construct the CME of 0th, 1st, 2nd order, and birth-death processes, and use it to obtain statistical measures of the population time trajectories. Additionally, we will explore how the predictions of the CME

¹This is really a conditional probability that can be fully written as $P(n, t|n_0, t_0)$, reflecting the fact that the system is subject to initial population sizes at time t_0 . We use a simplified notation here to symbolize this probability.

relate to deterministic ones. However, we will not discuss in detail how the solution to the CME is obtained in the cases we will examine. The interested reader can consult McQuarrie (1967) or Lecca et al. (2013b) and the references therein for further study. We also point the reader to a rigorous derivation of the CME (Gillespie, 1992) from a minimal set of premises whose origin lies in microphysical considerations and elementary concepts of probability theory.

A.1 1st Order Process

We will now derive the CME as applied to the 1st order process



where k is the rate constant with units sec^{-1} . For notational simplicity and readability since there is only one species to keep track of, we will denote the population size of A as n .

Let $P_n(t)$ denote the probability of the system consisting of n agents of A at time t . Given knowledge of the population size at time t , we seek an expression for the probability $P_n(t + dt)$, where dt is a sufficiently small time interval such that only one occurrence of the process takes place.

There are two distinct ways in which the number of agents will be n at $t + dt$: i) there were $n + 1$ agents at time t and a transition event takes place in dt , and ii) there were n agents at time t and no occurrence of the process takes place in dt . The probability $P_n(t + dt)$ can then be written as

$$P_n(t + dt) = k(n + 1) dt P_{n+1}(t) + (1 - k n dt) P_n(t). \quad (\text{A.2})$$

Notice that the first term of the above expression is scaled by the rate constant, a statistical factor of the number of agents present at time t , and the time interval dt . The statistical factor signifies that there are $n + 1$ ways in which the said process can occur when the population size at time t is $n + 1$. Therefore, the overall factor that $P_{n+1}(t)$ is multiplied by is $k(n + 1) dt$, which can be thought of as the transition probability for the population size being reduced from $n + 1$ to n . In the second term on the right-hand side of the above equation, $1 - k n dt$ is the probability that the process does not occur given that n agents exist at time t .

We proceed by rearranging the above equation and dividing both sides by dt to obtain

$$\frac{P_n(t + dt) - P_n(t)}{dt} = k(n + 1)P_{n+1}(t) - k n P_n(t). \quad (\text{A.3})$$

Taking the limit as $dt \rightarrow 0$, we see that the left-hand side of this equation is the partial derivative of P with respect to time,

$$\frac{\partial P_n}{\partial t} = k[(n + 1)P_{n+1} - nP_n]. \quad (\text{A.4})$$

Note that we dropped the notation indicating the explicit dependence on time to minimize notational clutter. This last result is the CME for a 1st order process and it fits the general form of the CME (equation A.1).

The steady-state solution is $n^* = 0$ (i.e., the population of A is depleted). The reader can ascertain using equation A.4 that the probability of this state is $P_0 = 1$.

A.1.1 Relationship to Deterministic Prediction

We remind the reader that the ODE for a 1st order process is

$$\frac{dn}{dt} = -kn, \quad (\text{A.5})$$

and its solution given the initial condition $n(0) = n_0$ is

$$n(t) = n_0 e^{-kt}. \quad (\text{A.6})$$

We explore the relationship between the probabilistic and deterministic approaches by multiplying both sides of the CME (equation A.28) by n and then summing over all n .

$$\frac{\partial \left(\sum_{n=0}^{\infty} n P_n \right)}{\partial t} = k \left[\sum_{n=0}^{\infty} n(n+1) P_{n+1} - \sum_{n=0}^{\infty} n^2 P_n \right] \quad (\text{A.7})$$

Notice that the sum runs from $n = 0$ to infinity, which suggests that there are an infinite number of possible integer-valued population sizes. However, it must be true that $P_n(t) = 0$ when $n > n_0$ since this is a degradation process (i.e., there cannot be more than the initial number of A agents). Therefore, there is a finite number of values for n and we change the upper limit of the above sums to reflect the fact that $P_{n>n_0} = 0$.

One way to deal with the expression on the right-hand side of equation A.7 is to write out each of the summation terms.

$$\sum_{n=0}^{n_0-1} n(n+1) P_{n+1} = 2P_2 + 6P_3 + 12P_4 + \dots + (n_0 - 1)n_0 P_{n_0} \quad (\text{A.8})$$

$$\sum_{n=0}^{n_0} n^2 P_n = P_1 + 4P_2 + 9P_3 + 16P_4 + \dots + n_0^2 P_{n_0} \quad (\text{A.9})$$

Subtraction leads to

$$\begin{aligned} \sum_{n=0}^{n_0-1} n(n+1) P_{n+1} - \sum_{n=0}^{n_0} n^2 P_n &= -P_1 - 2P_2 - 3P_3 - 4P_4 - \dots - n_0 P_{n_0} \\ &= - \sum_{n=0}^{n_0} n P_n. \end{aligned} \quad (\text{A.10})$$

Substituting these results into equation A.7, we get

$$\frac{\partial \left(\sum_{n=0}^{n_0} n P_n \right)}{\partial t} = k \left(- \sum_{n=0}^{n_0} n P_n \right). \quad (\text{A.11})$$

Recognizing that $\sum_{n=0}^{n_0} n P_n = \langle n \rangle$, the above equation simplifies to

$$\frac{d \langle n \rangle}{dt} = -k \langle n \rangle. \quad (\text{A.12})$$

This is the same as the ODE for the deterministic formulation of the process (equation A.6). This means that the *average* time trajectory of the population of A is the same as deterministically predicted. Therefore, given the initial condition $\langle n(0) \rangle = n_0$, the explicit time dependence of $\langle n(t) \rangle$ is:

$$\langle n(t) \rangle = n_0 e^{-kt}. \quad (\text{A.13})$$

This result brings attention to the fact that the continuous deterministic approach to chemical kinetics, as empirically observed in the thermodynamic limit of large populations, is a consequence of the stochastic formulation; not the other way around.

A.1.2 Determining the Variance from the CME

We aim to find the time dependence of the variance given the CME for a 1st order process. We begin by multiplying both sides of the CME (equation A.4) by n^2 , and then summing over all n .

$$\frac{\partial \left(\sum_{n=0}^{\infty} n^2 P_n \right)}{\partial t} = k \left[\sum_{n=0}^{\infty} n^2 (n+1) P_{n+1} - \sum_{n=0}^{\infty} n^3 P_n \right] \quad (\text{A.14})$$

We again proceed by writing out the summation terms while keeping in mind that $P_{n > n_0} = 0$,

$$\sum_{n=0}^{n_0-1} n^2 (n+1) P_{n+1} = 2P_2 + 12P_3 + \dots + (n_0 - 1)^2 n_0 P_{n_0} \quad (\text{A.15})$$

$$\sum_{n=0}^{n_0} n^3 P_n = P_1 + 8P_2 + 27P_3 + \dots + n_0^3 P_{n_0}, \quad (\text{A.16})$$

which upon subtraction yields

$$\begin{aligned} \sum_{n=0}^{n_0-1} n^2 (n+1) P_{n+1} - \sum_{n=0}^{n_0} n^3 P_n &= -P_1 - 6P_2 - 15P_3 - \dots - (2n_0 - 1)n_0 P_{n_0} \\ &= - \sum_{n=0}^{n_0} (2n - 1)n P_n. \end{aligned} \quad (\text{A.17})$$

We can now substitute A.17 into equation A.14.

$$\frac{\partial \left(\sum_{n=0}^{n_0} n^2 P_n \right)}{\partial t} = k \left[- \sum_{n=0}^{n_0} (2n-1) n P_n \right] \quad (\text{A.18})$$

$$= -2k \left(\sum_{n=0}^{n_0} n^2 P_n \right) + k \left(\sum_{n=0}^{n_0} n P_n \right) \quad (\text{A.19})$$

Recognizing that $\sum_{n=0}^{n_0} n P_n = \langle n \rangle$ and $\sum_{n=0}^{n_0} n^2 P_n = \langle n^2 \rangle$, the above equation is equivalent to

$$\frac{d \langle n^2 \rangle}{dt} = -2k \langle n^2 \rangle + k \langle n \rangle. \quad (\text{A.20})$$

This is a 1st order linear ODE and can be solved by using the integrating factor method.² Given the initial condition $\langle n^2(0) \rangle = n_0^2$, its solution is

$$\langle n^2(t) \rangle = n_0 e^{-kt} [1 + (n_0 - 1) e^{-kt}]. \quad (\text{A.21})$$

We can now substitute the result from equation A.21 into the definition of variance.

$$\sigma^2 = \langle n^2(t) \rangle - \langle n(t) \rangle^2 \quad (\text{A.22})$$

$$\begin{aligned} &= n_0 e^{-kt} [1 + (n_0 - 1) e^{-kt}] - (n_0 e^{-kt})^2 \\ &= n_0 e^{-kt} (1 - e^{-kt}) \end{aligned} \quad (\text{A.23})$$

We have used this expression to calculate the standard deviation of a 1st order process and to compare it to the results of our ABK simulations (see Figure 2.3).

A.1.3 CME Solution

We have thus far derived the first and second moments of the probability distribution $P_n(t)$ without access to an explicit solution to the CME. One way to solve the CME is through the method of generating functions. We will merely state the result of the method's application here. The solution is

$$P_n(t) = \frac{n_0!}{n!(n_0 - n)!} (e^{-kt})^n (1 - e^{-kt})^{n_0 - n}. \quad (\text{A.24})$$

This is clearly the binomial distribution with $p = e^{-kt}$. Since this distribution has a mean of $\langle n(t) \rangle = n_0 p$ and a variance of $\sigma^2 = n_0 p(1 - p)$, we obtain the same results as in equations A.13 and A.23.

²The interested reader can try solving the ODE with the integrating factor $\mu(t) = e^{2kt}$.

A.2 Birth-Death and 0th Order Processes

We present here a derivation of the CME as applied to the birth-death process



where k_b and k_d are the rate constants for a 0th and 1st order process, respectively. We will also present relevant results for a 0th order process, which is a special case of the above reaction scheme when $k_d = 0$. As we did in the preceding section, we will denote the population size of A as n .

Let $P_n(t)$ denote the probability of the system consisting of n agents of A at time t . Given knowledge of the population size at t , we aim to find an expression for the probability $P_n(t + dt)$, where dt is a sufficiently small time interval such that at most one occurrence of either of the above processes takes place. There are three ways in which the number of agents will be n at $t + dt$: i) there were $n - 1$ agents at time t and the 0th order process (birth) takes place in dt ; ii) there were $n + 1$ agents at time t and the 1st order process (death or degradation) takes place in dt ; iii) there were n agents at time t and neither process takes place in dt . This leads to the expression

$$P_n(t + dt) = k_b dt P_{n-1}(t) + k_d (n + 1) dt P_{n+1}(t) + (1 - k_b dt - k_d n dt) P_n(t). \quad (\text{A.26})$$

As we noted earlier in this Appendix when deriving the CME for a 1st order process, each of the above three contributions to the expression is scaled by the appropriate rate constant, time interval dt , and a statistical factor of the number of agents present at time t . For instance, the birth of an additional agent of A can occur in only one way, hence the factor is 1 and $P_{n-1}(t)$ is scaled by $k_b dt$. Similarly, the degradation of A can occur in $n + 1$ ways when the population size at time t is $n + 1$, and the overall factor is $k_d (n + 1) dt$. In the last term of equation A.26, the expression in parentheses, $1 - k_b dt - k_d n dt$, is the probability that neither reaction takes place in dt given n agents at time t .

We proceed by rearranging the above equation and dividing both sides by dt , to obtain

$$\frac{P_n(t + dt) - P_n(t)}{dt} = k_b P_{n-1}(t) + k_d (n + 1) P_{n+1}(t) - k_b P_n(t) - k_d n P_n(t). \quad (\text{A.27})$$

Taking the limit as $dt \rightarrow 0$, we see that the left side of this equation is clearly the partial derivative of P with respect to time,

$$\frac{\partial P_n}{\partial t} = k_b (P_{n-1} - P_n) + k_d [(n + 1) P_{n+1} - n P_n], \quad (\text{A.28})$$

where we omitted explicit mention of time dependence to minimize notational clutter. This result is the CME for a birth-death process.

CME for a 0th order process. The above arguments can be repeated when describing the birth of A in the absence of a degradation process. Setting $k_d = 0$

in equation A.28 gives the CME for a 0th order process,

$$\frac{\partial P_n}{\partial t} = k_b (P_{n-1} - P_n). \quad (\text{A.29})$$

A.2.1 Steady-State Solution

We now try to solve for the steady-state solution to the CME by setting $\partial P_n / \partial t = 0$.

$$k_b P_{n-1} + k_d (n+1) P_{n+1} = k_b P_n + k_d n P_n \quad (\text{A.30})$$

$$(n+1) P_{n+1} = \frac{k_b}{k_d} (P_n - P_{n-1}) + n P_n \quad (\text{A.31})$$

This last equation suggests that we can obtain a solution recursively. Assuming $P_{-1} = 0$, we initially set $n = 0$ to obtain:

$$P_1 = \frac{k_b}{k_d} P_0. \quad (\text{A.32})$$

Continuing this process and making successive substitutions to obtain solutions in terms of P_0 , we get

$$P_2 = \frac{1}{2} \left(\frac{k_b}{k_d} \right)^2 P_0 \quad (\text{A.33})$$

$$P_3 = \frac{1}{3!} \left(\frac{k_b}{k_d} \right)^3 P_0 \quad (\text{A.34})$$

⋮

$$P_n = \frac{1}{n!} \left(\frac{k_b}{k_d} \right)^n P_0. \quad (\text{A.35})$$

We can solve for P_0 by normalizing the distribution.

$$\sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{k_b}{k_d} \right)^n P_0 = P_0 \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{k_b}{k_d} \right)^n = P_0 e^{\frac{k_b}{k_d}} = 1 \quad (\text{A.36})$$

Therefore,

$$P_0 = e^{-\frac{k_b}{k_d}}, \quad (\text{A.37})$$

and the solution for P_n is

$$P_n = \frac{1}{n!} \left(\frac{k_b}{k_d} \right)^n e^{-\frac{k_b}{k_d}}. \quad (\text{A.38})$$

This result reveals that P_n is a Poisson distribution with an average rate of k_b/k_d .

A.2.2 Relationship to Deterministic Prediction

Interestingly, the probabilistic treatment we have presented yields a nonzero finite probability of species A becoming extinct: $P_0 = e^{-k_b/k_d}$. This is clearly different from the deterministic formulation of this birth-death process, where $n(t)$ changes monotonically and asymptotically approaches the steady state $n^* = k_b/k_d$.³ As a brief reminder, the ODE for a 1st order process is

$$\frac{dn}{dt} = k_b - k_d n, \quad (\text{A.39})$$

and its solution, given the initial condition $n(0) = 0$, is

$$n(t) = \frac{k_b}{k_d} (1 - e^{-k_d t}). \quad (\text{A.40})$$

Moreover, we see in the probabilistic formulation that as a result of P_n being Poissonian, the *average* steady-state value is $\langle n^* \rangle = k_b/k_d$. This is the same result as in the deterministic case.

We can further explore the relationship between the probabilistic and deterministic approaches by multiplying both sides of the CME (equation A.28) by n and then summing over all n .

$$\frac{\partial \left(\sum_{n=0}^{\infty} n P_n \right)}{\partial t} = k_b \left[\sum_{n=0}^{\infty} n P_{n-1} - \sum_{n=0}^{\infty} n P_n \right] + k_d \left[\sum_{n=0}^{\infty} n(n+1) P_{n+1} - \sum_{n=0}^{\infty} n^2 P_n \right] \quad (\text{A.41})$$

It is straightforward to determine the result of each of the bracketed groups of terms on the right-hand side of the above equation. Starting with the group multiplied by k_b , we see that

$$\sum_{n=0}^{\infty} n P_{n-1} = P_0 + 2P_1 + 3P_2 + 4P_3 + \dots \quad (\text{A.42})$$

$$\sum_{n=0}^{\infty} n P_n = P_1 + 2P_2 + 3P_3 + \dots, \quad (\text{A.43})$$

such that a subtraction of these two expressions yields

$$\begin{aligned} \sum_{n=0}^{\infty} n P_{n-1} - \sum_{n=0}^{\infty} n P_n &= P_0 + P_1 + P_2 + P_3 + \dots \\ &= \sum_{n=0}^{\infty} P_n. \end{aligned} \quad (\text{A.44})$$

³If $k_b \geq k_d$, then $n(t)$ increases monotonically for $n(0) = 0$ (or more generally, for $n(0) < n^*$). Similarly, if $k_b < k_d$, then $n(t)$ decreases monotonically for nonzero and positive values of $n(0)$.

The group of terms multiplied by k_d in A.41 is one we have previously examined in equation A.7. We determined the result to be

$$\sum_{n=0}^{\infty} n(n+1)P_{n+1} - \sum_{n=0}^{\infty} n^2 P_n = - \sum_{n=0}^{\infty} n P_n .$$

Substituting these results into equation A.41, we get

$$\frac{\partial \left(\sum_{n=0}^{\infty} n P_n \right)}{\partial t} = k_b \left(\sum_{n=0}^{\infty} P_n \right) + k_d \left(- \sum_{n=0}^{\infty} n P_n \right) . \quad (\text{A.45})$$

Finally, recognizing that $\sum_{n=0}^{\infty} P_n = 1$ and $\sum_{n=0}^{\infty} n P_n = \langle n \rangle$, the above equation is rewritten as

$$\frac{d \langle n \rangle}{dt} = k_b - k_d \langle n \rangle . \quad (\text{A.46})$$

This is the same as the ODE for the deterministic formulation of the process (equation A.39). This means that the *average* time trajectory of the population of A is the same as deterministically predicted. Accordingly, given the initial condition $\langle n(0) \rangle = 0$, the solution is

$$\langle n(t) \rangle = \frac{k_b}{k_d} (1 - e^{-k_d t}) . \quad (\text{A.47})$$

Corresponding results for a 0th order process.

The reader can verify that the above method can be applied to obtain the first moment of $n(t)$ for a birth process. Alternatively, we can use the result of equation A.46 and set $k_d = 0$, leading to

$$\frac{d \langle n \rangle}{dt} = k_b . \quad (\text{A.48})$$

Notice that the above differential equation has the same form as its deterministic counterpart (equation 2.107). We can easily obtain the solution,

$$\langle n(t) \rangle = k_b t , \quad (\text{A.49})$$

where we have used initial condition $\langle n(0) \rangle = 0$. Therefore, the CME predicts an *average* time trajectory for the population of A that is identical to that of the deterministic prediction.

A.2.3 Determining the Variance from the CME

We aim to find the time dependence of the variance for the birth-death process. We begin by multiplying both sides of the CME (equation A.28) by n^2 and then

summing over all n .

$$\frac{\partial \left(\sum_{n=0}^{\infty} n^2 P_n \right)}{\partial t} = k_b \left[\sum_{n=0}^{\infty} n^2 P_{n-1} - \sum_{n=0}^{\infty} n^2 P_n \right] + k_d \left[\sum_{n=0}^{\infty} n^2 (n+1) P_{n+1} - \sum_{n=0}^{\infty} n^3 P_n \right] \quad (\text{A.50})$$

We proceed as previously by analyzing each of the two groups of terms in the above equation.

$$\sum_{n=0}^{\infty} n^2 P_{n-1} = P_0 + 4P_1 + 9P_2 + 16P_3 + 25P_4 + \dots \quad (\text{A.51})$$

$$\sum_{n=0}^{\infty} n^2 P_n = P_1 + 4P_2 + 9P_3 + 16P_4 + \dots \quad (\text{A.52})$$

Subtraction of the above sums leads to

$$\begin{aligned} \sum_{n=0}^{\infty} n^2 P_{n-1} - \sum_{n=0}^{\infty} n^2 P_n &= P_0 + 3P_1 + 5P_2 + 7P_3 + 9P_4 + \dots \\ &= \sum_{n=0}^{\infty} (2n+1) P_n. \end{aligned} \quad (\text{A.53})$$

We encountered the second group of terms in equation A.50 when we presented the CME for a 1st order process (see equation A.14). For convenience, we rewrite the result (equation ??) below.

$$\sum_{n=0}^{\infty} n^2 (n+1) P_{n+1} - \sum_{n=0}^{\infty} n^3 P_n = - \sum_{n=0}^{\infty} (2n-1) n P_n$$

We now substitute the simplified forms of the two groups of terms into equation A.50.

$$\frac{\partial \left(\sum_{n=0}^{\infty} n^2 P_n \right)}{\partial t} = k_b \left[\sum_{n=0}^{\infty} (2n+1) P_n \right] + k_d \left[- \sum_{n=0}^{\infty} (2n-1) n P_n \right] \quad (\text{A.54})$$

We can then rewrite this expression as

$$\frac{\partial \left(\sum_{n=0}^{\infty} n^2 P_n \right)}{\partial t} = 2k_b \left(\sum_{n=0}^{\infty} n P_n \right) + k_b \left(\sum_{n=0}^{\infty} P_n \right) - 2k_d \left(\sum_{n=0}^{\infty} n^2 P_n \right) + k_d \left(\sum_{n=0}^{\infty} n P_n \right). \quad (\text{A.55})$$

Recognizing that $\sum_{n=0}^{\infty} P_n = 1$, $\sum_{n=0}^{\infty} nP_n = \langle n \rangle$, and the second raw moment is $\sum_{n=0}^{\infty} n^2 P_n = \langle n^2 \rangle$, the above equation becomes

$$\frac{d \langle n^2 \rangle}{dt} = 2k_b \langle n \rangle + k_b - 2k_d \langle n^2 \rangle + k_d \langle n \rangle. \quad (\text{A.56})$$

Finally, we rearrange the terms of the above result and substitute for $\langle n(t) \rangle$ (equation A.47). The differential equation for $\langle n^2 \rangle$ then is

$$\frac{d \langle n^2 \rangle}{dt} = k_b + \frac{k_b}{k_d} (1 - e^{-kat}) (2k_b + k_d) - 2k_d \langle n^2 \rangle. \quad (\text{A.57})$$

This is a 1st order linear ODE and can be solved by using the integrating factor method.⁴ Given the initial condition $\langle n^2(0) \rangle = 0$, the solution is

$$\langle n^2(t) \rangle = \left[\frac{k_b}{k_d} (1 - e^{-kat}) \right]^2 + \frac{k_b}{k_d} (1 - e^{-kat}). \quad (\text{A.58})$$

Reminding ourselves the solution to $\langle n(t) \rangle$ in equation A.47, our derived expression for $\langle n^2(t) \rangle$ can be equivalently written as

$$\langle n^2(t) \rangle = \langle n(t) \rangle^2 + \langle n(t) \rangle. \quad (\text{A.59})$$

Finally, we substitute the above result into the definition of variance to obtain

$$\begin{aligned} \sigma^2 &= \langle n^2(t) \rangle - \langle n(t) \rangle^2 \\ &= \langle n(t) \rangle^2 + \langle n(t) \rangle - \langle n(t) \rangle^2 \\ &= \langle n(t) \rangle. \end{aligned} \quad (\text{A.60})$$

The CME predicts that the variance has the same value as the average time trajectory of the population size. The standard deviation then has the form

$$\sigma = \sqrt{\langle n(t) \rangle} \quad (\text{A.61})$$

$$= \sqrt{\frac{k_b}{k_d} (1 - e^{-kat})}. \quad (\text{A.62})$$

A comparison of the standard deviation obtained through the CME and an ensemble of ABK simulations is shown in Figure 2.15.

Variance of a 0th order process. To obtain an expression for the variance and standard deviation of a birth process, we begin with equation A.56. Setting $k_d = 0$ and replacing $\langle n(t) \rangle$ with the result of equation A.49, we get

$$\frac{d \langle n^2(t) \rangle}{dt} = 2k_b \langle n(t) \rangle + k_b \quad (\text{A.63})$$

$$= 2k_b(k_b t) + k_b \quad (\text{A.64})$$

$$= 2k_b^2 t + k_b. \quad (\text{A.65})$$

⁴The interested reader can try solving the ODE with the integrating factor $\mu(t) = e^{2k_d t}$.

This differential equation, with initial condition $\langle n^2(0) \rangle = 0$, can be easily solved to obtain

$$\begin{aligned}\langle n^2(t) \rangle &= (k_b t)^2 + k_b t \\ &= \langle n(t) \rangle^2 + \langle n(t) \rangle.\end{aligned}\tag{A.66}$$

Finally, subtracting $\langle n(t) \rangle^2$ from both sides yields the variance,

$$\begin{aligned}\sigma^2 &= \langle n^2(t) \rangle - \langle n(t) \rangle^2 \\ &= \langle n(t) \rangle,\end{aligned}\tag{A.67}$$

and standard deviation,

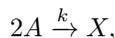
$$\sigma = \sqrt{\langle n(t) \rangle}\tag{A.68}$$

$$= \sqrt{k_b t}.\tag{A.69}$$

Note that the variance and standard deviation appear to have the same dependence on the average time trajectory of the population size as for a birth-death process (equations A.60 and A.61). However, their final algebraic form is different because $\langle n(t) \rangle = k_b t$ for a 0th order process. A comparison of the standard deviation obtained through the CME and an ensemble of ABK simulations is shown in Figure 2.13.

A.3 Homologous 2nd Order Process

We now move to the 2nd order process



with k being the microscopic rate constant for this process. We begin by writing the contributions to $P_n(t + dt)$.

$$P_n(t + dt) = k \frac{1}{2}(n+2)(n+1) dt P_{n+2} + \left[1 - k \frac{1}{2}n(n-1) dt \right] P_n \quad (\text{A.70})$$

The factor $\frac{1}{2}(n+2)(n+1)$ refers to the number of distinct pairs of A molecules, assuming they are indistinguishable from each other (hence the $1/2$ term). The same reasoning applies to the term in the factor in brackets. Rearranging, dividing by dt , and taking the limit at $dt \rightarrow 0$, we obtain the CME,

$$\frac{\partial P_n}{\partial t} = \frac{1}{2} k [(n+2)(n+1)P_{n+2} - n(n-1)P_n]. \quad (\text{A.71})$$

Solving the above equation to obtain the general solution, $P_n(t)$, is beyond the scope of our presentation in this Appendix (see McQuarrie, 1967, for more details).

A.3.1 Relationship to Deterministic Prediction

We begin by multiplying both sides of the CME by n and summing over all n .

$$\frac{\partial \left(\sum_{n=0}^{\infty} n P_n \right)}{\partial t} = \frac{1}{2} k \left[\sum_{n=0}^{\infty} n(n+2)(n+1)P_{n+2} - \sum_{n=0}^{\infty} n^2(n-1)P_n \right] \quad (\text{A.72})$$

We use an alternate and faster method in finding the result of the expression inside the brackets. We shift the index of the first term such that $n \mapsto n-2$.

$$\sum_{n=0}^{\infty} n(n+2)(n+1)P_{n+2} \mapsto \sum_{n=2}^{\infty} (n-2)n(n-1)P_n \quad (\text{A.73})$$

The summation index now starts at 2, but we can change that back to 0 because the terms in the new sum that correspond to $n=0$ and $n=1$ vanish anyway. The reader can assert that the new form of the sum generates the same series.

This means that we can rewrite equation A.72 as

$$\begin{aligned}
\frac{\partial \left(\sum_{n=0}^{\infty} n P_n \right)}{\partial t} &= \frac{1}{2} k \left[\sum_{n=0}^{\infty} n(n-1)(n-2) P_n - \sum_{n=0}^{\infty} n^2(n-1) P_n \right] \\
&= \frac{1}{2} k \left[\sum_{n=0}^{\infty} (n(n-1)(n-2) - n^2(n-1)) P_n \right] \\
&= \frac{1}{2} k \left[-2 \sum_{n=0}^{\infty} n(n-1) P_n \right]. \tag{A.74}
\end{aligned}$$

Finally, since $\sum_{n=0}^{\infty} n P_n = \langle n \rangle$ and the second factorial moment is $\sum_{n=0}^{\infty} n(n-1) P_n = \langle n(n-1) \rangle$, we get

$$\frac{d \langle n \rangle}{dt} = -k \langle n(n-1) \rangle, \tag{A.75}$$

which is reminiscent of equation 2.62. The CME leads to a differential rate law that is related to the agent-based interpretation of the process. This highlights the relationship between the deterministic and stochastic (CME) approaches: the ODE formalism is an exact mean-field predictor of population trajectories only in cases where the involved processes are linear (as we demonstrated for 1st and 0th order processes; see equations A.12 and A.46). The presence of any nonlinearity results in a discrepancy between the two approaches.

Notice that equation A.74 can be rewritten by distributing n inside the summation term, in which case $d \langle n \rangle / dt$ has an explicit dependence on $\langle n^2 \rangle$. That is,

$$\frac{d \langle n \rangle}{dt} = -k [\langle n^2 \rangle - \langle n \rangle]. \tag{A.76}$$

This equation cannot be solved unless we obtain an expression for the 2nd order raw moment by determining $d \langle n^2 \rangle / dt$. Using the same methods as we previously have, it can be shown that

$$\frac{d \langle n^2 \rangle}{dt} = -k [2 \langle n^3 \rangle - 4 \langle n^2 \rangle + 2 \langle n \rangle]. \tag{A.77}$$

Clearly, a dependence on a higher order raw moment appears inevitable. This is a paradigmatic case for the problem of finding a closed-form solution to the CME's moments: lower order moments depend on higher order ones in *non-linear systems*. Therefore, the set of ODEs for describing the distribution's raw moments becomes infinite. One way to resolve this issue is to assume that higher order moments are a product of lower ones (McQuarrie, 1967, p. 441). For instance, if we assume that $\langle n^2 \rangle = \langle n \rangle^2$, then the derivative of the first moment (equation A.76) becomes

$$\frac{d \langle n \rangle}{dt} = -k \langle n \rangle [\langle n \rangle - 1], \tag{A.78}$$

which is the same rate law as the deterministic agent-based interpretation of the process (equation 2.62).

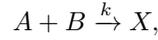
Furthermore, given the above assumption that $\langle n^2 \rangle = \langle n \rangle^2$, it becomes immediately clear that

$$\sigma^2 = \langle n^2 \rangle - \langle n \rangle^2 = 0. \quad (\text{A.79})$$

This means that insisting on the condition $\langle n^2 \rangle = \langle n \rangle^2$ essentially turns the probabilistic approach into a deterministic one, which obviates the primary purpose of the probabilistic formalism to obtain a measure of the fluctuations about the mean time trajectory. In any case, we will not discuss this issue any further here, other than to point out that resolving these problems remains an active area of research. For instance, Smadbeck and Kaznessis (2014; 2013) have proposed a CME closure scheme for dealing with this situation.

A.4 Heterologous 2nd Order Process

We consider the 2nd order heterologous process



where B is limiting (i.e., the initial populations are related by $N_{B,i} < N_{A,i}$) and k is the microscopic rate constant for this process. We define the parameter $Z = N_{A,i} - N_{B,i} = N_A(t) - N_B(t)$, and choose to represent the system state by the limiting population size $n = N_B(t)$. The CME then is

$$\frac{\partial P_n}{\partial t} = k[(n+1)(Z+n+1)P_{n+1} - n(Z+n)P_n], \quad (\text{A.80})$$

which can be (tediously) solved by the method of generating functions or Laplace transforms. We merely present the result here.

$$P_n(t) = \frac{N_{A,i}!N_{B,i}!}{(Z+n)!n!} \sum_{s=n}^{N_{B,i}} (-1)^{s-n} \frac{(Z+2s)(Z+n+s-1)!}{(N_{B,i}-s)!(s-n)!(N_{A,i}+s)!} e^{-s(Z+s)kt} \quad (\text{A.81})$$

This is not one of the common probability distributions that one may encounter in science. It is therefore best to numerically calculate $\langle n \rangle$ and σ^2 . We have indeed done this using Matlab and plot the results in Figures 2.11 and 2.12.

Appendix B

The Stochastic Simulation Algorithm (SSA)

A nonspatial population-based method for describing the time evolution of chemically reacting systems was developed by Gillespie (1977), and is applied to model the behavior of a discrete number of molecules in a well-stirred container. We hereby briefly explain how the algorithm works.

Consider discrete events (such as a chemical reaction) occurring at *constant* rate r_o over time T , so that the average number of events is $r_o T$. We begin by assuming this to be a homogeneous Poisson process. The justification for this is the following: if we divide T into s non-overlapping subintervals in each of which an event occurrence can be considered a binary, or Bernoulli, process (an event occurs or it does not), then the number of events in time T is binomially distributed. It can be shown that for large s the Poisson distribution emerges from the binomial distribution (the proof is short and can be found in any textbook on probability theory). Therefore, the probability of n events occurring within time T is

$$P(n) = \frac{(r_o T)^n}{n!} e^{-r_o T}. \quad (\text{B.1})$$

The time distribution between successive events can be analyzed by first supposing that an event occurred at time t . Then, the probability $P(\tau)$ that the next event takes place between $t + \tau$ and $t + \tau + \Delta\tau$ is

$$P(\tau) = P(\text{no event between } t \text{ and } t+\tau) \cdot P(1 \text{ event between } t+\tau \text{ and } t+\tau+\Delta\tau) \quad (\text{B.2})$$

$$P(\tau) = e^{-r_o \tau} (r_o \Delta\tau e^{-r_o \Delta\tau}), \quad (\text{B.3})$$

where we used equation B.1 to write each of the two terms in the expression for $P(\tau)$. For a sufficiently small time interval $\Delta\tau$,

$$\lim_{\Delta\tau \rightarrow 0} e^{-r_o \Delta\tau} = 1, \quad (\text{B.4})$$

and equation B.3 simplifies to

$$P(\tau) = e^{-r_o\tau} r_o d\tau, \quad (\text{B.5})$$

where $\Delta\tau$ has been replaced by the infinitesimal time interval dt . So, the waiting time between successive events (sometimes referred to as the *interoccurrence time*) is exponentially distributed with density function $\rho(\tau)$,

$$\rho(\tau) = r_o e^{-r_o\tau}, \quad (\text{B.6})$$

and average value $\langle \tau \rangle = 1/r_o$ and variance $\sigma^2 = \langle \tau^2 \rangle - \langle \tau \rangle^2 = 1/r_o^2$. It also follows that the cumulative distribution function (CDF) is $\mathcal{P}(\tau) = 1 - e^{-r_o\tau}$. Note that this result has the same algebraic form as the P_{ber} form of transition probabilities we used in our agent-based simulations. However, r_o is *not* the same as the KCF (symbol: Ω , see section 2.6). The CDF we mention here applies to the cumulative probability of a transition event for the *entire* population. In contrast, P_{ber} pertains to each agent in a population.

An interesting case to consider is when $\tau = 0$, and the exponential term in equation B.6 has its maximum value of 1. Then, the probability density becomes

$$\rho_{max}(\tau) = r_o \Rightarrow \frac{\rho_{max}(\tau)}{r_o} = 1. \quad (\text{B.7})$$

That is, the *maximum* value of $\rho(\tau)$ is r_o , and therefore $\rho(\tau)$ is always a fraction of r_o .

Alternative proof for $P(\tau)$. We offer an additional way to obtain the above result that τ is exponentially distributed, without initially assuming that we are dealing with a Poisson process.

We again begin by writing an expression for $P(\tau)$ (equation B.2). We can rewrite the first term on the right-hand side of B.2 as

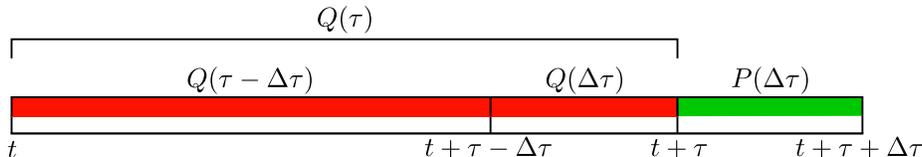
$$\begin{aligned} P(\text{no event between } t \text{ and } t + \tau) &= P(\text{no event between } t \text{ and } t + \tau - \Delta\tau) \\ &\cdot P(\text{no event between } t + \tau - \Delta\tau \text{ and } t + \tau). \end{aligned} \quad (\text{B.8})$$

Essentially, we have split up the time interval in which no event can occur into two time intervals: the first is of length $\tau - \Delta\tau$ and the second $\Delta\tau$. For notational convenience, we define $Q(\tau) = P(\text{no event between } t \text{ and } t + \tau)$. The above relationship can then be rewritten as

$$Q(\tau) = Q(\tau - \Delta\tau) Q(\Delta\tau). \quad (\text{B.9})$$

The figure below clarifies the time intervals we are considering and their respective probabilities. We have colored the intervals red if no reaction takes place,

and green if one event occurs within it.



Notice that we used the notation $P(\Delta\tau) = P(1 \text{ event between } t+\tau \text{ and } t+\tau+\Delta\tau)$.

Taking the natural logarithm of both sides and rearranging, we get

$$\ln Q(\tau) - \ln Q(\tau - \Delta\tau) = \ln Q(\Delta\tau). \quad (\text{B.10})$$

We recognize that $Q(\Delta\tau) = 1 - r_o \Delta\tau$, and the series expansion of the right-hand side of the above equation then is

$$\begin{aligned} \ln(1 - r_o \Delta\tau) &= - \left[r_o \Delta\tau + \frac{(r_o \Delta\tau)^2}{2} + \frac{(r_o \Delta\tau)^3}{3} + \dots \right] \\ &= - [r_o \Delta\tau + o(\Delta\tau)], \end{aligned} \quad (\text{B.11})$$

where $o(\Delta\tau)$ refers to terms that go to 0 faster than $\Delta\tau$. We then substitute this result into B.10, divide both sides by $\Delta\tau$ and take the limit as $\Delta\tau \rightarrow 0$. We have,

$$\lim_{\Delta\tau \rightarrow 0} \frac{\ln Q(\tau) - \ln Q(\tau - \Delta\tau)}{\Delta\tau} = -r_o, \quad (\text{B.12})$$

where the error term has vanished as $\Delta\tau \rightarrow 0$. The left-hand side is the derivative of $\ln Q(\tau)$. Thus,

$$\frac{d[\ln Q(\tau)]}{d\tau} = -r_o, \quad (\text{B.13})$$

whose solution, given the initial condition $Q(0) = 1$ (this follows from the definition of Q), is

$$Q(\tau) = e^{-r_o \tau}. \quad (\text{B.14})$$

We can then substitute this result into expression B.2, along with $P(d\tau) = r_o d\tau$, and continue as shown previously to get the exponential distribution of τ .

B.1 Algorithm Implementation

In general, if there are several chemical species participating in a total number of m reactions,¹ then let r_o represent the *sum* of the *microscopic rates*,² of all

¹The reactions are assumed to be Markovian processes.

²More accurately referred to as *propensity functions* in this context, which generally have the same algebraic form as the microscopic reaction rate law (discussed in section 2.1). The reader can refer to Gillespie (2007) for a more thorough review of the method.

reactions that can occur in the system,

$$r_o = \sum_{i=1}^m r_i, \quad (\text{B.15})$$

where r_i depend on the microscopic rate constants k_i and the number of molecules participating in a reaction.³

The numerical implementation of the algorithm requires a computing system to generate two (pseudo-)random numbers, κ_1 and κ_2 , drawn from the standard uniform distribution in the interval $(0, 1]$, which are used to determine:

1. *When* the next reaction will take place: as shown above, $0 < \frac{\rho(\tau)}{r_o} \leq 1$, and letting $\kappa_1 = \frac{\rho(\tau)}{r_o}$, we get

$$\kappa_1 = e^{-r_o \tau}, \quad (\text{B.16})$$

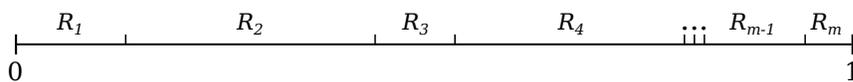
where we used equation B.6. Solving for τ , the time interval until the next reaction event occurs, we get

$$\tau = -\frac{\ln(\kappa_1)}{r_o}. \quad (\text{B.17})$$

2. *Which* reaction will take place next (after time interval τ): we normalize the individual reaction rates against the overall rate r_o

$$R_i = \frac{r_i}{r_o}. \quad (\text{B.18})$$

The interval $(0, 1]$ is therefore composed of non-overlapping, contiguous subintervals of length R_i such that $\sum R_i = 1$.



The reaction that occurs is selected by determining in which interval random number κ_2 falls in. For instance, if $0 < \kappa_2 < R_1$ then the first in the list of possible reactions occurs. If $R_1 \leq \kappa_2 < R_1 + R_2$, the second reaction occurs, etc.

3. Finally, the molecule numbers are adjusted based on which reaction was selected, and the next iteration of the algorithm takes place after the time variable has been advanced by τ sec.

It is evident from the above description that the algorithm does *not* consider the probability of reaction for each molecule, but instead determines in each

³For instance, for the 1st order process $A \rightarrow X$, $r = kN_A$, where N_A is the number of A molecules in the population.

iteration the time at which a single reaction event will occur in the entire population. The population is thereby treated as homogeneous in makeup (the rate constant parameterizes the probability of reaction for each species' entire population), and only productive reaction events are considered⁴ (i.e., the ones resulting in the formation of products). The rationale behind ignoring unproductive or nonreactive collisions has to do with the assumption that the population is homogeneous with respect to its spatial distribution. The reader has surely noted that none of the above arguments were made with spatial constraints in mind, and the molecules are assumed to be well-mixed or uniformly distributed within the reaction vessel's volume (or, more generally, the population density is constant throughout the space where agents reside).

B.2 Relation to CME

We set up the CME for a process that has n molecules of a species at a particular time t , and solve it for the case where the population remains at n at time $t + \tau$ (i.e., there are no reaction events in this time interval). We remind the reader that the general form of the CME is (Appendix A),

$$\frac{\partial P_n(t)}{\partial t} = \sum_{n'} T(n|n')P_{n'}(t) - \sum_{n''} T(n''|n)P_n(t).$$

Since we are assuming that the population size at the beginning of our time interval was n and not of some other size, the first term on the right-hand side of the above expression is zero (since $T(n|n') = 0$). Also, we adjust the notation to make it consistent with the above presentation of the SSA: the sum of all transition probabilities from state n is

$$\sum_{n''} T(n''|n) = r_o. \quad (\text{B.19})$$

The CME then becomes

$$\frac{\partial P_n(t)}{\partial t} = -r_o P_n(t). \quad (\text{B.20})$$

Upon separating variables and integrating with respect to the dummy time variable x , we get

$$\int_t^{t+\tau} \frac{dP_n(x)}{P_n(x)} dx = -r_o \int_t^{t+\tau} dx, \quad (\text{B.21})$$

whose solution, given that $P_n(t) = 1$ (since the initial population size was set), is

$$P_n(t + \tau) = e^{-r_o \tau}. \quad (\text{B.22})$$

⁴The method is sometimes referred to in the literature as "rejection-free," indicative of the fact that no potential reaction events are rejected as part of the algorithm. Notice that the ABK algorithm assesses the probability of all possible reaction events, many of which are rejected depending on the kinetic parameters in the modeled system.

This is the probability that the system remains at state n at time $t + \tau$, in agreement with B.14. Moreover, the probability of a transition event occurring in $[\tau, \tau + d\tau)$ is $r_o d\tau$.⁵ Therefore, the probability, $P(\tau)$, that the next transition event takes place between $t + \tau$ and $t + \tau + d\tau$ is

$$P(\tau) = e^{-r_o \tau} r_o d\tau, \tag{B.23}$$

which is the same result as in equation B.5 (which we obtained in two different ways in the first section of this appendix). We see that the CME predicts that τ is exponentially distributed, as we had previously concluded. As such, the SSA and CME approaches can be considered to be equivalent. Put another way, the SSA samples the probability distribution that is the solution to the CME, thusly producing an exact “realization” or trajectory of the population sizes.⁶ Similarly to the ABK method, an ensemble of such trajectories can be used to assess the expected fluctuations around the mean trajectory (i.e., the variance or standard deviation).

⁵We are assuming that r_o is the same at times t and $t + \tau$, since the system’s state had not changed in that interval: $r_o(t) = r_o(t + \tau)$.

⁶Formally referred to as a *Markov jump process*, since the population sizes can only jump from one integer value to another.

Appendix C

A Primer on Stability Analysis

We briefly present how to assess the stability of fixed points in 1- and 2-dimensional dynamical systems expressed as ODEs.¹ It is beyond the scope of this appendix to offer a detailed analysis of nonlinear dynamics (we refer the reader to Strogatz, 2015, or Kaplan and Glass, 1995, for lengthier treatments).

C.1 1-Dimensional Linear Systems

We begin with the general form for the ODE of an autonomous 1-dimensional system,

$$\frac{dx}{dt} = f(x), \quad (\text{C.1})$$

and will initially focus on the case where the function $f(x)$ is linear,

$$\frac{dx}{dt} = \lambda x, \quad (\text{C.2})$$

with λ being a real constant. We can generally determine the fixed points of a system by finding the values of x for which $\frac{dx}{dt} = 0$. Clearly, the only fixed point is $x^* = 0$. Moreover, it is easy to see that

$$\frac{d^2x}{dt^2} = \frac{d}{dt} \left(\frac{dx}{dt} \right) = \frac{d}{dt} (\lambda x) = \lambda \frac{dx}{dt} = \lambda^2 x. \quad (\text{C.3})$$

Similarly, the n^{th} derivative of x then is

$$\frac{d^n x}{dt^n} = \lambda^n x. \quad (\text{C.4})$$

¹As one may anticipate, these arguments can be extended to higher dimensional systems.

We can try to find the solution of the ODE by considering the Taylor expansion of $x(t)$ around $t = 0$.

$$x(t) = x(0) + \left. \frac{dx}{dt} \right|_{t=0} t + \left. \frac{d^2x}{dt^2} \right|_{t=0} \frac{t^2}{2!} + \cdots + \left. \frac{d^n x}{dt^n} \right|_{t=0} \frac{t^n}{n!} \quad (\text{C.5})$$

Upon substituting the expressions for the n^{th} derivative of x we already determined, we get

$$x(t) = x(0) + \lambda x(0)t + \lambda^2 x(0) \frac{t^2}{2!} + \cdots + \lambda^n x(0) \frac{t^n}{n!} \quad (\text{C.6})$$

$$= x(0) \left[1 + \lambda t + \frac{(\lambda t)^2}{2!} + \cdots + \frac{(\lambda t)^n}{n!} \right] \quad (\text{C.7})$$

$$= x(0) e^{\lambda t}. \quad (\text{C.8})$$

It is then clear that as long as $x(0) \neq 0$, $x(t)$ follows exponential growth or decay depending on if $\lambda > 0$ or $\lambda < 0$, respectively. The trivial case where $\lambda = 0$ means that $x(t) = x(0)$ for all values of t .

In summary, therefore, the stability of the fixed point $x^* = 0$ can be categorized according to the value of the parameter λ :

1. $x^* = 0$ is stable when $\lambda < 0$, and $x(t) \rightarrow 0$ as $t \rightarrow \infty$,
2. $x^* = 0$ is unstable when $\lambda > 0$, and $x(t) \rightarrow \infty$ as $t \rightarrow \infty$,
3. $x^* = 0$ is metastable when $\lambda = 0$, and $x(t) = x(0)$.

The term *metastable* in the above last case conveys the idea that any change in the value of λ will change the system's dynamics such that it follows one of the first two cases depending on the sign of λ .

If the original ODE had the more general form

$$\frac{dx}{dt} = \lambda x + b, \quad (\text{C.9})$$

then the analysis would be the same, except the fixed point has now shifted to $x^* = -\frac{b}{\lambda}$, or equivalently, $b = -\lambda x^*$. It can be more straightforward to rewrite the ODE by changing variables so that we can keep track of the *magnitude of a perturbation from the fixed point*. We can define

$$\chi(t) = x(t) - x^*. \quad (\text{C.10})$$

Clearly, since x^* is a constant,

$$\frac{d\chi}{dt} = \frac{dx}{dt} \quad (\text{C.11})$$

$$\begin{aligned} &= \lambda x + b \\ &= \lambda(\chi + x^*) - \lambda x^* \\ &= \lambda\chi. \end{aligned} \quad (\text{C.12})$$

Thus, the ODE for the variable χ has the same form as equation C.2, whose closed-form solution we have already determined (equation C.8). Solving in terms of our original variable $x(t)$, we get

$$\begin{aligned}\chi(t) &= \chi(0) e^{\lambda t} & (C.13) \\ x(t) - x^* &= (x(0) - x^*) e^{\lambda t} \\ x(t) &= x(0) e^{\lambda t} + x^* (1 - e^{\lambda t}). & (C.14)\end{aligned}$$

C.2 1-Dimensional Nonlinear Systems

We will now extend this discussion to ODEs of the form C.1, but with $f(x)$ being a nonlinear function. We first identify all of the fixed points and then linearize the function at each of them. Notice that this is a local approach: we aim to understand the stability of each fixed point by analyzing the system's behavior in its immediate vicinity. We accomplish this by considering a small perturbation $\chi(t) = x(t) - x^*$, and, as we have shown,

$$\frac{d\chi}{dt} = \frac{dx}{dt} = f(x) = f(\chi + x^*). \quad (C.15)$$

Taylor expanding the right-hand side of the above equation around $\chi = 0$ (which implies that $x = x^*$) yields

$$\frac{d\chi}{dt} = f(x^*) + \left. \frac{df(x)}{dx} \right|_{x=x^*} \chi + O(\chi^2), \quad (C.16)$$

where $O(\chi^2)$ represents quadratic and higher order terms in χ . Upon recognizing that $f(x^*) = 0$ since x^* is a fixed point, and simplifying the notation by rewriting $\left. \frac{df(x)}{dx} \right|_{x=x^*} = f'(x^*)$, the above equation becomes

$$\frac{d\chi}{dt} = f'(x^*) \chi + O(\chi^2). \quad (C.17)$$

We can rewrite this without the error term as

$$\frac{d\chi}{dt} \approx f'(x^*) \chi, \quad (C.18)$$

which is a 1-dimensional linear ODE with respect to χ . Therefore, $f'(x^*)$ plays the role of λ in our earlier discussion, and its sign will determine if the perturbation χ grows or decays (exponentially) as $t \rightarrow \infty$. Consequently, the stability at x^* can be determined. We note that this analysis does not hold when $f'(x^*) = 0$, in which case the $O(\chi^2)$ term cannot be ignored and a careful analysis of the nonlinear terms is needed.

Graphical analysis

It is sometimes convenient to analyze the behavior of the system graphically. One typically plots the derivative $\frac{dx}{dt} = f(x)$ as a function of x , and the x -intercepts correspond to fixed points whose stability is revealed by the slope of $f(x)$ as it crosses the x -axis. In accordance with our results upon linearization at x^* , if $f'(x^*) < 0$ then x^* is stable.

Another graphical method that is sometimes employed is the *rate-balance plot* when $f(x)$ is a polynomial with additive and subtractive terms representing a gain and loss in x , respectively. Then, the intersections between the graphs of the aggregate additive and subtractive terms denote the fixed points. Examples where we have used this method can be found in Figures 3.4a, 3.10a, 3.28, and 3.36a.

C.3 2-Dimensional Linear Systems

We now consider the general form for the ODEs of an autonomous 2-dimensional system

$$\frac{dx}{dt} = f(x, y) \quad (\text{C.19})$$

$$\frac{dy}{dt} = g(x, y). \quad (\text{C.20})$$

We will begin by presenting the case where the functions $f(x, y)$ and $g(x, y)$ are linear, and will subsequently generalize by considering nonlinearity.

A linear system can then be written as

$$\frac{dx}{dt} = ax + by \quad (\text{C.21})$$

$$\frac{dy}{dt} = cx + dy, \quad (\text{C.22})$$

where the coefficients a, b, c, d are real constants. It is convenient to rewrite this system of ODEs in matrix form,

$$\frac{d\mathbf{X}}{dt} = M\mathbf{X}, \quad (\text{C.23})$$

where bold-faced notation was used to indicate the column vector $\mathbf{X} = \begin{bmatrix} x \\ y \end{bmatrix}$ and

M is the square matrix of the coefficients, $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Notice that $\mathbf{X} = \mathbf{0}$ is always a fixed point of this linear system. The solution $\mathbf{X}(t)$ can be visualized parametrically by plotting it on the xy -plane, also known as the *phase plane* in dynamical systems analysis.

We can guess that the solution to this system of ODEs has the form $\mathbf{X}(t) = \mathbf{V}e^{\lambda t}$, with the parameter λ and the column vector \mathbf{V} to be determined. Upon

plugging this putative solution into equation C.23, we obtain

$$\frac{d}{dt} (\mathbf{V}e^{\lambda t}) = M\mathbf{V}e^{\lambda t} \quad (\text{C.24})$$

$$\lambda\mathbf{V}e^{\lambda t} = M\mathbf{V}e^{\lambda t}. \quad (\text{C.25})$$

Cancellation of the exponential terms leads to

$$M\mathbf{V} = \lambda\mathbf{V}, \quad (\text{C.26})$$

revealing that finding the solutions to the original system of ODEs is equivalent to an eigenvalue problem. The values of λ that we will solve for are the eigenvalues of their respective eigenvectors \mathbf{V} . Note that given the matrix notation we introduced, the above comments apply to systems of higher dimensionality as well.

We can find λ by solving the characteristic equation

$$\det(M - \lambda I) = 0, \quad (\text{C.27})$$

where $\det()$ stands for the determinant of a square matrix, and I is the identity matrix. Obviously, the order of the characteristic polynomial is the system's dimension. In the case of a 2-dimensional system, the characteristic polynomial is

$$(a - \lambda)(d - \lambda) - bc = 0 \quad (\text{C.28})$$

$$ad - (a + d)\lambda + \lambda^2 - bc = 0$$

$$\lambda^2 - (a + d)\lambda + (ad - bc) = 0. \quad (\text{C.29})$$

It is common to define the trace of M and its determinant as

$$\text{trace}(M) = \tau = a + d \quad (\text{C.30})$$

$$\det(M) = \Delta = ad - bc. \quad (\text{C.31})$$

Therefore, equation C.29 can be rewritten as

$$\lambda^2 - \tau\lambda + \Delta = 0, \quad (\text{C.32})$$

whose solutions are

$$\lambda_{1,2} = \frac{1}{2} \left(\tau \pm \sqrt{\tau^2 - 4\Delta} \right). \quad (\text{C.33})$$

Notice that the eigenvalues depend only on τ and Δ . It also follows that

$$\tau = \lambda_1 + \lambda_2 \quad (\text{C.34})$$

$$\Delta = \lambda_1\lambda_2. \quad (\text{C.35})$$

In the likely case that $\lambda_1 \neq \lambda_2$, their respective eigenvectors \mathbf{V}_1 and \mathbf{V}_2 are linearly independent, and the general solution for $\mathbf{X}(t)$ is the linear combination

$$\mathbf{X}(t) = C_1 e^{\lambda_1 t} \mathbf{V}_1 + C_2 e^{\lambda_2 t} \mathbf{V}_2,$$

where the constants C_1 and C_2 can be determined by applying the initial condition $\mathbf{X}(0) = \mathbf{X}_0$.

The stability of the fixed point $\mathbf{X}^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ is determined by the eigenvalues (or, equivalently, by τ and Δ). For real eigenvalues, if $\lambda_1 < 0$ and $\lambda_2 < 0$, then x^* is stable; on the other hand, if $\lambda_1 > 0$ and $\lambda_2 > 0$, then it is unstable. A *saddle point* occurs when the eigenvalues are of opposite sign (we show an example of a saddle in Figure 3.13b). In the case where the eigenvalues are complex, $\lambda_{1,2} = \alpha \pm \beta i$, then the exponential part of the solution can be written using Euler's formula,

$$e^{\lambda_{1,2}t} = e^{(\alpha \pm \beta i)t} = e^{\alpha t} e^{\pm i\beta t} = e^{\alpha t} (\cos \beta t \pm i \sin \beta t).$$

Therefore, the solution $\mathbf{X}(t)$ is periodic if $\alpha = 0$, or spirals into ($\alpha < 0$; stable spiral) or away ($\alpha > 0$; unstable spiral) from the fixed point (an example of a stable spiral is shown in Figure 3.38d).

An easy graphical way of classifying a fixed point's stability is by plotting Δ vs τ . We will not reproduce this stability diagram here because the plot can be found in the textbooks mentioned in the beginning of this appendix (for an online version, see https://en.wikipedia.org/wiki/Stability_theory).

C.4 2-Dimensional Nonlinear Systems

Armed with our treatment of linear systems, we can now extend our discussion to nonlinear ones. We will employ the same method as in the 1-dimensional case: linearization at each of the fixed points. These points can be found by setting the ODEs equal to zero,

$$\frac{dx}{dt} = f(x, y) = 0 \tag{C.36}$$

$$\frac{dy}{dt} = g(x, y) = 0, \tag{C.37}$$

and solving for any values of x and y that satisfy these conditions. A graphical analysis on the phase plane is often useful. We can solve for one variable with respect to the other using both of the above resulting algebraic expressions, and obtain the x and y *nullclines* (for instance, see Figure 3.13b). The intersections between these two curves represent fixed points whose stability can be determined using the method we describe below.

Given a fixed point (x^*, y^*) , we define the magnitude of a perturbation from each of these coordinates by

$$\chi(t) = x(t) - x^* \tag{C.38}$$

$$\psi(t) = y(t) - y^*. \tag{C.39}$$

It follows that

$$\frac{d\chi}{dt} = \frac{dx}{dt} \quad (\text{C.40})$$

$$\frac{d\psi}{dt} = \frac{dy}{dt}, \quad (\text{C.41})$$

since x^* and y^* are constants. We then do a Taylor expansion for both functions $f(x, y)$, $g(x, y)$ at $\chi = 0$ and $\psi = 0$, respectively.

$$\frac{d\chi}{dt} = \left. \frac{\partial f(x, y)}{\partial x} \right|_{(x^*, y^*)} \chi + \left. \frac{\partial f(x, y)}{\partial y} \right|_{(x^*, y^*)} \psi + O(\chi^2, \psi^2, \chi\psi) \quad (\text{C.42})$$

$$\frac{d\psi}{dt} = \left. \frac{\partial g(x, y)}{\partial x} \right|_{(x^*, y^*)} \chi + \left. \frac{\partial g(x, y)}{\partial y} \right|_{(x^*, y^*)} \psi + O(\chi^2, \psi^2, \chi\psi) \quad (\text{C.43})$$

Notice that each of the partial derivatives are evaluated at (x^*, y^*) and that we have omitted the 0th-order term in χ and ψ since $f(x^*, y^*) = g(x^*, y^*) = 0$. Thus, the dynamics of χ and ψ are described by

$$\begin{bmatrix} d\chi/dt \\ d\psi/dt \end{bmatrix} = \begin{bmatrix} \left. \frac{\partial f(x, y)}{\partial x} \right|_{(x^*, y^*)} & \left. \frac{\partial f(x, y)}{\partial y} \right|_{(x^*, y^*)} \\ \left. \frac{\partial g(x, y)}{\partial x} \right|_{(x^*, y^*)} & \left. \frac{\partial g(x, y)}{\partial y} \right|_{(x^*, y^*)} \end{bmatrix} \begin{bmatrix} \chi \\ \psi \end{bmatrix} + O(\chi^2, \psi^2, \chi\psi). \quad (\text{C.44})$$

The square matrix of the partial derivatives is the *Jacobian matrix* evaluated at the fixed point (x^*, y^*) .² Omitting the error term $O(\chi^2, \psi^2, \chi\psi)$ that includes quadratic and higher order terms, the linearized system of the perturbation ODEs becomes

$$\begin{bmatrix} d\chi/dt \\ d\psi/dt \end{bmatrix} \approx \begin{bmatrix} \left. \frac{\partial f(x, y)}{\partial x} \right|_{(x^*, y^*)} & \left. \frac{\partial f(x, y)}{\partial y} \right|_{(x^*, y^*)} \\ \left. \frac{\partial g(x, y)}{\partial x} \right|_{(x^*, y^*)} & \left. \frac{\partial g(x, y)}{\partial y} \right|_{(x^*, y^*)} \end{bmatrix} \begin{bmatrix} \chi \\ \psi \end{bmatrix}, \quad (\text{C.45})$$

which can be treated as a linear system of ODEs with the Jacobian matrix playing the role of M (see section C.3). Finally, the computed eigenvalues are used to classify the fixed point.

²The generality of this approach can be appreciated by connecting this discussion to 1-dimensional systems: the Jacobian of $f(x)$ is simply $df/dx = f'(x)$.

Appendix D

Selected Matlab Code

Listing D.1 – 1st order process for a homogeneous population.

```
1 % A --> B , Simulating 1st order kinetics.
2 % Considering individuals/agents of species A. Reporting time course
3 % and time distribution of reaction/transition events (extinction).
4
5 clear;      clc;
6 rng(0);
7
8 % *** Simulation Parameters ***
9 dt = 0.01;                                % Assume constant time step increment (sec)
10 totalTime = 10;                            % Simulation time (sec)
11 t_steps = totalTime / dt;
12 time = 0:dt:totalTime-dt;
13
14 agents = 10;                                % Initial number of A agents
15
16 lambda = 0.8;                               % Half-life in sec
17 k = log(2)/lambda;                          % Rate constant in 1/sec
18
19 reps = 100;                                 % Number of times to repeat simulation
20
21 % Preallocate memory.
22 S = zeros(reps,t_steps);                   % To monitor population size
23 extt = zeros(reps,agents);                 % To monitor extinction time of A agents
24
25 % *** Transition probability ***
26 % Choose one of the following two options:
27 P = 1 - exp(-k*dt);                        % P_ber or P_int
28 % P = k*dt;                                % P_dif
29 % Note that this probability is constant for fixed dt,
30 % so it can be calculated just once (outside of 'for' loops below).
31
32 %% ABK Simulation
33
34 for n = 1:reps
35
36     % Binary-valued (0/1) array for monitoring state of A agents; set to initial condition.
37     A = ones(1,agents);                    % 0 = dead , 1 = alive
38
39     S(n,1) = sum(A);                        % Initial population size
40
41     for t = 2:t_steps
42
43         tempA = find(A==1);                % Find alive A agents. This is done to
44                                             % speed up execution of 'for' loop below.
45
46         for i = 1:size(tempA,2)
47
48             if rand < P                    % * check probability condition *
49                 A(tempA(i)) = 0;          % A agent dies
50                 extt(n,tempA(i)) = time(t); % Record time of this agent's extinction/death
51             end
52
53         end                                % end 'for i' loop
54
55         S(n,t) = sum(A);
56
57     end                                    % end 'for t' loop
58
59 end
```

```

59 end % end 'for n' loop
60
61 %% Post-Processing
62 % Calculate the average and standard deviation of time trajectories.
63 avg = mean(S); % sdev = std(S);
64
65 % Remove zeros from extt (zeros represent agents that didn't die within totalTime).
66 for z=1:agents
67 extt_zeros = find(extt(:,z)==0);
68 for a = 1:size(extt_zeros,1)
69 extt(extt_zeros(a),z) = NaN;
70 end
71 end
72 extt = reshape(extt,[],1); % Reshape into column vector
73
74 % Calculate deterministic time trajectory.
75 y_theory = agents * exp(-k .* time);
76
77 %% Plot Time Trajectories
78 fig0 = figure; hold on;
79 set(fig0,'Position',[1 1 500 450]);
80 fig0.PaperUnits = 'inches';
81 fig0.PaperPosition = [0 0 6 5]; % Control the size of printed (eps) fig.
82
83 trial = 19; % For plotting a sample stochastic trajectory
84 % Choose a value (0 < trial <= reps)
85
86 p1 = plot(time,avg,'b');
87 p2 = plot(time,avg+sdev,'k','LineStyle','--','Color',[0.8 0.8 0.8]);
88 p3 = plot(time,avg-sdev,'k','LineStyle','--','Color',[0.8 0.8 0.8]);
89 p4 = plot(time,y_theory,'g');
90 p5 = plot(time,S(trial,:), 'c'); % Plot a sample stochastic trajectory
91
92 title(['N_A,i = ' num2str(agents)],'FontName','Times New Roman','FontSize',12);
93 axis([0 totalTime 0 agents]);
94 set(gca,'XMinorTick','on','Box','off');
95 xlabel('t (sec)');
96 ylabel('N_A');
97
98 leg0 = legend([p1 p4], 'ABK <N_A(t)>', 'N_A(t) = 10 e^{-kt}');
99 set(leg0,'FontName','Times New Roman','FontSize',9,...
100 'EdgeColor',[0.95 0.95 0.95],'Location','NorthEast');
101
102 % hgsave(['oi_TCdev_A=' num2str(agents) '.fig']);
103
104 %% Plot Extinction Time Distribution
105 bins = 20; % Number of bins for histogram
106
107 fig1 = figure; hold on;
108 [num1, c1] = hist(extt,bins); % Obtain histogram data
109 Num1 = num1 / (reps*agents); % Normalize
110
111 b = bar(c1',Num1');
112 set(b(1),'FaceColor','flat','EdgeColor',[0.70 0.78 1]);
113
114 % Calculate PDF of exponential distribution.
115 exp_PDF1 = 1 / (bins/totalTime) * k * exp(- k .* time);
116 % PDF needs to be adjusted for the number of bins in the histogram.
117 % Adjustment: divide by (bins/totalTime)
118
119 plot(time,exp_PDF1,'-','Color','g','LineWidth',2);
120
121 set(gca,'XMinorTick','on','YMinorTick','on','Box','off');
122 xlabel('t (sec)'); ylabel('Fraction of A\rightarrow X Transitions');
123
124 % Sample exponential PDF at the same time values as were used in histogram.
125 exp_PDF1d = 1 / (bins/totalTime) * k * exp(- k * c1);
126
127 % Coefficient of Determination R^2 for histogram fit to PDF.
128 disp(['R^2 = ' num2str(CoefDet(Num1,exp_PDF1d))]);
129 % See CoefDet function definition below.
130
131 %% Clean Up Unnecessary Variables; Save Data.
132 clear i t p* fig* leg* A tempA;
133 save(['oi_A=' num2str(agents) '.mat']);

```

```

1 function Rsq = CoefDet(data,theor)
2 % Find Coefficient of Determination, R^2. This applies to any curve fit, linear or nonlinear.
3 % Data and theoretical curve vectors should be of the same length.
4
5 SST = sum((data - mean(data)).^2); % Total sum of squares for simulation data
6 SSR = sum((data - theor).^2); % Sum of square residuals
7 Rsq = 1 - SSR ./ SST; % Definition of R^2

```

Listing D.2 – 2nd order process $2A \rightarrow X$ for a homogeneous population.

```

1  % A + A --> X
2  % Simulating 2nd order kinetics using Agent-Based Kinetics.
3  % Assuming a homogeneous population of A.
4  clear;
5  rng(0);
6
7  global k dif_canon;
8
9  agents = 10;
10
11  k = 0.01; % units: 1/sec
12
13  t_max = 100; % units: sec
14  dt = 1/100; % Fixed time step increment
15  steps = t_max / dt;
16  time = zeros(1,steps);
17
18  reps = 500; % Times to repeat simulation
19
20 % Preallocate memory:
21 Sa = zeros(reps,steps); Sc = zeros(reps,steps);
22 P = zeros(reps,steps);
23
24 %% ABK algorithm
25 for n=1:reps
26
27   % Initial number of reactant molecules; assume B is limiting.
28   Ao = agents; Cmax = Ao;
29   % Arrays for tracking reactant and product agent status
30   A = ones(1,Ao); C = zeros(1,Cmax);
31   % Initialize time-dependent sum of molecule numbers
32   At = zeros(1,steps); Ct = zeros(1,steps);
33   At(1) = sum(A); Ct(1) = sum(C);
34
35   time(1) = 0;
36
37   for t=2:steps
38
39     % dt = 1/(k*At(t-1)*(At(t-1)-1)); % Variable time step increment
40     % dt = exprnd(1/(k*At(t-1)*(At(t-1)-1))); % Exponentially-distributed time step increment
41
42     % *** Evaluate transition probability (choose one of the following forms):
43     % P(n,t) = 1 - 1 / (1 + k * (At(t-1)-1) * dt); % P_can
44     P(n,t) = 1 - 1 / (At(t-1)*(1-exp(-k*dt)) + exp(-k*dt)); % P_int
45     % P(n,t) = 1 - exp(-k * (At(t-1)-1) * dt); % P_ber
46     % P(n,t) = k * (At(t-1)-1) * dt; % PP_dif
47
48     tempAa = find(A==1); % Find alive A agents
49
50     for i = 1:size(tempAa,2)
51
52       if tempAa(i) ~= 0 && rand < P(n,t)
53
54         A(tempAa(i)) = 0; % First A agent dies
55
56         x = ceil(rand*size(tempAa,2));
57         while tempAa(x) == 0
58           x = ceil(rand*size(tempAa,2));
59         end
60
61         A(tempAa(x)) = 0; % Second A agent dies (chosen randomly)
62
63         tempAa(i) = 0; % Make sure the two A agents that reacted
64         tempAa(x) = 0; % are not reselected in this time step.
65
66         C(i) = 1; % C agent is born
67       end
68     end % end 'for i' loop
69
70     At(t) = sum(A); Ct(t) = sum(C);
71     time(t) = time(t-1) + dt;
72
73   end % end 'for t' loop
74
75   Sa(n,:) = At; Sc(n,:) = Ct;
76
77 end % end 'for n' loop
78
79 % Calculate mean and standard deviation of ABK trajectories:
80 avg = mean(Sa); sdev = std(Sa);
81
82 %% Solve ODE for 2nd order kinetics
83 dif_canon = 0; % use agent-based form of DE
84 [t_sol0, y_sol0] = ode45(@o2_homo_dif,0:t_max/200:t_max,[At(1) ; Ct(1)]);
85 dif_canon = 1; % use canonical form of DE
86 [t_sol1, y_sol1] = ode45(@o2_homo_dif,0:t_max/200:t_max,[At(1) ; Ct(1)]);
87
88 %% Plot

```

```

89 figure1 = figure('Name','2nd Order Rx Time course','NumberTitle','off');
90 set(figure1,'Position',[1 1 500 450]); hold on;
91
92 p1 = plot(time,avg,'b','MarkerSize',3,'DisplayName','ABK');
93 p1_dev1 = plot(time,avg+sdev,'LineStyle','--','Color',[0.8 0.8 0.8]);
94 p1_dev0 = plot(time,avg-sdev,'LineStyle','--','Color',[0.8 0.8 0.8]);
95
96 % Plot DE solutions for species A
97 p2 = plot(t_sol0,y_sol0(:,1),'g','DisplayName','DE_{ab}');
98 p3 = plot(t_sol1,y_sol1(:,1),'r','DisplayName','DE_{can}');
99
100 xlabel('t (sec)'); ylabel('N_A(t)');
101 axis([0 t_max 0 agents]); hold off;
102 title(['N_{A,i} = ' num2str(agents)],'FontName','Times New Roman','FontSize',12)
103 set(gca,'XMinorTick','on','YMinorTick','on','Box','off');
104 leg = legend([p1 p2 p3]);
105 set(leg,'Position',[0.673 0.374 0.182 0.123],'FontName','Times New Roman'...
106 , 'EdgeColor',[0.95 0.95 0.95]);

```

Listing D.3 – 2nd order process $A + B \rightarrow X$ for a homogeneous population.

```

1  % A + B --> C , Simulating 2nd order kinetics.
2  % Considering individuals/agents of all species, calculating
3  % transition probabilities with respect to the agents of species B.
4  % Reporting time course (average trajectory and standard deviation).
5
6  clear;      clc;
7  rng(0);
8
9  maxTime = 30;
10 dt = 1/100;           % Fixed time step increment
11 t_steps = ceil(maxTime / dt);
12
13 agents = 10;
14
15 global k;             % Bimolecular microscopic kinetic constant
16 k = 0.01;           % units: 1/sec
17
18 reps = 500;         % Repeat experiment this # of times
19
20 % Preallocate memory; Initialize time-dependent population sizes
21 time = zeros(1,t_steps);
22 At = zeros(reps,t_steps);      Bt = zeros(reps,t_steps);      Ct = zeros(reps,t_steps);
23
24 for n=1:reps
25
26     % Initial number of reactant molecules; assume B is limiting.
27     Ao = agents;           Bo = ceil(0.7*agents);           Cmax = Bo;
28     % Arrays for tracking reactant and product agent state: % 0 = dead , 1 = alive.
29     A = ones(1,Ao);       B = ones(1,Bo);           C = zeros(1,Cmax);
30     At(n,1) = sum(A);     Bt(n,1) = sum(B);           Ct(n,1) = sum(C);
31
32     for t=2:t_steps
33
34         temp1 = find(B==1);           % Find alive B agents
35
36         P = 1 - exp(-k * At(n,t-1) * dt); % *** Transition probability P_ber ***
37
38         for i = 1:size(temp1,2)
39
40             if rand < P               % * check probability condition *
41
42                 B(temp1(i)) = 0;     % B agent dies
43                 temp2 = find(A==1); % Find alive A agents
44                 x = ceil(rand*size(temp2,2)); % Randomly pick alive A agent
45                 A(temp2(x)) = 0;     % Picked A agent dies
46                 C(temp1(i)) = 1;     % C agent becomes alive
47
48             end
49
50         end
51
52         time(t) = time(t-1) + dt;
53         At(n,t) = sum(A);           Bt(n,t) = sum(B);           Ct(n,t) = sum(C);
54
55     end                               % end for t loop
56
57 end                                   % end for reps loop
58
59 % Population statistics of time trajectories
60 avgA = mean(At);           avgB = mean(Bt);           avgC = mean(Ct);
61 sdevA = std(At);          sdevB = std(Bt);          sdevC = std(Ct);
62
63 % Solve ODE for 2nd order kinetics
64 tmax = time(t-1);
65 [t_sol,y_sol] = ode45(@o2_dif,0:tmax/100:tmax,[At(1) ; Bt(1) ; Ct(1)]);
66
67 % Plot average time course and standard deviation.
68 figure('Name','2nd Order Rx Time course','NumberTitle','off'); hold on;
69 plot(time,avgA,time,avgB,time,avgC);
70 pl_devA1 = plot(time,avgA+sdevA,'LineStyle','--','Color',[0.8 0.8 0.8]);
71 pl_devA0 = plot(time,avgA-sdevA,'LineStyle','--','Color',[0.8 0.8 0.8]);
72 pl_devB1 = plot(time,avgB+sdevB,'LineStyle','--','Color',[0.8 0.8 0.8]);
73 pl_devB0 = plot(time,avgB-sdevB,'LineStyle','--','Color',[0.8 0.8 0.8]);
74 pl_devC1 = plot(time,avgC+sdevC,'LineStyle','--','Color',[0.8 0.8 0.8]);
75 pl_devC0 = plot(time,avgC-sdevC,'LineStyle','--','Color',[0.8 0.8 0.8]);
76
77 xlabel('t (sec)');
78 axis([0 maxTime 0 agents]);
79
80 % Plot DE Time Courses
81 plot(t_sol,y_sol(:,1),'b');           plot(t_sol,y_sol(:,2),'g');
82 plot(t_sol,y_sol(:,3),'r');
83
84 legend('ABK <N_A(t)>','ABK N_B(t)','ABK N_C(t)');

```

Listing D.4 – Irreversible mutual inhibition switch motif.

```

1  %      k_da      k_db
2  % <-- A -- B --> % For Report: B |-> R
3  %
4  %      k_a      k_b
5  %      | / \ |
6  %      | / \ |
7  %      | / \ |
8  %      | / \ |
9  %      S
10 % A and B are synthesized (0th order constants k_a, k_b respectively)
11 % A and B are degraded (k_da, k_db)
12 % A and B inhibit each other's synthesis (MUTUAL inhibition, alpha < 1):
13 % B influences the rate of the synthesis of A
14 % A influences the rate of the synthesis of B
15 % S inhibits the inhibition of B by A (assume complete repression)
16 % Assume S does not change.
17 % Considering individual agents.
18 %% Declare variables and functions
19
20 clear;          tic;          clc;
21 rng(0);
22 pb = waitbar(0, '0');
23
24 reps = 100;
25
26 maxTime = 5000;          % Maximum Simulation time (sec)
27 dt = 1/50;              % Constant (fixed) time step increment (sec)
28 t_steps = maxTime / dt;
29
30 k_a = 1;                % MICROSCOPIC basal rate of A synthesis (0th order)
31 k_b = 1;                % MICROSCOPIC basal rate of B synthesis (0th order)
32 k_da = 0.01;           % degradation rate constant (1st order) for A
33 k_db = 0.01;           % degradation rate constant (1st order) for B
34 % Populations producing half-maximal regulatory effect.
35 K_a = 50;                K_a = 50;                K_s = 50;
36
37 S_array = 0:20;        % Number of S molecules to do simulation for
38
39 % *** Feedback parameters ***
40 % Degree of activation: >1 activator, <1 repressor, =1 no regulation
41 alpha_a = 0;           % Effect of A on synthesis of B; = 0 means Complete Repression
42 alpha_b = 0;           % Effect of B on synthesis of A; = 0 means Complete Repression
43 % Hill Coefficients
44 n_b = 3;                n_a = 3;                n_s = 1;
45
46 % Specify initial population sizes here.
47 Ao = 10;                Bo = 10;
48
49 % Preallocate memory.
50 ABK_B = zeros(reps, size(S_array,2));
51
52 %% Run ABK simulation 'reps' number of times
53 for a=1:reps
54
55     progress = a / reps;
56     waitbar(progress, pb, sprintf('%0.0f%% %', progress*100));
57     fprintf(1, '\n');
58
59     for s=1:size(S_array,2)
60
61         fprintf(1, '.');
62
63         S = S_array(s);
64
65         % Initialize - Preallocate memory for variables; ** Initial conditions **
66         t = 1;          % Time counter variable
67         P_sa = zeros(1, t_steps); % For storing probability of A synthesis at each time step
68         P_sb = zeros(1, t_steps); % For storing probability of B synthesis at each time step
69
70         Ta = zeros(1, t_steps); % Sum of A agents in each time step
71         Tb = zeros(1, t_steps); % Sum of B agents in each time step
72
73         % ***** Initial conditions - Number of A, B Agents *****
74         agents = 200; % disp(['Agents = ' num2str(agents)]);
75         Ta(1) = Ao; % Tb(1) = Bo;
76         % *****
77
78         tempA = zeros(1, agents); % tempB = zeros(1, agents);
79         % Put 1 where agents are alive, then randomize the array
80         for c=1:Ta(1), tempA(c)=1; end
81         for d=1:Tb(1), tempB(d)=1; end
82         tempA = RandArray(tempA); % Randomize A array
83         tempB = RandArray(tempB); % Randomize B array
84         Av = [tempA ; tempA]; % Initialize vector storing state of A agents
85         Bv = [tempB ; tempB]; % Initialize vector storing state of B agents
86         % Notes on Av, Bv:
87         % - Markov process, so only previous and current time steps needed --> 2 rows:
88         clear c d tempA tempB;

```

```

89 % ABK Simulation
90 P_da = 1 - exp(-k_da * dt); % P_ber of degradation of A, 1st order wrt A
91 P_db = 1 - exp(-k_db * dt); % P_ber of degradation of B, 1st order wrt B
92
93 while t <= t_steps %&& Ta(t) > agents/100
94 % fprintf(1, '%');
95
96 k_sa = k_a * (K_b^n_b + alpha_a * Tb(t)^n_b) / (K_b^n_b + Tb(t)^n_b);
97 k_sb = k_b * (K_a^n_a + alpha_b * (Ta(t) * 1/(1+(S/K_s)^n_s))^n_a) / ...
98 (K_a^n_a + (Ta(t) * 1/(1+(S/K_s)^n_s))^n_a);
99
100 P_sa(t) = k_sa * dt; % 0th order synthesis of A (regulated by B)
101 P_sb(t) = k_sb * dt; % 0th order synthesis of B (regulated by A)
102
103 % Take care of 0th order processes first
104 if rand < P_sa(t)
105 tempA = find(Av(1,:)==0); % Randomly choose A agent synthesis
106 Av(2,tempA(ceil(rand * size(tempA,2)))) = 1;
107 end
108
109 if rand < P_sb(t)
110 tempB = find(Bv(1,:)==0); % Randomly choose B agent synthesis
111 Bv(2,tempB(ceil(rand * size(tempB,2)))) = 1;
112 end
113 % end of 0th order processes
114
115 tempA = find(Av(1,:)==1);
116 for i = 1:size(tempA,2)
117 if rand < P_da % Degradation of A, 1st order rx
118 Av(2,tempA(i)) = 0; % A agent is degraded
119 end
120 end
121
122 tempB = find(Bv(1,:)==1);
123 for j = 1:size(tempB,2)
124 if rand < P_db % Degradation of B, 1st order rx
125 Bv(2,tempB(j)) = 0; % A agent is degraded
126 end
127 end
128
129 Ta(t+1) = sum(Av(2,:)); % Tb(t+1) = sum(Bv(2,:));
130 Av(1,:) = Av(2,:); % Bv(1,:) = Bv(2,:);
131 t = t + 1;
132 end
133
134 % Remove unnecessary terminal 0's from arrays
135 % if t < t_steps
136 % Ta = Ta(1:t); % Tb = Tb(1:t);
137 % P_sa = P_sa(1:t); % P_sb = P_sb(1:t);
138 % P_da = P_da(1:t); % P_db = P_db(1:t);
139 % end
140 % finaltime = (t-1) * dt;
141
142 % Average B population over the last 1000 sec of the simulation
143 % Steady-state has been reached for this time period.
144 ABK_B(a,s) = mean(Tb(end-1000/dt:end));
145
146 end % end 'for s = 1:size(S_array,2)' loop
147 end % end 'for a=1:reps' loop
148
149 avg_Bss = mean(ABK_B); % sdev_Bss = std(ABK_B);
150
151 %% Plot signal-response (SR) curve
152 fig1 = figure('Name','ABK SR Curve','NumberTitle','off');
153 set(fig1,'Position',[1 500 406]); hold on;
154
155 % Plot Theoretical SR curve. Load data obtained from the next code listing.
156 SR = load('SR_curve.mat','S_array','B_ss');
157 for h=1:size(SR.B_ss,2)
158 if mod(h,2) == 0
159 p(h) = plot(SR.S_array,SR.B_ss(:,h),'--',...
160 'LineWidth',2,'Color',[1 0.69 0.39],'DisplayName','Unstable');
161 else
162 p(h) = plot(SR.S_array,SR.B_ss(:,h),'-b','LineWidth',2,'DisplayName','Stable');
163 end
164 end
165
166 % Plot ABK SR curve
167 p_ABK = errorbar(S_array,avg_Bss,sdev_Bss,'or','DisplayName','ABK <N_R*>');
168
169 xlabel('N_S');
170 ylabel('N_R*'); % Recall: B |-> R
171
172 axis([S_array(1) S_array(end) 0 120]);
173 set(gca,'XMinorTick','on','YMinorTick','on','Box','off'); hold off;
174
175 % leg1 = legend([p_ABK],'Location','SouthEast');
176 % set(leg1,'FontName','Times New Roman','FontSize',9,'EdgeColor',[0.95 0.95 0.95]);
177
178 %% Plot the percentage of simulations where the switch is in the ON state
179 for x=1:size(ABK_B,2)

```

```

180     ON(x) = size(find(ABK_B(:,x)>50),1) / reps * 100;
181 end
182
183 fig2 = figure('Name','% ON','NumberTitle','off');
184 set(fig2,'Position',[501 1 500 406]);
185 plot(12*ones(1,100),1:100,'--','Color',[1 0.69 0.39]); hold on;
186 plot(S_array,ON,'ob');
187 xlabel('N_S'); ylabel('% ON');
188 axis([S_array(1) S_array(end) 0 100]); hold off;
189 set(gca,'XMinorTick','on','YMinorTick','on','Box','off');
190
191 %% Finish
192 close(pb); toc

```

```

1 % This script calculates the deterministic SR curve for the mutual inhibition switch motif.
2 % All parameters are specified in preceding script.
3
4 agents = 110;
5
6 S_array = 0:0.5:20; % S values to do calculations for.
7
8 % Preallocate memory
9 A_ss = zeros(size(S_array,2),3); % Steady state (ss) values for A
10 B_ss = zeros(size(S_array,2),3); % Steady state (ss) values for B ( I-> R )
11
12 for n=1:size(S_array,2)
13
14     S = S_array(n);
15     fprintf(1,['-----\nS = ' num2str(S) '\n']);
16
17     % *** Set up differential equations of system using symbolic variables ***
18     syms A B k_sa_sym k_sb_sym positive;
19
20     k_sa_sym = k_a * (K_b^n_b + alpha_b * B^n_b) / (K_b^n_b + B^n_b);
21     k_sb_sym = k_b * (K_a^n_a + alpha_a * (A * 1/(1+(S/K_s)^n_s))^n_a) / ...
22     (K_a^n_a + (A * 1/(1+(S/K_s)^n_s))^n_a);
23
24     dA_sym = + k_sa_sym - k_da * A;
25     dB_sym = + k_sb_sym - k_db * B;
26
27     % Nullclines
28     A_nc_sym = solve(dA_sym == 0,A,'MaxDegree',4,'Real',true);
29     B_nc_sym = solve(dB_sym == 0,B,'MaxDegree',4,'Real',true);
30
31     % Calculate nullclines
32     A_nc = double(subs(A_nc_sym,B,0:agents));
33     B_nc = double(subs(B_nc_sym,A,0:agents));
34
35     % Nullcline values must be positive!
36     if size(B_nc,1) > 1
37         for j=size(B_nc,1):-1:1
38             if isempty(find(sign(B_nc(j,:)) == -1, 1,'first')) == false
39                 B_nc(j,:) = [];
40             end
41         end
42     end
43
44     % Nullcline intersections provide graphically-determined fixed points.
45     [A_ss_int, B_ss_int] = intersections(A_nc,0:agents,0:agents,B_nc);
46
47     % Clean up duplicate ss data
48     for z=size(B_ss_int,1):-1:2
49         temp = B_ss_int(z) - B_ss_int(z-1);
50         if abs(temp) < 0.0001
51             B_ss_int(z) = [];
52             A_ss_int(z) = [];
53         end
54     end
55
56     % * Calculate Jacobian matrix *
57     Jac = jacobian([dA_sym,dB_sym],[A,B]);
58
59     fprintf(1,'Fixed Points:\t\tEigenvalues\n');
60
61     for w=1:size(A_ss_int,1)
62         J = double(subs(Jac,[A,B],[A_ss_int(w),B_ss_int(w)]));
63         TrJ = J(1,1) + J(2,2);
64         DetJ = det(J);
65         DiscrJ = TrJ^2 - 4*DetJ;
66         lambda_plus(w) = (TrJ + sqrt(DiscrJ)) / 2;
67         lambda_minus(w) = (TrJ - sqrt(DiscrJ)) / 2;
68         fprintf(1,['A= %4.2f , B= %4.2f :\t%+6.4f , %+6.4f \n'],...
69             A_ss_int(w),B_ss_int(w),lambda_plus(w),lambda_minus(w));
70     end
71
72     if isempty(A_ss_int) == 0
73         A_ss(n,:) = A_ss_int(:); B_ss(n,:) = B_ss_int(:);
74     else
75         A_ss(n,:) = NaN; B_ss(n,:) = NaN;

```

```

76     end
77 end
78 end
79
80 % Remove duplicate fixed points from monostable systems
81 for i=1:size(A_ss,1)
82     if A_ss(i,1) == A_ss(i,2)
83         A_ss(i,2:3) = NaN;           B_ss(i,2:3) = NaN;
84     end
85 end
86
87 % Plot deterministic SR curve
88 fig1 = figure('Name','SR curve','NumberTitle','off');
89 set(fig1,'Position',[501 1 500 406]);           hold on;
90
91 for h=1:size(B_ss,2)
92     if mod(h,2) == 0
93         p(h) = plot(S_array,B_ss(:,h),'--r','LineWidth',2,'DisplayName','Unstable');
94     else
95         p(h) = plot(S_array,B_ss(:,h),'-b','LineWidth',2,'DisplayName','Stable');
96     end
97 end
98
99 axis([S_array(1) S_array(end) 0 max(max(B_ss))+10]);           hold off;
100 set(gca,'XMinorTick','on','YMinorTick','on','Box','off');
101 xlabel('N_S');           ylabel('N_R'*');
102
103 leg1 = legend([p(1) p(2)]);
104 set(leg1,'Location','SouthEast');
105 set(leg1,'FontName','Times New Roman','FontSize',9,...
106     'EdgeColor',[0.95 0.95 0.95]);
107
108 % Save results
109 save('SR_curve.mat');

```

Listing D.5 – 1st order process for a heterogeneous population.

```

1  % A --> B ; Simulating 1st order kinetics.
2  % *** Heterogeneous population ***
3
4  clear; clc;
5  rng(0);
6
7  dt = 0.01; % Assume constant time step increment (sec)
8  totalTime = 10; % Total simulation time (sec)
9  t_steps = totalTime / dt;
10
11  reps = 500; % Number of times to repeat simulation
12  % k_calc = zeros(size(dt_array,2),n);
13  S = zeros(reps,t_steps); % Preallocate memory for population size of A
14
15  agents = 10; % Initial population size
16  % Note: # of agents should be integrally divisible by # of subspecies
17
18  extt = zeros(reps,agents); % To monitor extinction time of agents
19
20  % ***** Set population's k values *****
21  % Specify subspecies k values [should be of size (1,subspecies)]
22  k_sub = [0.3 , 0.7];
23  subspecies = size(k_sub,2); % number of subspecies
24
25  k = [];
26  for h=1:subspecies
27      k = [k , k_sub(h)*ones(1,agents/subspecies)];
28  end
29  ko = mean(k);
30  % *****
31
32  % Time-independent probability for a given agent (vectorized)
33  P = 1 - exp(-k .* dt); % P from integrated rate law
34  % P = k .* dt; % P from differential rate law
35
36  for n = 1:reps
37
38      A = ones(2,agents); % Only 2 rows because of Markov property
39      S(n,1) = agents; % Set initial population size
40      time = zeros(1,t_steps);
41
42      for t = 1:t_steps
43
44          tempAa = find(A(1,:)==1); % Find 'alive' A agents
45
46          for i = 1:size(tempAa,2)
47              if rand < P(tempAa(i)) % ** check probability condition **
48                  A(2,tempAa(i)) = 0; % Agent dies
49                  extt(n,tempAa(i)) = time(t)+dt; % Record extinction time
50              end % 'if rand' statement
51          end % 'for i' loop
52
53          S(n,t+1) = sum(A(2,:)); % Sum up 'alive' aganets
54
55          A(1,:) = A(2,:);
56          time(t+1) = time(t) + dt;
57      end % 'for t' loop
58  end % 'for n' loop
59
60  avg = mean(S); % sdev = std(S);
61
62  %% ** Group extinction times by subspecies **
63  % Remove zeros from extt (zeros represent agents that didn't die within totalTime)
64  for z=1:agents
65      extt_zeros = find(extt(:,z)==0);
66      for a = 1:size(extt_zeros,1)
67          extt(extt_zeros(a),z) = NaN;
68      end
69  end
70
71  extt2 = zeros(reps*agents/subspecies,subspecies);
72  s = 1; w = 1;
73
74  for j=1:agents
75      extt2((w-1)*reps+1:w*reps,s) = extt(:,j);
76
77      if j < agents && k(j) == k(j+1)
78          s = s + 1;
79          w = 1;
80      else
81          w = w + 1;
82      end
83  end
84
85  %% Calculate theoretical curves
86  t = 0:dt:totalTime;
87  y_theory_ko = agents * exp(- ko * t);
88  % y_theory_sub1 = agents/2*exp(- k_sub(1) * t);

```

```

89 % y_theory_sub2 = agents/2*exp(- k_sub(2) * t);
90
91 %% Plot average time trajectory
92 figure;
93 p1 = plot(time,avg,'b','LineWidth',1); hold on;
94 p2 = plot(time,avg+sdev,'LineStyle','--','Color',[0.8 0.8 0.8],'LineWidth',2);
95 p3 = plot(time,avg-sdev,'LineStyle','--','Color',[0.8 0.8 0.8],'LineWidth',2);
96 p4 = plot(t,y_theory_ko,'g','LineStyle',':','LineWidth',2);
97 % p5 = plot(t,y_theory_sub1,'-r','MarkerSize',5);
98 % p6 = plot(t,y_theory_sub2,'-k','MarkerSize',5);
99
100 set(gca,'XMinorTick','on','Box','off');
101 title(['N_A,i = ' num2str(agents)],'FontName','Times New Roman','FontSize',12);
102 axis([0 10 0 agents]);
103 xlabel('t (sec)'); ylabel('N_A'); hold off;
104
105 leg0 = legend([p1 p4],'ABK <N_A(t)>','N_A(t) = 10 e^{-<k> t}');
106 set(leg0,'FontName','Times New Roman','FontSize',9,...
107 'EdgeColor',[0.95 0.95 0.95],'Location','NorthEast');
108
109 % hgsave(['oi_TCdev_A=' num2str(agents) '.fig']);
110
111 % **** Plot sample time trajectories ****
112 % trial = [1 2];
113 % plot(time,S(trial(1),:),'ob','MarkerSize',3); hold on;
114 % plot(time,S(trial(2),:),'+c','MarkerSize',3);
115 % xlabel('t (sec)'); ylabel('N_A');
116
117 %% Plot Extinction time distribution
118 bins = 10;
119
120 figure;
121 % hist(extt(:,:)); % histogram for all agents
122 % hist(extt2(:,:),bins); % histogram for subspecies
123
124 exp_PDF1 = k_sub(1) / (bins/totalTime) * exp(- k_sub(1) * t);
125 exp_PDF2 = k_sub(2) / (bins/totalTime) * exp(- k_sub(2) * t);
126 % Note: PDFs need to be adjusted for the number of bins in the histogram
127 % in order to obtain proper fits. Adjustment: divide by (bins/totalTime)
128
129 [num1, c1] = hist(extt2(:,1),bins); % histogram data for subspecies 1
130 [num2, c2] = hist(extt2(:,2),bins); % histogram data for subspecies 2
131 Num1 = num1 / (reps*agents/subspecies); % Normalize
132 Num2 = num2 / (reps*agents/subspecies); % Normalize
133
134 % b = bar([c1' c2'],[Num1' Num2']); % Works in Matlab 2013a (but not in 2017a)
135 b = bar(totalTime/bins/2:totalTime/bins:totalTime,[Num1' Num2']);
136 hold on;
137 set(b(1),'FaceColor','flat','EdgeColor',[0.70 0.78 1]); % For subspecies 1
138 set(b(2),'FaceColor',[1 0.1 0.1],'EdgeColor',[1 0.6 0.6]); % For subspecies 2
139 plot(t,exp_PDF1,'-','Color',[0.15 0.15 1],'LineWidth',1);
140 plot(t,exp_PDF2,'-','Color',[1 0.15 0.15],'LineWidth',1);
141
142 set(gca,'XTick',[0 1 2 3 4 5 6 7 8 9 10],...
143 'XMinorTick','on','YMinorTick','on','Box','off');
144 xlabel('t (sec)'); ylabel('Fraction of A\rightarrowX Transitions');
145
146 leg1 = legend(['Subspecies A1, k_{A1} = ' num2str(k_sub(1)) ' sec^{-1}',...
147 ['Subspecies A2, k_{A2} = ' num2str(k_sub(2)) ' sec^{-1}']);
148 set(leg1,'FontName','Times New Roman','FontSize',9,...
149 'EdgeColor',[0.95 0.95 0.95],'Location','NorthEast');
150
151 % Coefficient of Determination R^2 for histogram fits to exponential dist PDFs.
152 % Discretize exp PDF for time points c1, c2
153 exp_PDF1d = k_sub(1) / (bins/totalTime) * exp(- k_sub(1) * c1);
154 exp_PDF2d = k_sub(2) / (bins/totalTime) * exp(- k_sub(2) * c2);
155
156 disp(['Subspecies A1: R^2 = ' num2str(CoefDet(Num1,exp_PDF1d))]);
157 disp(['Subspecies A2: R^2 = ' num2str(CoefDet(Num2,exp_PDF2d))]);

```

Listing D.6 – 2nd order process $A + B \rightarrow X$ for a heterogeneous population.

```

1  % A + B --> C
2  % Simulating 2nd order kinetics using ABK.
3
4  % Heterogeneous Population! Assuming 2 subinteractions.
5  % Here, I evaluate each possible interaction with respect to B.
6  % (ie, I draw a random number for each possible B-A pair).
7  % Also, I randomize the order of sampled B-A interactions
8  % (see report section 4.1.4 for details).
9
10 clear; clc;
11 rng(0);
12
13 agents = 10;
14
15 maxTime = 200;
16 dt = 1/100; % Fixed time step increment
17 t_steps = ceil(maxTime / dt);
18
19 reps = 1000; % Repeat experiment reps times
20
21 % Initial number of reactant molecules; assume B is limiting.
22 Ao = agents; Bo = ceil(0.7*agents);
23 Co = 0; Cmax = Bo;
24
25 % ***** 2 Distinct Subinteractions *****
26 % ***** Set up k-matrix wrt B (size: Bo x Ao) *****
27 k_sub = [0.001, 0.009]; % units: 1/sec
28 subints = size(k_sub,2); % number of subinteractions
29 k1 = k_sub(1)*ones(Bo,Ao/2); k2 = k_sub(2)*ones(Bo,Ao/2);
30 k = [k1, k2];
31 ko = mean(mean(k)); % units: 1/sec
32 % *****
33
34 % ***** # Subints = Bo*Ao *****
35 % ***** Normally-Distributed k *****
36 % ***** Set up k-matrix wrt B (size: Bo x Ao) *****
37 % ko = 0.001; % Mean k value
38 % sigma = 0.0001; % standard deviation of k
39 %
40 % repeat = 1;
41 % while repeat == 1
42 % k = ko + sigma*randn(Bo,Ao); % Normally distributed with mean=ko and std=sigma
43 % repeat = 0;
44 % negk = find(k<=0);
45 % if isempty(negk)==0 % Ensure there are no negative k values
46 % repeat = 1;
47 % end
48 % end
49 % *****
50
51 extt1 = zeros(reps, Bo); % To monitor extinction time of 1st subinteraction
52 extt2 = zeros(reps, Bo); % To monitor extinction time of 2nd subinteraction
53
54 %% Preallocate memory; Initialize time-dependent sum of molecule numbers
55 At = zeros(reps,t_steps+1); Bt = zeros(reps,t_steps+1); Ct = zeros(reps,t_steps+1);
56
57 for n=1:reps
58
59     fprintf(1, '% ');
60
61     % Arrays for tracking reactant and product agent status
62     A = ones(1,Ao); B = ones(1,Bo); C = zeros(1,Cmax);
63     At(n,1) = sum(A); Bt(n,1) = sum(B); Ct(n,1) = sum(C);
64     time = zeros(1,t_steps);
65
66     for t=1:t_steps
67
68         tempA = find(A==1); % Find 'alive' agents A
69         tempB = find(B==1); % Find 'alive' agents B
70
71         tempA = RandArray(tempA); % Randomize order of alive A agents
72
73         for i = 1:size(tempB,2)
74             for j = 1:size(tempA,2) % Then see if it reacts with an A agent
75
76                 P = 1 - exp(- k(tempB(i),tempA(j)) * dt); % P_int
77                 P = k(tempB(i),tempA(j)) * dt; % P_dif
78
79                 if rand < P
80
81                     B(tempB(i)) = 0;
82                     A(tempA(j)) = 0;
83                     C(tempB(i)) = 1;
84
85                     % Assume two subinteractions:
86                     if tempA(j) <= Ao/2 % Extinction time for
87                         extt1(n,tempB(i)) = time(t)+dt; % 1st subinteraction
88                     else % Extinction time for

```

```

89             extt2(n,tempB(i)) = time(t)+dt; % 2nd subinteraction
90         end
91         break; % B(tempB(i)) agent has reacted in this time step.
92     end
93     end
94     end
95     end
96     end
97     end
98     end
99     At(n,t+1) = sum(A);    Bt(n,t+1) = sum(B);    Ct(n,t+1) = sum(C);
100    time(t+1) = time(t) + dt;
101
102    end
103
104    end % end 'for reps' loop
105
106    % Population statistics of time trajectories
107    avgA = mean(At);        avgB = mean(Bt);        avgC = mean(Ct);
108    sdevA = std(At);       sdevB = std(Bt);       sdevC = std(Ct);
109
110    % Solve ODE for 2nd order kinetics
111    [t_sol,y_sol] = ode45(@(t,y) o2_dif(t,y,ko),0:maxTime/100:maxTime,[Ao ; Bo ; Co]);
112
113    %% ** Group extinction times by subspecies ** - Works for subspecies = 2 simulation
114    % Remove zeros from extt (zeros represent agents that didn't die within totalTime)
115    f = size(extt1,2);
116    for z=1:f
117        extt1_zeros = find(extt1(:,z)==0);
118        for a = 1:size(extt1_zeros,1)
119            extt1(extt1_zeros(a),z) = NaN;
120        end
121    end
122
123    g = size(extt2,2);
124    for q=1:g
125        extt2_zeros = find(extt2(:,q)==0);
126        for b = 1:size(extt2_zeros,1)
127            extt2(extt2_zeros(b),q) = NaN;
128        end
129    end
130
131    extt1 = reshape(extt1,[],1); % Reshape to column vector
132    extt2 = reshape(extt2,[],1); % Reshape to column vector
133
134    %% Plot time trajectories
135    fig0 = figure('Name','2nd Order Rx Time course','NumberTitle','off');
136    fig0.PaperUnits = 'inches';
137    fig0.PaperPosition = [0 0 6 5]; % Control the size of printed (eps) fig.
138
139    plot(time,avgA,time,avgB); % plot(time,avgC); hold on;
140
141    % Plot DE Time Courses
142    plot(t_sol,y_sol(:,1),'c','LineWidth',2);
143    plot(t_sol,y_sol(:,2),'m','LineWidth',2);
144    % plot(t_sol,y_sol(:,3),'k','LineWidth',2);
145
146    p1_devA1 = plot(time,avgA+sdevA,'LineStyle','--','Color',[0.8 0.8 0.8],'LineWidth',2);
147    p1_devA0 = plot(time,avgA-sdevA,'LineStyle','--','Color',[0.8 0.8 0.8],'LineWidth',2);
148    p1_devB1 = plot(time,avgB+sdevB,'LineStyle','--','Color',[0.8 0.8 0.8],'LineWidth',2);
149    p1_devB0 = plot(time,avgB-sdevB,'LineStyle','--','Color',[0.8 0.8 0.8],'LineWidth',2);
150    % p1_devC1 = plot(time,avgC+sdevC,'LineStyle','--','Color',[0.8 0.8 0.8],'LineWidth',2);
151    % p1_devC0 = plot(time,avgC-sdevC,'LineStyle','--','Color',[0.8 0.8 0.8],'LineWidth',2);
152
153    title(['k_{BA-1} = ' num2str(k_sub(1)) ' sec^{-1} , k_{BA-2} = ' num2str(k_sub(2)) ' sec^{-1}'],...
154          'FontName','Times New Roman','FontSize',12,'FontWeight','Normal');
155
156    xlabel('t (sec)'); % ylabel('N(t)'); hold off;
157    axis([0 maxTime 0 agents]);
158    set(gca,'XMinorTick','on','Box','off');
159
160    leg0 = legend('ABK <N_A(t)>','ABK <N_B(t)>','DE N_A(t)','DE N_B(t)');
161    set(leg0,'FontName','Times New Roman','FontSize',9,...
162          'EdgeColor',[0.95 0.95 0.95],'Location','NorthEast');
163
164    %% Plot Extinction time distribution
165    bins = 20;
166    fig1 = figure;
167    fig1.PaperUnits = 'inches';
168    fig1.PaperPosition = [0 0 6 5]; % Control the size of printed (eps) fig.
169
170    [num1, c1] = hist(extt1,bins); % histogram data for subinteraction 1
171    [num2, c2] = hist(extt2,bins); % histogram data for subinteraction 2
172    Num1 = num1 / (reps*Bo); % Normalize
173    Num2 = num2 / (reps*Bo); % Normalize
174
175    % Generate values for PDF based on ODE solution.
176    timeD = 0:maxTime/bins:maxTime;
177    [t_sol1,y_sol1] = ode45(@(t,y) o2_dif(t,y,k_sub(1)),timeD,[Ao ; Bo ; Co]);
178    [t_sol2,y_sol2] = ode45(@(t,y) o2_dif(t,y,k_sub(2)),timeD,[Ao ; Bo ; Co]);
179    % dA = abs(diff(y_sol2(:,1))) / Ao;

```

```

180 dB1 = abs(diff(y_sol1(:,2))) / Bo / 2;
181 dB2 = abs(diff(y_sol2(:,2))) / Bo / 2;
182
183 timeDm = maxTime/bins/2:maxTime/bins:maxTime;
184
185 b = bar(timeDm,[Num1' Num2']);
186 set(b(1),'FaceColor','flat','EdgeColor',[0.70 0.78 1]); hold on;
187 set(b(2),'FaceColor',[1 0.1 0.1],'EdgeColor',[1 0.6 0.6]); % For subinteraction 1
188 % For subinteraction 2
189 plot(timeDm,dB1,'xb');
190 plot(timeDm,dB2,'or'); hold off;
191
192 axis([0 maxTime 0 0.3]);
193 set(gca,'Ytick',0:0.05:0.3,...
194 'YtickLabel',{'0' '0.05' '0.10' '0.15' '0.20' '0.25' '0.30'});
195 set(gca,'XMinorTick','on','YMinorTick','on','Box','off');
196 xlabel('t (sec)'); ylabel('Fraction of A+B\rightarrowX Transitions');
197
198 leg1 = legend(['Subinteraction BA-1, k_{BA-1} = ' num2str(k_sub(1)) ' sec^{-1}',...
199 ['Subinteraction BA-2, k_{BA-2} = ' num2str(k_sub(2)) ' sec^{-1}']);
200 set(leg1,'FontName','Times New Roman','FontSize',9,...
201 'EdgeColor',[0.95 0.95 0.95],'Location','NorthEast');
202
203 % Coefficient of Determination R^2 for histogram fit to distribution PDF.
204 disp(['Subint-1: R^2 = ' num2str(CoefDet(Num1,dB1'))]);
205 disp(['Subint-2: R^2 = ' num2str(CoefDet(Num2,dB2'))]);
206
207 err = [Num1' - dB1 , Num2' - dB2]; % calculate Deviation
208
209 ax2 = axes('Parent',fig1,'Position',[0.62 0.45 0.25 0.25],'FontSize',8);
210 p_res = plot(timeDm,err,'.','Parent',ax2); hold on;
211 set(p_res(2),'Color','r');
212 plot(timeDm,zeros(1,size(timeDm,2)),'--','Color',[0.5 0.5 0.5],'Parent',ax2);
213
214 axis([0 maxTime -0.04 0.04]);
215 set(gca,'XMinorTick','on','YMinorTick','on','Box','on');
216 xlabel('t (sec)'); ylabel('Deviation');

```

Listing D.7 – Function for fitting a curve to the Hill equation.

```

1 function [estimates, Rj] = FitCurve_Hill(xdata, ydata)
2 % This function fits the data to the Hill equation and the results
3 % are printed on standard output (A graph is NOT produced). . .
4 % Parameters n (the Hill coefficient, nH) and K are reported.
5 % ** Input arguments xdata, ydata, must be row vectors. **
6
7 % Call fminsearch with a random starting point.
8 start_point = rand(2);
9 model = @fitfun;
10 estimates = fminsearch(model, start_point);
11
12 % fitfun accepts curve parameters as inputs, and outputs sse,
13 % the sum of squares error for [xdata.^n / (xdata.^n + K)] - ydata,
14 % and the FittedCurve. FMINSEARCH only needs sse.
15
16     function [sse, FittedCurve] = fitfun(params)
17         n = params(1);
18         K = params(2);
19         FittedCurve = xdata.^n ./ (xdata.^n + K);
20         ErrorVector = FittedCurve - ydata;
21         sse = sum(ErrorVector.^2);
22     end
23
24 % Calculate Correlation Coefficient
25 Fit = xdata.^estimates(1) ./ (xdata.^estimates(1) + estimates(2));
26 Compare = [(ydata)' (Fit)'];
27 [r p] = corrcoef(Compare);
28 R = r(1,2);          p_value = p(1,2);
29
30 % Calculate sensitivity coefficient (Rj): Ratio of concentration
31 % of effector needed to produce a 90% response to that producing
32 % a 10% response.
33 x90 = (9 * estimates(2)).^(1/estimates(1));
34 x10 = (1/9 * estimates(2)).^(1/estimates(1));
35 Rj = x90 / x10;
36
37 % Print results to std output
38 disp(['nH=' num2str(estimates(1)) ', K=' num2str(estimates(2)) ...
39      ', R_j=' num2str(Rj)]);
40 disp(['R=' num2str(R) ', p=' num2str(p_value)]);
41
42 end
43
44 % This function can serve as a template for fitting data to a
45 % user-specified equation.

```

Bibliography

- Agate, R. J., Grisham, W., Wade, J., Mann, S., Wingfield, J., Schanen, C., Palotie, A., and Arnold, A. P. 2003. Neural, not gonadal, origin of brain sex differences in a gynandromorphic finch. *Proceedings of the National Academy of Sciences*, 100(8):4873–4878.
- Albayrak, C., Jordi, C. A., Zechner, C., Lin, J., Bichsel, C. A., Khammash, M., and Tay, S. 2016. Digital quantification of proteins and mRNA in single mammalian cells. *Molecular Cell*, 61(6):914–924.
- Apostol, T. M. 1967. *Calculus, Vol. I*. John Wiley and Sons, Inc., 2nd ed. edition.
- Arkin, A., Ross, J., and McAdams, H. H. 1998. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics*, 149(4):1633–1648.
- Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L., and Leibler, S. 2004. Bacterial persistence as a phenotypic switch. *Science*, 305(5690):1622–1625.
- Becksei, A., Séraphin, B., and Serrano, L. 2001. Positive feedback in eukaryotic gene networks: Cell differentiation by graded to binary response conversion. *EMBO Journal*, 20(10):2528–2535.
- Bratsun, D. A., Volfson, D. N., Hasty, J., and Tsimring, L. S. 2005. Non-markovian processes in gene regulation. In *Noise in Complex Systems and Stochastic Dynamics III*, volume 5845, pages 210–219.
- Bray, D. 1995. Protein molecules as computational elements in living cells. *Nature*, 376(6538):307–312.
- Cagatay, T., Turcotte, M., Elowitz, M. B., Garcia-Ojalvo, J., and Süel, G. M. 2009. Architecture-dependent noise discriminates functionally analogous differentiation circuits. *Cell*, 139(3):512–522.
- Cai, C. R., Wu, Z. X., and Guan, J. Y. 2013. Behavior of susceptible-vaccinated-infected-recovered epidemics with diversity in the infection rate of individuals. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 88(6):062805.

- Charlebois, D. A., Intosalmi, J., Fraser, D., and Kærn, M. 2011. An algorithm for the stochastic simulation of gene expression and heterogeneous population dynamics. *Communications in Computational Physics*, 9(01):89–112.
- Di Talia, S., Skotheim, J. M., Bean, J. M., Siggia, E. D., and Cross, F. R. 2007. The effects of molecular noise and size control on variability in the budding yeast cell cycle. *Nature*, 448(7156):947–951.
- Domijan, M., Brown, P. E., Shulgin, B. V., and Rand, D. A. 2016. PeTTSy: a computational tool for perturbation analysis of complex systems biology models. *BMC Bioinformatics*, 17(124):1–19.
- Donnan, F. G. 1927. Concerning the applicability of thermodynamics to the phenomena of life. *The Journal of General Physiology*, 8(6):685–688.
- Drawert, B., Hellander, A., Bales, B., Banerjee, D., Bellesia, G., Daigle, B. J., Douglas, G., Gu, M., Gupta, A., Hellander, S., Horuk, C., Nath, D., Takkar, A., Wu, S., Lötstedt, P., Krintz, C., and Petzold, L. R. 2016. Stochastic Simulation Service: Bridging the gap between the computational expert and the biologist. *PLOS Computational Biology*, 12(12):e1005220.
- El-Samad, H. and Khammash, M. 2006. Coherence resonance: a mechanism for noise induced stable oscillations in gene regulatory networks. In *Decision and Control, 45th IEEE Conference*, pages 2382–2387. IEEE.
- Eldar, A. and Elowitz, M. B. 2010. Functional roles for noise in genetic circuits. *Nature*, 467(7312):167–173.
- Elowitz, M. B. and Leibler, S. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338.
- Epstein, J. M. 2008. Why model? *Journal of Artificial Societies and Social Simulation*, 11(4):12.
- Ferrell, J. E. 1996. Tripping the switch fantastic: How a protein kinase cascade can convert graded inputs into switch-like outputs. *Trends in Biochemical Sciences*, 21(12):460–466.
- Ferrell, J. E. 2016. Perfect and near-perfect adaptation in cell signaling. *Cell Systems*, 2(2):62–67.
- Ferrell, J. E. and Ha, S. H. 2014. Ultrasensitivity part I: Michaelian responses and zero-order ultrasensitivity. *Trends in Biochemical Sciences*, 39(10):496–503.
- Ferrell, J. E. and Machleder, E. M. 1998. The biochemical basis of an all-or-none cell fate switch in *Xenopus* oocytes. *Science*, 280(5365):895–898.

- Fredrick, N. D., Berges, J. A., Twining, B. S., Nuñez-Milland, D., and Hellweger, F. L. 2013. Use of agent-based modeling to explore the mechanisms of intracellular phosphorus heterogeneity in cultured phytoplankton. *Applied and Environmental Microbiology*, 79(14):4359–4368.
- Gardner, M. 1970. Mathematical games: The fantastic combinations of John Conway's new solitaire game "Life". *Scientific American*, 223(October):120–123.
- Gardner, T. S., Cantor, C. R., and Collins, J. J. 2000. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403(6767):339–342.
- Geva-Zatorsky, N., Dekel, E., Batchelor, E., Lahav, G., and Alon, U. 2010. Fourier analysis and systems identification of the p53 feedback loop. *Proceedings of the National Academy of Sciences of the United States of America*, 107(30):13550–13555.
- Gillespie, D. T. 1977. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361.
- Gillespie, D. T. 1992. A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications*, 188(1-3):404–425.
- Gillespie, D. T. 2007. Stochastic simulation of chemical kinetics. *Annual Review of Physical Chemistry*, 58:35–55.
- Gillespie, D. T., Hellander, A., and Petzold, L. R. 2013. Perspective: Stochastic algorithms for chemical kinetics. *Journal of Chemical Physics*, 138(17):170901.
- Glasserman, P. and Yao, D. D. 1992. Some guidelines and guarantees for common random numbers. *Management Science*, 38(6):884–908.
- Goldbeter, A. and Koshland, D. E. 1981. An amplified sensitivity arising from covalent modification in biological systems. *Proceedings of the National Academy of Sciences*, 78(11):6840–6844.
- Gonzalez-Garcia, I., Sole, R. V., and Costa, J. 2002. Metapopulation dynamics and spatial heterogeneity in cancer. *Proceedings of the National Academy of Sciences*, 99(20):13085–13089.
- Gorochofski, T. E. 2016. Agent-based modelling in synthetic biology. *Essays In Biochemistry*, 60(4):325–336.
- Griffith, J. S. 1968. Mathematics of cellular control Processes: I. Negative feedback to one gene. *Journal of Theoretical Biology*, 20:202–208.
- Gunawardena, J. 2012. Some lessons about models from Michaelis and Menten. *Molecular Biology of the Cell*, 23(4):517–519.

- Halling, P. J. 1989. Do the laws of chemistry apply to living cells? *Trends in Biochemical Sciences*, 14(8):317–318.
- Hellweger, F. L. and Bucci, V. 2009. A bunch of tiny individuals - Individual-based modeling for microbes. *Ecological Modelling*, 220(1):8–22.
- Hellweger, F. L. and Kianirad, E. 2007. Accounting for intrapopulation variability in biogeochemical models using agent-based methods. *Environmental Science and Technology*, 41(8):2855–2860.
- Hill, A. V. 1910. The possible effects of the aggregation of the molecules of hæmoglobin on its dissociation curves. *The Journal of Physiology*, 40(Supplement - Proceedings of the Physiological Society):iv–vii.
- Hoops, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., and Kummer, U. 2006. COPASI - A COMplex PATHway SIMulator. *Bioinformatics*, 22(24):3067–3074.
- Hung, H.-C., Chiu, Y.-C., Huang, H.-C., and Wu, M.-C. 2017. An enhanced application of Lotka-Volterra model to forecast the sales of two competing retail formats. *Computers and Industrial Engineering*, 109:325–334.
- Ishida, K. 1964. Stochastic model for bimolecular reaction. *The Journal of Chemical Physics*, 41(8):2472–2478.
- Ju, Y. S., Martincorena, I., Gerstung, M., Petljak, M., Alexandrov, L. B., Rahbari, R., Wedge, D. C., Davies, H. R., Ramakrishna, M., Fullam, A., Martin, S., Alder, C., Patel, N., Gamble, S., O'Meara, S., Giri, D. D., Sauer, T., Pinder, S. E., Purdie, C. A., Borg, Å., Stunnenberg, H., Van De Vijver, M., Tan, B. K., Caldas, C., Tutt, A., Ueno, N. T., Van 'T Veer, L. J., Martens, J. W., Sotiriou, C., Knappskog, S., Span, P. N., Lakhani, S. R., Eyfjörd, J. E., Børresen-Dale, A. L., Richardson, A., Thompson, A. M., Viari, A., Hurles, M. E., Nik-Zainal, S., Campbell, P. J., and Stratton, M. R. 2017. Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. *Nature*, 543(7647):714–718.
- Kaplan, D. and Glass, L. 1995. *Understanding Nonlinear Dynamics*. Springer-Verlag.
- Karr, J. R., Takahashi, K., and Funahashi, A. 2015. The principles of whole-cell modeling. *Current Opinion in Microbiology*, 27:18–24.
- Kendall, D. G. 1950. An artificial realization of a simple "birth-and-death" process. *Journal of the Royal Statistical Society. Series B (Methodological)*, 12(1):116–119.
- Kermack, W. O. and McKendrick, A. G. 1927. Contributions to the mathematical theory of epidemics. *Proceedings of the Royal Society of London*, 115(772):700–721.

- Kolch, W., Halasz, M., Granovskaya, M., and Kholodenko, B. N. 2015. The dynamic control of signal transduction networks in cancer cells. *Nature Reviews Cancer*, 15(9):515–527.
- Krebs, C. J. 1972. *Ecology: The Experimental Analysis of Distribution and Abundance*. Harper and Row, New York.
- Kussell, E. and Leibler, S. 2005. Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, 309(5743):2075–2078.
- LaPorte, D. C. and Koshland, D. E. 1983. Phosphorylation of isocitrate dehydrogenase as a demonstration of enhanced sensitivity in covalent regulation. *Nature*, 305(5932):286–290.
- Lau, M. S. Y., Dalziel, B. D., Funk, S., McClelland, A., Tiffany, A., Riley, S., Metcalf, C. J. E., and Grenfell, B. T. 2017. Spatial and temporal dynamics of superspreading events in the 2014–2015 West Africa Ebola epidemic. *Proceedings of the National Academy of Sciences*, 114(9):2337–2342.
- Le Novère, N. and Shimizu, T. S. 2001. STOCHSIM: modelling of stochastic biomolecular processes. *Bioinformatics*, 17(6):575–576.
- Lecca, P., Laurenzi, I., and Jordan, F. 2013a. Simulation of ecodynamics: key nodes in food webs. In *Deterministic Versus Stochastic Modelling in Biochemistry and Systems Biology*, chapter 9, pages 348–372.
- Lecca, P., Laurenzi, I., and Jordan, F. 2013b. The stochastic approach to biochemical kinetics. In *Deterministic Versus Stochastic Modelling in Biochemistry and Systems Biology*, chapter 2, pages 35–82.
- Levin, S. A. 1998. Ecosystems and the biosphere as complex adaptive systems. *Ecosystems*, 1(5):431–436.
- Lim, E. T., Uddin, M., De Rubeis, S., Chan, Y., Kamumbu, A. S., Zhang, X., D’Gama, A. M., Kim, S. N., Hill, R. S., Goldberg, A. P., Poultney, C., Minshew, N. J., Kushima, I., Aleksic, B., Ozaki, N., Parellada, M., Arango, C., Penzol, M. J., Carracedo, A., Kolevzon, A., Hultman, C. M., Weiss, L. A., Fromer, M., Chiocchetti, A. G., Freitag, C. M., Church, G. M., Scherer, S. W., Buxbaum, J. D., and Walsh, C. A. 2017. Rates, distribution and implications of postzygotic mosaic mutations in autism spectrum disorder. *Nature Neuroscience*, 20(9):1217–1224.
- Liu, Y., Beyer, A., and Aebersold, R. 2016. On the dependency of cellular protein levels on mRNA abundance. *Cell*, 165(3):535–550.
- Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E., and Getz, W. M. 2005. Super-spreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359.

- Lodato, M. A., Woodworth, M. B., Lee, S., Evrony, G. D., Mehta, B. K., Karger, A., Lee, S., Chittenden, T. W., D’Gama, A. M., Cai, X., Luquette, L. J., Lee, E., Park, P. J., and Walsh, C. A. 2015. Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science*, 350(6256):94–98.
- Lotka, A. J. 1925. *Elements of Physical Biology*. Williams and Wilkins, Baltimore.
- Loveland, W., Morrissey, D., and Seaborg, G. T. 2006. *Modern Nuclear Chemistry*. Wiley-Interscience.
- Lumsden, C. J. and Rowe, G. W. 1984. Macroscopic behavior in finite Brusselators: Stochastic simulation of fluctuations and ensemble statistics. *The Journal of Chemical Physics*, 81(4):1851–1860.
- Macal, C. M. and North, M. J. 2010. Tutorial on agent-based modelling and simulation. *Journal of Simulation*, 4(3):151–162.
- Mantzaris, N. V. 2006. Stochastic and deterministic simulations of heterogeneous cell population dynamics. *Journal of Theoretical Biology*, 241(3):690–706.
- Marusyk, A. and Polyak, K. 2010. Tumor heterogeneity: Causes and consequences. *Biochimica et Biophysica Acta - Reviews on Cancer*, 1805(1):105–117.
- McDonnell, M. D. and Abbott, D. 2009. What is stochastic resonance? Definitions, misconceptions, debates, and its relevance to biology. *PLoS Computational Biology*, 5(5):e1000348.
- McKane, A. J., Nagy, J. D., Newman, T. J., and Stefanini, M. O. 2007. Amplified biochemical oscillations in cellular systems. *Journal of Statistical Physics*, 128(1-2):165–191.
- McKane, A. J. and Newman, T. J. 2005. Predator-prey cycles from resonant amplification of demographic stochasticity. *Physical Review Letters*, 94(21):1–4.
- McQuarrie, D. A. 1967. Stochastic Approach to Chemical Kinetics. *Journal of Applied Probability*, 4(3):413–478.
- Michaelis, L. and Menten, M. L. 1913. Die Kinetik der Invertinwirkung. *Biochem Z*, 49(February):333–369.
- Milo, R., Jorgensen, P., Moran, U., Weber, G., and Springer, M. 2010. BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic acids research*, 38(Database issue):D750–753.

- Morton-Firth, C. J. and Bray, D. 1998. Predicting temporal fluctuations in an intracellular signalling pathway. *Journal of Theoretical Biology*, 192(1):117–128.
- Nguyen, A., Yoshida, M., Goodarzi, H., and Tavazoie, S. F. 2016. Highly variable cancer subpopulations that exhibit enhanced transcriptome variability and metastatic fitness. *Nature Communications*, 7(1):11246.
- Novák, B. and Tyson, J. J. 2008. Design principles of biochemical oscillators. *Nature Reviews. Molecular Cell Biology*, 9(12):981–991.
- Paulsson, J., Berg, O. G., and Ehrenberg, M. 2000. Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation. *Proceedings of the National Academy of Sciences*, 97(13):7148–7153.
- Pogson, M., Smallwood, R., Qvarnstrom, E., and Holcombe, M. 2006. Formal agent-based modelling of intracellular chemical interactions. *Biosystems*, 85(1):37–45.
- Priest, J. R., Gawad, C., Kahlig, K. M., Yu, J. K., O’Hara, T., Boyle, P. M., Rajamani, S., Clark, M. J., Garcia, S. T. K., Ceresnak, S., Harris, J., Boyle, S., Dewey, F. E., Malloy-Walton, L., Dunn, K., Grove, M., Perez, M. V., Neff, N. F., Chen, R., Maeda, K., Dubin, A., Belardinelli, L., West, J., Antolik, C., Macaya, D., Quertermous, T., Trayanova, N. A., Quake, S. R., and Ashley, E. A. 2016. Early somatic mosaicism is a rare cause of long-QT syndrome. *Proceedings of the National Academy of Sciences*, 113(41):11555–11560.
- Prigogine, I. and Lefever, R. 1968. Symmetry breaking instabilities in dissipative systems. II. *Journal of Chemical Physics*, 48(4):1695–1700.
- Raj, A. and van Oudenaarden, A. 2008. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135(2):216–226.
- Rao, C. V. and Arkin, A. P. 2003. Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. *Journal of Chemical Physics*, 118(11):4999–5010.
- Rathinam, M., Sheppard, P. W., and Khammash, M. 2010. Efficient computation of parameter sensitivities of discrete stochastic chemical reaction networks. *The Journal of Chemical Physics*, 132(3):034103.
- Roy, A. and Klumpp, S. 2018. Simulating Genetic Circuits in Bacterial Populations with Growth Heterogeneity. *Biophysical Journal*, 114(2):484–492.
- Rudge, T. J., Steiner, P. J., Phillips, A., and Haseloff, J. 2012. Computational modeling of synthetic microbial biofilms. *ACS Synthetic Biology*, 1(8):345–352.
- Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H., Nes, E. H. V., Rietkerk, M., and Sugihara, G. 2009. Early-warning signals for critical transitions. *Nature*, 461(7260):53–59.

- Scheffer, M., Carpenter, S. R., Lenton, T. M., Bascompte, J., Brock, W., Dakos, V., Van De Koppel, J., Van De Leemput, I. A., Levin, S. A., Van Nes, E. H., Pascual, M., and Vandermeer, J. 2012. Anticipating critical transitions. *Science*, 338(6105):344–348.
- Segel, L. A. and Slemrod, M. 1989. The quasi-steady-state assumption: A case study in perturbation. *SIAM Review*, 31(3):446–477.
- Sharma, S. V., Lee, D. Y., Li, B., Quinlan, M. P., Takahashi, F., Maheswaran, S., McDermott, U., Azizian, N., Zou, L., Fischbach, M. A., Wong, K. K., Brandstetter, K., Wittner, B., Ramaswamy, S., Classon, M., and Settleman, J. 2010. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell*, 141(1):69–80.
- Simpson, E. H. 1949. Measurement of diversity. *Nature*, 163(4148):688.
- Smadbeck, P. and Kaznessis, Y. N. 2013. A closure scheme for chemical master equations. *Proceedings of the National Academy of Sciences*, 110(35):14261–14265.
- Smadbeck, P. and Kaznessis, Y. N. 2014. Solution of chemical master equations for nonlinear stochastic reaction networks. *Current Opinion in Chemical Engineering*, 5(1):90–95.
- Stefan, M. I., Bartol, T. M., Sejnowski, T. J., and Kennedy, M. B. 2014. Multi-state modeling of biomolecules. *PLoS Computational Biology*, 10(9):e1003844.
- Strogatz, S. H. 2015. *Nonlinear Dynamics and Chaos*. Westview Press, 2nd ed. edition.
- Süel, G. M., Garcia-Ojalvo, J., Liberman, L. M., and Elowitz, M. B. 2006. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature*, 440(7083):545–550.
- Swanton, C., McGranahan, N., Starrett, G. J., and Harris, R. S. 2015. APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. *Cancer Discovery*, 5(7):704–712.
- Taniguchi, Y., Choi, P. J., Li, G. W., Chen, H., Babu, M., Hearn, J., Emili, A., and Sunney Xie, X. 2010. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 329(5991):533–538.
- Thomas, P., Straube, A. V., Timmer, J., Fleck, C., and Grima, R. 2013. Signatures of nonlinearity in single cell noise-induced oscillations. *Journal of Theoretical Biology*, 335:222–234.
- Tsimring, L. S. 2014. Noise in biology. *Reports on Progress in Physics*, 77(2):026601.

- Turner, T. E., Schnell, S., and Burrage, K. 2004. Stochastic approaches for modelling in vivo reactions. *Computational Biology and Chemistry*, 28(3):165–178.
- Tyson, J. J. 1973. Some further studies of nonlinear oscillations in chemical systems. *The Journal of Chemical Physics*, 58(9):3919–3930.
- Tyson, J. J., Chen, K. C., and Novak, B. 2003. Sniffers, buzzers, toggles and blinkers: Dynamics of regulatory and signaling pathways in the cell. *Current Opinion in Cell Biology*, 15(2):221–231.
- Van Nes, E. H. and Scheffer, M. 2005. Implications of spatial heterogeneity for regime shifts in ecosystems. *Ecology*, 86(7):1797–1807.
- Vargas-Garcia, C. A., Torre, E. A., Herlyn, M., Krepler, C., Nathanson, K. L., Emert, B., Raj, A., Brafford, P. A., Singh, A., Beqiri, M., Sproesser, K., Eggan, E., Anastopoulos, I. N., Torborg, S. R., Xiao, M., Shaffer, S. M., and Dunagin, M. C. 2017. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature*, 546(7658):431–435.
- Vestergaard, C. L. and Géniois, M. 2015. Temporal Gillespie algorithm: Fast simulation of contagion processes on time-varying networks. *PLOS Computational Biology*, 11(10):e1004579.
- Vilar, J. M. G., Kueh, H. Y., Barkai, N., and Leibler, S. 2002. Mechanisms of noise-resistance in genetic oscillators. *Proceedings of the National Academy of Sciences*, 99(9):5988–5992.
- Volterra, V. 1926. Fluctuations in the abundance of a species considered mathematically. *Nature*, 118(2972):558–560.
- Wagner, G. P. and Altenberg, L. 1996. Complex adaptations and the evolution of evolvability. *Evolution*, 50(3):967–976.
- Winfree, A. T. 1984. The prehistory of the Belousov-Zhabotinsky oscillator. *Journal of Chemical Education*, 61(8):661.

Nomenclature

ABK	Agent-Based Kinetics
ABM	Agent-Based Modeling
CDF	Cumulative Distribution Function
CME	Chemical Master Equation
CRN	Common Random Numbers procedure
DDE	Delay Differential Equation
iFFL	Incoherent Feedforward Loop
KCF	Kinetic Complementarity Factor (symbol: Ω)
ODE	Ordinary Differential Equation
PDF	Probability Density Function
PHAF	Persistent High-Amplitude Fluctuations
PHM	Population Heterogeneity Matrix
RSF	Regulation Scaling Function (symbol: F)
SR	Signal-Response
SSA	Stochastic Simulation Algorithm (Gillespie's algorithm)
SSR	Sum of Squared Residuals