

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Mathematical Modeling and Analysis of Stochastic Chemical Reaction Networks Motivated by Applications to Epigenetic Cell Memory

Permalink

<https://escholarship.org/uc/item/9bz3m4n8>

Author

Fu, Yi

Publication Date

2025

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Mathematical Modeling and Analysis of Stochastic Chemical Reaction Networks
Motivated by Applications to Epigenetic Cell Memory

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Bioinformatics and Systems Biology

by

Yi Fu

Committee in charge:

Professor Ruth J. Williams, Chair
Professor James T. Kadonaga, Co-Chair
Professor Domitilla Del Vecchio
Professor Jeff M. Hasty
Professor Nathan E. Lewis
Professor Bo Li

2025

Copyright

Yi Fu, 2025

All rights reserved.

The Dissertation of Yi Fu is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2025

DEDICATION

To my parents.

TABLE OF CONTENTS

Dissertation Approval Page	iii
Dedication	iv
Table of Contents	v
List of Figures	viii
Acknowledgements	x
Vita	xii
Abstract of the Dissertation	xiii
Chapter 1 Introduction	1
1.1 Overview	1
1.1.1 Comparison Theorems	3
1.1.2 Singular Perturbations	3
1.1.3 Coclique Level Structure	4
1.2 Structure of the dissertation	5
1.3 Acknowledgement	5
Bibliography for Chapter 1	6
Chapter 2 Comparison Theorems	9
2.1 Abstract	9
2.2 Introduction	10
2.2.1 Overview	10
2.2.2 Related work	11
2.2.3 Notation and terminology	12
2.3 Stochastic Chemical Reaction Networks	13
2.4 Main Results	16
2.4.1 Preorders in $\mathbb{R}^d$	16
2.4.2 Stochastic comparison theorems	18
2.4.3 Monotonicity properties for (mean) first passage times and stationary distributions	25
2.5 Examples	29
2.6 Proofs of the Main Results	48
2.6.1 Proof of Theorem 2.4.1.	48
2.6.2 Proof of Theorem 2.4.2.	58
2.6.3 Proof of Theorem 2.4.3.	59
2.7 Conclusion	64
2.8 Supplementary Information (SI)	66

2.8.1	Criteria for positive recurrence and exponential ergodicity with application to Examples 2.5.2 and 2.5.5	66
2.8.2	Derivation of Markov chain transition directions, v_j	71
2.8.3	A generalization of Theorem 2.4.3	72
2.8.4	An algorithm for coupled stochastic simulation	82
2.9	Acknowledgement	85
	Bibliography for Chapter 2	85
Chapter 3	Singular Perturbations	89
3.1	Abstract	89
3.2	Introduction	90
3.2.1	Background	90
3.2.2	Focus of our work	90
3.2.3	Related work	91
3.2.4	Outline	93
3.2.5	Preliminaries and notation	93
3.3	Motivating Example: Chromatin Modification Circuit	95
3.3.1	1D model	97
3.3.2	2D model	102
3.4	Basic Setup and Definitions	104
3.4.1	Singularly perturbed, finite state, continuous time Markov chains ..	105
3.4.2	Stochastic Chemical Reaction Networks	108
3.5	Main Results: Stationary Distributions	110
3.5.1	The zeroth order term	110
3.5.2	Illustrative examples: 1D and 2D model	116
3.6	Main Results: Mean First Passage Times (MFPTs)	120
3.6.1	Algorithm to find the order of the poles for MFPTs	120
3.6.2	Leading coefficient in MFPT series expansion	122
3.6.3	Illustrative examples: 1D and 2D models	123
3.7	Main Results: Monotonic Dependence on Parameters	126
3.7.1	Comparison theorems for continuous time Markov chains	127
3.7.2	Illustrative example: 2D model	129
3.8	Further Applications	131
3.8.1	3D chromatin modification circuit model	131
3.8.2	4D chromatin modification circuit model	137
3.9	Conclusion	143
3.10	Supplementary Information (SI)	145
3.10.1	Some results in probability	145
3.10.2	Additional results for stationary distribution	151
3.10.3	Algorithm to find the order of the pole of the MFPT	157
3.10.4	Graphs for the algorithm to find the order of the MFPT	158
3.10.5	Justification for the algorithm to find the order of the MFPT	163
3.10.6	Application of the algorithm to the 2D, 3D and 4D models	177
3.10.7	Leading coefficient for the MFPT	188

3.10.8	1D Model: additional mathematical details	192
3.10.9	2D Model: additional mathematical details	196
3.10.10	3D Model: additional mathematical details	199
3.10.11	4D Model: additional mathematical details	206
3.11	Acknowledgement	217
	Bibliography for Chapter 3	217
Chapter 4	Coclique Level Structure	222
4.1	Abstract	222
4.2	Introduction	223
4.2.1	Overview	223
4.2.2	Related work	224
4.2.3	Terminology and Notation	225
4.3	Stochastic Chemical Reaction Networks (SCRNs)	226
4.4	Coclique level structure for SCRNs	228
4.5	Main Results	235
4.5.1	Existence and characterization of coclique level structures	235
4.5.2	Algorithm to determine all coclique level functions associated with a SCRN satisfying Assumption 4.4.1 and whose associated graph $\mathcal{G}$ is weakly connected	241
4.5.3	Using coclique level structure to bound MFPTs	242
4.5.4	Generalization to the non-weakly-connected case	247
4.6	Examples	254
4.6.1	Chromatin modification circuit including only histone modifications	254
4.6.2	Full chromatin modification circuit	259
4.6.3	Bi-parallel network motif	265
4.7	Conclusion	269
4.8	Supplementary Information (SI)	270
4.8.1	Proof of Lemma 4.4.1	270
4.8.2	Proof of Theorem 4.5.4	271
4.8.3	One-dimensional birth-death process: Mean first passage time.	273
4.8.4	Theorems 3.3 and 3.4 from Campos et al. [8]	275
4.9	Acknowledgement	278
	Bibliography for Chapter 4	278

LIST OF FIGURES

Figure 2.1.	Pictorial representation of conditions (2.4.4), (2.4.5) for a certain $(K_A + x) \cap \mathcal{X}$ in a two-dimensional lattice.	20
Figure 2.2.	Reaction model and corresponding Markov chain for enzymatic kinetics I example.	30
Figure 2.3.	Reaction model and corresponding Markov chain for enzymatic kinetics II example.	35
Figure 2.4.	Circuit inspired by Braess' paradox and corresponding Markov chain.	39
Figure 2.5.	Histone modification circuit and corresponding Markov chain.	42
Figure 2.6.	Histone modification circuit with positive TF-enabled autoregulation and corresponding Markov chain.	46
Figure 2.7.	A typical coupled realization of sample paths for Example 2.5.3.	81
Figure 3.1.	1D model.	98
Figure 3.2.	Histograms for the stationary distribution $\pi(\varepsilon)$ of the Markov chain X^ε.	100
Figure 3.3.	2D model.	103
Figure 3.4.	Graphs for the one-step transitions of X^ε, $\tilde{X}$ and $\hat{X}_A$ for the (a) 1D model and (b) 2D model.	115
Figure 3.5.	Key steps of the algorithm for the 1D model.	124
Figure 3.6.	3D Model and associated Markov chain.	132
Figure 3.7.	4D Model and associated Markov chain.	138
Figure 3.8.	Key steps of the algorithm for the 2D model.	183
Figure 3.9.	Key steps of the algorithm for the 3D model (from the fully active state to the fully repressed state).	184
Figure 3.10.	Key steps of the algorithm for the 3D model (from the fully repressed state to the fully active state).	185

Figure 3.11.	Key steps of the algorithm for the 4D model (from the fully active state to the fully repressed state).	186
Figure 3.12.	Key steps of the algorithm for the 4D model (from the fully repressed state to the fully active state).	187
Figure 4.1.	Three-species cascade motif: reaction diagram, graph $\mathcal{G}$ and associated Markov chain (Example 4.4.1).	233
Figure 4.2.	Key steps of the algorithm for identifying all coclique level functions for the projected continuous time Markov chain $\tilde{X}$ associated with a SCRN under Assumption 4.4.1 and assuming that $\mathcal{G}$ is weakly connected.	241
Figure 4.3.	Example of a SCRN whose associated graph $\mathcal{G}$ is not weakly connected (Example 4.5.2).	248
Figure 4.4.	Chromatin modification circuit including only histone modifications: reaction diagram, graph $\mathcal{G}$ and associated Markov chain.	256
Figure 4.5.	Full chromatin modification circuit: reaction diagram, graph $\mathcal{G}$ and associated Markov chain.	261
Figure 4.6.	Bi-parallel network motif: reaction diagram, graph $\mathcal{G}$ and associated Markov chain.	266

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my doctoral advisor, Professor Ruth J. Williams, for her invaluable advice, generous support and continuous guidance throughout the course of my PhD endeavor. I would like to extend my deepest thanks to my doctoral co-advisor, Professor James T. Kadonaga, for his expertise in biology and guidance on professional development. I am very grateful to Professor Domitilla Del Vecchio, Professor Jeff M. Hasty, Professor Nathan E. Lewis and Professor Bo Li for serving on my dissertation committee. My sincere thanks also goes to my wonderful collaborators, Dr. Simone Bruno and Dr. Felipe A. Campos. Last but not least, I would like to thank many people I came across along my life and academic journey.

Chapter 2 is reproduced with minor enumeration and formatting differences from the paper “Comparison theorems for stochastic chemical reaction networks” by Campos, F.A., Bruno, S., Fu, Y., Del Vecchio, D. and Williams, R.J., which has appeared in the *Bulletin of Mathematical Biology*. The dissertation author proposed and proved a generalization of the basic comparison theorem, developed a novel example inspired by Braess’ paradox, provided many details for examples used to illustrate the theory in the paper, and contributed to the writing of the manuscript.

Chapter 3 is extracted, with suitable enumeration and formatting differences, from the paper “Analysis of singularly perturbed stochastic chemical reaction networks motivated by applications to epigenetic cell memory” by Bruno, S., Campos, F.A., Fu, Y., Del Vecchio, D. and Williams, R.J., which has appeared in the *SIAM Journal on Applied Dynamical Systems*. The dissertation author developed systematic methodology to determine leading orders and leading coefficients for asymptotic expansions of mean first passage times, assisted in deriving higher order terms for stationary distribution series expansions, provided many details for examples used to illustrate the theory in the paper, and contributed to the writing of the manuscript.

Chapter 4 is reproduced with minor enumeration and formatting differences from

the paper “Coclique Level Structure for Stochastic Chemical Reaction Networks.” by Bruno, S., Fu, Y., Campos, F.A., Del Vecchio, D. and Williams, R.J., which has been submitted for publication and is currently under review. The dissertation author is an equal first author with S. Bruno on this paper.

This research was supported in part by NSF Collaborative Research grant MCB-2027947 (PI: R.J.W.) and by the Charles Lee Powell Foundation (PI: R.J.W.). This PhD was supported in part by the Bioinformatics and Systems Biology Block Grant.

VITA

- 2020 Bachelor of Science, University of California San Diego
Distinction in Biology with Specialization in Bioinformatics
High Distinction in Applied Mathematics
- 2021–2024 Teaching Assistant, University of California San Diego
- 2020–2025 Graduate Student Researcher, University of California San Diego
- 2025 Doctor of Philosophy, University of California San Diego
Bioinformatics and Systems Biology Graduate Program

PUBLICATIONS

- A.R. Massarat, A. Sen, J. Jauregui, S.T. Tyndale, Y. Fu, G. Erikson, G. McVicker (2021). Discovering Single Nucleotide Variants and Indels from Bulk and Single-Cell ATAC-Seq. *Nucleic Acids Research*, vol. 49, no. 14, pp. 7986-7994.
- E. Evert, Y. Fu, J.W. Helton, J. Yin (2021). Empirical Properties of Optima in Free Semidefinite Programs. *Experimental Mathematics*, vol. 32, no. 3, pp. 1-25.
- F.A. Campos, S. Bruno, Y. Fu, D. Del Vecchio, R.J. Williams (2023). Comparison Theorems for Stochastic Chemical Reaction Networks. *Bulletin of Mathematical Biology*, vol. 85, no. 39, pp. 1-41.
- S. Bruno, F.A. Campos, Y. Fu, D. Del Vecchio, R.J. Williams (2024). Analysis of Singularly Perturbed Stochastic Chemical Reaction Networks Motivated by Applications to Epigenetic Cell Memory. *SIAM Journal on Applied Dynamical Systems*, vol. 23, no. 4, pp. 2695-2731.
- S. Bruno, Y. Fu, F.A. Campos, D. Del Vecchio, R.J. Williams. Coclique Level Structure for Stochastic Chemical Reaction Networks, *Journal of Mathematical Biology*, under review, 36 pages, SI: 6 pages.

ABSTRACT OF THE DISSERTATION

Mathematical Modeling and Analysis of Stochastic Chemical Reaction Networks
Motivated by Applications to Epigenetic Cell Memory

by

Yi Fu

Doctor of Philosophy in Bioinformatics and Systems Biology

University of California San Diego, 2025

Professor Ruth J. Williams, Chair
Professor James T. Kadonaga, Co-Chair

Epigenetic cell memory (ECM), the inheritance of gene expression patterns across subsequent cell divisions, is a critical property of multi-cellular organisms in cell differentiation and cell fate determination. The theoretical study presented in this dissertation is motivated by the desire to obtain mechanistic insights into the duration of this epigenetic cell memory and the dynamics of mediators (such as histone modifications and DNA methylation) that contribute to this memory.

Stochastic dynamical systems involved in ECM can be described using Stochastic Chemical Reaction Networks (SCRNs), a class of continuous time Markov chain models

frequently used to study biochemical systems undergoing a series of reactions which change the numbers of molecules of a finite set of species in a probabilistic manner over time. This dissertation studies three aspects of the long-run and transient behavior of SCRN. First, comparison theorems are developed which provide stochastic ordering results for SCRN. These results provide sufficient conditions for establishing monotonic dependence on parameters of (mean) first passage times and stationary distributions of SCRN. Second, some results on singular perturbations for continuous time Markov chains are developed. The leading terms in rigorous series expansions of stationary distributions and mean first passage times are characterized using a reduced Markov chain. Finally, we introduce the concept of coclique level structure and derive closed form formulas for both upper and lower bounds for the mean first passage times for SCRN having such structures.

These newly developed theoretical tools not only allow us to set a rigorous mathematical basis for studying the effects of chromatin modification dynamics on epigenetic cell memory, but they can also be applied to other continuous time Markov chain models, especially those associated with (bio)chemical reaction networks.

Chapter 1

Introduction

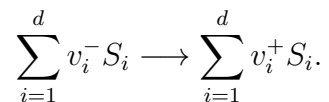
1.1 Overview

Epigenetic Cell Memory (ECM) is the ability of cells to pass on and maintain cell-type-specific information through generations of cell division without changes in DNA sequence. It plays an important role in cell differentiation and cell fate determination, with potential import for cell-based therapies. In order to understand the process of cell differentiation and to intervene in aberrant cells to rescue phenotype loss, it is important to quantitatively understand the duration of cell memory and the dynamics of mediators (such as chromatin states determined by histone modifications and DNA methylation) that contribute to this memory.

To model and study this behavior, we use Stochastic Chemical Reaction Networks (SCRNs). These are continuous time Markov chain models used to describe the stochastic dynamics of a chemical system undergoing a series of reactions which change the numbers of molecules of a finite set of species over time. While the theoretical study of SCRNs presented in this dissertation is motivated by the desire to understand the dynamics of interaction networks involved in ECM, our analysis is more general and provides a systematic framework for studying biochemical systems arising in other applications, such as intracellular viral kinetics (see Srivastava et al. [16] and Haseltine and Rawlings [9]) and enzymatic kinetics (see for example Kang et al. [12]).

(Stochastic) chemical reaction network theory started in the 1970s as a (stochastic) dynamical systems study of the reactions that take place in a chemical reactor. More recently, this theory has been used in systems biology to develop a quantitative understanding of system dynamics, in that realm.

To describe the setup for a SCRN, we consider a finite non-empty set of species $\mathcal{S} = \{S_1, \dots, S_d\}$, and a finite non-empty set $\mathcal{R} \subset \mathbb{Z}_+^d \times \mathbb{Z}_+^d$ that represents chemical reactions. The set $\mathcal{S}$ represents d species in a system subject to n reactions in $\mathcal{R}$, which change the number of copies of some species. For each reaction $(v^-, v^+) \in \mathcal{R}$, the d -dimensional vector v^- counts how many copies of each species are consumed by the reaction, and v^+ counts how many copies of each species are produced by the reaction. The reaction associated with (v^-, v^+) is usually written as



The pair $(\mathcal{S}, \mathcal{R})$ describes the stoichiometry of a Chemical Reaction Network (CRN). The stochastic dynamics can then be described by a Stochastic Chemical Reaction Network (SCRN), which is a continuous time Markov chain $X = (X_1, \dots, X_d)$ with a state space $\mathcal{X} \subset \mathbb{Z}_+^d$. The stochastic process X tracks the number of copies of each species in the system as a function of time and evolves as follows. Given a current state $x = (x_1, \dots, x_d) \in \mathcal{X}$ of X , for each reaction $(v^-, v^+) \in \mathcal{R}$, there is a clock which will ring at an exponentially distributed time with a state-dependent mean of $\frac{1}{\Lambda_{(v^-, v^+)}(x)}$. The first clock to ring triggers the Markov chain to take a jump according to the net change in the number of copies of each species $v^+ - v^-$ associated with that reaction firing. The rate functions $\Lambda_{(v^-, v^+)}(x) : (v^-, v^+) \in \mathcal{R}$ are called propensity functions and are often given by mass action kinetics.

This dissertation studies three aspects of SCRNs, briefly described in the following. Applications of these results to stochastic models of ECM are given in multiple examples.

1.1.1 Comparison Theorems

First, we focus on the question of how changing reaction rate parameters can affect the system dynamics. To address this, we develop comparison methods to establish stochastic ordering between two SCRN with similar structures. This type of result provides inequalities for stochastic processes. The reader may consult Muller and Stoyan [15] for an introduction to this subject. To the best of our knowledge, there has not been a previous systematic study of stochastic ordering for SCRN. At a more abstract level, some theorems have been established previously for stochastic processes and have been specialized to particular classes of stochastic processes such as those for queueing systems and point processes. For continuous time Markov chains, the work of Massey [14] is of special interest but is still very abstract. In this dissertation, we exploit the special structure of SCRN in order to develop comparison theorems under verifiable sufficient conditions.

In Chapter 2, we develop novel theorems for comparing two SCRN with similar structures, by constructing versions of them on the same probability space and in a comparable manner. These theoretical results allow us to directly compare (mean) first passage times and stationary distributions between SCRN with different rate parameters or initial conditions. Several examples are presented to illustrate these results. In addition, we provide an explicit algorithm for simultaneously simulating two SCRN in such a way that their sample paths are monotonically related.

1.1.2 Singular Perturbations

It has been reported that chromatin modifications, such as DNA methylation and histone modifications, are key mediators of ECM (see for example Carey [8], Heard and Martienssen [11], Allis et al. [1], Kim and Costello [13], and references in Bruno et al. [6] for more biological background), and the time scale separation between establishment (fast)

and erasure (slow) of these modifications extends the duration of cell memory (see Bruno et al. [6]). We model these chromatin modification circuits as SCRN, where the continuous time Markov chains are singularly perturbed. The perturbation parameter is often denoted by ε and is regarded as small. There is a substantial change in system behavior as ε transitions from being strictly positive to zero. There is some literature on discrete and continuous time, singularly perturbed Markov chains, especially by Avrachenkov et al. [2], Hassin and Haviv [10], Beltrán and Landim [3, 4], and Yin and Zhang [17]. However, we directly deal with the singularly perturbed continuous time Markov chains and provide more concrete theoretical results for both the stationary distribution and the mean first passage times.

In Chapter 3, we study the long-run and transient behavior of singularly perturbed, finite state, continuous time Markov chains as a small parameter ε tends to zero. We show that the stationary distributions of these singularly perturbed Markov chains admit power series expansions in ε and the mean first passage times admit Laurent series expansions in ε . An algorithm is provided to determine the orders of the poles of mean first passage times. Moreover, the leading terms in series expansions of stationary distributions and mean first passage times are characterized using a reduced Markov chain. We then apply these results to the chromatin modification models proposed in Bruno et al. [6]. These theoretical results can also be used to analyze other models that respect the same assumptions.

1.1.3 Coclique Level Structure

Formulas for mean first passage times (MFPTs) between states of a Markov chain are helpful in studying the transient behavior for a SCRN. While calculating an explicit formula is relatively straightforward for certain one-dimensional models, such as birth-death processes, this typically becomes significantly more difficult for higher-dimensional SCRN, especially when symbolic parameter dependencies are retained. We introduce the concept of *coclique level structure* and demonstrate how the presence of such a structure

in a SCRN allows us to derive closed form formulas for both upper and lower bounds for the MFPTs. Our methods can be applied to SCRNs taking values in a generic finite state space and can also be applied to models with non-mass-action kinetics.

In Chapter 4, we first introduce the concept of coclique level structure and develop theorems to determine whether certain SCRNs have this feature by studying associated graphs. Additionally, we develop an algorithm to identify, under specific assumptions, all possible coclique level structures associated with a given SCRN. For a SCRN with a coclique level structure, we present upper and lower bound formulas for the MFPTs. We further illustrate our results with examples from the biological areas of epigenetics, neurobiology and ecology.

1.2 Structure of the dissertation

This dissertation consists of an introduction, followed by three free-standing chapters, each one related to a paper. This document is designed so that the reader can read each one of these free-standing chapters independently. Chapter 2 is a slightly modified reproduction of Campos et al. [7]. Chapter 3 consists of a slightly modified extract from Bruno et al. [5]. Chapter 4 is a slightly modified reproduction of a paper submitted for publication. The modifications are in enumeration and format, to conform to the required structure for this dissertation.

1.3 Acknowledgement

Part of the abstract has been taken from Chapters 2, 3 and 4. Part of the introduction has taken text in a summarized way from Chapters 2, 3 and 4.

Bibliography for Chapter 1

- [1] C. David Allis, Marie-Laure Caparros, Thomas Jenuwein, and Danny Reinberg. *Epigenetics (Second Edition)*. Cold Spring Harbor Laboratory Press, 2015.
- [2] Konstantin E. Avrachenkov, Jerzy A. Filar, and Phil G. Howlett. *Analytic Perturbation Theory and Its Applications*. Society for Industrial and Applied Mathematics, 2013.
- [3] Johel Beltrán and Claudio Landim. “Tunneling and metastability of continuous time Markov chains”. In: *Journal of Statistical Physics* 140.6 (2010), pp. 1065–1114. DOI: 10.1007/s10955-010-0030-9.
- [4] Johel Beltrán and Claudio Landim. “Tunneling and metastability of continuous time Markov chains II, the nonreversible case”. In: *Journal of Statistical Physics* 149.4 (2012), pp. 598–618. DOI: 10.1007/s10955-010-0030-9.
- [5] Simone Bruno, Felipe A. Campos, Yi Fu, Domitilla Del Vecchio, and Ruth J. Williams. “Analysis of singularly perturbed stochastic chemical reaction networks motivated by applications to epigenetic cell memory”. In: *SIAM Journal on Applied Dynamical Systems* 23.4 (2024), pp. 2695–2731. DOI: 10.1137/23M1592389.
- [6] Simone Bruno, Ruth J. Williams, and Domitilla Del Vecchio. “Epigenetic cell memory: the gene’s inner chromatin modification circuit”. In: *PLOS Computational Biology* 18.4 (2022), pp. 1–27. DOI: 10.1371/journal.pcbi.1009961.
- [7] Felipe A. Campos, Simone Bruno, Yi Fu, Domitilla Del Vecchio, and Ruth J. Williams. “Comparison theorems for stochastic chemical reaction networks”. In: *Bulletin of Mathematical Biology* 85.39 (2023), pp. 1–39. DOI: 10.1007/s11538-023-01136-5.
- [8] Nessa Carey. *The Epigenetic Revolution*. Columbia University Press, 2012.

- [9] Eric L. Haseltine and James B. Rawlings. “Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics”. In: *The Journal of Chemical Physics* 117 (2002), pp. 6959–6969. DOI: 10.1063/1.1505860.
- [10] Refael Hassin and Moshe Haviv. “Mean passage times and nearly uncoupled Markov chains”. In: *SIAM Journal on Discrete Mathematics* 5.3 (1992), pp. 386–397. DOI: 10.1137/0405030.
- [11] Edith Heard and Robert A. Martienssen. “Transgenerational epigenetic inheritance: myths and mechanisms”. In: *Cell* 157.1 (2014), pp. 95–109. DOI: 10.1016/j.cell.2014.02.045.
- [12] Hye-Won Kang, Wasiur Rahman KhudaBukhsh, Heinz Koepl, and Grzegorz Rempala. “Quasi-steady-state approximations derived from the stochastic model of enzyme kinetics”. In: *Bulletin of Mathematical Biology* 81 (2019), pp. 1303–1336. DOI: 10.1007/s11538-019-00574-4.
- [13] Mirang Kim and Joseph Costello. “DNA methylation: an epigenetic mark of cellular memory”. In: *Experimental & Molecular Medicine* 49.4 (2017), pp. 1–8. DOI: 10.1038/emmm.2017.10.
- [14] William A. Massey. “Stochastic orderings for Markov processes on partially ordered spaces”. In: *Mathematics of Operations Research* 12.2 (1987), pp. 350–367. DOI: 10.1287/moor.12.2.350.
- [15] Alfred Müller and Dietrich Stoyan. *Comparison Methods for Stochastic Models and Risks*. Wiley Series in Probability and Statistics. John Wiley & Sons Inc., 2002.
- [16] Rishi K. Srivastava, Lingchong You, Jesse W. Summers, and John Yin. “Stochastic vs. deterministic modeling of intracellular viral kinetics”. In: *Journal of Theoretical Biology* 218.3 (2002), pp. 309–321. DOI: 10.1006/jtbi.2002.3078.

- [17] G. George Yin and Qing Zhang. *Continuous-Time Markov Chains and Applications: A Two-Time-Scale Approach*. Stochastic Modelling and Applied Probability. Springer New York, 2012.

Chapter 2

Comparison Theorems

2.1 Abstract

Continuous time Markov chains are frequently used as stochastic models for chemical reaction networks, especially in the growing field of systems biology. A fundamental problem for these Stochastic Chemical Reaction Networks (SCRNs) is to understand the dependence of the stochastic behavior of these systems on the chemical reaction rate parameters. Towards solving this problem, in this work we develop theoretical tools called comparison theorems that provide stochastic ordering results for SCRNs. These theorems give sufficient conditions for monotonic dependence on parameters in these network models, which allow us to obtain, under suitable conditions, information about transient and steady state behavior. These theorems exploit structural properties of SCRNs, beyond those of general continuous time Markov chains. Furthermore, we derive two theorems to compare stationary distributions and mean first passage times for SCRNs with different parameter values, or with the same parameters and different initial conditions. These tools are developed for SCRNs taking values in a generic (finite or countably infinite) state space and can also be applied for non-mass-action kinetics models. When propensity functions are bounded, our method of proof gives an explicit method for coupling two comparable SCRNs, which can be used to simultaneously simulate their sample paths in a comparable manner. We illustrate our results with applications to models of enzymatic

kinetics and epigenetic regulation by chromatin modifications.

2.2 Introduction

2.2.1 Overview

Stochastic Chemical Reaction Networks (SCRNs) are a class of continuous time Markov chain models used to describe the stochastic dynamics of a chemical system undergoing a series of reactions which change the numbers of molecules of a finite set of species over time. These models provide a framework for the theoretical study of biochemical systems in areas such as intracellular viral kinetics (see Srivastava et al. [26] and Haseltine and Rawlings [16]), enzymatic kinetics (see for example Kang et al. [19]) and epigenetic regulation by chromatin modifications (see Bruno et al. [7] for a recently developed model of chromatin regulation).

One of the most interesting questions for biochemical system models is: "What effect does changing reaction rate parameters have on system dynamics?" Indeed, different rate parameters for chemical processes can lead to different stochastic behaviors. One possible approach to evaluate the effect of parameter variations on system dynamics is through comparison theorems for stochastic processes. More precisely, this type of theorem provides inequalities between stochastic processes (see Müller and Stoyan [24] for a general reference on this topic).

In this work, we employ uniformization and coupling methods (see Grassmann [13] and Keilson [20]) to derive comparison theorems for SCRNs under verifiable sufficient conditions. These theoretical results enable us to develop two novel theorems yielding a direct comparison of mean first passage times and stationary distributions between SCRNs with different rate parameters or initial conditions. We apply these theorems to several examples to illustrate how they can be used to understand how key biological parameters affect stochastic behavior. While a major motivator for our work has been the study

of SCRN, we state our theorems in the context of continuous time Markov chains, for which the state space is a subset of $\mathbb{Z}_+^d$ (the set of d -dimensional vectors with non-negative integer entries), and the set of all possible transition vectors is a finite set. This thereby allows for other applications that have similar characteristics to SCRN. In addition, for the case of bounded transition intensities satisfying our conditions, we give an explicit concrete coupling of two comparable Markov chains, which can be used to simultaneously simulate them in such a way that their sample paths are monotonically related.

This work is structured as follows. In Section 2.3, we introduce some background on Stochastic Chemical Reaction Networks needed for this article. We present the main results in Section 2.4, with proofs provided in Section 2.6. In Section 2.5, we apply our theoretical tools to several examples, such as epigenetic regulation by chromatin modifications and enzymatic kinetics. Concluding remarks are presented in Section 2.7. SI-Section 2.8 contains some further details and extensions of the main results and examples in this work.

2.2.2 Related work

Due to the growing field of systems biology, the mathematical study of chemical reaction networks has seen a wealth of activity lately. Concerning comparison results, considerable work has been conducted on monotonicity properties for deterministic models of chemical reaction networks, i.e., systems of ordinary differential equations describing the dynamics of species concentrations. For example, Angeli et al. [3] proposed a graphical method, based on the monotonicity properties of the reaction rates with respect to species concentrations, to determine global stability properties for the models. More recently, Gori et al. [12] introduced sufficient conditions to verify the existence of a monotonicity property for the concentrations of species for any positive time with respect to their initial concentrations. However, these works do not address how changing parameters affects the behavior of stochastic models.

To the best of our knowledge, no systematic study of stochastic ordering has been

conducted for stochastic chemical reaction networks. On a more general level, theorems have been established for stochastic processes and have been specialized for particular classes such as for queueing systems and point processes (see Müller and Stoyan [24] for an introduction to the topic). For Markov chains, the work of Massey [21] is of special interest, since he establishes criteria for comparison of continuous time Markov chains in terms of their infinitesimal generators. For relevant work prior to Massey, there is a nice summary in [21]. In particular, Kamae et al. [18] showed that for Markov processes, a comparison between transition probability functions, at all fixed times and for all partially ordered starting points, can be realized in a pathwise stochastic comparison between versions of the Markov processes. In relation to Massey's work, our results provide simplified conditions and extended results for stochastic comparisons, which exploit the structure of stochastic chemical reaction networks. Furthermore, unlike Massey, we do not require a uniform bound on the rates of leaving each state. In addition, under the latter assumption, we explicitly construct versions of the stochastic processes on the same probability space that have comparable sample paths. More detail on the relationship of our work to that of Massey is given in Remark 2.4.2. In contrast to work on sensitivity analysis of distributions at a finite set of times and which considers only local changes in parameters (see for example Gunawan et al. [14], Gupta and Khammash [15] and references therein), our work provides a sample path comparison between stochastic processes for global changes in their parameters.

2.2.3 Notation and terminology

Denote by $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$ the set of non-negative integers. For an integer $d \geq 1$ we denote by $\mathbb{Z}_+^d$ the set of d -dimensional vectors with entries in $\mathbb{Z}_+$. For any integer $d \geq 1$, let $\mathbb{R}^d$ denote the d -dimensional Euclidean space. We usually write $\mathbb{R}$ for $\mathbb{R}^1$. We denote by $\mathbb{R}_+^d$ the set of vectors $x \in \mathbb{R}^d$ such that $x_i \geq 0$ for every $1 \leq i \leq d$. For $x \in \mathbb{R}^d$, let $\|x\|_\infty = \sup_{1 \leq i \leq d} |x_i|$ be the supremum norm. In this work, the sum over the empty

set is considered to be 0.

A binary relation $\preccurlyeq$ on a set $\mathcal{X}$ will be called reflexive if $x \preccurlyeq x$ for every $x \in \mathcal{X}$, transitive if $x \preccurlyeq y$ and $y \preccurlyeq z$ implies $x \preccurlyeq z$ for every $x, y, z \in \mathcal{X}$ and antisymmetric if $x \preccurlyeq y$ and $y \preccurlyeq x$ implies $x = y$ for every $x, y \in \mathcal{X}$. A preorder is a binary relation that is reflexive and transitive. A partial order is a preorder that is antisymmetric.

In this work, a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ will consist of a sample space Ω , a σ -algebra of events $\mathcal{F}$ and a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$. We will say that two real-valued random variables Y, Y' (defined on possibly different probability spaces) are equal in distribution, denoted as $Y' \stackrel{\text{dist}}{=} Y$, if their cumulative distribution functions agree. All stochastic processes considered in this work will have right-continuous sample paths that also have finite left-limits.

2.3 Stochastic Chemical Reaction Networks

In this section we provide necessary background on Stochastic Chemical Reaction Networks. The reader is referred to Anderson and Kurtz [2] for an introduction to this subject.

We assume there is a finite non-empty set $\mathcal{S} = \{S_1, \dots, S_d\}$ of d **species**, and a finite non-empty set $\mathcal{R} \subseteq \mathbb{Z}_+^d \times \mathbb{Z}_+^d$ that represents chemical **reactions**. We assume that $(w, w) \notin \mathcal{R}$ for every $w \in \mathbb{Z}_+^d$. The set $\mathcal{S}$ represents d different molecular species in a system subject to reactions $\mathcal{R}$ which change the number of molecules of each species. For each $(v^-, v^+) \in \mathcal{R}$, the d -dimensional vector v^- (the **reactant vector**) counts how many molecules of each species are consumed in the reaction, while v^+ (the **product vector**) counts how many molecules of each species are produced. The reaction is usually written as

$$\sum_{i=1}^d (v^-)_i S_i \rightarrow \sum_{i=1}^d (v^+)_i S_i.$$

To avoid the use of unnecessary symbols, we will assume that for each $1 \leq i \leq d$, there

exists a vector $w = (a_1, \dots, a_d)^T \in \mathbb{Z}_+^d$ with $a_i > 0$ such that (w, v) or (v, w) is in $\mathcal{R}$ for some $v \in \mathbb{Z}_+^d$, i.e., each species is either a reactant or a product in some reaction.

The net change in the quantity of molecules of each species due to a reaction $(v^-, v^+) \in \mathcal{R}$ is described by $v^+ - v^-$ and it is called the associated **reaction vector**. We denote the set of reaction vectors $\mathcal{V} := \{v \in \mathbb{Z}^d \mid v = v^+ - v^- \text{ for some } (v^-, v^+) \in \mathcal{R}\}$, let $n := |\mathcal{V}|$ the size of $\mathcal{V}$ and enumerate the members of $\mathcal{V}$ as $\{v_1, \dots, v_n\}$. Note that $\mathcal{V}$ does not contain the zero vector because $\mathcal{R}$ has no elements of the form (w, w) . Different reactions might have the same reaction vector. For each $v_j \in \mathcal{V}$ we consider the set $\mathcal{R}_{v_j} := \{(v^-, v^+) \in \mathcal{R} \mid v_j = v^+ - v^-\}$. The reaction vectors generate the **stoichiometric subspace** $\mathcal{L} := \text{span}(\mathcal{V})$. For $z \in \mathbb{R}^d$, we call $z + \mathcal{L}$ a **stoichiometric compatibility class**.

Given $(\mathcal{S}, \mathcal{R})$ we will consider an associated continuous time Markov chain $X = (X_1, \dots, X_d)$, with a state space $\mathcal{X}$ contained in $\mathbb{Z}_+^d$, which tracks the number of molecules of each species over time. Roughly speaking, the dynamics of X will be given by the following: given a current state $x = (x_1, \dots, x_d) \in \mathcal{X} \subseteq \mathbb{Z}_+^d$, for each reaction $(v^-, v^+) \in \mathcal{R}$, there is a clock which will ring at an exponentially distributed time (with rate $\Lambda_{(v^-, v^+)}(x)$). The clocks for distinct reactions are independent of one another. If the clock corresponding to $(v^-, v^+) \in \mathcal{R}$ rings first, the system moves from x to $x + v^+ - v^-$ at that time, and then the process repeats. We now define the Markov chain in more detail.

Consider a set of species $\mathcal{S}$ and of reactions $\mathcal{R}$, a set $\mathcal{X} \subseteq \mathbb{Z}_+^d$ and a collection of functions $\{\Lambda_{(v^-, v^+)} : \mathcal{X} \rightarrow \mathbb{R}_+\}_{(v^-, v^+) \in \mathcal{R}}$ such that for each $x \in \mathcal{X}$ and $(v^-, v^+) \in \mathcal{R}$, if $x + v^+ - v^- \notin \mathcal{X}$, then $\Lambda_{(v^-, v^+)}(x) = 0$. Now, for $1 \leq j \leq n$, $v_j \in \mathcal{V}$, define

$$\Upsilon_j(x) := \sum_{(v^-, v^+) \in \mathcal{R}_{v_j}} \Lambda_{(v^-, v^+)}(x).$$

Note that for each $x \in \mathcal{X}$ and $1 \leq j \leq n$, if $x + v_j \notin \mathcal{X}$, then $\Upsilon_j(x) = 0$. A **stochastic chemical reaction network (SCRN)** is a Markov chain X with state space $\mathcal{X}$ and

infinitesimal generator¹ Q given for $x, y \in \mathcal{X}$ by

$$Q_{x,y} = \begin{cases} \Upsilon_j(x) & \text{if } y - x = v_j \text{ for some } 1 \leq j \leq n, \\ -\sum_{j=1}^n \Upsilon_j(x) & \text{if } y = x, \\ 0 & \text{otherwise.} \end{cases}$$

The functions $\{\Lambda_{(v^-, v^+)} : \mathcal{X} \rightarrow \mathbb{R}_+\}_{(v^-, v^+) \in \mathcal{R}}$ are called **propensity** or **intensity** functions. A common form for the propensity functions is the following associated with **mass action kinetics**:

$$\Lambda_{(v^-, v^+)}(x) = \kappa_{(v^-, v^+)} \prod_{i=1}^d (x_i)_{(v^-)_i}, \quad (2.3.1)$$

where $\{\kappa_{(v^-, v^+)}\}_{(v^-, v^+) \in \mathcal{R}}$ are positive constants and for $m, \ell \in \mathbb{Z}_+$, the quantity $(m)_\ell$ is the falling factorial, i.e., $(m)_0 := 1$ and $(m)_\ell := m(m-1)\dots(m-\ell+1)$.

Remark 2.3.1. Our definition of SCRN allows for some model flexibility. Notice that the propensity functions are not necessarily defined on the whole lattice $\mathbb{Z}_+^d$ and they are not necessarily of the form (2.3.1). Indeed, in some of our applications, mass-conservation laws restrict the possible values that X may take (see Example 2.5.4). In addition, there may be other types of kinetics, such as those described by Hill functions (see Example 2.5.5).

A convenient way to represent such a Markov chain is given in Theorem 6.4.1 of Ethier and Kurtz [11]. For this, consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with independent unit rate Poisson processes $N_1, \dots, N_n$. There is a version of X defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that

$$X(t) = X(0) + \sum_{j=1}^n v_j N_j \left(\int_0^t \Upsilon_j(X(s)) ds \right), \quad (2.3.2)$$

¹Note that Q is sometimes called an infinitesimal transition matrix although it may have countably many “rows” and “columns”. The entries $Q_{x,y}$ for $x \neq y$ are the infinitesimal transition rates of going from x to y : $\mathbb{P}[X(t+h) = y | X(t) = x] = Q_{x,y}h + o(h)$ as $h \rightarrow 0$.

for every $0 \leq t < \tau$, where τ is the explosion time for X (which may be $+\infty$). From (2.3.2), it is easy to see that for a SCRN X with initial state $z \in \mathcal{X}$, $X(t)$ will stay in the stoichiometric compatibility class $z + \mathcal{L}$ intersected with $\mathbb{Z}_+^d$ for all time $0 \leq t < \tau$, with probability one. For this reason, sometimes it will be convenient to choose $\mathcal{X} = (z + \mathcal{L}) \cap \mathbb{Z}_+^d$, for a fixed $z \in \mathbb{Z}_+^d$.

While our work was initially motivated by questions for SCRNs, we will first develop our results in a more general context of continuous time Markov chains, for which the state space is contained in $\mathbb{Z}_+^d$ and the set of all possible transition vectors is a finite set, and then illustrate them for SCRNs.

2.4 Main Results

The general stochastic ordering results provided in this work are relative to a preorder relation on a state space $\mathcal{X} \subseteq \mathbb{Z}_+^d \subseteq \mathbb{R}^d$. We will define the preorder on all of $\mathbb{R}^d$ and then restrict it to various subsets. We introduce this notation and related notation in Section 2.4.1. In Section 2.4.2 we present the main results of this article, and in Section 2.4.3 we discuss relevant consequences for the comparison of (mean) first passage times and stationary distributions.

2.4.1 Preorders in $\mathbb{R}^d$

Let $m, d \geq 1$ be integers. Denote by $\leq$ the usual componentwise partial order on $\mathbb{R}^d$, i.e., for $x, y \in \mathbb{R}^d$, $x \leq y$ whenever $x_i \leq y_i$ for every $1 \leq i \leq d$. Additionally, we write $x < y$ whenever $x_i < y_i$ for every $1 \leq i \leq d$. For the rest of this work, we consider a matrix $A \in \mathbb{R}^{m \times d}$, where no row of A is identically zero.

Definition 2.4.1. For $x, y \in \mathbb{R}^d$, we say that $x \preceq_A y$ whenever $A(y - x) \geq 0$.

For the matrix A , consider the convex cone $K_A := \{x \in \mathbb{R}^d \mid Ax \geq 0\}$. Note that $x \preceq_A y$ holds if and only if $y - x \in K_A$. Moreover, the relation $\preceq_A$ is reflexive and

transitive, and therefore a preorder on $\mathbb{R}^d$. Also, for this relation,

$$\text{if } x \preccurlyeq_A y, \text{ then } x + z \preccurlyeq_A y + z \text{ for any } z \in \mathbb{R}^d. \quad (2.4.1)$$

For any $x \in \mathbb{R}^d$ consider the set

$$K_A + x = \{y \in \mathbb{R}^d \mid A(y - x) \geq 0\} = \{y \in \mathbb{R}^d \mid x \preccurlyeq_A y\}.$$

In the coming sections, we will consider the notions of increasing and decreasing sets with respect to $\preccurlyeq_A$ in a given subset of $\mathbb{Z}_+^d$. More concretely, consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$. We will say that a set $\Gamma \subseteq \mathcal{X}$ is **increasing** in $\mathcal{X}$ with respect to $\preccurlyeq_A$ if for every $x \in \Gamma$ and $y \in \mathcal{X}$, $x \preccurlyeq_A y$ implies that $y \in \Gamma$. We observe that, for $x \in \mathcal{X}$, the set

$$(K_A + x) \cap \mathcal{X} = \{y \in \mathcal{X} \mid x \preccurlyeq_A y\}$$

is increasing in $\mathcal{X}$ by the transitivity property of $\preccurlyeq_A$. On the other hand, we will say that a set $\Gamma \subseteq \mathcal{X}$ is **decreasing** in $\mathcal{X}$ with respect to $\preccurlyeq_A$ if for every $x \in \Gamma$ and $y \in \mathcal{X}$, $y \preccurlyeq_A x$ implies that $y \in \Gamma$. We will say that a point x is **maximal** (resp. **minimal**) in $\mathcal{X}$ if for every $y \in \mathcal{X}$, $x \preccurlyeq_A y$ (resp. $y \preccurlyeq_A x$) implies that $x = y$. In this case, the set $\Gamma = \{x\}$ would be increasing (resp. decreasing) in $\mathcal{X}$.

Remark 2.4.1. If $\text{rank}(A) = d$, then the relation $\preccurlyeq_A$ will be antisymmetric and therefore a partial order on $\mathbb{R}^d$. Indeed, if $\text{rank}(A) = d$, then $A(y - x) = 0$ implies that $x = y$. In addition, $\preccurlyeq_A$ will then be a partial order when restricted to $\mathcal{X} \subset \mathbb{Z}_+^d$. Throughout this article, we will not assume that $\text{rank}(A) = d$ and therefore, the relation $\preccurlyeq_A$ might not be a partial order on $\mathcal{X}$ (see Examples 2.5.1, 2.5.2, and 2.5.3).

2.4.2 Stochastic comparison theorems

The fundamental objects in the following results are a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$ and a pair of continuous time Markov chains X and $\check{X}$ with the same state space $\mathcal{X}$ and where it is assumed that the set of all possible transition vectors for X or $\check{X}$ is a finite set. We denote the size of this set by n . A primary example of this setup is two stochastic chemical reaction networks as described in Section 2.3 with different propensity functions. We will now formally introduce the notation for stating our results.

Consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, an integer $n \geq 1$ and a collection of distinct vectors $v_1, \dots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$, where 0 is the origin in $\mathbb{Z}^d$. Consider two collections of functions $\Upsilon = (\Upsilon_1, \dots, \Upsilon_n)$ and $\check{\Upsilon} = (\check{\Upsilon}_1, \dots, \check{\Upsilon}_n)$ defined on $\mathcal{X}$ and taking values in $\mathbb{R}_+$, such that for every $1 \leq j \leq n$ and $x \in \mathcal{X}$:

$$\text{if } x + v_j \notin \mathcal{X}, \text{ then } \Upsilon_j(x) = \check{\Upsilon}_j(x) = 0. \quad (2.4.2)$$

Consider a continuous time Markov chain X on the state space $\mathcal{X}$ with infinitesimal generator $Q = (Q_{x,y})_{x,y \in \mathcal{X}}$ defined for $x, y \in \mathcal{X}$ by

$$Q_{x,y} := \begin{cases} \Upsilon_j(x) & \text{if } y - x = v_j \text{ for some } 1 \leq j \leq n, \\ -\sum_{j=1}^n \Upsilon_j(x) & \text{if } x = y, \\ 0 & \text{otherwise.} \end{cases} \quad (2.4.3)$$

Consider the analogous continuous time Markov chain $\check{X}$ with infinitesimal generator $\check{Q}$ as in (2.4.3) but with functions $\check{\Upsilon}_1, \dots, \check{\Upsilon}_n$ instead of $\Upsilon_1, \dots, \Upsilon_n$. We call X and $\check{X}$ the continuous time Markov chains associated with Υ and $\check{\Upsilon}$ respectively. We will assume that X and $\check{X}$ do not explode in finite time. The following is our main result.

Theorem 2.4.1. *Consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, a collection of distinct vectors*

$v_1, \dots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$ and two collections of non-negative functions on $\mathcal{X}$, $\Upsilon = (\Upsilon_1, \dots, \Upsilon_n)$ and $\check{\Upsilon} = (\check{\Upsilon}_1, \dots, \check{\Upsilon}_n)$, such that (2.4.2) holds and the associated continuous time Markov chains do not explode in finite time. Consider a matrix $A \in \mathbb{R}^{m \times d}$ with non-zero rows and suppose that for every $x, y \in \mathcal{X}$ such that $x \preceq_A y$ the following hold:

$$\check{\Upsilon}_j(y) \leq \Upsilon_j(x), \quad \text{for each } 1 \leq j \leq n \text{ such that } y + v_j \in \mathcal{X} \setminus (K_A + x), \quad (2.4.4)$$

and

$$\check{\Upsilon}_j(y) \geq \Upsilon_j(x), \quad \text{for each } 1 \leq j \leq n \text{ such that } x + v_j \in \mathcal{X} \text{ and } y \notin K_A + x + v_j. \quad (2.4.5)$$

Then, for each pair $x^\circ, \check{x}^\circ \in \mathcal{X}$ such that $x^\circ \preceq_A \check{x}^\circ$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with two continuous time Markov chains $X = \{X(t), t \geq 0\}$ and $\check{X} = \{\check{X}(t), t \geq 0\}$ defined there, each having state space $\mathcal{X} \subseteq \mathbb{Z}_+^d$, with infinitesimal generators Q and $\check{Q}$, associated with Υ and $\check{\Upsilon}$, respectively, with initial conditions $X(0) = x^\circ$ and $\check{X}(0) = \check{x}^\circ$ and such that:

$$\mathbb{P} \left[X(t) \preceq_A \check{X}(t) \text{ for every } t \geq 0 \right] = 1. \quad (2.4.6)$$

An example of checking conditions (2.4.4) and (2.4.5) is given in Figure 2.1. The proof of Theorem 2.4.1 is given in Section 2.6.1. The main idea in the construction of the processes X and $\check{X}$ is *uniformization* (see Chapter 2 in Keilson [20]) together with a suitable coupling. Our proof uses a single Poisson process together with a sequence of i.i.d. uniform random variables to determine potential jumps for the two continuous time Markov chains, where for X and $\check{X}$, potential jumps in the same direction v_j are coupled together, and their probabilities of acceptance are given by normalized versions of their infinitesimal transition rates Υ_j and $\check{\Upsilon}_j$. Uniformization can be done provided the diagonal terms of the infinitesimal generators are uniformly bounded in size. In the proof of Theorem 2.4.1, we initially make this assumption on Q and $\check{Q}$ in order to construct X

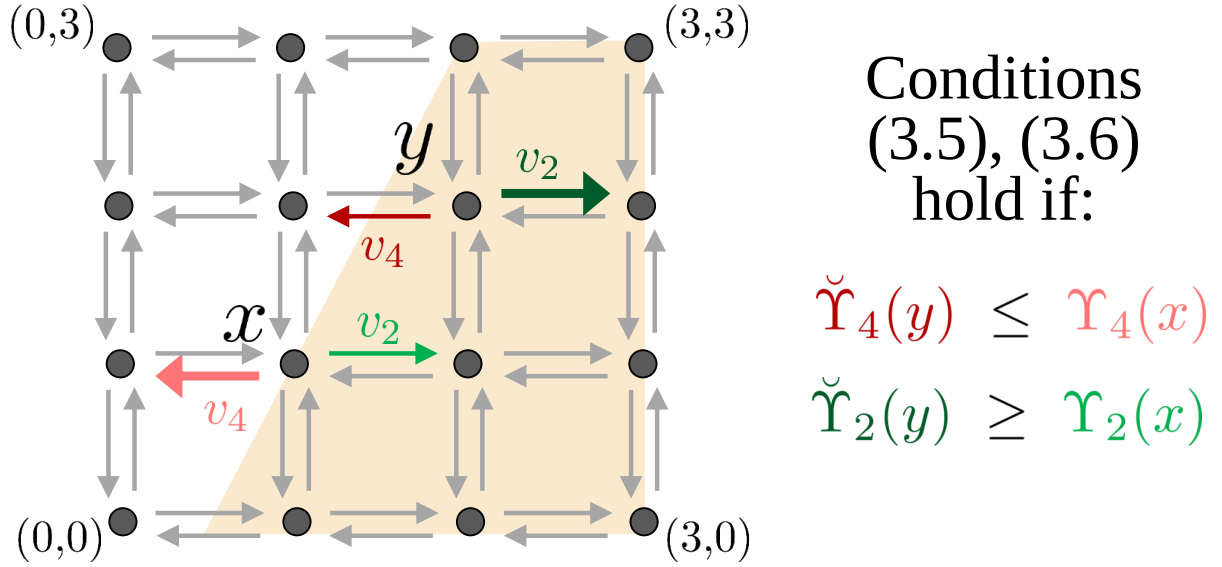


Figure 2.1. Pictorial representation of conditions (2.4.4), (2.4.5) for a certain $(K_A + x) \cap \mathcal{X}$ in a two-dimensional lattice. Here, $\mathcal{X} = \{0, 1, 2, 3\} \times \{0, 1, 2, 3\}$, $n = 4$, $v_1 = (0, 1)^T$, $v_2 = (1, 0)^T$, $v_3 = (0, -1)^T$, $v_4 = (-1, 0)^T$, where T denotes transpose, $A = \begin{bmatrix} 2 & -1 \end{bmatrix}$, and $(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid [2 \ -1](w - x) \geq 0\}$. In the graph, $(K_A + x) \cap \mathcal{X}$ consists of the states (black dots) that lie in the light orange region and the arrows represent possible transitions along v_1, v_2, v_3, v_4 between states. For the exhibited states $x, y \in \mathcal{X}$ with $x \preceq_A y$, the light green (dark green) and light red (dark red) arrows represent the transitions with rates $\Upsilon_2(x)$ ($\check{\Upsilon}_2(y)$) and $\Upsilon_4(x)$ ($\check{\Upsilon}_4(y)$) for the Markov chain X ($\check{X}$). Higher transitions rates are associated with thicker arrows. To check the conditions (2.4.4) and (2.4.5), since $y + v_4 \notin K_A + x$ and $y \notin K_A + x + v_2$, we need to check that $\check{\Upsilon}_4(y) \leq \Upsilon_4(x)$ and $\check{\Upsilon}_2(y) \geq \Upsilon_2(x)$.

and $\check{X}$. We then generalize the result to Markov chains that do not explode in finite time by using a truncation and limiting procedure. The construction mentioned here, for the case where the diagonal terms of the infinitesimal generators are uniformly bounded in size, besides playing a key role in our proofs, is also the basis for an algorithm described in SI-Section 2.8.4, which provides a way to simultaneously simulate the processes X and $\check{X}$ in a comparable manner.

Remark 2.4.2. In Theorem 5.3 of Massey [21], the author provides a necessary and sufficient condition for stochastic comparison of continuous time Markov chains at each fixed time for all partially ordered initial conditions. By the work of Kamae et al.

[18], the conditions in Massey [21] imply the existence of a coupling of continuous time Markov chains so that a relation such as (2.4.6) holds. Massey's condition requires that $\sum_{w \in \Gamma} Q_{xw} \leq \sum_{w \in \Gamma} \check{Q}_{yw}$ for every $x \preccurlyeq_A y$ and every set $\Gamma \subseteq \mathcal{X}$ that is increasing in $\mathcal{X}$ with respect to $\preccurlyeq_A$ and such that either $x \in \Gamma$ or $y \notin \Gamma$. These inequalities can often be difficult to check since first, they involve computing sums of terms in the infinitesimal generators and second, the form of all increasing sets can be hard to determine. In Theorem 2.4.1 we overcome these obstacles by providing simplified sufficient conditions that involve only pointwise comparison of entries in the infinitesimal generators associated to each of the transition vectors v_j . Besides this practical value, in our context, our results go beyond the work of Massey [21], since he assumes that $\preccurlyeq_A$ is a partial order (we only assume preorder) and he assumes that the diagonal entries of the infinitesimal generators are bounded (we generalize to non-exploding Markov chains). Our proof has a commonality with the work of Massey in the sense that we also use uniformization. It is different in the sense that, when infinitesimal transition rates are bounded, we construct an explicit coupling for all time, exploiting the simplified nature of our conditions, while Massey does not provide an explicit coupling. Instead, he proves existence of a stochastic comparison for each fixed time, using a semigroup approach.

Conditions (2.4.4) and (2.4.5) may be simplified if we consider a particular relation between the matrix A and the vectors $v_1, \dots, v_n$ in which $A \in \mathbb{Z}^{m \times d}$ and Av_j has entries taking values only in $\{-1, 0, 1\}$ for every $1 \leq j \leq n$. More concretely, let us consider a class of continuous time Markov chains such that, for a given matrix A with non-zero rows, if the Markov chain starts within the set $K_A + x$, then to go outside of it, the process will necessarily hit its boundary. In this case, we can derive a theorem whose conditions must be checked only on the boundary of $K_A + x$ because the only transitions that can lead the Markov chain outside or inside the set $K_A + x$ are ones starting on the boundary of $K_A + x$. Before stating the theorem, let us introduce the

sets $\partial_i(K_A + x) := \{y \in K_A + x \mid \langle A_{i\bullet}, y \rangle = \langle A_{i\bullet}, x \rangle\}$ ² for each $1 \leq i \leq m$. We can then characterize³ the boundary of $K_A + x$ as follows:

$$\partial(K_A + x) = \bigcup_{i=1}^m \partial_i(K_A + x). \quad (2.4.7)$$

Theorem 2.4.2. *Consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, a collection of distinct vectors $v_1, \dots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$ and two collections of non-negative functions on $\mathcal{X}$, $\Upsilon = (\Upsilon_1, \dots, \Upsilon_n)$ and $\check{\Upsilon} = (\check{\Upsilon}_1, \dots, \check{\Upsilon}_n)$ such that (2.4.2) holds and the associated continuous time Markov chains do not explode in finite time. Consider a matrix $A \in \mathbb{Z}^{m \times d}$ with non-zero rows and suppose that both of the following conditions hold:*

(i) *For each $1 \leq j \leq n$, the vector Av_j has entries in $\{-1, 0, 1\}$ only.*

(ii) *For each $x \in \mathcal{X}$, $1 \leq i \leq m$ and $y \in \partial_i(K_A + x) \cap \mathcal{X}$ we have that*

$$\check{\Upsilon}_j(y) \leq \Upsilon_j(x), \quad \text{for each } 1 \leq j \leq n \text{ such that } \langle A_{i\bullet}, v_j \rangle < 0, \quad (2.4.8)$$

and

$$\check{\Upsilon}_j(y) \geq \Upsilon_j(x), \quad \text{for each } 1 \leq j \leq n \text{ such that } \langle A_{i\bullet}, v_j \rangle > 0. \quad (2.4.9)$$

Then, for each pair $x^\circ, \check{x}^\circ \in \mathcal{X}$ such that $x^\circ \preceq_A \check{x}^\circ$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with two continuous time Markov chains $X = \{X(t), t \geq 0\}$ and $\check{X} = \{\check{X}(t), t \geq 0\}$ defined there, each having state space $\mathcal{X} \subseteq \mathbb{Z}_+^d$, with infinitesimal generators given by

²Here, for convenience of notation, let $A_{i\bullet}$ denote the row vector corresponding to the i -th row of A , for $1 \leq i \leq m$. In this article we will adopt the convention of considering the inner product $\langle \cdot, \cdot \rangle$ as a function of a row vector in its first entry and as a function of a column vector in the second entry. In particular, $\langle A_{i\bullet}, x \rangle = \sum_{k=1}^d A_{ik}x_k$.

³The fact that A does not contain zero rows allows for equation (2.4.7) to hold. In fact, let $A \in \mathbb{R}^{m \times d}$ be a matrix that is not identically zero and let $\bar{A}$ be the matrix obtained from A by erasing any rows that contain all zeros. Then, for $x, y \in \mathbb{R}^d$, $A(y-x) \geq 0$ if and only if $\bar{A}(y-x) \geq 0$, and so $K_A + x = K_{\bar{A}} + x$ and $x \preceq_A y$ if and only if $x \preceq_{\bar{A}} y$. However, if A contains a row $A_{i\bullet}$ such that $A_{i\bullet} = 0$, then $\partial_i(K_A + x) = K_A + x$ and if $K_A + x$ has nonempty interior, then equation (2.4.7) will not hold. Consequently, we have made the assumption that A has no zero rows.

Q and $\check{Q}$, associated with Υ and $\check{\Upsilon}$ respectively, with initial conditions $X(0) = x^\circ$ and $\check{X}(0) = \check{x}^\circ$, and such that:

$$\mathbb{P} \left[X(t) \preceq_A \check{X}(t) \text{ for every } t \geq 0 \right] = 1. \quad (2.4.10)$$

The proof of this theorem is given in Section 2.6.2 and involves checking that (2.4.4) and (2.4.5) of Theorem 2.4.1 hold, using conditions (i) and (ii) of Theorem 2.4.2.

Remark 2.4.3. In the context of Theorem 2.4.2, it is possible that for $x \in \mathcal{X}$, and $y \in \partial_{i_1}(K_A + x) \cap \partial_{i_2}(K_A + x) \cap \mathcal{X}$ with $i_1 \neq i_2$, it happens that $\langle A_{i_1 \bullet}, v_j \rangle < 0$ and $\langle A_{i_2 \bullet}, v_j \rangle > 0$ for some $1 \leq j \leq n$. For condition (ii) to hold, we must then have $\check{\Upsilon}_j(y) = \Upsilon_j(x)$.

When there are multiple vectors v_j with a common value for Av_j , the pointwise comparison in j , for $1 \leq j \leq n$, in conditions (2.4.8) and (2.4.9) in Theorem 2.4.2, can be weakened. To this end, let us introduce the set of distinct vectors $\{\eta^1, \dots, \eta^s\}$ formed by Av_j , for $1 \leq j \leq n$, where s denotes the cardinality of this set. Consider the subsets of indices

$$G^k := \{j \mid 1 \leq j \leq n \text{ and } Av_j = \eta^k\}, \quad \text{for } 1 \leq k \leq s. \quad (2.4.11)$$

Then we have the following theorem.

Theorem 2.4.3. Consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, a collection of distinct vectors $v_1, \dots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$ and two collections of non-negative functions on $\mathcal{X}$, $\Upsilon = (\Upsilon_1, \dots, \Upsilon_n)$ and $\check{\Upsilon} = (\check{\Upsilon}_1, \dots, \check{\Upsilon}_n)$ such that (2.4.2) holds and the associated continuous time Markov chains do not explode in finite time. Consider a matrix $A \in \mathbb{Z}^{m \times d}$ with non-zero rows and suppose that both of the following conditions hold:

(i) For each $1 \leq j \leq n$, the vector Av_j has entries in $\{-1, 0, 1\}$ only.

(ii) For each $x \in \mathcal{X}$, $1 \leq i \leq m$ and $y \in \partial_i(K_A + x) \cap \mathcal{X}$ we have that

$$\sum_{j \in G^k} \check{\Upsilon}_j(y) \leq \sum_{j \in G^k} \Upsilon_j(x), \quad \text{for each } k \text{ such that } \eta_i^k < 0,$$

and

$$\sum_{j \in G^k} \check{\Upsilon}_j(y) \geq \sum_{j \in G^k} \Upsilon_j(x), \quad \text{for each } k \text{ such that } \eta_i^k > 0.$$

Then, for each pair $x^\circ, \check{x}^\circ \in \mathcal{X}$ such that $x^\circ \preceq_A \check{x}^\circ$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with two continuous time Markov chains $X = \{X(t), t \geq 0\}$ and $\check{X} = \{\check{X}(t), t \geq 0\}$ defined there, each having state space $\mathcal{X} \subseteq \mathbb{Z}_+^d$, with infinitesimal generators Q and $\check{Q}$, associated with Υ and $\check{\Upsilon}$ respectively, with initial conditions $X(0) = x^\circ$ and $\check{X}(0) = \check{x}^\circ$ and such that:

$$\mathbb{P} \left[X(t) \preceq_A \check{X}(t) \text{ for every } t \geq 0 \right] = 1. \quad (2.4.12)$$

The proof of this theorem is given in Section 2.6.3.

Remark 2.4.4. If $\Upsilon = \check{\Upsilon}$, Theorems 2.4.1, 2.4.2 and 2.4.3 give sufficient conditions for monotonic dependence of the stochastic dynamic behavior on the initial condition. In the sense of Massey [21], this notion corresponds to constructing a *strongly monotone Markov chain*.

Remark 2.4.5. For *deterministic* dynamical systems, there is a considerable literature giving monotonicity conditions with respect to initial conditions (see e.g., Hirsch and Smith [17]). Furthermore, Angeli and Sontag [4] extended the concept of monotone systems to systems having external inputs (i.e., $\dot{x} = f(x, u)$, with x representing the state and u representing the input). More precisely, they developed tools to prove monotonic dependence of the deterministic dynamic behavior on the initial condition and external input, provided that certain sign conditions on the first partial derivatives of the function $f(x, u)$ are satisfied on the entire state and input space. These theoretical tools can be

used also to study how changing a system parameter affects the deterministic behavior of the system, by viewing u as the system parameter of interest.

Remark 2.4.6. Checking the conditions in Theorems 2.4.2 and 2.4.3 (if they hold) is less cumbersome than checking the conditions in Theorem 2.4.1. In fact, compared to Theorem 2.4.1, for Theorems 2.4.2 and 2.4.3, the conditions must be checked only on the boundaries of $K_A + x$, given that condition (i) there is assumed to hold. Furthermore, Theorem 2.4.3 has less restrictive conditions (i.e., comparing sums of infinitesimal rates associated with transitions inwards or outwards with respect to the hyperplanes $\{z \in \mathbb{R}^d \mid \langle A_{i\bullet}, z \rangle = \langle A_{i\bullet}, x \rangle = \langle A_{i\bullet}, y \rangle\}$, $1 \leq i \leq m$, instead of comparing transition rates one-by-one for $1 \leq j \leq n$).

2.4.3 Monotonicity properties for (mean) first passage times and stationary distributions

The first consequence of our main results is for first passage times and it is related to stochastic orderings of real-valued random variables. Let Y and Z be real-valued random variables with cumulative distribution functions F_Y and F_Z respectively. We say that Y is smaller than Z in the **usual stochastic order**, written $Y \preceq_{st} Z$ if $F_Y(t) \geq F_Z(t)$ for every $t \in \mathbb{R}$. The relation $Y \preceq_{st} Z$ is equivalent to the existence of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with random variables $Y' \stackrel{\text{dist}}{=} Y$ and $Z' \stackrel{\text{dist}}{=} Z$ defined there such that $\mathbb{P}(Y' \leq Z') = 1$. The reader may consult Chapter 1 in Müller and Stoyan [24] for the corresponding proofs and further properties of this notion.

Theorem 2.4.4. *Consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, a collection of distinct vectors $v_1, \dots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$ and two collections of non-negative functions on $\mathcal{X}$, $\Upsilon = (\Upsilon_1, \dots, \Upsilon_n)$ and $\check{\Upsilon} = (\check{\Upsilon}_1, \dots, \check{\Upsilon}_n)$, such that (2.4.2) holds and the associated continuous time Markov chains do not explode in finite time. Consider a matrix $A \in \mathbb{R}^{m \times d}$ with non-zero rows and suppose that at least one of the following holds:*

(i) For every $x, y \in \mathcal{X}$ such that $x \preceq_A y$, conditions (2.4.4) and (2.4.5) are satisfied.

(ii) The matrix A has integer-valued entries and conditions (i) and (ii) in Theorem 2.4.2 are satisfied.

(iii) The matrix A has integer-valued entries and conditions (i) and (ii) in Theorem 2.4.3 are satisfied.

Let $x^\circ, \check{x}^\circ \in \mathcal{X}$ be such that $x^\circ \preceq_A \check{x}^\circ$ and let $X = \{X(t), t \geq 0\}$ and $\check{X} = \{\check{X}(t), t \geq 0\}$ be two continuous time Markov chains (possibly defined on different probability spaces), each having state space $\mathcal{X} \subseteq \mathbb{Z}_+^d$, with infinitesimal generators Q and $\check{Q}$, associated with Υ and $\check{\Upsilon}$ respectively, and with initial conditions $X(0) = x^\circ$ and $\check{X}(0) = \check{x}^\circ$. For a non-empty set $\Gamma \subseteq \mathcal{X}$, consider $T_\Gamma := \inf\{t \geq 0 \mid X(t) \in \Gamma\}$ and $\check{T}_\Gamma := \inf\{t \geq 0 \mid \check{X}(t) \in \Gamma\}$. If Γ is increasing in $\mathcal{X}$ with respect to the relation $\preceq_A$, then

$$\check{T}_\Gamma \preceq_{st} T_\Gamma, \quad (2.4.13)$$

and the mean first passage time of $\check{X}$ from $\check{x}^\circ$ to Γ is dominated by the mean first passage time of X from x° to Γ . If Γ is decreasing in $\mathcal{X}$ with respect to the relation $\preceq_A$, then

$$T_\Gamma \preceq_{st} \check{T}_\Gamma,$$

and the mean first passage time of X from x° to Γ is dominated by the mean first passage time of $\check{X}$ from $\check{x}^\circ$ to Γ .

Proof. By Theorem 2.4.1, 2.4.2 or 2.4.3, we can construct two versions of the processes X and $\check{X}$ on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with initial conditions x° and $\check{x}^\circ$, respectively, and such that (2.4.6) or (2.4.10) or (2.4.12) hold. We denote these versions again by X and $\check{X}$, and we observe that to show (2.4.13), it suffices to show that for an increasing set Γ , $\mathbb{P}[\check{T}_\Gamma \leq T_\Gamma] = 1$ for T_Γ and $\check{T}_\Gamma$ associated with these versions of X and $\check{X}$.

To see that this holds, let $\tilde{\Omega}$ be a set of probability one on which

$$X(t) \preceq_A \check{X}(t), \quad \text{for all } t \geq 0 \quad (2.4.14)$$

(this exists by (2.4.6), (2.4.10) or (2.4.12)). On $\{T_\Gamma = +\infty\}$, it is clear that $\check{T}_\Gamma \leq T_\Gamma$. For each $\omega \in \{T_\Gamma < +\infty\} \cap \tilde{\Omega}$ and $\varepsilon > 0$ there is $\tau_\varepsilon(\omega) \in [T_\Gamma(\omega), T_\Gamma(\omega) + \varepsilon)$ such that $X(\tau_\varepsilon(\omega)) \in \Gamma$ and by (2.4.14), $X(\tau_\varepsilon(\omega)) \preceq_A \check{X}(\tau_\varepsilon(\omega))$. And then, since Γ is increasing, $\check{X}(\tau_\varepsilon(\omega)) \in \Gamma$. It follows that $\check{T}_\Gamma(\omega) \leq T_\Gamma(\omega) + \varepsilon$ and letting $\varepsilon \rightarrow 0$ we obtain that $\check{T}_\Gamma(\omega) \leq T_\Gamma(\omega)$. It follows that $\mathbb{P}[\check{T}_\Gamma \leq T_\Gamma] = 1$. For the result on mean first passage times, let $\bar{F}_{T_\Gamma} := 1 - F_{T_\Gamma}$ and $\bar{F}_{\check{T}_\Gamma} := 1 - F_{\check{T}_\Gamma}$ represent the complementary cumulative distribution functions for T_Γ and $\check{T}_\Gamma$, respectively. Observe that (2.4.13) implies that $\bar{F}_{\check{T}_\Gamma} \leq \bar{F}_{T_\Gamma}$. For a non-negative random variable, the mean of the random variable is given by the Lebesgue integral of the complementary cumulative distribution function. Consequently, the mean first passage time for $\check{X}$ from $\check{x}^\circ$ to Γ is given by $\int_0^\infty \bar{F}_{\check{T}_\Gamma}(t) dt \leq \int_0^\infty \bar{F}_{T_\Gamma}(t) dt$, where the latter is the mean first passage time for X from x° to Γ . If Γ is decreasing, analogous arguments yield the results stated for that case. $\square$

The second consequence of our results provides a comparison result for stationary distributions.

Theorem 2.4.5. *Consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, a collection of distinct vectors $v_1, \dots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$ and two collections of non-negative functions on $\mathcal{X}$, $\Upsilon = (\Upsilon_1, \dots, \Upsilon_n)$ and $\check{\Upsilon} = (\check{\Upsilon}_1, \dots, \check{\Upsilon}_n)$, such that (2.4.2) holds and the associated continuous time Markov chains do not explode in finite time. Consider a matrix $A \in \mathbb{R}^{m \times d}$ with non-zero rows and suppose that at least one of the following holds:*

- (i) *For every $x, y \in \mathcal{X}$ such that $x \preceq_A y$, conditions (2.4.4) and (2.4.5) are satisfied.*
- (ii) *The matrix A has integer-valued entries and conditions (i) and (ii) in Theorem 2.4.2 are satisfied.*

(iii) The matrix A has integer-valued entries and conditions (i) and (ii) in Theorem 2.4.3 are satisfied.

Assume that the two continuous time Markov chains on the set $\mathcal{X}$ with infinitesimal generators Q and $\check{Q}$, associated with Υ and $\check{\Upsilon}$ respectively, are irreducible and positive recurrent on $\mathcal{X}$, and denote the associated stationary distributions by π and $\check{\pi}$, respectively. If $\Gamma \subseteq \mathcal{X}$ is a non-empty set that is increasing in $\mathcal{X}$ with respect to $\preceq_A$, then

$$\sum_{x \in \Gamma} \pi_x \leq \sum_{x \in \Gamma} \check{\pi}_x. \quad (2.4.15)$$

If $\Gamma \subseteq \mathcal{X}$ is a non-empty set that is decreasing in $\mathcal{X}$ with respect to $\preceq_A$, then

$$\sum_{x \in \Gamma} \check{\pi}_x \leq \sum_{x \in \Gamma} \pi_x. \quad (2.4.16)$$

Proof. As in the proof of Theorem 2.4.4, we can construct two versions of the processes X and $\check{X}$ on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ for some pair of initial conditions $x^\circ \preceq_A \check{x}^\circ$. If $\Gamma \subseteq \mathcal{X}$ is increasing, equation (2.4.6) or (2.4.10) or (2.4.12) yields that $\mathbb{P}(X(t) \in \Gamma) \leq \mathbb{P}(\check{X}(t) \in \Gamma)$ for every $t \geq 0$. By letting $t \rightarrow \infty$ and observing that the stationary distribution is the steady-state distribution under our assumptions of irreducibility and positive recurrence, we obtain (2.4.15). If Γ is decreasing, an analogous argument yields (2.4.16). $\square$

Remark 2.4.7. A special case of Theorems 2.4.4 and 2.4.5 is when $\Gamma = \{x\}$ for some maximal or minimal element $x \in \mathcal{X}$.

In the next section, we give examples which illustrate Theorem 2.4.2 (see Examples 2.5.1, 2.5.2, 2.5.4 and 2.5.5), Theorem 2.4.3 (see Example 2.5.3), Theorem 2.4.4 and Theorem 2.4.5 for continuous time Markov chains that are stochastic chemical reaction networks. For Examples 2.5.1, 2.5.2 and 2.5.3, the state space $\mathcal{X}$ will be a stoichiometric

compatibility class $z + \mathcal{L}$ intersected with $\mathbb{Z}_+^d$. For Examples 2.5.4 and 2.5.5, we work with reduced Markov chains and the state space $\mathcal{X}$ will be a projection of a suitable higher dimensional stoichiometric compatibility class $z + \mathcal{L}$ intersected with $\mathbb{Z}_+^d$.

2.5 Examples

In this section, we apply the theoretical tools developed in this work to several examples. While in Examples 2.5.1, 2.5.3 and 2.5.4 the Markov chains analyzed have a finite state space, in Examples 2.5.2 and 2.5.5 the Markov chains have a countably infinite state space, but it is straightforward to verify that they do not explode (see SI-Section 2.8.1). The choice of matrix A in each example is based on the specific monotonicity relationship of interest. While for simpler cases the choice of A is straightforward, for more complicated systems the choice can be more subtle. In many cases, in order to study the monotonicity properties for the stochastic behavior of our system, we can rely on Theorem 2.4.2, which provides a reasonable approach to narrow down the choices for suitable A . The approach consists in solving, for each row i , the system of equations $\sum_{k=1}^d A_{ik}(v_j)_k = b_{ij}$, with b_{ij} equal to 1, -1 , or 0 depending, based on the monotonicity relationship of interest, whether we expect that the Markov chain transition in the direction v_j leads inside, outside, or is parallel to the boundary of the region $K_A + x$. Finally, it is worth noticing that, while all the following examples compare two identical reaction networks with different rate constants, our theory can also be applied to compare two different reaction networks as long as they have the same net reaction vectors $\{v_j\}_{j=1}^n$.

Example 2.5.1. Enzyme kinetics I

Let us consider a classic model of enzyme kinetics (see Michaelis and Menten [23] and Kang et al. [19]), where an enzyme catalyzes the conversion of a substrate to a product. The species considered here are substrate (S), enzyme (E), intermediate enzyme-substrate complex (SE), and product (P), and the chemical reaction system is depicted in Figure

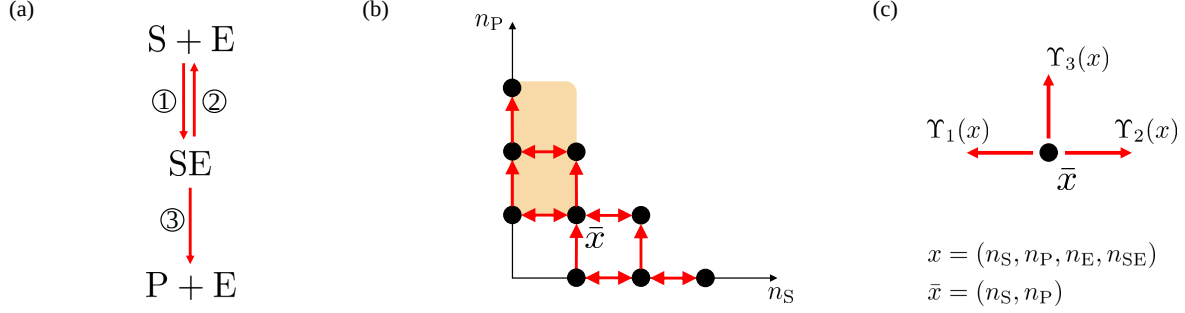


Figure 2.2. Reaction model and corresponding Markov chain for enzymatic kinetics I example. (a) Chemical reaction system. The numbers on the arrows correspond to the associated reactions. (b) Projected Markov chain graph for one stoichiometric compatibility class with two conservation laws $n_S + n_P + n_{SE} = S_{\text{tot}} = 3$ and $n_E + n_{SE} = E_{\text{tot}} = 2$. The projection takes a state $x = (n_S, n_P, n_E, n_{SE}) = (n_S, n_P, E_{\text{tot}} - S_{\text{tot}} + n_S + n_P, S_{\text{tot}} - n_S - n_P)$ to $\bar{x} = (n_S, n_P)$. We use black dots to represent the states, red double-ended arrows to represent transitions in both directions and red single-ended arrows to represent transitions in one direction. Note that $\bar{x} = (0, 0)$ is not a vertex in the graph because $0 \leq n_E = E_{\text{tot}} - S_{\text{tot}} + n_S + n_P$, and so $n_S + n_P \geq 3 - 2 = 1$. We use orange to highlight the projection of the region $K_A + x$ intersected with the stoichiometric compatibility class, where A is defined in (2.5.2). (c) The projections of the directions of the possible transitions of the Markov chain. The transition rates $\Upsilon_1(x)$, $\Upsilon_2(x)$, and $\Upsilon_3(x)$ are defined in (2.5.1).

2.2(a). We are interested in how the rate constant κ_3 affects the time to convert the substrate to the final product.

To this end, let us first introduce the set of species $\mathcal{S} = \{S, P, E, SE\}$, and the set of reactions $\mathcal{R} = \{(v_1^-, v_1^+), (v_2^-, v_2^+), (v_3^-, v_3^+)\}$, where $v_1^- = v_2^+ = (1, 0, 1, 0)^T$, $v_1^+ = v_2^- = v_3^- = (0, 0, 0, 1)^T$, $v_3^+ = (0, 1, 1, 0)^T$, where T denotes transpose. At a given time, let the counts of each of the species S, P, E and SE be denoted by n_S , n_P , n_E and n_{SE} , respectively. The state of the associated Markov chain is (n_S, n_P, n_E, n_{SE}) . The potential transitions of the Markov chain are in three possible directions:

$$v_1 = v_1^+ - v_1^- = (-1, 0, -1, 1)^T, v_2 = v_2^+ - v_2^- = (1, 0, 1, -1)^T, v_3 = v_3^+ - v_3^- = (0, 1, 1, -1)^T.$$

Fixing integers $S_{\text{tot}}, E_{\text{tot}} > 0$, we have a stoichiometric compatibility class $z + \mathcal{L}$ with $z = (S_{\text{tot}}, 0, E_{\text{tot}}, 0)$ and $\mathcal{L} := \text{span}\{v_1, v_2, v_3\}$, which is contained in a two-dimensional

affine subspace of four dimensional space. Then, the state space of the Markov chain is

$$\mathcal{X} = (z + \mathcal{L}) \cap \mathbb{Z}_+^4 = \{(x_1, x_2, x_3, x_4) \in \mathbb{Z}_+^4 | x_1 + x_2 + x_4 = S_{\text{tot}}, x_3 + x_4 = E_{\text{tot}}\}.$$

The two constraints described in the last expression for $\mathcal{X}$ characterize the two linearly independent conservation laws for this chemical reaction system: $n_S + n_P + n_{SE} = S_{\text{tot}}$ and $n_E + n_{SE} = E_{\text{tot}}$.

Given a state $x = (x_1, x_2, x_3, x_4) \in \mathcal{X}$, following mass-action kinetics, the infinitesimal transition rates are

$$\Upsilon_1(x) = \kappa_1 x_1 x_3, \quad \Upsilon_2(x) = \kappa_2 x_4, \quad \Upsilon_3(x) = \kappa_3 x_4, \quad (2.5.1)$$

for constants $\kappa_1, \kappa_2, \kappa_3 > 0$. Here, we have used κ_j as an abbreviation for $\kappa_{(v_j^-, v_j^+)}$, $j = 1, 2, 3$. We will use similar abbreviations in the other examples too.

We note that the projected process $(X_1, X_2)(\cdot)$ is still a continuous time Markov chain, and we could apply our theory to it. However, when the functions Υ_j , $j = 1, 2, 3$, are written in terms of these two components, they will have a more complex, non-mass action form. Here we apply our theory directly to our four dimensional Markov chain. For the purpose of visualization, Figure 2.2(b) shows the two dimensional projection of the four dimensional Markov chain graph for one stoichiometric compatibility class. In Examples 2.5.2 and 2.5.3, we also analyze Markov chains without projections, and in Examples 2.5.4 and 2.5.5, we analyze projected Markov chains.

In order to study how the rate constant κ_3 affects the time to convert the substrate to the final product, let us define the state $(0, S_{\text{tot}}, E_{\text{tot}}, 0)$ associated with $n_P = S_{\text{tot}}$ as p , the state $(S_{\text{tot}}, 0, E_{\text{tot}}, 0)$ associated with $n_S = S_{\text{tot}}$ as s , and the mean first passage time to reach the state p , starting from s , as $\mathbb{E}_s[T_p]$. We will verify that the assumptions of Theorems 2.4.2, 2.4.4 hold and exploit them to determine how κ_3 affects $\mathbb{E}_s[T_p]$. To this

end, define the matrix

$$A = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad (2.5.2)$$

and consider the preorder $x \preceq_A y$, defined by $A(y - x) \geq 0$, and the set $K_A + x = \{w \in \mathbb{R}^4 \mid x \preceq_A w\}$. Let us also consider the infinitesimal transition rates $\check{\Upsilon}_1(x)$, $\check{\Upsilon}_2(x)$ and $\check{\Upsilon}_3(x)$ defined as for $\Upsilon_1(x)$, $\Upsilon_2(x)$ and $\Upsilon_3(x)$, but with $\check{\kappa}_1 = \kappa_1$, $\check{\kappa}_2 = \kappa_2$, $\check{\kappa}_3 > \kappa_3$ in place of κ_1 , κ_2 , κ_3 , respectively. Condition (i) of Theorem 2.4.2 (i.e., for every $1 \leq j \leq n$, the vector Av_j has entries in $\{-1, 0, 1\}$) holds since $Av_1 = (1, 0)^T$, $Av_2 = (-1, 0)^T$ and $Av_3 = (0, 1)^T$. Condition (ii) of Theorem 2.4.2 also holds, as shown in the paragraph below.

Verification of condition (ii) of Theorem 2.4.2. We first consider $x \in \mathcal{X}$ and $y \in \partial_1(K_A + x) \cap \mathcal{X}$, where

$$\begin{aligned} & \partial_1(K_A + x) \cap \mathcal{X} \\ = & \{w \in \mathbb{Z}_+^4 \mid x_1 = w_1, x_2 \leq w_2\} \cap \mathcal{X} \\ = & \{w \in \mathbb{Z}_+^4 \mid x_1 = w_1, x_2 \leq w_2, x_1 + x_2 + x_4 = w_1 + w_2 + w_4 = S_{\text{tot}}, \\ & \quad x_3 + x_4 = w_3 + w_4 = E_{\text{tot}}\} \\ = & \{w \in \mathbb{Z}_+^4 \mid x_1 = w_1, x_2 \leq w_2, x_3 \leq w_3, x_4 \geq w_4, w_1 + w_2 + w_4 = S_{\text{tot}}, w_3 + w_4 = E_{\text{tot}}\} \\ = & \{w \in \mathcal{X} \mid x_1 = w_1, x_2 \leq w_2, x_3 \leq w_3, x_4 \geq w_4\}. \end{aligned}$$

Since $\langle A_{1\bullet}, v_1 \rangle = 1$, $\langle A_{1\bullet}, v_2 \rangle = -1$, we need to check that $\Upsilon_1(x) \leq \check{\Upsilon}_1(y)$ and $\Upsilon_2(x) \geq \check{\Upsilon}_2(y)$. The first inequality holds because $y \in \partial_1(K_A + x) \cap \mathcal{X}$ implies $x_1 = y_1$ and $x_3 \leq y_3$ so that $\Upsilon_1(x) = \kappa_1 x_1 x_3 \leq \kappa_1 y_1 y_3 = \check{\kappa}_1 y_1 y_3 = \check{\Upsilon}_1(y)$. The second inequality holds because $y \in \partial_1(K_A + x) \cap \mathcal{X}$ implies $x_4 \geq y_4$ so that $\Upsilon_2(x) = \kappa_2 x_4 \geq \kappa_2 y_4 = \check{\kappa}_2 y_4 = \check{\Upsilon}_2(y)$. Secondly, we consider $x \in \mathcal{X}$, $y \in \partial_2(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 \geq w_1, x_2 = w_2, x_3 \geq w_3, x_4 \leq w_4\}$. Then, since $\langle A_{2\bullet}, v_3 \rangle = 1$, we need to check that $\Upsilon_3(x) \leq \check{\Upsilon}_3(y)$. This holds because $y \in \partial_2(K_A + x) \cap \mathcal{X}$ implies $x_4 \leq y_4$ so that $\Upsilon_3(x) = \kappa_3 x_4 \leq \kappa_3 y_4 \leq \check{\kappa}_3 y_4 = \check{\Upsilon}_3(y)$.

Since all of the hypotheses of Theorem 2.4.2 hold, we can conclude that, for each

$x^\circ, \check{x}^\circ \in \mathcal{X}$ with $x^\circ \preccurlyeq_A \check{x}^\circ$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with two Markov chains $X = \{X(t), t \geq 0\}$ and $\check{X} = \{\check{X}(t), t \geq 0\}$ associated with Υ and $\check{\Upsilon}$, respectively, such that $X(0) = x^\circ$, $\check{X}(0) = \check{x}^\circ$ and

$$\mathbb{P} \left[X(t) \preccurlyeq_A \check{X}(t) \text{ for every } t \geq 0 \right] = 1.$$

Furthermore, applying Theorem 2.4.4 with the set $\Gamma = \{p\} = \{(0, S_{\text{tot}}, E_{\text{tot}}, 0)\}$, which is increasing in $\mathcal{X}$ with respect to $\preccurlyeq_A$, we see that the mean first passage time from s to p , $\mathbb{E}_s[T_p]$, is a decreasing function of κ_3 .

Because the Markov chain has one absorbing state, p , per stoichiometric compatibility class, the stationary distribution on a given stoichiometric compatibility class is trivial, and hence so too are its monotonicity properties.

Example 2.5.2. Enzyme kinetics II

Let us consider an extension of the enzymatic kinetics model introduced in the previous example, in which the substrate S can enter and leave the system and the product can revert to the substrate. This is a simplified version of the enzymatic kinetics considered by Anderson et al. [1]. The chemical reaction system is depicted in Figure 2.3(a).

Now, for this case study, we first determine how the reaction rate constant κ_5 affects the stochastic behavior of the system and then we will study properties of the system with respect to initial conditions. To this end, let us introduce the set of species $\mathcal{S} = \{S, P, E, SE\}$, and, similar to Example 2.5.1, we let (n_S, n_P, n_E, n_{SE}) be the state of the Markov chain that records the number of molecules of each species. The potential transitions of the Markov chain are in six possible directions, v_j for $j = 1, \dots, 6$, where $v_1 = -v_2 = (-1, 0, -1, 1)^T$, $v_3 = -v_4 = (0, 1, 1, -1)^T$, and $v_5 = -v_6 = (1, 0, 0, 0)^T$ (see SI-Section 2.8.2 for the derivation of the v_j , $j = 1, \dots, 6$). Since there is one linearly independent conservation law in this chemical reaction system: $n_E + n_{SE} = E_{\text{tot}}$, each stoichiometric compatibility class is contained in a three-dimensional affine subspace of

four dimensional space, denoted as $z + \mathcal{L}$, where $z = (0, 0, E_{\text{tot}}, 0)$ and $\mathcal{L} := \text{span}\{v_1, v_3, v_5\}$, with fixed integer $E_{\text{tot}} > 0$. Then, we can choose the state space of the Markov chain to be $\mathcal{X} = (z + \mathcal{L}) \cap \mathbb{Z}_+^4 = \{(x_1, x_2, x_3, x_4) \in \mathbb{Z}_+^4 | x_3 + x_4 = E_{\text{tot}}\}$. Furthermore, given a state $x = (x_1, x_2, x_3, x_4) \in \mathcal{X}$, following mass-action kinetics, the associated infinitesimal transition rates are given by

$$\begin{aligned} \Upsilon_1(x) &= \kappa_1 x_1 x_3, & \Upsilon_2(x) &= \kappa_2 x_4, & \Upsilon_3(x) &= \kappa_3 x_4, \\ \Upsilon_4(x) &= \kappa_4 x_2 x_3, & \Upsilon_5(x) &= \kappa_5, & \Upsilon_6(x) &= \kappa_6 x_1, \end{aligned} \tag{2.5.3}$$

for $\kappa_1, \kappa_2, \kappa_3, \kappa_4, \kappa_5, \kappa_6 > 0$. As in Example 2.5.1, we apply our theory directly to our four dimensional Markov chain, but, for the purpose of illustration, Figure 2.3(b) shows the three dimensional projection of the Markov chain graph for one stoichiometric compatibility class.

Now, for the first analysis (determining how κ_5 affects the stochastic behavior of the system), we verify that the assumptions of Theorems 2.4.2 and 2.4.5 hold and use them to determine how κ_5 affects the stationary distribution. To this end, define the matrix

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix} \tag{2.5.4}$$

and consider the preorder $x \preceq_A y$, defined by $A(y - x) \geq 0$. For $x \in \mathcal{X}$, $K_A + x = \{w \in \mathbb{R}^4 | x \preceq_A w\}$. Furthermore, let us consider the infinitesimal transition rates $\check{\Upsilon}_1(x)$, $\check{\Upsilon}_2(x)$, $\check{\Upsilon}_3(x)$, $\check{\Upsilon}_4(x)$, $\check{\Upsilon}_5(x)$ and $\check{\Upsilon}_6(x)$ defined as for $\Upsilon_1(x)$, $\Upsilon_2(x)$, $\Upsilon_3(x)$, $\Upsilon_4(x)$, $\Upsilon_5(x)$ and $\Upsilon_6(x)$, but with $\check{\kappa}_i$ in place of κ_i , where $\check{\kappa}_i = \kappa_i$, for $i = 1, 2, 3, 4, 6$, and $\check{\kappa}_5 \geq \kappa_5$. Given that $Av_1 = (-1, 0, 1)^T$, $Av_2 = (1, 0, -1)^T$, $Av_3 = (0, 1, -1)^T$, $Av_4 = (0, -1, 1)^T$, $Av_5 = (1, 0, 0)^T$ and $Av_6 = (-1, 0, 0)^T$, we have that condition (i) of Theorem 2.4.2 holds. Condition (ii) of that theorem also holds, as shown in the next paragraph.

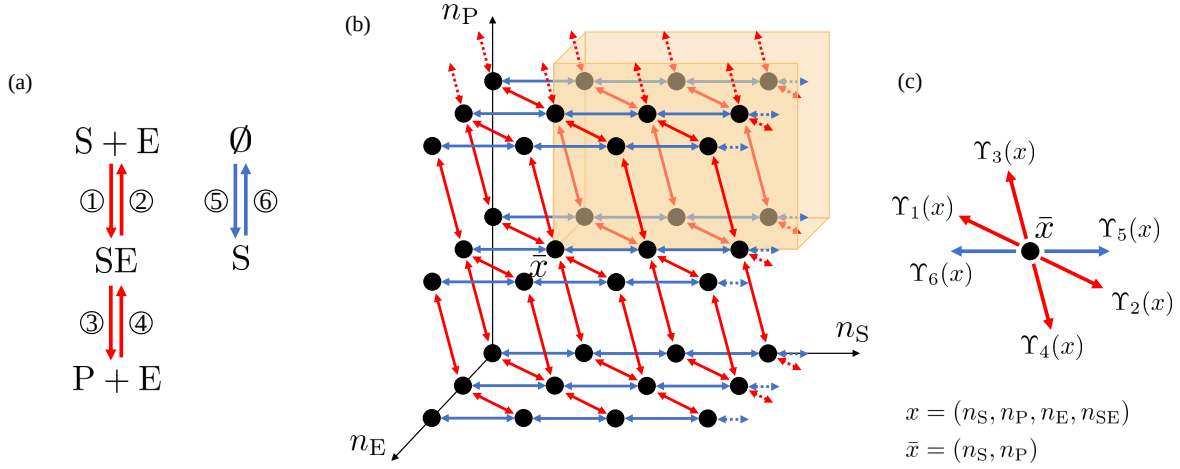


Figure 2.3. Reaction model and corresponding Markov chain for enzymatic kinetics II example. (a) Chemical reaction system. The numbers on the arrows correspond to the associated reactions. (b) Projected Markov chain graph for one stoichiometric compatibility class with the conservation law $n_E + n_{SE} = E_{\text{tot}} = 2$. The projection takes a state $x = (n_S, n_P, n_E, n_{SE}) = (n_S, n_P, n_E, E_{\text{tot}} - n_E) \in \mathcal{X}$ to $\bar{x} = (n_S, n_P, n_E) \in \mathbb{Z}_+^3 : 0 \leq n_E \leq 2$. Here, we use black dots to represent the states, red double-ended arrows to represent transitions in both directions associated with the reactions represented by the red arrows in (a) and blue double-ended arrows to represent transitions in both directions associated with the reactions represented by the blue arrows in (a). We use dotted arrowed-lines to indicate that the pattern of Markov chain transitions extends to infinity. We use orange to highlight the projections of the region $K_A + x$ intersected with the stoichiometric compatibility class, where A is defined in (2.5.4). (c) The projections of the directions of the possible transitions of the Markov chain within a stoichiometric compatibility class. The transition rates $\Upsilon_i(x)$, $i = 1, 2, 3, 4, 5, 6$, are defined in (2.5.3).

Verification of condition (ii) of Theorem 2.4.2. First consider $x \in \mathcal{X}$ and $y \in \partial_1(K_A + x) \cap \mathcal{X}$, where $\partial_1(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 = w_1, x_2 \leq w_2, x_3 \geq w_3, x_4 \leq w_4\}$. Since $\langle A_{1\bullet}, v_2 \rangle = \langle A_{1\bullet}, v_5 \rangle = 1$ and $\langle A_{1\bullet}, v_1 \rangle = \langle A_{1\bullet}, v_6 \rangle = -1$, we need to check that $\Upsilon_1(x) \geq \check{\Upsilon}_1(y)$, $\Upsilon_6(x) \geq \check{\Upsilon}_6(y)$, $\Upsilon_2(x) \leq \check{\Upsilon}_2(y)$, and $\Upsilon_5(x) \leq \check{\Upsilon}_5(y)$. Given that $y \in \partial_1(K_A + x) \cap \mathcal{X}$, the first inequality holds because $\Upsilon_1(x) = \kappa_1 x_1 x_3 \geq \kappa_1 y_1 y_3 = \check{\kappa}_1 y_1 y_3 = \check{\Upsilon}_1(y)$, the second inequality holds because $\Upsilon_6(x) = \kappa_6 x_1 = \kappa_6 y_1 = \check{\kappa}_6 y_1 = \check{\Upsilon}_6(y)$, the third inequality holds because $\Upsilon_2(x) = \kappa_2 x_4 \leq \kappa_2 y_4 = \check{\kappa}_2 y_4 = \check{\Upsilon}_2(y)$, and the fourth inequality holds because $\Upsilon_5(x) = \kappa_5 \leq \check{\kappa}_5 = \check{\Upsilon}_5(y)$. Secondly, we consider $x \in \mathcal{X}$ and $y \in \partial_2(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 \leq w_1, x_2 = w_2, x_3 \geq w_3, x_4 \leq w_4\}$. Given that

$\langle A_{3\bullet}, v_3 \rangle = 1$ and $\langle A_{3\bullet}, v_4 \rangle = -1$, we need to check that $\Upsilon_4(x) \geq \check{\Upsilon}_4(y)$ and $\Upsilon_3(x) \leq \check{\Upsilon}_3(y)$. The first inequality holds because $\Upsilon_4(x) = \kappa_4 x_2 x_3 \geq \kappa_4 y_2 y_3 = \check{\kappa}_4 y_2 y_3 = \check{\Upsilon}_4(y)$ and the second inequality holds because $\Upsilon_3(x) = \kappa_3 x_4 \leq \kappa_3 y_4 = \check{\kappa}_3 y_4 = \check{\Upsilon}_3(y)$. Finally, consider $x \in \mathcal{X}$ and $y \in \partial_3(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 \leq w_1, x_2 \leq w_2, x_3 = w_3, x_4 = w_4\}$. Since $\langle A_{3\bullet}, v_1 \rangle = \langle A_{3\bullet}, v_4 \rangle = 1$ and $\langle A_{3\bullet}, v_2 \rangle = \langle A_{3\bullet}, v_3 \rangle = -1$, we need to check that $\Upsilon_2(x) \geq \check{\Upsilon}_2(y)$, $\Upsilon_3(x) \geq \check{\Upsilon}_3(y)$, $\Upsilon_1(x) \leq \check{\Upsilon}_1(y)$, and $\Upsilon_4(x) \leq \check{\Upsilon}_4(y)$. Indeed, we have that $\Upsilon_2(x) = \kappa_2 x_4 = \kappa_2 y_4 = \check{\kappa}_2 y_4 = \check{\Upsilon}_2(y)$, $\Upsilon_3(x) = \kappa_3 x_4 = \kappa_3 y_4 = \check{\kappa}_3 y_4 = \check{\Upsilon}_3(y)$, $\Upsilon_1(x) = \kappa_1 x_1 x_3 \leq \kappa_1 y_1 y_3 = \check{\kappa}_1 y_1 y_3 = \check{\Upsilon}_1(y)$, and $\Upsilon_4(x) = \kappa_4 x_2 x_3 \leq \kappa_4 y_2 y_3 = \check{\kappa}_4 y_2 y_3 = \check{\Upsilon}_4(y)$.

Thus, all of the hypothesis of Theorem 2.4.2 are verified, and so, for each pair $x^\circ, \check{x}^\circ \in \mathcal{X}$ satisfying $x^\circ \preccurlyeq_A \check{x}^\circ$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with two Markov chains $X = \{X(t), t \geq 0\}$ and $\check{X} = \{\check{X}(t), t \geq 0\}$ associated with Υ and $\check{\Upsilon}$, respectively, such that $X(0) = x^\circ$, $\check{X}(0) = \check{x}^\circ$ and $\mathbb{P}[X(t) \preccurlyeq_A \check{X}(t) \text{ for every } t \geq 0] = 1$. The Markov chains $X, \check{X}$ are irreducible and positive recurrent (see SI-Section 2.8.1). Furthermore, for the increasing set in $\mathcal{X}$ with respect to $\preccurlyeq_A$ defined as $\Gamma(x) = \{w \in \mathcal{X} \mid x_1 \leq w_1, x_2 \leq w_2, x_3 \geq w_3, x_4 \leq w_4\}$, we can apply Theorem 2.4.5 and obtain that $\sum_{w \in \Gamma(x)} \pi_w \leq \sum_{w \in \Gamma(x)} \check{\pi}_w$. Loosely speaking, this means that increasing κ_5 causes the stationary distribution $\pi(x)$ to shift mass towards states characterized by lower x_3 and higher x_1, x_2 and x_4 .

For this specific case, in which we have a stochastic chemical reaction network associated with a complex balanced dynamical system, an explicit expression for the stationary distribution can be obtained by applying Theorem 4.1 in Anderson et al. [1]. Analysis of this formula would provide results in agreement with the ones obtained by applying the theoretical tools developed in this work. Specifically, π_x can be written as a product of two Poisson distributions and a binomial distribution, i.e.,

$$\pi_x = \left(e^{-c_1} \frac{c_1^{x_1}}{x_1!} \right) \left(e^{-c_2} \frac{c_2^{x_2}}{x_2!} \right) \left(\text{E}_{\text{tot}}! \frac{c_3^{x_3} c_4^{x_4}}{x_3! x_4!} \right), \quad x \in \mathcal{X},$$

in which (c_1, c_2, c_3, c_4) represents the complex balanced equilibrium for the deterministic model, where

$$c_1 = \frac{\kappa_5}{\kappa_6}, \quad c_2 = \frac{\kappa_1 \kappa_3 \kappa_5}{\kappa_2 \kappa_4 \kappa_6}, \quad c_3 = \frac{1}{1 + \frac{\kappa_1 \kappa_5}{\kappa_2 \kappa_6}}, \quad \text{and} \quad c_4 = \frac{\frac{\kappa_1 \kappa_5}{\kappa_2 \kappa_6}}{1 + \frac{\kappa_1 \kappa_5}{\kappa_2 \kappa_6}}.$$

In most cases, it is not possible to derive an analytical formula for the stationary distribution, but our theorems can still be applied and then monotonicity properties for π can still be determined even without an explicit expression for π . For instance, in the context of the above example, if the infinitesimal transition rates Υ_i do not follow mass-action kinetics, the deficiency zero theorem and Theorem 4.1 in Anderson et al. [1] do not apply. Nevertheless, our theory can still be easily applied to study monotonicity properties for sample paths and stationary distributions.

As pointed out in Remark 2.4.4, we can also exploit our theoretical tools to determine monotonicity properties of the system with respect to the initial conditions. For this, suppose that $\check{\kappa}_i = \kappa_i$ for $i = 1, 2, 3, 4, 5, 6$. Then, by the analysis above, Theorem 2.4.2 holds and yields monotonically (with preorder induced by the matrix A) with respect to the initial conditions.

Example 2.5.3. A network topology arising in Braess' paradox

A natural question in synthetic biology may involve the prediction of whether an engineered biological circuit with additional reactions will lead to the desired effect of accelerating the process or unexpected behaviors. Now, we consider an example inspired by Braess' paradox, which arises from transportation networks, where adding one or more roads to a road network can slow down overall traffic flow through the network (see Braess [6] and see also a related state-dependent queuing network model in Calvert et al. [8]). A simple network of this type is one where there are two routes to get from the start to the final destination, and adding a linkage road between the routes can in some cases increase travel times. Figure 2.4(a) shows a reaction network analogue of the Braess' network

topology. Of course, our chemical reaction network is a little different from a road network since there is no congestion nor competition between molecules and pathways are chosen randomly with certain probabilities instead of routing decisions being based on the number of cars on the routes. Nevertheless, the example considered here is interesting because adding a reaction to cross-link two pathways might intuitively be interpreted as a detour and be expected to increase the time to the final destination, while this is sometimes not the case in this example.

The chemical reaction system is depicted in Figure 2.4(a), which involves four species $\mathcal{S} = \{S_1, S_2, S_3, S_4\}$. The state of the Markov chain is $(n_{S_1}, n_{S_2}, n_{S_3}, n_{S_4})$ where n_{S_i} is the number of copies of S_i for $i = 1, 2, 3, 4$. The potential transitions of the Markov chain are in five possible directions, v_j , $j = 1, \dots, 5$, where $v_1 = (-1, 1, 0, 0)^T$, $v_2 = (0, -1, 0, 1)^T$, $v_3 = (-1, 0, 1, 0)^T$, $v_4 = (0, 0, -1, 1)^T$ and $v_5 = (0, -1, 1, 0)^T$ (see SI-Section 2.8.2 for the derivation of the v_j , $j = 1, \dots, 5$). Fixing an integer $S_{\text{tot}} > 0$, the associated stoichiometric compatibility class is $z + \mathcal{L}$ with $z = (S_{\text{tot}}, 0, 0, 0)$ and $\mathcal{L} := \text{span}\{v_1, v_2, v_3, v_4, v_5\}$. The set $z + \mathcal{L}$ is a three-dimensional affine subspace of four dimensional space. We choose the state space of our Markov chain to be

$$\mathcal{X} = (z + \mathcal{L}) \cap \mathbb{Z}_+^4 = \{(x_1, x_2, x_3, x_4) \in \mathbb{Z}_+^4 | x_1 + x_2 + x_3 + x_4 = S_{\text{tot}}\}.$$

The constraint introduced in the last expression for $\mathcal{X}$ follows from the conservation law in this chemical reaction system, that is $n_{S_1} + n_{S_2} + n_{S_3} + n_{S_4} = S_{\text{tot}}$. Given a generic state $x = (x_1, x_2, x_3, x_4)$, following mass-action kinetics, the infinitesimal transition rates are

$$\Upsilon_1(x) = \kappa_1 x_1, \quad \Upsilon_2(x) = \kappa_2 x_2, \quad \Upsilon_3(x) = \kappa_3 x_1, \quad \Upsilon_4(x) = \kappa_4 x_3, \quad \Upsilon_5(x) = \kappa_5 x_2. \quad (2.5.5)$$

For the purpose of illustration, Figure 2.4(b) shows the three dimensional projection of the Markov chain graph for one stoichiometric compatibility class.

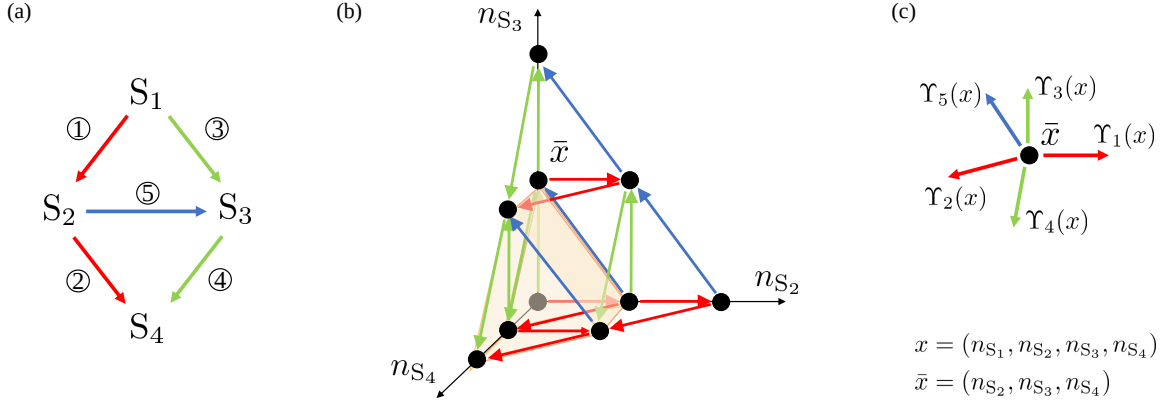


Figure 2.4. Circuit inspired by Braess' paradox and corresponding Markov chain. (a) Chemical reaction system. The numbers on the arrows correspond to the associated reactions. (b) Projected Markov chain graph for one stoichiometric compatibility class with the conservation law $n_{S_1} + n_{S_2} + n_{S_3} + n_{S_4} = S_{\text{tot}} = 2$. The projection takes a state $x = (n_{S_1}, n_{S_2}, n_{S_3}, n_{S_4}) = (S_{\text{tot}} - n_{S_2} - n_{S_3} - n_{S_4}, n_{S_2}, n_{S_3}, n_{S_4}) \in \mathcal{X}$ to $\bar{x} = (n_{S_2}, n_{S_3}, n_{S_4})$. Here, we use black dots to represent the states and red (blue, green) arrows to represent transitions in directions associated with the reactions represented by the red (blue, green) arrows in (a). We use orange to highlight the projection of the region $K_A + x$ intersected with the stoichiometric compatibility class, where A is defined in (2.5.6). (c) The projections of the directions of the possible transitions of the Markov chain within a stoichiometric compatibility class. The transition rates $\Upsilon_i(x)$, $i = 1, 2, 3, 4, 5$, are given in (2.5.5).

A natural question is how the time $T_{(0,0,0,S_{\text{tot}})}$ to reach the state $(0, 0, 0, S_{\text{tot}})$ from $(S_{\text{tot}}, 0, 0, 0)$ depends on the rate constants $\kappa_1, \kappa_2, \kappa_3, \kappa_4$ and κ_5 . For this, we use Theorem 2.4.4. Let

$$A = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 \end{bmatrix}. \quad (2.5.6)$$

The matrix A here defines a preorder that is not a partial order of $\mathcal{X}$. For $x \in \mathcal{X}$, consider infinitesimal transition rates $\check{\Upsilon}_1(x), \check{\Upsilon}_2(x), \check{\Upsilon}_3(x), \check{\Upsilon}_4(x)$ and $\check{\Upsilon}_5(x)$ defined as for $\Upsilon_1(x), \Upsilon_2(x), \Upsilon_3(x), \Upsilon_4(x)$ and $\Upsilon_5(x)$, but with $\check{\kappa}_i$ in place of κ_i where $\check{\kappa}_i = \kappa_i$, for $i = 1, 2, 3, 4$, and $\check{\kappa}_5 \neq \kappa_5$. Suppose that $\kappa_2 = \kappa_4$. Now, let us verify that the assumptions of Theorem 2.4.3 hold. Condition (i) holds since $Av_1 = (1, -1)^T$, $Av_2 = (0, 1)^T$, $Av_3 = (1, -1)^T$, $Av_4 = (0, 1)^T$ and $Av_5 = (0, 0)^T$. Condition (ii) of Theorem 2.4.3 also

holds, as shown in the paragraph below.

Verification of condition (ii) of Theorem 2.4.3. Let $x \in \mathcal{X}$, and first consider $x \in \mathcal{X}$ and $y \in \partial_1(K_A + x) \cap \mathcal{X}$, where $\partial_1(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 = w_1, x_2 + x_3 \geq w_2 + w_3, x_4 \leq w_4\}$. Given that $Av_2 = Av_4, Av_1 = Av_3$, and $\langle A_{1\bullet}, v_1 \rangle = \langle A_{1\bullet}, v_3 \rangle = 1$, we need to check that $\Upsilon_1(x) + \Upsilon_3(x) \leq \check{\Upsilon}_1(y) + \check{\Upsilon}_3(y)$. Since $y \in \partial_1(K_A + x) \cap \mathcal{X}$, then $\Upsilon_1(x) = \kappa_1 x_1 = \kappa_1 y_1 = \check{\kappa}_1 y_1 = \check{\Upsilon}_1(y)$ and $\Upsilon_3(x) = \kappa_3 x_1 = \kappa_3 y_1 = \check{\kappa}_3 y_1 = \check{\Upsilon}_3(y)$, and so the desired inequality holds with equality. Secondly, consider $y \in \partial_2(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 \geq w_1, x_2 + x_3 = w_2 + w_3, x_4 \leq w_4\}$. Given that $Av_2 = Av_4, Av_1 = Av_3$, and $\langle A_{1\bullet}, v_1 \rangle = \langle A_{1\bullet}, v_3 \rangle = -1$ and $\langle A_{1\bullet}, v_2 \rangle = \langle A_{1\bullet}, v_4 \rangle = 1$, we need to check that $\Upsilon_2(x) + \Upsilon_4(x) \leq \check{\Upsilon}_2(y) + \check{\Upsilon}_4(y)$ and $\Upsilon_1(x) + \Upsilon_3(x) \geq \check{\Upsilon}_1(y) + \check{\Upsilon}_3(y)$. For $x \in \mathcal{X}$ and $y \in \partial_2(K_A + x) \cap \mathcal{X}$, we have that $\Upsilon_2(x) + \Upsilon_4(x) = \kappa_2 x_2 + \kappa_4 x_3 = \kappa_2(x_2 + x_3) \leq \kappa_2(y_2 + y_3) = \check{\kappa}_2(y_2 + y_3) = \check{\Upsilon}_2(y) + \check{\Upsilon}_4(y)$ and $\Upsilon_1(x) = \kappa_1 x_1 \geq \kappa_1 y_1 = \check{\kappa}_1 y_1 = \check{\Upsilon}_1(y)$, $\Upsilon_3(x) = \kappa_3 x_1 \geq \kappa_3 y_1 = \check{\kappa}_3 y_1 = \check{\Upsilon}_3(y)$.

Thus, all hypotheses of Theorem 2.4.3 hold, and so for every $x^\circ, \check{x}^\circ \in \mathcal{X}$ where $x^\circ \preccurlyeq_A \check{x}^\circ$ there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with two Markov chains $X = \{X(t), t \geq 0\}$ and $\check{X} = \{\check{X}(t), t \geq 0\}$ associated with Υ and $\check{\Upsilon}$, respectively, such that $X(0) = x^\circ, \check{X}(0) = \check{x}^\circ$ and $\mathbb{P} \left[X(t) \preccurlyeq_A \check{X}(t) \text{ for every } t \geq 0 \right] = 1$. Let $\Gamma = \{(0, 0, 0, S_{\text{tot}})\}$. This is an increasing set in $\mathcal{X}$ with respect to the relation $\preccurlyeq_A$. Let $T_{(0,0,0,S_{\text{tot}})}$, respectively $\check{T}_{(0,0,0,S_{\text{tot}})}$ be the first time that the Markov chain X , respectively $\check{X}$, reaches the set Γ . Then, by Theorem 2.4.4, if $X(0) = \check{X}(0) = (S_{\text{tot}}, 0, 0, 0)$, we have that $\check{T}_{(0,0,0,S_{\text{tot}})} \preccurlyeq_{st} T_{(0,0,0,S_{\text{tot}})}$. By interchanging $\check{\Upsilon}_5$ and κ_5 , we can conclude that $\check{T}_{(0,0,0,S_{\text{tot}})}$ and $T_{(0,0,0,S_{\text{tot}})}$ are stochastically equivalent (equal in distribution). It follows that the mean first passage time from $(S_{\text{tot}}, 0, 0, 0)$ to $(0, 0, 0, S_{\text{tot}})$ is insensitive to κ_5 when $\kappa_2 = \kappa_4$. This is naively counter-intuitive: since the fifth reaction re-routes some samples to another state where the last reaction has the same rate constant as the final reaction without re-routing, it should take a longer expected time since re-routing also takes some time. However, in reality, the presence of the fifth reaction also fastens the rate to transition from S_2 , and

this balances the time of re-routing. Most importantly, our theorem is able to capture this result without explicitly calculating the mean first passage time and allows us to reach the conclusion easily. We expect that in more complex situations, our method will be a valuable tool to establish monotonicity and insensitivity results.

Given that the Markov chain has one absorbing state per stoichiometric compatibility class, the stationary distribution for a given stoichiometric compatibility class is trivial, and hence so too are its monotonicity properties.

SI-Theorem 2.8.2 allows us to conclude further interesting properties for this network. Using two other A matrices (see SI-Section 2.8.3), we can conclude that adding reaction ⑤ (changing from $\kappa_5 = 0$ to $\kappa_5 > 0$) causes the mean first passage time from $(S_{\text{tot}}, 0, 0, 0)$ to $(0, 0, 0, S_{\text{tot}})$ to increase if $\kappa_2 > \kappa_4$ or to decrease if $\kappa_2 < \kappa_4$. More explicitly, this shows that there can be opposing effects on the mean first passage time with different choices of κ_2 and κ_4 when reaction ⑤ is added.

Example 2.5.4. Epigenetic regulation by chromatin modifications

Epigenetic regulation is the modification of the DNA structure, due to chromatin modifications, that determines if a gene is active or repressed. There are several chromatin modifications that can affect the DNA structure. Here, we will focus only on histone modifications. More precisely, we consider a ubiquitous model for a histone modification circuit (see Dodd et al. [10] and Bruno et al. [7]). The species considered are nucleosomes that are unmodified (D), modified with repressive modifications (D^R), and modified with activating modifications (D^A), and, in terms of molecular interactions, each histone modification autocatalyzes itself and promotes the erasure of the other one. The chemical reaction system considered is depicted in Figure 2.5(a). The amount of each species is represented by n_D , n_{D^R} and n_{D^A} , respectively, and their sum is conserved, that is $n_D + n_{D^R} + n_{D^A} = D_{\text{tot}}$, with D_{tot} representing the total number of nucleosomes within the gene.

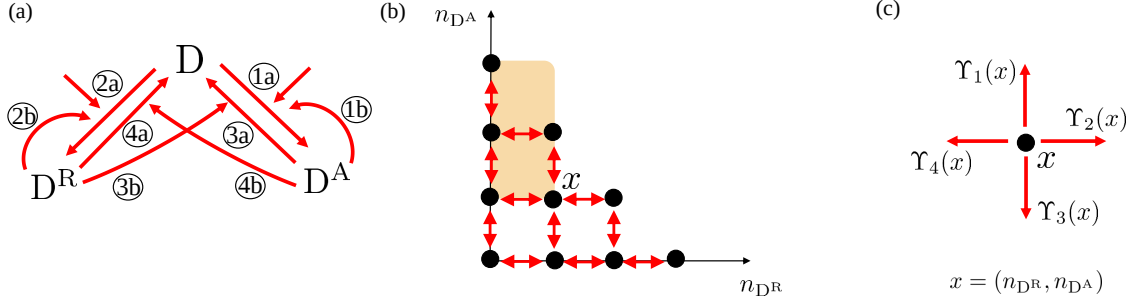
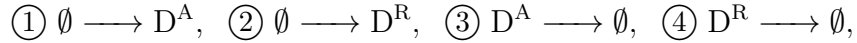


Figure 2.5. Histone modification circuit and corresponding Markov chain. (a) Original chemical reaction system. The numbers on the arrows correspond to the associated reactions. (b) Markov chain graph associated with the reduced chemical reaction system. Here, we consider $D_{\text{tot}} = 3$, we use black dots to represent the states and red double-ended arrows to represent transitions in both directions. We use orange to highlight the region $(K_A + x) \cap \mathcal{X}$, with A defined in (2.5.9). (c) Direction of the possible transitions of the Markov chains, whose rates are given in equation (2.5.8).

By fixing an integer $D_{\text{tot}} > 0$, we fix one stoichiometric compatibility class. The projected process $(X_1, X_2)(\cdot) = (n_{D^R}, n_{D^A})$ is still a continuous time Markov chain, and in this example we choose to apply our theory to this reduced system. This is the same as studying the reduced chemical reaction system defined as follows:



with two species $\mathcal{S} = \{D^R, D^A\}$ and four reactions $\mathcal{R} = \{(v_1^-, v_1^+), (v_2^-, v_2^+), (v_3^-, v_3^+), (v_4^-, v_4^+)\}$, where $v_1^- = v_2^- = v_3^+ = v_4^+ = (0, 0)^T$, $v_2^+ = v_4^- = (1, 0)^T$, $v_1^+ = v_3^- = (0, 1)^T$, and with associated propensity functions of non mass-action type defined as follows:

$$\begin{aligned} \Lambda_{(v_1^-, v_1^+)}(x) &= (D_{\text{tot}} - (x_1 + x_2)) (\kappa_{1a} + \kappa_{1b}x_2), \\ \Lambda_{(v_2^-, v_2^+)}(x) &= (D_{\text{tot}} - (x_1 + x_2)) (\kappa_{2a} + \kappa_{2b}x_1), \\ \Lambda_{(v_3^-, v_3^+)}(x) &= x_2 (\kappa_{3a} + x_1\kappa_{3b}), \quad \Lambda_{(v_4^-, v_4^+)}(x) = x_1\mu (c\kappa_{3a} + x_2\kappa_{3b}), \end{aligned} \tag{2.5.7}$$

in which κ_{1a} , κ_{1b} , κ_{3a} , κ_{3b} , κ_{2a} , κ_{2b} , $\kappa_{4a} = \mu c\kappa_{3a}$, $\kappa_{4b} = \mu\kappa_{3b}$ are the rate constants that go with each of the reactions shown in Figure 2.5(a), respectively.

The state space for the Markov chain is $\mathcal{X} = \{(x_1, x_2) \in \mathbb{Z}_+^2 \mid x_1 + x_2 \leq D_{\text{tot}}\}$. Given a generic state $x = (x_1, x_2) \in \mathcal{X}$, the potential transitions of the Markov chain are in four possible directions $v_j = v_j^+ - v_j^-$, $j = 1, 2, 3, 4$, that can be written as $v_1 = (0, 1)^T$, $v_2 = (1, 0)^T$, $v_3 = (0, -1)^T$ and $v_4 = (-1, 0)^T$, with associated infinitesimal transition rates

$$\Upsilon_1(x) = \Lambda_{(v_1^-, v_1^+)}(x), \quad \Upsilon_2(x) = \Lambda_{(v_2^-, v_2^+)}(x), \quad \Upsilon_3(x) = \Lambda_{(v_3^-, v_3^+)}(x), \quad \Upsilon_4(x) = \Lambda_{(v_4^-, v_4^+)}(x). \quad (2.5.8)$$

We are interested in determining how the asymmetry of the system, represented by the parameter μ affects the stochastic behavior of the system. In particular, we will focus on studying the stationary distribution and the *time to memory loss* of the active and repressed state, defined as the mean first passage time to reach the fully repressed state ($r = (n_{\text{DR}}, n_{\text{DA}}) = (D_{\text{tot}}, 0)$), starting from the fully active state ($a = (n_{\text{DR}}, n_{\text{DA}}) = (0, D_{\text{tot}})$), and vice versa (i.e., $h_{a,r} = \mathbb{E}_a[T_r]$ and $h_{r,a} = \mathbb{E}_r[T_a]$). To this end, we first verify that we can apply Theorem 2.4.2. Let

$$A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (2.5.9)$$

For $x \in \mathcal{X}$, $K_A + x = \{w \in \mathbb{R}^2 \mid x \preceq_A w\}$ and $(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x \preceq_A w\}$. See Figure 2.5(b) for an example of $\mathcal{X}$ and $(K_A + x) \cap \mathcal{X}$ for $D_{\text{tot}} = 3$. We introduce infinitesimal transition rates $\check{\Upsilon}_1(x)$, $\check{\Upsilon}_2(x)$, $\check{\Upsilon}_3(x)$ and $\check{\Upsilon}_4(x)$ defined as for $\Upsilon_1(x)$, $\Upsilon_2(x)$, $\Upsilon_3(x)$ and $\Upsilon_4(x)$, with all the parameters having the same values except that μ is replaced by $\check{\mu}$, where $\check{\mu} \geq \mu$. Since $Av_1 = (0, 1)^T$, $Av_2 = (-1, 0)^T$, $Av_3 = (0, -1)^T$ and $Av_4 = (1, 0)^T$, we have that condition (i) of Theorem 2.4.2 holds. Condition (ii) also holds, as shown in the paragraph below.

Verification of condition (ii) of Theorem 2.4.2. Consider $x \in \mathcal{X}$ and $y \in \partial_1(K_A + x) \cap \mathcal{X}$, where $\partial_1(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 = w_1, x_2 \leq w_2\}$. Since $\langle A_{1\bullet}, v_4 \rangle = 1$ and $\langle A_{1\bullet}, v_2 \rangle = -1$, we must check that $\Upsilon_2(x) \geq \check{\Upsilon}_2(y)$ and $\Upsilon_4(x) \leq$

$\check{\Upsilon}_4(y)$. Since $y \in \partial_1(K_A + x) \cap \mathcal{X}$ implies $x_1 = y_1$ and $x_2 \leq y_2$, we have $\Upsilon_2(x) = (D_{\text{tot}} - (x_1 + x_2))(\kappa_{2a} + \kappa_{2b}x_1) \geq (D_{\text{tot}} - (y_1 + y_2))(\kappa_{2a} + \kappa_{2b}y_1) = \check{\Upsilon}_2(y)$ and $\Upsilon_4(x) = x_1\mu(c\kappa_{3a} + x_2\kappa_{3b}) \leq y_1\mu(c\kappa_{3a} + y_2\kappa_{3b}) \leq y_1\check{\mu}(c\kappa_{3a} + y_2\kappa_{3b}) = \check{\Upsilon}_4(y)$, and so both inequalities hold. Similarly, for $x \in \mathcal{X}$ and $y \in \partial_2(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 \geq w_1, x_2 = w_2\}$, since $\langle A_{2\bullet}, v_1 \rangle = 1$ and $\langle A_{2\bullet}, v_3 \rangle = -1$, we need to check that $\Upsilon_1(x) \leq \check{\Upsilon}_1(y)$ and $\Upsilon_3(x) \geq \check{\Upsilon}_3(y)$. Indeed, $\Upsilon_1(x) = (D_{\text{tot}} - (x_1 + x_2))(\kappa_{1a} + \kappa_{1b}x_2) \leq (D_{\text{tot}} - (y_1 + y_2))(\kappa_{1a} + \kappa_{1b}y_2) = \check{\Upsilon}_1(y)$ and $\Upsilon_3(x) = x_2(\kappa_{3a} + x_1\kappa_{3b}) \geq y_2(\kappa_{3a} + y_1\kappa_{3b}) = \check{\Upsilon}_3(y)$.

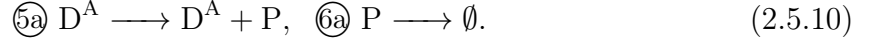
Since all of the hypotheses of Theorem 2.4.2 hold, for each pair $x^\circ, \check{x}^\circ \in \mathcal{X}$ satisfying $x^\circ \preceq_A \check{x}^\circ$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with two Markov chains $X = \{X(t), t \geq 0\}$ and $\check{X} = \{\check{X}(t), t \geq 0\}$ associated with Υ and $\check{\Upsilon}$, respectively, such that $X(0) = x^\circ$, $\check{X}(0) = \check{x}^\circ$ and $\mathbb{P}[X(t) \preceq_A \check{X}(t) \text{ for every } t \geq 0] = 1$.

We can also apply Theorem 2.4.5. The Markov chains X and $\check{X}$ are irreducible and, having only finitely many states, are positive recurrent. Based on the order $\preceq_A$ we introduced, the fully active state $a = (0, D_{\text{tot}})$ is maximal in $\mathcal{X}$ and the fully repressed state $r = (D_{\text{tot}}, 0)$ is minimal in $\mathcal{X}$. Then, by Theorem 2.4.5, we can conclude that $\pi_a \leq \check{\pi}_a$ and $\pi_r \geq \check{\pi}_r$. This implies that increasing μ increases the probability of the system in steady-state to be in the active state a to the detriment of the repressed state r (and vice versa for decreasing μ). We can also apply Theorem 2.4.4. Since $\{a\}$ is increasing and $\{r\}$ is decreasing, then by Theorem 2.4.4, $\check{h}_{r,a} = \mathbb{E}_r[\check{T}_a] \leq \mathbb{E}_r[T_a] = h_{r,a}$ and $h_{a,r} = \mathbb{E}_a[T_r] \leq \mathbb{E}_a[\check{T}_r] = \check{h}_{a,r}$. Since the only difference between the two systems was that $\mu \leq \check{\mu}$, these results imply that the time to memory loss of the active state increases for higher values of μ , while the time to memory loss of the repressed state decreases for higher values of μ .

Example 2.5.5. Epigenetic regulation by chromatin modifications with positive TF-enabled autoregulation

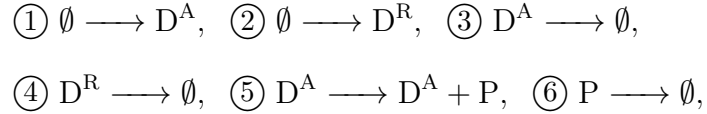
Now, we consider the histone modification circuit considered in the previous example

with an additional positive autoregulation loop. For this, we assume that a protein expressed by the gene of interest recruits writers for the activating histone modifications. Consequently, we introduce the gene product P as an additional species for our system and add the following reactions to the ones shown in Figure 2.5(a):



Furthermore, given the P-enabled autoregulation loop (see Figure 2.6(a)), let us consider the rate constant that goes with $\textcircled{1a}$ in Figure 2.5(a) as $\kappa_{1a} = \kappa_{1a}^0 + \kappa_{1a}^1 g(n_P)$, with κ_{1a}^0 and κ_{1a}^1 representing the rate constants that go with the D^A basal *de-novo* establishment process and with the D^A *de-novo* establishment process enhanced by n_P , respectively, and $g(n_P)$ representing a non-negative, bounded, monotonically increasing function of n_P (see Section 3.4 in Bruno et al. [7]).

Here, we are interested in determining how the reaction rate constant κ_{5a} affects the reactivation time of the gene. As before, we have the conservation law $n_D + n_{D^R} + n_{D^A} = D_{\text{tot}}$, with D_{tot} representing the total number of nucleosomes within the gene, and by fixing $D_{\text{tot}} > 0$, we fix one stoichiometric compatibility class and the projected process $(X_1, X_2, X_3)(\cdot) = (n_{D^R}, n_{D^A}, n_P)$ is a continuous time Markov chain. This is the same as studying the reduced chemical reaction system:



with set of species $\mathcal{S} = \{D^R, D^A, P\}$, set of reactions $\mathcal{R} = \{(v_1^-, v_1^+), (v_2^-, v_2^+), (v_3^-, v_3^+), (v_4^-, v_4^+), (v_5^-, v_5^+), (v_6^-, v_6^+)\}$, where $v_1^- = v_2^- = v_3^+ = v_4^+ = v_6^+ = (0, 0, 0)^T$, $v_2^+ = v_4^- = (1, 0, 0)^T$, $v_1^+ = v_3^- = v_5^- = (0, 1, 0)^T$, $v_5^+ = (0, 1, 1)^T$, $v_6^- = (0, 0, 1)^T$, and with associated

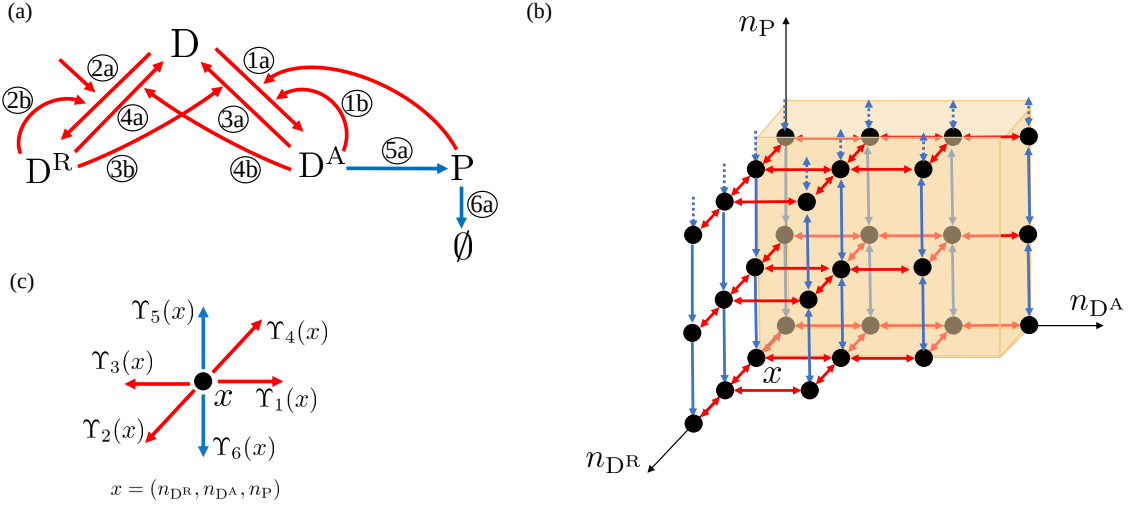


Figure 2.6. Histone modification circuit with positive TF-enabled autoregulation and corresponding Markov chain. (a) Original chemical reaction system. The numbers on the arrows correspond to the associated reactions. (b) Markov chain graph. Here, we consider $D_{\text{tot}} = 3$, we use black dots to represent the states and red double-ended arrows to represent transitions in both directions associated with the reactions represented by the red arrows in (a). Similarly we use blue double-ended (single-ended) arrows to represent transitions in both directions (in one direction) associated with the reactions represented by the blue arrows in (a). We use blue dotted lines to show that, in the vertical direction, the Markov chain has countably infinitely many states, connected by transitions in both directions. Finally, we use orange to highlight the region $K_A + x$ intersected with the state space $\mathcal{X}$, with A defined in (2.5.12). (c) Direction of the possible transitions of the Markov chain starting from a state x , whose rates are defined in equation (2.5.11).

propensity functions of non mass-action type defined as follows:

$$\begin{aligned} \Lambda_{(v_1^-, v_1^+)}(x) &= (D_{\text{tot}} - (x_1 + x_2)) (\kappa_{1a}^0 + \kappa_{1a}^1 g(x_3) + \kappa_{1b} x_2), \\ \Lambda_{(v_2^-, v_2^+)}(x) &= (D_{\text{tot}} - (x_1 + x_2)) (\kappa_{2a} + \kappa_{2b} x_1), \quad \Lambda_{(v_3^-, v_3^+)}(x) = x_2 (\kappa_{3a} + x_1 \kappa_{3b}), \\ \Lambda_{(v_4^-, v_4^+)}(x) &= x_1 \mu (c \kappa_{3a} + x_2 \kappa_{3b}), \quad \Lambda_{(v_5^-, v_5^+)}(x) = \kappa_{5a} x_2, \quad \Lambda_{(v_6^-, v_6^+)}(x) = \kappa_{6a} x_3, \end{aligned}$$

in which κ_{5a} and κ_{6a} are the rate constants that go with reactions $\textcircled{5a}$ and $\textcircled{6a}$ in (2.5.10), respectively, and all the other rate constants are defined as for (2.5.7).

The state space for the Markov chain is $\mathcal{X} = \{(x_1, x_2, x_3) \in \mathbb{Z}_+^3 \mid x_1 + x_2 \leq D_{\text{tot}}\}$. Given a generic state $x = (x_1, x_2, x_3)$, the transitions of the Markov chain are in six possible directions $v_j = v_j^+ - v_j^-$, $j \in \{1, \dots, 6\}$, that can be written as $v_1 = (0, 1, 0)^T$,

$v_2 = (1, 0, 0)^T$, $v_3 = (0, -1, 0)^T$, $v_4 = (-1, 0, 0)^T$, $v_5 = (0, 0, 1)^T$, $v_6 = (0, 0, -1)^T$, with associated infinitesimal transition rates:

$$\begin{aligned}\Upsilon_1(x) &= \Lambda_{(v_1^-, v_1^+)}(x), \quad \Upsilon_2(x) = \Lambda_{(v_2^-, v_2^+)}(x), \quad \Upsilon_3(x) = \Lambda_{(v_3^-, v_3^+)}(x), \\ \Upsilon_4(x) &= \Lambda_{(v_4^-, v_4^+)}(x), \quad \Upsilon_5(x) = \Lambda_{(v_5^-, v_5^+)}(x), \quad \Upsilon_6(x) = \Lambda_{(v_6^-, v_6^+)}(x).\end{aligned}\tag{2.5.11}$$

As mentioned before, we are interested in determining how the protein production rate κ_{5a} affects the reactivation time of the gene, defined as $h_{r,\Theta} = \mathbb{E}_r[T_\Theta]$, where $r = (D_{\text{tot}}, 0, 0)$ and $\Theta = \{w \in \mathcal{X} \mid w = (0, D_{\text{tot}}, i), i \in \mathbb{Z}_+\}$ corresponds to the set of states characterized by the fully active state $n_{\text{DA}} = D_{\text{tot}}$. We first check that the assumptions of Theorem 2.4.2 hold. Let

$$A = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.\tag{2.5.12}$$

For $x \in \mathcal{X}$, $x \preceq_A y$ and the set $K_A + x = \{y \in \mathbb{R}^3 \mid x \preceq_A y\}$. For our example, the region $(K_A + x) \cap \mathcal{X}$ is depicted in orange in Figure 2.6(b). We introduce infinitesimal transition rates $\check{\Upsilon}_1(x), \check{\Upsilon}_2(x), \check{\Upsilon}_3(x), \check{\Upsilon}_4(x), \check{\Upsilon}_5(x)$ and $\check{\Upsilon}_6(x)$ defined as for $\Upsilon_1(x), \Upsilon_2(x), \Upsilon_3(x), \Upsilon_4(x), \Upsilon_5(x)$ and $\Upsilon_6(x)$, with all the parameters having the same values except that κ_{5a} is replaced by $\check{\kappa}_{5a} > \kappa_{5a}$. Condition (i) of Theorem 2.4.2 holds since $Av_1 = (0, 1, 0)^T$, $Av_2 = (-1, 0, 0)^T$, $Av_3 = (0, -1, 0)^T$, $Av_4 = (1, 0, 0)^T$, $Av_5 = (0, 0, 1)^T$, $Av_6 = (0, 0, -1)^T$. Condition (ii) also holds, as shown in the paragraph below.

Verification of condition (ii) of Theorem 2.4.2. First consider $x \in \mathcal{X}$ and $y \in \partial_1(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 = w_1, x_2 \leq w_2, x_3 \leq w_3\}$. Since $\langle A_{1\bullet}, v_4 \rangle = 1$ and $\langle A_{1\bullet}, v_2 \rangle = -1$, we need to check that $\Upsilon_4(x) \leq \check{\Upsilon}_4(y)$ and $\Upsilon_2(x) \geq \check{\Upsilon}_2(y)$. Since $x_1 = y_1, x_2 \leq y_2, x_3 \leq y_3$, we have that $\Upsilon_4(x) = x_1\mu(c\kappa_{3a} + x_2\kappa_{3b}) \leq y_1\mu(c\kappa_{3a} + y_2\kappa_{3b}) = \check{\Upsilon}_4(y)$ and $\Upsilon_2(x) = (D_{\text{tot}} - (x_1 + x_2))(\kappa_{2a} + \kappa_{2b}x_1) \geq (D_{\text{tot}} - (y_1 + y_2))(\kappa_{2a} + \kappa_{2b}y_1) = \check{\Upsilon}_2(y)$. Secondly, consider $x \in \mathcal{X}$ and $y \in \partial_2(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 \geq w_1, x_2 = w_2, x_3 \leq$

$w_3\}$. Since $\langle A_{2\bullet}, v_1 \rangle = 1$ and $\langle A_{2\bullet}, v_3 \rangle = -1$, we need to check that $\Upsilon_1(x) \leq \check{\Upsilon}_1(y)$ and $\Upsilon_3(x) \geq \check{\Upsilon}_3(y)$. Since $x_1 \geq y_1, x_2 = y_2, x_3 \leq y_3$, we have $\Upsilon_1(x) = (D_{\text{tot}} - (x_1 + x_2))(\kappa_{1a}^0 + \kappa_{1a}^1 g(x_3) + \kappa_{1b} x_2) \leq (D_{\text{tot}} - (y_1 + y_2))(\kappa_{1a}^0 + \kappa_{1a}^1 g(y_3) + \kappa_{1b} y_2) = \check{\Upsilon}_1(y)$ and $\Upsilon_3(x) = x_2(\kappa_{3a} + x_1 \kappa_{3b}) \geq y_2(\kappa_{3a} + y_1 \kappa_{3b}) = \check{\Upsilon}_3(y)$. Finally, consider $x \in \mathcal{X}$ and $y \in \partial_3(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 \geq w_1, x_2 \leq w_2, x_3 = w_3\}$. Since $\langle A_{3\bullet}, v_5 \rangle = 1$ and $\langle A_{3\bullet}, v_6 \rangle = -1$, we must check that $\Upsilon_5(x) \leq \check{\Upsilon}_5(y)$ and $\Upsilon_6(x) \geq \check{\Upsilon}_6(y)$. Since $x_1 \geq y_1, x_2 \leq y_2, x_3 = y_3$, we obtain $\Upsilon_5(x) = \kappa_{5a} x_2 \leq \kappa_{5a} y_2 \leq \check{\kappa}_{5a} y_2 = \check{\Upsilon}_5(y)$ and $\Upsilon_6(x) = \kappa_{6a} x_3 = \kappa_{6a} y_3 = \check{\Upsilon}_6(y)$.

Since all the hypotheses of Theorem 2.4.2 hold, for each $x^\circ, \check{x}^\circ \in \mathcal{X}$ satisfying $x^\circ \preccurlyeq_A \check{x}^\circ$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with two Markov chains $X = \{X(t), t \geq 0\}$ and $\check{X} = \{\check{X}(t), t \geq 0\}$ associated with Υ and $\check{\Upsilon}$, respectively, such that $X(0) = x^\circ$, $\check{X}(0) = \check{x}^\circ$ and $\mathbb{P}[X(t) \preccurlyeq_A \check{X}(t) \text{ for every } t \geq 0] = 1$. Furthermore, since the hypotheses of Theorem 2.4.2 hold, we can also apply Theorem 2.4.4. Specifically, for $r = (D_{\text{tot}}, 0, 0)$ and $\Theta = \{y \in \mathcal{X} \mid y = (0, D_{\text{tot}}, i), i \in \mathbb{Z}_+\}$, since Θ is an increasing set in $\mathcal{X}$ with respect to the relation $\preccurlyeq_A$, then $h_{r, \Theta} \geq \check{h}_{r, \Theta}$. This implies that, assuming that the only difference between the two systems is in the value of the protein production rate parameter, κ_{5a} , higher protein production rates reduce the mean reaction time for the gene.

2.6 Proofs of the Main Results

2.6.1 Proof of Theorem 2.4.1.

Consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, a collection of distinct vectors $v_1, \dots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$ and two collections of non-negative functions on $\mathcal{X}$, $\Upsilon = (\Upsilon_1, \dots, \Upsilon_n)$ and $\check{\Upsilon} = (\check{\Upsilon}_1, \dots, \check{\Upsilon}_n)$, such that (2.4.2) holds. Let $Q = (Q_{x,y})_{x,y \in \mathcal{X}}$ and $\check{Q} = (\check{Q}_{x,y})_{x,y \in \mathcal{X}}$ denote the infinitesimal generators for the continuous time Markov chains associated with Υ and $\check{\Upsilon}$, respectively. In the following, let $A \in \mathbb{R}^{m \times d}$ be a matrix with non-zero rows and consider the relation $\preccurlyeq_A$ as defined in Definition 2.4.1.

For the proof of Theorem 2.4.1, we first assume that

$$\sup_{x \in \mathcal{X}} \Upsilon_j(x) < \infty \quad \text{and} \quad \sup_{x \in \mathcal{X}} \check{\Upsilon}_j(x) < \infty \quad \text{for every } 1 \leq j \leq n. \quad (2.6.1)$$

This restriction will be relaxed later. Then, we define a constant $\lambda > 0$ and a pair of functions Φ_λ and $\check{\Phi}_\lambda$, which will be key to our construction of the coupled processes X and $\check{X}$. Let $\lambda > 0$ such that:

$$\lambda > n \max \left\{ \sup_{x \in \mathcal{X}} \sum_{j=1}^n \Upsilon_j(x), \sup_{x \in \mathcal{X}} \sum_{j=1}^n \check{\Upsilon}_j(x) \right\}. \quad (2.6.2)$$

Note that both $\frac{\Upsilon_j(x)}{\lambda}$ and $\frac{\check{\Upsilon}_j(x)}{\lambda}$ are less than $\frac{1}{n}$ for every $x \in \mathcal{X}$ and $1 \leq j \leq n$. For $x \in \mathcal{X}$, consider the sets

$$I_j(x) := \left[\frac{j-1}{n}, \frac{j-1}{n} + \frac{\Upsilon_j(x)}{\lambda} \right), \quad 1 \leq j \leq n. \quad (2.6.3)$$

If $\Upsilon_j(x) = 0$, then $I_j(x)$ is the empty set. On the other hand, if $\Upsilon_j(x) > 0$, then $I_j(x)$ is an interval that is a strict subset of $[\frac{j-1}{n}, \frac{j}{n})$. Define the function $\Phi_\lambda(\cdot, \cdot) : \mathcal{X} \times [0, 1] \rightarrow \mathcal{X}$ by

$$\Phi_\lambda(x, u) := x + \sum_{j=1}^n v_j \mathbf{1}_{I_j(x)}(u), \quad x \in \mathcal{X}, \quad u \in [0, 1]. \quad (2.6.4)$$

For $x \in \mathcal{X}$, the sets $I_1(x), \dots, I_n(x)$ are mutually disjoint and so for any $u \in [0, 1]$ either $\Phi_\lambda(x, u) = x$ or $\Phi_\lambda(x, u) = x + v_j$ for some $1 \leq j \leq n$. In the second case, this will happen if and only if $u \in I_j(x)$ for the corresponding index j . The latter condition implies that $I_j(x) \neq \emptyset$, hence by (2.6.3), $\Upsilon_j(x) > 0$ and by (2.4.2), $x + v_j \in \mathcal{X}$.

This shows that $\Phi_\lambda(\cdot, \cdot)$ is well-defined as an $\mathcal{X}$ -valued function. We define intervals $\check{I}_j(x)$, $1 \leq j \leq n$, $x \in \mathcal{X}$ and a function $\check{\Phi}_\lambda : \mathcal{X} \times [0, 1] \rightarrow \mathcal{X}$ in an analogous manner to that above, where $\check{\Phi}_\lambda$ is defined as in (2.6.4), but with the intervals $I_j(x)$ replaced by $\check{I}_j(x)$, where these are defined as in (2.6.3), but with $\Upsilon_j(x)$ replaced by $\check{\Upsilon}_j(x)$.

Lemma 2.6.1. *Suppose that $x, y \in \mathcal{X}$ are such that $x \preceq_A y$ and the following hold:*

$$\check{\Upsilon}_j(y) \leq \Upsilon_j(x), \quad \text{for each } 1 \leq j \leq n \text{ such that } y + v_j \in \mathcal{X} \setminus (K_A + x), \quad (2.6.5)$$

and

$$\check{\Upsilon}_j(y) \geq \Upsilon_j(x), \quad \text{for each } 1 \leq j \leq n \text{ such that } x + v_j \in \mathcal{X} \text{ and } y \notin K_A + x + v_j. \quad (2.6.6)$$

Then, for each $u \in [0, 1]$,

$$\Phi_\lambda(x, u) \preceq_A \check{\Phi}_\lambda(y, u). \quad (2.6.7)$$

Proof. First, we note that $\Phi_\lambda, \check{\Phi}_\lambda$ have the following property: for every $u \in [0, 1]$ and $1 \leq j \leq n$,

$$\text{if } \Phi_\lambda(x, u) = x + v_j, \text{ then } \check{\Phi}_\lambda(y, u) \in \{y, y + v_j\}, \quad (2.6.8)$$

since $I_j(x), \check{I}_j(y) \subseteq [\frac{j-1}{n}, \frac{j}{n})$. Similarly,

$$\text{if } \check{\Phi}_\lambda(y, u) = y + v_j, \text{ then } \Phi_\lambda(x, u) \in \{x, x + v_j\}. \quad (2.6.9)$$

Furthermore, if $\check{\Upsilon}_j(y) \geq \Upsilon_j(x)$, then

$$\Phi_\lambda(x, u) = x + v_j \text{ implies that } \check{\Phi}_\lambda(y, u) = y + v_j, \quad (2.6.10)$$

since under this condition, $I_j(x) \subseteq \check{I}_j(y)$. Similarly, if $\check{\Upsilon}_j(y) \leq \Upsilon_j(x)$, then

$$\check{\Phi}_\lambda(y, u) = y + v_j \text{ implies that } \Phi_\lambda(x, u) = x + v_j. \quad (2.6.11)$$

Now, to prove (2.6.7), fix $u \in [0, 1]$. We consider two cases.

Case 1: $\check{\Phi}_\lambda(y, u) = y + v_j$ for some $1 \leq j \leq n$.

Fix such an index j . Then, by (2.6.9), either $\Phi_\lambda(x, u) = x + v_j$ or $\Phi_\lambda(x, u) = x$.

- a) If $\Phi_\lambda(x, u) = x + v_j$, then by (2.4.1), $x + v_j \preceq_A y + v_j$ and thus $\Phi_\lambda(x, u) \preceq_A \check{\Phi}_\lambda(y, u)$.
- b) If $\Phi_\lambda(x, u) = x$, then $y + v_j \in K_A + x$. To see this, we note that $y + v_j \in \mathcal{X}$ by (2.4.2) and since $\check{\Upsilon}_j(y) > 0$ because $\check{I}_j(y) \neq \emptyset$. Then, if $y + v_j \notin K_A + x$, by (2.6.5), we would have $\check{\Upsilon}_j(y) \leq \Upsilon_j(x)$, which would imply that $\Phi_\lambda(x, u) = x + v_j$ by (2.6.11). But this contradicts the assumption that $\Phi_\lambda(x, u) = x$. Thus, $y + v_j \in K_A + x$ and so $\Phi_\lambda(x, u) = x \preceq_A y + v_j = \check{\Phi}_\lambda(y, u)$.

Case 2: $\check{\Phi}_\lambda(y, u) = y$. Again, we consider two subcases.

- a) If $\Phi_\lambda(x, u) = x$, then (2.6.7) holds, since $x \preceq_A y$ by assumption.
- b) If $\Phi_\lambda(x, u) = x + v_j$ for some $1 \leq j \leq n$, then $y \in K_A + x + v_j$ for the corresponding value of j . To see this, fix the value of j for which $\Phi_\lambda(x, u) = x + v_j$ and notice that $x + v_j \in \mathcal{X}$ by (2.4.2) and since $\Upsilon_j(x) > 0$. If $y \notin K_A + x + v_j$, then by (2.6.6) we would have $\Upsilon_j(x) \leq \check{\Upsilon}_j(y)$, which would imply that $\check{\Phi}_\lambda(y, u) = y + v_j$. This contradicts the assumption that $\check{\Phi}_\lambda(y, u) = y$. Thus, we must have $y \in K_A + x + v_j$ and then $\Phi_\lambda(x, u) = x + v_j \preceq_A y = \check{\Phi}_\lambda(y, u)$.

□

Now that all these preliminaries have been established under (2.6.1), we proceed with the main part of the proof of Theorem 2.4.1 with this assumption. For this proof, we assume that all of the conditions of Theorem 2.4.1 hold and in addition that condition (2.6.1) holds. The latter ensures that the pair of continuous time Markov chains with infinitesimal generators Q and $\check{Q}$ are *uniformizable* (see Chapter 2 in Keilson [20]). With $\lambda > 0$ as in (2.6.2), the (possibly infinite) matrices⁴ $P_\lambda(Q) := \frac{1}{\lambda}Q + I$ and $P_\lambda(\check{Q}) :=$

⁴These “matrices” may have countably many rows and columns, in which case they could be considered as operators on ℓ^∞ . For convenience, we still call them matrices here.

$\frac{1}{\lambda}\check{Q} + I$ are stochastic⁵, where $I = (I_{x,y})_{x,y \in \mathcal{X}}$ is the identity matrix. Indeed, for $x \in \mathcal{X}$, $(P_\lambda(Q))_{x,x} = \frac{Q_{x,x}}{\lambda} + 1 = 1 - \frac{|Q_{x,x}|}{\lambda} \in [1 - \frac{1}{n}, 1]$, for $y \neq x$, $(P_\lambda(Q))_{x,y} = \frac{Q_{x,y}}{\lambda} \in [0, \frac{1}{n}]$ and $\sum_{y \in \mathcal{X}} (P_\lambda(Q))_{x,y} = \sum_{y \in \mathcal{X}} \frac{1}{\lambda} Q_{x,y} + 1 = 1$.

Now, let $x^\circ, \check{x}^\circ \in \mathcal{X}$ be such that $x^\circ \preccurlyeq_A \check{x}^\circ$. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where the following are defined:

- (i) A Poisson process $N = \{N(t), 0 \leq t < \infty\}$ of rate $\lambda > 0$.
- (ii) A sequence of independent and identically distributed (i.i.d.) random variables $U = (U_k)_{k \geq 1}$ where each U_k has the uniform distribution on $[0, 1]$.

Additionally, choose N to be independent of U . We construct two discrete-time processes, $Y = (Y_k)_{k \geq 0}$ and $\check{Y} = (\check{Y}_k)_{k \geq 0}$, by defining $Y_0 := x^\circ$, $\check{Y}_0 := \check{x}^\circ$, and for $k \geq 0$,

$$Y_{k+1} := \Phi_\lambda(Y_k, U_{k+1}), \quad \check{Y}_{k+1} := \check{\Phi}_\lambda(\check{Y}_k, U_{k+1}). \quad (2.6.12)$$

Then Y and $\check{Y}$ are discrete-time Markov chains with transition matrices $P_\lambda(Q)$ and $P_\lambda(\check{Q})$, respectively. Now, define the processes

$$X(t) := Y_{N(t)}, \quad \check{X}(t) := \check{Y}_{N(t)}, \quad t \geq 0. \quad (2.6.13)$$

According to Section 2.1 in Keilson [20] (see the discussion around Equation 2.1.6 there), X and $\check{X}$ are continuous time Markov chains with infinitesimal generators Q and $\check{Q}$, respectively, and with initial conditions $X(0) = x^\circ$ and $\check{X}(0) = \check{x}^\circ$.

In order to prove (2.4.6), it suffices to check that the following holds:

$$\mathbb{P}[Y_k \preccurlyeq_A \check{Y}_k] = 1, \quad \text{for every } k \geq 0. \quad (2.6.14)$$

Indeed, if this is true, then $\mathbb{P}[Y_k \preccurlyeq_A \check{Y}_k \text{ for every } k \geq 0] = 1$ and therefore $\mathbb{P}[Y_{N(t)} \preccurlyeq_A$

⁵Stochastic here means that all entries take values in $[0, 1]$ and all row sums equal one.

$\check{Y}_{N(t)}$ for every $t \geq 0] = 1$. We will prove (2.6.14) by induction on k . We already know that $x^\circ \preccurlyeq_A \check{x}^\circ$ and so (2.6.14) holds for $k = 0$. Now, assume $\mathbb{P}[Y_k \preccurlyeq_A \check{Y}_k] = 1$ for some $k \geq 0$. Since conditions (2.4.4) and (2.4.5) hold for every $x, y \in \mathcal{X}$ such that $x \preccurlyeq_A y$, by Lemma 2.6.1 we obtain that on a set of probability one, on which $Y_k \preccurlyeq_A \check{Y}_k$,

$$Y_{k+1} = \Phi_\lambda(Y_k, U_{k+1}) \preccurlyeq_A \check{\Phi}_\lambda(\check{Y}_k, U_{k+1}) = \check{Y}_{k+1},$$

and so (2.6.14) holds with $k + 1$ in place of k . This completes the induction step and so Theorem 2.4.1 is proved whenever (2.6.1) holds.

For the case where (2.6.1) does not hold, we construct the corresponding continuous time Markov chains as a limit in distribution of appropriately coupled continuous time Markov chains with truncated propensity functions for which (2.6.1) holds. Many elements for this case are similar to the previous case, although the use of Lemma 2.6.1 is different. We provide the details below, where we assume that the hypotheses of Theorem 2.4.1 hold.

We consider truncations of the propensity functions Υ and $\check{\Upsilon}$. More concretely, for $x^\circ, \check{x}^\circ \in \mathcal{X}$ such that $x^\circ \preccurlyeq_A \check{x}^\circ$, let $M_0 \geq 1$ be an integer such that $\|x^\circ\|_\infty, \|\check{x}^\circ\|_\infty \leq M_0$. For every integer $M \geq M_0$, consider the finite set $\mathcal{X}_M := \{x \in \mathcal{X} \mid \|x\|_\infty \leq M\}$, together with the functions $\Upsilon_j^M, \check{\Upsilon}_j^M : \mathcal{X} \rightarrow \mathbb{R}_+$ defined by $\Upsilon_j^M(x) := \Upsilon_j(x)\mathbb{1}_{\mathcal{X}_M}(x)$ and $\check{\Upsilon}_j^M(x) := \check{\Upsilon}_j(x)\mathbb{1}_{\mathcal{X}_M}(x)$ for $1 \leq j \leq n$ and $x \in \mathcal{X}$. We see that for every $M \geq M_0$, (2.4.2) holds with the functions $\Upsilon^M = (\Upsilon_1^M, \dots, \Upsilon_n^M)$ and $\check{\Upsilon}^M = (\check{\Upsilon}_1^M, \dots, \check{\Upsilon}_n^M)$ in place of Υ and $\check{\Upsilon}$. Also, since $\mathcal{X}_M$ is a finite set, $\sup_{x \in \mathcal{X}} \Upsilon_j^M(x) = \sup_{x \in \mathcal{X}_M} \Upsilon_j(x) < \infty$ and $\sup_{x \in \mathcal{X}} \check{\Upsilon}_j^M(x) = \sup_{x \in \mathcal{X}_M} \check{\Upsilon}_j(x) < \infty$ for every $1 \leq j \leq n$. Furthermore, by (2.4.4) and (2.4.5), we have that for every pair $x, y \in \mathcal{X}_M$ such that $x \preccurlyeq_A y$,

$$\begin{aligned} \check{\Upsilon}_j^M(y) &\leq \Upsilon_j^M(x), \quad \text{for every } 1 \leq j \leq n \text{ such that } y + v_j \in \mathcal{X} \setminus (K_A + x), \text{ and} \\ \check{\Upsilon}_j^M(y) &\geq \Upsilon_j^M(x), \quad \text{for every } 1 \leq j \leq n \text{ such that } x + v_j \in \mathcal{X} \text{ and } y \notin K_A + x + v_j. \end{aligned} \tag{2.6.15}$$

Let Q^M and $\check{Q}^M$ denote the infinitesimal generators associated with Υ^M and $\check{\Upsilon}^M$, respectively. We define an increasing sequence $\{\lambda_M\}_{M \geq M_0}$ of positive numbers such that $\lambda_M \rightarrow \infty$ as $M \rightarrow \infty$ and $\lambda_M > n \max \left\{ \sup_{x \in \mathcal{X}} \sum_{j=1}^n \Upsilon_j^M(x), \sup_{x \in \mathcal{X}} \sum_{j=1}^n \check{\Upsilon}_j^M(x) \right\}$ for every $M \geq M_0$. Define $\Phi_{\lambda_M}(\cdot, \cdot), \check{\Phi}_{\lambda_M}(\cdot, \cdot) : \mathcal{X} \times [0, 1] \rightarrow \mathcal{X}$ as in (2.6.4), but with Υ^M and $\check{\Upsilon}^M$ in place of Υ and $\check{\Upsilon}$, respectively. Since (2.6.15) holds, applying Lemma 2.6.1 with $\Upsilon^M, \check{\Upsilon}^M, \lambda_M, \Phi_{\lambda_M}, \check{\Phi}_{\lambda_M}$ in place of $\Upsilon, \check{\Upsilon}, \lambda, \Phi_\lambda, \check{\Phi}_\lambda$ yields that

$$\Phi_{\lambda_M}(x, u) \preceq_A \check{\Phi}_{\lambda_M}(y, u) \quad \text{for every } x, y \in \mathcal{X}_M \text{ such that } x \preceq_A y \text{ and } u \in [0, 1]. \quad (2.6.16)$$

Now, for each $M \geq M_0$ consider a probability space $(\Omega^M, \mathcal{F}^M, \mathbb{P}^M)$ where the following are defined:

- (i) A Poisson process $N^M = \{N^M(t), 0 \leq t < \infty\}$ of rate $\lambda_M > 0$.
- (ii) An i.i.d. sequence $U^M = (U_k^M)_{k \geq 1}$ of uniform $[0, 1]$ random variables.

Additionally, choose N^M to be independent of U^M . For every $M \geq M_0$, we construct two discrete-time processes, $Y^M = (Y_k^M)_{k \geq 0}$ and $\check{Y}^M = (\check{Y}_k^M)_{k \geq 0}$, by defining $Y_0^M := x^\circ$, $\check{Y}_0^M := \check{x}^\circ$ and for $k \geq 0$,

$$Y_{k+1}^M := \Phi_{\lambda_M}(Y_k^M, U_{k+1}^M), \quad \check{Y}_{k+1}^M := \check{\Phi}_{\lambda_M}(\check{Y}_k^M, U_{k+1}^M).$$

Similarly to the previous case, Y^M and $\check{Y}^M$ are discrete-time Markov chains with transition matrices $P_{\lambda_M}(Q^M) := \frac{1}{\lambda_M} Q^M + I$ and $P_{\lambda_M}(\check{Q}^M) := \frac{1}{\lambda_M} \check{Q}^M + I$, respectively.

Now, we claim that for each $M \geq M_0$:

$$\mathbb{P}^M \left[Y_{k \wedge S^M}^M \preceq_A \check{Y}_{k \wedge S^M}^M \text{ for every } k \geq 0 \right] = 1. \quad (2.6.17)$$

where $S^M := \inf \{k \geq 0 \mid Y_k^M \notin \mathcal{X}_M \text{ or } \check{Y}_k^M \notin \mathcal{X}_M\}$. In fact, (2.6.17) is equivalent to proving that $\mathbb{P}^M \left[Y_{k \wedge S^M}^M \preceq_A \check{Y}_{k \wedge S^M}^M \right] = 1$ for every $k \geq 0$, which we do by induction. We already

know that $Y_0^M \preceq_A \check{Y}_0^M$. Assuming the statement is true for some $k \geq 0$, to establish it for $k+1$ we distinguish between two cases. First, on $\{S^M \leq k\}$, $Y_{(k+1) \wedge S^M}^M = Y_{k \wedge S^M}^M \preceq_A \check{Y}_{k \wedge S^M}^M = \check{Y}_{(k+1) \wedge S^M}^M$, $\mathbb{P}^M$ -a.s.. Second, on $\{S^M > k\}$, $Y_k^M \in \mathcal{X}_M, \check{Y}_k^M \in \mathcal{X}_M$, and by the induction assumption, $Y_k^M \preceq_A \check{Y}_k^M$, $\mathbb{P}^M$ -a.s.. Applying Lemma 2.6.1, we obtain $\mathbb{P}^M$ -a.s. on $\{S^M > k\}$ that

$$Y_{(k+1) \wedge S^M}^M = Y_{k+1}^M = \Phi_{\lambda_M}(Y_k^M, U_{k+1}^M) \preceq_A \check{\Phi}_{\lambda_M}(\check{Y}_k^M, U_{k+1}^M) = \check{Y}_{(k+1) \wedge S^M}^M,$$

where we have used (2.6.16).

Now, for each $M \geq M_0$, we define the processes

$$X^M(t) := Y_{N^M(t)}^M, \quad \check{X}^M(t) := \check{Y}_{N^M(t)}^M, \quad t \geq 0.$$

Then, X^M and $\check{X}^M$ are continuous time Markov chains with infinitesimal generators Q^M and $\check{Q}^M$ respectively, and with initial conditions $X^M(0) = x^\circ$ and $\check{X}^M(0) = \check{x}^\circ$. Define $T^M := \inf\{t \geq 0 \mid X^M(t) \notin \mathcal{X}_M \text{ or } \check{X}^M(t) \notin \mathcal{X}_M\}$ and, because Y^M and $\check{Y}^M$ are the discrete time skeletons for X^M and $\check{X}^M$, we have that $\mathbb{P}^M$ -a.s.

$$T^M = \inf\{t \geq 0 \mid N^M(t) = S^M\}.$$

Then, it follows from (2.6.17) that

$$\mathbb{P}^M \left[X^M(t \wedge T^M) \preceq_A \check{X}^M(t \wedge T^M) \text{ for every } t \geq 0 \right] = 1. \quad (2.6.18)$$

We now prove that for every $t \geq 0$,

$$\mathbb{P}^M[T^M < t] \rightarrow 0 \quad \text{as} \quad M \rightarrow \infty. \quad (2.6.19)$$

For this, let $T_{X^M}^M := \inf\{t \geq 0 \mid X^M(t) \notin \mathcal{X}_M\}$ and $T_{\check{X}^M}^M := \inf\{t \geq 0 \mid \check{X}^M(t) \notin \mathcal{X}_M\}$. Since $T^M = T_{X^M}^M \wedge T_{\check{X}^M}^M$, then

$$\mathbb{P}^M[T^M < t] \leq \mathbb{P}^M[T_{X^M}^M < t] + \mathbb{P}^M[T_{\check{X}^M}^M < t], \quad \text{for every } t \geq 0. \quad (2.6.20)$$

Now, since $Q_{x,y}^M = Q_{x,y}$ for $x \in \mathcal{X}_M$ and $y \in \mathcal{X}$, $X^M(\cdot \wedge T_{X^M}^M)$ will have the same distribution as a Markov chain with infinitesimal generator Q and initial condition x° , stopped at the first time it leaves $\mathcal{X}_M$. Because of this, $T_{X^M}^M$ has the same distribution as the first time a continuous time Markov chain with infinitesimal generator Q leaves $\mathcal{X}_M$. Since a continuous time Markov chain with infinitesimal generator Q has been assumed to not explode in finite time, we obtain that $\mathbb{P}^M[T_{X^M}^M < t] \rightarrow 0$ as $M \rightarrow \infty$. Similar reasoning holds for $T_{\check{X}^M}^M$. Combining with (2.6.20), we obtain (2.6.19).

Denote by $\mathcal{D}([0, \infty), \mathcal{X}^2)$ the space of right-continuous functions from $[0, \infty)$ into $\mathcal{X}^2$ that also have finite left-limits. As usual, this space is endowed with Skorokhod's J_1 topology. The pair $(X^M, \check{X}^M)$ have paths in $\mathcal{D}([0, \infty), \mathcal{X}^2)$ and we obtain $(X, \check{X})$ as a limit in distribution of $(X^M, \check{X}^M)$ as $M \rightarrow \infty$. We first verify that the sequence of processes $\{(X^M, \check{X}^M)\}_{M \geq M_0}$ is tight. For this, it suffices to check that each sequence $\{X^M\}_{M \geq M_0}$ and $\{\check{X}^M\}_{M \geq M_0}$ is tight, which we do by means of Theorem 7.2 in Chapter 3 of Ethier and Kurtz [11]. Condition (a) there (compact containment) is satisfied, because of (2.6.19) and because for $\tilde{M} \geq M \geq M_0$ we have that $X^{\tilde{M}}(\cdot \wedge T_{X^M}^M)$ under $\mathbb{P}^{\tilde{M}}$ has the same law as $X^M(\cdot \wedge T_{X^M}^M)$ under $\mathbb{P}^M$, where $T_{X^M}^{\tilde{M}} := \inf\{t \geq 0 \mid X^{\tilde{M}}(t) \notin \mathcal{X}_M\}$. To verify condition (b) in Theorem 7.2 of Ethier and Kurtz [11], for $t_0 > 0$ fixed and $\eta > 0$, let $M_\eta \geq M_0$ be such that $\mathbb{P}^M[T_{X^M}^M < t_0] \leq \frac{\eta}{2}$ for all $M \geq M_\eta$. Then,

$$\begin{aligned} \mathbb{P}^M[w'(X^M, \delta, t_0) \geq \eta] &\leq \mathbb{P}^M[w'(X^M, \delta, t_0) \geq \eta; T_{X^M}^M \geq t_0] + \mathbb{P}^M[T_{X^M}^M < t_0] \\ &\leq \tilde{\mathbb{P}}[w'(\tilde{X}, \delta, t_0) \geq \eta; \tau_{\tilde{X}}^M \geq t_0] + \frac{\eta}{2} \\ &\leq \tilde{\mathbb{P}}[w'(\tilde{X}, \delta, t_0) \geq \eta] + \frac{\eta}{2}, \end{aligned}$$

where $w'(\cdot, \cdot, \cdot)$ is the modulus of continuity, as defined in Equation (6.2) in Chapter 3 of Ethier and Kurtz [11], $\tilde{X}$ under $\tilde{\mathbb{P}}$ is a realization of the Markov chain associated with the infinitesimal generator Q that starts with x° , and $\tau_X^M := \inf\{t \geq 0 \mid \tilde{X}(t) \notin \mathcal{X}_M\}$. Since $\tilde{X}$ under $\tilde{\mathbb{P}}$ is a single process with right-continuous paths having finite left-limits, the tightness applies to it and so the term $\tilde{\mathbb{P}}[w'(\tilde{X}, \delta, t_0) \geq \eta]$ can be made less than $\frac{\eta}{2}$ by choosing δ sufficiently small and so condition (b) of Theorem 7.2 of Ethier and Kurtz [11] is satisfied. It follows that $\{X^M\}_{M \geq M_0}$ is tight. Similar reasoning yields tightness for $\{\check{X}^M\}_{M \geq M_0}$.

It follows that there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with two processes X and $\check{X}$ defined there, having paths that are right-continuous with finite left-limits, and a subsequence $\{M_k\}_{k \geq 1}$ such that $M_k \rightarrow \infty$ as $k \rightarrow \infty$, and the sequence $\{(X^{M_k}, \check{X}^{M_k})\}_{k \geq 1}$ converges in distribution to the pair of processes $(X, \check{X})$. To identify the law of the limit, note that since $\{Q^{M_k}\}_{k \geq 1}$ converges pointwise to Q , for any function f with bounded support in $\mathcal{X}$, $f(X(t)) - \int_0^t Qf(X(s))ds$ will inherit the martingale property of $f(X^{M_k}(t)) - \int_0^t Q^{M_k}f(X^{M_k}(s))ds$. It follows from the martingale characterization that X is a continuous time Markov chain with infinitesimal generator Q (see Chapter 4 in Ethier and Kurtz [11]). Similarly, $\check{X}$ will be a continuous time Markov chain with infinitesimal generator $\check{Q}$. In addition, the processes have inherited initial conditions $X(0) = x^\circ$ and $\check{X}(0) = \check{x}^\circ$.

Finally, to show that (2.4.6) holds, consider the set

$$F = \{(f, g) \in \mathcal{D}([0, \infty), \mathcal{X}^2) \mid f(t) \preceq_A g(t) \text{ for all } t \geq 0\},$$

which is closed in the Skorokhod topology. From (2.6.18) we know that the stopped processes satisfy $\mathbb{P}^{M_k}[(X^{M_k}(\cdot \wedge T^{M_k}), \check{X}^{M_k}(\cdot \wedge T^{M_k})) \in F] = 1$ for every $k \geq 1$. Furthermore, from (2.6.19) we know that $T^{M_k} \rightarrow \infty$ in probability as $k \rightarrow \infty$. The reader may verify that this last fact, along with the convergence of $(X^{M_k}, \check{X}^{M_k})$ to $(X, \check{X})$, implies that $(X^{M_k}(\cdot \wedge T^{M_k}), \check{X}^{M_k}(\cdot \wedge T^{M_k}))$ converges in distribution to $(X, \check{X})$ as $k \rightarrow \infty$. By the

Portmanteau Theorem (see Theorem 2.1 in Billingsley [5]),

$$1 = \limsup_{k \rightarrow \infty} \mathbb{P}^{M_k}[(X^{M_k}(\cdot \wedge T^{M_k}), \check{X}^{M_k}(\cdot \wedge T^{M_k})) \in F] \leq \mathbb{P}[(X, \check{X}) \in F]$$

and we obtain (2.4.6).

Remark 2.6.1. The proof of Theorem 2.4.1 provides a method to simulate the sample paths for the continuous time Markov chains X and $\check{X}$ in a coupled manner for the case where (2.6.1) holds. Roughly speaking, the procedure consists of determining $\lambda > 0$ as in (2.6.2), $\Phi_\lambda, \check{\Phi}_\lambda$ as in (2.6.4), $Y, \check{Y}$ as in (2.6.12) and $X, \check{X}$ as in (2.6.13). For the benefit of the reader, this method is described as an algorithm in SI-Section 2.8.4, which yields coupled sample paths under the assumptions of Theorems 2.4.2, 2.4.3 and 2.8.2.

2.6.2 Proof of Theorem 2.4.2

By Theorem 2.4.1, it suffices to prove that for every $x, y \in \mathcal{X}$ such that $x \preceq_A y$, conditions (2.4.4) and (2.4.5) hold. For this, we make some observations first. Consider $x, y \in \mathcal{X}$ such that $x \preceq_A y$ and let $1 \leq j \leq n$. Observe that $x \preceq_A y + v_j$ will hold if and only if $A(y + v_j - x) \geq 0$ which is equivalent to:

$$\langle A_{i\bullet}, y - x \rangle + \langle A_{i\bullet}, v_j \rangle \geq 0, \quad \text{for every } 1 \leq i \leq m. \quad (2.6.21)$$

Similarly, $x + v_j \preceq_A y$ will hold if and only if

$$\langle A_{i\bullet}, y - x \rangle - \langle A_{i\bullet}, v_j \rangle \geq 0, \quad \text{for every } 1 \leq i \leq m. \quad (2.6.22)$$

Since $x \preceq_A y$, then $\langle A_{i\bullet}, y - x \rangle \geq 0$ for every $1 \leq i \leq m$. Now, consider $i \in \{1, \dots, m\}$ such that $\langle A_{i\bullet}, y - x \rangle > 0$. Since $A \in \mathbb{Z}^{m \times d}$ and $y - x \in \mathbb{Z}^d$, then $\langle A_{i\bullet}, y - x \rangle \geq 1$. This yields that

$$\langle A_{i\bullet}, y - x \rangle + \langle A_{i\bullet}, v_j \rangle \geq 1 + \langle A_{i\bullet}, v_j \rangle \geq 0, \quad (2.6.23)$$

since $\langle A_{i\bullet}, v_j \rangle \in \{-1, 0, 1\}$. Similarly, $\langle A_{i\bullet}, y - x \rangle - \langle A_{i\bullet}, v_j \rangle \geq 1 - \langle A_{i\bullet}, v_j \rangle \geq 0$. By observing that the interior of $K_A + x$ is of the form $\text{int}(K_A + x) = \{y \in \mathbb{R}^d \mid Ax < Ay\}$, the latter argument shows that for every $x \in \mathcal{X}$ and $y \in \text{int}(K_A + x) \cap \mathcal{X}$, we have

$$x \preceq_A y + v_j \text{ and } x + v_j \preceq_A y, \quad \text{for every } 1 \leq j \leq n. \quad (2.6.24)$$

Now, let us check condition (2.4.4). For this, let $x, y \in \mathcal{X}$ be such that $x \preceq_A y$ and let $1 \leq j \leq n$ be such that $y + v_j \in \mathcal{X} \setminus (K_A + x)$. By (2.6.24), $y \notin \text{int}(K_A + x)$ and since $y \in K_A + x$, we must have $y \in \partial(K_A + x) = \{z \in K_A + x \mid \langle A_{i\bullet}, z \rangle = \langle A_{i\bullet}, x \rangle \text{ for some } 1 \leq i \leq m\}$, the boundary of $K_A + x$. Consider the set of indices $\mathbf{K}_y := \{i \mid \langle A_{i\bullet}, y \rangle = \langle A_{i\bullet}, x \rangle, 1 \leq i \leq m\}$, which is non-empty. Observe that for every $i \notin \mathbf{K}_y$, $\langle A_{i\bullet}, y - x \rangle > 0$ and from (2.6.23), $\langle A_{i\bullet}, (y + v_j) - x \rangle \geq 0$, while for $i \in \mathbf{K}_y$, $\langle A_{i\bullet}, (y + v_j) - x \rangle = \langle A_{i\bullet}, v_j \rangle$. From this, we can infer that there exists an $i_k \in \mathbf{K}_y$ such that $\langle A_{i_k\bullet}, v_j \rangle < 0$. Indeed, if this was not the case, then $\langle A_{i\bullet}, (y + v_j) - x \rangle \geq 0$ for every $i \in \mathbf{K}_y$ and consequently (2.6.21) would hold. This contradicts the fact that $y + v_j \notin K_A + x$. By (2.4.8), we know that $\langle A_{i_k\bullet}, v_j \rangle < 0$ implies $\check{\Upsilon}_j(y) \leq \Upsilon_j(x)$ and we conclude that (2.4.4) holds.

To check condition (2.4.5), let $x, y \in \mathcal{X}$ be such that $x \preceq_A y$ and let $1 \leq j \leq n$ be such that $x + v_j \in \mathcal{X}$ and $y \notin K_A + x + v_j$. Again, by (2.6.24), we obtain that $y \in \partial(K_A + x)$ and $\mathbf{K}_y \neq \emptyset$. For every $i \notin \mathbf{K}_y$, $\langle A_{i\bullet}, y - (x + v_j) \rangle \geq 0$, while for $i \in \mathbf{K}_y$, $\langle A_{i\bullet}, y - (x + v_j) \rangle = -\langle A_{i\bullet}, v_j \rangle$. From this, we can infer that there exists an $i_k \in \mathbf{K}_y$ such that $\langle A_{i_k\bullet}, v_j \rangle > 0$. By (2.4.9), we know that $\langle A_{i_k\bullet}, v_j \rangle > 0$ implies $\check{\Upsilon}_j(y) \geq \Upsilon_j(x)$ and we conclude that (2.4.5) holds.

2.6.3 Proof of Theorem 2.4.3

The proof of this result uses similar general ideas to the ones used in the proof of Theorem 2.4.1. However, since the conditions involve sums, the construction is somewhat

different and more complex and we provide the details below. Let us consider again a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, a collection of distinct vectors $v_1, \dots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$ and two collections of non-negative functions on $\mathcal{X}$, $\Upsilon = (\Upsilon_1, \dots, \Upsilon_n)$ and $\check{\Upsilon} = (\check{\Upsilon}_1, \dots, \check{\Upsilon}_n)$ such that (2.4.2) holds. In the following, let $A \in \mathbb{Z}^{m \times d}$ be a matrix with non-zero rows such that condition (i) of Theorem 2.4.3 holds.

We initially assume that $\sup_{x \in \mathcal{X}} \Upsilon_j(x) < \infty$ and $\sup_{x \in \mathcal{X}} \check{\Upsilon}_j(x) < \infty$ for every $1 \leq j \leq n$, and let $\lambda > 0$ such that (2.6.2) holds. We shall relax these assumptions later. We start by defining functions analogous to Φ_λ and $\check{\Phi}_\lambda$ as defined in (2.6.4), although this time, the construction is more involved.

Recall that s denotes the size of the set $\{Av_j \mid 1 \leq j \leq n\}$ and that the index sets $G^k \neq \emptyset$, $1 \leq k \leq s$, defined in (2.4.11), are such that $Av_j = \eta^k$ for all $j \in G^k$, $1 \leq k \leq s$. Consider a bijection $\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ such that the vectors $v_{\sigma(1)}, \dots, v_{\sigma(n)}$ have the property that the first $|G^1|$ vectors have indices in G^1 , the next $|G^2|$ vectors have indices in G^2 , and so on. More precisely, the bijection σ is such that for $1 \leq k \leq s$, $Av_{\sigma(q)} = \eta^k$, whenever $\sum_{\ell=1}^{k-1} |G^\ell| + 1 \leq q \leq \sum_{\ell=1}^k |G^\ell|$. Recall for this that a sum over an empty set is taken to equal zero.

For $x \in \mathcal{X}$, we define a family of intervals $\{I^k(x) \mid 1 \leq k \leq s\}$ as follows. Let $p_0 := 0$, and for $1 \leq k \leq s$, inductively define $p_k := \sum_{\ell=1}^k |G^\ell|$, and

$$I^k(x) := \bigcup_{q=p_{k-1}+1}^{p_k} I_q^k(x), \quad (2.6.25)$$

where for $p_{k-1} + 1 \leq q \leq p_k$,

$$I_q^k(x) := \left[\frac{p_{k-1}}{n} + \sum_{\ell=p_{k-1}+1}^{q-1} \frac{\Upsilon_{\sigma(\ell)}(x)}{\lambda}, \frac{p_{k-1}}{n} + \sum_{\ell=p_{k-1}+1}^q \frac{\Upsilon_{\sigma(\ell)}(x)}{\lambda} \right). \quad (2.6.26)$$

The sets $I_q^k(x)$, with $1 \leq k \leq s$ and $p_{k-1} + 1 \leq q \leq p_k$, are mutually disjoint, and by (2.6.2), the length of $I^k(x)$ is less than $\frac{p_k - p_{k-1}}{n} = \frac{|G^k|}{n}$, and so the sum of the lengths of

$\{I^k(x) | 1 \leq k \leq s\}$ is less than $\frac{1}{n} \sum_{k=1}^s |G^k| = 1$. Now, let us define $\Psi_\lambda(\cdot, \cdot) : \mathcal{X} \times [0, 1] \rightarrow \mathcal{X}$ by

$$\Psi_\lambda(x, u) := x + \sum_{k=1}^s \sum_{q=p_{k-1}+1}^{p_k} v_{\sigma(q)} \mathbb{1}_{I_q^k(x)}(u), \quad x \in \mathcal{X}, u \in [0, 1]. \quad (2.6.27)$$

Note that $Av_{\sigma(q)} = \eta^k$ for $p_{k-1} + 1 \leq q \leq p_k$, $1 \leq k \leq s$. From the above properties of $I_q^k(x)$, we have that for any $u \in [0, 1]$, either $u \notin \bigcup_{k=1}^s I_q^k(x)$ or $u \in I_q^k(x)$ for exactly one k and q such that $I_q^k(x) \neq \emptyset$. The latter condition implies, by (2.6.26), that $\Upsilon_{\sigma(q)}(x) > 0$ and then, by (2.4.2), $x + v_{\sigma(q)} \in \mathcal{X}$. This shows that $\Psi_\lambda(\cdot, \cdot)$ is well-defined as an $\mathcal{X}$ -valued function.

In an analogous manner to that above, we can define intervals $\check{I}^k(x)$, $\check{I}_q^k(x)$, $1 \leq k \leq s$, $p_{k-1} + 1 \leq q \leq p_k$, $x \in \mathcal{X}$ and a function $\check{\Psi}_\lambda : \mathcal{X} \times [0, 1] \rightarrow \mathcal{X}$, as in (2.6.25)–(2.6.27), but with $\check{\Upsilon}_j(x)$, $\check{I}^k(x)$, $\check{I}_q^k(x)$, $\check{\Psi}_\lambda$ in place of $\Upsilon_j(x)$, $I^k(x)$, $I_q^k(x)$, Ψ_λ .

Lemma 2.6.2. *Suppose that $x, y \in \mathcal{X}$ are such that $x \preceq_A y$ and the following hold: whenever $y \in \partial_i(K_A + x) \cap \mathcal{X}$ for some $1 \leq i \leq m$, we have*

$$\sum_{j \in G^k} \check{\Upsilon}_j(y) \leq \sum_{j \in G^k} \Upsilon_j(x), \quad \text{for every } k \text{ such that } \eta_i^k < 0, \quad (2.6.28)$$

and

$$\sum_{j \in G^k} \check{\Upsilon}_j(y) \geq \sum_{j \in G^k} \Upsilon_j(x), \quad \text{for every } k \text{ such that } \eta_i^k > 0. \quad (2.6.29)$$

Then, for each $u \in [0, 1]$,

$$\Psi_\lambda(x, u) \preceq_A \check{\Psi}_\lambda(y, u). \quad (2.6.30)$$

Proof. First, we note that $\Psi_\lambda, \check{\Psi}_\lambda$ have the following properties: for every $u \in [0, 1]$, $1 \leq k \leq s$, $j \in G^k$,

$$\text{if } \Psi_\lambda(x, u) = x + v_j, \text{ then } \check{\Psi}_\lambda(y, u) \in \{y + v_\ell : \ell \in G^k\} \cup \{y\}, \quad (2.6.31)$$

since $I_{\sigma^{-1}(j)}^k(x), \check{I}_{\sigma^{-1}(\ell)}^k(y) \subseteq [\frac{p_{k-1}}{n}, \frac{p_k}{n})$ for $\ell \in G^k$. Similarly,

$$\text{if } \check{\Psi}_\lambda(y, u) = y + v_j, \text{ then } \Psi_\lambda(x, u) \in \{x + v_\ell : \ell \in G^k\} \cup \{x\}. \quad (2.6.32)$$

Furthermore, for $1 \leq k \leq s$, $j \in G^k$, if $\sum_{\ell \in G^k} \check{\Upsilon}_\ell(y) \geq \sum_{\ell \in G^k} \Upsilon_\ell(x)$, then

$$\Psi_\lambda(x, u) = x + v_j \text{ implies that } \check{\Psi}_\lambda(y, u) = y + v_\ell \text{ for some } \ell \in G^k,$$

since under the condition, $I^k(x) \subseteq \check{I}^k(y)$. Similarly, if $\sum_{\ell \in G^k} \check{\Upsilon}_\ell(y) \leq \sum_{\ell \in G^k} \Upsilon_\ell(x)$, then

$$\check{\Psi}_\lambda(y, u) = y + v_j \text{ implies that } \Psi_\lambda(x, u) = x + v_\ell \text{ for some } \ell \in G^k.$$

We also have that, for $1 \leq k \leq s$ and $j \in G^k$, $x \preccurlyeq_A y + v_j$ if and only if

$$\langle A_{i\bullet}, y - x \rangle + \langle A_{i\bullet}, v_j \rangle \geq 0, \quad \text{for every } 1 \leq i \leq m.$$

Similarly, $x + v_j \preccurlyeq_A y$ if and only if

$$\langle A_{i\bullet}, y - x \rangle - \langle A_{i\bullet}, v_j \rangle \geq 0, \quad \text{for every } 1 \leq i \leq m.$$

Furthermore, for $1 \leq k \leq s$ and $j, \ell \in G^k$, since $Av_j = Av_\ell$ and $x \preccurlyeq_A y$, then

$$\langle A_{i\bullet}, y - x \rangle + \langle A_{i\bullet}, (v_j - v_\ell) \rangle = \langle A_{i\bullet}, y - x \rangle \geq 0, \quad \text{for every } 1 \leq i \leq m. \quad (2.6.33)$$

To prove (2.6.30), we first consider the situation where $y \in \text{int}(K_A + x) = \{w \in \mathbb{R}^d \mid Ax < Aw\}$. Then, for each $1 \leq i \leq m$, $\langle A_{i\bullet}, y - x \rangle > 0$ and since $A \in \mathbb{Z}^{m \times d}$ and $y - x \in \mathbb{Z}^d$, we have $\langle A_{i\bullet}, y - x \rangle \geq 1$. This implies that for $1 \leq k \leq s$ and $j \in G^k$,

$$\langle A_{i\bullet}, y - x \rangle + \langle A_{i\bullet}, v_j \rangle \geq 1 + \langle A_{i\bullet}, v_j \rangle \geq 0, \quad \text{for every } 1 \leq i \leq m, \quad (2.6.34)$$

since $\langle A_{i\bullet}, v_j \rangle \in \{-1, 0, 1\}$ by condition (i) of Theorem 2.4.3. Similarly, for $1 \leq k \leq s$ and $j \in G^k$,

$$\langle A_{i\bullet}, y - x \rangle - \langle A_{i\bullet}, v_j \rangle \geq 1 - \langle A_{i\bullet}, v_j \rangle \geq 0, \quad \text{for every } 1 \leq i \leq m. \quad (2.6.35)$$

It follows from (2.6.33)–(2.6.35) that if $y \in \text{int}(K_A + x) \cap \mathcal{X}$, then for any $1 \leq k \leq s$ and $j, \ell \in G^k$:

$$x \preceq_A y + v_j, \quad x + v_j \preceq_A y \quad \text{and} \quad x + v_\ell \preceq_A y + v_j.$$

We also have, by assumption, that $x \preceq_A y$. Thus, (2.6.30) holds for all $u \in [0, 1]$.

Now, we turn to the other situation where $y \in \partial_i(K_A + x) \cap \mathcal{X}$ for some $1 \leq i \leq m$. Then $\mathbf{K}_y := \{i \mid \langle A_{i\bullet}, y \rangle = \langle A_{i\bullet}, x \rangle, 1 \leq i \leq m\}$ is non-empty. Let $u \in [0, 1]$. We consider two cases.

Case 1: $\check{\Psi}_\lambda(y, u) = y + v_j$ for some $1 \leq j \leq n$.

Fix such an index j . Consider the unique $1 \leq k \leq s$ such that $j \in G^k$. Then, by (2.6.32), either $\Psi_\lambda(x, u) = x + v_\ell$ for some $\ell \in G^k$, or $\Psi_\lambda(x, u) = x$.

a) If $\Psi_\lambda(x, u) = x + v_\ell$ for some $\ell \in G^k$, then, since $x \preceq_A y$ and $Av_j = Av_\ell$, we have $x + v_\ell \preceq_A y + v_j$. Hence, $\Psi_\lambda(x, u) \preceq_A \check{\Psi}_\lambda(y, u)$ and (2.6.30) holds.

b) If $\Psi_\lambda(x, u) = x$, we claim that $y + v_j \in K_A + x$. To see this, observe that for every $i \notin \mathbf{K}_y$, $\langle A_{i\bullet}, y - x \rangle > 0$ and as for (2.6.34), $\langle A_{i\bullet}, (y + v_j) - x \rangle \geq 0$, while for $i \in \mathbf{K}_y$, $\langle A_{i\bullet}, (y + v_j) - x \rangle = \langle A_{i\bullet}, v_j \rangle \in \{-1, 0, 1\}$. For each $i \in \mathbf{K}_y$, if $\langle A_{i\bullet}, v_j \rangle = -1$, then by (2.6.28), we would have $\sum_{\ell \in G^k} \check{\Upsilon}_\ell(y) \leq \sum_{\ell \in G^k} \Upsilon_\ell(x)$, which would imply that $\Psi_\lambda(x, u) = x + v_\ell$ for some $\ell \in G^k$, but this contradicts the assumption that $\Psi_\lambda(x, u) = x$. So we must have $\langle A_{i\bullet}, v_j \rangle \geq 0$ and hence $\langle A_{i\bullet}, (y + v_j) - x \rangle \geq 0$ for all $i \in \mathbf{K}_y$. Thus, $y + v_j \in K_A + x$ and so $\Psi_\lambda(x, u) = x \preceq_A y + v_j = \check{\Psi}_\lambda(y, u)$ holds.

Case 2: $\check{\Psi}_\lambda(y, u) = y$. Again, we consider two subcases.

- a) If $\Psi_\lambda(x, u) = x$, then (2.6.30) holds, because $x \preceq_A y$.
- b) If $\Psi_\lambda(x, u) = x + v_j$ for some $1 \leq j \leq n$, we claim that $y \in K_A + x + v_j$ for the corresponding value of j . To see this, fix the value of j for which $\Psi_\lambda(x, u) = x + v_j$, let $1 \leq k \leq s$ be such that $j \in G^k$, and observe that for every $i \notin \mathbf{K}_y$, $\langle A_{i\bullet}, y - x \rangle > 0$ and as for (2.6.35), $\langle A_{i\bullet}, y - (x + v_j) \rangle \geq 0$, while for $i \in \mathbf{K}_y$, $\langle A_{i\bullet}, y - (x + v_j) \rangle = -\langle A_{i\bullet}, v_j \rangle \in \{-1, 0, 1\}$. For each $i \in \mathbf{K}_y$, if $\langle A_{i\bullet}, v_j \rangle = 1$, then by (2.6.29), we would have $\sum_{\ell \in G^k} \check{\Upsilon}_\ell(y) \geq \sum_{\ell \in G^k} \Upsilon_\ell(x)$, which would imply that $\check{\Psi}_\lambda(y, u) = y + v_\ell$ for some $\ell \in G^k$. This would contradict the assumption that $\check{\Psi}_\lambda(y, u) = y$. So we must have $\langle A_{i\bullet}, v_j \rangle \leq 0$ and hence $\langle A_{i\bullet}, y - (x + v_j) \rangle = \langle A_{i\bullet}, y - x \rangle - \langle A_{i\bullet}, v_j \rangle \geq 0$ for all $i \in \mathbf{K}_y$. Thus, we have $y \in K_A + x + v_j$ and then $\Psi_\lambda(x, u) = x + v_j \preceq_A y = \check{\Psi}_\lambda(y, u)$.

□

In order to prove Theorem 2.4.3, from here on we can follow a similar procedure to the one used in the proof of Theorem 2.4.1 after Lemma 2.6.1 was proved there. For the case where (2.6.1) holds, we define two discrete-time processes, $Y = (Y_k)_{k \geq 0}$ and $\check{Y} = (\check{Y}_k)_{k \geq 0}$, by defining $Y_0 := x^\circ$, $\check{Y}_0 := \check{x}^\circ$, and for $k \geq 0$,

$$Y_{k+1} := \Psi_\lambda(Y_k, U_{k+1}), \quad \check{Y}_{k+1} := \check{\Psi}_\lambda(\check{Y}_k, U_{k+1}),$$

and define X and $\check{X}$ using these and an independent Poisson process N as in (2.6.13). For the case where (2.6.1) does not hold, we can use a truncation procedure similar to that for Theorem 2.4.1. In both cases, we use Lemma 2.6.2 instead of Lemma 2.6.1.

2.7 Conclusion

In this work, we first reviewed the concept of Stochastic Chemical Reaction Networks (SCRNs), a class of continuous time Markov chain models frequently used to describe the stochastic behavior of chemical reaction systems. We also gave the definitions of

preorder and increasing set considered in this work. In Section 2.4.2, we presented the main theoretical results of this work. We first derived, by exploiting uniformization and then coupling of stochastic processes (see Grassmann [13] and Keilson [20]), three theorems which give practical sufficient conditions for stochastic dominance of one continuous time Markov chain over another. More precisely, these theorems provide conditions under which, when one or more parameters is changed monotonically, the system is almost surely “higher” with respect to a certain preorder. While the first theorem (Theorem 2.4.1) can be used for any SCRNs, it has extensive conditions to check. The second set of theorems (Theorems 2.4.2 and 2.4.3) can be used for more specific SCRNs classes, but they have assumptions that only need to be checked at the boundary of certain translated convex cones. All these theorems can be applied to SCRNs with either finite or countably many states. In Section 2.4.3, we exploited these tools to develop two theorems to specifically study the monotonicity properties of stationary distributions and mean first passage times depending on system parameters.

Subsequently, in Section 2.5, we presented some illustrative examples to highlight the advantages of using our theoretical tools in order to study the stochastic behavior of SCRNs. Specifically, we focused on two common models for enzymatic kinetics (see for example Michaelis and Menten [23], Anderson et al. [1], Del Vecchio and Murray [9] and Kang et al. [19]), on a model inspired by Braess’s paradox (see Calvert et al. [8]) and on a recently developed model describing the main interactions among histone modifications alone, and together with an expressed protein (see Bruno et al. [7]). In these illustrative examples we see that our sufficient conditions can be easy to check and our results can be also used to study networks with a countably infinite number of states. Furthermore, the conclusions obtained by using our theorems are true for trajectories of the Markov chains, yielding results for both transient and steady state behavior.

Overall, in this work we derived and presented theorems that can be used for the theoretical study of monotonicity of SCRNs associated to a variety of chemical reaction

systems. Future work will include the adaptation of our theoretical tools to other forms of monotonicity for SCRN (see Definition 5.1.1 in Müller and Stoyan [24] as an example), the investigation of possible correlations between the network graph properties and the monotonicity properties of the SCRN (extension of the work of Angeli et al. [3] to SCRN), and the application of our results to deterministic chemical reaction network through appropriate limits.

2.8 Supplementary Information (SI)

2.8.1 Criteria for positive recurrence and exponential ergodicity with application to Examples 2.5.2 and 2.5.5

Here we recall fairly general conditions for positive recurrence and exponential ergodicity of a continuous time Markov chain. Such conditions are well known and are usually referred to as Foster-Lyapunov-type conditions. We also apply these to the Markov chains in Examples 2.5.2 and 2.5.5 later in this section.

Theorem 2.8.1. *Let X be an irreducible continuous time Markov chain with state space $\mathcal{X}$ and infinitesimal generator Q . Suppose $V : \mathcal{X} \rightarrow \mathbb{R}_+$ is norm-like, that is $\{x \in \mathcal{X} : V(x) \leq a\}$ is compact ⁶ for each $a \in \mathbb{R}_+$. Further assume that for some $c > 0$, $d > 0$ and a compact set C ,*

$$QV(x) \leq -c + d\mathbf{1}_C(x), \quad \text{for all } x \in \mathcal{X}. \quad (2.8.1)$$

Then, X is non-explosive and positive recurrent, and has a unique stationary distribution π . If instead of (2.8.1), we have that for some $c' > 0$ and $d' > 0$,

$$QV(x) \leq -c'V(x) + d', \quad \text{for all } x \in \mathcal{X}, \quad (2.8.2)$$

⁶Since X is a Markov chain, $\mathcal{X}$ is finite or countable, and we endow it with the usual discrete topology consisting of all subsets of $\mathcal{X}$.

then (2.8.1) automatically holds and the consequences stated above hold, and in addition, the stationary distribution satisfies

$$\pi(V) = \sum_{x \in \mathcal{X}} \pi_x V(x) < \infty,$$

and there is $0 < B < \infty$ and $0 < \beta < 1$ such that for all $t \geq 0$ and $x \in \mathcal{X}$,

$$\sum_{y \in \mathcal{X}} |P_{xy}(t) - \pi_y| \leq \|P_{x\bullet}(t) - \pi\|_{V+1} \leq B(V(x) + 1)\beta^t, \quad (2.8.3)$$

where $P_{xy}(t) = \mathbb{P}[X(t) = y | X(0) = x]$ and

$$\|P_{x\bullet}(t) - \pi\|_{V+1} = \sup_{|g| \leq V+1} \left| \sum_{y \in \mathcal{X}} (P_{xy}(t) - \pi_y) g(y) \right|.$$

Remark 2.8.1. When the second inequality in (2.8.3) holds, we say that the Markov chain is exponentially ergodic in the $(V + 1)$ -norm.

Proof. We will verify the sufficient conditions for each of non-explosion, positive recurrence and exponential ergodicity given in Meyn and Tweedie [22]. Note that X is a Borel right-process under the definition in Sharpe [25]. In addition, since the state space is discrete, each compact set is finite and therefore petite.

For $m \in \mathbb{Z}_+$, if Q_m is the infinitesimal generator for the Markov chain X killed upon exit from $O_m = \{x \in \mathcal{X} : V(x) \leq m\}$ ⁷, then $Q_m V(x) \leq QV(x)$ for $x \in O_m$. It then follows from (2.8.1) that conditions (CD0) and (CD2) (with $f = 1$) in Meyn and Tweedie [22] hold with Q_m in place of $\mathcal{A}_m$ there. By Theorem 2.1 and Theorem 4.2 in Meyn and Tweedie [22], the Markov chain is non-explosive and positive recurrent, and it has a unique stationary distribution π .

On the other hand, if (2.8.2) holds, then (2.8.1) holds, using the norm like property

⁷Upon exit from O_m , the killed process goes to a cemetery state Δ_m in $\mathcal{X} \setminus O_m$ where $V(\Delta_m) = \min\{V(x) : x \in \mathcal{X} \setminus O_m\}$.

of V . Furthermore, (2.8.2) implies that conditions (CD0), (CD2) (with $f = V + 1$) and (CD3) in Meyn and Tweedie [22] hold with Q_m in place of $\mathcal{A}_m$ there. By Theorem 2.1, Theorem 4.2 and Theorem 6.1 in Meyn and Tweedie [22], the Markov chain is non-explosive and positive recurrent, with a unique stationary distribution π such that $\pi(V) < \infty$, and it is exponentially ergodic in the $(V + 1)$ -norm, that is the second inequality in (2.8.3) holds for all $t \geq 0$ and $x \in \mathcal{X}$. For fixed $t \geq 0$ and $x \in \mathcal{X}$, setting $g(y) = \text{sgn}(P_{xy}(t) - \pi_y)$, for $y \in \mathcal{X}$, we have that $|g| \leq 1 \leq V + 1$, and

$$\sum_{y \in \mathcal{X}} |P_{xy}(t) - \pi_y| = \left| \sum_{y \in \mathcal{X}} (P_{xy}(t) - \pi_y) g(y) \right| \leq \|P_{x\bullet}(t) - \pi\|_{V+1},$$

yielding the first inequality in (2.8.3). $\square$

Application to Example 2.5.2

For Example 2.5.2, we first show that the Markov chain is irreducible. For this, consider $x^\circ = (0, 0, E_{\text{tot}}, 0)$ and any fixed state $x = (x_1, x_2, x_3, E_{\text{tot}} - x_3) : 0 \leq x_3 \leq E_{\text{tot}}$. Starting at x° , by having reaction ⑤ fire $x_1 + E_{\text{tot}} - x_3 + x_2$ times in succession, then having reaction ①, immediately followed by reaction ③, fire x_2 times in succession and then reaction ① fire $E_{\text{tot}} - x_3$ times, without any other reactions firing, we see that the Markov chain can transition with positive probability from x° to x . Since each reaction is reversible, it also follows that the Markov chain can transition from x to x° with positive probability. Thus, the Markov chain is irreducible.

Next we will introduce a norm-like function V and show that (2.8.1) holds. For each $x \in \mathcal{X}$, let

$$V(x) = x_1^2 + ((2E_{\text{tot}} - 1)b + 1)x_2 + bx_4^2,$$

where

$$b = \frac{1 + (\kappa_5 + \kappa_2 E_{\text{tot}} + \kappa_3 E_{\text{tot}}) + \frac{(2\kappa_5 + \kappa_6 + 2\kappa_2 E_{\text{tot}})^2}{8\kappa_6}}{\kappa_2 E_{\text{tot}}(2E_{\text{tot}} - 1)}. \quad (2.8.4)$$

Notice that $b > 0$ since $E_{\text{tot}} \geq 1$. Then, for each $a \in \mathbb{R}_+$, $\{x \in \mathcal{X} : V(x) \leq a\}$ consists of finitely many states, and for any $x \in \mathcal{X}$,

$$\begin{aligned}
QV(x) &= \sum_{j=1}^6 \Upsilon_j(x) \cdot (V(x + v_j) - V(x)) \\
&= \kappa_1 x_1 x_3 \cdot (((x_1 - 1)^2 + b(x_4 + 1)^2) - (x_1^2 + bx_4^2)) \\
&\quad + \kappa_2 x_4 \cdot (((x_1 + 1)^2 + b(x_4 - 1)^2) - (x_1^2 + bx_4^2)) \\
&\quad + \kappa_3 x_4 \cdot (((2E_{\text{tot}} - 1)b + 1)(x_2 + 1) + b(x_4 - 1)^2) \\
&\quad \quad \quad - (((2E_{\text{tot}} - 1)b + 1)x_2 + bx_4^2)) \\
&\quad + \kappa_4 x_2 x_3 \cdot (((2E_{\text{tot}} - 1)b + 1)(x_2 - 1) + b(x_4 + 1)^2) \\
&\quad \quad \quad - (((2E_{\text{tot}} - 1)b + 1)x_2 + bx_4^2)) \\
&\quad + \kappa_5 \cdot ((x_1 + 1)^2 - x_1^2) + \kappa_6 x_1 \cdot ((x_1 - 1)^2 - x_1^2) \\
&= \kappa_1 x_1 x_3 \cdot (-2x_1 + 1 + b(2x_4 + 1)) + \kappa_2 x_4 \cdot (2x_1 + 1 + b(-2x_4 + 1)) \\
&\quad + \kappa_3 x_4 \cdot (((2E_{\text{tot}} - 1)b + 1) + b(-2x_4 + 1)) \\
&\quad + \kappa_4 x_2 x_3 \cdot (-((2E_{\text{tot}} - 1)b + 1) + b(2x_4 + 1)) \\
&\quad + \kappa_5 \cdot (2x_1 + 1) + \kappa_6 x_1 \cdot (-2x_1 + 1) \\
&= -(2\kappa_1 x_3 + 2\kappa_6) \cdot x_1^2 + (2\kappa_5 + \kappa_6 + \kappa_1(1 + b)x_3 + 2b\kappa_1 x_3 x_4 + 2\kappa_2 x_4) \cdot x_1 \\
&\quad + \kappa_4 (-2bE_{\text{tot}} x_3 + 2bx_3 x_4 + (2b - 1)x_3) \cdot x_2 \\
&\quad + \kappa_5 + (\kappa_2(1 + b) + \kappa_3(2bE_{\text{tot}} + 1))x_4 - 2b(\kappa_2 + \kappa_3)x_4^2 \\
&= -(2\kappa_1 x_3 + 2\kappa_6) \cdot x_1^2 + (2\kappa_5 + \kappa_6 + \kappa_1(1 + b)x_3 + 2b\kappa_1 x_3 x_4 + 2\kappa_2 x_4) \cdot x_1 \\
&\quad + \kappa_4 (-2bx_3^2 + (2b - 1)x_3) \cdot x_2 \\
&\quad + \kappa_5 + (\kappa_2(1 + b) + \kappa_3(2bE_{\text{tot}} + 1))x_4 - 2b(\kappa_2 + \kappa_3)x_4^2.
\end{aligned}$$

The last equality uses the fact that $x_3 = E_{\text{tot}} - x_4$. For the following, we note that $-2bx_3^2 + (2b - 1)x_3 = (-2bx_3 + (2b - 1))x_3 \leq -1$ when $x_3 \geq 1$.

We now consider two cases for $QV(x)$: when $x_3 = 0$ and $x_3 > 0$. For the first case,

when $x_3 = 0$, we have $x_4 = E_{\text{tot}}$ and

$$QV(x) = -2\kappa_6 \cdot x_1^2 + (2\kappa_5 + \kappa_6 + 2\kappa_2 E_{\text{tot}}) \cdot x_1 + \kappa_5 + (\kappa_2(1+b) + \kappa_3)E_{\text{tot}} - 2b\kappa_2 E_{\text{tot}}^2.$$

As a quadratic function, the last expression is bounded above by

$$b\kappa_2 E_{\text{tot}}(1 - 2E_{\text{tot}}) + (\kappa_5 + \kappa_2 E_{\text{tot}} + \kappa_3 E_{\text{tot}}) + \frac{(2\kappa_5 + \kappa_6 + 2\kappa_2 E_{\text{tot}})^2}{8\kappa_6} = -1,$$

since b is chosen as in (2.8.4). For the second case when $x_3 \in \{1, \dots, E_{\text{tot}}\}$, we have

$$\begin{aligned} QV(x) &\leq -2\kappa_6 \cdot x_1^2 + (2\kappa_5 + \kappa_6 + \kappa_1(1+b)x_3 + 2b\kappa_1 x_3 x_4 + 2\kappa_2 x_4) \cdot x_1 \\ &\quad - \kappa_4 \cdot x_2 + \kappa_5 + (\kappa_2(1+b) + \kappa_3(2bE_{\text{tot}} + 1))x_4 - 2b(\kappa_2 + \kappa_3)x_4^2 \\ &\leq -2\kappa_6 \cdot x_1^2 + \kappa_6 \cdot x_1^2 + \left(\frac{2\kappa_5 + \kappa_6 + \kappa_1(1+b)x_3 + 2b\kappa_1 x_3 x_4 + 2\kappa_2 x_4}{2\sqrt{\kappa_6}} \right)^2 \\ &\quad - \kappa_4 \cdot x_2 + \kappa_5 + (\kappa_2(1+b) + \kappa_3(2bE_{\text{tot}} + 1))x_4 - 2b(\kappa_2 + \kappa_3)x_4^2 \\ &\leq -\kappa_6 \cdot x_1^2 - \kappa_4 \cdot x_2 + \frac{(2\kappa_5 + \kappa_6 + \kappa_1(1+b)E_{\text{tot}} + 2b\kappa_1 E_{\text{tot}}^2 + 2\kappa_2 E_{\text{tot}})^2}{4\kappa_6} \\ &\quad + \kappa_5 + (\kappa_2(1+b) + \kappa_3(2bE_{\text{tot}} + 1))E_{\text{tot}}, \end{aligned}$$

where we have used the fact that $2ab \leq a^2 + b^2$ for the second inequality and the last expression will be less than or equal to -1 whenever x_1 or x_2 is sufficiently large.

Let $C = \{x \in \mathcal{X} : QV(x) > -1\}$. Then, by the above, C consists of finitely many points, which implies that C is a compact set. Then, (2.8.1) holds with $c = 1$ and $d = 1 + \max_{x \in C} QV(x) \vee 0$. It follows by SI-Theorem 2.8.1 that the Markov chain in Example 2.5.2 is non-explosive and positive recurrent, and has a unique stationary distribution π .

Application to Example 2.5.5

For Example 2.5.5, we first check that the Markov chain is irreducible. For this, consider $x^\circ = (0, D_{\text{tot}}, 0)$ and any fixed state $x = (x_1, x_2, x_3) : 0 \leq x_1 + x_2 \leq D_{\text{tot}}$. Starting

at x° , by having reaction ⑤ fire x_3 times, then reaction ③ fire $D_{\text{tot}} - x_2$ times, and finally, reaction ② fire x_1 times, without any other reactions firing, we see that the Markov chain can transition from x° to x with positive probability. For the reverse transition, by having reaction ⑥ fire x_3 times, then reaction ④ fire x_1 times, and finally, reaction ① fire $D_{\text{tot}} - x_2$ times, we see that the Markov chain can transition from x to x° with positive probability. Thus, the Markov chain is irreducible.

Next we will introduce a norm-like function V and show that (2.8.2) holds. For each $x \in \mathcal{X}$, let

$$V(x) = x_3.$$

Then $\{x \in \mathcal{X} : V(x) \leq a\} = \{x \in \mathbb{Z}_+^3 : x_1 + x_2 \leq D_{\text{tot}}, x_3 \leq a\}$ consists of finitely many states for each $a \in \mathbb{R}_+$, and for any $x \in \mathcal{X}$,

$$\begin{aligned} QV(x) &= \sum_{j=1}^6 \Upsilon_j(x) \cdot (V(x + v_j) - V(x)) = \sum_{j=1}^6 \Upsilon_j(x) \cdot V(v_j) \\ &= \Upsilon_1(x) \cdot 0 + \Upsilon_2(x) \cdot 0 + \Upsilon_3(x) \cdot 0 + \Upsilon_4(x) \cdot 0 + \kappa_{5a}x_2 \cdot 1 + \kappa_{6a}x_3 \cdot (-1) \\ &\leq -\kappa_{6a}x_3 + \kappa_{5a}D_{\text{tot}} = -c'V(x) + d', \end{aligned}$$

where $c' = \kappa_{6a}$ and $d' = \kappa_{5a}$. Therefore, we conclude by SI-Theorem 2.8.1 that the Markov chain in Example 2.5.5 is non-explosive and positive recurrent with a unique stationary distribution π such that $\pi(V) < \infty$, and it is exponentially ergodic in the $(V + 1)$ -norm.

2.8.2 Derivation of Markov chain transition directions, v_j

Example 2.5.2: The set of reactions associated to the chemical reaction system in Figure 2.3(a) is given by $\mathcal{R} = \{(v_1^-, v_1^+), (v_2^-, v_2^+), (v_3^-, v_3^+), (v_4^-, v_4^+), (v_5^-, v_5^+), (v_6^-, v_6^+)\}$, where (v_1^-, v_1^+) , (v_2^-, v_2^+) , (v_3^-, v_3^+) are defined as in Example 2.5.1, and $v_4^- = (0, 1, 1, 0)^T$, $v_4^+ = (0, 0, 0, 1)^T$, $v_5^- = v_6^+ = (0, 0, 0, 0)^T$, $v_5^+ = v_6^- = (1, 0, 0, 0)^T$. Then, the potential transitions of the Markov chain are in six possible directions, $v_j = v_j^+ - v_j^-$ for $j = 1, \dots, 6$,

where $v_1 = -v_2 = (-1, 0, -1, 1)^T$, $v_3 = -v_4 = (0, 1, 1, -1)^T$, and $v_5 = -v_6 = (1, 0, 0, 0)^T$.

Example 2.5.3: The set of five reactions associated to the chemical reaction system in Figure 2.4(a) is given by $\mathcal{R} = \{(v_1^-, v_1^+), (v_2^-, v_2^+), (v_3^-, v_3^+), (v_4^-, v_4^+), (v_5^-, v_5^+)\}$, where $v_1^- = v_3^- = (1, 0, 0, 0)^T$, $v_1^+ = v_2^- = v_5^- = (0, 1, 0, 0)^T$, $v_3^+ = v_4^- = v_5^+ = (0, 0, 1, 0)^T$, $v_2^+ = v_4^+ = (0, 0, 0, 1)^T$. Then, the potential transitions of the Markov chain are in five possible directions, $v_j = v_j^+ - v_j^-$, $j = 1, \dots, 5$, where $v_1 = (-1, 1, 0, 0)^T$, $v_2 = (0, -1, 0, 1)^T$, $v_3 = (-1, 0, 1, 0)^T$, $v_4 = (0, 0, -1, 1)^T$ and $v_5 = (0, -1, 1, 0)^T$.

2.8.3 A generalization of Theorem 2.4.3

Here, we provide a more general version of Theorem 2.4.3. The simpler form given as Theorem 2.4.3 in the main text was used there because it is more straightforward to state and the conditions are easier to verify. However, the more general version of the theorem provided in this section can be useful in some cases, such as Example 2.5.3, which we will discuss at the end of this section.

The generalization relies on the idea that grouping of vectors can be more general than what is described in (2.4.11), and so we introduce the following assumption.

Assumption 2.8.1. Consider a collection of distinct vectors $v_1, \dots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$ and a matrix $A \in \mathbb{Z}^{m \times d}$ with non-zero rows. Suppose that there exists a partition⁸ $\{G^1, \dots, G^s\}$ of $\{1, \dots, n\}$ and an associated bijection $\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ such that for $p_0 := 0$ and $1 \leq k \leq s$, with $p_k := \sum_{\ell=1}^k |G^\ell|$, we have $\sigma(q) \in G^k$ for $p_{k-1} + 1 \leq q \leq p_k$, and whenever $p_{k-1} + 2 \leq q \leq p_k$, we also have for each $1 \leq i \leq m$ that either $\langle A_{i\bullet}, v_{\sigma(q)} \rangle$ is equal to $\langle A_{i\bullet}, v_{\sigma(q-1)} \rangle$ or it is 0.

Theorem 2.8.2. Consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, suppose that SI-Assumption 2.8.1 holds and consider two collections of non-negative functions on $\mathcal{X}$, $\Upsilon = (\Upsilon_1, \dots, \Upsilon_n)$ and $\check{\Upsilon} = (\check{\Upsilon}_1, \dots, \check{\Upsilon}_n)$, such that (2.4.2) holds and the associated continuous time Markov

⁸In particular, $\{G^1, \dots, G^s\}$ is a finite collection of non-empty disjoint sets of distinct numbers, the union of which is $\{1, \dots, n\}$.

chains do not explode in finite time. Further suppose that both of the following conditions hold:

(i) For each $1 \leq j \leq n$, the vector Av_j has entries in $\{-1, 0, 1\}$ only.

(ii) For each $x \in \mathcal{X}$, $1 \leq i \leq m$ and $y \in \partial_i(K_A + x) \cap \mathcal{X}$ we have that for each $1 \leq k \leq s$,

$$\sum_{j \in G_i^{k,-}} \check{\Upsilon}_j(y) \leq \sum_{j \in G_i^{k,-}} \Upsilon_j(x), \quad \text{where } G_i^{k,-} = \{j \in G^k \mid \langle A_{i\bullet}, v_j \rangle = -1\},$$

and

$$\sum_{j \in G_i^{k,+}} \check{\Upsilon}_j(y) \geq \sum_{j \in G_i^{k,+}} \Upsilon_j(x), \quad \text{where } G_i^{k,+} = \{j \in G^k \mid \langle A_{i\bullet}, v_j \rangle = 1\}.$$

Then, for each pair $x^\circ, \check{x}^\circ \in \mathcal{X}$ such that $x^\circ \preceq_A \check{x}^\circ$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with two continuous time Markov chains $X = \{X(t), t \geq 0\}$ and $\check{X} = \{\check{X}(t), t \geq 0\}$ defined there, each having state space $\mathcal{X} \subseteq \mathbb{Z}_+^d$, with infinitesimal generators Q and $\check{Q}$, associated with Υ and $\check{\Upsilon}$ respectively, with initial conditions $X(0) = x^\circ$ and $\check{X}(0) = \check{x}^\circ$ and such that:

$$\mathbb{P} \left[X(t) \preceq_A \check{X}(t) \text{ for every } t \geq 0 \right] = 1.$$

Remark 2.8.2. Under SI-Assumption 2.8.1, for a given $1 \leq k \leq s$ and $1 \leq i \leq m$, at most one of $G_i^{k,-}$ and $G_i^{k,+}$ is non-empty. If $G_i^{k,-} \neq \emptyset$, then $G_i^{k,-} = \{\sigma(q) \mid p_{k-1}+1 \leq q \leq q^*\}$ where $q^* = \max\{q \mid p_{k-1}+1 \leq q \leq p_k \text{ and } \langle A_{i\bullet}, v_{\sigma(q)} \rangle = -1\}$. On the other hand, if $G_i^{k,+} \neq \emptyset$, then $G_i^{k,+} = \{\sigma(q) \mid p_{k-1}+1 \leq q \leq q^*\}$ where $q^* = \max\{q \mid p_{k-1}+1 \leq q \leq p_k \text{ and } \langle A_{i\bullet}, v_{\sigma(q)} \rangle = 1\}$. Furthermore, in either case, for $q^* < q \leq p_k$, we have $\langle A_{i\bullet}, v_{\sigma(q)} \rangle = 0$.

The proof for SI-Theorem 2.8.2 can be found below. With SI-Theorem 2.8.2 in place, we can extend Theorems 2.4.4 and 2.4.5 by adding an additional alternative condition (iv): SI-Assumption 2.8.1 holds, and conditions (i) and (ii) in SI-Theorem 2.8.2 are satisfied. If

this is satisfied instead of one of (i)–(iii) in Theorems 2.4.4 and 2.4.5, then the conclusions of these theorems about (mean) first passage times and stationary distributions will still hold.

Proof of SI-Theorem 2.8.2

We initially assume that $\sup_{x \in \mathcal{X}} \Upsilon_j(x) < \infty$ and $\sup_{x \in \mathcal{X}} \check{\Upsilon}_j(x) < \infty$ for every $1 \leq j \leq n$, and let $\lambda > 0$ such that (2.6.2) holds. We shall relax these assumptions later. Further suppose that SI-Assumption 2.8.1 and condition (i) of SI-Theorem 2.8.2 both hold. For $x \in \mathcal{X}$, define $I^k(x)$, $I_q^k(x)$, Ψ_λ , $\check{I}^k(x)$, $\check{I}_q^k(x)$ and $\check{\Psi}_\lambda$ in the same manner as in the proof of Theorem 2.4.3 (see (2.6.25)–(2.6.27)), with $\{G^k \mid 1 \leq k \leq s\}$, $\{p_k \mid 0 \leq k \leq s\}$ and σ as in SI-Assumption 2.8.1. Our proof of SI-Theorem 2.8.2 has some elements that are the same as those for the proof of Theorem 2.4.3. However, some additional elements are needed. We give the details for completeness. As for Theorem 2.4.3, $\Psi_\lambda(\cdot, \cdot)$ and $\check{\Psi}_\lambda(\cdot, \cdot)$ are well-defined as $\mathcal{X}$ -valued functions.

Lemma 2.8.1. *Suppose that $x, y \in \mathcal{X}$ are such that $x \preccurlyeq_A y$ and the following hold: whenever $y \in \partial_i(K_A + x) \cap \mathcal{X}$ for some $1 \leq i \leq m$, we have that for each $1 \leq k \leq s$,*

$$\sum_{j \in G_i^{k,-}} \check{\Upsilon}_j(y) \leq \sum_{j \in G_i^{k,-}} \Upsilon_j(x), \quad \text{where } G_i^{k,-} = \{j \in G^k \mid \langle A_{i\bullet}, v_j \rangle = -1\}, \quad (2.8.5)$$

and

$$\sum_{j \in G_i^{k,+}} \check{\Upsilon}_j(y) \geq \sum_{j \in G_i^{k,+}} \Upsilon_j(x), \quad \text{where } G_i^{k,+} = \{j \in G^k \mid \langle A_{i\bullet}, v_j \rangle = 1\}. \quad (2.8.6)$$

Then, for each $u \in [0, 1]$:

$$\Psi_\lambda(x, u) \preccurlyeq_A \check{\Psi}_\lambda(y, u). \quad (2.8.7)$$

Proof. First, we note that $\Psi_\lambda, \check{\Psi}_\lambda$ have the following properties: for every $u \in [0, 1]$,

$$1 \leq k \leq s, j \in G^k,$$

$$\text{if } \Psi_\lambda(x, u) = x + v_j, \text{ then } \check{\Psi}_\lambda(y, u) \in \{y + v_\ell : \ell \in G^k\} \cup \{y\}, \quad (2.8.8)$$

since $I_{\sigma^{-1}(j)}^k(x), \check{I}_{\sigma^{-1}(\ell)}^k(y) \subseteq [\frac{p_{k-1}}{n}, \frac{p_k}{n})$ for $\ell \in G^k$. Similarly,

$$\text{if } \check{\Psi}_\lambda(y, u) = y + v_j, \text{ then } \Psi_\lambda(x, u) \in \{x + v_\ell : \ell \in G^k\} \cup \{x\}. \quad (2.8.9)$$

Furthermore, for $1 \leq k \leq s, 1 \leq i \leq m, j \in G_i^{k,+}$, if $\sum_{\ell \in G_i^{k,+}} \check{\Upsilon}_\ell(y) \geq \sum_{\ell \in G_i^{k,+}} \Upsilon_\ell(x)$, then

$$\Psi_\lambda(x, u) = x + v_j \text{ implies that } \check{\Psi}_\lambda(y, u) = y + v_\ell \text{ for some } \ell \in G_i^{k,+}, \quad (2.8.10)$$

since under the condition, we have $\cup_{q=p_{k-1}+1}^{q^*} I_q^k(x) \subseteq \cup_{q=p_{k-1}+1}^{q^*} \check{I}_q^k(y)$ where

$$q^* = \max\{q \mid p_{k-1} + 1 \leq q \leq p_k \text{ and } \langle A_{i\bullet}, v_{\sigma(q)} \rangle = 1\}$$

and, by SI-Assumption 2.8.1, $G_i^{k,+} = \{\sigma(q) \mid p_{k-1} + 1 \leq q \leq q^*\}$. Similarly, for $1 \leq k \leq s, 1 \leq i \leq m, j \in G_i^{k,-}$, if $\sum_{\ell \in G_i^{k,-}} \check{\Upsilon}_\ell(y) \leq \sum_{\ell \in G_i^{k,-}} \Upsilon_\ell(x)$, then

$$\check{\Psi}_\lambda(y, u) = y + v_j \text{ implies that } \Psi_\lambda(x, u) = x + v_\ell \text{ for some } \ell \in G_i^{k,-}. \quad (2.8.11)$$

We also have that, for $1 \leq k \leq s$ and $j \in G^k, x \preceq_A y + v_j$ if and only if

$$\langle A_{i\bullet}, y - x \rangle + \langle A_{i\bullet}, v_j \rangle \geq 0, \quad \text{for every } 1 \leq i \leq m.$$

Similarly, $x + v_j \preceq_A y$ if and only if

$$\langle A_{i\bullet}, y - x \rangle - \langle A_{i\bullet}, v_j \rangle \geq 0, \quad \text{for every } 1 \leq i \leq m.$$

Furthermore, for $1 \leq k \leq s$ and $j, \ell \in G^k$, $x + v_\ell \preceq_A y + v_j$ if and only if

$$\langle A_{i\bullet}, y - x \rangle + \langle A_{i\bullet}, v_j - v_\ell \rangle \geq 0, \quad \text{for every } 1 \leq i \leq m.$$

To prove (2.8.7), we first consider the situation where $y \in \text{int}(K_A + x) = \{w \in \mathbb{R}^d \mid Ax < Aw\}$. Then, for each $1 \leq i \leq m$, $\langle A_{i\bullet}, y - x \rangle > 0$ and since $A \in \mathbb{Z}^{m \times d}$ and $y - x \in \mathbb{Z}^d$, we have $\langle A_{i\bullet}, y - x \rangle \geq 1$. This implies that for each $1 \leq k \leq s$ and $j \in G^k$,

$$\langle A_{i\bullet}, y - x \rangle + \langle A_{i\bullet}, v_j \rangle \geq 1 + \langle A_{i\bullet}, v_j \rangle \geq 0, \quad \text{for every } 1 \leq i \leq m, \quad (2.8.12)$$

since $\langle A_{i\bullet}, v_j \rangle \in \{-1, 0, 1\}$ by condition (i) of SI-Theorem 2.8.2. Similarly, for each $1 \leq k \leq s$ and $j \in G^k$,

$$\langle A_{i\bullet}, y - x \rangle - \langle A_{i\bullet}, v_j \rangle \geq 1 - \langle A_{i\bullet}, v_j \rangle \geq 0, \quad \text{for every } 1 \leq i \leq m. \quad (2.8.13)$$

In addition, for $1 \leq k \leq s$ and $j, \ell \in G^k$,

$$\langle A_{i\bullet}, y - x \rangle + \langle A_{i\bullet}, v_j - v_\ell \rangle \geq 1 + \langle A_{i\bullet}, v_j - v_\ell \rangle \geq 0, \quad \text{for every } 1 \leq i \leq m, \quad (2.8.14)$$

since, by SI-Assumption 2.8.1, if $\langle A_{i\bullet}, v_j \rangle \neq 0$, then either $\langle A_{i\bullet}, v_\ell \rangle = \langle A_{i\bullet}, v_j \rangle$ or $\langle A_{i\bullet}, v_\ell \rangle = 0$. It follows from (2.8.12)–(2.8.14) that if $y \in \text{int}(K_A + x) \cap \mathcal{X}$, then for any $1 \leq k \leq s$ and $j, \ell \in G^k$:

$$x \preceq_A y + v_j, \quad x + v_j \preceq_A y \quad \text{and} \quad x + v_\ell \preceq_A y + v_j.$$

We also have, by assumption, that $x \preceq_A y$. It follows that if $y \in \text{int}(K_A + x) \cap \mathcal{X}$, then $\{x, x + v_\ell \mid \ell \in G^k\} \preceq_A \{y, y + v_j \mid j \in G^k\}$ for $1 \leq k \leq s$ and consequently (2.8.7) holds for all $u \in [0, 1]$.

Now, we turn to the other situation where $y \in \partial_i(K_A + x) \cap \mathcal{X}$ for some $1 \leq i \leq m$.

Then $\mathbf{K}_y := \{i \mid \langle A_{i\bullet}, y \rangle = \langle A_{i\bullet}, x \rangle, 1 \leq i \leq m\}$ is non-empty. Let $u \in [0, 1]$. We consider two cases.

Case 1: $\check{\Psi}_\lambda(y, u) = y + v_j$ for some $1 \leq j \leq n$.

Fix such an index j . Consider the unique $1 \leq k \leq s$ such that $j \in G^k$. Then, by (2.8.9), either $\Psi_\lambda(x, u) = x + v_\ell$ for some $\ell \in G^k$, or $\Psi_\lambda(x, u) = x$.

a) Suppose $\Psi_\lambda(x, u) = x + v_\ell$ for some $\ell \in G^k$. Observe that for every $i \notin \mathbf{K}_y$, $\langle A_{i\bullet}, y - x \rangle > 0$ and as for (2.8.14), $\langle A_{i\bullet}, (y + v_j) - (x + v_\ell) \rangle \geq 0$, while for $i \in \mathbf{K}_y$, $\langle A_{i\bullet}, (y + v_j) - (x + v_\ell) \rangle = \langle A_{i\bullet}, v_j \rangle - \langle A_{i\bullet}, v_\ell \rangle$. For each $i \in \mathbf{K}_y$,

i) if $\langle A_{i\bullet}, v_j \rangle = -1$ and $\langle A_{i\bullet}, v_\ell \rangle = 0$, then $j \in G_i^{k,-}$ and $\ell \notin G_i^{k,-}$. By (2.8.5), we would then have $\sum_{r \in G_i^{k,-}} \check{\Upsilon}_r(y) \leq \sum_{r \in G_i^{k,-}} \Upsilon_r(x)$, which would imply by (2.8.11) that $\Psi_\lambda(x, u) = x + v_r$ for some $r \in G_i^{k,-}$. Since we are assuming that $\Psi_\lambda(x, u) = x + v_\ell$ and we know the vectors $v_1, \dots, v_n$ are distinct, we obtain that $\ell = r \in G_i^{k,-}$. This contradicts $\ell \notin G_i^{k,-}$.

ii) if $\langle A_{i\bullet}, v_j \rangle = 0$ and $\langle A_{i\bullet}, v_\ell \rangle = 1$, then $j \notin G_i^{k,+}$ and $\ell \in G_i^{k,+}$. By (2.8.6), we would then have $\sum_{r \in G_i^{k,+}} \check{\Upsilon}_r(y) \geq \sum_{r \in G_i^{k,+}} \Upsilon_r(x)$, which would imply by (2.8.10) that $j \in G_i^{k,+}$. This contradicts $j \notin G_i^{k,+}$.

iii) in all the other cases, that is when

$$(\langle A_{i\bullet}, v_j \rangle, \langle A_{i\bullet}, v_\ell \rangle) \in \{(1, 0), (0, -1), (1, 1), (0, 0), (-1, -1)\},$$

we have $\langle A_{i\bullet}, (y + v_j) - (x + v_\ell) \rangle \geq 0$.

Combining the above for case a), we see that $\langle A_{i\bullet}, (y + v_j) - (x + v_\ell) \rangle \geq 0$ for each $1 \leq i \leq m$, which implies that $\Psi_\lambda(x, u) = x + v_\ell \preceq_A y + v_j = \check{\Psi}_\lambda(y, u)$.

b) Suppose $\Psi_\lambda(x, u) = x$. We claim that $y + v_j \in K_A + x$. To see this, observe that for every $i \notin \mathbf{K}_y$, $\langle A_{i\bullet}, y - x \rangle > 0$ and as for (2.8.12), $\langle A_{i\bullet}, (y + v_j) - x \rangle \geq 0$, while for

$i \in \mathbf{K}_y$, $\langle A_{i\bullet}, (y + v_j) - x \rangle = \langle A_{i\bullet}, v_j \rangle \in \{-1, 0, 1\}$. For each $i \in \mathbf{K}_y$, if $\langle A_{i\bullet}, v_j \rangle = -1$, which means $j \in G_i^{k,-}$, then by (2.8.5) we would have $\sum_{\ell \in G_i^{k,-}} \check{\Upsilon}_\ell(y) \leq \sum_{\ell \in G_i^{k,-}} \Upsilon_\ell(x)$, which would imply by (2.8.11) that $\Psi_\lambda(x, u) = x + v_\ell$ for some $\ell \in G_i^{k,-}$, but this contradicts the assumption that $\Psi_\lambda(x, u) = x$. So we must have $\langle A_{i\bullet}, v_j \rangle \geq 0$ and hence $\langle A_{i\bullet}, (y + v_j) - x \rangle \geq 0$ for all $i \in \mathbf{K}_y$. Thus, $y + v_j \in K_A + x$ and so $\Psi_\lambda(x, u) = x \preceq_A y + v_j = \check{\Psi}_\lambda(y, u)$.

Case 2: $\check{\Psi}_\lambda(y, u) = y$. Again, we consider two subcases.

- a) If $\Psi_\lambda(x, u) = x$, then (2.8.7) holds, because $x \preceq_A y$.
- b) If $\Psi_\lambda(x, u) = x + v_j$ for some $1 \leq j \leq n$, we claim that $y \in K_A + x + v_j$ for the corresponding value of j . To see this, fix the value of j for which $\Psi_\lambda(x, u) = x + v_j$, let $1 \leq k \leq s$ be such that $j \in G^k$, and observe that for every $i \notin \mathbf{K}_y$, $\langle A_{i\bullet}, y - x \rangle > 0$ and as for (2.8.13), $\langle A_{i\bullet}, y - (x + v_j) \rangle \geq 0$, while for $i \in \mathbf{K}_y$, $\langle A_{i\bullet}, y - (x + v_j) \rangle = -\langle A_{i\bullet}, v_j \rangle \in \{-1, 0, 1\}$. For each $i \in \mathbf{K}_y$, if $\langle A_{i\bullet}, v_j \rangle = 1$, which means $j \in G_i^{k,+}$, then by (2.8.6), we would have $\sum_{\ell \in G_i^{k,+}} \check{\Upsilon}_\ell(y) \geq \sum_{\ell \in G_i^{k,+}} \Upsilon_\ell(x)$, which would imply by (2.8.10) that $\check{\Psi}_\lambda(y, u) = y + v_\ell$ for some $\ell \in G_i^{k,+}$, but this contradicts the assumption that $\check{\Psi}_\lambda(y, u) = y$. So we must have $\langle A_{i\bullet}, v_j \rangle \leq 0$ and hence $\langle A_{i\bullet}, y - (x + v_j) \rangle \geq 0$ for all $i \in \mathbf{K}_y$. Thus, we have $y \in K_A + x + v_j$ and then $\Psi_\lambda(x, u) = x + v_j \preceq_A y = \check{\Psi}_\lambda(y, u)$.

□

In order to prove SI-Theorem 2.8.2, from here on we can follow a similar procedure to the one used in the proof of Theorem 2.4.1 after Lemma 2.6.1 was proved there. For the case where (2.6.1) holds, we define two discrete-time processes, $Y = (Y_k)_{k \geq 0}$ and $\check{Y} = (\check{Y}_k)_{k \geq 0}$, by defining $Y_0 := x^\circ$, $\check{Y}_0 := \check{x}^\circ$, and for $k \geq 0$,

$$Y_{k+1} := \Psi_\lambda(Y_k, U_{k+1}), \quad \check{Y}_{k+1} := \check{\Psi}_\lambda(\check{Y}_k, U_{k+1}), \quad (2.8.15)$$

and define X and $\check{X}$ using these and an independent Poisson process N as in (2.6.13). For the case where (2.6.1) does not hold, we can use a truncation procedure similar to that for Theorem 2.4.1. In both cases, we use SI-Lemma 2.8.1 instead of Lemma 2.6.1.

Two other A matrices for Example 2.5.3

Let

$$A = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & -1 & 0 \end{bmatrix}. \quad (2.8.16)$$

For $x \in \mathcal{X}$, consider infinitesimal transition rates $\check{\Upsilon}_1(x), \check{\Upsilon}_2(x), \check{\Upsilon}_3(x), \check{\Upsilon}_4(x)$ and $\check{\Upsilon}_5(x)$ defined as for $\Upsilon_1(x), \Upsilon_2(x), \Upsilon_3(x), \Upsilon_4(x)$ and $\Upsilon_5(x)$ in (2.5.5), but with $\check{\kappa}_i$ in place of κ_i where $\check{\kappa}_i = \kappa_i$, for $i = 1, 2, 3, 4$, and $\check{\kappa}_5 \leq \kappa_5$. Suppose that $\kappa_2 > \kappa_4$. Now, let us verify that the assumptions of SI-Theorem 2.8.2 hold. Condition (i) holds since $Av_1 = (1, 0, -1)^T$, $Av_2 = (0, 0, 1)^T$, $Av_3 = (1, -1, -1)^T$, $Av_4 = (0, 1, 1)^T$ and $Av_5 = (0, -1, 0)^T$. SI-Assumption 2.8.1 holds with $G^1 = \{3, 1\}$, $G^2 = \{4, 2\}$, $G^3 = \{5\}$ and $\sigma(1) = 3$, $\sigma(2) = 1$, $\sigma(3) = 4$, $\sigma(4) = 2$, $\sigma(5) = 5$. To verify that condition (ii) of SI-Theorem 2.8.2 holds, fix $x \in \mathcal{X}$ and first consider $y \in \partial_1(K_A + x) \cap \mathcal{X}$, where $\partial_1(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 = w_1, x_3 \geq w_3, x_2 + x_3 \geq w_2 + w_3, x_4 \leq w_4\}$. Given that $\langle A_{1\bullet}, v_1 \rangle = \langle A_{1\bullet}, v_3 \rangle = 1$, we need to check that $\Upsilon_1(x) + \Upsilon_3(x) \leq \check{\Upsilon}_1(y) + \check{\Upsilon}_3(y)$. Since $y \in \partial_1(K_A + x) \cap \mathcal{X}$, then $\Upsilon_1(x) = \kappa_1 x_1 = \kappa_1 y_1 = \check{\kappa}_1 y_1 = \check{\Upsilon}_1(y)$ and $\Upsilon_3(x) = \kappa_3 x_1 = \kappa_3 y_1 = \check{\kappa}_3 y_1 = \check{\Upsilon}_3(y)$, and so the desired inequality holds with equality. Secondly, consider $y \in \partial_2(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 \geq w_1, x_3 = w_3, x_2 \geq w_2, x_4 \leq w_4\}$. Given that $\langle A_{2\bullet}, v_3 \rangle = \langle A_{2\bullet}, v_5 \rangle = -1$ and $\langle A_{2\bullet}, v_4 \rangle = 1$, we need to check that $\Upsilon_3(x) \geq \check{\Upsilon}_3(y)$, $\Upsilon_4(x) \leq \check{\Upsilon}_4(y)$ and $\Upsilon_5(x) \geq \check{\Upsilon}_5(y)$. Since $y \in \partial_2(K_A + x) \cap \mathcal{X}$, then $\Upsilon_3(x) = \kappa_3 x_1 \geq \kappa_3 y_1 = \check{\kappa}_3 y_1 = \check{\Upsilon}_3(y)$, $\Upsilon_4(x) = \kappa_4 x_3 = \kappa_4 y_3 = \check{\kappa}_4 y_3 = \check{\Upsilon}_4(y)$ and $\Upsilon_5(x) = \kappa_5 x_2 \geq \kappa_5 y_2 \geq \check{\kappa}_5 y_2 = \check{\Upsilon}_5(y)$, and so the desired inequality holds. Lastly, consider $y \in \partial_3(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 \geq w_1, x_3 \geq w_3, x_2 + x_3 = w_2 + w_3, x_4 \leq w_4\}$. Given that $\langle A_{1\bullet}, v_1 \rangle = \langle A_{1\bullet}, v_3 \rangle = -1$ and $\langle A_{1\bullet}, v_2 \rangle = \langle A_{1\bullet}, v_4 \rangle = 1$, we need to check

that $\Upsilon_1(x) + \Upsilon_3(x) \geq \check{\Upsilon}_1(y) + \check{\Upsilon}_3(y)$ and $\Upsilon_2(x) + \Upsilon_4(x) \leq \check{\Upsilon}_2(y) + \check{\Upsilon}_4(y)$. For $y \in \partial_3(K_A + x) \cap \mathcal{X}$, since $\kappa_2 > \kappa_4$ was assumed, we have that $\Upsilon_2(x) + \Upsilon_4(x) = \kappa_2 x_2 + \kappa_4 x_3 = (\kappa_2 - \kappa_4)x_2 + \kappa_4(x_2 + x_3) \leq (\kappa_2 - \kappa_4)y_2 + \kappa_4(y_2 + y_3) = \check{\kappa}_2 y_2 + \check{\kappa}_4 y_3 = \check{\Upsilon}_2(y) + \check{\Upsilon}_4(y)$ and $\Upsilon_1(x) = \kappa_1 x_1 \geq \kappa_1 y_1 = \check{\kappa}_1 y_1 = \check{\Upsilon}_1(y)$, $\Upsilon_3(x) = \kappa_3 x_1 \geq \kappa_3 y_1 = \check{\kappa}_3 y_1 = \check{\Upsilon}_3(y)$. Thus, the conditions of SI-Theorem 2.8.2 are satisfied and so the conclusion of that theorem holds.

Let $\Gamma = \{(0, 0, 0, S_{\text{tot}})\}$. This is an increasing set in $\mathcal{X}$ with respect to the relation $\preccurlyeq_A$. Let $T_{(0,0,0,S_{\text{tot}})}$, respectively $\check{T}_{(0,0,0,S_{\text{tot}})}$ be the first time that the Markov chain X , respectively $\check{X}$, reaches the set Γ . Then, by the generalization of Theorem 2.4.4, if $X(0) = \check{X}(0) = (S_{\text{tot}}, 0, 0, 0)$, we have that $\check{T}_{(0,0,0,S_{\text{tot}})} \preccurlyeq_{st} T_{(0,0,0,S_{\text{tot}})}$. It follows that increasing κ_5 will increase the mean first passage time from $(S_{\text{tot}}, 0, 0, 0)$ to $(0, 0, 0, S_{\text{tot}})$ when $\kappa_2 > \kappa_4$ (see Figure 2.7). Indeed, when $\kappa_2 > \kappa_4$, it takes a longer time to get to $(0, 0, 0, S_{\text{tot}})$ from $(S_{\text{tot}}, 0, 0, 0)$ if reaction $\textcircled{5}$ is added to the system without that reaction.

On the other hand, suppose

$$A = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & -1 & -1 & 0 \end{bmatrix}$$

and infinitesimal transition rates $\check{\Upsilon}_1(x), \check{\Upsilon}_2(x), \check{\Upsilon}_3(x), \check{\Upsilon}_4(x)$ and $\check{\Upsilon}_5(x)$ are defined as for $\Upsilon_1(x), \Upsilon_2(x), \Upsilon_3(x), \Upsilon_4(x)$ and $\Upsilon_5(x)$ in (2.5.5), but with $\check{\kappa}_i$ in place of κ_i where $\check{\kappa}_i = \kappa_i$, for $i = 1, 2, 3, 4$, $\check{\kappa}_5 \geq \kappa_5$, and $\kappa_2 < \kappa_4$. We can verify that the assumptions of SI-Theorem 2.8.2 hold, as follows. Condition (i) holds since $Av_1 = (1, -1, -1)^T$, $Av_2 = (0, 1, 1)^T$, $Av_3 = (1, 0, -1)^T$, $Av_4 = (0, 0, 1)^T$ and $Av_5 = (0, 1, 0)^T$. SI-Assumption 2.8.1 holds with $G^1 = \{1, 3\}$, $G^2 = \{2, 4\}$, $G^3 = \{5\}$ and $\sigma(1) = 1$, $\sigma(2) = 3$, $\sigma(3) = 2$, $\sigma(4) = 4$, $\sigma(5) = 5$. To verify that condition (ii) of SI-Theorem 2.8.2 holds, let $x \in \mathcal{X}$, and first consider $y \in \partial_1(K_A + x) \cap \mathcal{X}$, where $\partial_1(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 = w_1, x_2 \geq w_2, x_2 + x_3 \geq w_2 + w_3, x_4 \leq w_4\}$. Given that $\langle A_{1\bullet}, v_1 \rangle = \langle A_{1\bullet}, v_3 \rangle = 1$, we

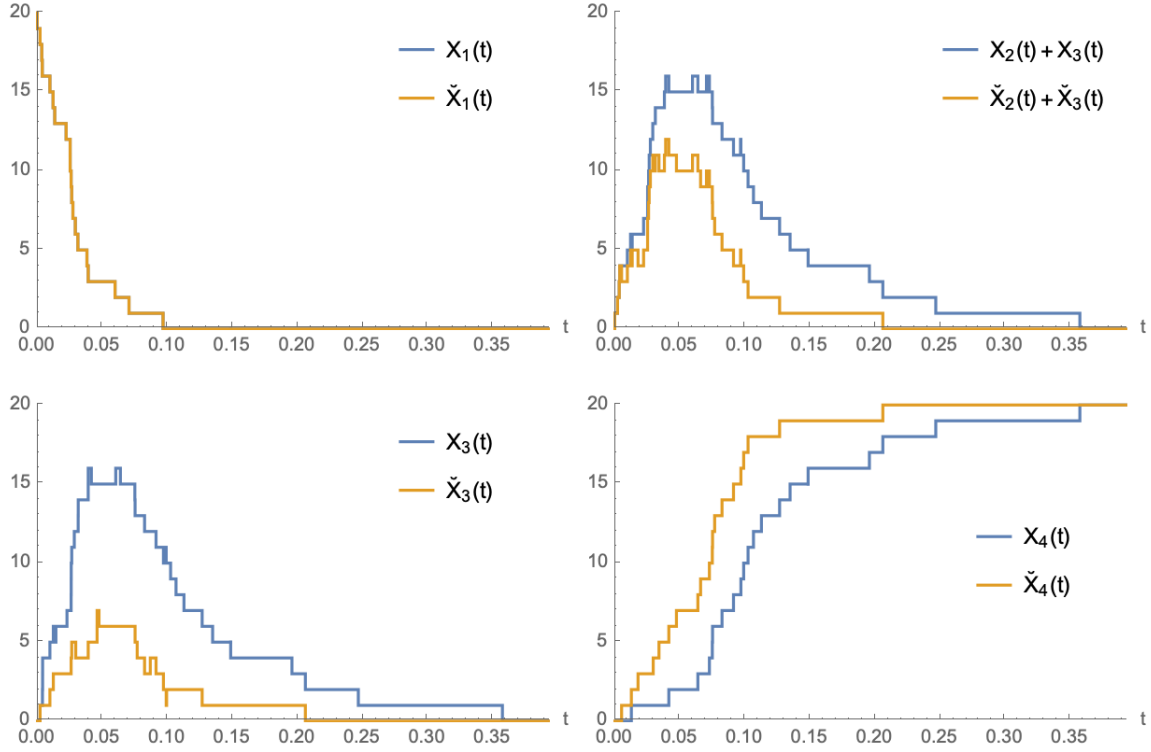


Figure 2.7. A typical coupled realization of sample paths for Example 2.5.3. Here $S_{\text{tot}} = 20$, $\kappa_1 = \check{\kappa}_1 = 30$, $\kappa_2 = \check{\kappa}_2 = 50$, $\kappa_3 = \check{\kappa}_3 = 10$, $\kappa_4 = \check{\kappa}_4 = 10$, $\kappa_5 = 1000$ and $\check{\kappa}_5 = 10$. Both processes X and $\check{X}$ start at $(S_{\text{tot}}, 0, 0, 0)$. As shown in SI-Section 2.8.3, for these parameters, we have, almost surely, $X(t) \preceq_A \check{X}(t)$ for every $t \geq 0$ where the matrix A is given in (2.8.16). A coupled realization using the algorithm described in SI-Section 2.8.4 is plotted to illustrate this result. In particular, we see in this sample that $X_1(t) \geq \check{X}_1(t)$, $X_2(t) + X_3(t) \geq \check{X}_2(t) + \check{X}_3(t)$ and $X_3(t) \geq \check{X}_3(t)$ for all times t . Moreover, the first passage time to $(0, 0, 0, S_{\text{tot}})$ for X (which is equivalent to the first time to get to the state where $X_4 = S_{\text{tot}}$) is larger than for $\check{X}$. Since this first passage time result is true for all coupled samples, we can conclude that the mean first passage time from $(S_{\text{tot}}, 0, 0, 0)$ to $(0, 0, 0, S_{\text{tot}})$ is larger for X than for $\check{X}$.

need to check that $\Upsilon_1(x) + \Upsilon_3(x) \leq \check{\Upsilon}_1(y) + \check{\Upsilon}_3(y)$. Since $y \in \partial_1(K_A + x) \cap \mathcal{X}$, then $\Upsilon_1(x) = \kappa_1 x_1 = \kappa_1 y_1 = \check{\kappa}_1 y_1 = \check{\Upsilon}_1(y)$ and $\Upsilon_3(x) = \kappa_3 x_1 = \kappa_3 y_1 = \check{\kappa}_3 y_1 = \check{\Upsilon}_3(y)$, and so the desired inequality holds with equality. Secondly, consider $y \in \partial_2(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 \geq w_1, x_2 = w_2, x_2 + x_3 \geq w_2 + w_3, x_4 \leq w_4\}$. Given that $\langle A_{2\bullet}, v_2 \rangle = \langle A_{2\bullet}, v_5 \rangle = 1$ and $\langle A_{2\bullet}, v_1 \rangle = -1$, we need to check that $\Upsilon_1(x) \geq \check{\Upsilon}_1(y)$, $\Upsilon_2(x) \leq \check{\Upsilon}_2(y)$ and $\Upsilon_5(x) \leq \check{\Upsilon}_5(y)$. Since $y \in \partial_2(K_A + x) \cap \mathcal{X}$, then $\Upsilon_1(x) = \kappa_1 x_1 \geq \kappa_1 y_1 = \check{\kappa}_1 y_1 = \check{\Upsilon}_1(y)$,

$\Upsilon_2(x) = \kappa_2 x_2 = \kappa_2 y_2 = \check{\kappa}_2 y_2 = \check{\Upsilon}_2(y)$ and $\Upsilon_5(x) = \kappa_5 x_2 = \kappa_5 y_2 \leq \check{\kappa}_5 y_2 = \check{\Upsilon}_5(y)$, and so the desired inequality holds. Lastly, consider $y \in \partial_3(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} \mid x_1 \geq w_1, x_2 \geq w_2, x_2 + x_3 = w_2 + w_3, x_4 \leq w_4\}$. Given that $\langle A_{1\bullet}, v_1 \rangle = \langle A_{1\bullet}, v_3 \rangle = -1$ and $\langle A_{1\bullet}, v_2 \rangle = \langle A_{1\bullet}, v_4 \rangle = 1$, we need to check that $\Upsilon_1(x) + \Upsilon_3(x) \geq \check{\Upsilon}_1(y) + \check{\Upsilon}_3(y)$ and $\Upsilon_2(x) + \Upsilon_4(x) \leq \check{\Upsilon}_2(y) + \check{\Upsilon}_4(y)$. For $y \in \partial_3(K_A + x) \cap \mathcal{X}$, since $\kappa_2 < \kappa_4$, we have that $\Upsilon_1(x) + \Upsilon_3(x) = (\kappa_1 + \kappa_3)x_1 \geq (\kappa_1 + \kappa_3)y_1 = (\check{\kappa}_1 + \check{\kappa}_3)y_1 = \check{\Upsilon}_1(y) + \check{\Upsilon}_3(y)$ and $\Upsilon_2(x) + \Upsilon_4(x) = \kappa_2 x_2 + \kappa_4 x_3 = (\kappa_4 - \kappa_2)x_3 + \kappa_2(x_2 + x_3) \leq (\kappa_4 - \kappa_2)y_3 + \kappa_2(y_2 + y_3) = \check{\kappa}_2 y_2 + \check{\kappa}_4 y_3 = \check{\Upsilon}_2(y) + \check{\Upsilon}_4(y)$. Thus, the conditions of SI-Theorem 2.8.2 are satisfied. In particular, we can conclude when $\kappa_2 < \kappa_4$, that it takes less time to get to $(0, 0, 0, S_{\text{tot}})$ from $(S_{\text{tot}}, 0, 0, 0)$ if reaction $\textcircled{5}$ is added to the system without that reaction⁹.

2.8.4 An algorithm for coupled stochastic simulation

We now provide an algorithm for stochastic simulation of the coupled continuous time Markov chains X and $\check{X}$ under the conditions of Theorems 2.4.1, 2.4.2, 2.4.3 or SI-Theorem 2.8.2, when the transitions rates are bounded on the state space, i.e., when (2.6.1) holds.

⁹The system without reaction $\textcircled{5}$ can be obtained from the system with reaction $\textcircled{5}$ by setting $\kappa_5 = 0$. While strictly speaking a zero rate constant is not within our definition of mass action kinetics, our theory does cover propensity functions with such a zero rate constant.

Algorithm: Stochastic simulation for coupled continuous time Markov chains

X and $\check{X}$.

Data: Integer $n \geq 1$, real $T > 0$, set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, vectors $v_1, \dots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$,
 $x^\circ, \check{x}^\circ$ in $\mathcal{X}$, functions $\Upsilon = (\Upsilon_1, \dots, \Upsilon_n)$ and $\check{\Upsilon} = (\check{\Upsilon}_1, \dots, \check{\Upsilon}_n)$ and
integer $a \in \{1, 2, 3, 4\}$ to indicate which theorem is invoked (2.4.1, 2.4.2,
2.4.3 or 2.8.2).

Result: Sample of initial time and subsequent potential jump times

$T_0, T_1, \dots, T_N$ and associated states $X(T_0), X(T_1), \dots, X(T_N)$ and
 $\check{X}(T_0), \check{X}(T_1), \dots, \check{X}(T_N)$ for the continuous time Markov chains X
and $\check{X}$ in the time interval $[0, T]$.

$\lambda \leftarrow 1 + n \max \left\{ \sup_{x \in \mathcal{X}} \sum_{j=1}^n \Upsilon_j(x), \sup_{x \in \mathcal{X}} \sum_{j=1}^n \check{\Upsilon}_j(x) \right\};$

$K \leftarrow 0;$

$T_0 \leftarrow 0;$

while $T_K \leq T$ **do**

$T_{K+1} \leftarrow T_K + \text{Exponential}(\lambda);$

$K \leftarrow K + 1;$

$Y_0, \check{Y}_0 \leftarrow x^\circ, \check{x}^\circ;$

$X(T_0), \check{X}(T_0) \leftarrow x^\circ, \check{x}^\circ;$

$N \leftarrow K - 1;$

if $N \geq 1$ **then**

for $k \leftarrow 0$ **to** $N - 1$ **do**

$U \leftarrow \text{Uniform}([0, 1]);$

$Y_{k+1}, \check{Y}_{k+1} \leftarrow \text{TransitionDTMC}(\Upsilon, \lambda, Y_k, U, a),$

$\text{TransitionDTMC}(\check{\Upsilon}, \lambda, \check{Y}_k, U, a);$

$X(T_{k+1}), \check{X}(T_{k+1}) \leftarrow Y_{k+1}, \check{Y}_{k+1};$

The random variables $T_1, \dots, T_N$ are called potential jump times because it could be that $X(T_k) = X(T_{k+1})$ or $\check{X}(T_k) = \check{X}(T_{k+1})$ for some $0 \leq k \leq N-1$. Letting $T_{N+1} := T$, the trajectories of X are given by $X(t) = X(T_k)$ for $T_k \leq t < T_{k+1}$, $0 \leq k \leq N$, and similarly for the trajectories of $\check{X}$.

The function **TransitionDTMC** can be found below. This function is meant to replicate $\Phi_\lambda(x, u)$ in (2.6.4) for the case of Theorem 2.4.1 and Theorem 2.4.2, and $\Psi_\lambda(x, u)$ in (2.6.27) for the case of Theorem 2.4.3 and SI-Theorem 2.8.2.

Function **TransitionDTMC**($\Upsilon, \lambda, x, u, a$)

Data: Integer $n \geq 1$, set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, vectors $v_1, \dots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$.

Input: Function $\Upsilon = (\Upsilon_1, \dots, \Upsilon_n)$, $\lambda > 0$, $x \in \mathcal{X}$, $u \in [0, 1]$. Integer

$a \in \{1, 2, 3, 4\}$ to indicate which theorem is invoked (2.4.1, 2.4.2, 2.4.3 or 2.8.2). For the case of Theorem 2.4.3 or SI-Theorem 2.8.2, include partition $\{G^1, \dots, G^s\}$ and bijection σ .

Output: State $x + v \in \mathcal{X}$.

$v \leftarrow 0$;

if $a \in \{1, 2\}$ **then** // The case of Theorem 2.4.1 or Theorem 2.4.2.

for $j \leftarrow 1$ **to** n **do**

if $\frac{j-1}{n} \leq u < \frac{j-1}{n} + \frac{\Upsilon_j(x)}{\lambda}$ **then**

$v \leftarrow v_j$;

else // The case of Theorem 2.4.3 or SI-Theorem 2.8.2.

$p_0 \leftarrow 0$;

for $k \leftarrow 1$ **to** s **do**

$p_k \leftarrow p_{k-1} + |G^k|$;

for $q \leftarrow p_{k-1} + 1$ **to** p_k **do**

if $\frac{p_{k-1}}{n} + \sum_{\ell=p_{k-1}+1}^{q-1} \frac{\Upsilon_{\sigma(\ell)}(x)}{\lambda} \leq u < \frac{p_{k-1}}{n} + \sum_{\ell=p_{k-1}+1}^q \frac{\Upsilon_{\sigma(\ell)}(x)}{\lambda}$ **then**

$v \leftarrow v_{\sigma(q)}$;

return $x + v$

Remark 2.8.3. The above algorithm can be adapted to provide simultaneous stochastic simulation for X and $\check{X}$ when transition rates are not bounded on the state space, by applying the algorithm on a sequence of bounded sets, which expand to the whole state space. This employs a sequence of successively defined stopping times $\tau_0 = 0$, $\tau_\ell = \inf\{t \geq \tau_{\ell-1} \mid X(t) \notin C_\ell \text{ or } \check{X}(t) \notin C_\ell\}$, $\ell = 1, 2, \dots$, where the C_ℓ are compact, $C_\ell \subseteq C_{\ell+1}$ for $\ell = 1, 2, \dots$ and $\cup_{\ell=1}^\infty C_\ell = \mathcal{X}$. The simulation of the pair $(X, \check{X})(t)$ for $\tau_\ell \leq t < \tau_{\ell+1}$ uses the above algorithm on C_ℓ for $\ell = 1, 2, \dots$.

2.9 Acknowledgement

Chapter 2 is reproduced with minor enumeration and formatting differences from the paper “Comparison theorems for stochastic chemical reaction networks” by Campos, F.A., Bruno, S., Fu, Y., Del Vecchio, D. and Williams, R.J., which has appeared in the *Bulletin of Mathematical Biology*. The dissertation author proposed and proved a generalization of the basic comparison theorem, developed a novel example inspired by Braess’ paradox, provided many details for examples used to illustrate the theory in the paper, and contributed to the writing of the manuscript.

Bibliography for Chapter 2

- [1] David F. Anderson, Gheorghe Craciun, and Thomas G. Kurtz. “Product-Form stationary distributions for deficiency zero chemical reaction networks”. In: *Bulletin of Mathematical Biology* 72 (2010), pp. 1947–1970.
- [2] David F. Anderson and Thomas G. Kurtz. *Stochastic Analysis of Biochemical Systems*. Mathematical Biosciences Institute Graduate Lecture Series. Springer International Publishing, 2015.

- [3] David Angeli, Patrick De Leenheer, and Eduardo D. Sontag. “On the structural monotonicity of chemical reaction networks”. In: *Proceedings of the 45th IEEE Conference on Decision and Control* (2006), pp. 7–12. DOI: 10.1109/CDC.2006.376698.
- [4] David Angeli and Eduardo D. Sontag. “Monotone control systems”. In: *IEEE Transactions on Automatic Control* 48.10 (2003), pp. 1684–1698. DOI: 10.1109/TAC.2003.817920.
- [5] Patrick Billingsley. *Convergence of Probability Measures (Second Edition)*. Wiley Series in Probability and Statistics. John Wiley & Sons Inc., 1999.
- [6] Dietrich Braess. “Über ein Paradoxon aus der Verkehrsplanung”. In: *Unternehmensforschung* 12 (1968), pp. 258–268.
- [7] Simone Bruno, Ruth J. Williams, and Domitilla Del Vecchio. “Epigenetic cell memory: the gene’s inner chromatin modification circuit”. In: *PLOS Computational Biology* 18.4 (2022), pp. 1–27. DOI: 10.1371/journal.pcbi.1009961.
- [8] Bruce Calvert, Wiremu Solomon, and Ilze Ziedins. “Braess’s paradox in a queueing network with state-dependent routing”. In: *Journal of Applied Probability* 34.1 (1997), pp. 134–154. DOI: 10.2307/3215182.
- [9] Domitilla Del Vecchio and Richard M. Murray. *Biomolecular Feedback Systems*. Princeton University Press, 2014.
- [10] Ian Dodd, Mille A. Micheelsen, Kim Sneppen, and Geneviève Thon. “Theoretical analysis of epigenetic cell memory by nucleosome modification”. In: *Cell* 129 (2007), pp. 813–822. DOI: 10.1016/j.cell.2007.02.053.
- [11] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes: Characterization and Convergence*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., 1986.

- [12] Roberta Gori, Paolo Milazzo, and Lucia Nasti. “Towards an efficient verification method for monotonicity properties of chemical reaction networks”. In: *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC)* 3 (2019), pp. 250–257. DOI: 10.5220/0007522002500257.
- [13] Winfried K. Grassmann. “Transient solutions in Markovian queueing systems”. In: *Computers & Operations Research* 4.1 (1977), pp. 47–53. DOI: 10.1016/0305-0548(77)90007-7.
- [14] Rudiyanto Gunawan, Yang Cao, Linda Petzold, and Francis J. Doyle. “Sensitivity analysis of discrete stochastic systems”. In: *Biophysical Journal* 88.4 (2005), pp. 2530–2540. DOI: 10.1529/biophysj.104.053405.
- [15] Ankit Gupta and Mustafa Khammash. “Sensitivity analysis for stochastic chemical reaction networks with multiple time-scales”. In: *Electronic Journal of Probability* 19 (2014), pp. 1–53. DOI: 10.1214/EJP.v19-3246.
- [16] Eric L. Haseltine and James B. Rawlings. “Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics”. In: *The Journal of Chemical Physics* 117 (2002), pp. 6959–6969. DOI: 10.1063/1.1505860.
- [17] Morris W. Hirsch and Hal Smith. “Chapter 4 Monotone Dynamical Systems”. In: *Handbook of Differential Equations: Ordinary Differential Equations (Volume 2)*. 2006, pp. 239–357.
- [18] Teturo Kamae, Ulrich Krengel, and George L. O’Brien. “Stochastic inequalities on partially ordered spaces”. In: *The Annals of Probability* 5.6 (1977), pp. 899–912. DOI: 10.1214/aop/1176995659.
- [19] Hye-Won Kang, Wasiur Rahman KhudaBukhsh, Heinz Koepl, and Grzegorz Rempala. “Quasi-steady-state approximations derived from the stochastic model of

- enzyme kinetics”. In: *Bulletin of Mathematical Biology* 81 (2019), pp. 1303–1336. DOI: 10.1007/s11538-019-00574-4.
- [20] Julian Keilson. *Markov Chain Models—rarity and Exponentiality*. Applied mathematical sciences. Springer New York, 1979.
 - [21] William A. Massey. “Stochastic orderings for Markov processes on partially ordered spaces”. In: *Mathematics of Operations Research* 12.2 (1987), pp. 350–367. DOI: 10.1287/moor.12.2.350.
 - [22] Sean P. Meyn and Richard L. Tweedie. “Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes”. In: *Advances in Applied Probability* 25.3 (1993), pp. 518–548.
 - [23] Leonor Michaelis and Maud L. Menten. “Die kinetik der invertinwirkung”. In: *Biochemische Zeitschrift* 49 (1913), pp. 333–369.
 - [24] Alfred Müller and Dietrich Stoyan. *Comparison Methods for Stochastic Models and Risks*. Wiley Series in Probability and Statistics. John Wiley & Sons Inc., 2002.
 - [25] Michael Sharpe. *General Theory of Markov Processes*. Pure and applied mathematics. Academic Press, 1988.
 - [26] Rishi K. Srivastava, Lingchong You, Jesse W. Summers, and John Yin. “Stochastic vs. deterministic modeling of intracellular viral kinetics”. In: *Journal of Theoretical Biology* 218.3 (2002), pp. 309–321. DOI: 10.1006/jtbi.2002.3078.

Chapter 3

Singular Perturbations

3.1 Abstract

Epigenetic cell memory, the inheritance of gene expression patterns across subsequent cell divisions, is a critical property of multi-cellular organisms. In recent work [12], a subset of the authors observed in a simulation study how the stochastic dynamics and time-scale differences between establishment and erasure processes in chromatin modifications (such as histone modifications and DNA methylation) can have a critical effect on epigenetic cell memory. In this work, we provide a mathematical framework to rigorously validate and extend beyond these computational findings. Viewing our stochastic model of a chromatin modification circuit as a singularly perturbed, finite state, continuous time Markov chain, we extend beyond existing theory in order to characterize the leading coefficients in the series expansions of stationary distributions and mean first passage times. In particular, we characterize the limiting stationary distribution in terms of a reduced Markov chain, provide an algorithm to determine the orders of the poles of mean first passage times, and determine how changing erasure rates affects system behavior. The theoretical tools developed in this work not only allow us to set a rigorous mathematical basis for the computational findings of our prior work, highlighting the effect of chromatin modification dynamics on epigenetic cell memory, but they can also be applied to other singularly perturbed Markov chains beyond the applications in this work, especially those

associated with chemical reaction networks.

3.2 Introduction

3.2.1 Background

Epigenetic cell memory, the inheritance of gene expression patterns across subsequent cell divisions (see for example Huang et al. [25]), is a critical property of multi-cellular organisms of intense interest in the field of systems biology (see for example Ptashne [37] and Xiong and Ferrell [38]). It has previously been discovered that chromatin modifications, such as DNA methylation and histone modifications, are key mediators of epigenetic cell memory (see for example Carey [15], Heard and Martienssen [23], Allis et al. [1], Kim and Costello [30], and references in Bruno et al. [12] for more biological background). More precisely, it was found via simulations of stochastic models that the time scale separation between establishment (fast) and erasure (slow) of these modifications extends the duration of cell memory, and that different asymmetries between erasure rates of chromatin modifications can lead to different gene expression patterns (see Bruno et al. [12, 13, 11]). Here, we provide a mathematical framework to rigorously validate these computational findings and to further explore models of chromatin modification circuits. We do this in a way that the results obtained and the tools developed can be applied to other mathematical models beyond the applications in this work, especially stochastic models of chemical reaction networks.

3.2.2 Focus of our work

In this work, we consider different versions of the chromatin modification circuit proposed in Bruno et al. [12]. In particular, we start with simpler circuits that include histone modifications only and then we consider more elaborate circuits that include also DNA methylation. All of these circuits can be viewed as examples of Stochastic Chemical

Reaction Networks (SCRNs). A SCRN is a continuous time Markov chain living in the non-negative integer lattice in d -dimensions, where the components of the Markov chain track the number of molecules of each of d species in the network over time, and each jump of the Markov chain corresponds to the firing of a reaction in the network (see for example Anderson and Kurtz [3]). A more precise description is given in Section 3.4.2.

In order to analyze these stochastic models, we first determine how the stationary distributions and mean first passage times between states vary when a small parameter ε (non-dimensional parameter that scales the speed of the basal erasure of all the chromatin modifications) tends to zero. To this end, we show that the stationary distributions and the mean first passage times of these singularly perturbed Markov chains admit series expansions in ε and we develop theoretical tools to determine the coefficients in these expansions. Then, we focus on determining how the different erasure rates of chromatin modifications affect the behavior of the chromatin modification circuit models. This latter study is conducted by exploiting comparison theorems for Markov chains recently developed in Campos et al. [14].

One of the key features of our work is that these tools and the associated mathematical results are not only applicable to the chromatin modification models, but they can also be used to analyze other models that respect the same set of assumptions.

3.2.3 Related work

As mentioned in the previous paragraph, the stochastic behavior of the chromatin modification circuit models can be described by singularly perturbed continuous time Markov chains. There is some literature on discrete and continuous time, singularly perturbed Markov chains, especially by Avrachenkov et al. [6], Hassin and Haviv [22], Beltrán and Landim [8, 9], and Yin and Zhang [39]. Avrachenkov et al. [6] gave general characterizations of series expansions for the stationary distribution and mean first passage times of a singularly perturbed *discrete time* Markov chain with finite state space. While

their theory can be in principle translated to continuous time Markov chains, our work mostly deals directly with the singularly perturbed continuous time Markov chains and provides more concrete theoretical results for the leading coefficients of the stationary distribution series expansion and the orders of the poles of the mean first passage times. For the leading coefficients in the series expansion for the mean first passage times, we use in part the results of Avrachenkov and Haviv [7] and Avrachenkov et al. [6] and adapt their work to the continuous time Markov chain setting. We treat in detail the case where the chain for $\varepsilon = 0$ has more than one absorbing state and at least one transient state. Furthermore, we also provide an interpretation of leading coefficients in the series expansion of the stationary distribution in terms of a certain restricted Markov chain. An algorithm we give to determine the order of the pole of the mean first passage time extends the work of Hassin and Haviv [22] from discrete time to continuous time. We also extend the original algorithm's scope to treat mean first passage times to a subset of states, instead of just a single state. Beltrán and Landim [8, 9] study metastable and tunneling behavior for a sequence $\{\eta^N\}_{N=1}^\infty$ of time-homogeneous continuous time Markov chains with countable state spaces. Under an acceleration of time by a factor θ_N , they give conditions under which the trace of the accelerated process on the metastates is asymptotically Markovian as $N \rightarrow \infty$. For our case, this would correspond to accelerating time for $\eta^N = X^\varepsilon$ by $\theta_N \approx \frac{1}{\varepsilon}$. Beltrán and Landim identified the transition rates for the limiting Markov chain and proved that its stationary distribution can be obtained as a limit from the stationary distribution for η^N . While this work is potentially related to what we did, it requires knowing the stationary distribution for η^N *a priori*. Our approach does not need to know that stationary distribution explicitly and we also study mean first passage times, giving explicit asymptotics for both. Finally, Yin and Zhang [39] conducted an extensive study focused on determining matched asymptotic expansions for the marginal distributions at time t of singularly perturbed continuous time Markov chains. Their infinitesimal generators, generalizing those of Phillips and Kokotovic [36] and Pan

and Basar [35], are of the form $Q(\varepsilon) = \frac{1}{\varepsilon}Q^{(0)} + Q^{(1)}$, and can be time dependent. For the time independent case, this would correspond to studying the marginal distributions of our Markov chain X^ε in the "linear" case and at time $\frac{t}{\varepsilon}$ as $\varepsilon \rightarrow 0$, i.e., $\lim_{\varepsilon \rightarrow 0} X^\varepsilon(\frac{t}{\varepsilon})$. Thus, while their work potentially might provide information about stationary distributions as $\varepsilon \rightarrow 0$, we directly study the power series expansion (in ε) of the stationary distribution of X^ε , and we also study series expansions of mean first passage times for X^ε , and we develop more concrete analyses for both.

3.2.4 Outline

In Section 3.3 we introduce two simplified models for the chromatin modification circuit that do not include DNA methylation. Through these examples, we introduce the mathematical setting and questions we address in this work. We describe the basic setup and definitions needed for this work in Section 3.4. We present our main results in Section 3.5–3.7. Some proofs are given there, whilst others are in SI-Section 3.10. Further applications of the theoretical tools developed in Section 3.5–3.7 for chromatin modification circuits that include DNA methylation are presented in Section 3.8. Concluding remarks are given in Section 3.9.

3.2.5 Preliminaries and notation

Denote the set of integers by $\mathbb{Z}$. For an integer $d \geq 2$ we denote by $\mathbb{Z}^d$ the set of d -dimensional vectors with entries in $\mathbb{Z}$. Denote by $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$, the set of non-negative integers. For an integer $d \geq 2$ we denote by $\mathbb{Z}_+^d$ the set of d -dimensional vectors with entries in $\mathbb{Z}_+$. We denote by $\mathbf{1}$ a vector of any dimension where all entries are 1's. The size of $\mathbf{1}$ will be understood from the context. The set of real numbers will be denoted by $\mathbb{R}$, $\mathbb{R}_+ = [0, \infty)$, $\mathbb{R}_{>0} = (0, \infty)$, and d -dimensional Euclidean space will be denoted by $\mathbb{R}^d$ for $d \geq 2$. For integers $n, m \geq 1$, the set of $n \times m$ matrices with real-valued entries will be denoted by $\mathbb{R}^{n \times m}$. The set of complex numbers will be denoted by $\mathbb{C}$.

Let $\mathcal{X}$ be a finite set. If needed, we will enumerate the entries of $\mathcal{X}$ by $\{1, \dots, |\mathcal{X}|\}$. For a matrix $A = (A_{x,y})_{x,y \in \mathcal{X}}$ with real-valued entries, we denote the kernel of A by $\ker(A) := \{x \in \mathbb{R}^{|\mathcal{X}|} : Ax = 0\}$ and the nullity of A by $\text{nullity}(A) := \dim(\ker(A))$. We denote the spectrum of A by $\text{sp}(A)$ and the spectral radius by $\text{spr}(A) = \max\{|\lambda| : \lambda \in \text{sp}(A)\}$. A matrix $Q = (Q_{x,y})_{x,y \in \mathcal{X}}$ will be called a Q -matrix if $Q_{x,y} \geq 0$ for every $x \neq y \in \mathcal{X}$ and $Q\mathbf{1} = 0$. We denote the identity matrix, which has 1's on the diagonal and zeros elsewhere, by $I = (I_{x,y})_{x,y \in \mathcal{X}}$. For a vector $v = (v_x)_{x \in \mathcal{X}}$ we denote by $\text{diag}((v_x)_{x \in \mathcal{X}})$ the diagonal matrix in $\mathcal{X}$ with entries given by v . Vectors are column vectors unless indicated otherwise and a superscript of T will denote the transpose of a vector or matrix. For integers $n, m \geq 1$ and a matrix $A \in \mathbb{R}^{n \times m}$, we denote by $\|A\| = (\sum_{i=1}^n \sum_{j=1}^m |A_{i,j}|^2)^{1/2}$ the Frobenius norm of A ¹. For a vector $v \in \mathbb{R}^n$, we denote the Euclidean norm of v by $\|v\| = (\sum_{i=1}^n |v_i|^2)^{1/2}$.

Definition 3.2.1. Given a matrix $A^{(0)}$ in $\mathbb{R}^{n \times m}$, a **real-analytic perturbation** of $A^{(0)}$ is a matrix-valued function $A : [0, \varepsilon_0) \rightarrow \mathbb{R}^{n \times m}$, where $\varepsilon_0 > 0$, and

$$A(\varepsilon) = \sum_{k=0}^{\infty} \varepsilon^k A^{(k)}, \quad 0 \leq \varepsilon < \varepsilon_0,$$

in which $\{A^{(k)} : k \geq 0\}$ is a sequence of matrices in $\mathbb{R}^{n \times m}$ such that

$$\sum_{k=0}^{\infty} \varepsilon^k \|A^{(k)}\| < \infty, \quad \text{for every } 0 \leq \varepsilon < \varepsilon_0. \quad (3.2.1)$$

Such a perturbation is called **linear** if $A(\varepsilon) = A^{(0)} + \varepsilon A^{(1)}$ for $0 \leq \varepsilon < \varepsilon_0$.

By (3.2.1), a real-analytic perturbation of $A^{(0)}$ can be extended to a function $F(z) := \sum_{k=0}^{\infty} z^k A^{(k)}$ defined on $B(0, \varepsilon_0) = \{z \in \mathbb{C} : |z| < \varepsilon_0\}$. The function F will be called an **analytic perturbation** or **complex-analytic perturbation** of $A^{(0)}$. This

¹Here, we chose to fix a particular norm on $\mathbb{R}^{n \times m}$, although other choices of norm will often work.

extension will allow us to invoke results in complex analysis in order to study real-analytic perturbations. An example of this is the following result.

Proposition 3.2.1. *Let $A : [0, \varepsilon_0) \rightarrow \mathbb{R}^{n \times n}$ be a real-analytic perturbation of $A^{(0)}$ such that $A^{-1}(\varepsilon)$ exists for every $0 < \varepsilon < \varepsilon_0$. Then, there is $\varepsilon_1 \in (0, \varepsilon_0)$ and $p \in \mathbb{Z}_+$ such that*

$$A^{-1}(\varepsilon) = \sum_{k=-p}^{\infty} \varepsilon^k B^{(k)}, \quad 0 < \varepsilon < \varepsilon_1, \quad (3.2.2)$$

where $\sum_{k=-p}^{\infty} \varepsilon^k \|B^{(k)}\| < \infty$ for every $0 < \varepsilon < \varepsilon_1$, $\{B^{(k)} : k \geq -p\}$ is a sequence of matrices in $\mathbb{R}^{n \times n}$, $B^{(-p)}$ is not the identically zero matrix and p is called the **order of the pole** at $\varepsilon = 0$.

This result is given in the analytic setting as Theorem 2.4 in Avrachenkov et al. [6]. Proposition 3.2.1 follows by extending $A(\cdot)$ to a complex disk, then using Cramer's rule as in the proof of Theorem 2.4 in Avrachenkov et al. [6] and checking that the matrices $\{B^{(k)} : k \geq -p\}$ obtained are real-valued.

3.3 Motivating Example: Chromatin Modification Circuit

In order to understand how the interactions among known chromatin modifications influence epigenetic cell memory, we consider the chemical reaction model of the gene's inner chromatin modification circuit introduced in Bruno et al. [12]. This model has the nucleosome with DNA wrapped around it, denoted by D , as a basic unit that can be modified either with activating marks, such as H3K4 methylation (H3K4me3) or H3K4 acetylation (H3K4ac), or repressive marks, such as H3K9 methylation (H3K9me3) or DNA methylation. H3K4me3 and H3K4ac are two histone modifications that promote a less compact DNA around the nucleosomes and they are then associated with gene activation (see for example Chapter 3 in Allis et al. [1], and Zhang et al. [40]). In

the model, it is assumed that H3K4me3 and H3K4ac co-exist and the nucleosome with either of these modifications is represented by D^A . On the contrary, both the histone modification H3K9me3 and DNA methylation cause the DNA to be tightly wrapped around the nucleosome and therefore, they are associated with gene repression (see for example Huang et al. [25]). A nucleosome with DNA methylation only, H3K9 methylation (H3K9me3) only or both is represented by D_1^R , D_2^R and D_{12}^R , respectively.

One of the key parameters of the system is $\varepsilon > 0$, a non-dimensional parameter that scales the speed of basal erasure of all chromatin modifications. We are interested in studying the behavior of the system in the limiting regime $\varepsilon \rightarrow 0$, in which the chromatin modification system has a bimodal limiting stationary distribution (see Bruno et al. [12]). One peak corresponds to the active chromatin state (most of the nucleosomes are modified with activating marks) and the other one is in the repressed chromatin state (most of the nucleosomes are modified with repressive marks). We aim to derive formulas that characterize, as ε goes to 0, the behavior of the stationary distribution and the “time to memory loss” of the active (repressed) state, defined as the mean first passage time to reach the repressed (active) state, starting from the active (repressed) state.

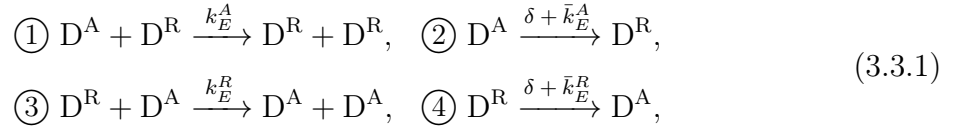
Two other critical parameters of the system are μ and μ' : they capture the asymmetry between the erasure rates of repressive and activating chromatin modifications. More precisely, μ (μ') quantifies the asymmetry between erasure rates of repressive histone modifications (DNA methylation) and activating histone modifications. Part of our study is to analytically determine how μ and μ' affect the stationary distribution and the time to memory loss of the active and repressed states.

In this section, we introduce two simplified models of the chromatin modification circuit in which, compared to the full model described above, DNA methylation is not included and the only chromatin marks are histone modifications. We will use these simpler models in Section 3.5–3.7 to directly apply and then better understand the theory developed in this work. Then, in Section 3.8 we deal with more elaborate models that also

include DNA methylation. Note that, for consistency, we use the same notation for the species and the reaction rate constants as the one used in the paper where these models were introduced in Bruno et al. [12].

3.3.1 1D model

We first consider a simplified model in which a gene has a total of $D_{\text{tot}} \geq 2$ nucleosomes, where each nucleosome either has an activating histone modification, D^A , or a repressive histone modification, D^R , and there are no unmodified nucleosomes in this simplified model. If the amounts of nucleosomes having repressive (D^R) and activating (D^A) modifications are denoted as n_{D^R} and n_{D^A} , respectively, then we have the conservation law $n_{D^R} + n_{D^A} = D_{\text{tot}}$. We call this the 1D model because it suffices to keep track of the amount of D^R (for example), since the amount of D^A can be deduced by the conservation law. Furthermore, the basal and recruited erasure of D^A (D^R) coincide with the basal *de-novo* establishment and maintenance of D^R (D^A). The chemical reaction system for this 1D model, shown in Figure 3.1(a), is the following:



where $\delta, k_E^A, \bar{k}_E^A, k_E^R, \bar{k}_E^R > 0$. Here, the form of the reaction rate constants is due to the fact that reactions with the same reactants and products have been combined. We denote the reaction volume by V , and let $\varepsilon := \frac{\delta + \bar{k}_E^A}{k_E^A(D_{\text{tot}}/V)} = \frac{\delta_A}{\bar{k}_E^A(D_{\text{tot}}/V)}$, where $\delta_A := \delta + \bar{k}_E^A$. We also consider the constant $\mu := \frac{k_E^R}{k_E^A}$, which captures the asymmetry between the erasure rates of repressive and activating histone modifications. We introduce the constant b such that $\mu b = \frac{\delta_R}{\delta_A}$, with $\delta_R := \delta + \bar{k}_E^R$. Then, $\delta_A = \varepsilon \frac{k_E^A D_{\text{tot}}}{V}$ and $\delta_R := \delta_A \mu b = \varepsilon \frac{k_E^R D_{\text{tot}}}{V} \mu b$. So, as $\varepsilon \rightarrow 0$, both δ_A and δ_R go to 0, with $D_{\text{tot}}, \frac{k_E^A}{V}, \mu$ and b fixed.

Now, consider a continuous time Markov chain X^ε , with state space $\mathcal{X} = \{0, \dots, D_{\text{tot}}\}$,

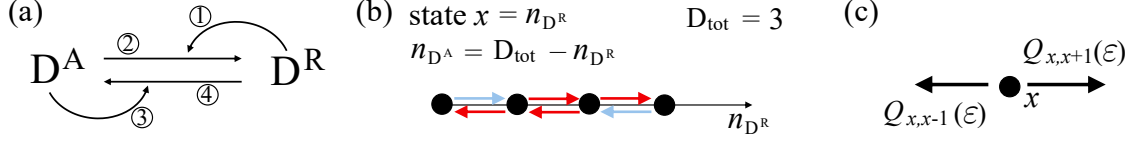


Figure 3.1. 1D model. (a) Chemical reaction system. The numbers on the arrows correspond to the chemical reactions associated with the arrows as described in (3.3.1) in the main text. (b) Markov chain graph. Here, we consider $D_{\text{tot}} = 3$ and we use black dots to represent the states, red arrows to represent transitions that are $O(1)$ and blue arrows to represent transitions that are $O(\varepsilon)$. (c) Direction of the potential transitions of X^ε starting from a state x , whose rates are given in (3.3.2).

where $D_{\text{tot}} \geq 2$ is an integer, which keeps track of n_{DR} through time. Given that we have the conservation law $n_{\text{DR}} + n_{\text{DA}} = D_{\text{tot}}$, n_{DA} can be obtained as a function of n_{DR} , that is $n_{\text{DA}} = D_{\text{tot}} - n_{\text{DR}}$. Assuming stochastic mass-action kinetics (including the usual volume scaling of rate constants, see for example Del Vecchio et al. [17]), the infinitesimal generator $Q(\varepsilon)$ ² for X^ε is given by:

$$Q_{x,x+\ell}(\varepsilon) = \begin{cases} \left(\frac{k_E^A}{V}x + \varepsilon \frac{k_E^A}{V}D_{\text{tot}} \right) (D_{\text{tot}} - x) & \text{if } \ell = 1 \\ \mu \left(\frac{k_E^A}{V}(D_{\text{tot}} - x) + b\varepsilon \frac{k_E^A}{V}D_{\text{tot}} \right) x & \text{if } \ell = -1 \\ 0 & \text{otherwise,} \end{cases} \quad (3.3.2)$$

for $x \in \mathcal{X}$, $\ell \in \mathbb{Z} \setminus \{0\}$ and $x + \ell \in \mathcal{X}$, and $Q_{x,x}(\varepsilon) = -\sum_{y \in \mathcal{X} \setminus \{x\}} Q_{x,y}(\varepsilon)$ for $x \in \mathcal{X}$. We extend this definition to $\varepsilon = 0$ by defining $Q_{x,y}(0) := \lim_{\varepsilon \rightarrow 0} Q_{x,y}(\varepsilon)$ for $x, y \in \mathcal{X}$. We will follow a similar convention for other examples. A representation of the Markov chain graph for $D_{\text{tot}} = 3$ and the possible transitions, with associated rates, are given in Figure 3.1(b)–(c). We consider X^0 to be the continuous time Markov chain with infinitesimal generator given by $Q(0)$. The process X^0 corresponds to a SCRN model associated with

²Note that $Q(\varepsilon)$ is sometimes called an infinitesimal transition matrix. The entries $Q_{x,y}(\varepsilon)$ for $x \neq y$ are the infinitesimal transition rates of going from x to y : $\mathbb{P}[X^\varepsilon(t+h) = y | X^\varepsilon(t) = x] = Q_{x,y}(\varepsilon)h + o(h)$ as $h \rightarrow 0$.

the autocatalytic reactions ① and ③ in (3.3.1), alone. Note that

$$Q(\varepsilon) = Q^{(0)} + \varepsilon Q^{(1)}, \quad \varepsilon \geq 0, \quad (3.3.3)$$

for appropriate matrices $Q^{(0)}$ and $Q^{(1)}$ in $\mathbb{R}^{|\mathcal{X}| \times |\mathcal{X}|}$. By (3.3.3), we can see that $Q(\cdot)$ is a *real-analytic* (and moreover *linear*) perturbation of $Q^{(0)}$ (see Section 3.2.5 for definitions). Note that for every $\varepsilon > 0$, X^ε is irreducible, while X^0 has a transient communicating class $\{1, \dots, D_{\text{tot}} - 1\}$ and two absorbing states (0 and D_{tot}) (see SI-Section 3.10.8). Because of this discontinuity at $\varepsilon = 0$, we say that $Q(\cdot)$ is a *singular perturbation* of $Q^{(0)}$ (see Section 3.4.1 for a precise definition).

We first want to determine the probability for the gene to be in the active state a ($x = 0$), repressed state r ($x = D_{\text{tot}}$) or one of the intermediate states ($x \in \{1, \dots, D_{\text{tot}} - 1\}$) after a long time (life-time of the organism), as a function of ε . We are especially interested in the limit of the stationary distribution for the system, $\pi(\varepsilon)$, as $\varepsilon \rightarrow 0$ (i.e., the basal erasure rate of the chromatin modifications is much lower than their maintenance rate). Since X^ε is irreducible for $\varepsilon > 0$ (and it has a finite state space), it has a unique stationary distribution $\pi(\varepsilon)$. In Section 3.4.1, we show that $\pi(0) := \lim_{\varepsilon \rightarrow 0} \pi(\varepsilon)$ exists and the function $\pi(\cdot)$ admits a convergent power series expansion:

$$\pi(\varepsilon) = \sum_{k=0}^{\infty} \varepsilon^k \pi^{(k)} \quad \text{for } 0 \leq \varepsilon < \varepsilon_1, \quad (3.3.4)$$

for some $\varepsilon_1 > 0$. In order to determine $\pi(0)$, we can take limits and observe that $\pi(0)Q(0) = 0$ and so $\pi(0)$ is a stationary distribution for $Q(0)$. Indeed, $\pi(0)$ is a specific mixture of atoms on the two absorbing states (0 and D_{tot}) for X^0 .

In Figure 3.2, we see how the function $\pi(\varepsilon)$ changes as $\varepsilon \rightarrow 0$ for several values of μ with D_{tot} , $\frac{k_E^A}{V}$ and b fixed. Furthermore, for this simpler chromatin modification circuit, because of the birth-death structure of X^ε , we can obtain explicit formulas for $\pi(\varepsilon)$ when

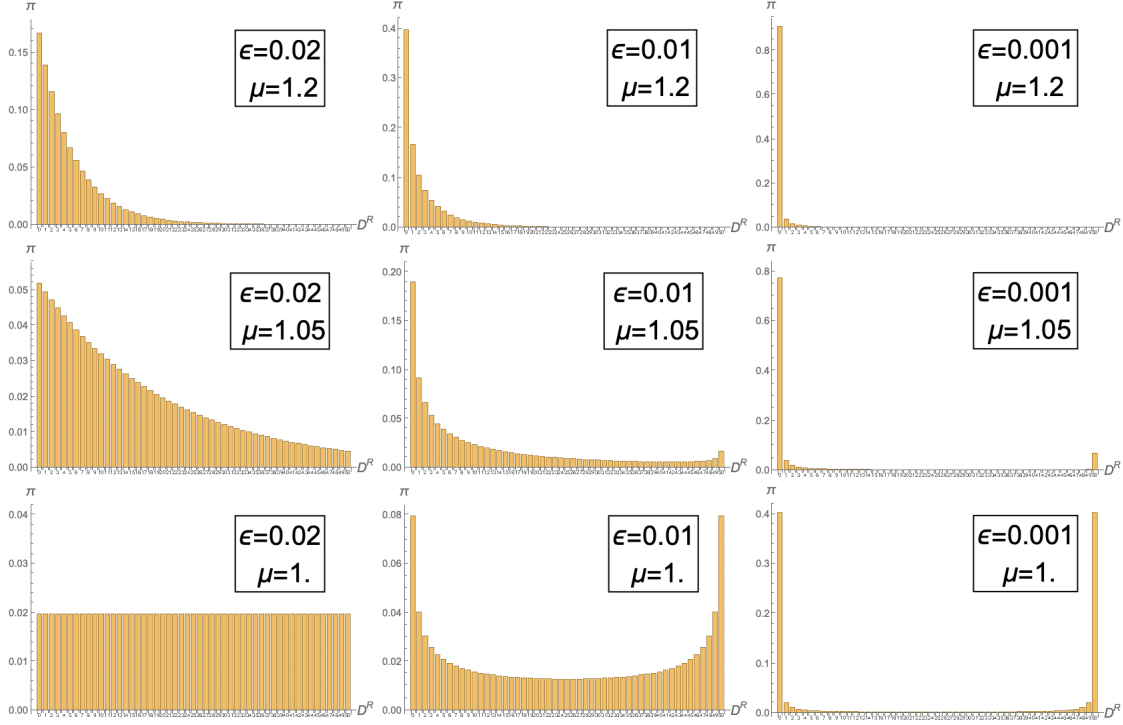


Figure 3.2. Histograms for the stationary distribution $\pi(\varepsilon)$ of the Markov chain X^ε defined by (3.3.2), for different values of ε and μ . The plot was generated by numerically solving $\pi(\varepsilon)Q(\varepsilon) = 0$ using the *Eigenvector* function in Mathematica. The parameters used were $D_{\text{tot}} = 50$, $k_E^A/V = 1$, and $b = 1$.

$\varepsilon > 0$ (see SI-Section 3.10.8). On letting $\varepsilon \rightarrow 0$, we obtain:

$$\pi_x(0) = \begin{cases} \frac{b\mu^{D_{\text{tot}}}}{1+b\mu^{D_{\text{tot}}}} & \text{if } x = 0 \\ 0 & \text{if } x \in \{1, \dots, D_{\text{tot}} - 1\} \\ \frac{1}{1+b\mu^{D_{\text{tot}}}} & \text{if } x = D_{\text{tot}}. \end{cases} \quad (3.3.5)$$

Thus, $\pi_x(0) \neq 0$ only for $x = 0$ and $x = D_{\text{tot}}$ and $\pi_0(0)$ increases as μ increases, while $\pi_{D_{\text{tot}}}(0)$ decreases as μ increases.

For continuous time Markov chains beyond the one-dimensional birth-death processes seen here, determining $\pi(0)$ will be a considerable task. In Section 3.5, we address the problem of determining $\pi(0)$, together with the whole expansion (3.3.4), in a system-

atic way, for a class of singularly perturbed Markov chains that includes our models of chromatin modification circuits. For the 1D model considered here, the derivation of the first two terms in the expansion is given in Section 3.5.2.

Now, in order to evaluate the time to memory loss of the active and repressed states, let us define the first passage time as $\tau_y^\varepsilon = \inf\{t \geq 0 : X^\varepsilon(t) = y\}$ for a state $y \in \mathcal{X}$. We will see in (3.4.2) that the mean first passage time (MFPT) for X^ε starting from $x \in \mathcal{X}$, $h_{x,y}(\varepsilon) = \mathbb{E}_x[\tau_y^\varepsilon]$, has a Laurent series expansion of the form:

$$h_{x,y}(\varepsilon) = \frac{c_{-p}}{\varepsilon^p} + \dots + \frac{c_{-1}}{\varepsilon} + c_0 + \varepsilon c_1 + \dots \quad \text{for } 0 < \varepsilon < \varepsilon_{\{y\}},$$

for some $\varepsilon_{\{y\}} > 0$, for some $p \in \mathbb{Z}_+$ (depending on x, y) and where $c_{-p} > 0$. Then, considering the repressed state $r = D_{\text{tot}}$ and the active state $a = 0$, we define the time to memory loss of the repressed state as $h_{r,a}(\varepsilon)$ and the time to memory loss of the active state as $h_{a,r}(\varepsilon)$. Now, we are interested in the derivation of analytical formulas for $h_{r,a}(\varepsilon)$ and $h_{a,r}(\varepsilon)$. This will allow us to understand how the time to memory loss changes as $\varepsilon \rightarrow 0$, as well as how the asymmetry of the system (represented by μ) affects this limit. For this case study, exploiting its birth-death structure, we can directly derive relevant formulas (see SI-Section 3.10.8 and SI-Equations (3.10.55)–(3.10.56)). In particular, defining $\lambda_x^\varepsilon = Q_{x,x+1}(\varepsilon)$, $\gamma_x^\varepsilon = Q_{x,x-1}(\varepsilon)$, with $Q_{x,x+1}(\varepsilon)$ and $Q_{x,x-1}(\varepsilon)$ defined in (3.3.2), and $r_j^\varepsilon = \frac{\lambda_1^\varepsilon \lambda_2^\varepsilon \dots \lambda_j^\varepsilon}{\gamma_1^\varepsilon \gamma_2^\varepsilon \dots \gamma_j^\varepsilon}$, for $j = 1, 2, \dots, D_{\text{tot}} - 1$, the time to memory loss of the repressed state is given by

$$h_{r,a}(\varepsilon) = \frac{r_{D_{\text{tot}}-1}^\varepsilon}{\gamma_{D_{\text{tot}}}^\varepsilon} \left(1 + \sum_{i=1}^{D_{\text{tot}}-1} \frac{1}{r_i^\varepsilon} \right) + \sum_{i=2}^{D_{\text{tot}}-1} \left[\frac{r_{i-1}^\varepsilon}{\gamma_i^\varepsilon} \left(1 + \sum_{j=1}^{i-1} \frac{1}{r_j^\varepsilon} \right) \right] + \frac{1}{\gamma_1^\varepsilon}. \quad (3.3.6)$$

Similarly, defining $\tilde{r}_j^\varepsilon = \frac{\gamma_{D_{\text{tot}}-1}^\varepsilon \gamma_{D_{\text{tot}}-2}^\varepsilon \dots \gamma_{D_{\text{tot}}-j}^\varepsilon}{\lambda_{D_{\text{tot}}-1}^\varepsilon \lambda_{D_{\text{tot}}-2}^\varepsilon \dots \lambda_{D_{\text{tot}}-j}^\varepsilon}$, for $j = 1, 2, \dots, D_{\text{tot}} - 1$, the time to memory

loss of the active state is given by

$$h_{a,r}(\varepsilon) = \frac{\tilde{r}_{D_{\text{tot}}-1}^\varepsilon}{\lambda_0^\varepsilon} \left(1 + \sum_{j=1}^{D_{\text{tot}}-1} \frac{1}{\tilde{r}_j^\varepsilon} \right) + \sum_{i=2}^{D_{\text{tot}}-1} \left[\frac{\tilde{r}_{i-1}^\varepsilon}{\lambda_{D_{\text{tot}}-i}^\varepsilon} \left(1 + \sum_{j=1}^{i-1} \frac{1}{\tilde{r}_j^\varepsilon} \right) \right] + \frac{1}{\lambda_{D_{\text{tot}}-1}^\varepsilon}. \quad (3.3.7)$$

Since λ_0^ε and $\gamma_{D_{\text{tot}}}^\varepsilon$ are the only transition rates that are $O(\varepsilon)$ with the rest being $O(1)$, the time to memory loss of both the active and repressed states are $O(\varepsilon^{-1})$, that is, $p = 1$, and as $\varepsilon \rightarrow 0$, these mean times tend to infinity.

Furthermore, γ_x^ε , with $x \in \{1, 2, \dots, D_{\text{tot}}\}$, are the only rates that depend on μ (they are linear in μ). Examining (3.3.6)–(3.3.7) with this observation in mind, we see that, if μ is increased (that is, the erasure rate of the repressive histone modification is increased compared to that of the active histone modification), $h_{a,r}(\varepsilon)$ increases, while $h_{r,a}(\varepsilon)$ decreases. The opposite happens when μ is decreased.

More complicated situations arise when we do not have a birth-death structure to work with, as in the model of the next example. To evaluate how critical system parameters affect the time to memory loss for such more elaborate systems, we develop an algorithm to determine p (see Section 3.6.1), we give an expression for the leading term in the series expansion of the mean first passage time (see Section 3.6.2), and we exploit theoretical results developed in Campos et al. [14] for comparing continuous time Markov chains, to determine how the asymmetry of the system affects the time to memory loss (see Section 3.7).

3.3.2 2D model

Let us consider a model in which, compared to the previous one, we assume that a nucleosome can also be unmodified. More precisely, in this case we denote the number of nucleosomes unmodified (D), modified with repressive modifications (D^R), and modified with activating modifications (D^A) by n_D , n_{D^R} and n_{D^A} , respectively, and we have that $n_D + n_{D^R} + n_{D^A} = D_{\text{tot}}$, with D_{tot} representing the total number of nucleosomes within

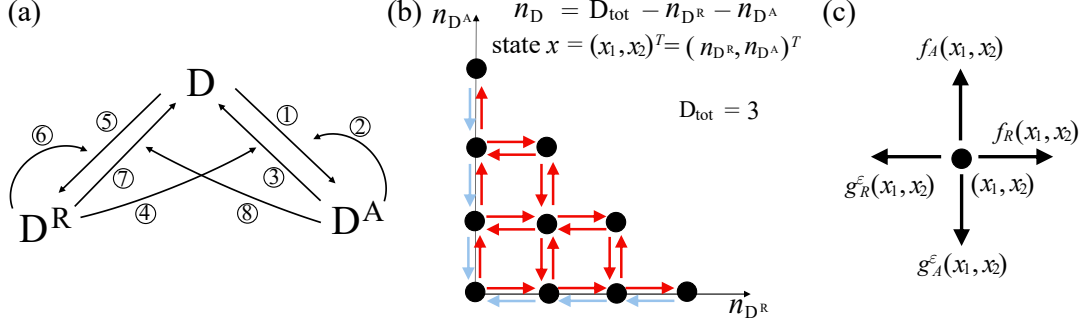
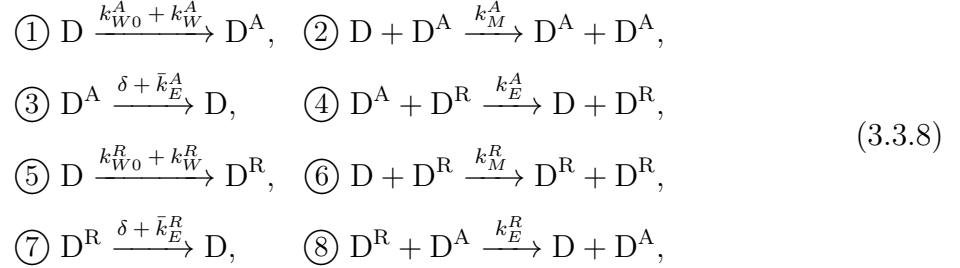


Figure 3.3. 2D model. (a) Chemical reaction system. The numbers on the arrows correspond to the reactions associated with the arrows as described in (3.3.8) in the main text. (b) Markov chain graph. Here, we consider $D_{\text{tot}} = 3$ and we use black dots to represent the states, red arrows to represent transition rates that are $O(1)$, and blue arrows to represent transition rates that are $O(\varepsilon)$. (c) Directions of the possible one step transitions for X^ε starting from a state $x = (x_1, x_2)^T$, whose rates are given in (3.3.9).

the gene. Furthermore, each histone modification autocatalyzes its own production and promotes the erasure of the other one (see for example Dodd et al. [18] and Bruno et al. [12]). The chemical reaction system for the 2D model, shown in Figure 3.3(a), is the following:



where $k_{W0}^A, k_W^A, k_M^A, \delta, \bar{k}_E^A, k_E^A, k_{W0}^R, k_W^R, k_M^R, \bar{k}_E^R, k_E^R > 0$. Here, the form of the reaction rate constants is due to the fact that reactions with the same reactants and products have been combined. Now, similar to what we did for the previous model, let us denote the reaction volume by V , and let $\varepsilon := \frac{\delta + \bar{k}_E^A}{k_M^A(D_{\text{tot}}/V)} = \frac{\delta_A}{k_M^A(D_{\text{tot}}/V)}$, with $\delta_A := \delta + \bar{k}_E^A$, and $\mu := \frac{k_E^R}{k_E^A}$. Additionally, consider the constant b such that $\mu b = \frac{\delta_R}{\delta_A}$, with $\delta_R := \delta + \bar{k}_E^R$. Then $\delta_R = \delta_A \mu b = \varepsilon \frac{k_M^A D_{\text{tot}}}{V} \mu b$. So, as $\varepsilon \rightarrow 0$, both δ_A and δ_R go to 0 with $D_{\text{tot}}, \frac{k_M^A D_{\text{tot}}}{V}, \mu$ and b fixed.

We consider the continuous time Markov chain $X^\varepsilon = \{(X_1^\varepsilon(t), X_2^\varepsilon(t))^T, t \geq 0\}$,

which keeps track of $(n_{\text{DR}}, n_{\text{DA}})$ through time. Since the total number of nucleosomes D_{tot} is constant, the state space is $\mathcal{X} = \{x = (x_1, x_2)^T \in \mathbb{Z}_+^2 : x_1 + x_2 \leq D_{\text{tot}}\}$. The potential one step transitions for X^ε from $x \in \mathcal{X}$ are shown in Figure 3.3(c), where the associated transition vectors are given by $v_1 = -v_2 = (0, 1)^T$ and $v_3 = -v_4 = (1, 0)^T$ and the infinitesimal transition rates (assuming mass-action kinetics with the usual volume scaling of rate constants) are given by

$$\begin{aligned}
Q_{x, x+v_1}(\varepsilon) &= f_A(x) = (D_{\text{tot}} - (x_1 + x_2)) \left(k_{W0}^A + k_W^A + \frac{k_M^A}{V} x_2 \right), \\
Q_{x, x+v_3}(\varepsilon) &= f_R(x) = (D_{\text{tot}} - (x_1 + x_2)) \left(k_{W0}^R + k_W^R + \frac{k_M^R}{V} x_1 \right), \\
Q_{x, x+v_2}(\varepsilon) &= g_A^\varepsilon(x) = x_2 \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} + x_1 \frac{k_E^A}{V} \right), \\
Q_{x, x+v_4}(\varepsilon) &= g_R^\varepsilon(x) = x_1 \mu \left(\varepsilon \frac{k_M^R}{V} D_{\text{tot}} + x_2 \frac{k_E^R}{V} \right).
\end{aligned} \tag{3.3.9}$$

This is a more complicated model compared to the previous example. In order to study its stationary distribution and mean first passage times, we will exploit the theory developed in this work, as shown in Section 3.5–3.7.

3.4 Basic Setup and Definitions

In Section 3.4.1 we provide basic definitions for singularly perturbed continuous time Markov chains and describe some key properties for them. In particular, we describe the form of series expansions for their stationary distributions and mean first passage times. We will study these quantities and apply our results to a class of continuous time Markov chains called Stochastic Chemical Reaction Networks (SCRNs) which are defined in Section 3.4.2. Our models of chromatin modification circuits will be SCRNs. All of the models considered will have a finite state space.

3.4.1 Singularly perturbed, finite state, continuous time Markov chains

Suppose $\mathcal{X}$ is a finite set and $|\mathcal{X}| > 1$. For a value $\varepsilon_0 > 0$, consider a family $\{X^\varepsilon : 0 \leq \varepsilon < \varepsilon_0\}$ of continuous time Markov chains with state space $\mathcal{X}$ and infinitesimal generators $\{Q(\varepsilon) : 0 \leq \varepsilon < \varepsilon_0\}$ where $\varepsilon \mapsto Q(\varepsilon)$ is a real-analytic perturbation of $Q(0)$. Thus,

$$Q(\varepsilon) = Q^{(0)} + \varepsilon Q^{(1)} + \varepsilon^2 Q^{(2)} + \dots,$$

where $\{Q^{(k)} : k \geq 0\}$ is a family of $|\mathcal{X}| \times |\mathcal{X}|$ real-valued matrices such that $\sum_{k=0}^{\infty} \varepsilon^k \|Q^{(k)}\| < \infty$ for every $0 \leq \varepsilon < \varepsilon_0$. Assume that the continuous time Markov chains X^ε are irreducible for $0 < \varepsilon < \varepsilon_0$. In this context, the perturbation is **singular** when X^0 has more than one recurrent class. This notion of singular will be the focus of our attention although some of our work applies for the regular (non-singular) case too. All of our chromatin modification circuit models have associated singular continuous time Markov chains, where the perturbation is **linear**, i.e., $Q^{(k)} = 0$ for every $k \geq 2$.

When $0 < \varepsilon < \varepsilon_0$, there is an equivalent characterization of X^ε using holding times with exponential parameters $\{q_x(\varepsilon)\}_{x \in \mathcal{X}}$ and a transition probability matrix $P(\varepsilon)$ for the **embedded discrete time Markov chain**. Specifically, for each $x \in \mathcal{X}$, $q_x(\varepsilon) = -Q_{x,x}(\varepsilon) \neq 0$, since X^ε is irreducible, and for all $x, y \in \mathcal{X}$, $P_{x,x}(\varepsilon) = 0$, $P_{x,y}(\varepsilon) = \frac{Q_{x,y}(\varepsilon)}{q_x(\varepsilon)}$, for $y \neq x$ in $\mathcal{X}$. Note that $Q(\varepsilon) = \text{diag}(q(\varepsilon))(P(\varepsilon) - I)$. The matrix $P(\varepsilon)$ has a power series expansion in ε for sufficiently small $0 \leq \varepsilon < \varepsilon_P$ for some $\varepsilon_P > 0$ (the justification is similar to that for (3.4.5) below).

The first quantities we are interested in studying are mean first passage times. Consider a nonempty set $\mathcal{B} \subseteq \mathcal{X}$ such that $\mathcal{B} \neq \mathcal{X}$ and let

$$\tau_B^\varepsilon := \inf\{t \geq 0 : X^\varepsilon(t) \in \mathcal{B}\}.$$

We define the **mean first passage time (MFPT)** (for X^ε) from $x \in \mathcal{X}$ to $\mathcal{B}$ as

$$h_{x,\mathcal{B}}(\varepsilon) = \mathbb{E}[\tau_{\mathcal{B}}^\varepsilon \mid X^\varepsilon(0) = x].$$

If $\mathcal{B} = \{y\}$ for some $y \in \mathcal{X}$, we adopt the notation: $h_{x,y}(\varepsilon) := h_{x,\{y\}}(\varepsilon)$. Using first step analysis (see for example Equation (3.1) in Norris [33]), for $0 < \varepsilon < \varepsilon_0$,

$$h_{x,\mathcal{B}}(\varepsilon) = \begin{cases} 0 & \text{if } x \in \mathcal{B} \\ \frac{1}{q_x(\varepsilon)} + \sum_{y \in \mathcal{X}} P_{x,y}(\varepsilon) h_{y,\mathcal{B}}(\varepsilon) & \text{if } x \in \mathcal{B}^c. \end{cases} \quad (3.4.1)$$

Now, define $P^{\mathcal{B}^c}(\varepsilon)$ and $Q^{\mathcal{B}^c}(\varepsilon)$ as the matrices obtained by removing the columns and rows of $P(\varepsilon)$ and $Q(\varepsilon)$, respectively, corresponding to states in $\mathcal{B}$. Then, by noting that $I - P^{\mathcal{B}^c}(\varepsilon)$ is invertible (see SI-Lemma 3.10.2) and that $Q^{\mathcal{B}^c}(\varepsilon) = -\text{diag}((q_x(\varepsilon))_{x \in \mathcal{B}^c})(I - P^{\mathcal{B}^c}(\varepsilon))$ is invertible, we obtain from (3.4.1) that

$$h_{\mathcal{B}}(\varepsilon) = -(Q^{\mathcal{B}^c}(\varepsilon))^{-1} \mathbf{1},$$

where $h_{\mathcal{B}}(\varepsilon) := (h_{x,\mathcal{B}}(\varepsilon))_{x \in \mathcal{B}^c}$, I is the identity matrix of dimension $|\mathcal{B}^c|$, and $\mathbf{1}$ is the vector of all 1's, of size $|\mathcal{B}^c|$. Proposition 3.2.1 yields that there is $0 < \varepsilon_{\mathcal{B}} < \varepsilon_0$ such that $-(Q^{\mathcal{B}^c}(\varepsilon))^{-1}$ can be expanded as a matrix-valued Laurent series as in (3.2.2) for $0 < \varepsilon < \varepsilon_{\mathcal{B}}$, and then for each $x \in \mathcal{B}^c$,

$$\mathbb{E}_x[\tau_{\mathcal{B}}^\varepsilon] = h_{x,\mathcal{B}}(\varepsilon) = \sum_{k=-p(x)}^{\infty} \rho_x^{(k)} \varepsilon^k, \quad 0 < \varepsilon < \varepsilon_{\mathcal{B}}, \quad (3.4.2)$$

where $p(x) \geq 0$ is an integer, $\rho_x^{(-p(x))} > 0$, $\rho_x^{(k)} \in \mathbb{R}$ for $k > -p(x)$, and the convergence is absolute convergence for $0 < \varepsilon < \varepsilon_{\mathcal{B}}$. The quantity $p(x)$ will be called the **order of the pole** of (3.4.2). In Section 3.6.1, we will show how to find $p(x)$ by using an algorithm that uses the order, with respect to ε , of the transitions of the Markov chain X^ε .

A second quantity of interest is the stationary distribution for X^ε . For $0 < \varepsilon < \varepsilon_0$, since X^ε is assumed to be irreducible and has finite state space, there is a unique **stationary distribution** $\pi(\varepsilon) = (\pi_x(\varepsilon))_{x \in \mathcal{X}}$, which is the unique probability row vector satisfying $\pi(\varepsilon)Q(\varepsilon) = 0$. We are interested in studying $\pi(\varepsilon)$ as $\varepsilon \rightarrow 0$. For this, first consider $\eta_x^\varepsilon = \inf\{t \geq 0 : X^\varepsilon(t) \neq x\}$ and $\zeta_x^\varepsilon = \inf\{t > \eta_x^\varepsilon : X^\varepsilon(t) = x\}$, $x \in \mathcal{X}$. Note that $\mathbb{E}_y[\zeta_x^\varepsilon] = h_{y,x}(\varepsilon)$ for $y \neq x$. For each $x \in \mathcal{X}$, $\mathbb{E}_x[\zeta_x^\varepsilon]$ is called the **mean return time** to the state x , and for $0 < \varepsilon < \varepsilon_0$ satisfies

$$\mathbb{E}_x[\zeta_x^\varepsilon] = \frac{1}{q_x(\varepsilon)} + \sum_{y \neq x} P_{xy}(\varepsilon) \mathbb{E}_y[\zeta_x^\varepsilon] = \frac{1}{q_x(\varepsilon)} + \sum_{y \neq x} P_{xy}(\varepsilon) h_{y,x}(\varepsilon). \quad (3.4.3)$$

It is well known (see for example Theorem 3.8.1 in Norris [33]) that for $0 < \varepsilon < \varepsilon_0$,

$$\pi_x(\varepsilon) = \frac{1}{\mathbb{E}_x[\zeta_x^\varepsilon]} \cdot \frac{1}{q_x(\varepsilon)}, \quad x \in \mathcal{X}. \quad (3.4.4)$$

From (3.4.2)–(3.4.3), we can see that $\varepsilon \mapsto q_x(\varepsilon) \mathbb{E}_x[\zeta_x^\varepsilon]$ can be extended to an analytic function on a punctured disk about 0 in $\mathbb{C}$, with a Laurent series expansion having at most a pole of finite order at 0. The radius of the punctured disk may be smaller than ε_0 . This, together with (3.4.4), implies that $\varepsilon \mapsto \pi_x(\varepsilon)$ can be extended to an analytic function on a punctured disk about 0 in $\mathbb{C}$, also with a Laurent series expansion. Since this function is bounded by one when restricted to sufficiently small positive values of ε , we can remove the singularity at 0 and obtain that $\pi(0) := \lim_{\varepsilon \rightarrow 0} \pi(\varepsilon)$ exists and furthermore $\varepsilon \mapsto \pi(\varepsilon)$ is a real-analytic perturbation of $\pi(0)$. In other words,

$$\pi(\varepsilon) = \sum_{k=0}^{\infty} \varepsilon^k \pi^{(k)}, \quad 0 \leq \varepsilon < \varepsilon_1, \quad (3.4.5)$$

for sufficiently small $\varepsilon_1 > 0$ and where $\{\pi^{(k)} : k \geq 0\}$ is a sequence of real-valued $|\mathcal{X}|$ -dimensional vectors such that $\sum_{k=0}^{\infty} \varepsilon^k \|\pi^{(k)}\| < \infty$ for every $0 \leq \varepsilon < \varepsilon_1$.

3.4.2 Stochastic Chemical Reaction Networks

In this section, we provide some background on Stochastic Chemical Reaction Networks. The reader is referred to Anderson and Kurtz [4] for a more in depth introduction to this subject.

We assume there is a finite non-empty set $\mathcal{S} = \{S_1, \dots, S_d\}$ of d **species**, and a finite non-empty set $\mathcal{R} \subseteq \mathbb{Z}_+^d \times \mathbb{Z}_+^d$ that represents chemical **reactions**. We assume that $(w, w) \notin \mathcal{R}$ for every $w \in \mathbb{Z}_+^d$. The set $\mathcal{S}$ represents d different molecular species in a system subject to reactions $\mathcal{R}$ which change the number of molecules of some species. For each $(v^-, v^+) \in \mathcal{R}$, the d -dimensional vector v^- (the **reactant vector**) counts how many molecules of each species are consumed in the reaction, while v^+ (the **product vector**) counts how many molecules of each species are produced. The reaction is usually written as

$$\sum_{i=1}^d (v^-)_i S_i \longrightarrow \sum_{i=1}^d (v^+)_i S_i.$$

To avoid the use of unnecessary species, we will assume that for each $1 \leq i \leq d$, there exists a vector $w = (w_1, \dots, w_d)^T \in \mathbb{Z}_+^d$ with $w_i > 0$ such that (w, v) or (v, w) is in $\mathcal{R}$ for some $v \in \mathbb{Z}_+^d$, i.e., each species is either a reactant or a product in some reaction.

The net change in the quantity of molecules of each species due to a reaction $(v^-, v^+) \in \mathcal{R}$ is described by $v^+ - v^-$ and it is called the associated **reaction vector**. We denote the set of reaction vectors by $\mathcal{V} := \{v \in \mathbb{Z}^d \mid v = v^+ - v^- \text{ for some } (v^-, v^+) \in \mathcal{R}\}$, we let $n := |\mathcal{V}|$ the size of $\mathcal{V}$ and we enumerate the members of $\mathcal{V}$ as $\{v_1, \dots, v_n\}$. Note that $\mathcal{V}$ does not contain the zero vector because $\mathcal{R}$ has no elements of the form (w, w) . Different reactions might have the same reaction vector. For each $v_j \in \mathcal{V}$ we consider the set $\mathcal{R}_{v_j} := \{(v^-, v^+) \in \mathcal{R} \mid v_j = v^+ - v^-\}$. The matrix $S \in \mathbb{R}^{d \times |\mathcal{R}|}$ whose columns are the elements $v^+ - v^-$ for $(v^-, v^+) \in \mathcal{R}$ will be called the **stoichiometric matrix**.

Given $(\mathcal{S}, \mathcal{R})$ we will consider an associated continuous time Markov chain $X = (X_1, \dots, X_d)^T$, with a state space $\mathcal{X}$ contained in $\mathbb{Z}_+^d$, which tracks the number of molecules

of each species over time. Roughly speaking, the dynamics of X will be given by the following: given a current state $x = (x_1, \dots, x_d)^T \in \mathcal{X} \subseteq \mathbb{Z}_+^d$, for each reaction $(v^-, v^+) \in \mathcal{R}$, there is a clock which will ring at an exponentially distributed time (with rate $\Lambda_{(v^-, v^+)}(x)$). The clocks for distinct reactions are independent of one another. If the clock corresponding to $(v^-, v^+) \in \mathcal{R}$ rings first, the system moves from x to $x + v^+ - v^-$ at that time, and then the process repeats. We now define the continuous time Markov chain in more detail.

Consider sets of species $\mathcal{S}$ and reactions $\mathcal{R}$, a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$ and a collection of functions $\Lambda = \{\Lambda_{(v^-, v^+)} : \mathcal{X} \rightarrow \mathbb{R}_+\}_{(v^-, v^+) \in \mathcal{R}}$ such that for each $x \in \mathcal{X}$ and $(v^-, v^+) \in \mathcal{R}$, if $x + v^+ - v^- \notin \mathcal{X}$, then $\Lambda_{(v^-, v^+)}(x) = 0$. Now, for $1 \leq j \leq n$, define

$$\Upsilon_j(x) := \sum_{(v^-, v^+) \in \mathcal{R}_{v_j}} \Lambda_{(v^-, v^+)}(x).$$

Note that for each $x \in \mathcal{X}$ and $1 \leq j \leq n$, if $x + v_j \notin \mathcal{X}$, then $\Upsilon_j(x) = 0$. The functions $\{\Lambda_{(v^-, v^+)} : \mathcal{X} \rightarrow \mathbb{R}_+\}_{(v^-, v^+) \in \mathcal{R}}$ are called **propensity** or **intensity** functions. A common form for the propensity functions is the following, which is associated with **mass action kinetics**:

$$\Lambda_{(v^-, v^+)}(x) = \kappa_{(v^-, v^+)} \prod_{i=1}^d (x_i)_{(v^-)_i},$$

where $\{\kappa_{(v^-, v^+)}\}_{(v^-, v^+) \in \mathcal{R}}$ are non-negative constants and for $m, \ell \in \mathbb{Z}_+$, the quantity $(m)_\ell$ is the falling factorial, i.e., $(m)_0 := 1$ and $(m)_\ell := m(m-1) \dots (m-\ell+1)$. A **stochastic chemical reaction network (SCRN)** (associated with $(\mathcal{S}, \mathcal{R}, \mathcal{X}, \Lambda)$) is a continuous time Markov chain X with state space $\mathcal{X}$ and infinitesimal generator Q given for $x, y \in \mathcal{X}$ by

$$Q_{x,y} = \begin{cases} \Upsilon_j(x) & \text{if } y - x = v_j \text{ for some } 1 \leq j \leq n, \\ -\sum_{j=1}^n \Upsilon_j(x) & \text{if } y = x, \\ 0 & \text{otherwise.} \end{cases} \quad (3.4.6)$$

A SCRN associated with $(\mathcal{S}, \mathcal{R}, \mathcal{X}, \Lambda)$ is said to satisfy a **conservation law** if there is a d -dimensional non-zero vector m such that $m^T S = 0$, and hence $m^T X(t) = x_{\text{tot}}$ for every $t \geq 0$, for some constant x_{tot} . Consequently, we can reduce the dimension of the continuous time Markov chain describing the system by one. For example, if $m = (1, \dots, 1)^T$, then the projected process $(X_1, \dots, X_{d-1})^T$ is again a continuous time Markov chain with state space $\{(x_1, \dots, x_{d-1})^T \in \mathbb{Z}_+^{d-1} \mid (x_1, \dots, x_{d-1}, x_{\text{tot}} - \sum_{i=1}^{d-1} x_i)^T \in \mathcal{X}\}$. In our examples, we will often use this type of reduction.

3.5 Main Results: Stationary Distributions

This section focuses on characterizing the terms in the series expansion (3.4.5). In Section 3.5.1, we focus on determining the term $\pi^{(0)} = \pi(0)$, while in SI-Section 3.10.2 we provide a result which enables computation of all of the higher order terms $\pi^{(k)}$ for $k > 0$, under additional assumptions. In Section 3.5.2, we apply these results to the examples introduced in Section 3.3. Further examples for higher dimensional models of the chromatin modification circuits will be given in Section 3.8. We remind the reader that to ease notation, we have adopted the convention that stationary distribution vectors will be row vectors, even though we do not use the transpose notation T to indicate this.

3.5.1 The zeroth order term

As in Section 3.4.1, consider a family $\{X^\varepsilon : 0 \leq \varepsilon < \varepsilon_0\}$ of continuous time Markov chains on a finite state space $\mathcal{X}$, with infinitesimal generators $\{Q(\varepsilon) : 0 \leq \varepsilon < \varepsilon_0\}$ where $\varepsilon \mapsto Q(\varepsilon)$ is a real-analytic perturbation of $Q(0)$ with coefficients $\{Q^{(k)} : k \geq 0\}$ and additionally $Q(\varepsilon)$ is irreducible for every $0 < \varepsilon < \varepsilon_0$. The matrix $Q(0) = Q^{(0)}$ is a Q -matrix for which $\mathcal{X}$ decomposes into recurrent (or ergodic) states $\mathcal{A}$ and transient states $\mathcal{T}$. From now on, we assume the following.

Assumption 3.5.1. The set $\mathcal{A}$ consists of $|\mathcal{A}| \geq 1$ absorbing states for $Q(0)$, while $\mathcal{T}$

consists of $|\mathcal{T}| \geq 1$ transient states for $Q(0)$.

In other words, in the dynamics of $Q(0)$ there is at least one transient state, at least one recurrent state and all the recurrent states are absorbing. Now, we label the state space starting with the states in $\mathcal{A}$ and followed by the ones in $\mathcal{T}$. For every $k \geq 0$, we can write $Q^{(k)}$ as

$$Q^{(k)} = \left[\begin{array}{c|c} A_k & S_k \\ \hline R_k & T_k \end{array} \right], \quad (3.5.1)$$

where $A_k \in \mathbb{R}^{|\mathcal{A}| \times |\mathcal{A}|}$, $S_k \in \mathbb{R}^{|\mathcal{A}| \times |\mathcal{T}|}$, $R_k \in \mathbb{R}^{|\mathcal{T}| \times |\mathcal{A}|}$ and $T_k \in \mathbb{R}^{|\mathcal{T}| \times |\mathcal{T}|}$. In a similar fashion, we can write

$$Q(\varepsilon) = \left[\begin{array}{c|c} A(\varepsilon) & S(\varepsilon) \\ \hline R(\varepsilon) & T(\varepsilon) \end{array} \right], \quad (3.5.2)$$

for $0 \leq \varepsilon < \varepsilon_0$, where $A(\varepsilon) \in \mathbb{R}^{|\mathcal{A}| \times |\mathcal{A}|}$, $S(\varepsilon) \in \mathbb{R}^{|\mathcal{A}| \times |\mathcal{T}|}$, $R(\varepsilon) \in \mathbb{R}^{|\mathcal{T}| \times |\mathcal{A}|}$ and $T(\varepsilon) \in \mathbb{R}^{|\mathcal{T}| \times |\mathcal{T}|}$.

From Assumption 3.5.1, we obtain that

$$Q^{(0)} = Q(0) = \left[\begin{array}{c|c} 0 & 0 \\ \hline R_0 & T_0 \end{array} \right],$$

where T_0 is an invertible matrix (see SI-Lemma 3.10.4).

For each $0 < \varepsilon < \varepsilon_0$, we denote by $\pi(\varepsilon) = (\pi_x(\varepsilon))_{x \in \mathcal{X}}$ the stationary distribution for $Q(\varepsilon)$. In Section 3.4.1, we showed that the limit $\pi(0) := \lim_{\varepsilon \rightarrow 0} \pi(\varepsilon)$ exists and that $\varepsilon \mapsto \pi(\varepsilon)$ is a real-analytic perturbation of $\pi(0)$ with expansion given by (3.4.5) for $0 \leq \varepsilon < \varepsilon_1$. For convenience, decompose the row vector $\pi(\varepsilon)$ as $\pi(\varepsilon) = [\alpha(\varepsilon), \beta(\varepsilon)]$ for $0 \leq \varepsilon < \varepsilon_1$ where $\alpha(\varepsilon) \in \mathbb{R}^{|\mathcal{A}|}$ and $\beta(\varepsilon) \in \mathbb{R}^{|\mathcal{T}|}$. From (3.4.5), letting $\pi^{(k)} = [\alpha^{(k)}, \beta^{(k)}]$, we have

$$\alpha(\varepsilon) = \sum_{k=0}^{\infty} \varepsilon^k \alpha^{(k)} \quad \text{and} \quad \beta(\varepsilon) = \sum_{k=0}^{\infty} \varepsilon^k \beta^{(k)}$$

for $0 \leq \varepsilon < \varepsilon_1$. Since $\pi(\varepsilon)$ is a probability distribution for every $0 \leq \varepsilon < \varepsilon_1$, we have that

$\sum_{k=0}^{\infty} \varepsilon^k (\pi^{(k)} \mathbf{1}) = 1$, which yields that $\pi(0) \mathbf{1} = 1$ and $\pi^{(k)} \mathbf{1} = 0$ for every $k \geq 1$. Since $\pi(0)Q(0) = 0$, $\pi(0)$ is a stationary distribution for X^0 and so, by Assumption 3.5.1, it must be supported on $\mathcal{A}$ and so $\beta^{(0)} = 0$. In the next result we establish an equation that is satisfied by $\alpha^{(0)} = \alpha(0)$ and introduce a key matrix for our analysis. For convenience, let $\alpha := \alpha(0)$.

Lemma 3.5.1. *Under Assumption 3.5.1, $\pi(0) = [\alpha, 0]$, where 0 is the zero row vector of size $|\mathcal{T}|$ and α is an $|\mathcal{A}|$ -dimensional probability vector satisfying the equation:*

$$\alpha(A_1 + S_1(-T_0)^{-1}R_0) = 0. \quad (3.5.3)$$

In addition,

$$\beta^{(1)} = \alpha S_1(-T_0)^{-1}. \quad (3.5.4)$$

See SI-Section 3.10.2 for the proof of Lemma 3.5.1. For convenience, we adopt the notation:

$$Q_{\mathcal{A}} := A_1 + S_1(-T_0)^{-1}R_0. \quad (3.5.5)$$

In SI-Lemma 3.10.7, we show that $Q_{\mathcal{A}}$ is a Q -matrix of size $|\mathcal{A}| \times |\mathcal{A}|$. As a consequence, there exists a continuous time Markov chain with state space $\mathcal{A}$ and infinitesimal generator $Q_{\mathcal{A}}$. In general, a probability vector satisfying (3.5.3) needs not be unique. The following condition will imply uniqueness.

Assumption 3.5.2. The Markov chain associated with $Q_{\mathcal{A}}$ has a single recurrent class.

By SI-Lemma 3.10.1, Assumption 3.5.2 is equivalent to the condition $\dim(\ker(Q_{\mathcal{A}}^T)) = 1$. The next result then follows from Lemma 3.5.1.

Theorem 3.5.1. *Suppose Assumptions 3.5.1 and 3.5.2 hold. Then, $\pi(0) = [\alpha, 0]$, where α is the unique probability vector on $\mathcal{A}$ such that $\alpha Q_{\mathcal{A}} = 0$.*

As we will see, all of the chromatin modification circuit models presented in this work satisfy both Assumptions 3.5.1 and 3.5.2. Also note that Lemma 3.5.1 yields a characterization of $\beta^{(1)}$ by means of (3.5.4).

Theorem 3.5.1 is simple to state, yet less easy to use since simple formulas for $Q_{\mathcal{A}}$ can be seldom obtained, making Assumption 3.5.2 hard to verify directly using (3.5.5). In this regard, we now introduce an auxiliary continuous time Markov chain $\tilde{X}$ and use it to construct (via time-change) a realization $\hat{X}_{\mathcal{A}}$ of the continuous time Markov chain with infinitesimal generator $Q_{\mathcal{A}}$. This will enable us to give assumptions on $\tilde{X}$ that will imply Assumption 3.5.2 and which can sometimes be easier to verify. Also, this explicit realization for $\hat{X}_{\mathcal{A}}$ can lead to alternative ways to verify Assumption 3.5.2. Under Assumption 3.5.1, consider the matrix

$$\tilde{Q} = \left[\begin{array}{c|c} A_1 & S_1 \\ \hline R_0 & T_0 \end{array} \right]. \quad (3.5.6)$$

In SI-Lemma 3.10.7 we prove that $\tilde{Q}$ is a Q -matrix. Let $\tilde{X}$ be a continuous time Markov chain with infinitesimal generator $\tilde{Q}$. For the purpose of illustration, if we assume that the perturbation is linear (as in (3.3.3)) and $\varepsilon_0 > 1$, then the transitions of $\tilde{X}$ consist of the transitions of X^0 augmented by the transitions of X^1 that emanate from $\mathcal{A}$. See Figure 3.4(a)–(b) for an illustration related to the 1D and 2D models, respectively, introduced in Section 3.3.

Now, consider the **occupation time** of $\mathcal{A}$ by the Markov chain $\tilde{X}$ up to time $t \geq 0$, given by $\chi_{\mathcal{A}}(t) := \int_0^t \mathbf{1}_{\mathcal{A}}(\tilde{X}(s))ds$ for $t \geq 0$. Denote by $\chi_{\mathcal{A}}(\infty) = \lim_{t \rightarrow \infty} \chi_{\mathcal{A}}(t) = \int_0^\infty \mathbf{1}_{\mathcal{A}}(\tilde{X}(s))ds$. Since T_0 is invertible, SI-Lemmas 3.10.4 and 3.10.5 yield that $\mathbb{P}_x[\chi_{\mathcal{A}}(\infty) = \infty] = 1$ for all $x \in \mathcal{X}$. Additionally, consider the right-continuous inverse of $\chi_{\mathcal{A}}$, $\tau(s) := \inf\{t \geq 0 : \chi_{\mathcal{A}}(t) > s\}$, defined for $s \geq 0$. We define the **restriction** process $\hat{X}_{\mathcal{A}}$ as

$$\hat{X}_{\mathcal{A}}(s) := \tilde{X}(\tau(s)), \quad s \geq 0. \quad (3.5.7)$$

By properties of the right-continuous inverse (see Problem 4.5 in [28], for example), the reader may verify that $\hat{X}_{\mathcal{A}}$ corresponds to observing $\tilde{X}$ only on the time intervals where $\tilde{X}$ is in $\mathcal{A}$. Roughly speaking, we are *erasing* the times where $\tilde{X}$ is outside of $\mathcal{A}$. In the language of Blumenthal and Gettoor [10], $\chi_{\mathcal{A}}$ is a continuous additive functional for $\tilde{X}$, and by Exercise V.2.11 in Blumenthal and Gettoor [10], we obtain that $\hat{X}_{\mathcal{A}}$ is a continuous time Markov chain with state space $\mathcal{A}$. In the next result, we prove that $\hat{X}_{\mathcal{A}}$ is a realization of the continuous time Markov chain associated with $Q_{\mathcal{A}}$. See Figure 3.4(a)–(b) for a representation of $\hat{X}_{\mathcal{A}}$ associated with the 1D and 2D models, respectively.

Lemma 3.5.2. *Suppose Assumption 3.5.1 holds. Then, $\hat{X}_{\mathcal{A}}$ has infinitesimal generator $Q_{\mathcal{A}}$.*

The proof of Lemma 3.5.2 is given in SI-Section 3.10.2. We now introduce some assumptions that imply that Assumption 3.5.2 holds.

Assumption 3.5.3. For $\tilde{X}$, there exists a communicating class $\mathcal{C}$ such that $\mathcal{A} \subseteq \mathcal{C}$.

We note that, if such a class $\mathcal{C}$ exists, then it has to be recurrent. In fact, if it was transient then $\chi_{\mathcal{A}}(\infty) < \infty$ with positive probability under $\mathbb{P}_x$, $x \in \mathcal{A}$, which is a contradiction.

Assumption 3.5.4. The Markov chain $\tilde{X}$ is irreducible.

We note that Assumption 3.5.4 implies Assumption 3.5.3. Moreover, they are both related to Assumption 3.5.2 in the following way.

Lemma 3.5.3. *Suppose Assumptions 3.5.1 and 3.5.3 hold. Then, the process $\hat{X}_{\mathcal{A}}$ is irreducible. As a consequence, either Assumption 3.5.4 or 3.5.3 implies Assumption 3.5.2.*

The proof of Lemma 3.5.3 is given in SI-Section 3.10.2. The next result follows from Lemmas 3.5.2, 3.5.3 and Theorem 3.5.1.

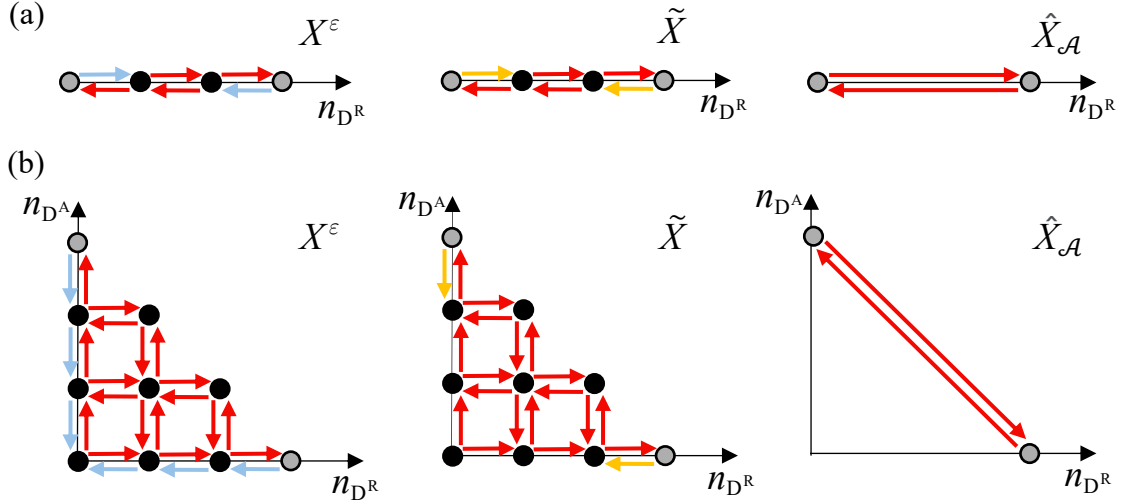


Figure 3.4. Graphs for the one-step transitions of X^ε , $\tilde{X}$ and $\hat{X}_\mathcal{A}$ for the (a) 1D model and (b) 2D model. Here, we consider $D_{\text{tot}} = 3$ and we use gray dots to represent the states belonging to $\mathcal{A}$ and black dots to represent all the other states, red arrows to represent transitions that are $O(1)$ for X^ε , $\tilde{X}$ and $\hat{X}_\mathcal{A}$, blue arrows to represent transitions that are $O(\varepsilon)$ for X^ε , and golden arrows to represent the transitions for $\tilde{X}$ that were $O(\varepsilon)$ for X^ε and became $O(1)$ for $\tilde{X}$.

Theorem 3.5.2. *Suppose Assumptions 3.5.1 and 3.5.3 hold. Then, $\pi(0) = [\alpha, 0]$ where α is the unique stationary distribution for $\hat{X}_\mathcal{A}$ and all entries of α are strictly positive.*

Assumptions 3.5.3 and 3.5.4 can be understood graphically in some cases. For example, Figure 3.4 illustrates that for the 1D model, Assumption 3.5.4 is satisfied. For the 2D model, we can see that while Assumption 3.5.4 is not satisfied (since the state $(0, 0)^T$ forms its own (transient) class for $\tilde{X}$), Assumption 3.5.3 does indeed hold. In Section 3.8, we will see that neither Assumption 3.5.4 nor 3.5.3 is satisfied by the 3D or 4D model. However, the weaker Assumption 3.5.2 does hold.

In the SI-Section 3.10.2, we give recursive formulae for the higher order terms $\pi^{(k)}$, $k = 1, 2, \dots$, under the following additional assumption (see SI-Theorem 3.10.1).

Assumption 3.5.5. The perturbation is linear, i.e., $Q(\varepsilon) = Q^{(0)} + \varepsilon Q^{(1)}$ for $0 \leq \varepsilon < \varepsilon_0$.

3.5.2 Illustrative examples: 1D and 2D model

1D model. We use the tools developed in the preceding section to derive the terms $\pi^{(0)}$ and $\pi^{(1)}$ in the expansion (3.3.4) for the 1D model introduced in Section 3.3.1. Fix $D_{\text{tot}} \geq 2$ and let X^ε with infinitesimal generator $Q(\varepsilon)$ be as in Section 3.3.1, with the expression for $Q^{(k)}$ given in (3.5.1). By (3.3.3), Assumption 3.5.5 holds. Moreover, for each $0 < \varepsilon < \varepsilon_0$, with ε_0 being a fixed, positive constant, $Q(\varepsilon)$ is irreducible, while $Q(0)$ has a non-empty set of transient states $\mathcal{T} = \{1, \dots, D_{\text{tot}} - 1\}$ and a set of two absorbing states $\mathcal{A} = \{a, r\}$, with $a = 0$ representing the fully active state ($n_{\text{DA}} = D_{\text{tot}}$) and $r = D_{\text{tot}}$ representing the fully repressed state ($n_{\text{DR}} = D_{\text{tot}}$). Then, Assumption 3.5.1 holds (see SI-Section 3.10.8). Furthermore, by defining $f(x) := x(D_{\text{tot}} - x)$ for $x \in \mathcal{X}$, we can write the matrices R_0 and T_0 in the matrix $Q^{(0)}$ as follows:

$$R_0 = \begin{bmatrix} \mu \frac{k_A}{V} f(1) & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & \frac{k_E}{V} f(1) \end{bmatrix},$$

$$T_0 = \begin{bmatrix} -(\lambda_1^0 + \gamma_1^0) & \lambda_1^0 & 0 & \dots & \dots & \dots & 0 \\ 0 & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 & \vdots & \vdots \\ \dots & 0 & \gamma_x^0 & -(\lambda_x^0 + \gamma_x^0) & \lambda_x^0 & 0 & \dots \\ \vdots & 0 & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & \dots & 0 & \gamma_{D_{\text{tot}}-1}^0 & -(\lambda_{D_{\text{tot}}-1}^0 + \gamma_{D_{\text{tot}}-1}^0) \end{bmatrix},$$

where R_0 is a $(D_{\text{tot}} - 1) \times 2$ matrix and T_0 is a $(D_{\text{tot}} - 1) \times (D_{\text{tot}} - 1)$ tridiagonal matrix, and $\gamma_x^0 = \mu \frac{k_E^A}{V} f(x)$, $\lambda_x^0 = \frac{k_E^A}{V} f(x)$, and $f(D_{\text{tot}} - 1) = f(1) = (D_{\text{tot}} - 1)$. In addition, we can write A_1 and S_1 of $Q^{(1)}$ as follows:

$$A_1 = \begin{bmatrix} -\frac{k_E^A}{V} D_{\text{tot}}^2 & 0 \\ 0 & -b\mu \frac{k_E^A}{V} D_{\text{tot}}^2 \end{bmatrix}, \quad S_1 = \begin{bmatrix} \frac{k_E^A}{V} D_{\text{tot}}^2 & 0 & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 & b\mu \frac{k_E^A}{V} D_{\text{tot}}^2 \end{bmatrix}.$$

The process $\tilde{X}$, whose infinitesimal generator is defined in (3.5.6), is irreducible (see SI-Section 3.10.8). This is illustrated in Figure 3.4(a). Thus, Assumption 3.5.4 holds. Then, Assumption 3.5.3 is also satisfied and Theorem 3.5.2 can be applied. This yields that $\pi(0) = \pi^{(0)} = [\alpha, 0] = [\alpha_a, \alpha_r, 0, \dots, 0]$ where α is the unique stationary distribution for the restriction process $\hat{X}_{\mathcal{A}}$ as defined by (3.5.7), whose infinitesimal generator is $Q_{\mathcal{A}} = A_1 + S_1(-T_0)^{-1}R_0$ by Lemma 3.5.2 and (3.5.5). Now,

$$Q_{\mathcal{A}} = \frac{1 - \mu}{1 - \mu^{D_{\text{tot}}}} \frac{k_E^A}{V} D_{\text{tot}}^2 \begin{bmatrix} -1 & 1 \\ b\mu^{D_{\text{tot}}} & -b\mu^{D_{\text{tot}}} \end{bmatrix}, \quad (3.5.8)$$

and since α is the unique probability vector satisfying $\alpha Q_{\mathcal{A}} = 0$, we have

$$\alpha_a = \frac{b\mu^{D_{\text{tot}}}}{1 + b\mu^{D_{\text{tot}}}}, \quad \alpha_r = \frac{1}{1 + b\mu^{D_{\text{tot}}}}.$$

These results are in agreement with (3.3.5) in Section 3.3.1, where we explicitly computed the stationary distribution $\pi(\varepsilon)$ and let $\varepsilon \rightarrow 0$ (see SI-Section 3.10.8).

Now, since Assumptions 3.5.1, 3.5.3, and 3.5.5 hold, we can apply SI-Theorem 3.10.1 to derive an expression for $\beta^{(1)}$. For the transient states $\mathcal{T} = \{1, \dots, D_{\text{tot}} - 1\}$, we have $\beta^{(1)} = [\pi_1^{(1)}, \dots, \pi_{D_{\text{tot}}-1}^{(1)}] = \alpha S_1(-T_0)^{-1}$, and so for $x \in \mathcal{T}$,

$$\beta_x^{(1)} = \frac{b\mu^{D_{\text{tot}}}}{1 + b\mu^{D_{\text{tot}}}} \frac{k_E^A}{V} D_{\text{tot}}^2 (-T_0)_{1,x}^{-1} + \frac{1}{1 + b\mu^{D_{\text{tot}}}} b\mu \frac{k_E^A}{V} D_{\text{tot}}^2 (-T_0)_{D_{\text{tot}}-1,x}^{-1},$$

in which $(-T_0)_{1,x}^{-1}$ and $(-T_0)_{D_{\text{tot}}-1,x}^{-1}$, for $x \in \mathcal{T}$, are the elements indexed by $(1, x)$ and $(D_{\text{tot}} - 1, x)$ of the matrix $(-T_0)^{-1}$, respectively. After some calculations, we obtain

$$(-T_0)_{1,x}^{-1} = \frac{\left(1 + \prod_{i=1}^{D_{\text{tot}}-1-x} \mu^i\right)}{\frac{k_E^A}{V} f(x) \left(1 + \prod_{i=1}^{D_{\text{tot}}-1} \mu^i\right)}, \quad (-T_0)_{D_{\text{tot}}-1,x}^{-1} = \frac{\left(1 + \prod_{i=1}^{x-1} \mu^i\right) \mu^{D_{\text{tot}}-1-x}}{\frac{k_E^A}{V} f(x) \left(1 + \prod_{i=1}^{D_{\text{tot}}-1} \mu^i\right)},$$

and then $\beta_x^{(1)}, x \in \mathcal{T}$, can be written as follows:

$$\beta_x^{(1)} = \frac{D_{\text{tot}}^2}{f(x)} \frac{b\mu^{D_{\text{tot}}-x}}{1 + b\mu^{D_{\text{tot}}}} = \frac{D_{\text{tot}}^2}{x(D_{\text{tot}} - x)} \frac{b\mu^{D_{\text{tot}}-x}}{1 + b\mu^{D_{\text{tot}}}}.$$

2D model. We analyze the stationary distribution for the 2D model introduced in Section 3.3.2. Fix $D_{\text{tot}} \geq 2$ and let X^ε with infinitesimal generator $Q(\varepsilon)$ be as in Section 3.3.1, with the expression for the $Q^{(k)}$ given by (3.5.1). By (3.3.9), Assumption 3.5.5 holds for this model. Furthermore, $Q(0)$ has a non-empty set of transient states $\mathcal{T} = \{i_1, \dots, i_m\}$ where $m = \frac{(D_{\text{tot}}+2)(D_{\text{tot}}+1)}{2} - 2$, $i_1 = (0, D_{\text{tot}} - 1)^T$, $i_m = (D_{\text{tot}} - 1, 0)^T$, and absorbing states $\mathcal{A} = \{a, r\}$, with $a = (0, D_{\text{tot}})^T$ corresponding to the fully active state ($n_{\text{DA}} = D_{\text{tot}}$) and with $r = (D_{\text{tot}}, 0)^T$ corresponding to the fully repressed state ($n_{\text{DR}} = D_{\text{tot}}$), respectively. Then, Assumption 3.5.1 holds (see SI-Section 3.10.9). From (3.3.9), we see that $A_0 = 0$, $S_0 = 0$ and

$$A_1 = \begin{bmatrix} -\frac{k_M^A}{V} D_{\text{tot}}^2 & 0 \\ 0 & -\frac{k_M^A}{V} D_{\text{tot}}^2 \mu b \end{bmatrix}, \quad S_1 = \begin{bmatrix} \frac{k_M^A}{V} D_{\text{tot}}^2 & 0 & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 & \frac{k_M^A}{V} D_{\text{tot}}^2 \mu b \end{bmatrix}.$$

Furthermore, $R_0 \in \mathbb{R}^{m \times 2}$ is given by

$$R_0 = \begin{bmatrix} k_{W0}^A + k_W^A + \frac{k_M^A}{V}(D_{\text{tot}} - 1) & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & k_{W0}^R + k_W^R + \frac{k_M^R}{V}(D_{\text{tot}} - 1) \end{bmatrix},$$

and $R_1 = 0$. The matrices T_0 and T_1 are more complex and examples of them, for $D_{\text{tot}} = 2$, are provided in SI-Section 3.10.9. For $\tilde{X}$, $\mathcal{C} = \mathcal{X} \setminus \{(0, 0)\}$ is a communicating class such that $\mathcal{A} \subseteq \mathcal{C}$. This implies that Assumption 3.5.3 is satisfied. Given that Assumptions 3.5.1 and 3.5.3 are satisfied, Theorem 3.5.2 can be applied and we obtain that $\pi(0) = \pi^{(0)} = [\alpha, 0] = [\alpha_a, \alpha_r, 0 \dots, 0]$ where α is the unique stationary distribution for the process $\hat{X}_{\mathcal{A}}$, whose infinitesimal generator is $Q_{\mathcal{A}} = A_1 + S_1(-T_0)^{-1}R_0$. This means that α is the unique probability vector such that $\alpha(A_1 + S_1(-T_0)^{-1}R_0) = 0$. Furthermore, given that Assumption 3.5.5 is satisfied, we can apply SI-Theorem 3.10.1 to derive an expression for $\beta^{(1)} = [\pi_{i_1}^{(1)}, \dots, \pi_{i_m}^{(1)}] = \alpha S_1(-T_0)^{-1}$.

As an example, if $D_{\text{tot}} = 2$, the matrix $Q_{\mathcal{A}}$ is given by

$$Q_{\mathcal{A}} = \frac{4}{K} \frac{k_M^A}{V} \begin{bmatrix} -(k_{W0}^R + k_W^R)(k_{W0}^R + k_W^R + \frac{k_M^R}{V}) & (k_{W0}^R + k_W^R)(k_{W0}^R + k_W^R + \frac{k_M^R}{V}) \\ b\mu^2(k_{W0}^A + k_W^A)(k_{W0}^A + k_W^A + \frac{k_M^A}{V}) & -b\mu^2(k_{W0}^A + k_W^A)(k_{W0}^A + k_W^A + \frac{k_M^A}{V}) \end{bmatrix}, \quad (3.5.9)$$

with

$$K = (k_{W0}^A + k_W^A + \frac{k_M^A}{V} + k_{W0}^R + k_W^R)(k_{W0}^R + k_W^R + \frac{k_M^R}{V}) + \mu(k_{W0}^R + k_W^R + \frac{k_M^R}{V} + k_{W0}^A + k_W^A)(k_{W0}^A + k_W^A + \frac{k_M^A}{V}), \quad (3.5.10)$$

and then $\pi^{(0)}$ is given by

$$\pi_x^{(0)} = \begin{cases} \frac{b\mu^2(k_{W0}^A + k_W^A)(k_{W0}^A + k_W^A + \frac{k_M^A}{V})}{b\mu^2(k_{W0}^A + k_W^A)(k_{W0}^A + k_W^A + \frac{k_M^A}{V}) + (k_{W0}^R + k_W^R)(k_{W0}^R + k_W^R + \frac{k_M^R}{V})} & \text{if } x = (0, D_{\text{tot}})^T, \\ 0 & \text{if } x \in \mathcal{T}, \\ \frac{(k_{W0}^R + k_W^R)(k_{W0}^R + k_W^R + \frac{k_M^R}{V})}{b\mu^2(k_{W0}^A + k_W^A)(k_{W0}^A + k_W^A + \frac{k_M^A}{V}) + (k_{W0}^R + k_W^R)(k_{W0}^R + k_W^R + \frac{k_M^R}{V})} & \text{if } x = (D_{\text{tot}}, 0)^T. \end{cases} \quad (3.5.11)$$

See SI-Section 3.10.9 for the evaluation of $\pi_x^{(1)}$ for the transient states $x \in \mathcal{T}$ when $D_{\text{tot}} = 2$. For this value of D_{tot} , we see from (3.5.11) that $\pi_x^{(0)}$ depends monotonically on μ for each fixed x . As D_{tot} increases, the algebraic complexity of a full parameter representation of $\pi_x^{(0)}$ increases very rapidly. Thus, to investigate monotonic dependence on parameters for biologically relevant values of D_{tot} (of the order of 50, considering an average gene length of 10,000 bp [16] and one nucleosome per 200 bp [19]), we shall use comparison theorems developed in Campos et al. [14], without calculating any explicit formula (see Section 3.7).

3.6 Main Results: Mean First Passage Times (MFPTs)

In this section we develop a theoretical framework to study mean first passage times for continuous time Markov chains. We first develop an algorithm to determine the order of the pole of MFPTs for singularly perturbed Markov chains in Section 3.6.1. In Section 3.6.2, we focus on determining the leading coefficient for MFPTs, under some assumptions introduced in Section 3.5. In Section 3.6.3, we apply these results to the examples introduced in Section 3.3.

3.6.1 Algorithm to find the order of the poles for MFPTs

Our algorithm is adapted from an algorithm developed by Hassin and Haviv [22] for discrete time Markov chains. The idea used in Hassin and Haviv [22] was to consider

transitions between subsets of states and to keep track of the sojourn times in the sets of states. This is used to define a coarser version of the process, which may not be a Markov process and which moves between groups of states of the original Markov chain. This idea can be adapted to the continuous time setting as well. For this, we introduce stopping times to more explicitly track the sojourn times than was done in Hassin and Haviv [22]. In addition, we extend the original algorithm's scope to consider the mean first passage time to a subset of states, instead of just a single state. Hassin and Haviv [22] uses r-cycles and notes that these could be replaced by more general r-components. Here, we focus on using the latter and call the set of vertices in such an r-component an r-connected set.

In this section, we consider a singularly perturbed, finite-state, continuous time Markov chain X^ε on $\mathcal{X}$ with infinitesimal generator $Q(\varepsilon)$ as described in Section 3.4.1. We provide an algorithm for finding the orders $\{p(v) : v \in \mathcal{B}^c\}$ of the poles for the mean first passage times to $\mathcal{B} \subset \mathcal{X}$ for X^ε starting from states in $\mathcal{B}^c$, where $\mathcal{B} \neq \emptyset$ is a strict subset of $\mathcal{X}$. We begin with a few definitions and some notation and then present the algorithm.

Definition 3.6.1. Given $\varepsilon_0 > 0$ and a function $f : (0, \varepsilon_0) \rightarrow \mathbb{R}_{>0}$, we say $f = \Theta(\varepsilon^k)$ if there exist $k \in \mathbb{Z}$ and strictly positive $m, M \in \mathbb{R}_{>0}$ such that, for all $0 < \varepsilon < \varepsilon_0$,

$$m\varepsilon^k \leq f(\varepsilon) \leq M\varepsilon^k.$$

If $f = \Theta(\varepsilon^k)$ for some $k \in \mathbb{Z}$, we say the **order** (at the origin) of f is k . If $f = \Theta(\varepsilon^{-k})$ where $k \in \mathbb{Z}_+$, we say that the **order of the pole** of f is k .

Because the perturbation of X^ε is real analytic, $|\mathcal{X}| > 1$ and X^ε is irreducible for $\varepsilon > 0$, there exists $\varepsilon_{\max} > 0$ such that for each $x \neq y \in \mathcal{X}$, either $Q_{x,y}(\varepsilon) \equiv 0$ for all $\varepsilon \in (0, \varepsilon_{\max})$ or $Q_{x,y}(\varepsilon) > 0$ for all $\varepsilon \in (0, \varepsilon_{\max})$. In the latter case, the order of $Q_{x,y}(\varepsilon)$ is a non-negative integer, which we denote by k_{xy} . We let $E_0 = \{(x, y) : Q_{x,y}(\varepsilon) > 0 \text{ for all } \varepsilon \in (0, \varepsilon_{\max})\}$. As the algorithm progresses, states of $\mathcal{X}$ are gathered together to form composite nodes and the graph of the states of X^ε progresses through a series

of reduced graphs. If u is a node in one of the graphs, then $S(u) \subset \mathcal{X}$ consists of the states in $\mathcal{X}$ that are collapsed to form the (reduced) node u . In Steps 2 and 3 of the algorithm, the function $\mathcal{K}$ and the initial values of p are inductively determined for all of these graphs. The final values of p for nodes in $\mathcal{B}^c$ are then determined in Step 4. With $\mathcal{K}_{uv}$ being defined, a directed edge (u, v) in one of the graphs is called an **r-edge**, where r is for regular, if $\mathcal{K}_{uv} = 0$, and an **r-path** is a directed path in the graph consisting of r -edges only. A set C in one of the graphs is called an **r-connected set** if $|C| > 1$ and there exists an r -path from u to v for any $u \neq v \in C$. The **order of the pole of the expected sojourn time** spent in an r -connected set C depends only on the set C and is denoted by $p(c)$ where c is a node representing the set C . For any node w outside of C , $\mathcal{K}_{cw}$ and $\mathcal{K}_{wc}$ are the **the order of the probabilities of a one-step transition** from c to w and from w to c , respectively. In Step 4 of the algorithm, $p(\cdot)$ keeps being updated but will stay finite and eventually fixate. The algorithm statement and related proof can be found in SI-Sections 3.10.3–3.10.5.

3.6.2 Leading coefficient in MFPT series expansion

In Section 3.4.1, we have shown that for each $0 < \varepsilon < \varepsilon_0$, the unique stationary distribution $\pi(\varepsilon)$ for X^ε admits a real-analytic expansion in powers of ε . By (3.4.3)–(3.4.4), for $x \in \mathcal{X}$,

$$\frac{1}{\pi_x(\varepsilon)} = q_x(\varepsilon) \mathbb{E}_x[\zeta_x^\varepsilon] = 1 + \sum_{y \neq x} Q_{x,y}(\varepsilon) h_{y,x}(\varepsilon).$$

Recall that $E_0 = \{(x, y) : Q_{x,y}(\varepsilon) > 0 \text{ for all } \varepsilon \in (0, \varepsilon_{\max})\}$ and k_{xy} is the order of $Q_{x,y}(\varepsilon)$ for each $(x, y) \in E_0$. Using the algorithm in Section 3.6.1, we can obtain the order of the pole, $p_x(y)$, of the mean first passage time $h_{y,x}(\varepsilon)$ from y to x for all $y \neq x \in \mathcal{X}$. Therefore, for each $x \in \mathcal{X}$, the order of $\pi_x(\varepsilon)$ is

$$k_x = \max\{p_x(y) - k_{xy} : (x, y) \in E_0; 0\} \geq 0, \quad (3.6.1)$$

and then

$$\pi_x(\varepsilon) = \sum_{k=k_x}^{\infty} \varepsilon^k \pi_x^{(k)}.$$

The following theorem is for continuous time and builds on discrete time results of Avrachenkov et al. [7, 6].

Theorem 3.6.1. *Suppose Assumptions 3.5.1, 3.5.2 and 3.5.5 hold. Let $Q_{\mathcal{A}}$ be given by (3.5.5), $\hat{X}_{\mathcal{A}}$ be as defined in (3.5.7), and α be the unique stationary distribution for $\hat{X}_{\mathcal{A}}$ defined in Theorem 3.5.1. Let $D = (-Q_{\mathcal{A}} + \mathbb{1}\alpha)^{-1} - \mathbb{1}\alpha$. For $y \in \mathcal{X}$, let k_y be the order of the stationary distribution $\pi_y(\varepsilon)$ of X^ε , defined by (3.6.1). Then, for $x, y \in \mathcal{A}$, the mean first passage time from x to y for X^ε is*

$$h_{x,y}(\varepsilon) = \frac{D_{y,y} - D_{x,y}}{\pi_y^{(k_y)}} \frac{1}{\varepsilon^{k_y+1}} + O\left(\frac{1}{\varepsilon^{k_y}}\right). \quad (3.6.2)$$

Moreover, if $\hat{X}_{\mathcal{A}}$ is an irreducible Markov chain, then the order of the pole of $h_{x,y}(\varepsilon)$ is one, i.e., $k_y = 0$, and the coefficient of ε^{-1} in (3.6.2) is equal to the mean first passage time from x to y for the process $\hat{X}_{\mathcal{A}}$.

The proof of Theorem 3.6.1 is given in SI-Section 3.10.7.

Remark 3.6.1. It may be possible that $D_{y,y} - D_{x,y} = 0$. In this case,

$$h_{x,y}(\varepsilon) = 0 \cdot \frac{1}{\varepsilon^{k_y+1}} + O\left(\frac{1}{\varepsilon^{k_y}}\right) = O\left(\frac{1}{\varepsilon^{k_y}}\right).$$

However, if we find that the order of the pole of $h_{x,y}(\varepsilon)$ is $k_y + 1$, using the algorithm in Section 3.6.1, then we can rule out the possibility of $D_{y,y} - D_{x,y}$ being zero.

3.6.3 Illustrative examples: 1D and 2D models

We first apply the algorithm given in Section 3.6.1 to find the order of the pole of the time to memory loss in the 1D and 2D models introduced in Section 3.3. For the 1D

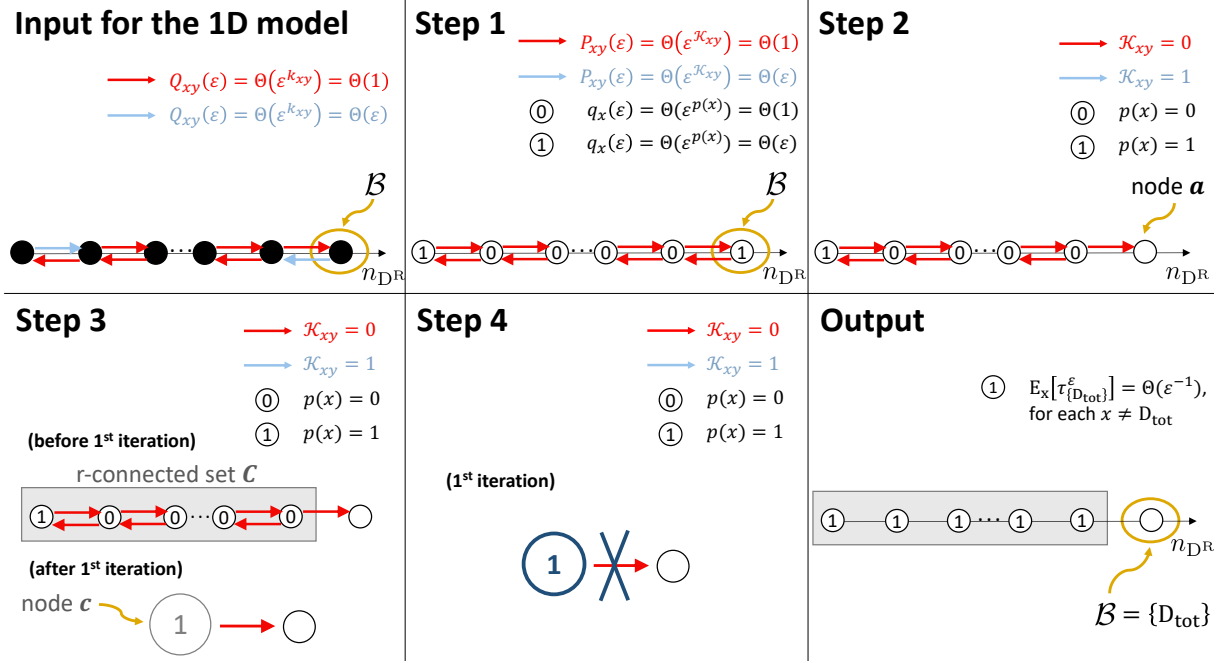


Figure 3.5. Key steps of the algorithm for the 1D model. The algorithm is described in Section 3.6.1, and it finds the order of the pole of the mean first passage time to $\mathcal{B} \subset \mathcal{X}$ from each state in $\mathcal{B}^c$. In our 1D model, the input for the algorithm is the order of each of the non-zero off-diagonal entries in $Q(\varepsilon)$ and the set $\mathcal{B} = \{D_{\text{tot}}\}$. The order of the non-zero entries in $Q(\varepsilon)$ is represented by colored arrows in the graph in the “Input” panel. Step 1 transforms the orders in the $Q(\varepsilon)$ -matrix into the orders in the $P(\varepsilon)$ -matrix and the exponential parameters $q(\varepsilon)$ to give an equivalent construction for the continuous time Markov chain. The order of the non-zero entries in $P(\varepsilon)$ is represented by colored arrows in the graph, and the number in the circle at a state $x \in \mathcal{B}^c$ is the order of the pole $p(x)$ of $\frac{1}{q_x(\varepsilon)}$ (the mean sojourn time at the state x). In Step 2, the set $\mathcal{B}$ is relabeled as the node a , and then all transitions from a to $\mathcal{B}^c$ are removed. Step 3 for the 1D model involves only one iteration, where the collection of all nodes except the node a (called an r -connected set $\mathcal{C}$) is condensed to a single node c , and the order of the pole at c is $p(c) = \max_{u \in \mathcal{C}} p(u) + \min\{\mathcal{K}_{uv} : u \in \mathcal{C}, v \notin \mathcal{C} \text{ and } (u, v) \in E\} = 1 + 0 = 1$, where E denotes the edge set of the graph in Step 3 before the 1st iteration. Moreover, $\mathcal{K}_{ca} = \min\{\mathcal{K}_{ua} : u \in \mathcal{C} \text{ and } (u, a) \in E\} - \min\{\mathcal{K}_{uv} : u \in \mathcal{C}, v \notin \mathcal{C} \text{ and } (u, v) \in E\} = 0 - 0 = 0$. Step 4 involves one iteration. In this iteration, the node c is the only node other than a , so its value of p is fixed, and then any edges leading to or from c are removed. When all of the nodes other than a have been fixed, the order of the pole of the mean first passage time from each state in $\mathcal{B}^c$ to $\mathcal{B}$ is given by the fixed value of the node to which the state belongs.

model, we could also directly derive the analytical expression for the time to memory loss by exploiting first step analysis (see for example Equation (3.1) in Norris [33]) and solve the system (3.4.1) introduced in Section 3.4.1 (see SI-Section 3.10.8). Figure 3.5 illustrates the key steps of the algorithm for the 1D model, which lead to the conclusion that the time to memory loss for the active state is $\Theta(\varepsilon^{-1})$. Because of the symmetry in the input graph in Figure 3.5, the time to memory loss for the repressed state is also $\Theta(\varepsilon^{-1})$. These orders found by the algorithm are consistent with what can be directly derived by first step analysis. Similarly, SI-Figure 3.8 illustrates the key steps of the algorithm for the 2D model, which leads to the conclusion that the time to memory loss of both the active and the repressed states is $\Theta(\varepsilon^{-1})$.

Next, we find the leading coefficient for the time to memory loss in the 1D and 2D models, which is the coefficient of the ε^{-1} term in all cases. Recall from Section 3.5.2 that Assumptions 3.5.1, 3.5.3 and 3.5.5 hold for both 1D and 2D models and hence by Lemma 3.5.3, so does Assumption 3.5.2 and $\hat{X}_{\mathcal{A}}$ is irreducible. For the 1D model, $Q_{\mathcal{A}}$ is given by (3.5.8). Thus, by Theorem 3.6.1, the leading coefficient of the time to memory loss for the active state is the mean first passage time from the fully active state a to the fully repressed state r in $\hat{X}_{\mathcal{A}}$, which has an exponential distribution with parameter $(Q_{\mathcal{A}})_{a,r} = \frac{1-\mu}{1-\mu^{D_{\text{tot}}}} \frac{k_E^A}{V} D_{\text{tot}}^2$ since $\hat{X}_{\mathcal{A}}$ has only two states. Thus, the time to memory loss for the active state is

$$h_{a,r}(\varepsilon) = \frac{1-\mu^{D_{\text{tot}}}}{1-\mu} \frac{V}{k_E^A} \frac{1}{D_{\text{tot}}^2} \varepsilon^{-1} + O(1),$$

and the time to memory loss for the repressed state is

$$h_{r,a}(\varepsilon) = \frac{1-\mu^{D_{\text{tot}}}}{1-\mu} \frac{V}{k_E^A} \frac{1}{b\mu^{D_{\text{tot}}} D_{\text{tot}}^2} \varepsilon^{-1} + O(1).$$

Similarly, in the 2D model, by Theorem 3.6.1,

$$h_{a,r}(\varepsilon) = \frac{1}{(Q_{\mathcal{A}})_{a,r}} \varepsilon^{-1} + O(1) \quad \text{and} \quad h_{r,a}(\varepsilon) = \frac{1}{(Q_{\mathcal{A}})_{r,a}} \varepsilon^{-1} + O(1).$$

As an example, when $D_{\text{tot}} = 2$, $Q_{\mathcal{A}}$ is shown in (3.5.9) and we obtain that

$$h_{a,r}(\varepsilon) = \frac{V}{k_M^A} \frac{K}{4(k_{W0}^R + k_W^R)(k_{W0}^R + k_W^R + \frac{k_M^R}{V})} \varepsilon^{-1} + O(1)$$

and

$$h_{r,a}(\varepsilon) = \frac{V}{k_M^A} \frac{K}{4b\mu^2(k_{W0}^A + k_W^A)(k_{W0}^A + k_W^A + \frac{k_M^A}{V})} \varepsilon^{-1} + O(1),$$

with K defined in (3.5.10).

3.7 Main Results: Monotonic Dependence on Parameters

An important aspect to consider in the study of the stochastic behavior of the chromatin modification circuit is that the erasure rate is different for each type of chromatin modification. These differences can introduce asymmetries in the system that can affect the stationary distribution and the time to memory loss of the active state and repressed state. These asymmetries are captured by the two parameters μ and μ' . In particular, μ quantifies the asymmetry between erasure rates of repressive and activating histone modifications and μ' quantifies the asymmetry between erasure rates of DNA methylation and activating histone modifications. In order to determine how the different chromatin modification erasure rates affect the stochastic behavior of the system, we study how μ and μ' affect the stationary distribution and the time to memory loss of the active and repressed gene states.

For the 1D model of the chromatin modification circuit, that does not include DNA methylation, we have an analytical expression for the stationary distribution in (3.3.5)

and the time to memory loss in (3.3.6)–(3.3.7) and we can understand the effect of μ by directly studying the formulas. However, for the higher-than-1D models we do not have an explicit expression for the stationary distribution or time to memory loss. This is the reason why for these models we exploit the comparison theory developed in Campos et al. [14] that allows to determine how μ and μ' affect the stochastic behavior of the system through the construction of a coupling between processes with different values for these parameters. In the next subsection, we briefly summarize the relevant theory from Campos et al. [14].

3.7.1 Comparison theorems for continuous time Markov chains

Denote by $\leq$ the usual componentwise partial order on $\mathbb{R}^d$, i.e., for $x, y \in \mathbb{R}^d$, $x \leq y$ whenever $x_i \leq y_i$ for every $1 \leq i \leq d$. Let $m, d \geq 1$ be integers, consider a matrix $A \in \mathbb{R}^{m \times d}$, where no row of A is identically zero, and consider the following definition.

Definition 3.7.1 (Definition 3.1 from Campos et al. [14]). For $x, y \in \mathbb{R}^d$, we say that $x \preceq_A y$ whenever $A(y - x) \geq 0$.

For the matrix A , consider the convex cone $K_A := \{x \in \mathbb{R}^d : Ax \geq 0\}$, and, for any $x \in \mathbb{R}^d$, consider the set $K_A + x = \{y \in \mathbb{R}^d : A(y - x) \geq 0\} = \{y \in \mathbb{R}^d : x \preceq_A y\}$ and the sets $\partial_i(K_A + x) := \{y \in K_A + x : \langle A_{i\bullet}, y \rangle = \langle A_{i\bullet}, x \rangle\}$ ³ for $1 \leq i \leq m$. Then, the boundary of $K_A + x$ can be expressed as

$$\partial(K_A + x) = \bigcup_{i=1}^m \partial_i(K_A + x).$$

Consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, we will say that a set $\Gamma \subseteq \mathcal{X}$ is **increasing** with respect to $\preceq_A$ if for every $x \in \Gamma$ and $y \in \mathcal{X}$, $x \preceq_A y$ implies that $y \in \Gamma$. We observe that,

³Here, for convenience of notation, let $A_{i\bullet}$ denote the row vector corresponding to the i -th row of A , for $1 \leq i \leq m$. In this article, we will adopt the convention of considering the inner product $\langle \cdot, \cdot \rangle$ as a function of a row vector in its first entry and as a function of a column vector in the second entry. In particular, $\langle A_{i\bullet}, x \rangle = \sum_{k=1}^d A_{ik}x_k$.

for $x \in \mathcal{X}$, the set

$$(K_A + x) \cap \mathcal{X} = \{y \in \mathcal{X} : x \preceq_A y\}$$

is increasing by the transitivity property of $\preceq_A$. On the other hand, we will say that a set $\Gamma \subseteq \mathcal{X}$ is **decreasing** with respect to $\preceq_A$ if for every $x \in \Gamma$ and $y \in \mathcal{X}$, $y \preceq_A x$ implies that $y \in \Gamma$.

Now, consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$ and a finite set of distinct nonzero possible transition vectors for a pair of continuous time Markov chains on $\mathcal{X}$. We denote the set of vectors by $\{v_1, \dots, v_n\} \subseteq \mathbb{Z}^d \setminus \{0\}$, where 0 is the origin in $\mathbb{Z}^d$. Consider two collections of functions $\Upsilon = (\Upsilon_1, \dots, \Upsilon_n)$ and $\check{\Upsilon} = (\check{\Upsilon}_1, \dots, \check{\Upsilon}_n)$ from $\mathcal{X}$ into $\mathbb{R}_+$ such that $\Upsilon_j(x) = \check{\Upsilon}_j(x) = 0$ if $x + v_j \notin \mathcal{X}$. Assume that $Q = (Q_{x,y})_{x,y \in \mathcal{X}}$, given by (3.4.6), is the infinitesimal generator for a continuous time Markov chain X and $\check{Q}$, defined by (3.4.6) but with functions $\check{\Upsilon}_1, \dots, \check{\Upsilon}_n$ in place of $\Upsilon_1, \dots, \Upsilon_n$, is the infinitesimal generator for a continuous time Markov chain $\check{X}$. We call X and $\check{X}$ the continuous time Markov chains associated with Υ and $\check{\Upsilon}$ respectively. The following stochastic comparison result was proved in Campos et al. [14]. The condition (i) of the theorem and $A \in \mathbb{Z}^{n \times d}$ ensure that to go outside of $K_A + x$, the Markov chains will necessarily hit the boundary of $K_A + x$.

Theorem 3.7.1 (Theorems 3.2, 3.4 and 3.5 from Campos et al. [14]). *With $\mathcal{X}$, $v_1, \dots, v_n$, Υ and $\check{\Upsilon}$ as described above, assume that the continuous time Markov chains associated with Υ and $\check{\Upsilon}$ do not explode in finite time. Consider a matrix $A \in \mathbb{Z}^{m \times d}$ with nonzero rows and suppose that both of the following conditions hold:*

(i) *For each $1 \leq j \leq n$, the vector Av_j has entries in $\{-1, 0, 1\}$ only.*

(ii) *For each $x \in \mathcal{X}$, $1 \leq i \leq m$ and $y \in \partial_i(K_A + x) \cap \mathcal{X}$ we have that*

$$\check{\Upsilon}_j(y) \leq \Upsilon_j(x), \quad \text{for each } 1 \leq j \leq n \text{ such that } \langle A_{i\bullet}, v_j \rangle < 0,$$

and

$$\check{\Upsilon}_j(y) \geq \Upsilon_j(x), \quad \text{for each } 1 \leq j \leq n \text{ such that } \langle A_{i\bullet}, v_j \rangle > 0.$$

Then, for each pair $x^\circ, \check{x}^\circ \in \mathcal{X}$ such that $x^\circ \preccurlyeq_A \check{x}^\circ$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with realizations of the two continuous time Markov chains $X = \{X(t) : t \geq 0\}$ and $\check{X} = \{\check{X}(t) : t \geq 0\}$ defined there, each having state space $\mathcal{X} \subseteq \mathbb{Z}_+^d$, with infinitesimal generators given by Q and $\check{Q}$, associated with Υ and $\check{\Upsilon}$, respectively, with initial conditions $X(0) = x^\circ$ and $\check{X}(0) = \check{x}^\circ$, and such that:

$$\mathbb{P} \left[X(t) \preccurlyeq_A \check{X}(t) \text{ for every } t \geq 0 \right] = 1.$$

Furthermore, for a non-empty set $\Gamma \subseteq \mathcal{X}$, consider $\tau_\Gamma := \inf\{t \geq 0 : X(t) \in \Gamma\}$ and $\check{\tau}_\Gamma := \inf\{t \geq 0 : \check{X}(t) \in \Gamma\}$. If Γ is increasing with respect to the relation $\preccurlyeq_A$, then $\mathbb{E}[\check{\tau}_\Gamma] \leq \mathbb{E}[\tau_\Gamma]$. If Γ is decreasing with respect to the relation $\preccurlyeq_A$, then $\mathbb{E}[\tau_\Gamma] \leq \mathbb{E}[\check{\tau}_\Gamma]$. Finally, suppose that the two continuous time Markov chains are irreducible and positive recurrent on $\mathcal{X}$, and denote the associated stationary distributions by π and $\check{\pi}$, respectively. Then, if $\Gamma \subseteq \mathcal{X}$ is a non-empty set that is increasing with respect to $\preccurlyeq_A$, we have $\sum_{x \in \Gamma} \pi_x \leq \sum_{x \in \Gamma} \check{\pi}_x$, or if $\Gamma \subseteq \mathcal{X}$ is a non-empty set that is decreasing with respect to $\preccurlyeq_A$, we have $\sum_{x \in \Gamma} \check{\pi}_x \leq \sum_{x \in \Gamma} \pi_x$.

3.7.2 Illustrative example: 2D model

We are interested in determining how the asymmetry of the system (represented by the parameter $\mu = k_E^R/k_E^A$) affects the stationary distribution $\pi(\varepsilon)$ and the times to memory loss, $h_{a,r}(\varepsilon)$ and $h_{r,a}(\varepsilon)$, of the active state $a = (0, D_{\text{tot}})^T$ and the repressed state $r = (D_{\text{tot}}, 0)^T$, respectively, for the continuous time Markov chain X^ε described in Section 3.3.2. For this, we use Theorem 3.7.1. For $\varepsilon \in (0, \varepsilon_0)$, let X^ε be the continuous time

Markov chain with

$$\Upsilon_1(x) = f_A(x), \quad \Upsilon_2(x) = g_A^\varepsilon(x), \quad \Upsilon_3(x) = f_R(x), \quad \Upsilon_4(x) = g_R^\varepsilon(x), \quad x \in \mathcal{X}$$

with $\mathcal{X}$, $v_1, \dots, v_4$, and $f_A(x)$, $g_A^\varepsilon(x)$, $f_R(x)$, $g_R^\varepsilon(x)$ as defined in Section 3.3.2, and introduce the continuous time Markov chain $\check{X}^\varepsilon$ defined on $\mathcal{X}$, having the same transition vectors of X^ε , and having infinitesimal transition rates $\check{\Upsilon}_1(x), \dots, \check{\Upsilon}_4(x)$ defined as for $\Upsilon_1(x), \dots, \Upsilon_4(x)$, with all the parameters having the same values except that μ is replaced by $\check{\mu}$, where $\mu \geq \check{\mu}$. Let

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad (3.7.1)$$

and let us consider the partial order $x \preceq_A y$. A similar example was analyzed by Example 4.4 in Campos et al. [14], using the results of Theorem 3.7.1. The only differences are that, in Campos et al. [14], the matrix A is the negative of the matrix given in (3.7.1) and the inequality between μ and $\check{\mu}$ is the opposite compared to the one considered here. The relationship between the notation in Campos et al. [14] and our notation is $\kappa_{1a} = k_{W0}^A + k_W^A$, $\kappa_{1b} = (k_M^A/V)$, $\kappa_{2a} = k_{W0}^R + k_W^R$, $\kappa_{2b} = (k_M^R/V)$, $\kappa_{3a} = \varepsilon(k_M^A/V)$, $\kappa_{3b} = (k_E^A/V)$, $c = b$.

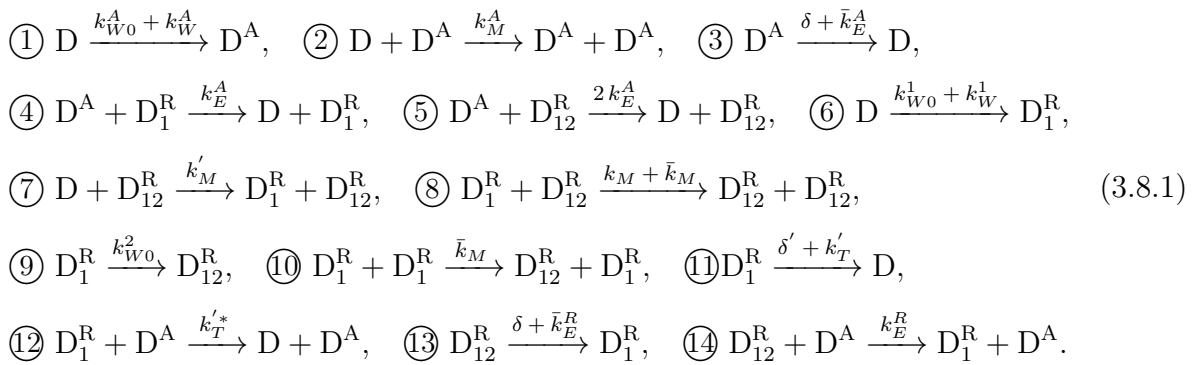
From the analysis in Campos et al. [14], we can directly conclude that, if $\pi(\varepsilon)$ is the stationary distribution for X^ε and $\check{\pi}(\varepsilon)$ the stationary distribution for $\check{X}^\varepsilon$, then $\check{\pi}_a(\varepsilon) \leq \pi_a(\varepsilon)$ and $\check{\pi}_r(\varepsilon) \geq \pi_r(\varepsilon)$. This implies that increasing μ increases the probability of the system in steady-state being in the active state a to the detriment of the repressed state r (and vice versa for decreasing μ). We can also conclude, using natural notation for quantities associated with X^ε and $\check{X}^\varepsilon$, that, defining $\tau_y^\varepsilon = \inf\{t \geq 0 : X^\varepsilon(t) = y\}$ and $\check{\tau}_y^\varepsilon = \inf\{t \geq 0 : \check{X}^\varepsilon(t) = y\}$, $h_{r,a}(\varepsilon) = \mathbb{E}_r[\tau_a^\varepsilon] \leq \mathbb{E}_r[\check{\tau}_a^\varepsilon] = \check{h}_{r,a}(\varepsilon)$ and $\check{h}_{a,r}(\varepsilon) = \mathbb{E}_a[\check{\tau}_r^\varepsilon] \leq \mathbb{E}_a[\tau_r^\varepsilon] = h_{a,r}(\varepsilon)$, implying that the time to memory loss of the repressed state decreases for higher values of μ , while the time to memory loss of the active state increases for higher values of μ .

3.8 Further Applications

In this section, compared to the models of the chromatin modification circuit introduced in Section 3.3, which do not include DNA methylation, we introduce more elaborate models that include DNA methylation and we study their stochastic behavior by exploiting the theory developed in this work.

3.8.1 3D chromatin modification circuit model

We now introduce a model in which DNA methylation is also a possible chromatin mark. The species involved are D (unmodified nucleosome), D_1^R (nucleosome with CpGme only), D_{12}^R (nucleosome with both H3K9me3 and CpGme) and D^A (nucleosome with an activating histone modification). In particular, we assume that, in order to be modified with both repressive modifications, D is first modified with DNA methylation, obtaining D_1^R , and then with a repressive histone modification, obtaining D_{12}^R . The opposite order of modifications is not allowed. This enables us to simplify the model and the related analysis. This assumption will be removed in the 4D model analyzed in Section 3.8.2. The chemical reaction system for the 3D model, shown in Figure 3.6(a), is the following:



where $k_{W0}^A, k_W^A, k_M^A, \delta, \bar{k}_E^A, k_E^A, k_{W0}^1, k_W^1, k_{W0}^2, k_W^2, k'_M, \bar{k}_M, k_M, \delta', k'_T, k_T'^*, \bar{k}_E^R, k_E^R > 0$ and the form of the reaction rate constants is due to the fact that reactions with the same reactants and products have been combined. As we did for the 2D model, define parameters

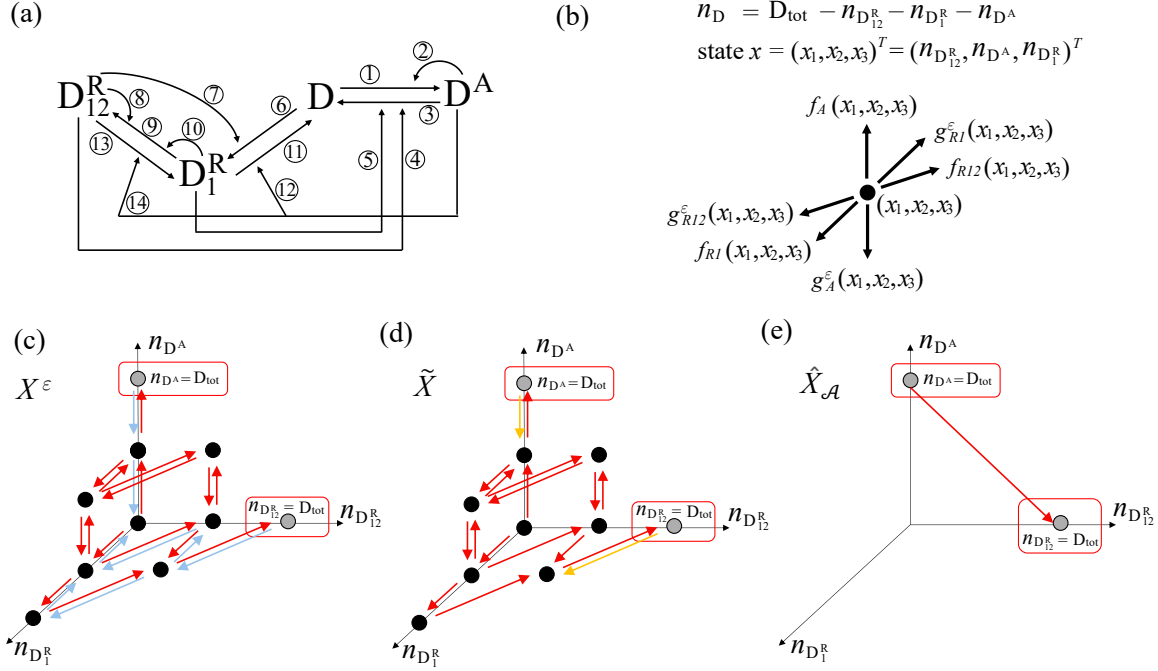


Figure 3.6. 3D Model and associated Markov chain. (a) Original chemical reaction system. The numbers on the arrows correspond to the reactions associated with the arrows as described in (3.8.1) in the main text. (b) Directions of the possible transitions of the Markov chain X^ε , starting from a state $x = (x_1, x_2, x_3)^T$ and whose rates are given in (3.8.2). (c) Graph for X^ε . Here, the red (blue) arrows correspond to $O(1)$ ($O(\varepsilon)$) transition rates. (d) Graph for the Markov chain $\tilde{X}$. Here, the gold arrows correspond to transitions that were $O(\varepsilon)$ in X^ε and became $O(1)$ in $\tilde{X}$. (e) Graph for the Markov chain $\hat{X}_\mathcal{A}$. For (c), (d), and (e) the state of the Markov chain is $x = (n_{D_{12}^R}, n_{D^A}, n_{D_1^R})^T$ and we consider $D_{\text{tot}} = 2$. In panels (c)–(e), we use gray dots to represent the states belonging to $\mathcal{A}$ and black dots to represent all the other states.

$\varepsilon = \frac{\delta + \bar{k}_E^A}{k_A^M D_{\text{tot}}}$ and $\mu = \frac{k_E^R}{k_E^A}$, with a constant b such that $\frac{\delta + \bar{k}_E^R}{\delta + \bar{k}_E^A} = b\mu$. Furthermore, since this model includes DNA methylation, we also define $\mu' = \frac{k_T'^*}{k_E^A}$ and a constant β such that $\frac{\delta' + k_T'}{\delta + \bar{k}_E^A} = \beta\mu'$. The parameter μ' quantifies the asymmetry between the erasure rates of DNA methylation and activating histone modifications.

The Markov chain associated with the system is a linearly perturbed finite state continuous time Markov chain with the state x tracking $n_{D_{12}^R}$, n_{D^A} , $n_{D_1^R}$, that is, the total number of nucleosomes of types D_{12}^R , D^A , and D_1^R , respectively. If the total number of modifiable nucleosomes is D_{tot} , which is conserved, the state space is $\mathcal{X} = \{(x_1, x_2, x_3) \in$

$\mathbb{Z}_+^3 : x_1 + x_2 + x_3 \leq D_{\text{tot}}\}$. The transition vectors for X^ε are given by $v_1 = (0, 1, 0)^T$, $v_2 = (0, -1, 0)^T$, $v_3 = (0, 0, 1)^T$, $v_4 = (0, 0, -1)^T$, $v_5 = (1, 0, -1)^T$ and $v_6 = (-1, 0, 1)^T$.

The infinitesimal transition rates are

$$\begin{aligned}
Q_{x, x+v_1}(\varepsilon) &= f_A(x) = (D_{\text{tot}} - (x_1 + x_2 + x_3)) \left(k_{W0}^A + k_W^A + \frac{k_M^A}{V} x_2 \right), \\
Q_{x, x+v_2}(\varepsilon) &= g_A^\varepsilon(x) = x_2 \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} + \frac{k_E^A}{V} (x_3 + 2x_1) \right), \\
Q_{x, x+v_3}(\varepsilon) &= f_{R1}(x) = (D_{\text{tot}} - (x_1 + x_2 + x_3)) \left(k_{W0}^1 + k_W^1 + \frac{k_M'}{V} x_1 \right), \\
Q_{x, x+v_4}(\varepsilon) &= g_{R1}^\varepsilon(x) = x_3 \mu' \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} \beta + x_2 \frac{k_E^A}{V} \right), \\
Q_{x, x+v_5}(\varepsilon) &= f_{R12}(x) = x_3 \left(k_{W0}^2 + \frac{k_M}{V} x_1 + \frac{\bar{k}_M}{V} \left(x_1 + \frac{x_3 - 1}{2} \right) \right), \\
Q_{x, x+v_6}(\varepsilon) &= g_{R12}^\varepsilon(x) = x_1 \mu \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} b + x_2 \frac{k_E^A}{V} \right).
\end{aligned} \tag{3.8.2}$$

A representation of the possible transitions, with associated rates, and the Markov chain graph for $D_{\text{tot}} = 2$ are given in Figure 3.6(b)–(c). Each rate depends on the state x .

Stationary distribution

We now focus on the expansion as a function of ε of the stationary distribution for the 3D model. In SI-Section 3.10.10, we show that, when $\varepsilon = 0$, the continuous time Markov chain associated with the 3D model has transient states $\mathcal{T} = \{i_1, \dots, i_m\}$ where $m = \sum_{j=0}^{D_{\text{tot}}} \left(\frac{(j+2)(j+1)}{2} \right) - 2$, $i_1 = (0, D_{\text{tot}} - 1, 0)^T$, $i_m = (D_{\text{tot}} - 1, 0, 1)^T$, and absorbing states $\mathcal{A} = \{a, r\}$, with $a = (0, D_{\text{tot}}, 0)^T$ corresponding to the fully active state ($n_{D^A} = D_{\text{tot}}$) and $r = (D_{\text{tot}}, 0, 0)^T$ corresponding to the fully repressed state ($n_{D_{12}^R} = D_{\text{tot}}$), respectively. Then, Assumption 3.5.1 holds (see SI-Section 3.10.10). Furthermore, $\mathcal{X} = \mathcal{A} \cup \mathcal{T}$ and from (3.8.2) we see that $Q(\varepsilon)$ can be written in the form (3.5.2), where $Q(\varepsilon)$ is a linear perturbation of $Q(0)$. Hence, Assumption 3.5.5 holds. Assumption 3.5.2 also holds, where the recurrent class is $\{r\}$ (see SI-Section 3.10.10). Then, we can apply SI-Theorem 3.10.1. We first obtain that $\pi(0) = \pi^{(0)} = [\alpha, 0] = [\alpha_a, \alpha_r, 0, \dots, 0]$ where α is the unique stationary

distribution for the process $\hat{X}_{\mathcal{A}}$ with infinitesimal generator $Q_{\mathcal{A}} = A_1 + S_1(-T_0)^{-1}R_0$. Since the recurrent class $\{r\}$ is a singleton and α is supported on $\{r\}$, we have

$$\alpha_a = 0 \quad \text{and} \quad \alpha_r = 1.$$

We now derive an expression for $\pi^{(1)}$. For the transient states $\mathcal{T} = \{i_1, \dots, i_m\}$, $\beta^{(1)} = [\pi_{i_1}^{(1)}, \dots, \pi_{i_m}^{(1)}] = \alpha S_1(-T_0)^{-1} = [0, \dots, 0, \pi_{i_m}^{(1)}]$, with

$$\pi_{i_m}^{(1)} = \frac{\mu b \frac{k_M^A}{V} D_{\text{tot}}^2}{k_{W0}^2 + (\frac{k_M}{V} + \frac{k_M}{V})(D_{\text{tot}} - 1)}.$$

See SI-Section 3.10.10 for the detailed mathematical derivation. Moreover, $\alpha^{(1)} = [\pi_a^{(1)}, \pi_r^{(1)}]$ is the unique vector such that $\alpha^{(1)}Q_{\mathcal{A}} = -\beta^{(1)}[R_1 + T_1(-T_0)^{-1}R_0]$, $\alpha^{(1)}\mathbf{1} = -\beta^{(1)}\mathbf{1}$.

As an illustration, suppose $D_{\text{tot}} = 2$. Then,

$$Q_{\mathcal{A}} = \frac{K_1 + \mu K_2}{K_3 + \mu K_4 + \mu' K_5 + \mu \mu' K_6} \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}, \quad (3.8.3)$$

$$\pi_a^{(1)} = \frac{\mu^2 \mu'^2 K_7}{K_8(K_9 + K_{10}\mu)}, \quad \pi_r^{(1)} = -\frac{\mu^2 \mu'^2 K_7 + \mu K_{11}(K_9 + K_{10}\mu)}{K_8(K_9 + K_{10}\mu)}, \quad (3.8.4)$$

with K_i , $i = 1, \dots, 11$, are non-negative constants independent of ε , μ and μ' (see SI-Section 3.10.10 for their precise definitions and for the detailed mathematical derivation). Hence, the stationary distribution for $D_{\text{tot}} = 2$ satisfies

$$\pi_x(\varepsilon) = \begin{cases} \varepsilon \frac{(\mu \mu')^2 K_7}{K_8(K_9 + K_{10}\mu)} + O(\varepsilon^2) & \text{if } x = a = (0, 2, 0)^T, \\ O(\varepsilon^2) & \text{if } x \in \mathcal{T} \setminus \{i_m\}, \\ \varepsilon \frac{K_{11}}{K_8} \mu + O(\varepsilon^2) & \text{if } x = i_m = (1, 0, 1)^T, \\ 1 - \varepsilon \frac{(\mu \mu')^2 K_7 + \mu K_{11}(K_9 + K_{10}\mu)}{K_8(K_9 + K_{10}\mu)} + O(\varepsilon^2) & \text{if } x = r = (2, 0, 0)^T. \end{cases}$$

For small $\varepsilon > 0$, the stationary distribution is concentrated around the active and repressed states, although more mass is concentrated around the repressed state. However, higher values of μ' increase the probability of being in the active state, while decreasing the probability of being in the repressed state.

Time to memory loss

In this section, we determine how the leakage of the system ε and the asymmetry between activating histone modifications and DNA methylation μ' affect the time to memory loss of the active state $h_{a,r}(\varepsilon)$ and the time to memory loss of the repressed state $h_{r,a}(\varepsilon)$.

Firstly, by the algorithm in Section 3.6.1, we have that $h_{a,r}(\varepsilon)$ is $O(\varepsilon^{-1})$ and $h_{r,a}(\varepsilon)$ is $O(\varepsilon^{-2})$ (see SI-Section 3.10.6). This means that decreasing the leakage extends the memory of both the active and repressed chromatin states, but the effect is stronger for the repressed state. This difference is influenced by the co-existence and cooperation between DNA methylation and repressive histone marks that introduce a structural bias in the 3D chromatin modification circuit towards a repressed chromatin state.

These results are consistent with the ones obtained by applying Theorem 3.6.1, which allows us not only to find the order of $h_{a,r}(\varepsilon)$ and $h_{r,a}(\varepsilon)$, but also to find an expression for their leading coefficients (see SI-Section 3.10.10 for the detailed mathematical derivation). As an example, when $D_{\text{tot}} = 2$, $Q_{\mathcal{A}}$ and $\pi_a^{(1)}$ are shown in (3.8.3)–(3.8.4), and we obtain from them that

$$h_{a,r}(\varepsilon) = \frac{K_3 + \mu K_4 + \mu' K_5 + \mu \mu' K_6}{K_1 + \mu K_2} \frac{1}{\varepsilon} + O(1),$$

and

$$h_{r,a}(\varepsilon) = \frac{K_3 + \mu K_4 + \mu' K_5 + \mu \mu' K_6}{K_1 + \mu K_2} \frac{K_8(K_9 + K_{10}\mu)}{\mu^2 \mu'^2 K_7} \frac{1}{\varepsilon^2} + O\left(\frac{1}{\varepsilon}\right),$$

where K_i , $i = 1, \dots, 11$, are non-negative constants independent of ε , μ and μ' , defined in SI-Section 3.10.10.

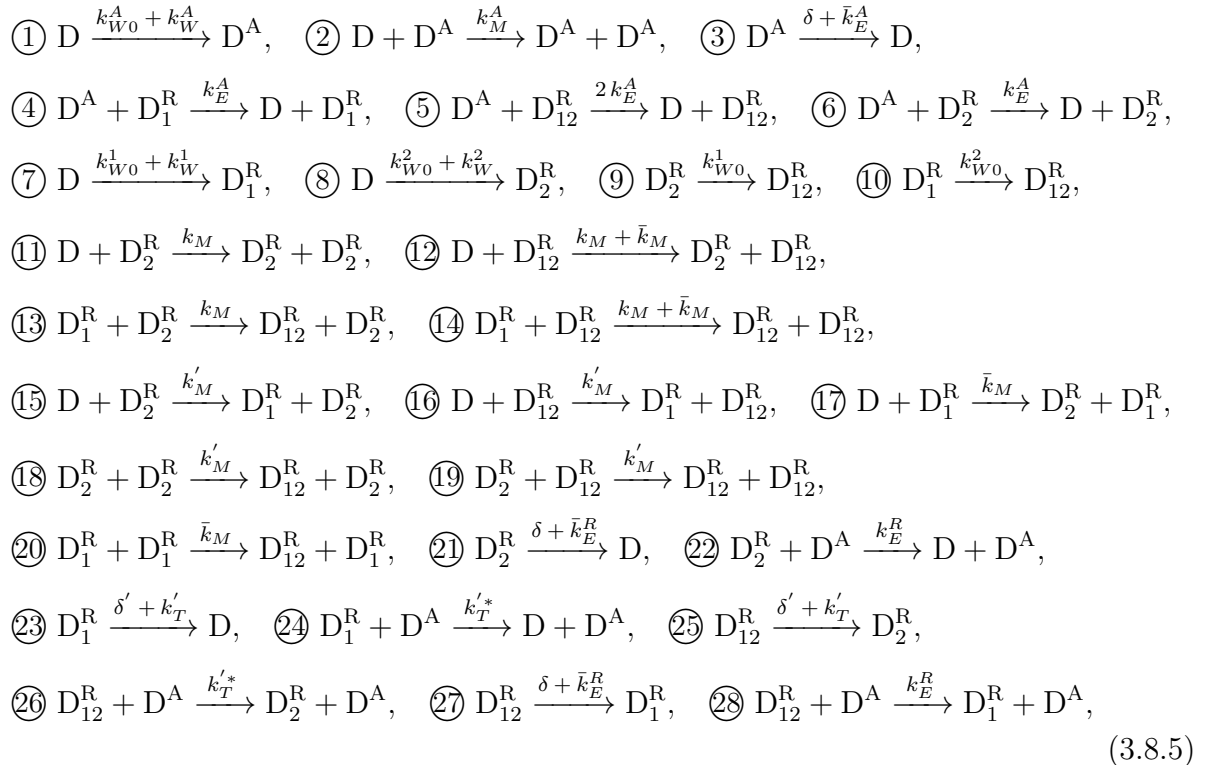
Now, we focus on understanding how the asymmetry between chromatin modification erasure rates affects the time to memory loss. In particular, since experimental data suggest that the asymmetry between the erasure rates of DNA methylation and activating histone modifications is more pronounced than the asymmetry between erasure rates of opposite histone modifications, in this analysis we focus only on studying the effect of μ' , but a similar procedure to the one presented in the next paragraph could be applied to study the effect of μ . To this end, we exploit the comparison Theorem 3.7.1 to determine directly how μ' affects $h_{a,r}(\varepsilon)$ and $h_{r,a}(\varepsilon)$, without deriving an explicit expression for them. To this end, we first note that the transitions of the Markov chain $X^\varepsilon(t)$ are in six possible directions, that can be written as $v_1 = (0, 1, 0)^T$, $v_2 = (0, -1, 0)^T$, $v_3 = (0, 0, 1)^T$, $v_4 = (0, 0, -1)^T$, $v_5 = (1, 0, -1)^T$, $v_6 = (-1, 0, 1)^T$, with the associated infinitesimal transition rates that can be written as $\Upsilon_1(x) = f_A(x)$, $\Upsilon_2(x) = g_\varepsilon^A(x)$, $\Upsilon_3(x) = f_{R1}(x)$, $\Upsilon_4(x) = g_{R1}^\varepsilon(x)$, $\Upsilon_5(x) = f_{R12}(x)$, $\Upsilon_6(x) = g_{R12}^\varepsilon(x)$. Define the matrix

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

and, for $x \in \mathcal{X}$, $(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} : x \preccurlyeq_A w\}$. Let us also introduce infinitesimal transition rates $\check{\Upsilon}_i(x)$, $i = 1, 2, \dots, 6$, defined as for $\Upsilon_i(x)$, $i = 1, 2, \dots, 6$, with all the parameters having the same values except that μ' is replaced by $\check{\mu}'$, with $\mu' \geq \check{\mu}'$. All of the conditions of Theorem 3.7.1 hold (see SI-Section 3.10.10) and so we can apply the theorem. This allows us to establish that, since $a = (0, D_{\text{tot}}, 0)^T \preccurlyeq_A r = (D_{\text{tot}}, 0, 0)^T$, then $\check{h}_{a,r}(\varepsilon) \leq h_{a,r}(\varepsilon)$ and $h_{r,a}(\varepsilon) \leq \check{h}_{r,a}(\varepsilon)$, where $\check{\cdot}$ indicates quantities associated with $\check{\Upsilon}$. Thus, we can conclude that, given that the only difference between the two systems was that $\mu' \geq \check{\mu}'$, the time to memory loss of the active state is monotonically increasing with μ' , while the time to memory loss of the repressed state is monotonically decreasing with μ' .

3.8.2 4D chromatin modification circuit model

Now, we consider a complete model in which the species involved are D , D_1^R , D_{12}^R , D^A and D_2^R (nucleosome with H3K9me3 only). Compared to the 3D model, we assume that, in order to be modified with both repressive modifications, D can be also modified first with a repressive histone modification (H3K9me3), obtaining D_2^R , and then with DNA methylation (CpGme), obtaining D_{12}^R . The chemical reaction system for the 4D model, shown in Figure 3.7(a), is



in which the form of the reaction rate constants is due to the fact that reactions with the same reactants and products have been combined. As we did for the 3D model, let us define the parameter $\varepsilon = \frac{\delta + \bar{k}_E^A}{k_M^A \frac{D_{\text{tot}}}{V}}$, the parameter $\mu = \frac{k_E^R}{k_E^A}$, with a constant b such that $\frac{\delta + \bar{k}_E^R}{\delta + \bar{k}_E^A} = b\mu$, and the parameter $\mu' = \frac{k_T'^*}{k_E^A}$, with a constant β such that $\frac{\delta' + k_T'}{\delta + \bar{k}_E^A} = \beta\mu'$.

The Markov chain X^ε associated with the system is a linearly perturbed finite state continuous time Markov chain with the state x representing the number of each

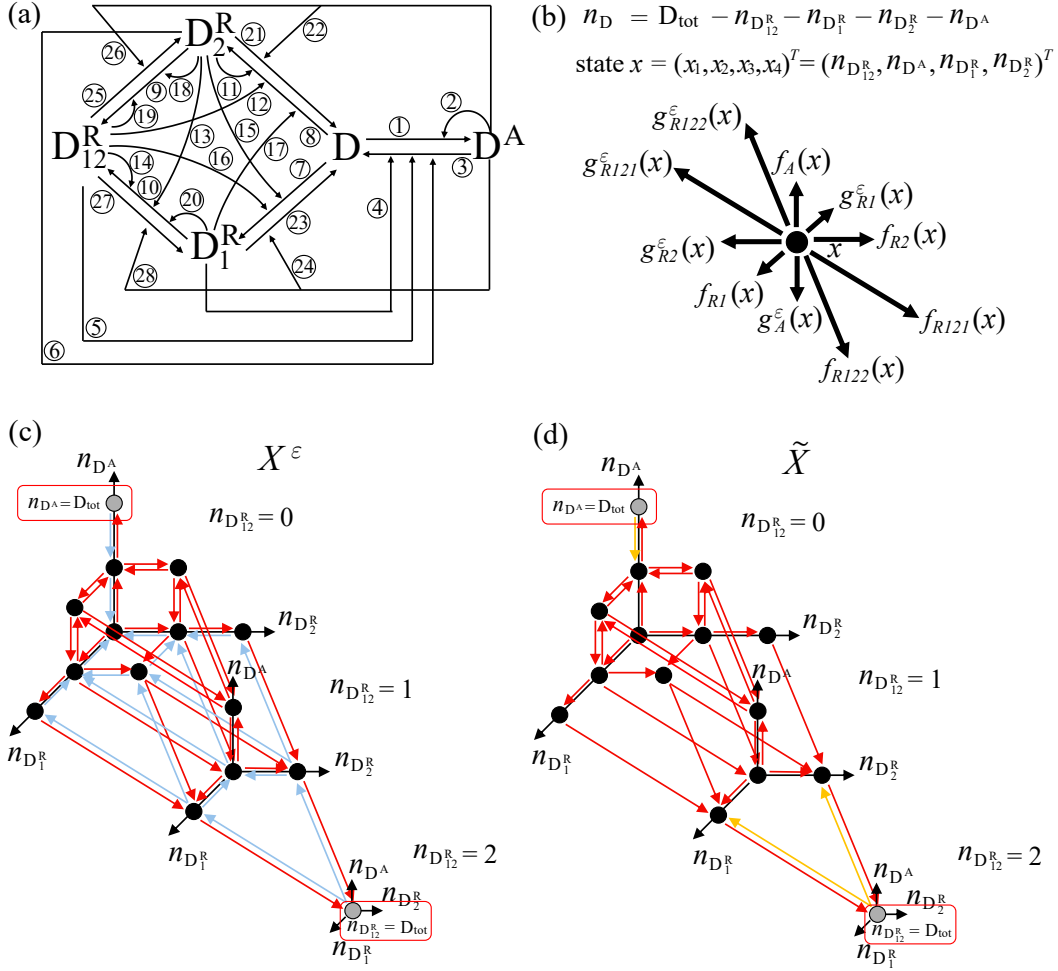


Figure 3.7. 4D Model and associated Markov chain. (a) Chemical reaction system. The numbers on the arrows correspond to the reactions associated with the arrows as described in (3.8.5) in the main text. (b) Directions of the possible transitions of the Markov chain X^ε associated with the reduced SCRN, starting from a state $x = (x_1, x_2, x_3, x_4)^T$ and whose rates are given in (3.8.6). (c) Graph for X^ε . Here, the red (blue) arrows correspond to $O(1)$ ($O(\varepsilon)$) transition rates. (d) Graph for the Markov chain $\tilde{X}$. Here, the golden arrows correspond to the transitions that were $O(\varepsilon)$ in X^ε and became $O(1)$ in $\tilde{X}$. For (c) and (d) the state of the Markov chain is $x = (n_{D_{12}^R}, n_{D^A}, n_{D_1^R}, n_{D_2^R})^T$, we consider $D_{\text{tot}} = 2$, and we show three interconnected slices ($n_{D_{12}^R} = 0, 1, 2$) of the Markov chain state space. In panels (c) and (d), we use gray dots to represent the states belonging to $\mathcal{A}$ and black dots to represent all the other states.

type of modified nucleosome, i.e., $x = (n_{D_{12}^R}, n_{D_A}, n_{D_1^R}, n_{D_2^R})^T = (x_1, x_2, x_3, x_4)^T$. If the total number of nucleosomes is D_{tot} , which is conserved, then the state space is $\mathcal{X} = \{(x_1, x_2, x_3, x_4)^T \in \mathbb{Z}_+^4 : x_1 + x_2 + x_3 + x_4 \leq D_{\text{tot}}\}$. The transition vectors for X^ε are given by $v_1 = (0, 1, 0, 0)^T$, $v_2 = (0, -1, 0, 0)^T$, $v_3 = (0, 0, 1, 0)^T$, $v_4 = (0, 0, -1, 0)^T$, $v_5 = (0, 0, 0, 1)^T$, $v_6 = (0, 0, 0, -1)^T$, $v_7 = (1, 0, -1, 0)^T$, $v_8 = (-1, 0, 1, 0)^T$, $v_9 = (1, 0, 0, -1)^T$ and $v_{10} = (-1, 0, 0, 1)^T$. The infinitesimal transition rates are

$$\begin{aligned}
Q_{x,x+v_1}(\varepsilon) &= f_A(x) = (D_{\text{tot}} - (x_1 + x_2 + x_3 + x_4)) \left(k_{W0}^A + k_W^A + \frac{k_M^A}{V} x_2 \right), \\
Q_{x,x+v_2}(\varepsilon) &= g_A^\varepsilon(x) = x_2 \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} + \frac{k_E^A}{V} (x_3 + x_4 + 2x_1) \right), \\
Q_{x,x+v_3}(\varepsilon) &= f_{R1}(x) = (D_{\text{tot}} - (x_1 + x_2 + x_3 + x_4)) \left(k_{W0}^1 + k_W^1 + \frac{k_M'}{V} (x_1 + x_4) \right), \\
Q_{x,x+v_4}(\varepsilon) &= g_{R1}^\varepsilon(x) = x_3 \mu' \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} \beta + x_2 \frac{k_E^A}{V} \right), \\
Q_{x,x+v_5}(\varepsilon) &= f_{R2}(x) = (D_{\text{tot}} - (x_1 + x_2 + x_3 + x_4)) \\
&\quad \left(k_{W0}^2 + k_W^2 + \frac{k_M}{V} (x_1 + x_4) + \frac{\bar{k}_M}{V} (x_1 + x_3) \right), \\
Q_{x,x+v_6}(\varepsilon) &= g_{R2}^\varepsilon(x) = x_4 \mu \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} b + x_2 \frac{k_E^A}{V} \right), \\
Q_{x,x+v_7}(\varepsilon) &= f_{R121}(x) = x_3 \left(k_{W0}^2 + \frac{k_M}{V} (x_1 + x_4) + \frac{\bar{k}_M}{V} \left(x_1 + \frac{x_3 - 1}{2} \right) \right), \\
Q_{x,x+v_8}(\varepsilon) &= g_{R121}^\varepsilon(x) = x_1 \mu \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} b + x_2 \frac{k_E^A}{V} \right), \\
Q_{x,x+v_9}(\varepsilon) &= f_{R122}(x) = x_4 \left(k_{W0}^1 + \frac{k_M'}{V} \left(x_1 + \frac{x_4 - 1}{2} \right) \right), \\
Q_{x,x+v_{10}}(\varepsilon) &= g_{R122}^\varepsilon(x) = x_1 \mu' \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} \beta + x_2 \frac{k_E^A}{V} \right).
\end{aligned} \tag{3.8.6}$$

A representation of the transition vectors and the Markov chain graph for $D_{\text{tot}} = 2$ are given in Figure 3.7(b)–(c), respectively. As before, each rate depends on the state x , but in the rest of the section we will not show this dependency to simplify the notation. Now, we determine the stochastic behavior of the full chromatin modification circuit

model in terms of its stationary distribution and time to memory loss. For this study, we will consider $k_W^A = k_W^1 = k_W^2 = 0$ (i.e., there are no external transcription factors enhancing the establishment of chromatin modifications). This assumption will not change the qualitative nature of the results focused on studying the effect of ε , μ , and μ' on the stochastic behavior of the chromatin modification circuit model.

Stationary distribution

We now determine the zeroth and first order terms of the stationary distribution expansion for the 4D model. As shown in SI-Section 3.10.11, when $\varepsilon = 0$, the continuous time Markov chain associated with the 4D model has transient states $\mathcal{T} = \{i_1, \dots, i_m\}$ where $m = \sum_{j=0}^{D_{\text{tot}}} \sum_{k=0}^j \binom{(k+2)(k+1)}{2} - 2$, $i_1 = (0, D_{\text{tot}} - 1, 0, 0)^T$, $i_{m-1} = (D_{\text{tot}} - 1, 0, 0, 1)^T$, $i_m = (D_{\text{tot}} - 1, 0, 1, 0)^T$, and absorbing states $\mathcal{A} = \{a, r\}$, with $a = (0, D_{\text{tot}}, 0, 0)^T$ corresponding to the fully active state ($n_{D^A} = D_{\text{tot}}$) and $r = (D_{\text{tot}}, 0, 0, 0)^T$ corresponding to the fully repressed state ($n_{D^R} = D_{\text{tot}}$), respectively. Then, Assumption 3.5.1 holds (see SI-Section 3.10.11), so that $\mathcal{X} = \mathcal{A} \cup \mathcal{T}$, and we can rewrite the infinitesimal generator $Q(\varepsilon)$ in the form of (3.5.2) (see SI-Section 3.10.11), where the perturbation is linear and so Assumption 3.5.5 holds. We can also verify Assumption 3.5.2 (see SI-Section 3.10.11). Hence, we can apply SI-Theorem 3.10.1, as was done for the 3D model. In particular, we obtain $\pi(0) = \pi^{(0)} = [\alpha, 0] = [\alpha_a, \alpha_r, 0, \dots, 0]$, with $\alpha_a = 0$, $\alpha_r = 1$, and $\beta^{(1)} = [\pi_{i_1}^{(1)}, \dots, \pi_{i_m}^{(1)}] = [0, \dots, 0, \pi_{i_{m-1}}^{(1)}, \pi_{i_m}^{(1)}]$, with

$$\pi_{i_{m-1}}^{(1)} = \frac{\mu' \beta \frac{k_M^A}{V} D_{\text{tot}}^2}{k_{W0}^1 + \frac{k'_M}{V} (D_{\text{tot}} - 1)} \quad \text{and} \quad \pi_{i_m}^{(1)} = \frac{\mu b \frac{k_M^A}{V} D_{\text{tot}}^2}{k_{W0}^2 + (\frac{k_M}{V} + \frac{\bar{k}_M}{V}) (D_{\text{tot}} - 1)}.$$

See SI-Section 3.10.11 for the detailed mathematical derivation. Moreover, $\alpha^{(1)} = [\pi_a^{(1)}, \pi_r^{(1)}]$ is the unique vector such that $\alpha^{(1)} Q_{\mathcal{A}} = -\beta^{(1)} [R_1 + T_1(-T_0)^{-1} R_0]$, $\alpha^{(1)} \mathbf{1} = -\beta^{(1)} \mathbf{1}$.

As an example, suppose $D_{\text{tot}} = 2$ and assume $\beta = b$, $k_{W0}^1 = k_{W0}^2 = k_{W0}^A$ and $k'_M = \bar{k}_M = k_M = k_M^A$. These assumptions do not affect the final qualitative conclusions

related to the effect of ε, μ and μ' on the stationary distribution. Then,

$$Q_{\mathcal{A}} = \frac{\bar{K}_1(\mu, \mu')}{\bar{K}_2(\mu, \mu')} \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}, \quad (3.8.7)$$

$$\pi_a^{(1)} = \frac{\bar{K}_3(\mu, \mu')}{\bar{K}_4(\mu, \mu')}, \quad \pi_r^{(1)} = -\frac{\bar{K}_3(\mu, \mu')}{\bar{K}_4(\mu, \mu')} - \mu' K_{18} - \mu K_{19}, \quad (3.8.8)$$

with

$$\begin{aligned} \bar{K}_1(\mu, \mu') &= K_1((\mu')^2 K_2 + (\mu)^2 K_3 + \mu\mu' K_4 + \mu' K_5 + \mu K_6 + K_7), \\ \bar{K}_2(\mu, \mu') &= \mu' \mu (\mu' + \mu) K_8 + (\mu')^2 K_9 + (\mu)^2 K_{10} + \mu\mu' K_{11} + \mu' K_{12} + \mu K_{13} + K_{14}, \\ \bar{K}_3(\mu, \mu') &= (\mu\mu')^2 K_{15}((\mu + \mu') K_{16} + K_{17}), \\ \bar{K}_4(\mu, \mu') &= K_{20}((\mu')^2 K_2 + (\mu)^2 K_3 + \mu\mu' K_4 + \mu' K_5 + \mu K_6 + K_7), \end{aligned} \quad (3.8.9)$$

in which K_i , $i = 1, \dots, 20$, are non-negative functions independent of ε , μ and μ' (see SI-Section 3.10.11 for their precise definitions and for the detailed mathematical derivation).

We then have

$$\pi_x(\varepsilon) = \begin{cases} \varepsilon \frac{\bar{K}_3(\mu, \mu')}{\bar{K}_4(\mu, \mu')} + O(\varepsilon^2) & \text{if } x = a = (0, 2, 0, 0)^T \\ O(\varepsilon^2) & \text{if } x \in \mathcal{T} \setminus \{i_{m-1}, i_m\} \\ \varepsilon \mu' K_{18} + O(\varepsilon^2) & \text{if } x = i_{m-1} = (1, 0, 0, 1)^T \\ \varepsilon \mu K_{19} + O(\varepsilon^2) & \text{if } x = i_m = (1, 0, 1, 0)^T \\ 1 - \varepsilon \left(\frac{\bar{K}_3(\mu, \mu')}{\bar{K}_4(\mu, \mu')} + \mu' K_{18} + \mu K_{19} \right) + O(\varepsilon^2) & \text{if } x = r = (2, 0, 0, 0)^T. \end{cases}$$

For small $\varepsilon > 0$, the stationary distribution is concentrated around the active and repressed states, and higher values of μ' or μ shift the distribution towards the active state.

Time to memory loss

As was done for the 3D model, we determine for the 4D model how the parameters ε and μ' affect the time to memory loss of the active state, $h_{a,r}(\varepsilon)$, and the time to memory loss of the repressed state, $h_{r,a}(\varepsilon)$. Firstly, by the algorithm in Section 3.6.1, $h_{a,r}(\varepsilon)$ is $O(\varepsilon^{-1})$ and $h_{r,a}(\varepsilon)$ is $O(\varepsilon^{-2})$ (see SI-Section 3.10.6). Then, by applying Theorem 3.6.1 we can obtain expressions for the leading coefficients of $h_{a,r}(\varepsilon)$ and $h_{r,a}(\varepsilon)$ (see SI-Section 3.10.11 for the detailed mathematical derivation). As an example, when $D_{\text{tot}} = 2$, Q_A and $\pi_a^{(1)}$ are shown in (3.8.7) and (3.8.8), and we obtain that

$$h_{a,r}(\varepsilon) = \frac{\bar{K}_2(\mu, \mu')}{\bar{K}_1(\mu, \mu')} \frac{1}{\varepsilon} + O(1) \quad \text{and} \quad h_{r,a}(\varepsilon) = \frac{\bar{K}_2(\mu, \mu') K_4(\mu, \mu')}{\bar{K}_1(\mu, \mu') \bar{K}_3(\mu, \mu')} \frac{1}{\varepsilon^2} + O\left(\frac{1}{\varepsilon}\right),$$

where $\bar{K}_i(\mu, \mu')$, $i = 1, \dots, 4$, are defined in (3.8.9).

Now, we determine how μ' , the parameter encapsulating the asymmetry between the DNA methylation erasure rate and the activating histone modification erasure rate, affects the time to memory loss. To this end, we seek to determine directly how μ' affects $h_{a,r}(\varepsilon)$ and $h_{r,a}(\varepsilon)$, without deriving an explicit expression for them. To this end, we would like to exploit two theorems from Campos et al. [14], namely, Theorem S.2 and Theorem 3.4 there. The transitions of the Markov chain X^ε are in ten possible directions, $v_1 = -v_2 = (1, 0, -1, 0)^T$, $v_3 = -v_4 = (1, 0, 0, -1)^T$, $v_5 = -v_6 = (0, 1, 0, 0)^T$, $v_7 = -v_8 = (0, 0, 1, 0)^T$, and $v_9 = -v_{10} = (0, 0, 0, 1)^T$, with the associated infinitesimal transition rates $\Upsilon_1(x) = f_{R121}(x)$, $\Upsilon_2(x) = g_{R121}^\varepsilon(x)$, $\Upsilon_3(x) = f_{R122}(x)$, $\Upsilon_4(x) = g_{R122}^\varepsilon(x)$, $\Upsilon_5(x) = f_A(x)$, $\Upsilon_6(x) = g_A^\varepsilon(x)$, $\Upsilon_7(x) = f_{R1}(x)$, $\Upsilon_8(x) = g_{R1}^\varepsilon(x)$, $\Upsilon_9(x) = f_{R2}(x)$, $\Upsilon_{10}(x) = g_{R2}^\varepsilon(x)$. Consider infinitesimal transition rates $\check{\Upsilon}_i(x)$, $i = 1, 2, \dots, 10$, defined as for $\Upsilon_i(x)$, $i = 1, 2, \dots, 10$, with all the parameters having the same values except that μ' is replaced by $\check{\mu}'$, with $\mu' \geq \check{\mu}'$. While we have not been able to see how to exploit Theorems S.2 and 3.4 from Campos et al. [14] for these exact rates, we have been able to do this for closely related rates. If we introduce a small approximation in the transition rates of X^ε ,

namely, $\frac{x_3-1}{2} \approx x_3$ and $\frac{x_4-1}{2} \approx x_4$ in $f_{R121}(x)$ and $f_{R122}(x)$, respectively, then Theorems S.2 and 3.4 in Campos et al. [14] apply with

$$A = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix},$$

and $(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} : x \preceq_A w\}$ (see SI-Section 3.10.11). This approximation can be justified by introducing the reasonable assumption that each nucleosome characterized by a repressive modification (D_1^R and D_2^R) has the ability to catalyze the establishment of the opposite repressive mark on itself. With this approximation, since $a = (0, D_{\text{tot}}, 0, 0)^T \preceq_A r = (D_{\text{tot}}, 0, 0, 0)^T$, then $\check{h}_{a,r}(\varepsilon) \leq h_{a,r}(\varepsilon)$ and $h_{r,a}(\varepsilon) \leq \check{h}_{r,a}(\varepsilon)$. Thus, the time to memory loss of the active state increases with higher values of μ' , while the time to memory loss of the repressed state decreases with higher values of μ' .

3.9 Conclusion

In this work, we provided a mathematical formulation and rigorous proofs to validate the computational findings in Bruno et al. [12], showing how the time scale separation between establishment and erasure processes of chromatin modifications affects epigenetic cell memory. To this end, we developed and adapted theory for singularly perturbed continuous time Markov chains and we analyzed the behavior of stationary distributions and mean first passage times as functions of the singular perturbation parameter ε .

We first showed that $\pi(\varepsilon)$ can be expressed as a series expansion (see Section 3.4.1) for sufficiently small ε . We then proved that the limit $\pi(0) = \lim_{\varepsilon \rightarrow 0} \pi(\varepsilon)$ is unique and we determined an expression for it (see Section 3.5.1). We also provided an iterative procedure for computing all of the higher order terms in the expansion of $\pi(\varepsilon)$ (see SI-Section 3.10.2).

Similarly, for the mean first passage time (MFPT) between states, we first showed there is a Laurent series expansion for sufficiently small ε (see (3.4.2) in Section 3.4.1). We then developed a graph based algorithm to identify the order of the leading term in the series expansion (see Section 3.6.1), and we also determined the leading coefficient there (see Section 3.6.2). We then applied these tools to the chromatin modification circuit models proposed in Bruno et al. [12], to provide a rigorous basis for the computational findings given there (see Sections 3.5.2, 3.6.3, 3.7.2 and 3.8).

Our rigorous derivations of the analytical expressions for the stationary distributions and time to memory loss, and our results on monotonic dependence on parameters, lead to a mechanistic understanding of how ε , μ and μ' affect the stochastic behavior of chromatin modification systems. As an example, our results suggest that higher values of μ and μ' shift mass of the stationary distribution more towards the active state. This finding is consistent with recent experimental results demonstrating that transfection of the DNA methylation eraser enzyme TET1, represented in our model by higher μ' (see Bruno et al. [12]), into Chinese hamster ovary (CHO-K1) repressed cells causes them to shift towards the active state (see Palacios et al. [34]). More generally, the mechanistic understanding of how ε , μ , and μ' affect the stochastic behavior of chromatin modification systems, as derived in our study, is crucial for determining experimental interventions on molecular players, such as chromatin modifier enzymes, to modulate cell memory. This mechanistic insight is expected to be extremely valuable for applications such as cell fate reprogramming and engineering approaches to cell therapy. Furthermore, the mathematical results and theoretical tools developed in this work can be applied beyond the scope of the epigenetic cell memory models analyzed in this research work. In fact, they can be applied to all stochastic models that respect the assumptions considered. Future work will investigate how to generalize these results by removing some of these assumptions, including allowing the Markov chain to have countably many states and $Q(0)$ to have ergodic classes as well as absorbing states. While there is some theory for countably many states, such as in

Altman et al. [2], the continuous time Markov chains for further applications that we have in mind are not uniformizable and have many transient states for $Q(0)$, and the theory in Altman et al. [2] needs to be generalized for them.

3.10 Supplementary Information (SI)

3.10.1 Some results in probability

Let $\mathcal{X}$ be a finite set. Recall the notation for matrices introduced in Section 3.2.5.

Lemma 3.10.1. *Let $X = \{X(t) : t \geq 0\}$ be a continuous time Markov chain with state space $\mathcal{X}$ and infinitesimal generator $Q = (Q_{x,y})_{x,y \in \mathcal{X}}$. Then, the number of recurrent classes for X is equal to $\text{nullity}(Q^T) = \text{nullity}(Q)$.*

Proof. Since Q is a square matrix, the Rank plus Nullity Theorem yields that $\text{nullity}(Q^T) = \text{nullity}(Q)$. Now, consider $\lambda > \max_{x \in \mathcal{X}} |Q_{x,x}|$ and define $P := I + Q/\lambda$, where I is the identity matrix of size $|\mathcal{X}| \times |\mathcal{X}|$. The matrix P is stochastic and such that for every $x \neq y$ in $\mathcal{X}$, $P_{x,y} > 0$ if and only if $Q_{x,y} > 0$. As a consequence, the recurrent classes of X are the same as the recurrent classes of P . By Theorem IV.2.4 in Isaacson and Madsen [26], the number of recurrent classes of P is equal to the maximum number of linearly independent left eigenvectors satisfying $\pi P = \pi$. By observing that $\pi P = \pi$ if and only if $\pi Q = 0$, we see that this latter quantity is equal to $\text{nullity}(Q^T)$. $\square$

The following is Proposition 6.3 in Asmussen [5].

Proposition 3.10.1. *Let $(P_{x,y})_{x,y \in \mathcal{X}}$ be a nonnegative substochastic matrix ($P\mathbf{1} \leq \mathbf{1}$) such that for each $x \in \mathcal{X}$ there are $z_1, \dots, z_m, y \in \mathcal{X}$ such that $P_{x,z_1}P_{z_1,z_2} \dots P_{z_m,y} > 0$ and $\sum_{z \in \mathcal{X}} P_{y,z} < 1$. Then, $\text{spr}(P) < 1$.*

We use SI-Proposition 3.10.1 in order to obtain invertibility for some matrices, as in the next result.

Lemma 3.10.2. *Let $X = \{X(t) : t \geq 0\}$ be an irreducible continuous time Markov chain with state space $\mathcal{X}$ and with an embedded discrete time Markov chain with transition matrix P . Consider a nonempty set $\mathcal{B} \subseteq \mathcal{X}$ such that $\mathcal{B} \neq \mathcal{X}$ and consider $P^{\mathcal{B}^c}$ to be the matrix obtained by removing the columns and rows of P corresponding to states in $\mathcal{B}$. Then, $I - P^{\mathcal{B}^c}$ is invertible and its inverse is given by the absolutely convergent series $\sum_{k=0}^{\infty} (P^{\mathcal{B}^c})^k$, where $(P^{\mathcal{B}^c})^0 = I$.*

Proof. Observe that $P^{\mathcal{B}^c} = (P_{x,y})_{x,y \in \mathcal{B}^c}$ is a nonnegative substochastic matrix. Since X is an irreducible continuous time Markov chain, its embedded discrete time Markov chain is also irreducible. Thus, for each $x \in \mathcal{B}^c$, there exist $z_1, \dots, z_m, y \in \mathcal{B}^c$ and $\tilde{y} \in \mathcal{B}$ such that $P_{x,z_1} P_{z_1,z_2} \dots P_{z_m,y} P_{y,\tilde{y}} > 0$. Then, $P_{x,z_1}^{\mathcal{B}^c} P_{z_1,z_2}^{\mathcal{B}^c} \dots P_{z_m,y}^{\mathcal{B}^c} > 0$ and $\sum_{z \in \mathcal{B}^c} P_{y,z}^{\mathcal{B}^c} = \sum_{z \in \mathcal{B}^c} P_{y,z} < 1$ since $P_{y,\tilde{y}} > 0$ and $\sum_{z \in \mathcal{X}} P_{y,z} = 1$. By SI-Proposition 3.10.1, $\text{spr}(P^{\mathcal{B}^c}) < 1$. This fact, together with Theorem 5.6.15 in Horn and Johnson [24] yields the convergence of $\sum_{k=0}^{\infty} (P^{\mathcal{B}^c})^k$. Moreover, $(I - P^{\mathcal{B}^c}) \sum_{k=0}^{\infty} (P^{\mathcal{B}^c})^k = \sum_{k=0}^{\infty} (P^{\mathcal{B}^c})^k (I - P^{\mathcal{B}^c}) = I$, which yields the desired result. $\square$

We will use the following continuous time analogue of SI-Proposition 3.10.1.

Lemma 3.10.3. *Let $(Q_{x,y})_{x,y \in \mathcal{X}}$ be a matrix such that $Q\mathbf{1} \leq 0$ and such that $Q_{x,x} \leq 0$ for each $x \in \mathcal{X}$ and $Q_{x,y} \geq 0$ for each $x \neq y \in \mathcal{X}$. In addition, suppose that for each $x \in \mathcal{X}$ there are distinct $z_1, \dots, z_m, y \in \mathcal{X}$ different from x such that $Q_{x,z_1} Q_{z_1,z_2} \dots Q_{z_m,y} > 0$ and $\sum_{z \in \mathcal{X}} Q_{y,z} < 0$. Then, for every $v \in \text{sp}(Q)$, the real part of v is negative. In particular, Q is invertible.*

Proof. Consider $\lambda > \max_{x \in \mathcal{X}} |Q_{x,x}|$ and define $P := I + Q/\lambda$, where I is the identity matrix of size $|\mathcal{X}| \times |\mathcal{X}|$. The matrix P is nonnegative, substochastic and such that $P_{x,y} = \frac{1}{\lambda} Q_{x,y}$ for every $x \neq y \in \mathcal{X}$. With these elements, we obtain that for each $x \in \mathcal{X}$ there are distinct $z_1, \dots, z_m, y \in \mathcal{X}$ such that $P_{x,z_1} P_{z_1,z_2} \dots P_{z_m,y} > 0$ and where $\sum_{z \in \mathcal{X}} P_{y,z} = 1 + \frac{1}{\lambda} \sum_{z \in \mathcal{X}} Q_{y,z} < 1$. SI-Proposition 3.10.1 yields that $\text{spr}(P) < 1$. By

observing that $v \in \text{sp}(Q)$ implies that $1 + \frac{v}{\lambda} \in \text{sp}(P)$, we obtain that $1 > |1 + \frac{v}{\lambda}| \geq |1 + \frac{\Re(v)}{\lambda}|$ where $\Re(v)$ is the real part of v . The latter inequality implies that $\Re(v) < 0$. $\square$

Consider a nonempty set $\mathcal{B} \subseteq \mathcal{X}$ such that $\mathcal{B} \neq \mathcal{X}$ and a Q -matrix written as

$$Q = \left[\begin{array}{c|c} \overbrace{Q^{\mathcal{B}}}^{\mathcal{B}} & \overbrace{S}^{\mathcal{B}^c} \\ \hline R & Q^{\mathcal{B}^c} \end{array} \right] \begin{matrix} \mathcal{B} \\ \mathcal{B}^c \end{matrix}.$$

Consider a process $X = \{X(t) : t \geq 0\}$ defined on a measurable space $(\Omega, \mathcal{F})$ and a collection of probability measures $\{\mathbb{P}_x : x \in \mathcal{X}\}$ on $(\Omega, \mathcal{F})$ such that for every $x \in \mathcal{X}$, X is a continuous time Markov chain under $\mathbb{P}_x$ with infinitesimal generator given by Q and such that $\mathbb{P}_x[X(0) = x] = 1$. Consider the stopping time $\tau_{\mathcal{B}} := \inf\{t \geq 0 : X(t) \in \mathcal{B}\}$.

Lemma 3.10.4. *The following are equivalent:*

(i) *For every $x \in \mathcal{B}^c$, there exists a $z \in \mathcal{B}$ such that x leads to z under Q , i.e., there are distinct $x_1, \dots, x_m \in \mathcal{B}^c$, different from x , such that $Q_{x,x_1}, Q_{x_1,x_2}, \dots, Q_{x_m,z} > 0$.*

(ii) *$Q^{\mathcal{B}^c}$ is invertible.*

(iii) *$\mathbb{E}_x[\tau_{\mathcal{B}}] < \infty$ for every $x \in \mathcal{B}^c$.*

If any of (i) – (iii) hold, then

$$\mathbb{P}_x[X(\tau_{\mathcal{B}}) = y] = -(Q^{\mathcal{B}^c})^{-1}R_{x,y} \quad (3.10.1)$$

for every $x \in \mathcal{B}^c$ and $y \in \mathcal{B}$. Moreover, if $\mathcal{B}^c$ is a set of transient states for X , then

(i) – (iii) hold.

Part of the results in SI-Lemma 3.10.4 appear as Lemma 1 in Gaver et al. [21] for the case where $Q^{\mathcal{B}}$ and S are the zero matrix. For completeness, we provide a self-contained proof here.

Proof. The implication (i) $\Rightarrow$ (ii) is a consequence of SI-Lemma 3.10.3 with $\mathcal{B}^c, Q^{\mathcal{B}^c}$ in place of $\mathcal{X}, Q$ there.

In the following, recall that for every $x \in \mathcal{X}$ and function $f : \mathcal{X} \rightarrow \mathbb{R}$, the process

$$M_f^{\mathcal{B}}(t) := f(X(t \wedge \tau_{\mathcal{B}})) - f(X(0)) - \int_0^{t \wedge \tau_{\mathcal{B}}} f(X(s)) ds, \quad t \geq 0, \quad (3.10.2)$$

is a martingale under $\mathbb{P}_x$ (see Theorem 3.32 in Liggett [31]), where $f(y) := \sum_{z \in \mathcal{X}} Q_{yz} f(z)$ for $y \in \mathcal{X}$.

For (ii) $\Rightarrow$ (iii), consider the function $f(y) := -[(Q^{\mathcal{B}^c})^{-1} \mathbf{1}]_y \mathbf{1}_{\mathcal{B}^c}(y)$ for $y \in \mathcal{X}$. The reader may verify that $f(y) = -1$ for every $y \in \mathcal{B}^c$. Therefore, for an $x \in \mathcal{B}^c$, taking expectations in (3.10.2) yields $f(x) - \mathbb{E}_x[f(X(t \wedge \tau_{\mathcal{B}}))] = \mathbb{E}_x[t \wedge \tau_{\mathcal{B}}]$ for every $t \geq 0$. Hence, $\mathbb{E}_x[t \wedge \tau_{\mathcal{B}}] \leq 2 \sup_{x \in \mathcal{X}} |f(x)|$ for every $t \geq 0$ and we conclude the desired result by letting $t \rightarrow \infty$.

For (iii) $\Rightarrow$ (i), we prove that not (i) implies not (iii). Suppose there exists $x \in \mathcal{B}^c$ such that no point of $\mathcal{B}$ is accessible from x . Then, $\tau_{\mathcal{B}} = \infty$ $\mathbb{P}_x$ -a.s., so (iii) does not hold.

For (3.10.1), consider $x \in \mathcal{B}^c, y \in \mathcal{B}$ and the martingale $M_f^{\mathcal{B}}$ with

$$f(z) = -(Q^{\mathcal{B}^c})^{-1} R_{z,y} \mathbf{1}_{\mathcal{B}^c}(z) + \mathbf{1}_{\{y\}}(z)$$

for $z \in \mathcal{X}$. The reader may verify that $f(x) = 0$ for $x \in \mathcal{B}^c$, which yields that $\mathbb{E}_x[f(X(t \wedge \tau_{\mathcal{B}}))] = f(x)$ for every $t \geq 0$. If any of (i) – (iii) hold, then $\tau_{\mathcal{B}} < \infty$, $\mathbb{P}_x$ -a.s., and on letting $t \rightarrow \infty$ and using bounded convergence, we obtain $\mathbb{E}_x[f(X(\tau_{\mathcal{B}}))] = f(x)$, which implies $\mathbb{P}_x[X(\tau_{\mathcal{B}}) = y] = -(Q^{\mathcal{B}^c})^{-1} R_{x,y}$.

Now, suppose that every $x \in \mathcal{B}^c$ is transient. Then, $\mathbb{P}_x(\tau_{\mathcal{B}} < \infty) = 1$ for each $x \in \mathcal{B}^c$. For a proof by contradiction, suppose that $Q^{\mathcal{B}^c}$ is not invertible, which implies the existence of a nonzero vector $v = (v(x))_{x \in \mathcal{B}^c} \neq 0$ such that $Q^{\mathcal{B}^c} v = 0$. Then, consider the function $f(y) = v(y) \mathbf{1}_{\mathcal{B}^c}(y)$, for which $f(y) = 0$ for $y \in \mathcal{B}^c$. Consider an $x \in \mathcal{B}^c$ such

that $v(x) \neq 0$, then $M_f^{\mathcal{B}}(t) = f(X(t \wedge \tau_{\mathcal{B}})) - v(x)$ is a bounded martingale. On taking expectations we have $\mathbb{E}_x[f(X(t \wedge \tau_{\mathcal{B}}))] = v(x)$. Since the states in $\mathcal{B}^c$ are transient, $X(\cdot)$ will $\mathbb{P}_x$ -a.s. leave $\mathcal{B}^c$. Then letting $t \rightarrow \infty$ and using bounded convergence yields $0 = v(x)$ which is the desired contradiction. Hence (ii) (as well as (i), (iii)) must hold. $\square$

SI-Lemma 3.10.4 has a useful consequence in terms of occupations times. In the above context, consider the occupation time of $\mathcal{B}$ by the Markov chain X up to time $t \geq 0$: $\chi_{\mathcal{B}}(t) = \int_0^t \mathbb{1}_{\mathcal{B}}(X(s))ds$. Denote by $\chi_{\mathcal{B}}(\infty) = \lim_{t \rightarrow \infty} \chi_{\mathcal{B}}(t) = \int_0^{\infty} \mathbb{1}_{\mathcal{B}}(X(s))ds$.

Lemma 3.10.5. *Suppose that*

$$\mathbb{P}_y[\tau_{\mathcal{B}} < \infty] = 1 \text{ for all } y \in \mathcal{B}^c. \quad (3.10.3)$$

Then $\mathbb{P}_x[\chi_{\mathcal{B}}(\infty) = \infty] = 1$ for every $x \in \mathcal{X}$.

Remark 3.10.1. If any of the conditions (i)–(iii) in SI-Lemma 3.10.4 holds, then (3.10.3) holds.

Proof. Fix $x \in \mathcal{X}$. Let $\sigma_{-1} = 0$ and $\sigma_0 = \inf\{t \geq \sigma_{-1} : X(t) \in \mathcal{B}\}$. Then, inductively define for $k = 0, 1, 2, \dots$, $\sigma_{2k+1} = \inf\{t \geq \sigma_{2k} : X(t) \in \mathcal{B}^c\}$ and $\sigma_{2k+2} = \inf\{t \geq \sigma_{2k+1} : X(t) \in \mathcal{B}\}$. Using (3.10.3) and the strong Markov property of X , we have

$$\sigma_{2k} < \infty \text{ } \mathbb{P}_x\text{-a.s. on } \{\sigma_{2k-1} < \infty\} \quad (3.10.4)$$

for $k = 0, 1, 2, \dots$, and

$$\chi_{\mathcal{B}}(\infty) = \sum_{k=0}^{\infty} \mathbb{1}_{\{\sigma_{2k} < \infty\}}(\sigma_{2k+1} - \sigma_{2k}) = \lim_{N \rightarrow \infty} \sum_{k=0}^N \mathbb{1}_{\{\sigma_{2k} < \infty\}}(\sigma_{2k+1} - \sigma_{2k}),$$

where terms in the sum indexed by $k : \sigma_{2k} = \infty$ are taken to be zero. Now, $\mathbb{P}_x$ -a.s.,

$$\prod_{k=0}^N \exp(-\mathbb{1}_{\{\sigma_{2k} < \infty\}}(\sigma_{2k+1} - \sigma_{2k})) = \prod_{k=0}^N \mathbb{1}_{\{\sigma_{2k} < \infty\}} \exp(-(\sigma_{2k+1} - \sigma_{2k})),$$

where we used (3.10.4) and the fact that $\sigma_{-1} = 0$, to conclude that the product is zero $\mathbb{P}_x$ -a.s., if $\{\sigma_{2k} = \infty\}$ for any $k \in \{0, 1, \dots, N\}$. Hence,

$$\begin{aligned} & \mathbb{E}_x[\exp(-\chi_{\mathcal{B}}(\infty))] \\ &= \lim_{N \rightarrow \infty} \mathbb{E}_x \left[\prod_{k=0}^N \mathbb{1}_{\{\sigma_{2k} < \infty\}} \exp(-(\sigma_{2k+1} - \sigma_{2k})) \right] \\ &= \lim_{N \rightarrow \infty} \mathbb{E}_x \left[\prod_{k=0}^{N-1} \mathbb{1}_{\{\sigma_{2k} < \infty\}} \exp(-(\sigma_{2k+1} - \sigma_{2k})) \mathbb{1}_{\{\sigma_{2N} < \infty\}} \mathbb{E}_x [\exp(-(\sigma_{2N+1} - \sigma_{2N})) | X(\sigma_{2N})] \right]. \end{aligned}$$

On $\{\sigma_{2N} < \infty\}$, we have $X(\sigma_{2N}) \in \mathcal{B}$ and $X(t) \in \mathcal{B}$ for $\sigma_{2N} \leq t < \sigma_{2N+1}$. Hence, for $a > \max_{y \in \mathcal{B}} |Q_{y,y}|$, using the strong Markov property, on $\{\sigma_{2N} < \infty\}$, conditioned on $X(\sigma_{2N})$, $\sigma_{2N+1} - \sigma_{2N}$ stochastically dominates an exponential random variable with parameter a and so

$$\mathbb{E}_x [\exp(-(\sigma_{2N+1} - \sigma_{2N})) | X(\sigma_{2N})] \leq \int_0^\infty e^{-t} a e^{-at} dt = \frac{a}{1+a}.$$

Similarly, for $k = N-1, \dots, 0$, on $\{\sigma_{2k} < \infty\}$,

$$\mathbb{E}_x [\exp(-(\sigma_{2k+1} - \sigma_{2k})) | X(\sigma_{2k})] \leq \frac{a}{1+a}. \quad (3.10.5)$$

Then,

$$\begin{aligned} \mathbb{E}_x[\exp(-\chi_{\mathcal{B}}(\infty))] &\leq \limsup_{N \rightarrow \infty} \mathbb{E}_x \left[\prod_{k=0}^{N-1} \mathbb{1}_{\{\sigma_{2k} < \infty\}} \exp(-(\sigma_{2k+1} - \sigma_{2k})) \mathbb{1}_{\{\sigma_{2N} < \infty\}} \frac{a}{1+a} \right] \\ &\leq \limsup_{N \rightarrow \infty} \mathbb{E}_x \left[\prod_{k=0}^{N-1} \mathbb{1}_{\{\sigma_{2k} < \infty\}} \exp(-(\sigma_{2k+1} - \sigma_{2k})) \frac{a}{1+a} \right]. \end{aligned}$$

Repeatedly conditioning on $\{\sigma_{2k} < \infty\}$, for $k = N - 1, \dots, 0$ and using (3.10.5), we obtain

$$\mathbb{E}_x[\exp(-\chi_{\mathcal{B}}(\infty))] \leq \limsup_{N \rightarrow \infty} \left(\frac{a}{1+a} \right)^{N+1} = 0.$$

Hence, $\mathbb{P}_x[\chi_{\mathcal{B}}(\infty) = \infty] = 1$. $\square$

Lemma 3.10.6. *Let $X = \{X(t) : t \geq 0\}$ be a continuous time Markov chain with state space $\mathcal{X}$ and infinitesimal generator $Q = (Q_{x,y})_{x,y \in \mathcal{X}}$. Suppose there is an absorbing state $y \in \mathcal{X}$. If there are distinct states $z_1, \dots, z_m \in \mathcal{X}$ different from x and y such that $Q_{x,z_1} Q_{z_1,z_2} \dots Q_{z_m,y} > 0$, then x is a transient state for X .*

Proof. Since $Q_{x,z_1} Q_{z_1,z_2} \dots Q_{z_m,y} > 0$, we have $\mathbb{P}_x[X(t_0) = y] > 0$ for some $t_0 > 0$ (see Theorem 3.2.1 in Norris [33]). Thus $\mathbb{P}_x[X(t) \neq x \text{ for all } t \geq t_0] \geq \mathbb{P}_x[X(t_0) = y] \mathbb{P}[X(t) = y \text{ for all } t > t_0 | X(t_0) = y] = \mathbb{P}_x[X(t_0) = y] > 0$, which means x is a transient state. $\square$

3.10.2 Additional results for stationary distribution

Higher order terms for linear perturbations

Under the assumption of the perturbation being linear (which is the case for our chromatin modification circuit models), we now provide an iterative procedure for computing all of the terms in the series expansion of $\pi(\cdot)$.

Theorem 3.10.1. *Suppose Assumptions 3.5.1, 3.5.2 and 3.5.5 hold. Then, the following hold for the sequence $\{\pi^{(k)} : k \geq 0\}$ in (3.4.5):*

(i) $\pi^{(0)} = [\alpha^{(0)}, \beta^{(0)}] = [\alpha, 0]$ where α is the unique probability vector on $\mathcal{A}$ such that

$$\alpha Q_{\mathcal{A}} = 0,$$

(ii) for every $k \geq 1$, $\pi^{(k)} = [\alpha^{(k)}, \beta^{(k)}]$, where

$$\beta^{(k)} = (\alpha^{(k-1)} S_1 + \beta^{(k-1)} T_1)(-T_0)^{-1} \tag{3.10.6}$$

and $\alpha^{(k)}$ is the unique vector such that

$$\alpha^{(k)} Q_{\mathcal{A}} = -\beta^{(k)} [R_1 + T_1(-T_0)^{-1} R_0], \quad (3.10.7)$$

$$\alpha^{(k)} \mathbf{1} = -\beta^{(k)} \mathbf{1}. \quad (3.10.8)$$

Moreover, if $|\mathcal{A}| \geq 2$, for every $k \geq 1$ we obtain

$$\alpha^{(k)} = \tilde{\alpha}^{(k)} + (-\beta^{(k)} \mathbf{1} - \tilde{\alpha}^{(k)} \mathbf{1}) \alpha, \quad (3.10.9)$$

where $\tilde{\alpha}^{(k)} := -\beta^{(k)} (R_1 + T_1(-T_0)^{-1} R_0) Q_{\mathcal{A}}^\dagger$ for $k \geq 1$ and $Q_{\mathcal{A}}^\dagger$ is a generalized inverse⁴ of $Q_{\mathcal{A}}$.

The proof of SI-Theorem 3.10.1 is given further below.

Proofs of Lemmas 3.5.1–3.5.3 and SI-Theorem 3.10.1

Proof of Lemma 3.5.1

Proof. It has already been established before Lemma 3.5.1 that $\pi(0) = [\alpha, 0]$. By equating to zero the coefficients of the terms $\varepsilon^m : m = 0, 1, \dots$ in the series $(\sum_{k=0}^{\infty} \varepsilon^k \pi^{(k)}) (\sum_{k=0}^{\infty} \varepsilon^k Q^{(k)})$, we obtain that $\sum_{k=0}^m \pi^{(k)} Q^{(m-k)} = 0$ for every $m \geq 0$. In particular, $\pi^{(1)} Q^{(0)} + \pi^{(0)} Q^{(1)} = 0$, which yields,

$$\begin{bmatrix} \alpha^{(1)} & \beta^{(1)} \end{bmatrix} \left[\begin{array}{c|c} 0 & 0 \\ \hline R_0 & T_0 \end{array} \right] + \begin{bmatrix} \alpha & 0 \end{bmatrix} \left[\begin{array}{c|c} A_1 & S_1 \\ \hline R_1 & T_1 \end{array} \right] = 0.$$

⁴A generalized inverse $Q_{\mathcal{A}}^\dagger$ of $Q_{\mathcal{A}}$ is such that $Q_{\mathcal{A}} Q_{\mathcal{A}}^\dagger Q_{\mathcal{A}} = Q_{\mathcal{A}}$. The Moore-Penrose inverse is a generalized inverse. The deviation matrix for $\hat{X}_{\mathcal{A}}$ is $D = (-Q_{\mathcal{A}} + \mathbf{1}\alpha)^{-1} - \mathbf{1}\alpha$, and $-D$ is also a generalized inverse of $Q_{\mathcal{A}}$. Meyer [32] suggested that $-D$ is a better generalized inverse to use than the Moore-Penrose inverse since it can be computed efficiently and embeds answers concerning the transitory behavior of the Markov chain. Avrachenkov et al. [6], in the context of discrete time Markov chains, use a suitable deviation matrix when they need a generalized inverse. Here, if we take $Q_{\mathcal{A}}^\dagger = -D$, then the term $\tilde{\alpha}^{(k)} \mathbf{1}$ in (3.10.9) is equal to zero since $D\mathbf{1} = 0$ and then $\alpha^{(k)} = \beta^{(k)} ((R_1 + T_1(-T_0)^{-1} R_0) D - \mathbf{1}\alpha)$.

From this, we obtain two equations:

$$\beta^{(1)}R_0 + \alpha A_1 = 0 \quad (3.10.10)$$

and

$$\beta^{(1)}T_0 + \alpha S_1 = 0. \quad (3.10.11)$$

Since T_0 is invertible, from (3.10.11) we obtain (3.5.4). We conclude by substituting this formula for $\beta^{(1)}$ in (3.10.10). $\square$

Proof of Lemma 3.5.2

Proof. By following the proof of Lemma 2 in Gaver et al. [21] and using formula (3.10.1), we can prove that the transition rates between $x \neq y \in \mathcal{A}$ for $\hat{X}_{\mathcal{A}}$ are given by $(Q_{\mathcal{A}})_{x,y}$. In essence, the argument is as follows. From the state $x \in \mathcal{A}$, the Markov chain $\tilde{X}$ may move to $y \in \mathcal{A}$ in two ways that lead to a one-step transition for $\hat{X}_{\mathcal{A}}$. First, it could happen that $\tilde{X}$ jumps directly to y at a rate of $(A_1)_{x,y}$. Second, the chain $\tilde{X}$ may go to some state $z \in \mathcal{T}$ at a rate $(S_1)_{x,z}$ and from there, jump between states in $\mathcal{T}$ until getting back to $\mathcal{A}$ at the state $y \in \mathcal{A}$. By (3.10.1), this happens with probability $((-T_0)^{-1}R_0)_{z,y}$. Putting this all together, the rate of transition for $\hat{X}_{\mathcal{A}}$ from x to y will be

$$(A_1)_{x,y} + \sum_{z \in \mathcal{T}} (S_1)_{x,z} ((-T_0)^{-1}R_0)_{z,y} = (Q_{\mathcal{A}})_{x,y}. \quad (3.10.12)$$

$\square$

Proof of Lemma 3.5.3

Proof. Consider $x \neq y \in \mathcal{A}$. Then, there exists a sequence of states $x_0 = x, x_1, \dots, x_m = y$ in $\mathcal{C}$ such that $\tilde{Q}_{x,x_1} \tilde{Q}_{x_1,x_2} \dots \tilde{Q}_{x_{m-1},y} > 0$. Roughly speaking, the proof follows by erasing the times that $\tilde{X}$ is outside of $\mathcal{A}$. We now give the details. Consider $i \in \{0, 1, \dots, m-1\}$

with $x_i \in \mathcal{A}$. If $x_{i+1} \in \mathcal{A}$, then, by (3.10.12), $(Q_{\mathcal{A}})_{x_i, x_{i+1}} \geq (A_1)_{x_i, x_{i+1}} = \tilde{Q}_{x_i, x_{i+1}} > 0$. If $x_{i+1} \in \mathcal{T}$, consider the path of states $x_i, x_{i+1}, \dots, x_k$ for $0 \leq i < k \leq m$ such that $x_i, x_k \in \mathcal{A}$ and $x_{i+1}, \dots, x_{k-1} \in \mathcal{T}$. Since x_{i+1} leads to x_k for $\tilde{X}$, then $\mathbb{P}_{x_{i+1}}[\tilde{X}(\tau_{\mathcal{A}}) = x_k] > 0$ where $\tau_{\mathcal{A}} := \inf\{t \geq 0 : \tilde{X}(t) \in \mathcal{A}\}$. By (3.10.1), this means that $((-T_0)^{-1}R_0)_{x_{i+1}, x_k} > 0$ which yields $(Q_{\mathcal{A}})_{x_i, x_k} \geq (S_1)_{x_i, x_{i+1}}((-T_0)^{-1}R_0)_{x_{i+1}, x_k} = \tilde{Q}_{x_i, x_{i+1}}((-T_0)^{-1}R_0)_{x_{i+1}, x_k} > 0$. These observations yield a sequence of states $x_0 = x, x_{i_1}, \dots, x_{i_j} = y$ in $\mathcal{A}$ such that $(Q_{\mathcal{A}})_{x, x_{i_1}}(Q_{\mathcal{A}})_{x_{i_1}, x_{i_2}} \dots (Q_{\mathcal{A}})_{x_{i_{j-1}}, y} > 0$. $\square$

Proof of SI-Theorem 3.10.1

Proof. Point (i) was established in Theorem 3.5.1. For (ii), we equate to zero the coefficients of the terms ε^m for $m = 0, 1, 2, \dots$ in the terms of the series, $(\sum_{k=0}^{\infty} \varepsilon^k \pi^{(k)})(Q^{(0)} + \varepsilon Q^{(1)})$ to obtain that $\pi^{(0)}Q^{(0)} = 0$ and $\pi^{(k)}Q^{(0)} + \pi^{(k-1)}Q^{(1)} = 0$ for every $k \geq 1$. The latter requires that for all $k \geq 1$,

$$\begin{bmatrix} \alpha^{(k)} & \beta^{(k)} \end{bmatrix} \left[\begin{array}{c|c} 0 & 0 \\ \hline R_0 & T_0 \end{array} \right] + \begin{bmatrix} \alpha^{(k-1)} & \beta^{(k-1)} \end{bmatrix} \left[\begin{array}{c|c} A_1 & S_1 \\ \hline R_1 & T_1 \end{array} \right] = 0.$$

Now, this yields two equations:

$$\beta^{(k)}R_0 + \alpha^{(k-1)}A_1 + \beta^{(k-1)}R_1 = 0, \quad (3.10.13)$$

$$\beta^{(k)}T_0 + \alpha^{(k-1)}S_1 + \beta^{(k-1)}T_1 = 0. \quad (3.10.14)$$

For $\beta^{(k)}$, we obtain the relation (3.10.6) directly from (3.10.14) for all $k \geq 1$. For $\alpha^{(k)}$, let's see first that it satisfies (3.10.7). From (3.10.6), for all $k \geq 1$, $\beta^{(k+1)} = (\alpha^{(k)}S_1 + \beta^{(k)}T_1)(-T_0)^{-1}$ and using this in (3.10.13) (with k replaced by $k+1$) we obtain for all $k \geq 1$

$$(\alpha^{(k)}S_1 + \beta^{(k)}T_1)(-T_0)^{-1}R_0 + \alpha^{(k)}A_1 + \beta^{(k)}R_1 = 0. \quad (3.10.15)$$

By rearranging (3.10.15) and using (3.5.5), we obtain (3.10.7) for all $k \geq 1$. On the other hand, since $\langle \pi^{(k)}, \mathbf{1} \rangle = 0$ for every $k \geq 1$, we obtain (3.10.8).

For the uniqueness of $\alpha^{(k)}$, for all $k \geq 1$, if $|\mathcal{A}| = 1$, then $\alpha^{(k)}$ has only one entry and it is determined uniquely by (3.10.8). If $|\mathcal{A}| \geq 2$, consider another solution $\gamma^{(k)}$ of (3.10.7)–(3.10.8), where $(\gamma^{(k)})^T \in \mathbb{R}^{|\mathcal{A}|}$. By Assumption 3.5.2 and SI-Lemma 3.10.1, $\dim(\ker((Q_{\mathcal{A}})^T)) = 1$ and therefore, by (3.10.7), $\alpha^{(k)} - \gamma^{(k)} = c\alpha$ for some $c \in \mathbb{R}$. Using (3.10.8), then $0 = \alpha^{(k)}\mathbf{1} - \gamma^{(k)}\mathbf{1} = c\alpha\mathbf{1} = c$, and therefore $c = 0$, and $\alpha^{(k)} = \gamma^{(k)}$.

For existence of a solution $\alpha^{(k)}$ of (3.10.7)–(3.10.8), using the properties

$$R_0\mathbf{1} + T_0\mathbf{1} = 0 \quad \text{and} \quad R_1\mathbf{1} + T_1\mathbf{1} = 0,$$

we have that

$$\begin{aligned} (R_1 + T_1(-T_0)^{-1}R_0)\mathbf{1} &= R_1\mathbf{1} + T_1(-T_0)^{-1}R_0\mathbf{1} \\ &= -T_1\mathbf{1} + T_1(-T_0)^{-1}R_0\mathbf{1} \\ &= T_1(-T_0)^{-1}(T_0\mathbf{1} + R_0\mathbf{1}) \\ &= 0. \end{aligned}$$

Then, since $\dim(\ker(Q_{\mathcal{A}})) = \dim(\ker((Q_{\mathcal{A}})^T)) = 1$ and $\mathbf{1} \in \ker(Q_{\mathcal{A}})$, we have

$$(-\beta^{(k)}(R_1 + T_1(-T_0)^{-1}R_0))^T \in \ker(Q_{\mathcal{A}})^{\perp} = \text{range}((Q_{\mathcal{A}})^T),$$

and so (3.10.7) has a solution and (3.10.8) will determine the multiple of α to add to any particular solution to obtain the unique solution $\alpha^{(k)}$ of both equations.

Furthermore, if $Q_{\mathcal{A}}^{\dagger}$ is a generalized inverse of $Q_{\mathcal{A}}$, then

$$\tilde{\alpha}^{(k)} := -\beta^{(k)}(R_1 + T_1(-T_0)^{-1}R_0)Q_{\mathcal{A}}^{\dagger}$$

is a solution to (3.10.7) (see James [27] for an exposition). Similar to the uniqueness argument, $\alpha^{(k)} - \tilde{\alpha}^{(k)} = c\alpha$ for some $c \in \mathbb{R}$. By (3.10.8), $c = -\beta^{(k)}\mathbf{1} - \tilde{\alpha}^{(k)}\mathbf{1}$ and we obtain (3.10.9). $\square$

SI-Lemma 3.10.7

Lemma 3.10.7. *Under Assumption 3.5.1, the matrices $Q_{\mathcal{A}}$ and $\tilde{Q}$ are Q -matrices of sizes $|\mathcal{A}| \times |\mathcal{A}|$ and $|\mathcal{X}| \times |\mathcal{X}|$ respectively.*

Proof. First, observe that $\lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} Q_{x,y}(\varepsilon) = (A_1)_{x,y}$ if $x, y \in \mathcal{A}$, while $\lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} Q_{x,y}(\varepsilon) = (S_1)_{x,y}$ if $x \in \mathcal{A}$ and $y \in \mathcal{T}$. Then, since $Q(\varepsilon)$ is a Q -matrix, S_1 has nonnegative entries, $(A_1)_{x,y} \geq 0$ for $x \neq y \in \mathcal{A}$, and

$$\sum_{y \in \mathcal{A}} (A_1)_{x,y} + \sum_{y \in \mathcal{T}} (S_1)_{x,y} = 0 \quad \text{for every } x \in \mathcal{A}. \quad (3.10.16)$$

For $x \neq y \in \mathcal{A}$, $(Q_{\mathcal{A}})_{x,y} = (A_1)_{x,y} + \sum_{z \in \mathcal{T}} (S_1)_{x,z} ((-T_0)^{-1} R_0)_{z,y}$ is nonnegative since $(A_1)_{x,y} \geq 0$ and by (3.10.1), $((-T_0)^{-1} R_0)_{z,y} = 0$ for $z \in \mathcal{T}$. For $x \in \mathcal{A}$, $\sum_{y \in \mathcal{A}} (Q_{\mathcal{A}})_{x,y}$ is equal to

$$\begin{aligned} & \sum_{y \in \mathcal{A}} (A_1)_{x,y} + \sum_{y \in \mathcal{A}} \sum_{z \in \mathcal{T}} (S_1)_{x,z} ((-T_0)^{-1} R_0)_{z,y} \\ &= \sum_{y \in \mathcal{A}} (A_1)_{x,y} + \sum_{z \in \mathcal{T}} (S_1)_{x,z} \sum_{y \in \mathcal{A}} ((-T_0)^{-1} R_0)_{z,y} \\ &= \sum_{y \in \mathcal{A}} (A_1)_{x,y} + \sum_{z \in \mathcal{T}} (S_1)_{x,z} = 0, \end{aligned}$$

where we used (3.10.1) and (3.10.16). Hence $Q_{\mathcal{A}}$ is a Q -matrix.

For $\tilde{Q}$, for $x \neq y \in \mathcal{X}$, if $x \in \mathcal{A}$, then $\tilde{Q}_{x,y}$ corresponds to an off diagonal term in A_1 or a term in S_1 , both of which are nonnegative. If $x \in \mathcal{T}$, then $\tilde{Q}_{x,y}$ corresponds to an off diagonal term in T_0 or a term in R_0 , which are both nonnegative since $Q(0)$ is a Q -matrix. To check that the row-sums of $\tilde{Q}$ are zero, first consider when $x \in \mathcal{A}$.

Then, $\sum_{y \in \mathcal{X}} \tilde{Q}_{x,y} = \sum_{y \in \mathcal{A}} (A_1)_{x,y} + \sum_{y \in \mathcal{T}} (S_1)_{x,y} = 0$ by (3.10.16). If $x \in \mathcal{T}$, then $\sum_{y \in \mathcal{X}} \tilde{Q}_{x,y} = \sum_{y \in \mathcal{A}} (R_0)_{x,y} + \sum_{y \in \mathcal{T}} (T_0)_{x,y} = 0$, since this corresponds to summing across a row of $Q(0)$. $\square$

3.10.3 Algorithm to find the order of the pole of the MFPT

Input: $\mathcal{B} \subset \mathcal{X}$, and k_{xy} , the order of $Q_{x,y}(\varepsilon)$ for each $(x, y) \in E_0$.

Output: $p(x)$, the order of the pole of the mean first passage from $x \in \mathcal{B}^c$ to $\mathcal{B}$.

(p will also be defined for condensed nodes in the course of the algorithm)

Step 1 (Set up the initial graph (V, E))

Construct a directed graph $G = (V, E)$ where $V = \mathcal{X}$ and $E = E_0$.

Set, for each $u \in V$, $p(u) \leftarrow \min\{k_{uv} : (u, v) \in E\}$.

Set, for each $(u, v) \in E$, $\mathcal{K}_{uv} \leftarrow k_{uv} - p(u)$.

Step 2 (Condense $\mathcal{B}$ into a single node a)

Introduce a new node a .

Set, for each $w \in \mathcal{B}^c$ such that $(w, v) \in E$ for some $v \in \mathcal{B}$,

$$\mathcal{K}_{wa} \leftarrow \min\{\mathcal{K}_{wv} : v \in \mathcal{B} \text{ and } (w, v) \in E\}.$$

Update $V \leftarrow \mathcal{B}^c \cup \{a\}$ and

$$E \leftarrow \{(u, v) \in E : u \in \mathcal{B}^c \text{ and } v \in \mathcal{B}^c\} \\ \cup \{(w, a) : (w, v) \in E \text{ for some } w \in \mathcal{B}^c \text{ and } v \in \mathcal{B}\}.$$

Set, for each $u \in V \setminus \{a\}$, $S(u) \leftarrow \{u\}$, and $S(a) \leftarrow \mathcal{B}$.

Step 3 (Condense r-connected sets)

Repeat the following until G contains no r-connected sets:

Let $C \subset V$ be an r-connected set

Let c be a new node representing the r-connected set C .

Set $p(c) \leftarrow \max_{u \in C} p(u) + \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u, v) \in E\}$.

Set, for each $w \in V \setminus C$ such that $(u, w) \in E$ for some $u \in C$,

$$\mathcal{K}_{cw} \leftarrow \min\{\mathcal{K}_{uw} : u \in C \text{ and } (u, w) \in E\}$$

$$- \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u, v) \in E\}$$

Set, for each $w \in V \setminus C$ such that $(w, v) \in E$ for some $v \in C$,

$$\mathcal{K}_{wc} \leftarrow \min\{\mathcal{K}_{wv} : v \in C \text{ and } (w, v) \in E\}.$$

Update $V \leftarrow (V \setminus C) \cup \{c\}$ and

$$E \leftarrow \{(u, v) \in E : u \notin C \text{ and } v \notin C\}$$

$$\cup \{(c, w) : (u, w) \in E \text{ for some } u \in C \text{ and } w \notin C\}$$

$$\cup \{(w, c) : (w, v) \in E \text{ for some } w \notin C \text{ and } v \in C\}.$$

Set $S(c) \leftarrow \cup_{u \in C} S(u)$.

Step 4 (Compute $p(x)$ where $x \in \mathcal{B}^c$)

Repeat the following until $V = \{a\}$:

Let $v^* \in V \setminus \{a\}$ be such that $p(v^*) = \max_{u \in V \setminus \{a\}} p(u)$ and break the tie arbitrarily.

For each $x \in S(v^*) \subset \mathcal{B}^c$, the order of the pole of the mean first passage time from

$$x \text{ to } \mathcal{B} \text{ is } p(x) \leftarrow p(v^*).$$

Update, for each $u \in V$ such that $(u, v^*) \in E$, $p(u) \leftarrow \max\{p(u), p(v^*) - \mathcal{K}_{uv^*}\}$.

Update $V \leftarrow V \setminus \{v^*\}$ and $E \leftarrow \{(u, v) \in E : u \neq v^* \text{ and } v \neq v^*\}$.

3.10.4 Graphs for the algorithm to find the order of the MFPT

Here we elaborate on the graphs that we use in the algorithm and the definitions that we gave in Section 3.6.1 and SI-Section 3.10.3. While, in the algorithm statement, we used the same notation for the updated graphs as in the original graph, it will be clearer for the justification given in SI-Section 3.10.5 if we specify which copy of the graph we are looking at for each step. Accordingly, we provide a more detailed version of the definitions of these graphs and associated notation in this section.

Step 1: Graph G

For each $\varepsilon \in (0, \varepsilon_{\max})$ and $x \in \mathcal{X}$, the exponential parameter satisfies

$$q_x(\varepsilon) = -Q_{x,x}(\varepsilon) = \sum_{y \neq x \in X} Q_{x,y}(\varepsilon) = \sum_{(x,y) \in E_0} Q_{x,y}(\varepsilon) > 0.$$

Since the order of $Q_{x,y}(\varepsilon)$ is k_{xy} for each $y \in \mathcal{X}$ such that $(x, y) \in E_0$, the order of $q_x(\varepsilon)$ is $p_0(x) = \min\{k_{xy} : (x, y) \in E_0\}$. For each $(x, y) \in E_0$, the transition probability $P_{x,y}(\varepsilon)$ for the embedded discrete time Markov chain is

$$P_{x,y}(\varepsilon) = \frac{Q_{x,y}(\varepsilon)}{q_x(\varepsilon)},$$

the order of which is

$$\mathcal{K}_{xy} = k_{xy} - \min\{k_{xy} : (x, y) \in E_0\} = k_{xy} - p_0(x). \quad (3.10.17)$$

We start with a weighed graph $G = (V, E)$ where $V = \mathcal{X}$ and $E = E_0$. For each $u \in V$, the node weight of u is $p_0(u)$, which is the **order of the pole of the expected sojourn time** at state x until escape from x for X^ε . For each $(u, v) \in E$, the edge weight of (u, v) is $\mathcal{K}_{uv}$, which is the **order of the transition probability** from u to v .

Step 2: Graph $G^{(0)}$

If $x \in \mathcal{B}^c$ is such that $(x, y) \in E$ for some $y \in \mathcal{B}$, then the transition probability from x to $\mathcal{B}$ is positive and is given by

$$P_{x,\mathcal{B}}(\varepsilon) = \sum_{\substack{y \in \mathcal{B}: \\ (x,y) \in E}} P_{x,y}(\varepsilon).$$

For such x , the order of $P_{x,\mathcal{B}}(\varepsilon)$ is

$$\mathcal{K}_{xa} = \min\{\mathcal{K}_{xy} : y \in \mathcal{B}, (x, y) \in E\}, \quad (3.10.18)$$

where $\mathcal{K}_{xy}$ is the order of $P_{x,y}(\varepsilon)$ for each $(x, y) \in E$.

Now, we are ready to specify the graph $G^{(0)} = (V^{(0)}, E^{(0)})$ which serves as the base case for Step 3. We group the nodes in $\mathcal{B}$ into a single node, denoted by a , so the set of nodes becomes $V^{(0)} = (V \setminus \mathcal{B}) \cup \{a\} = \mathcal{B}^c \cup \{a\}$. All of the edges starting from or going to a node in $\mathcal{B}$ are then removed. If there was an edge from $x \in \mathcal{B}^c$ to a node in $\mathcal{B}$, then we add back an edge (x, a) . We leave out all edges from a node in $\mathcal{B}$ to a node in $\mathcal{B}^c$, since we are interested in the mean first passage time to the set $\mathcal{B}$. The resulting edge set is $E^{(0)} = \{(u, v) \in E : u \in \mathcal{B}^c \text{ and } v \in \mathcal{B}^c\} \cup \{(w, a) : (w, v) \in E \text{ for some } w \in \mathcal{B}^c \text{ and } v \in \mathcal{B}\}$.

Let $S(u) = \{u\}$ for all $u \in V^{(0)} \setminus \{a\}$ and $S(a) = \mathcal{B}$. Note that $\{S(u) : u \in V^{(0)}\}$ is a partition of the state space $\mathcal{X}$, denoting the grouping of nodes in $V^{(0)}$. For each $u \in V^{(0)} \setminus \{a\}$ and $x \in S(u) = \{u\}$, we define $p_0^x(u)$ to be the **order of the pole of the expected sojourn time** in $S(u)$ before exiting $S(u)$ when starting the process at the state x . For each $(u, v) \in E^{(0)}$ and $x \in S(u) = \{u\}$, we define $\mathcal{K}_{uv}^x$ to be the **order of the probability of a transition** to $S(v)$ upon exiting from $S(u)$ when the process is started at the state x . For these terms, $p_0^x(u) = p_0(u)$ and $\mathcal{K}_{uv}^x = \mathcal{K}_{uv}$, which is the base case for SI-Lemma 3.10.9.

Step 3: Graphs $\{G^{(N)}\}_{N=0}^M$

In Step 3, we define a sequence of graphs $\{G^{(N)} = (V^{(N)}, E^{(N)})\}_{N=0}^M$ recursively, where the exact value of $M \geq 0$ is not pre-determined and is only revealed when an exit condition for the recursion is satisfied. We know this recursion will end after a finite number of iterations because the number of nodes in $V^{(N)}$ is strictly decreasing with N . The weight p_0 of each node and the weight $\mathcal{K}$ of each edge are also defined iteratively, and each is defined only once.

We have already defined $G^{(0)}$ in Step 2. Fix $N \in \{1, 2, \dots, M+1\}$, where the value of $M < \infty$ is defined below. At the N^{th} iteration, an edge $(u, v) \in E^{(N-1)}$ is called an **r-edge** if its edge weight $\mathcal{K}_{uv}$ is 0; a directed path in $G^{(N-1)}$ is called an **r-path** if it consists of r-edges only. A set $C \subset V^{(N-1)}$ is called an **r-connected set** in $G^{(N-1)}$ if

$|C| > 1$ and there exists an r-path from u to v for any $u \neq v \in C$. Here we use the qualifier “r” to indicate that these edges, paths and cycles are “regular”. If there is no r-connected set in $G^{(N-1)}$, the iteration stops. We set the value of M to the first value of $N - 1$ such that $G^{(N-1)}$ does not have any r-connected set. At that time point, the iteration stops and we move to Step 4 where $G^{(M)}$ will be the initial graph for Step 4. Otherwise, $N \in \{1, \dots, M\}$, and we let C_N be an r-connected set in $G^{(N-1)}$, which is condensed to a new node c_N in $G^{(N)}$. Then, we define the graph $G^{(N)} = (V^{(N)}, E^{(N)})$, where $V^{(N)} = (V^{(N-1)} \setminus C_N) \cup \{c_N\}$, and $E^{(N)} = \{(u, v) \in E^{(N-1)} : u \notin C_N \text{ and } v \notin C_N\} \cup \{(c_N, w) : (u, w) \in E^{(N-1)} \text{ for some } u \in C_N \text{ and } w \notin C_N\} \cup \{(w, c_N) : (w, v) \in E^{(N-1)} \text{ for some } w \notin C_N \text{ and } v \in C_N\}$. Let $S(c_N) = \cup_{u \in C_N} S(u)$. Note that $\{S(u) : u \in V^{(N)}\}$ is again a partition of the state space $\mathcal{X}$, denoting the grouping of nodes in $V^{(N)}$.

We define $p_0^x(c_N)$ to be the **order of the pole of the expected sojourn time** in $S(c_N)$ until the first exit from $S(c_N)$ when the process is started at the state $x \in S(c_N)$. In SI-Lemma 3.10.9, we will show that the value of $p_0^x(c_N)$ is independent of the state $x \in S(c_N)$, and we define $p_0(c_N) = p_0^x(c_N)$. For each $w \in V^{(N-1)} \setminus C_N$ such that $(w, c_N) \in E^{(N)}$, define $\mathcal{K}_{wc_N}^x$ to be the **order of the probability of a transition** to $S(c_N)$ upon exiting from $S(w)$ when the process is started at the state $x \in S(w)$. In SI-Lemma 3.10.9, we will show that the value of $\mathcal{K}_{wc_N}^x$ is independent of the state $x \in S(w)$, and $\mathcal{K}_{wc_N}^x = \mathcal{K}_{wc_N}$ where

$$\mathcal{K}_{wc_N} = \min\{\mathcal{K}_{wv} : v \in C_N \text{ and } (w, v) \in E^{(N-1)}\}. \quad (3.10.19)$$

For each $w \in V^{(N-1)} \setminus C_N$ such that $(c_N, w) \in E^{(N)}$, define $\mathcal{K}_{c_N w}^x$ to be the **order of the probability of a transition** to $S(w)$ upon exiting from $S(c_N)$ when the process is started at the state $x \in S(c_N)$. In SI-Lemma 3.10.9, we will show that the value of $\mathcal{K}_{c_N w}^x$

is independent of the state $x \in S(c_N)$, and $\mathcal{K}_{c_N w}^x = \mathcal{K}_{c_N w}$ where

$$\begin{aligned} \mathcal{K}_{c_N w} &= \min\{\mathcal{K}_{uw} : u \in C_N \text{ and } (u, w) \in E^{(N-1)}\} \\ &\quad - \min\{\mathcal{K}_{uv} : u \in C_N, v \notin C_N \text{ and } (u, v) \in E^{(N-1)}\}. \end{aligned} \quad (3.10.20)$$

We note that for each $N \in \{1, \dots, M\}$, since there is no edge in $E^{(N-1)}$ that leads from a , the node a is never part of any r-connected set, and so $a \in V^{(N)}$ and there is at least one other node in $V^{(N)}$ besides a . Also, the irreducibility of X^ε when $0 < \varepsilon < \varepsilon_0$ implies that there is a path from x to y in G for each $x \in \mathcal{X} \setminus \mathcal{B}$ and $y \in \mathcal{X}$. This implies that if $u_N \neq v_N \in V^{(N)}$ for some $N \in \{0, 1, \dots, M\}$ such that $x \in S(u_N)$ and $y \in S(v_N)$, then there is a path from u_N to v_N in $G^{(N)}$. Therefore, for each $N \in \{0, 1, \dots, M\}$, there is always an outgoing edge from some $u' \in C_N$ (u' cannot be a) to some $v' \in V^{(N-1)} \setminus C_N$ in $G^{(N-1)}$. In addition, as can be seen from the definition of the $\mathcal{K}$'s in (3.10.17)–(3.10.20), for each $N = 0, 1, \dots, M$ and $u'' \in V^{(N)} \setminus \{a\}$, there exists an r-edge $(u'', v'') \in E^{(N)}$ for some $v'' \in V^{(N)}$. For $G^{(M)}$, $|V^{(M)}| \geq 2$. Furthermore, if we only look at r-edges (and ignore the other edges), $G^{(M)}$ is an acyclic graph (as it contains no r-connected set). It follows that the node a is the only sink because for each $u \in V^{(M)} \setminus \{a\}$, there is an outgoing r-edge, and thus there is an r-path from u to a for each $u \in V^{(M)} \setminus \{a\}$.

Step 4: Graphs $\{G^{(M,N)}\}_{N=0}^{|V^{(M)}|-1}$

In Step 4, we define a sequence of graphs $\{G^{(M,N)} = (V^{(M,N)}, E^{(M,N)})\}_{N=0}^{|V^{(M)}|-1}$ recursively, where $G^{(M,0)} = G^{(M)}$. In each iteration, the weight of one of the nodes in $V^{(M,N-1)} \subset V^{(M)}$ is finalized and determines the value of p there, and the weights p_{N-1} of other nodes in $V^{(M,N)}$ are updated to p_N .

Fix $N \in \{1, \dots, |V^{(M)}| - 1\}$. At the N^{th} iteration, let $v_N \in V^{(M,N-1)} \setminus \{a\}$ be such that

$$p_{N-1}(v_N) = \max_{u \in V^{(M,N-1)} \setminus \{a\}} p_{N-1}(u) =: p(v_N), \quad (3.10.21)$$

where we break the tie arbitrarily.

Now, we define the graph $G^{(M,N)} = (V^{(M,N)}, E^{(M,N)})$ where $V^{(M,N)} = V^{(M,N-1)} \setminus \{v_N\}$ and $E^{(M,N)} = \{(u, v) \in E^{(M,N-1)} : u \neq v_N \text{ and } v \neq v_N\}$. For each $u \in V^{(M,N)}$, let

$$p_N(u) = \begin{cases} \max\{p_{N-1}(u), p_{N-1}(v_N) - \mathcal{K}_{uv_N}\}, & \text{for } (u, v_N) \in E^{(M,N-1)}, \\ p_{N-1}(u), & \text{for } (u, v_N) \notin E^{(M,N-1)}. \end{cases} \quad (3.10.22)$$

In SI-Theorem 3.10.2, we will show that for each $x \in S(v_N)$, the order of the pole of the mean first passage time from x to the set $\mathcal{B}$ is $p(x) = p(v_N)$.

3.10.5 Justification for the algorithm to find the order of the MFPT

Recall that we defined $\{G^{(N)}\}_{N=0}^M = \{(V^{(N)}, E^{(N)})\}_{N=0}^M$ and $G^{(M,0)} = G^{(M)}$ in SI-Section 3.10.4. Each $G^{(N)}$ defines a partition $\{S(u) : u \in V^{(N)}\}$ of $\mathcal{X}$, which will be used in our proofs. Since X^ε is an irreducible continuous time Markov chain for all $\varepsilon \in (0, \varepsilon_0)$, each $G^{(N)}$ is weakly connected and has the property that any node in $V^{(N)} \setminus \{a\}$ has an out-going edge starting from the node.

In this section, we will provide the justification for the algorithm. Step 1 of the algorithm sets up the original continuous time Markov chain using a skeleton chain. Step 2 of the algorithm serves as the base case for Step 3, and SI-Lemma 3.10.9 justifies Steps 2 and 3. SI-Theorem 3.10.2 shows that Step 4 works, which gives our main result for the order of the pole of the mean first passage time from each state $x \in \mathcal{B}^c$ to $\mathcal{B}$.

We will start with describing in more detail the Big Theta notation used in this section and defining some useful stopping times that will be used in our proof.

More on Big Theta notation

In Section 3.6.1, we have defined orders for analytic functions using Big Theta notation. Here we give a few more definitions and remarks for inequalities involving the

Big Theta notation, on how to compare the orders of analytic functions and on arithmetic for orders. These conventions streamline the proofs in the following subsections.

Definition 3.10.1. Given $\varepsilon_0 > 0$ and a function $f : (0, \varepsilon_0) \rightarrow \mathbb{R}_{>0}$, we say $f \leq \Theta(\varepsilon^k)$ if there exist $k \in \mathbb{Z}$ and a strictly positive $M_f \in \mathbb{R}_{>0}$ such that, for all $0 < \varepsilon < \varepsilon_0$,

$$f(\varepsilon) \leq M_f \varepsilon^k.$$

We say $f \geq \Theta(\varepsilon^k)$ if there exist $k \in \mathbb{Z}$ and a strictly positive $m_f \in \mathbb{R}_{>0}$ such that, for all $0 < \varepsilon < \varepsilon_0$,

$$f(\varepsilon) \geq m_f \varepsilon^k.$$

Remark 3.10.2. Let $k, k_1, k_2 \in \mathbb{Z}$ and $k_1 \leq k \leq k_2$. If $f = \Theta(\varepsilon^k)$, then $f \leq \Theta(\varepsilon^{k_1})$ and $f \geq \Theta(\varepsilon^{k_2})$.

Remark 3.10.3. For functions f and g mapping $(0, \varepsilon_0)$ into $\mathbb{R}_{>0}$, we write $f = g \cdot \Theta(\varepsilon^k)$ if $\frac{f}{g} = \Theta(\varepsilon^k)$, $f \leq g \cdot \Theta(\varepsilon^k)$ if $\frac{f}{g} \leq \Theta(\varepsilon^k)$, $f \geq g \cdot \Theta(\varepsilon^k)$ if $\frac{f}{g} \geq \Theta(\varepsilon^k)$.

Lemma 3.10.8. Let $k_1, k_2 \in \mathbb{Z}$, $\varepsilon_0 > 0$ and $f, g : (0, \varepsilon_0) \rightarrow \mathbb{R}_{>0}$. If $f = \Theta(\varepsilon^{k_1})$ and $g = \Theta(\varepsilon^{k_2})$, then

$$\frac{1}{f} = \Theta(\varepsilon^{-k_1}), \quad f + g = \Theta(\varepsilon^{\min\{k_1, k_2\}}), \quad f \cdot g = \Theta(\varepsilon^{k_1+k_2}),$$

$$\max\{f, g\} = \Theta(\varepsilon^{\min\{k_1, k_2\}}), \quad \min\{f, g\} = \Theta(\varepsilon^{\max\{k_1, k_2\}}).$$

We leave the proof of SI-Lemma 3.10.8 to the reader.

Stopping times $\tau_n^{\varepsilon, N}$

For each graph $G^{(N)} = (V^{(N)}, E^{(N)})$, $N \in \{0, 1, \dots, M\}$, recall that $\{S(u) : u \in V^{(N)}\}$ is a partition of the state space $\mathcal{X}$. We define the series of stopping times $\{\tau_n^{\varepsilon, N}\}_{n=0}^\infty$, which captures times of transitions of X^ε between sets in the partition $\{S(u) : u \in V^{(N)}\}$

of $\mathcal{X}$. Formally, we let $\tau_0^{\varepsilon, N} = 0$, and for $n = 1, 2, \dots$, we successively define

$$\tau_n^{\varepsilon, N} = \inf \left\{ t \geq \tau_{n-1}^{\varepsilon, N} : X^\varepsilon(t) \notin S(v_{n-1}) \right\}, \quad (3.10.23)$$

where v_{n-1} is the element in $V^{(N)}$ such that $X^\varepsilon(\tau_{n-1}^{\varepsilon, N}) \in S(v_{n-1})$.

Justification for Step 3 of the algorithm

Lemma 3.10.9. (i) For $N = 0$ in Step 3,

(a) for each $u \in V^{(0)} \setminus \{a\}$ and $x \in S(u)$, $\mathbb{E}_x[\tau_1^{\varepsilon, 0}] = \Theta(\varepsilon^{-p_0(u)})$.

(b) for each $(u, v) \in E^{(0)}$ and $x \in S(u)$, $\mathbb{P}_x[X^\varepsilon(\tau_1^{\varepsilon, 0}) \in S(v)] = \Theta(\varepsilon^{\mathcal{K}_{uv}})$.

(ii) For $N \in \{1, 2, \dots, M\}$ in Step 3, let

$$k = \min\{\mathcal{K}_{uv} : u \in C_N, v \notin C_N \text{ and } (u, v) \in E^{(N-1)}\}.$$

(We note that k depends on N although we will not indicate that in the notation.)

(a) For each $x \in S(c_N)$, $\mathbb{E}_x[\tau_1^{\varepsilon, N}] = \Theta(\varepsilon^{-p_0^x(c_N)})$ where $p_0^x(c_N) = p_0(c_N)$ and

$$p_0(c_N) = \max\{p_0(u) : u \in C_N\} + k,$$

(b) For each $x \in S(c_N)$ and $w \in V^{(N-1)} \setminus C_N$ such that $(u, w) \in E^{(N-1)}$ for some $u \in C_N$, $\mathbb{P}_x[X^\varepsilon(\tau_1^{\varepsilon, N}) \in S(w)] = \Theta(\varepsilon^{\mathcal{K}_{c_N w}^x})$ where $\mathcal{K}_{c_N w}^x = \mathcal{K}_{c_N w}$ and

$$\mathcal{K}_{c_N w} = \min\{\mathcal{K}_{uw} : u \in C_N \text{ and } (u, w) \in E^{(N-1)}\} - k,$$

(c) For each $x \in S(w)$ where $w \in V^{(N-1)} \setminus C_N$ is such that $(w, v) \in E^{(N-1)}$ for

some $v \in C_N$, $\mathbb{P}_x[X^\varepsilon(\tau_1^{\varepsilon,N}) \in S(c_N)] = \Theta(\varepsilon^{\mathcal{K}_{wc_N}^x})$ where $\mathcal{K}_{wc_N}^x = \mathcal{K}_{wc_N}$ and

$$\mathcal{K}_{wc_N} = \min\{\mathcal{K}_{wv} : v \in C_N \text{ and } (w, v) \in E^{(N-1)}\}.$$

Proof. Our proof proceeds by induction. The base case ($N = 0$) has been established in SI-Section 3.10.4.

For fixed $1 \leq N \leq M$, assume that (i) (a)–(b) and (ii) (a)–(c) hold with N replaced by $0, 1, \dots, N-1$. We abbreviate $\tau_n^{\varepsilon, N-1}$ as τ_n^ε for $n = 0, 1, 2, \dots$. Let

$$\delta_{out}(C_N) = \{(u, v) \in E^{(N-1)} : u \in C_N \text{ and } v \notin C_N\}$$

denote all out-going boundary edges of C_N so that

$$k = \min\{\mathcal{K}_{uv} : (u, v) \in \delta_{out}(C_N)\}.$$

First, consider the discrete time process $\{X^\varepsilon(\tau_n^\varepsilon)\}_{n=0}^\infty$, which is not necessarily a Markov process. We will derive a lower bound and an upper bound for

$$\mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_n^\varepsilon) \in S(c_N) \text{ for } 0 \leq m \leq n\}} \right], \quad (3.10.24)$$

which is the expected amount of time that $\{Y_n^\varepsilon = X^\varepsilon(\tau_n^\varepsilon)\}_{n=0}^\infty$ spends in $S(c_N)$ before exiting from there, when started from a fixed state $x \in S(c_N)$.

For the lower bound, let

$$\rho_1 = \max_{y \in S(c_N)} \mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \notin S(c_N)],$$

the maximum over $y \in S(c_N)$ of the probability that for X^ε started at y , when X^ε exits $S(u_y)$, where $u_y \in C_N$ such that $y \in S(u_y)$, X^ε exits outside of $S(c_N)$. By the induction

hypothesis, $\mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(v)] = \Theta(\varepsilon^{\mathcal{K}_{uv}})$ for each u, v such that $(u, v) \in \delta_{out}(C_N)$ and $y \in S(u)$. Thus, using SI-Lemma 3.10.8, we have

$$\rho_1 = \max_{y \in S(c_N)} \sum_{\substack{(u,v) \in \delta_{out}(C_N): \\ y \in S(u)}} \mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(v)] = \Theta(\varepsilon^{\min\{\mathcal{K}_{uv} : (u,v) \in \delta_{out}(C_N)\}}) = \Theta(\varepsilon^k).$$

For $x \in S(c_N)$, let $\phi_n(x) = \mathbb{P}_x[X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m \leq n]$ for $n = 0, 1, 2, \dots$. Then, $\phi_0(x) = 1$, $\phi_1(x) = \mathbb{P}_x[X^\varepsilon(\tau_1^\varepsilon) \in S(c_N)] \geq 1 - \rho_1$, and by the strong Markov property, for $n \geq 2$,

$$\begin{aligned} & \phi_n(x) \\ &= \sum_{y \in S(c_N)} \mathbb{P}_x[X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m \leq n-2; X^\varepsilon(\tau_{n-1}^\varepsilon) = y] \mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(c_N)] \\ &\geq \phi_{n-1}(x)(1 - \rho_1). \end{aligned}$$

Hence, $\phi_n(x) \geq (1 - \rho_1)^n$ for $n = 0, 1, 2, \dots$. Then, for $x \in S(c_N)$,

$$\mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m \leq n\}} \right] = \sum_{n=0}^{\infty} \phi_n(x) \geq \sum_{n=0}^{\infty} (1 - \rho_1)^n = \frac{1}{\rho_1} = \Theta(\varepsilon^{-k}). \quad (3.10.25)$$

and so (3.10.24) is bounded below by $\Theta(\varepsilon^{-k})$.

For the upper bound, let $w_0 \in C_N$ be such that

$$\min\{\mathcal{K}_{w_0 v} : (w_0, v) \in \delta_{out}(C_N)\} = \min\{\mathcal{K}_{uv} : (u, v) \in \delta_{out}(C_N)\} = k.$$

Since the order of the probability $\mathbb{P}_x[X^\varepsilon(\tau_1^\varepsilon) \notin S(c_N)]$ might equal $k' > k$ for some $w \neq w_0 \in C_N$ and $x \in S(w)$, such a smaller order probability of directly exiting $S(c_N)$ from $S(w)$ makes it seem possible that (3.10.24) could be $\Theta(\varepsilon^{-k'})$ for some $k' > k$. Indeed, using a similar approach to the one we used for the lower bound, we can show that (3.10.24) is bounded above by $\Theta(\varepsilon^{\max\{\mathcal{K}_{uv} : (u,v) \in \delta_{out}(C_N)\}}) \geq \Theta(\varepsilon^k)$. However, we would like a more

stringent upper bound. To achieve this, we will show below that from $S(w)$, X^ε can exit $S(c_N)$ at least as quickly by means of a transition from $S(w)$ to $S(w_0)$ via the r -connected set and then from $S(w_0)$ to $V^{(N-1)} \setminus S(c_N)$.

Let $\zeta_0^\varepsilon = 0$, and for $n = 1, 2, \dots$, we successively define

$$\eta_{n-1}^\varepsilon = \inf \left\{ t \geq \zeta_{n-1}^\varepsilon : X^\varepsilon(t) \notin S(v) \text{ where } v \in V^{(N-1)} \text{ and } X^\varepsilon(\zeta_{n-1}^\varepsilon) \in S(v) \right\},$$

$$\zeta_n^\varepsilon = \inf \left\{ t \geq \eta_{n-1}^\varepsilon : X^\varepsilon(t) \in S(w_0) \text{ or } X^\varepsilon(t) \notin S(c_N) \right\}.$$

Note that $\{\zeta_n^\varepsilon\}_{n=0}^\infty$ and $\{\eta_n^\varepsilon\}_{n=0}^\infty$ depend on N . Let

$$\rho_2 = \min_{y \in S(w_0)} \mathbb{P}_y[X^\varepsilon(\zeta_1^\varepsilon) \notin S(w_0)] = \min_{y \in S(w_0)} \mathbb{P}_y[X^\varepsilon(\zeta_1^\varepsilon) \notin S(c_N)],$$

By the induction hypothesis, $\mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(v)] = \Theta(\varepsilon^{\mathcal{K}_{w_0v}})$ for each v such that $(w_0, v) \in \delta_{out}(C_N)$ and $y \in S(w_0)$, and so

$$\rho_2 \geq \min_{y \in S(w_0)} \mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \notin S(c_N)] = \Theta(\varepsilon^{\min\{\mathcal{K}_{w_0v} : (w_0, v) \in \delta_{out}(C_N)\}}) = \Theta(\varepsilon^k), \quad (3.10.26)$$

where the inequality holds since starting from any $y \in S(w_0)$, if $X^\varepsilon(\tau_1^\varepsilon) \notin S(c_N)$, X^ε exits outside of $S(c_N)$ after leaving $S(w_0)$ and so $X^\varepsilon(\zeta_1^\varepsilon) \notin S(w_0)$.

For $x \in S(c_N)$, $\mathbb{P}_x$ -a.s., the sum

$$\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_n^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(w_0)\}}$$

counts the number of distinct visits to $S(w_0)$, including the initial start there if $x \in S(w_0)$, before X^ε escapes from $S(c_N)$. By the definition of the $\{\zeta_n^\varepsilon\}_{n=0}^\infty$, $\mathbb{P}_x$ -a.s., the sum

$$\mathbb{1}_{\{X^\varepsilon(0) \in S(w_0)\}} + \sum_{n=1}^{\infty} \mathbb{1}_{\{X^\varepsilon(\zeta_n^\varepsilon) \in S(w_0) \text{ for } 1 \leq m \leq n\}}$$

counts the same quantity. Thus, for $x \in S(w_0)$, using the strong Markov property and (3.10.26),

$$\begin{aligned}
\psi(x) &:= \mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(w_0)\}} \right] \\
&= 1 + \mathbb{E}_x \left[\sum_{n=1}^{\infty} \mathbb{1}_{\{X^\varepsilon(\zeta_m^\varepsilon) \in S(w_0) \text{ for } 1 \leq m \leq n\}} \right] \\
&= 1 + \mathbb{E}_x \left[\mathbb{1}_{\{X^\varepsilon(\zeta_1^\varepsilon) \in S(w_0)\}} \mathbb{E}_{X^\varepsilon(\zeta_1^\varepsilon)} \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\zeta_m^\varepsilon) \in S(w_0) \text{ for } 0 \leq m \leq n\}} \right] \right] \\
&\leq 1 + (1 - \rho_2) \max_{y \in S(w_0)} \psi(y). \tag{3.10.27}
\end{aligned}$$

Note that $\max_{y \in S(w_0)} \psi(y) < \infty$ because the state space is finite and X^ε is positive recurrent. Hence, by (3.10.27), $\max_{y \in S(w_0)} \psi(y) \leq \frac{1}{\rho_2}$. Then, for $x \in S(c_N) \setminus S(w_0)$, by the strong Markov property,

$$\psi(x) \leq \mathbb{P}_x[X^\varepsilon(\zeta_1^\varepsilon) \in S(w_0)] \max_{y \in S(w_0)} \psi(y) \leq \frac{1}{\rho_2}.$$

Thus, for any $x \in S(c_N)$,

$$\mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(w_0)\}} \right] \leq \frac{1}{\rho_2} = \Theta(\varepsilon^{-k}). \tag{3.10.28}$$

Let $w_1, w_2 \in C_N$ be such that $(w_1, w_2) \in E^{(N-1)}$ and $\mathcal{K}_{w_1 w_2} = 0$. By the induction

hypothesis, $\mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(w_2)] = \Theta(1)$ for all $y \in S(w_1)$. Then, for $x \in S(c_N)$,

$$\begin{aligned}
& \mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(w_2)\}} \right] \\
& \geq \sum_{n=0}^{\infty} \mathbb{P}_x[X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(w_1), X^\varepsilon(\tau_{n+1}^\varepsilon) \in S(w_2)] \\
& = \sum_{n=0}^{\infty} \sum_{y \in S(w_1)} \mathbb{P}_x[X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) = y] \cdot \mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(w_2)] \\
& \geq \sum_{n=0}^{\infty} \mathbb{P}_x[X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(w_1)] \cdot \min_{y \in S(w_1)} \mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(w_2)] \\
& = \mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(w_1)\}} \right] \cdot \Theta(1). \tag{3.10.29}
\end{aligned}$$

where the first equality holds from the strong Markov property of X^ε . Since C_N is an r -connected set, we can start from the node w_1 and the order inequality (3.10.29) can be passed from node to node in C_N and back to the node w_1 (w_0 is included in the path) so that we will actually have equality in (3.10.29) and for all $v \in C_N$,

$$\begin{aligned}
& \mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(v)\}} \right] \\
& = \mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(w_0)\}} \right] \cdot \Theta(1). \tag{3.10.30}
\end{aligned}$$

Therefore, combining (3.10.28) and (3.10.30), and since there are only finitely many nodes in C_N , we can obtain by summing over $v \in C_N$ that (3.10.24) is bounded above by $\Theta(\varepsilon^{-k})$. Combining with (3.10.25), we have that, for $x \in S(c_N)$, (3.10.24) is $\Theta(\varepsilon^{-k})$. Moreover, by (3.10.30), for each $x \in S(c_N)$ and each $v \in C_N$,

$$\mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(v)\}} \right] = \Theta(\varepsilon^{-k}). \tag{3.10.31}$$

To prove (ii) (a), fix $x \in S(c_N)$. By the induction hypothesis, $\mathbb{E}_y[\tau_1^\varepsilon] = \Theta(\varepsilon^{-p_0(u)})$

for each $y \in S(u)$ where $u \in C_N$. Thus, the expected sojourn time in $S(c_N)$ is

$$\begin{aligned}
\mathbb{E}_x[\tau_1^{\varepsilon, N}] &= \mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m \leq n\}} \cdot (\tau_{n+1}^\varepsilon - \tau_n^\varepsilon) \right] \\
&= \sum_{n=0}^{\infty} \sum_{u \in C_N} \sum_{y \in S(u)} \mathbb{E}_x[\mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) = y\}}] \cdot \mathbb{E}_y[\tau_1^\varepsilon - \tau_0^\varepsilon] \\
&= \sum_{u \in C_N} \mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(u)\}} \right] \cdot \Theta(\varepsilon^{-p_0(u)}) \\
&= \sum_{u \in C_N} \Theta(\varepsilon^{-k-p_0(u)}) = \Theta(\varepsilon^{-k-\max\{p_0(u): u \in C_N\}}) = \Theta(\varepsilon^{-p_0(c_N)}).
\end{aligned}$$

where the first equality holds from the strong Markov property of X^ε , we used (3.10.31) for the third equality, and we used SI-Lemma 3.10.8 for the fourth equality.

To prove (ii) (b), fix $x \in S(c_N)$ and $w \in V^{(N-1)} \setminus C_N$. By the induction hypothesis, $\mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(w)] = \Theta(\varepsilon^{\mathcal{K}_{uw}})$ for each $y \in S(u)$ where $u \in C_N$. Thus, starting from x , the probability of exiting $S(c_N)$ by means of a transition from a state in $S(c_N)$ to a state in $S(w)$ is given by

$$\begin{aligned}
\mathbb{P}_x[X^\varepsilon(\tau_1^{\varepsilon, N}) \in S(w)] &= \sum_{n=0}^{\infty} \mathbb{P}_x[X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m \leq n; X^\varepsilon(\tau_{n+1}^\varepsilon) \in S(w)] \\
&= \sum_{n=0}^{\infty} \sum_{u \in C_N} \sum_{y \in S(u)} \mathbb{E}_x[\mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) = y\}}] \cdot \mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(w)] \\
&= \sum_{\substack{u \in C_N: \\ (u, w) \in E^{(N-1)}}} \mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(u)\}} \right] \cdot \Theta(\varepsilon^{\mathcal{K}_{uw}}) \\
&= \sum_{\substack{u \in C_N: \\ (u, w) \in E^{(N-1)}}} \Theta(\varepsilon^{-k+\mathcal{K}_{uw}}) = \Theta(\varepsilon^{-k+\min\{\mathcal{K}_{uw}: u \in C_N \text{ and } (u, w) \in E^{(N-1)}\}}) = \Theta(\varepsilon^{\mathcal{K}_{c_N w}}),
\end{aligned}$$

where we used (3.10.31) for the third equality, and for the second equality, we used the fact that there must be an edge in $E^{(N-1)}$ between u and w if $\mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(w)] > 0$ for some and hence all $y \in S(u)$.

To prove (ii) (c), fix $x \in S(w)$ where $w \in V^{(N-1)} \setminus C_N$. By the induction hypothesis,

$\mathbb{P}_x[X^\varepsilon(\tau_1^\varepsilon) \in S(v)] = \Theta(\varepsilon^{\mathcal{K}_{wv}})$ for each $v \in C_N$. Thus, starting from x , the probability of entering $S(c_N)$ by means of a transition from a state in $S(w)$ to a state in $S(c_N)$ is

$$\begin{aligned} \mathbb{P}_x[X^\varepsilon(\tau_1^{\varepsilon,N}) \in S(c_N)] &= \sum_{v \in C_N} \mathbb{P}_x[X^\varepsilon(\tau_1^\varepsilon) \in S(v)] \\ &= \sum_{\substack{v \in C_N: \\ (w,v) \in E^{(N-1)}}} \Theta(\varepsilon^{\mathcal{K}_{wv}}) = \Theta(\varepsilon^{\min\{\mathcal{K}_{wv}: v \in C_N \text{ and } (w,v) \in E^{(N-1)}\}}) = \Theta(\varepsilon^{\mathcal{K}_{wc_N}}). \end{aligned}$$

□

Justification for Step 4 of the algorithm

Lemma 3.10.10. *Fix $w \in V^{(M,0)} \setminus \{a\}$. Let $\tau_n^\varepsilon = \tau_n^{\varepsilon,M}$, for $n = 0, 1, 2, \dots$, as defined in (3.10.23). Then, starting from $x \in S(w)$, the expected number of distinct visits to $S(w)$, including the initial start there, before X^ε enters $S(a)$ is*

$$\mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \notin S(a) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(w)\}} \right] = \Theta(1).$$

Proof. Let $\zeta_0^\varepsilon = 0$, and for $n = 1, 2, \dots$, successively define

$$\eta_{n-1}^\varepsilon = \inf \{t \geq \zeta_{n-1}^\varepsilon : X^\varepsilon(t) \notin S(v) \text{ where } v \in V^{(M,0)} \text{ and } X^\varepsilon(\zeta_{n-1}^\varepsilon) \in S(v)\},$$

$$\zeta_n^\varepsilon = \inf \{t \geq \eta_{n-1}^\varepsilon : X^\varepsilon(t) \in S(w) \cup S(a)\}.$$

Note that for $x \in S(w)$, $\mathbb{P}_x$ -a.s.,

$$\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \notin S(a) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(w)\}} = \sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\zeta_m^\varepsilon) \in S(w) \text{ for } 0 \leq m \leq n\}}, \quad (3.10.32)$$

since they both count the number of distinct visits to $S(w)$, including the initial start there, before X^ε enters $S(a)$.

Recall from SI-Section 3.10.4 that for each $u \in V^{(M,0)} \setminus \{a\}$, there is an r-path from

u to a . Let such an r-path from w to a be $w \rightarrow w_1 \dots \rightarrow w_d \rightarrow a$ where $w, w_1, \dots, w_d, a$ are distinct. By definition, an edge $(u, v) \in V^{(M,0)}$ is an r-edge implies that $\mathbb{P}_z[X^\varepsilon(\tau_1^\varepsilon) \in S(v)] = \Theta(1)$ for all $z \in S(u)$. Thus, for any $y \in S(w)$, using the strong Markov property of X^ε , we have

$$\begin{aligned}
\Theta(1) = 1 &\geq \mathbb{P}_y[X^\varepsilon(\zeta_1^\varepsilon) \notin S(w)] = \mathbb{P}_y[X^\varepsilon(\zeta_1^\varepsilon) \in S(a)] \\
&\geq \mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(w_1), \dots, X^\varepsilon(\tau_d^\varepsilon) \in S(w_d), X^\varepsilon(\tau_{d+1}^\varepsilon) \in S(a)] \\
&= \sum_{z \in S(w_d)} \mathbb{P}_x[X^\varepsilon(\tau_1^\varepsilon) \in S(w_1), \dots, X^\varepsilon(\tau_d^\varepsilon) \in S(w_d), X^\varepsilon(\tau_d^\varepsilon) = z] \cdot \mathbb{P}_z[X^\varepsilon(\tau_1^\varepsilon) \in S(a)] \\
&= \mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(w_1), \dots, X^\varepsilon(\tau_d^\varepsilon) \in S(w_d)] \cdot \Theta(1) = \dots = \Theta(1) \cdot \dots \cdot \Theta(1) = \Theta(1).
\end{aligned}$$

Using a similar approach to that used in SI-Lemma 3.10.9, we can show that for $x \in S(w)$,

$$\mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbf{1}_{\{X^\varepsilon(\zeta_m^\varepsilon) \in S(w) \text{ for } 0 \leq m \leq n\}} \right] \geq \frac{1}{\max_{y \in S(w)} \mathbb{P}_y[X^\varepsilon(\zeta_1^\varepsilon) \notin S(w)]} = \Theta(1),$$

and

$$\mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbf{1}_{\{X^\varepsilon(\zeta_m^\varepsilon) \in S(w) \text{ for } 0 \leq m \leq n\}} \right] \leq \frac{1}{\min_{y \in S(w)} \mathbb{P}_y[X^\varepsilon(\zeta_1^\varepsilon) \notin S(w)]} = \Theta(1).$$

Combining these inequalities with (3.10.32) yields the desired result. $\square$

Theorem 3.10.2. *Let $\tau_{\mathcal{B}}^\varepsilon = \inf\{t \geq 0 : X^\varepsilon(t) \in \mathcal{B}\}$ be the first passage time to $\mathcal{B}$ for X^ε .*

For each $N = 1, \dots, |V^{(M)}| - 1$ and $x \in S(v_N)$, we have

$$\mathbb{E}_x[\tau_{\mathcal{B}}^\varepsilon] = \Theta(\varepsilon^{-p(v_N)}). \tag{3.10.33}$$

Proof. It suffices by iteration to prove that for each fixed $1 \leq N \leq |V^{(M)}| - 1$, if

$$\mathbb{E}_y[\tau_{\mathcal{B}}^\varepsilon] = \Theta(\varepsilon^{-p(v_k)}) \text{ for all } y \in S(v_k) \text{ and } 1 \leq k \leq N - 1, \tag{3.10.34}$$

then (3.10.33) holds for all $x \in S(v_N)$. By convention, (3.10.34) holds automatically for $N = 1$.

For the iteration step, fix $1 \leq N \leq |V^{(M)}| - 1$, and assume that (3.10.34) holds. If $N = 1$, let $A = V^{(M,0)} \setminus \{a\}$, and if $N > 1$, let $A = V^{(M,0)} \setminus \{v_1, \dots, v_{N-1}, a\}$. Recall that for $w \in A$ and $1 \leq k \leq N - 1$, we have $w \in V^{(M,k)}$, and so by (3.10.22),

$$p_k(w) = \begin{cases} \max\{p_{k-1}(w), p_{k-1}(v_k) - \mathcal{K}_{wv_k}\} & \text{for } (w, v_k) \in E^{(M,k-1)}, \\ p_{k-1}(w), & \text{for } (w, v_k) \notin E^{(M,k-1)}. \end{cases} \quad (3.10.35)$$

Note that $(w, v_k) \in E^{(M,k-1)}$ if and only if $(w, v_k) \in E^{(M,0)}$. Since $p_{k-1}(v_k) =: p(v_k)$ for $1 \leq k \leq N - 1$, by iterating (3.10.35), we can obtain

$$p_{N-1}(w) = \max\{p_0(w), \max\{p(v_k) - \mathcal{K}_{wv_k} : 1 \leq k \leq N - 1 \text{ and } (w, v_k) \in E^{(M,0)}\}\}, \quad (3.10.36)$$

where we make the convention that a maximum over an empty set is $-\infty$. In particular, since $v_N \in A$, we have

$$p(v_N) := p_{N-1}(v_N) = \max\{p_0(v_N), \max\{p(v_k) - \mathcal{K}_{v_N v_k} : 1 \leq k \leq N - 1 \text{ and } (v_N, v_k) \in E^{(M,0)}\}\}. \quad (3.10.37)$$

Fix $x \in S(v_N)$. We will derive a lower bound and an upper bound for $\mathbb{E}_x[\tau_B^\varepsilon]$. For the lower bound, let $\tau^\varepsilon = \inf\{t \geq 0 : X^\varepsilon(t) \notin S(v_N)\}$. Recall that $\mathbb{P}_x$ -a.s., $\tau^\varepsilon = \tau_1^{\varepsilon, M}$ as defined in (3.10.23). By SI-Lemma 3.10.9, for each $y \in S(w)$ where $w \in V^{(M,0)} = V^{(M)}$ is such that $(v_N, w) \in E^{(M,0)} = E^{(M)}$, $\mathbb{P}_x[X^\varepsilon(\tau^\varepsilon) = y] = \Theta(\varepsilon^{\mathcal{K}_{v_N w}})$ and $\mathbb{E}_x[\tau^\varepsilon] = \Theta(\varepsilon^{-p_0(v_N)})$.

By first step analysis,

$$\begin{aligned}
\mathbb{E}_x[\tau_{\mathcal{B}}^\varepsilon] &= \mathbb{E}_x[\tau^\varepsilon] + \sum_{(v_N, w) \in E^{(M, 0)}} \sum_{y \in S(w)} \mathbb{P}_x[X^\varepsilon(\tau^\varepsilon) = y] \cdot \mathbb{E}_y[\tau_{\mathcal{B}}^\varepsilon] \\
&\geq \mathbb{E}_x[\tau^\varepsilon] + \sum_{\substack{1 \leq k \leq N-1: \\ (v_N, v_k) \in E^{(M, 0)}}} \sum_{y \in S(v_k)} \mathbb{P}_x[X^\varepsilon(\tau^\varepsilon) = y] \cdot \mathbb{E}_y[\tau_{\mathcal{B}}^\varepsilon] \\
&= \Theta(\varepsilon^{-p_0(v_N)}) + \sum_{\substack{1 \leq k \leq N-1: \\ (v_N, v_k) \in E^{(M, 0)}}} \Theta(\varepsilon^{\mathcal{K}_{v_N v_k}}) \cdot \Theta(\varepsilon^{-p(v_k)}) = \Theta(\varepsilon^{-p_{N-1}(v_N)}). \quad (3.10.38)
\end{aligned}$$

where we used (3.10.34) in the second last equality, and used SI-Lemma 3.10.8 and (3.10.37) for the last equality.

For the upper bound, let $\eta^\varepsilon = \inf\{t \geq 0 : X^\varepsilon(t) \notin \bigcup_{u \in A} S(u)\}$. Let $\tau_n^\varepsilon = \tau_n^{\varepsilon, M}$, for $n = 0, 1, 2, \dots$, as defined in (3.10.23). Then, using SI-Lemma 3.10.10 and the strong Markov property, for $w \in A \subset V^{(M, 0)} \setminus \{a\}$,

$$\begin{aligned}
&\mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_n^\varepsilon) \in \bigcup_{u \in A} S(u) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(w)\}} \right] \\
&\leq \mathbb{P}_x[X^\varepsilon(\zeta^\varepsilon) \in S(w)] \max_{y \in S(w)} \mathbb{E}_y \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_n^\varepsilon) \notin S(a) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(w)\}} \right] \\
&\leq \Theta(1). \quad (3.10.39)
\end{aligned}$$

where $\zeta^\varepsilon = \inf\{t \geq 0 : X^\varepsilon(t) \in S(w)\}$.

For $1 \leq k \leq N - 1$ such that there exists $w \in A$ where $(w, v_k) \in E^{(M,0)}$,

$$\begin{aligned}
\mathbb{P}_x[X^\varepsilon(\eta^\varepsilon) \in S(v_k)] &= \sum_{n=0}^{\infty} \mathbb{P}_x[X^\varepsilon(\tau_n^\varepsilon) \in \bigcup_{u \in A} S(u) \text{ for } 0 \leq m \leq n; X^\varepsilon(\tau_{n+1}^\varepsilon) \in S(v_k)] \\
&= \sum_{n=0}^{\infty} \sum_{w \in A} \sum_{y \in S(w)} \mathbb{P}_x[X^\varepsilon(\tau_n^\varepsilon) \in \bigcup_{u \in A} S(u) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) = y] \cdot \mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(v_k)] \\
&= \sum_{\substack{w \in A: \\ (w, v_k) \in E^{(M,0)}}} \mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in \bigcup_{u \in A} S(u) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(w)\}} \right] \cdot \Theta(\varepsilon^{\mathcal{K}_{wv_k}}) \\
&\leq \sum_{\substack{w \in A: \\ (w, v_k) \in E^{(M,0)}}} \Theta(1) \cdot \Theta(\varepsilon^{\mathcal{K}_{wv_k}}), \tag{3.10.40}
\end{aligned}$$

where the second equality holds from strong Markov property of X^ε , the third equality uses SI-Lemma 3.10.9, and we used (3.10.39) for the last inequality. Using SI-Lemma 3.10.9 and (3.10.39),

$$\begin{aligned}
\mathbb{E}_x[\eta^\varepsilon] &= \mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in \bigcup_{u \in A} S(u) \text{ for } 0 \leq m \leq n\}} \cdot (\tau_{n+1}^\varepsilon - \tau_n^\varepsilon) \right] \\
&= \sum_{n=0}^{\infty} \sum_{w \in A} \sum_{y \in S(w)} \mathbb{P}_x[X^\varepsilon(\tau_m^\varepsilon) \in \bigcup_{u \in A} S(u) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) = y] \cdot \mathbb{E}_y[\tau_1^\varepsilon - \tau_0^\varepsilon] \\
&= \sum_{w \in A} \mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in \bigcup_{u \in A} S(u) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(w)\}} \right] \cdot \Theta(\varepsilon^{-p_0(w)}) \\
&\leq \sum_{w \in A} \Theta(1) \cdot \Theta(\varepsilon^{-p_0(w)}) \leq \Theta(\varepsilon^{-p_{N-1}(v_N)}), \tag{3.10.41}
\end{aligned}$$

where we have used (3.10.36) and (3.10.21) to conclude that $p_0(w) \leq p_{N-1}(w) \leq p_{N-1}(v_N)$

for all $w \in A$. Therefore, using first step analysis, we have

$$\begin{aligned}
\mathbb{E}_x[\tau_{\mathcal{B}}^\varepsilon] &= \mathbb{E}_x[\eta^\varepsilon] + \sum_{1 \leq k \leq N-1} \sum_{y \in S(v_k)} \mathbb{P}_x[X^\varepsilon(\eta^\varepsilon) = y] \cdot \mathbb{E}_y[\tau_{\mathcal{B}}^\varepsilon] \\
&= \mathbb{E}_x[\eta^\varepsilon] + \sum_{1 \leq k \leq N-1} \mathbb{P}_x[X^\varepsilon(\eta^\varepsilon) \in S(v_k)] \cdot \Theta(\varepsilon^{-p(v_k)}) \\
&\leq \Theta(\varepsilon^{-p_{N-1}(v_N)}) + \sum_{\substack{w \in A, \\ 1 \leq k \leq N-1: \\ (w, v_k) \in E^{(M, 0)}}} \Theta(\varepsilon^{\mathcal{K}_{wv_k}}) \cdot \Theta(\varepsilon^{-p(v_k)}) \\
&\leq \Theta(\varepsilon^{-p_{N-1}(v_N)}) + \Theta(\varepsilon^{-\max\{p_{N-1}(w): w \in A\}}) = \Theta(\varepsilon^{-p_{N-1}(v_N)}), \tag{3.10.42}
\end{aligned}$$

where we used (3.10.34) for the second equality, (3.10.40) and (3.10.41) for the first inequality, and (3.10.36) and SI-Lemma 3.10.8 for the second inequality.

By (3.10.38) and (3.10.42), we conclude that $\mathbb{E}_x[\tau_{\mathcal{B}}^\varepsilon] = \Theta(\varepsilon^{-p_{N-1}(v_N)}) = \Theta(\varepsilon^{-p(v_N)})$.

□

3.10.6 Application of the algorithm to the 2D, 3D and 4D models

The algorithm is described in Section 3.6.1, and it finds the order of the pole of the mean first passage time to $\emptyset \neq \mathcal{B} \subset \mathcal{X}$ from each state in $\mathcal{B}^c$. In this section, we will apply the algorithm to the 2D, 3D and 4D models and find the order of the poles of the mean first passage times of interest to the fully repressed state and the fully active state (see SI-Figures 3.8–3.12). For each figure, the “Input” panel shows the order of each of the non-zero off-diagonal entries in $Q(\varepsilon)$ and the set $\mathcal{B}$ which contains a single state, which is either the fully repressed state or the fully active state. The orders of these non-zero entries in $Q(\varepsilon)$ are represented by colored arrows in the graph (red for order 0 and blue for order 1). Step 1 transforms the orders in the infinitesimal generator $Q(\varepsilon)$ into orders for the transition matrix $P(\varepsilon)$ and the exponential parameters $q(\varepsilon)$ to give an equivalent construction for the continuous time Markov chain. The orders of the non-zero entries in

$P(\varepsilon)$ are given by $\mathcal{K}$ and represented by colored arrows in the graph, and the number in the circle at a state $x \in \mathcal{B}^c$ is the order of the pole $p(x)$ of $\frac{1}{q_x(\varepsilon)}$ (the mean sojourn time at the state x). In Step 2, the set $\mathcal{B}$ contains only one state and is just relabeled as the node a . All transitions from a to $\mathcal{B}^c$ are then removed. While the Input, Step 1 and Step 2 are universal across all the figures in this section, we explain the Step 3, Step 4 and Output panels separately for each application below since they are more distinct.

2D model (from the fully active state to the fully repressed state): see SI-Figure 3.8. The explanation of the panels for Input, Step 1 and Step 2 is given above with $\mathcal{B} = \{(D_{\text{tot}}, 0)^T\}$. Step 3 for the 2D model involves only one iteration, where the collection of all nodes except the node a and the origin 0 (called an r-connected set C) is condensed to a single node c , and the order of the pole at c is $p(c) = \max_{u \in C} p(u) + \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u, v) \in E\} = 1 + 0 = 1$, where E denotes the edge set of the graph in Step 3 before the first iteration. Moreover, $\mathcal{K}_{c0} = \min\{\mathcal{K}_{u0} : u \in C \text{ and } (u, 0) \in E\} - \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u, v) \in E\} = 1 - 0 = 1$, $\mathcal{K}_{ca} = \min\{\mathcal{K}_{ua} : u \in C \text{ and } (u, a) \in E\} - \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u, v) \in E\} = 0 - 0 = 0$, and $\mathcal{K}_{0c} = \min\{\mathcal{K}_{0v} : v \in C \text{ and } (0, v) \in E\} = 0$. Step 4 involves two iterations. In the first iteration, we fix the node with the largest value of p , which is c in our case. At any node other than a that is connected to c (i.e., the origin 0), the value of p is updated to $p(0) = \max\{p(0), p(c) - \mathcal{K}_{0c}\} = \max\{0, 1 - 0\} = 1$, and then any edges leading to or from c are removed. In the second iteration, of the remaining nodes, we fix the node with the largest value of p , which is the origin. When all of the nodes other than a have been fixed, the order of the pole of the mean first passage time from each state in $\mathcal{B}^c$ to $\mathcal{B}$ is given by the fixed value of the node to which the state belongs.

2D model (from the fully repressed state to the fully active state) Because of the symmetry in the input graph in SI-Figure 3.8, the orders of the poles of the mean first passage times to the fully repressed state can be obtained in the same way as above.

3D model (from the fully active state to the fully repressed state): see SI-

Figure 3.9. The explanation of the panels for Input, Step 1 and Step 2 is given above with $\mathcal{B} = \{(D_{\text{tot}}, 0, 0)^T\}$. A state represents $(n_{D_{12}^R}, n_{D^A}, n_{D_1^R})^T$. Step 3 involves only one iteration, where the collection of all nodes except for $(0, 0, 0)^T$, $(0, 0, D_{\text{tot}})^T$, $(1, 0, D_{\text{tot}} - 1)^T$, $(2, 0, D_{\text{tot}} - 2)^T, \dots, (D_{\text{tot}} - 2, 0, 2)^T$, and $(D_{\text{tot}} - 1, 0, 1)^T$ (called an r-connected set C) is condensed to a single node c . The order of the pole of the sojourn time at C is $p(c) = \max_{u \in C} p(u) + \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u, v) \in E\} = 1 + 0 = 1$, where E denotes the edge set of the graph in Step 3 before the first iteration. Moreover, $\mathcal{K}_{c, (0, 0, 0)^T} = \min\{\mathcal{K}_{u, (0, 0, 0)^T} : u \in C \text{ and } (u, (0, 0, 0)^T) \in E\} - \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u, v) \in E\} = 1 - 0 = 1$, $\mathcal{K}_{(0, 0, 0)^T, c} = \min\{\mathcal{K}_{(0, 0, 0)^T, v} : v \in C \text{ and } ((0, 0, 0)^T, v) \in E\} = 0$, $\mathcal{K}_{c, (0, 0, D_{\text{tot}})^T} = \min\{\mathcal{K}_{u, (0, 0, D_{\text{tot}})^T} : u \in C \text{ and } (u, (0, 0, D_{\text{tot}})^T) \in E\} - \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u, v) \in E\} = 0 - 0 = 0$, $\mathcal{K}_{(0, 0, D_{\text{tot}})^T, c} = \min\{\mathcal{K}_{(0, 0, D_{\text{tot}})^T, v} : v \in C \text{ and } ((0, 0, D_{\text{tot}})^T, v) \in E\} = 1$, $\dots$, and $\mathcal{K}_{c, (D_{\text{tot}} - 1, 0, 1)^T} = \min\{\mathcal{K}_{u, (D_{\text{tot}} - 1, 0, 1)^T} : u \in C \text{ and } (u, (D_{\text{tot}} - 1, 0, 1)^T) \in E\} - \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u, v) \in E\} = 0 - 0 = 0$, $\mathcal{K}_{(D_{\text{tot}} - 1, 0, 1)^T, c} = \min\{\mathcal{K}_{(D_{\text{tot}} - 1, 0, 1)^T, v} : v \in C \text{ and } ((D_{\text{tot}} - 1, 0, 1)^T, v) \in E\} = 1$. Step 4 involves $(D_{\text{tot}} + 2)$ iterations. In the first iteration, we fix the node with the largest value of p , which is c in our case. At any node u other than a that is connected to c , the value of p is updated according to the formula $p(u) = \max\{p(u), p(c) - \mathcal{K}_{uc}\}$, and then any edges leading to or from c are removed. In the second iteration, the node $(0, 0, 0)^T$ has the largest value of p among the remaining nodes, and thus is fixed. There is no other nodes connected to $(0, 0, 0)^T$ at this point, so we move to the next iteration. In the third iteration, the node $(0, 0, D_{\text{tot}})^T$ is fixed. The node $(1, 0, D_{\text{tot}} - 1)^T$ is connected to it, and thus the $p((1, 0, D_{\text{tot}} - 1)^T)$ is updated to be $\max\{p((1, 0, D_{\text{tot}} - 1)^T), p((0, 0, D_{\text{tot}})^T) - \mathcal{K}_{(1, 0, D_{\text{tot}} - 1)^T, (0, 0, D_{\text{tot}})^T}\} = 0$. Then, any edges leading to or from $(0, 0, D_{\text{tot}})^T$ are removed. The remaining iterations will be similar to the third one. When all of the nodes other than a have been fixed, the order of the pole of the mean first passage time from each state in $\mathcal{B}^c$ to $\mathcal{B}$ is given by the fixed value of the node to which the state belongs.

3D model (from the fully repressed state to the fully active state): see SI-Figure

3.10. The explanation of the panels for Input, Step 1 and Step 2 is given above with $\mathcal{B} = \{(0, D_{\text{tot}}, 0)^T\}$. Step 3 involves two iterations. In the first iteration, the collection of nodes consisting of $(D_{\text{tot}} - 1, 0, 1)^T$ and $(D_{\text{tot}}, 0, 0)^T$ (called an r-connected set C_1) is condensed into a single node c_1 . The order of the pole of the sojourn time in C_1 is $p(c_1) = \max_{u \in C_1} p(u) + \min\{\mathcal{K}_{uv} : u \in C_1, v \notin C_1 \text{ and } (u, v) \in E\} = 1 + 1 = 2$, where E denotes the edge set of the graph in Step 3 before the first iteration. Moreover, $\mathcal{K}_{c_1, (D_{\text{tot}}-2, 0, 2)^T} = \min\{\mathcal{K}_{u, (D_{\text{tot}}-2, 0, 2)^T} : u \in C_1 \text{ and } (u, (D_{\text{tot}} - 2, 0, 2)^T) \in E\} - \min\{\mathcal{K}_{uv} : u \in C_1, v \notin C_1 \text{ and } (u, v) \in E\} = 1 - 1 = 0$, $\mathcal{K}_{(D_{\text{tot}}-2, 0, 2)^T, c_1} = \min\{\mathcal{K}_{(D_{\text{tot}}-2, 0, 2)^T, v} : v \in C_1 \text{ and } ((D_{\text{tot}} - 2, 0, 2)^T, v) \in E\} = 0$, $\mathcal{K}_{c_1, (D_{\text{tot}}-1, 0, 0)^T} = \min\{\mathcal{K}_{u, (D_{\text{tot}}-1, 0, 0)^T} : u \in C_1 \text{ and } (u, (D_{\text{tot}} - 1, 0, 0)^T) \in E\} - \min\{\mathcal{K}_{uv} : u \in C_1, v \notin C_1 \text{ and } (u, v) \in E\} = 1 - 1 = 0$ and $\mathcal{K}_{(D_{\text{tot}}-1, 0, 0)^T, c_1} = \min\{\mathcal{K}_{(D_{\text{tot}}-1, 0, 0)^T, v} : v \in C_1 \text{ and } ((D_{\text{tot}} - 1, 0, 0)^T, v) \in E\} = 0$. In the second iteration of Step 3, the collection of all nodes except for $(0, 0, 0)^T$ and a (called an r-connected set C_2) is condensed to a single node c_2 . The order of the pole of the sojourn time in C_2 is $p(c_2) = \max_{u \in C_2} p(u) + \min\{\mathcal{K}_{uv} : u \in C_2, v \notin C_2 \text{ and } (u, v) \in E\} = 2 + 0 = 2$, where E denotes the edge set of the graph in Step 3 before the second iteration. Moreover, $\mathcal{K}_{c_2, (0, 0, 0)^T} = \min\{\mathcal{K}_{u, (0, 0, 0)^T} : u \in C_2 \text{ and } (u, (0, 0, 0)^T) \in E\} - \min\{\mathcal{K}_{uv} : u \in C_2, v \notin C_2 \text{ and } (u, v) \in E\} = 1 - 0 = 1$, $\mathcal{K}_{c_2, a} = \min\{\mathcal{K}_{ua} : u \in C_2 \text{ and } (u, a) \in E\} - \min\{\mathcal{K}_{uv} : u \in C_2, v \notin C_2 \text{ and } (u, v) \in E\} = 0 - 0 = 0$, and $\mathcal{K}_{(0, 0, 0)^T, c_2} = \min\{\mathcal{K}_{(0, 0, 0)^T, v} : v \in C_2 \text{ and } ((0, 0, 0)^T, v) \in E\} = 0$. Step 4 involves two iterations. In the first iteration, we fix the node with the largest value of p , which is c_2 in our case. At any node other than a that is connected to c_2 (i.e., the origin $(0, 0, 0)^T$), the value of p is updated to $p((0, 0, 0)^T) = \max\{p((0, 0, 0)^T), p(c_2) - \mathcal{K}_{(0, 0, 0)^T, c_2}\} = \max\{0, 2 - 0\} = 2$, and then any edges leading to or from c_2 are removed. In the second iteration, among the remaining nodes, we fix the node with the largest value of p , which is the origin. When all of the nodes other than a have been fixed, the order of the pole of the mean first passage time from each state in $\mathcal{B}^c$ to $\mathcal{B}$ is given by the fixed value of the node to which the state belongs.

4D model (from the fully active state to the fully repressed state): see SI-Figure 3.11. We illustrate how to use the algorithm for the 4D model when $D_{\text{tot}} = 2$; for larger D_{tot} , the methodology will be the same. A state represents $(n_{D_{12}^R}, n_{D^A}, n_{D_1^R}, n_{D_2^R})^T$. The explanation of the panels for Input, Step 1 and Step 2 is given above with $\mathcal{B} = \{(2, 0, 0, 0)^T\}$. Step 3 involves only one iteration, where the collection of all nodes except for $(0, 0, 0, 0)^T$, $(0, 0, 2, 0)^T$, $(0, 0, 1, 1)^T$, $(0, 0, 0, 2)^T$, $(1, 0, 1, 0)^T$ and $(1, 0, 0, 1)^T$ (called an r-connected set C) is condensed to a single node c . The order of the pole of the sojourn time in C is $p(c) = \max_{u \in C} p(u) + \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u, v) \in E\} = 1 + 0 = 1$, where E denotes the edge set of the graph in Step 3 before the first iteration. Moreover, the value of $\mathcal{K}$ for edges between c and an original node w that is not in C are defined according to the formula $\mathcal{K}_{c,w} = \min\{\mathcal{K}_{uw} : u \in C \text{ and } (u, w) \in E\} - \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u, v) \in E\}$, $\mathcal{K}_{w,c} = \min\{\mathcal{K}_{w,v} : v \in C \text{ and } (w, v) \in E\}$. Step 4 involves seven iterations. In the first iteration, we fix the node with the largest value of p , which is c in our case. At any node u other than a that is connected to c , the value of p is updated according to the formula $p(u) = \max\{p(u), p(c) - \mathcal{K}_{uc}\}$, and then any edges leading to or from c are removed. In the second iteration of Step 4, the node $(0, 0, 0, 0)^T$ has the largest value of p among the remaining nodes, and then is fixed. There are no other nodes connected to $(0, 0, 0, 0)^T$ at this point, so we move to the next iteration. In the third iteration, the node $(0, 0, 2, 0)^T$ is fixed. The node $(1, 0, 1, 0)^T$ is connected to it, and thus $p((1, 0, 1, 0)^T)$ is updated to be $\max\{p((1, 0, 1, 0)^T), p((0, 0, 2, 0)^T) - \mathcal{K}_{(1,0,1,0)^T,(0,0,2,0)^T}\} = 0$. Then, any edges leading to or from $(0, 0, 2, 0)^T$ are removed. The remaining iterations will be similar to the third one. When all of the nodes other than a have been fixed, the order of the pole of the mean first passage time from each state in $\mathcal{B}^c$ to $\mathcal{B}$ is given by the fixed value of the node to which the state belongs.

4D model (from the fully repressed state to the fully active state): see SI-Figure 3.12. We again illustrate how to use the algorithm for the 4D model when $D_{\text{tot}} = 2$; for larger D_{tot} , the methodology will be the same. The explanation of the panels for Input, Step

1 and Step 2 is given above with $\mathcal{B} = \{(0, 2, 0, 0)^T\}$. Step 3 involves two iterations. In the first iteration, the collection of the nodes $(1, 0, 1, 0)^T$, $(1, 0, 0, 1)^T$ and $(2, 0, 0, 0)^T$ (called an r-connected set C_1) is condensed into a single node c_1 . The order of the pole of the sojourn time in C_1 is $p(c_1) = \max_{u \in C_1} p(u) + \min\{\mathcal{K}_{uv} : u \in C_1, v \notin C_1 \text{ and } (u, v) \in E\} = 1 + 1 = 2$, where E denotes the edge set of the graph in Step 3 before the first iteration. Moreover, the value of $\mathcal{K}$ of edges between c_1 and an original node w that is not in C_1 are defined if there is an edge between some node $u \in C_1$ and w and according to the formula $\mathcal{K}_{c_1, w} = \min\{\mathcal{K}_{uw} : u \in C_1 \text{ and } (u, w) \in E\} - \min\{\mathcal{K}_{uv} : u \in C_1, v \notin C_1 \text{ and } (u, v) \in E\}$, $\mathcal{K}_{w, c_1} = \min\{\mathcal{K}_{w, v} : v \in C_1 \text{ and } (w, v) \in E\}$. In the second iteration of Step 3, the collection of all nodes except for $(0, 0, 0, 0)^T$ and a (called an r-connected set C_2) is condensed to a single node c_2 . The order of the pole of the sojourn time in C_2 is $p(c_2) = \max_{u \in C_2} p(u) + \min\{\mathcal{K}_{uv} : u \in C_2, v \notin C_2 \text{ and } (u, v) \in E\} = 2 + 0 = 2$, where E denotes the edge set of the graph in Step 3 before the second iteration. Moreover, $\mathcal{K}_{c_2, (0, 0, 0, 0)^T} = \min\{\mathcal{K}_{u, (0, 0, 0, 0)^T} : u \in C_2 \text{ and } (u, (0, 0, 0, 0)^T) \in E\} - \min\{\mathcal{K}_{uv} : u \in C_2, v \notin C_2 \text{ and } (u, v) \in E\} = 1 - 0 = 1$, $\mathcal{K}_{c_2, a} = \min\{\mathcal{K}_{ua} : u \in C_2 \text{ and } (u, a) \in E\} - \min\{\mathcal{K}_{uv} : u \in C_2, v \notin C_2 \text{ and } (u, v) \in E\} = 0 - 0 = 0$, and $\mathcal{K}_{(0, 0, 0, 0)^T, c_2} = \min\{\mathcal{K}_{(0, 0, 0, 0)^T, v} : v \in C_2 \text{ and } ((0, 0, 0, 0)^T, v) \in E\} = 0$. Step 4 involves two iterations. In the first iteration, we fix the node with the largest value of p , which is c_2 in our case. At any node other than a that is connected to c_2 (i.e., the origin $(0, 0, 0, 0)^T$), the value of p is updated to $p((0, 0, 0, 0)^T) = \max\{p((0, 0, 0, 0)^T), p(c_2) - \mathcal{K}_{(0, 0, 0, 0)^T, c_2}\} = \max\{0, 2 - 0\} = 2$, and then any edges leading to or from c_2 are removed. In the second iteration, of the remaining nodes, we fix the node with the largest value of p , which is the origin. When all of the nodes other than a have been fixed, the order of the pole of the mean first passage time from each state in $\mathcal{B}^c$ to $\mathcal{B}$ is given by the fixed value of the node to which the state belongs.

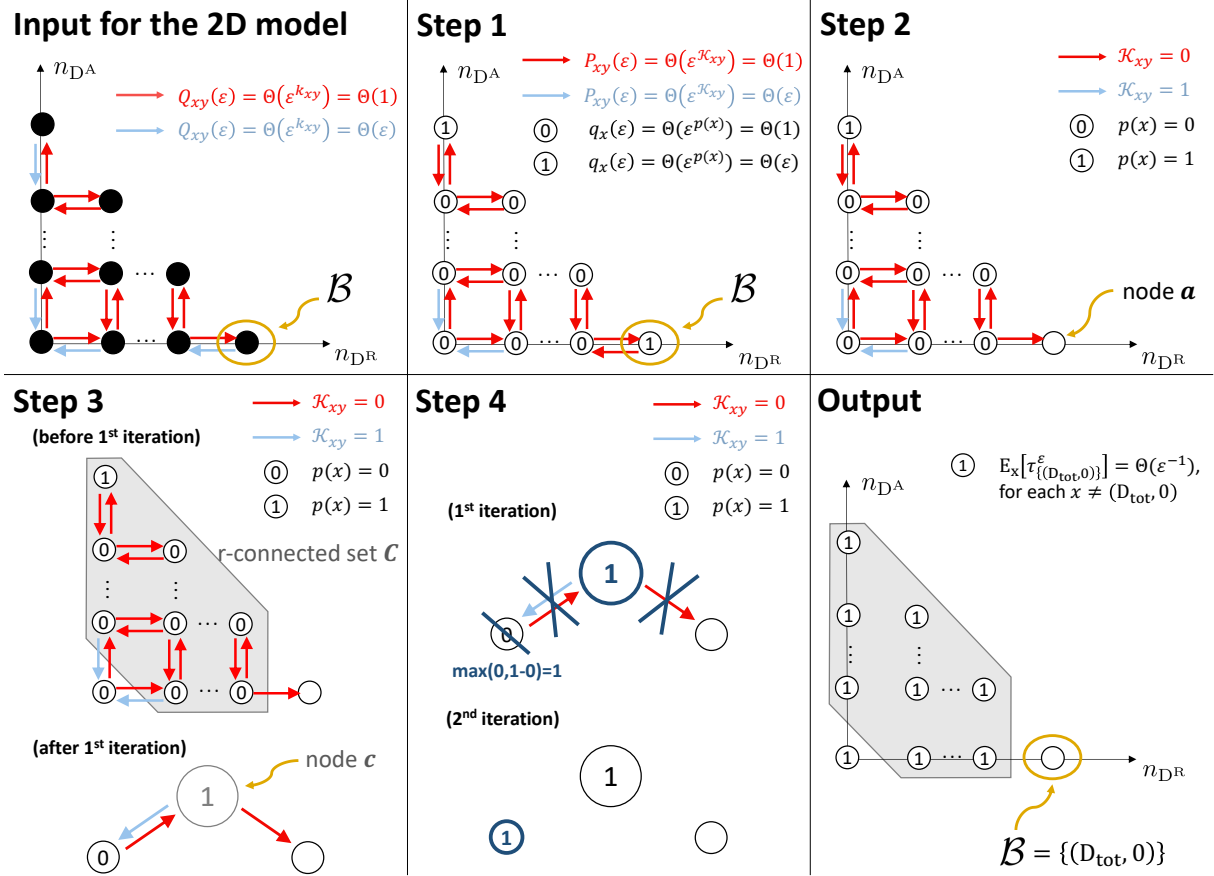
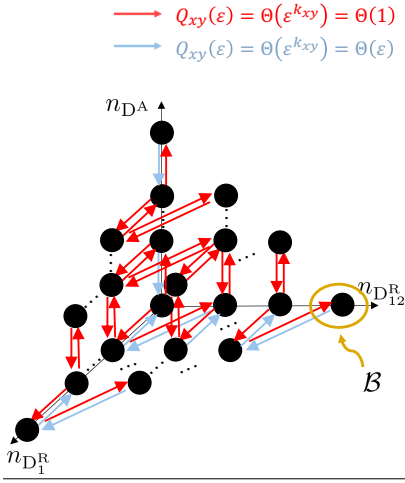
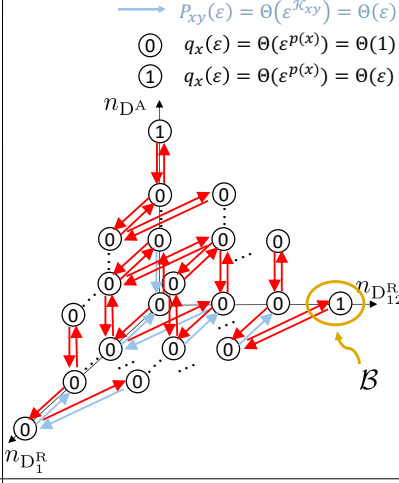


Figure 3.8. Key steps of the algorithm for the 2D model.

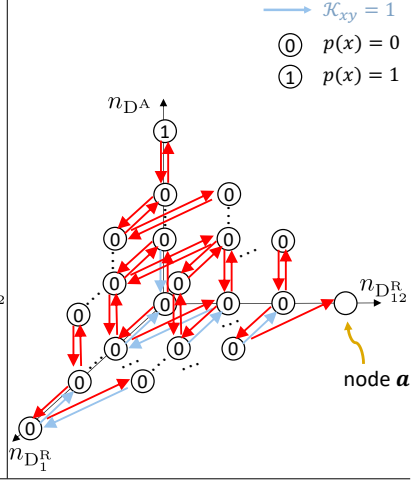
Input for the 3D model



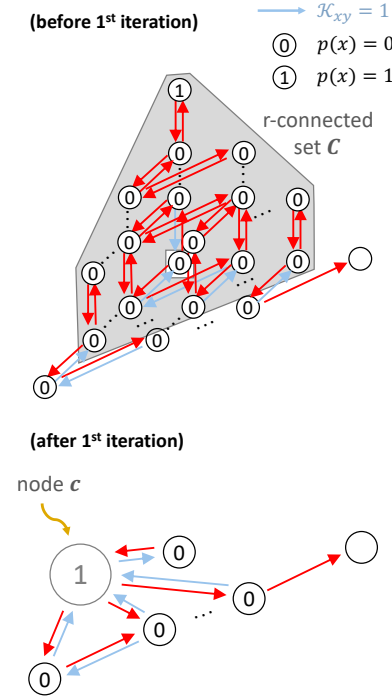
Step 1



Step 2



Step 3



Step 4

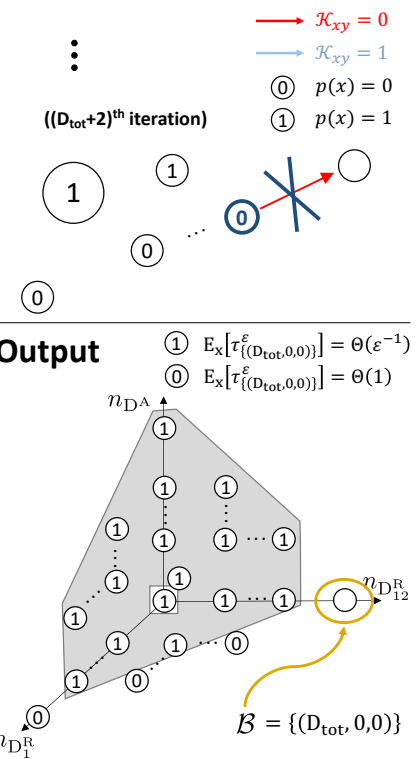
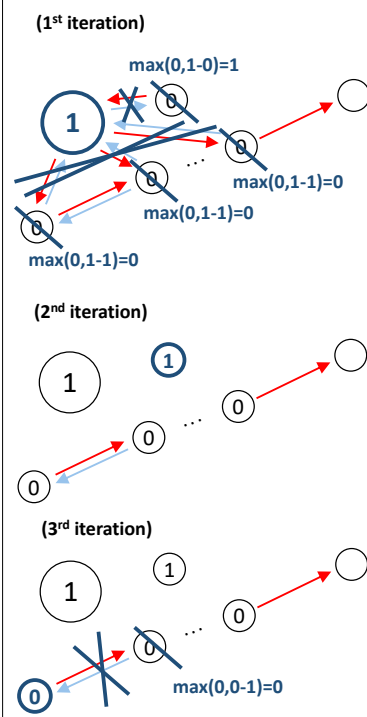
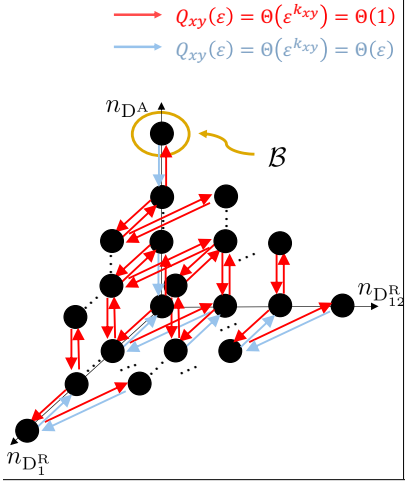
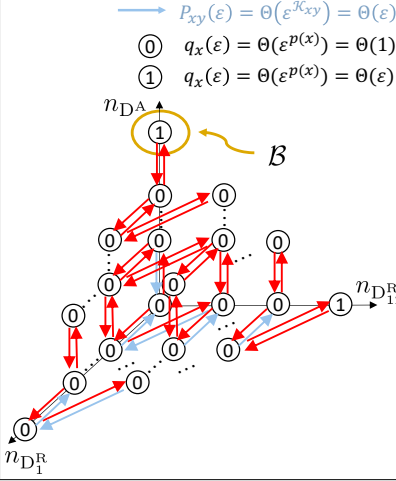


Figure 3.9. Key steps of the algorithm for the 3D model (from the fully active state to the fully repressed state).

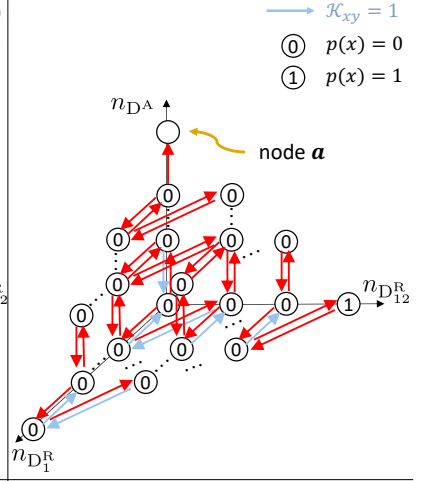
Input for the 3D model



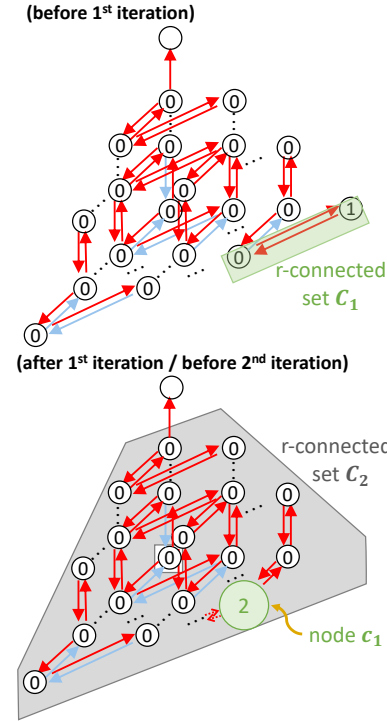
Step 1



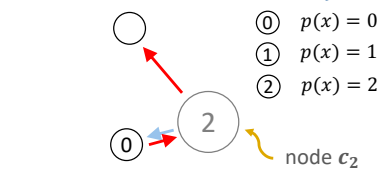
Step 2



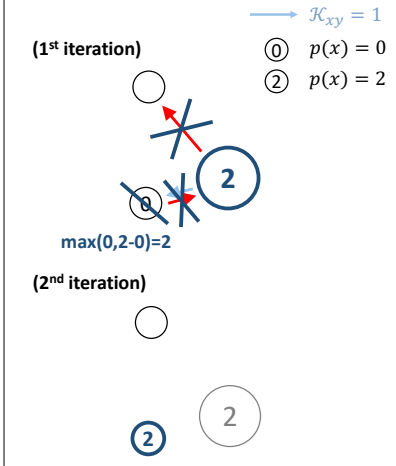
Step 3



(after 2nd iteration)



Step 4



Output

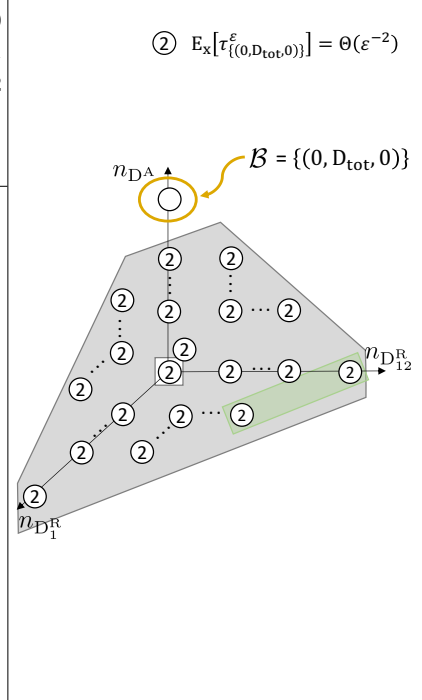
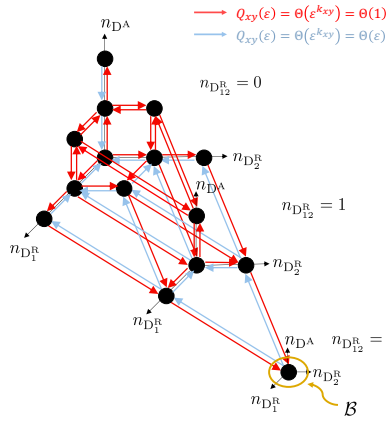
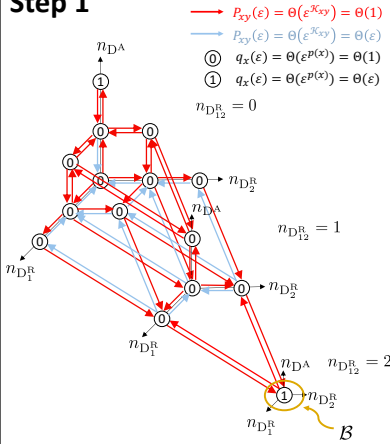


Figure 3.10. Key steps of the algorithm for the 3D model (from the fully repressed state to the fully active state).

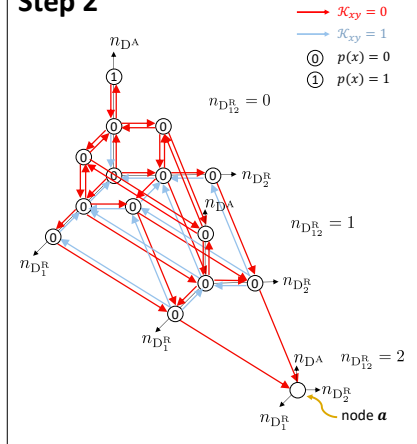
Input for the 4D model



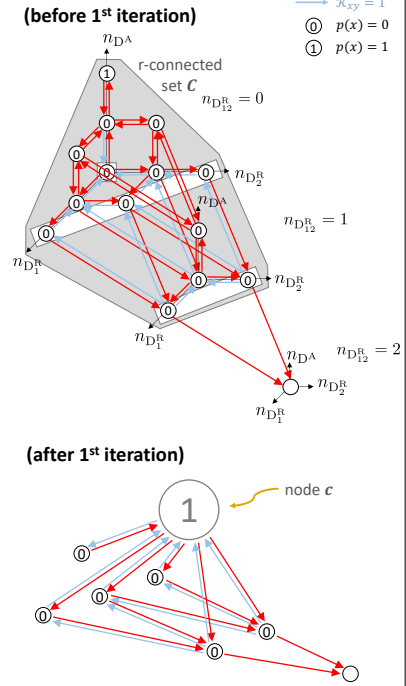
Step 1



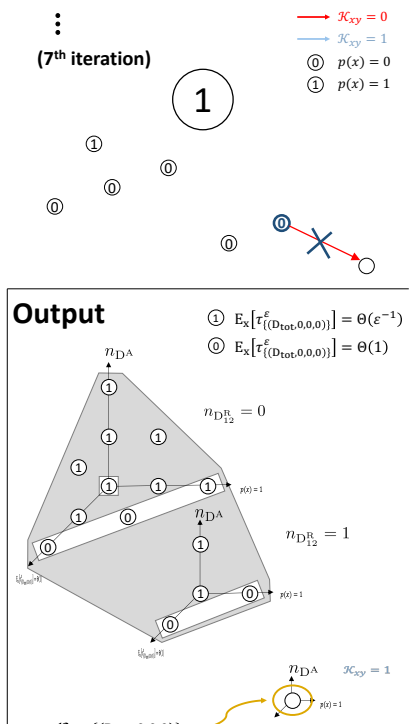
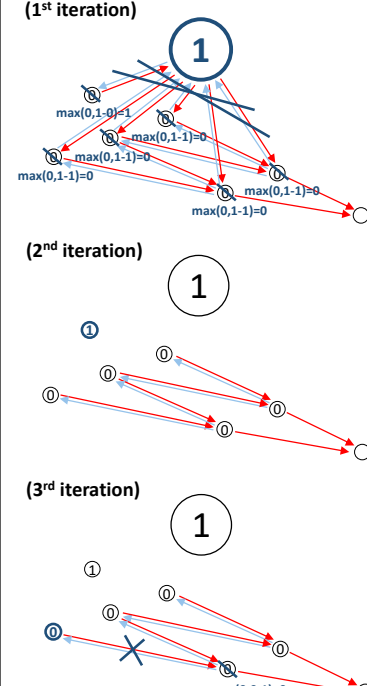
Step 2



Step 3



Step 4



Output

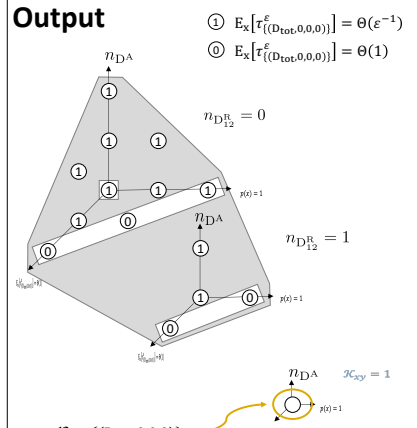
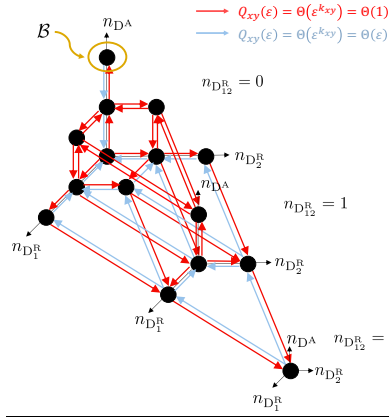
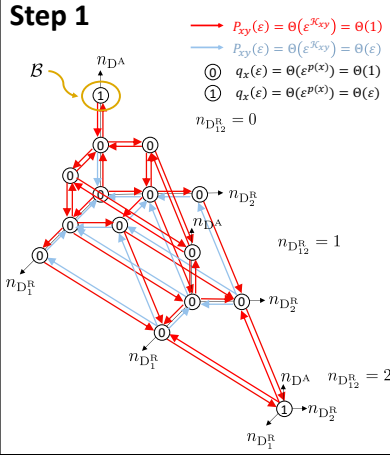


Figure 3.11. Key steps of the algorithm for the 4D model (from the fully active state to the fully repressed state).

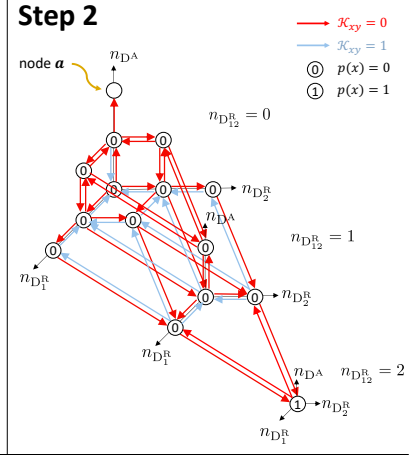
Input for the 4D model



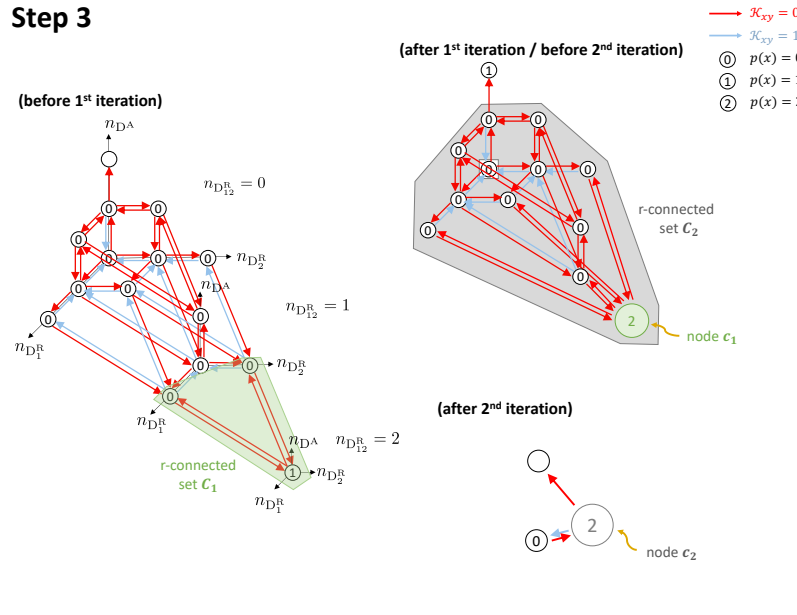
Step 1



Step 2



Step 3



Step 4

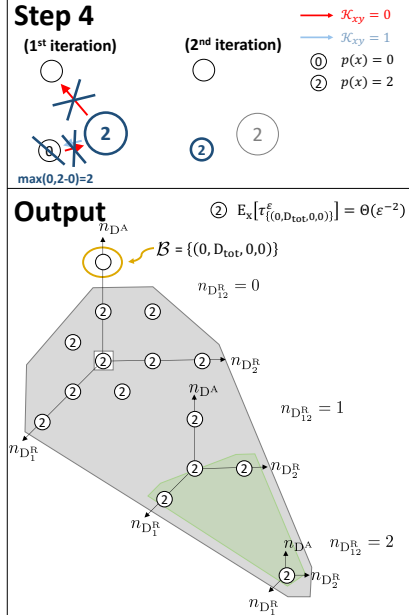


Figure 3.12. Key steps of the algorithm for the 4D model (from the fully repressed state to the fully active state).

3.10.7 Leading coefficient for the MFPT

Proof of Theorem 3.6.1

Proof. Fix $\lambda = \max\{q_x(\varepsilon) : x \in \mathcal{X}, 0 \leq \varepsilon < \varepsilon_0\}$. The λ here should not be confused with other rates λ with subscripts and/or superscripts used elsewhere of this work. In the following, we use the breve symbol to denote notation associated with discrete time Markov chains defined below.

Let $Y^\varepsilon = \{Y^\varepsilon(n) : n \in \mathbb{Z}_+\}$ be a discrete time Markov chain with transition matrix $\check{P}(\varepsilon) = I + \frac{1}{\lambda}Q(\varepsilon)$ for each $0 \leq \varepsilon < \varepsilon_0$ ⁵. Note that Y^ε is a singularly perturbed discrete time Markov chain under the definition of Avrachenkov et al. [6]. Let $\check{\Pi}(\varepsilon)$ be the ergodic projection of Y^ε and $\check{H}(\varepsilon)$ be the deviation matrix of Y^ε (as defined in (3.10.44)–(3.10.45)). The ergodic projection of Y^0 is $\check{\Pi}(0) = \check{W}\check{M}$, where I is the $|\mathcal{A}| \times |\mathcal{A}|$ identity matrix and

$$\check{W} = \left[\begin{array}{c} I \\ -T_0^{-1}R_0 \end{array} \right] \quad \text{and} \quad \check{M} = \left[\begin{array}{c|c} I & 0 \end{array} \right].$$

Then,

$$\check{M} \left(\frac{1}{\lambda} Q^{(1)} \right) \check{W} = \frac{1}{\lambda} \left[\begin{array}{c|c} I & 0 \end{array} \right] \left[\begin{array}{c|c} A_1 & S_1 \\ \hline R_1 & T_1 \end{array} \right] \left[\begin{array}{c} I \\ -T_0^{-1}R_0 \end{array} \right] = \frac{1}{\lambda} (A_1 + S_1(-T_0)^{-1}R_0) = \frac{1}{\lambda} Q_{\mathcal{A}}.$$

Assumptions 3.5.1, 3.5.2 and SI-Lemma 3.10.1 imply that the null space of this matrix is one dimensional.

Using the computational algorithm in Section 6.3.1 of Avrachenkov et al. [6], the generator⁶ for an aggregated discrete time Markov chain is $\check{M} \left(\frac{1}{\lambda} Q^{(1)} \right) \check{W} = \frac{1}{\lambda} Q_{\mathcal{A}}$, whose null space is one dimensional. Then, by the computational algorithm on page 176–177

⁵In general, when $0 < \varepsilon < \varepsilon_0$, the discrete time Markov chain Y^ε is different from the embedded discrete time Markov chain described in Section 3.4.1. In particular, the discrete time Markov chain used here can have self loops, whereas the embedded discrete time Markov chain has no self loops.

⁶The transition matrix for a discrete time Markov chain with generator $\mathcal{G}$ is $\mathcal{P} = I + \mathcal{G}$.

of Avrachenkov et al. [6] the deviation matrix $\check{H}(\varepsilon)$ has a Laurent series expansion with order of the pole equal to one:

$$\check{H}(\varepsilon) = \frac{1}{\varepsilon} \check{H}^{(-1)} + \check{H}^{(0)} + \varepsilon \check{H}^{(1)} + \dots$$

Since the aggregated Markov chain has a single recurrent class by Assumption 3.5.2, the ergodic projection of the aggregated Markov chain is $\mathbf{1}\alpha$, where α is a row vector denoting the unique stationary distribution of the aggregated discrete time Markov chain. The deviation matrix of this aggregated Markov chain is $\check{D} = (-\frac{1}{\lambda}Q_{\mathcal{A}} + \mathbf{1}\alpha)^{-1} - \mathbf{1}\alpha$. By Theorem 6.7 in Avrachenkov et al. [6],

$$\check{H}^{(-1)} = \check{W}\check{D}\check{M} = \begin{bmatrix} I \\ -T_0^{-1}R_0 \end{bmatrix} \check{D} \begin{bmatrix} I & | & 0 \end{bmatrix} = \begin{bmatrix} \check{D} & | & 0 \\ T_0^{-1}R_0\check{D} & | & 0 \end{bmatrix}.$$

For each $0 < \varepsilon < \varepsilon_0$, let $\check{h}_{x,y}(\varepsilon)$ be the mean first passage time from x to y in Y^ε .

Then, the mean first passage time from x to y in X^ε is

$$\begin{aligned} h_{x,y}(\varepsilon) = \frac{1}{\lambda} \check{h}_{x,y}(\varepsilon) &= \frac{1}{\lambda} \frac{\check{H}_{y,y}(\varepsilon) - \check{H}_{x,y}(\varepsilon)}{\pi_y(\varepsilon)} \\ &= \frac{1}{\lambda} \frac{\left(\frac{1}{\varepsilon}\check{D}_{y,y} + O(1)\right) - \left(\frac{1}{\varepsilon}\check{D}_{x,y} + O(1)\right)}{\pi_y^{(k_y)} \varepsilon^{k_y} + O(\varepsilon^{k_y+1})} \\ &= \frac{D_{y,y} - D_{x,y}}{\pi_y^{(k_y)}} \frac{1}{\varepsilon^{k_y+1}} + O\left(\frac{1}{\varepsilon^{k_y}}\right), \end{aligned} \quad (3.10.43)$$

where we used (3.10.47) in the second equality and used (3.10.46) to show that $\check{D} = \lambda D$ (these equations are proved independently further below).

When $\hat{X}_{\mathcal{A}}$ is irreducible, then $\pi^{(0)} = \alpha$ has all strictly positive entries. The mean first passage time from x to y in $\hat{X}_{\mathcal{A}}$ is finite and positive, and it is $\frac{1}{\lambda} \frac{\check{D}_{y,y} - \check{D}_{x,y}}{\alpha_y} = \frac{D_{y,y} - D_{x,y}}{\pi_y^{(0)}}$ by a similar reasoning to that in (3.10.43). In this case, the order of the pole of $h_{x,y}(\varepsilon)$ is

one and the leading coefficient is the mean first passage time from x to y in $\hat{X}_{\mathcal{A}}$. $\square$

Properties of the deviation matrix for a discrete time Markov chain

In this section, we will start with a few results stated in Section 6.1 of Avrachenkov et al. [6] about discrete time Markov chains with finite state space. These include the definitions and properties of the ergodic projection, the fundamental matrix and the deviation matrix. Then, we show one more fact about the deviation matrix. Lastly, Theorem 4.4.7 of Kemeny and Snell [29] gave a formula for mean first passage times for irreducible discrete time Markov chains in terms of the fundamental matrix and the stationary distribution, which is also briefly mentioned in Avrachenkov et al. [6]. We will write this in terms of the deviation matrix and the stationary distribution with a simple modification.

Suppose $Y = \{Y(n) : n \in \mathbb{Z}_+\}$ is a discrete time Markov chain with a finite state space $\mathcal{Y}$. Suppose the state space $\mathcal{Y}$ is partitioned into m ergodic classes (possibly including absorbing states) and a set of transient states, and accordingly, the transition matrix $\check{P}$ is

$$\check{P} = \left[\begin{array}{ccc|c} \check{A}_1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \check{A}_m & 0 \\ \hline \check{R}_1 & \dots & \check{R}_m & \check{T} \end{array} \right]. \quad (3.10.44)$$

The ergodic projection of Y is given by the *Cesaro* limit,

$$\check{\Pi} = \lim_{N \rightarrow \infty} \frac{1}{N+1} \sum_{n=0}^N \check{P}^n. \quad (3.10.45)$$

It follows that $\check{\Pi}(I - \check{P}) = 0$ and $\check{\Pi}\check{\Pi} = \check{\Pi}$. The ergodic projection $\check{\Pi}$ is the eigenprojection of the transition matrix $\check{P}$ corresponding to its maximal eigenvalue 1. That is, if $\check{\pi}_i$ is the unique stationary distribution for the discrete time Markov chain with transition matrix

$\check{A}_i$ for $1 \leq i \leq m$, then $\check{\Pi} = \check{W}\check{M}$ with

$$\check{W} = \left[\begin{array}{ccc|c} \mathbb{1} & \dots & 0 & \\ \vdots & \ddots & \vdots & \\ 0 & \dots & \mathbb{1} & \\ \hline (I - \check{T})^{-1}\check{R}_1\mathbb{1} & \dots & (I - \check{T})^{-1}\check{R}_m\mathbb{1} & \end{array} \right] \quad \text{and} \quad \check{M} = \left[\begin{array}{ccc|c} \check{\pi}_1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \check{\pi}_m & 0 \end{array} \right],$$

where $\check{W}$ and $\check{M}$ form bases for the right and left eigenspaces, respectively, which implies that $\check{P}\check{W} = \check{W}$ and $\check{M}\check{P} = \check{M}$. One can see that $v(I - \check{P} + \check{\Pi}) = 0$ implies that $v = 0$ and so $(I - \check{P} + \check{\Pi})$ is invertible. The fundamental matrix $\check{Z}$ and the deviation matrix $\check{H}$ of Y are well-defined:

$$\check{Z} = \lim_{N \rightarrow \infty} \frac{1}{N+1} \sum_{n=0}^N \sum_{\ell=0}^n (\check{P} - \check{\Pi})^\ell = (I - \check{P} + \check{\Pi})^{-1},$$

$$\check{H} = \check{Z} - \check{\Pi} = (I - \check{P} + \check{\Pi})^{-1} - \check{\Pi}.$$

We also have that $\check{H}\check{\Pi} = (\check{Z} - \check{\Pi})\check{\Pi} = 0$ since $\check{P}\check{\Pi} = \check{P}\check{W}\check{M} = \check{W}\check{M} = \check{\Pi}$, $\check{\Pi} = \check{Z}(I - \check{P} + \check{\Pi})\check{\Pi} = \check{Z}(\check{\Pi} - \check{\Pi} + \check{\Pi}) = \check{Z}\check{\Pi}$, and $\Pi^2 = \Pi$.

Now, we show a property of the deviation matrix that is not in Avrachenkov et al. [6] and is useful in Section 3.6.2. Suppose Q is an infinitesimal generator for a continuous time Markov chain on $\mathcal{Y}$ and $|Q_{y,y}| \leq \lambda$ for all $y \in \mathcal{Y}$. Then, $\check{P} = I + \frac{1}{\lambda}Q$ defines a transition matrix for a discrete time Markov chain. The associated ergodic projection and deviation matrix $\check{\Pi}$ and $\check{H}$ for $\check{P}$ satisfy $\check{\Pi}Q = \lambda(\check{\Pi}\check{P} - \check{\Pi}) = 0$ and

$$\check{H} = \left(I - \left(I + \frac{1}{\lambda}Q \right) + \check{\Pi} \right)^{-1} - \check{\Pi} = \left(-\frac{1}{\lambda}Q + \check{\Pi} \right)^{-1} - \check{\Pi},$$

and so

$$\begin{aligned}
I &= (\check{H} + \check{\Pi}) \left(-\frac{1}{\lambda} Q + \check{\Pi} \right) = -\frac{1}{\lambda} \check{H} Q + \check{H} \check{\Pi} - \frac{1}{\lambda} \check{\Pi} Q + \check{\Pi} \check{\Pi} \\
&= -\frac{1}{\lambda} \check{H} Q + \check{\Pi} = -\frac{1}{\lambda} \check{H} Q + \frac{1}{\lambda} \check{H} \check{\Pi} - \check{\Pi} Q + \check{\Pi} \check{\Pi} = \left(\frac{1}{\lambda} \check{H} + \check{\Pi} \right) (-Q + \check{\Pi}).
\end{aligned}$$

Thus,

$$\frac{1}{\lambda} \check{H} = ((-Q + \check{\Pi})^{-1} - \check{\Pi}), \quad (3.10.46)$$

where we have used the fact that $-Q + \check{\Pi}$ is invertible because $\check{\Pi}$ is the eigenprojection of Q corresponding to the eigenvalue 0.

Lastly, assume that Y is irreducible. Then, Y has a unique stationary distribution $\check{\pi}$, which is a row vector, and the ergodic projection of Y is $\mathbb{1}\check{\pi}$. By Theorem 4.4.7 in Kemeny and Snell [29], the mean first passage time from $x \in \mathcal{Y}$ to $y \in \mathcal{Y}$ is $\frac{\check{Z}_{y,y} - \check{Z}_{x,y}}{\check{\pi}_y}$. Since $\check{Z} = \check{H} + \mathbb{1}\check{\pi}$, the mean first passage time from $x \in \mathcal{Y}$ to $y \in \mathcal{Y}$ is

$$\frac{(\check{H}_{y,y} + (\mathbb{1}\check{\pi})_{y,y}) - (\check{H}_{x,y} + (\mathbb{1}\check{\pi})_{x,y})}{\check{\pi}_y} = \frac{(\check{H}_{y,y} + \check{\pi}_y) - (\check{H}_{x,y} + \check{\pi}_y)}{\check{\pi}_y} = \frac{\check{H}_{y,y} - \check{H}_{x,y}}{\check{\pi}_y}. \quad (3.10.47)$$

3.10.8 1D Model: additional mathematical details

Verification of Assumption 3.5.1. In order to show that Assumption 3.5.1 holds, consider the states $a = 0$ and $r = D_{\text{tot}}$ and the set $\mathcal{T} = \{1, \dots, D_{\text{tot}} - 1\}$ defined in Section 3.3.1. Since $D_{\text{tot}} \geq 2$, $\mathcal{T} \neq \emptyset$. From (3.3.2), we can see that $Q_{a,a+1}(0) = Q_{a,a-1}(0) = Q_{r,r+1}(0) = Q_{r,r-1}(0) = 0$. As a consequence, both a and r are absorbing states under $Q(0)$. To see that the states in $\mathcal{T}$ are transient under $Q(0)$, consider a state $x \in \mathcal{T}$. Since $Q_{z,z+1}(0) = \frac{k_E^A}{V} (D_{\text{tot}} - z)z > 0$ for all $z \in \{1, \dots, D_{\text{tot}} - 1\}$, we have $Q_{x,x+1}(0) \dots Q_{D_{\text{tot}}-1,D_{\text{tot}}}(0) > 0$. By SI-Lemma 3.10.6 and the fact that r is an absorbing state, we have that x is a transient state for X^0 .

Verification of Assumptions 3.5.4 and 3.5.2. By Lemma 3.5.3, it suffices to show Assumption 3.5.4 holds. From (3.3.2), we can see that $\tilde{Q}_{a,a+1} > 0$. From the analysis made to prove Assumption 3.5.1, we know that there is a positive probability for $\tilde{X}$ to transition from $x \in \mathcal{X} \setminus \{a, r\}$ to r . It follows that any state $x \in \mathcal{X} \setminus \{r\}$ leads to r under $\tilde{X}$. Now, we would like to show that there is a positive probability for transition from r to $x \neq r \in \mathcal{X}$ for the process $\tilde{X}$. This is because $\tilde{Q}_{r,r-1} = b \frac{k_E^A}{V} D_{\text{tot}}^2 > 0$ and $\tilde{Q}_{z,z-1} = Q_{z,z-1}(0) = \mu \frac{k_E^A}{V} (D_{\text{tot}} - z)z > 0$ for all $z \in \{1, \dots, D_{\text{tot}} - 1\}$. Thus, r leads to any state in $\mathcal{X} \setminus \{r\}$ under $\tilde{X}$. Combining the above, we see that $\tilde{X}$ is irreducible and Assumption 3.5.4 holds.

Stationary distribution. Let us consider a one-dimensional finite state continuous time Markov chain in which the state space $\mathcal{X} = \{0, 1, \dots, K\}$ and the off-diagonal entries of the infinitesimal generator Q are all zero except for the following positive rates:

$$\begin{aligned} Q_{x,x+1} &= \lambda_x & \text{if } x \in \{0, \dots, K-1\}, \\ Q_{x,x-1} &= \gamma_x & \text{if } x \in \{1, \dots, K\}. \end{aligned} \tag{3.10.48}$$

Thus, the continuous time Markov chain is a birth-and-death process, it satisfies detailed balance (see Gardiner [20]) and so the stationary distribution $\pi = (\pi_x)_{x \in \{0, 1, \dots, K\}}$ satisfies

$$\pi_x = \frac{\lambda_{x-1}}{\gamma_x} \pi_{x-1}, \quad \text{for } x \in \{1, \dots, K\}.$$

Applying this equality recursively, we can express π_x , $x \in \{1, \dots, K\}$, as a function of π_0 , obtaining

$$\pi_x = \pi_0 \prod_{i=1}^x \frac{\lambda_{i-1}}{\gamma_i}. \tag{3.10.49}$$

Using the fact that $\sum_{j=0}^K \pi_j = 1$, we obtain

$$\pi_0 = \frac{1}{1 + \sum_{j=1}^K \left(\prod_{i=1}^j \frac{\lambda_{i-1}}{\gamma_i} \right)}. \tag{3.10.50}$$

Substituting (3.10.50) in (3.10.49), we obtain

$$\pi_x = \frac{\prod_{i=1}^x \frac{\lambda_{i-1}}{\gamma_i}}{1 + \sum_{j=1}^K \left(\prod_{i=1}^j \frac{\lambda_{i-1}}{\gamma_i} \right)} \quad \text{for } x \in \{1, \dots, K\}. \quad (3.10.51)$$

Now, consider the one-dimensional continuous time Markov chain introduced in Section 3.3.1 with state space $\mathcal{X} = \{0, 1, \dots, D_{\text{tot}}\}$ and infinitesimal generator as defined in (3.3.2), which has nonzero off-diagonal entries given, for $\varepsilon > 0$, by

$$\begin{aligned} \lambda_x^\varepsilon &:= Q_{x,x+1}(\varepsilon) = \left(\frac{k_E^A}{V} x + \varepsilon \frac{k_E^A}{V} D_{\text{tot}} \right) (D_{\text{tot}} - x) & \text{if } x \in \{0, \dots, D_{\text{tot}} - 1\}, \\ \gamma_x^\varepsilon &:= Q_{x,x-1}(\varepsilon) = \mu \left(\frac{k_E^A}{V} (D_{\text{tot}} - x) + b\varepsilon \frac{k_E^A}{V} D_{\text{tot}} \right) x & \text{if } x \in \{1, \dots, D_{\text{tot}}\}. \end{aligned} \quad (3.10.52)$$

By substituting the expressions for the rates in (3.10.52) into (3.10.50)–(3.10.51) and suitably rearranging the terms, we obtain that

$$\pi_x(0) = \lim_{\varepsilon \rightarrow 0} \pi_x(\varepsilon) = \begin{cases} \frac{b\mu^{D_{\text{tot}}}}{1+b\mu^{D_{\text{tot}}}} & \text{if } x = 0 \\ 0 & \text{if } x \in \{1, \dots, D_{\text{tot}} - 1\} \\ \frac{1}{1+b\mu^{D_{\text{tot}}}} & \text{if } x = D_{\text{tot}}. \end{cases}$$

Mean first passage time. Consider the one-dimensional, finite state, continuous time Markov chain introduced in (3.10.48). We will determine an analytical expression for the MFPT from $x = K$ to $x = 0$ and from $x = 0$ to $x = K$ for this chain. We first focus on the former. For this, we exploit first step analysis (see for example Equation (3.1) in [33]), proceeding in a similar manner to that for (3.4.1), to obtain

$$\begin{cases} h_{0,0} = 0, \\ h_{x,0} = \frac{1}{\lambda_x + \gamma_x} + \frac{\lambda_x}{\lambda_x + \gamma_x} h_{x+1,0} + \frac{\gamma_x}{\lambda_x + \gamma_x} h_{x-1,0} & \text{if } x \in \{1, \dots, K-1\}, \\ h_{K,0} = \frac{1}{\gamma_K} + h_{K-1,0}, \end{cases} \quad (3.10.53)$$

where for $x, y \in \mathcal{X}$, $h_{x,y} = \mathbb{E}_x[\tau_y]$, $\tau_y = \inf\{t \geq 0 : X(t) = y\}$, X is the continuous time Markov chain with infinitesimal generator given by (3.10.48). Now, defining $\Delta h_{x,x-1} = h_{x,0} - h_{x-1,0}$ for $x \in \{1, \dots, K\}$, we can rewrite (3.10.53) in the following way:

$$\begin{cases} h_{0,0} = 0, \\ \Delta h_{x,x-1} = \frac{1}{\gamma_x} + \frac{\lambda_x}{\gamma_x} \Delta h_{x+1,x} & \text{if } x \in \{1, \dots, K-1\}, \\ \Delta h_{K,K-1} = \frac{1}{\gamma_K}. \end{cases} \quad (3.10.54)$$

From (3.10.54), we have an explicit formula for $\Delta h_{K,K-1}$ and any $\Delta h_{x,x-1}$ can be expressed as a function of $\Delta h_{x+1,x}$. Furthermore, if we sum the $\Delta h_{x,x-1}$ for $x = 1, \dots, K$, we obtain

$$h_{K,0} = h_{K,0} - h_{0,0} = \sum_{x=1}^K (\Delta h_{x,x-1}) = \Delta h_{1,0} + \Delta h_{2,1} + \dots + \Delta h_{K-1,K-2} + \Delta h_{K,K-1}.$$

Thus, to evaluate the MFPT from $x = K$ to $x = 0$, we can calculate $\Delta h_{x,x-1}$ for $x = K, K-1, \dots, 1$ and then sum all of the terms. Defining $r_j = \frac{\lambda_1 \lambda_2 \dots \lambda_j}{\gamma_1 \gamma_2 \dots \gamma_j}$, for $j = 1, \dots, K$, we obtain

$$\begin{aligned} h_{K,0} &= \frac{1}{\gamma_K} \left(1 + \frac{\lambda_{K-1}}{\gamma_{K-1}} + \frac{\lambda_{K-1} \lambda_{K-2}}{\gamma_{K-1} \gamma_{K-2}} + \dots + r_{K-1} \right) \\ &\quad + \frac{1}{\gamma_{K-1}} \left(1 + \frac{\lambda_{K-2}}{\gamma_{K-2}} + \frac{\lambda_{K-2} \lambda_{K-3}}{\gamma_{K-2} \gamma_{K-3}} + \dots + r_{K-2} \right) + \dots + \frac{1}{\gamma_1} \\ &= \frac{r_{K-1}}{\gamma_K} \left(1 + \sum_{i=1}^{K-1} \frac{1}{r_i} \right) + \sum_{i=2}^{K-1} \left[\frac{r_{i-1}}{\gamma_i} \left(1 + \sum_{j=1}^{i-1} \frac{1}{r_j} \right) \right] + \frac{1}{\gamma_1}. \end{aligned} \quad (3.10.55)$$

With a similar procedure, we can obtain the MFPT from $x = 0$ to $x = K$. More precisely,

defining $\tilde{r}_j = \frac{\gamma_{K-1}\gamma_{K-2}\dots\gamma_{K-j}}{\lambda_{K-1}\lambda_{K-2}\dots\lambda_{K-j}}$, we have

$$\begin{aligned} h_{0,K} &= \frac{1}{\lambda_0} \left(1 + \frac{\gamma_1}{\lambda_1} + \frac{\gamma_1\gamma_2}{\lambda_1\lambda_2} + \dots + \tilde{r}_{K-1} \right) \\ &\quad + \frac{1}{\lambda_1} \left(1 + \frac{\gamma_2}{\lambda_2} + \frac{\gamma_2\gamma_3}{\lambda_2\lambda_3} + \dots + \tilde{r}_{K-2} \right) + \dots + \frac{1}{\lambda_{K-1}} \\ &= \frac{\tilde{r}_{K-1}}{\lambda_0} \left(1 + \sum_{j=1}^{K-1} \frac{1}{\tilde{r}_j} \right) + \sum_{i=2}^{K-1} \left[\frac{\tilde{r}_{i-1}}{\lambda_{K-i}} \left(1 + \sum_{j=1}^{i-1} \frac{1}{\tilde{r}_j} \right) \right] + \frac{1}{\lambda_{K-1}}. \end{aligned} \quad (3.10.56)$$

A more detailed derivation of the $h_{0,K}$ and $h_{K,0}$ is given in Bruno et al. [12].

Let us consider the one-dimensional continuous time Markov chain introduced in Section 3.3.1, with state space $\mathcal{X} = \{0, 1, \dots, D_{\text{tot}}\}$ and infinitesimal transition rates that can be written as in (3.10.52). Substituting parameters in (3.10.55)–(3.10.56) yields (3.3.6)–(3.3.7), respectively. Since all of the transition rates are $O(1)$, except for λ_0^ε and $\gamma_{D_{\text{tot}}}^\varepsilon$ which are $O(\varepsilon)$, we have that both $h_{D_{\text{tot}},0}(\varepsilon)$ and $h_{0,D_{\text{tot}}}(\varepsilon)$ are $O(1/\varepsilon)$. This means that in the limit as $\varepsilon \rightarrow 0$, $h_{D_{\text{tot}},0}(\varepsilon)$ and $h_{0,D_{\text{tot}}}(\varepsilon)$, which correspond to the time to memory loss of the repressed and active states, respectively, tend to infinity.

3.10.9 2D Model: additional mathematical details

Verification of Assumption 3.5.1. In order to show that Assumption 3.5.1 holds, consider the states $a = (0, D_{\text{tot}})^T$ and $r = (D_{\text{tot}}, 0)^T$ and the set $\mathcal{T} = \{i_1, \dots, i_m\}$ defined in Section 3.5.2. From (3.3.9), we can see that $Q_{a,a+v_j}(0) = Q_{r,r+v_j}(0) = 0$ for every $1 \leq j \leq 4$. As a consequence, both a and r are absorbing states under $Q(0)$. To see that the states in $\mathcal{T}$ are transient under $Q(0)$, consider a state $x = (x_1, x_2)^T \in \mathcal{T}$. First, suppose $x_1 \neq 0$. By having the one-step transition along $v_2 = (0, -1)^T$ occurring x_2 times where $Q_{z,z+v_2}(0) = \frac{k_E^A}{V} z_2 x_1 > 0$ for all $z = (x_1, z_2)^T$ and $1 \leq z_2 \leq x_2$, and having one-step transition along $v_3 = (1, 0)^T$ occurring $D_{\text{tot}} - x_1$ times where $Q_{z,z+v_3}(0) = (D_{\text{tot}} - z_1) \left(k_{W0}^R + k_W^R + \frac{k_M^R}{V} z_1 \right) > 0$ for all $z = (z_1, 0)^T$ and $x_1 \leq z_1 \leq D_{\text{tot}} - 1$, we have a positive probability of transition from x to r under $Q(0)$. By SI-Lemma 3.10.6 and the

fact that r is an absorbing state, we have that x is a transient state for X^0 . On the other hand, suppose $x_1 = 0$. Since $x = (0, x_2) \in \mathcal{T}$, we have $0 \leq x_2 \leq D_{\text{tot}} - 1$. We can first have a one-step transition along $v_3 = (1, 0)^T$, where $Q_{x, x+v_3}(0) = (D_{\text{tot}} - x_2) (k_{W0}^R + k_W^R) > 0$, to reach the state $(1, x_2)^T$ and then take the steps for the $x_1 \neq 0$ case to reach r . In this way, there is a positive probability of transition from x to the absorbing state r under $Q(0)$, and thus x is transient by SI-Lemma 3.10.6.

Verification of Assumptions 3.5.3 and 3.5.2. By Lemma 3.5.3, it suffices to show Assumption 3.5.3 holds. From (3.3.9), we can see that $\tilde{Q}_{a, a+v_2} > 0$. From the analysis made to prove Assumption 3.5.1, we know that there is a positive probability to transition from all $x \in \mathcal{X} \setminus \{a, r\}$ to r . Now, we would like to show that there is a positive probability to transition from r to $x = (x_1, x_2)^T \in \mathcal{X} \setminus \{(0, 0)^T\}$ for the process $\tilde{X}$. We first can have a one-step transition along $v_4 = (-1, 0)^T$ where $Q_{r, r+v_4}^{(1)} = \mu b \frac{k_M^A}{V} D_{\text{tot}}^2 > 0$, then have a one-step transition along $v_1 = (0, 1)^T$ where $Q_{r+v_4, r+v_4+v_1}^{(0)} = k_{W0}^A + k_W^A > 0$, then have one-step transitions along $v_4 = (-1, 0)^T$ occurring $D_{\text{tot}} - x_1 - 1$ times where $Q_{z, z+v_4}^{(0)} = \mu \frac{k_E^A}{V} z_1 > 0$ for all $z = (z_1, 1)^T$ and $x_1 + 1 \leq z_1 \leq D_{\text{tot}} - 1$. If $x_2 \neq 0$, we finally have one-step transitions along $v_1 = (0, 1)^T$ occurring $x_2 - 1$ times where $Q_{z, z+v_1}^{(0)} = (D_{\text{tot}} - (x_1 + z_2)) \left(k_{W0}^A + k_W^A + \frac{k_M^A}{V} z_2 \right) > 0$ for all $z = (x_1, z_2)^T$ and $1 \leq z_2 \leq x_2 - 1$; if $x_2 = 0$ and $x_1 \neq 0$, we will make a one-step transition along $v_2 = (0, -1)^T$ to $(x_1, 0)^T$ where $Q_{z, z+v_2}^{(0)} = \frac{k_E^A}{V} x_1 > 0$ with $z = (x_1, 1)^T$ and $x_1 \geq 1$. Therefore, we have that there is a positive probability of transition from r to each $x \in \mathcal{X} \setminus \{(0, 0)^T\}$. Since $Q_{-v_j, (0, 0)^T}^{(0)} = 0$ for $j \in \{1, 2, 3, 4\}$ such that $-v_j \in \mathcal{X}$, we conclude that $\mathcal{C} = \mathcal{X} \setminus \{(0, 0)^T\}$ is a closed communicating class under $\tilde{Q}$ and since it contains $\mathcal{A}$, Assumption 3.5.3 holds. Note that Assumption 3.5.4 does not hold.

Stationary distribution. Here, we derive the expression for $\pi_x^{(1)}$, $x \in \mathcal{T} = \{i_1, \dots, i_m\}$,

for the case $D_{\text{tot}} = 2$. In this case $T(\varepsilon) = T_0 + \varepsilon T_1$, with

$$T_0 = \begin{bmatrix} -q_3 & k_{W0}^R + k_W^R & 0 & 0 \\ \mu \frac{k_E^A}{V} & -\frac{k_E^A}{V}(1 + \mu) & 0 & \frac{k_E^A}{V} \\ 2(k_{W0}^A + k_W^A) & 0 & -q_5 & 2(k_{W0}^R + k_W^R) \\ 0 & k_{W0}^A + k_W^A & 0 & -q_6 \end{bmatrix},$$

$$T_1 = \begin{bmatrix} -2\frac{k_M^A}{V} & 0 & 2\frac{k_M^A}{V} & 0 \\ 2\frac{k_M^A}{V}\mu b & -2\frac{k_M^A}{V}(1 + \mu b) & 0 & 2\frac{k_M^A}{V} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2\frac{k_M^A}{V}\mu b & -2\frac{k_M^A}{V} \end{bmatrix},$$

in which $q_3 = k_{W0}^A + k_W^A + \frac{k_M^A}{V} + k_{W0}^R + k_W^R$, $q_5 = 2(k_{W0}^A + k_W^A + k_{W0}^R + k_W^R)$ and $q_6 = k_{W0}^A + k_W^A + k_{W0}^R + k_W^R + \frac{k_M^R}{V}$. Then, by (3.5.4), $\beta_x^{(1)} = \pi_x^{(1)}$, $x \in \mathcal{T}$, is given by $\beta^{(1)} = \alpha S_1(-T_0)^{-1}$, where $\alpha = [\pi_{(0,2)}^{(0)}, \pi_{(2,0)}^{(0)}]$ is already given in (3.5.11). After some calculations, $\pi_x^{(1)}$ can be written for $D_{\text{tot}} = 2$ as

$$\begin{aligned} \pi_{i_1}^{(1)} &= \frac{4b\varepsilon^2 \left(\frac{k_M^A}{V}\right)^2 \bar{k}_W^A \mu^2 \left(\bar{k}_W^R (\bar{k}_W^R + \frac{k_M^R}{V}) + (\bar{k}_W^A + \frac{k_M^A}{V}) \left((1 + \mu)(\bar{k}_W^R + \frac{k_M^R}{V}) + \mu \frac{k_M^A}{V}\right)\right)}{d_1 d_2}, \\ \pi_{i_2}^{(1)} &= \frac{4b\varepsilon^2 \left(\frac{k_M^A}{V}\right)^2 \bar{k}_W^A \bar{k}_W^R \mu \left((\bar{k}_W^R + \frac{k_M^R}{V})(\bar{k}_W^R + \bar{k}_W^A + k_M^A) + \mu(\bar{k}_W^A + \frac{k_M^A}{V})(\bar{k}_W^R + \bar{k}_W^A + k_M^R)\right)}{\frac{k_E^A}{V} d_1 d_2}, \\ \pi_{i_3}^{(1)} &= 0, \\ \pi_{i_4}^{(1)} &= \frac{4b\varepsilon^2 \left(\frac{k_M^A}{V}\right)^2 \bar{k}_W^R \mu \left((\bar{k}_W^R + \frac{k_M^R}{V}) \left((1 + \mu)(\bar{k}_W^A + \frac{k_M^A}{V}) + \mu \bar{k}_W^R\right) + \mu \bar{k}_W^A (\bar{k}_W^A + \frac{k_M^A}{V})\right)}{d_1 d_2}, \end{aligned}$$

in which $\bar{k}_W^A = k_{W0}^A + k_W^A$, $\bar{k}_W^R = k_{W0}^R + k_W^R$, $d_1 = \varepsilon \frac{k_M^A}{V} (\bar{k}_W^R (\bar{k}_W^R + \frac{k_M^R}{V}) + b \bar{k}_W^A \mu^2 (\bar{k}_W^A + \frac{k_M^A}{V}))$, and $d_2 = ((\bar{k}_W^A + \frac{k_M^A}{V})((1 + \mu)(\bar{k}_W^R + \frac{k_M^R}{V}) + \mu \bar{k}_W^A) + \bar{k}_W^R (\bar{k}_W^R + \frac{k_M^R}{V}))$ and in which $i_1 = (0, 1)^T$, $i_2 = (1, 1)^T$, $i_3 = (0, 0)^T$, and $i_4 = (0, 2)^T$.

3.10.10 3D Model: additional mathematical details

Verification of Assumption 3.5.1. In order to show that Assumption 3.5.1 holds, consider the states $a = (0, D_{\text{tot}}, 0)^T$ and $r = (D_{\text{tot}}, 0, 0)^T$ and the set $\mathcal{T} = \{i_1, \dots, i_m\}$ defined in Section 3.8.1. From (3.8.2), we can see that $Q_{a,a+v_j}(0) = Q_{r,r+v_j}(0) = 0$ for every $1 \leq j \leq 6$. As a consequence, both a and r are absorbing states under $Q(0)$. To see that the states in $\mathcal{T}$ are transient under $Q(0)$, consider a state $x = (x_1, x_2, x_3)^T \in \mathcal{T}$. First, suppose $x_1 + x_3 \neq 0$. By having the one-step transitions along $v_2 = (0, -1, 0)^T$ occurring x_2 times where $Q_{z,z+v_2}(0) = \frac{k_E^A}{V}(x_3 + 2x_1)z_2 > 0$ for all $z = (x_1, z_2, x_3)^T$ and $1 \leq z_2 \leq x_2$, then having one-step transitions along $v_3 = (0, 0, 1)^T$ occurring $D_{\text{tot}} - x_1 - x_3$ times where $Q_{z,z+v_3}(0) = (D_{\text{tot}} - (x_1 + z_3)) \left(k_{W0}^1 + k_W^1 + \frac{k'_M}{V}x_1 \right) > 0$ for all $z = (x_1, 0, z_3)^T$ and $x_3 \leq z_3 \leq D_{\text{tot}} - x_1 - 1$ and finally having one-step transitions along $v_5 = (1, 0, -1)^T$ occurring $D_{\text{tot}} - x_1$ times where $Q_{z,z+v_5}(0) = (D_{\text{tot}} - z_1) \left(k_{W0}^2 + \frac{k_M}{V}z_1 + \frac{\bar{k}_M}{V} \frac{D_{\text{tot}}+z_1-1}{2} \right) > 0$ for all $z = (z_1, 0, D_{\text{tot}} - z_1)^T$ and $x_1 \leq z_1 \leq D_{\text{tot}} - 1$, we have a positive probability of transition from x to r under $Q(0)$. By SI-Lemma 3.10.6 and the fact that r is an absorbing state, we have that x is a transient state for X^0 . On the other hand, suppose $x_1 + x_3 = 0$. Since $x = (0, x_2, 0)^T \in \mathcal{T}$, we have $0 \leq x_2 \leq D_{\text{tot}} - 1$. We can first have a one-step transition along $v_3 = (0, 0, 1)^T$, where $Q_{x,x+v_3}(0) = (D_{\text{tot}} - x_2) (k_{W0}^1 + k_W^1) > 0$, to reach the state $(0, x_2, 1)^T$ and then take the steps in the $x_1 + x_3 \neq 0$ case. In this way, there is a positive probability of transition from x to the absorbing state r , and thus x is transient by SI-Lemma 3.10.6.

Verification of Assumption 3.5.2. To show that Assumption 3.5.2 holds, consider the continuous time Markov chain $\tilde{X}$ with infinitesimal generator $\tilde{Q}$ as described in (3.5.6) and shown in Figure 3.6(d). We will first see that $\{i_m, r\}$ forms a closed class under $\tilde{Q}$. For this, we see that $Q_{r,r+v_j}(\varepsilon)$ vanishes for every $1 \leq j \leq 5$ and $\varepsilon \geq 0$, while $Q_{r,r+v_6}(\varepsilon) = \varepsilon \mu b \frac{k_M^A}{V} D_{\text{tot}}^2$. Therefore, the only transition from r under $\tilde{Q}$ is given by $\tilde{Q}_{r,r+v_6} = \mu b \frac{k_M^A}{V} D_{\text{tot}}^2 > 0$, where $r + v_6 = i_m$. From (3.8.2), we can see that $Q_{i_m,i_m+v_j}(0) = 0$ for every $j \in \{1, 2, 3, 4, 6\}$

and $Q_{i_m, i_m + v_5}(0) = k_{W0}^2 + \frac{k_M}{V}(D_{\text{tot}} - 1) + \frac{\bar{k}_M}{V}(D_{\text{tot}} - 1) > 0$. Since $i_m + v_5 = r$ we see that $\tilde{Q}_{i_m, r} > 0$. Therefore, $\{i_m, r\}$ forms a closed class under $\tilde{Q}$. The fact that $\hat{X}_{\mathcal{A}}$, shown in Figure 3.6(e) consists of erasing the times from $\tilde{X}$ in which the process is in $\mathcal{T}$, together with Lemma 3.5.2, yields that r is an absorbing state under $Q_{\mathcal{A}}$. From (3.8.2), we can see that $\tilde{Q}_{a, i_1} > 0$. From the analysis made to prove Assumption 3.5.1 we obtain that i_1 leads to r under $\tilde{Q}$ which is part of a closed class. By interpreting $\hat{X}_{\mathcal{A}}$ again as a time-change of $\tilde{X}$, by Lemma 3.5.2 we obtain that a is transient under $Q_{\mathcal{A}}$. As a consequence, $Q_{\mathcal{A}}$ has a single recurrent class consisting of the state r , and so Assumption 3.5.2 holds, and furthermore, $\alpha = [\alpha_a, \alpha_r]$ with $\alpha_a = 0$ and $\alpha_r = 1$. In addition, the previous arguments show that neither Assumption 3.5.3 nor 3.5.4 holds for this model.

Stationary distribution. Here, we derive an expression for $\pi_x^{(1)}$, $x \in \mathcal{T} = \{i_1, \dots, i_m\}$. Matrices A_1 , S_1 , and R_0 can be written as

$$A_1 = \begin{bmatrix} -s_1 & 0 \\ 0 & -s_2 \end{bmatrix}, \quad S_1 = \begin{bmatrix} s_1 & 0 & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 & s_2 \end{bmatrix}, \quad R_0 = \begin{bmatrix} r_1 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & r_2 \end{bmatrix},$$

with $s_1 = \frac{k_M^A}{V}D_{\text{tot}}^2$, $s_2 = \mu b \frac{k_M^A}{V}D_{\text{tot}}^2$, $r_1 = k_{W0}^A + k_W^A + \frac{k_M^A}{V}(D_{\text{tot}} - 1)$ and $r_2 = k_{W0}^2 + (\frac{k_M}{V} + \frac{\bar{k}_M}{V})(D_{\text{tot}} - 1)$. From (3.5.4), $\beta^{(1)} = [\pi_{i_1}^{(1)}, \dots, \pi_{i_m}^{(1)}] = \alpha S_1 (-T_0)^{-1}$, and so, given that the last row of T_0 is made of all zeros except for the last element $(T_0)_{i_m, i_m}$ and given that $(T_0)_{i_m, i_m} = k_{W0}^2 + (\frac{k_M}{V} + \frac{\bar{k}_M}{V})(D_{\text{tot}} - 1)$, we obtain $\beta^{(1)} = [0, \dots, 0, \pi_{i_m}^{(1)}]$ with

$$\pi_{i_m}^{(1)} = \frac{\mu b \frac{k_M^A}{V}D_{\text{tot}}^2}{k_{W0}^2 + (\frac{k_M}{V} + \frac{\bar{k}_M}{V})(D_{\text{tot}} - 1)}.$$

Now, $\alpha^{(1)} = [\pi_a^{(1)}, \pi_r^{(1)}]$ is the unique vector such that $\alpha^{(1)}Q_{\mathcal{A}} = -\beta^{(1)}[R_1 + T_1(-T_0)^{-1}R_0]$

and $\alpha^{(1)}\mathbb{1} = -\beta^{(1)}\mathbb{1}$.

For an illustration, suppose $D_{\text{tot}} = 2$. Then $A_0 = 0$, $S_0 = 0$ and matrices $A_1 \in \mathbb{R}^{2 \times 2}$ and $S_1 \in \mathbb{R}^{2 \times 8}$ are given by

$$A_1 = \begin{bmatrix} -4\frac{k_M^A}{V} & 0 \\ 0 & -4\frac{k_M^A}{V}\mu b \end{bmatrix}, \quad S_1 = \begin{bmatrix} 4\frac{k_M^A}{V} & 0 & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 & 4\frac{k_M^A}{V}\mu b \end{bmatrix}.$$

Furthermore, $R_1 = 0$ and $R_0 \in \mathbb{R}^{8 \times 2}$ can be written as

$$R_0 = \begin{bmatrix} \bar{k}_W^A + \frac{k_M^A}{V} & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & k_{W0}^2 + \frac{k_M}{V} + \frac{\bar{k}_M}{V} \end{bmatrix}.$$

Finally, matrices $T_0, T_1 \in \mathbb{R}^{8 \times 8}$ can be written as

$$T_0 = \begin{bmatrix} -\bar{k}_W^A - \frac{k_M^A}{V} - \bar{k}_W^1 & 0 & 0 & 0 & \bar{k}_W^1 & 0 & 0 & 0 \\ 0 & -\frac{k_F^A}{V}(2 + \mu) & 0 & 2\frac{k_F^A}{V} & \mu\frac{k_F^A}{V} & 0 & 0 & 0 \\ 2\bar{k}_W^A & 0 & -2(\bar{k}_W^A + \bar{k}_W^1) & 0 & 0 & 2\bar{k}_W^1 & 0 & 0 \\ 0 & \bar{k}_W^A & 0 & -(\bar{k}_W^A + \bar{k}_W^1 + \frac{k_M'}{V}) & 0 & 0 & 0 & \bar{k}_W^1 + \frac{k_M'}{V} \\ \mu'\frac{k_F^A}{V} & k_{W0}^2 & 0 & 0 & -(\frac{k_F^A}{V}(1 + \mu') + k_{W0}^2) & \frac{k_F^A}{V} & 0 & 0 \\ 0 & 0 & 0 & k_{W0}^2 & \bar{k}_W^A & -(k_{W0}^2 + \bar{k}_W^A + \bar{k}_W^1) & \bar{k}_W^1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -2(k_{W0}^2 + \frac{\bar{k}_M}{V}) & 2(k_{W0}^2 + \frac{\bar{k}_M}{2V}) \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2(k_{W0}^2 + \frac{k_M}{V} + \frac{\bar{k}_M}{V}) \end{bmatrix},$$

$$T_1 = \begin{bmatrix} -2\frac{k_M^A}{V} & 0 & 2\frac{k_M^A}{V} & 0 & 0 & 0 & 0 & 0 \\ 0 & -2\frac{k_M^A}{V}(1 + b\mu) & 0 & 2\frac{k_M^A}{V} & 2b\mu\frac{k_M^A}{V} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2b\mu\frac{k_M^A}{V} & 0 & 2b\mu\frac{k_M^A}{V} & 0 & 0 \\ 2\mu'\beta\frac{k_M^A}{V} & 0 & 0 & 0 & -2\frac{k_M^A}{V}(1 + \beta\mu') & 2\frac{k_M^A}{V} & 0 & 0 \\ 0 & 0 & 2\mu'\beta\frac{k_M^A}{V} & 0 & 0 & -2\mu'\beta\frac{k_M^A}{V} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4\mu'\beta\frac{k_M^A}{V} & -4\mu'\beta\frac{k_M^A}{V} & 0 \\ 0 & 0 & 0 & 2\mu'\beta\frac{k_M^A}{V} & 0 & 0 & 2\mu b\frac{k_M^A}{V} & -2(\mu'\beta\frac{k_M^A}{V} + \mu b\frac{k_M^A}{V}) \end{bmatrix},$$

in which $\bar{k}_W^A = k_{W0}^A + k_W^A$, $\bar{k}_W^R = k_{W0}^R + k_W^R$. Now, by SI-Theorem 3.10.1, we first obtain that $\pi(0) = \pi^{(0)} = [\alpha, 0] = [\alpha_a, \alpha_r, 0, \dots, 0]$ where α is the unique probability vector such that $\alpha Q_{\mathcal{A}} = 0$. In this case,

$$Q_{\mathcal{A}} = \frac{K_1 + \mu K_2}{K_3 + \mu K_4 + \mu' K_5 + \mu\mu' K_6} \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix},$$

with

$$\begin{aligned} K_1 &= 8\bar{k}_W^1 \frac{k_M^A}{V} \left(\bar{k}_W^1 + \frac{k_M'}{V} \right) \left((\bar{k}_W^1 + k_{W0}^2) \left(\frac{k_E^A}{V} + k_{W0}^2 \right) + \bar{k}_W^A k_{W0}^2 \right), \\ K_2 &= 4\bar{k}_W^1 \frac{k_M^A}{V} \frac{k_E^A}{V} \left(\left(\bar{k}_W^1 + \frac{k_M'}{V} \right) (\bar{k}_W^1 + k_{W0}^2) + \bar{k}_W^1 \bar{k}_W^A \right), \\ K_3 &= \frac{k_E^A}{V} \left(2 \left(\bar{k}_W^1 + \bar{k}_W^A + k_{W0}^2 + \frac{k_M'}{V} + \frac{k_M^A}{V} \right) \right) (\bar{k}_W^1)^2 \\ &\quad + 2\bar{k}_W^1 \left(\bar{k}_W^1 \left(\bar{k}_W^1 + k_{W0}^2 \left(k_{W0}^2 + \frac{k_M'}{V} \right) \right) + \frac{k_M'}{V} (k_{W0}^2)^2 \right) \\ &\quad + 2\frac{k_M^A}{V} k_{W0}^2 \left(\frac{k_M'}{V} (k_{W0}^2 + \bar{k}_W^A) + \bar{k}_W^1 + (2\bar{k}_W^1 + \bar{k}_W^A + k_{W0}^2) \right) \\ &\quad + 2\frac{k_M^A}{V} \frac{k_M'}{V} k_{W0}^2 (\bar{k}_W^1 + \bar{k}_W^A) + 2\bar{k}_W^1 k_{W0}^2 \left(\frac{k_E^A}{V} \frac{k_M'}{V} + \bar{k}_W^A \frac{k_M^A}{V} + 2\bar{k}_W^A \frac{k_M'}{V} \right), \\ K_4 &= \frac{k_E^A}{V} \left(\left(\bar{k}_W^1 + \frac{k_M'}{V} \right) \left(\bar{k}_W^A + \frac{k_M^A}{V} \right) (\bar{k}_W^1 + k_{W0}^2) + (\bar{k}_W^1 + k_{W0}^2) \bar{k}_W^1 \frac{k_M'}{V} \right) \\ &\quad + \frac{k_E^A}{V} \bar{k}_W^1 \left(\bar{k}_W^1 (\bar{k}_W^A + k_{W0}^2 + 1) + \bar{k}_W^A \left(\bar{k}_W^A + \frac{k_M^A}{V} \right) \right), \\ K_5 &= 2\frac{k_E^A}{V} (\bar{k}_W^1 + k_{W0}^2 + \bar{k}_W^A) \left(\bar{k}_W^A + \frac{k_M^A}{V} \right) \left(\bar{k}_W^1 + \frac{k_M'}{V} \right), \\ K_6 &= \frac{k_E^A}{V} (\bar{k}_W^1 + k_{W0}^2 + \bar{k}_W^A) \left(\bar{k}_W^A + \frac{k_M^A}{V} \right) \left(\bar{k}_W^A + \bar{k}_W^1 + \frac{k_M'}{V} \right), \end{aligned} \tag{3.10.57}$$

and thus $\alpha_a = 0$ and $\alpha_r = 1$. Let us now derive an expression for $\beta^{(1)}$. Starting from the transient states $\mathcal{T} = \{i_1, \dots, i_8\}$, we obtain that $\beta^{(1)} = [\pi_{i_1}^{(1)}, \dots, \pi_{i_8}^{(1)}]$ can be determine by

$\beta^{(1)} = \alpha S_1(-T_0)^{-1}$, obtaining $\beta^{(1)} = [0, \dots, 0, \pi_{i8}^{(1)}]$ with

$$\pi_{i8}^{(1)} = \frac{4\mu b \frac{k_M^A}{V}}{k_{W0}^2 + (\frac{k_M}{V} + \frac{\bar{k}_M}{V})}.$$

Finally, $\alpha^{(1)} = [\pi_a^{(1)}, \pi_r^{(1)}]$ is the unique vector such that $\alpha^{(1)} Q_{\mathcal{A}} = -\beta^{(1)}(R_1 + T_1(-T_0)^{-1}R_0)$ and $\alpha^{(1)} \mathbf{1} = -\beta^{(1)} \mathbf{1}$. After some calculations, we obtain

$$\pi_a^{(1)} = \frac{\mu^2 \mu'^2 K_7}{K_8(K_9 + K_{10}\mu)}, \quad \pi_r^{(1)} = -\pi_a^{(1)} - \pi_{i8}^{(1)} = -\frac{\mu^2 \mu'^2 K_7 + \mu K_{11}(K_9 + K_{10}\mu)}{K_8(K_9 + K_{10}\mu)},$$

with $\bar{k}_W^1 = k_{W0}^1 + k_W^1$, $k_W^A = k_{W0}^A + k_W^A$,

$$\begin{aligned} K_7 &= 2\beta b \frac{k_M^A}{V} \bar{k}_W^A \left(\bar{k}_W^A + \frac{k_M^A}{V} \right) (\bar{k}_W^A + \bar{k}_W^1 + k_{W0}^2), \\ K_8 &= \bar{k}_W^1 \left(k_{W0}^2 + \frac{k_M}{V} + \frac{\bar{k}_M}{V} \right), \\ K_9 &= 2 \left(\bar{k}_W^1 + \frac{k_M'}{V} \right) \left(\frac{k_E^A}{V} (\bar{k}_W^1 + k_{W0}^2) + k_{W0}^2 (\bar{k}_W^2 + \bar{k}_W^1 + \bar{k}_W^A) \right), \\ K_{10} &= \frac{k_E^A}{V} \left((\bar{k}_W^1 + k_{W0}^2) \left(\bar{k}_W^1 + \frac{k_M'}{V} \right) + \bar{k}_W^A \bar{k}_W^1 \right), \\ K_{11} &= 4b \bar{k}_W^1 \frac{k_M^A}{V}. \end{aligned} \tag{3.10.58}$$

Time to memory loss. As a reminder, we denote the time to memory loss of the active state by $h_{a,r}(\varepsilon)$ and the time to memory loss of repressed state by $h_{r,a}(\varepsilon)$. Let us start by deriving the order and the leading coefficient of $h_{a,r}(\varepsilon)$ and $h_{r,a}(\varepsilon)$. By (3.6.1), we know the order of the stationary distribution at a and r are $k_a = -\min\{1 - 2, 0\} = 1$ and $k_r = -\min\{1 - 1, 0\} = 0$, respectively. This is consistent with the results in Section 3.8.1. Moreover, the leading coefficient in the stationary distribution for the fully repressed and fully active states are $\pi_r^{(0)} = 1$ and $\pi_a^{(1)} > 0$, respectively. Now, $\hat{X}_{\mathcal{A}}$ has the infinitesimal

generator in the form of

$$Q_{\mathcal{A}} = \begin{bmatrix} -(Q_{\mathcal{A}})_{a,r} & (Q_{\mathcal{A}})_{a,r} \\ 0 & 0 \end{bmatrix},$$

and $\hat{X}_{\mathcal{A}}$ has a unique stationary distribution $\alpha = [0, 1]$. The deviation matrix for $\hat{X}_{\mathcal{A}}$ is

$$D = \begin{bmatrix} \frac{1}{(Q_{\mathcal{A}})_{a,r}} & -\frac{1}{(Q_{\mathcal{A}})_{a,r}} \\ 0 & 0 \end{bmatrix}.$$

By Theorem 3.6.1,

$$h_{a,r}(\varepsilon) = \frac{D_{r,r} - D_{a,r}}{\pi_r^{(k_r)}} \frac{1}{\varepsilon^{k_r+1}} + O\left(\frac{1}{\varepsilon^{k_r}}\right) = \frac{D_{r,r} - D_{a,r}}{\alpha_r} \frac{1}{\varepsilon} + O(1) = \frac{1}{(Q_{\mathcal{A}})_{a,r}} \frac{1}{\varepsilon} + O(1),$$

and

$$h_{r,a}(\varepsilon) = \frac{D_{a,a} - D_{r,a}}{\pi_a^{(k_a)}} \frac{1}{\varepsilon^{k_a+1}} + O\left(\frac{1}{\varepsilon^{k_a}}\right) = \frac{D_{a,a} - D_{r,a}}{\pi_a^{(1)}} \frac{1}{\varepsilon^2} + O\left(\frac{1}{\varepsilon}\right) = \frac{1}{(Q_{\mathcal{A}})_{a,r} \cdot \pi_a^{(1)}} \frac{1}{\varepsilon^2} + O\left(\frac{1}{\varepsilon}\right).$$

As an example, when $D_{\text{tot}} = 2$, $Q_{\mathcal{A}}$ and $\pi_a^{(1)}$ are already given in (3.8.3)–(3.8.4), and we obtain that

$$h_{a,r}(\varepsilon) = \frac{K_3 + \mu K_4 + \mu' K_5 + \mu \mu' K_6}{K_1 + \mu K_2} \frac{1}{\varepsilon} + O(1),$$

and

$$h_{r,a}(\varepsilon) = \frac{K_3 + \mu K_4 + \mu' K_5 + \mu \mu' K_6}{K_1 + \mu K_2} \frac{K_8(K_9 + K_{10}\mu)}{\mu^2 \mu'^2 K_7} \frac{1}{\varepsilon^2} + O\left(\frac{1}{\varepsilon}\right),$$

where K_i , $i = 1, \dots, 11$, are non-negative constants independent of ε , μ and μ' as defined in (3.10.57)–(3.10.58).

Verification of conditions of Theorem 3.7.1. Let us verify that both conditions (i) and (ii) of Theorem 3.7.1 hold. To this end, let us first write the directions of the six possible transitions of the continuous time Markov chain $X^\varepsilon(t)$, which are $v_1 = (0, 1, 0)^T$,

$v_2 = (0, -1, 0)^T$, $v_3 = (0, 0, 1)^T$, $v_4 = (0, 0, -1)^T$, $v_5 = (1, 0, -1)^T$, $v_6 = (-1, 0, 1)^T$, with the associated infinitesimal transition rates that can be written as $\Upsilon_1(x) = f_A(x)$, $\Upsilon_2(x) = g_\varepsilon^A(x)$, $\Upsilon_3(x) = f_{R1}(x)$, $\Upsilon_4(x) = g_{R1}^\varepsilon(x)$, $\Upsilon_5(x) = f_{R12}(x)$, $\Upsilon_6(x) = g_{R12}^\varepsilon(x)$.

Define the matrix

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

and, for $x \in \mathcal{X}$, $(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} : x \preceq_A w\}$. Let us also introduce infinitesimal transition rates $\check{\Upsilon}_i(x)$, $i = 1, 2, \dots, 6$, defined as for $\Upsilon_i(x)$, $i = 1, 2, \dots, 6$, with all the parameters having the same values except that μ' is replaced by $\check{\mu}'$, with $\mu' \geq \check{\mu}'$. Given that $Av_1 = (0, -1, 0)^T$, $Av_2 = (0, 1, 0)^T$, $Av_3 = (0, 0, 1)^T$, $Av_4 = (0, 0, -1)^T$, $Av_5 = (1, 0, 0)^T$, $Av_6 = (-1, 0, 0)^T$, condition (i) of Theorem 3.7.1 holds.

To verify condition (ii) of Theorem 3.7.1, first consider $x \in \mathcal{X}$ and $y \in \partial_1(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} : x_1 = w_1, x_2 \geq w_2, x_1 + x_3 \leq w_1 + w_3\} = \{w \in \mathbb{R}^3 : x_1 = w_1, x_2 \geq w_2, x_3 \leq w_3\}$. Given that $\langle A_{1\bullet}, v_5 \rangle = 1$ and $\langle A_{1\bullet}, v_6 \rangle = -1$, we need to verify that $\Upsilon_5(x) \leq \check{\Upsilon}_5(y)$ and $\Upsilon_6(x) \geq \check{\Upsilon}_6(y)$. Since $x_1 = y_1, x_2 \geq y_2, x_3 \leq y_3$, then $\Upsilon_5(x) = x_3 \left(k_{W0}^2 + \frac{k_M}{V}x_1 + \frac{\bar{k}_M}{V}(x_1 + \frac{x_3-1}{2}) \right) \leq y_3 \left(k_{W0}^2 + \frac{k_M}{V}x_1 + \frac{\bar{k}_M}{V}(y_1 + \frac{y_3-1}{2}) \right) = \check{\Upsilon}_5(y)$ and $\Upsilon_6(x) = x_1\mu \left(\varepsilon \frac{k_M^A}{V}D_{\text{tot}}b + x_2 \frac{k_E^A}{V} \right) \geq y_1\mu \left(\varepsilon \frac{k_M^A}{V}D_{\text{tot}}b + y_2 \frac{k_E^A}{V} \right) = \check{\Upsilon}_6(y)$. Let us now consider $x \in \mathcal{X}$ and $y \in \partial_2(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} : x_1 \leq w_1, x_2 = w_2, x_1 + x_3 \leq w_1 + w_3\}$. Given that $\langle A_{2\bullet}, v_1 \rangle = -1$ and $\langle A_{2\bullet}, v_2 \rangle = 1$, we need to verify that $\Upsilon_1(x) \geq \check{\Upsilon}_1(y)$ and $\Upsilon_2(x) \leq \check{\Upsilon}_2(y)$. Since $x_1 \leq y_1, x_2 = y_2, x_1 + x_3 \leq y_1 + y_3$, then $\Upsilon_1(x) = (D_{\text{tot}} - (x_1 + x_2 + x_3)) \left(k_{W0}^A + k_W^A + \frac{k_M^A}{V}x_2 \right) \geq (D_{\text{tot}} - (y_1 + y_2 + y_3)) \left(k_{W0}^A + k_W^A + \frac{k_M^A}{V}y_2 \right) = \check{\Upsilon}_1(y)$ and $\Upsilon_2(x) = x_2 \left(\varepsilon \frac{k_M^A}{V}D_{\text{tot}} + \frac{k_E^A}{V}(x_3 + 2x_1) \right) \leq y_2 \left(\varepsilon \frac{k_M^A}{V}D_{\text{tot}} + \frac{k_E^A}{V}(y_3 + 2y_1) \right) = \check{\Upsilon}_2(y)$. Finally, consider $x \in \mathcal{X}$ and $y \in \partial_3(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} : x_1 \leq w_1, x_2 \geq w_2, x_1 + x_3 = w_1 + w_3\} = \{w \in \mathcal{X} : x_1 \leq w_1, x_2 \geq w_2, x_3 \geq w_3\}$. Given that $\langle A_{3\bullet}, v_3 \rangle = 1$ and $\langle A_{3\bullet}, v_4 \rangle = -1$, we need to check that $\Upsilon_3(x) \leq \check{\Upsilon}_3(y)$ and $\Upsilon_4(x) \geq \check{\Upsilon}_4(y)$. Since $x_1 \leq$

$y_1, x_2 \geq y_2, x_3 \geq y_3$, then $\Upsilon_3(x) = (D_{\text{tot}} - (x_1 + x_2 + x_3)) \left(k_{W0}^1 + k_W^1 + \frac{k'_M}{V} x_1 \right) \leq (D_{\text{tot}} - (y_1 + y_2 + y_3)) \left(k_{W0}^1 + k_W^1 + \frac{k'_M}{V} y_1 \right) = \check{\Upsilon}_3(y)$ and $\Upsilon_4(x) = x_3 \mu' \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} \beta + x_2 \frac{k_E^A}{V} \right) \geq y_3 \check{\mu}' \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} \beta + y_2 \frac{k_E^A}{V} \right) = \check{\Upsilon}_4(y)$.

We can then conclude that all of the conditions of Theorem 3.7.1 hold.

3.10.11 4D Model: additional mathematical details

Verification of Assumption 3.5.1. In order to show that Assumption 3.5.1 holds, consider the states $a = (0, D_{\text{tot}}, 0, 0)^T$ and $r = (D_{\text{tot}}, 0, 0, 0)^T$ and the set $\mathcal{T} = \{i_1, \dots, i_m\}$ defined in Section 3.8.2. From (3.8.6), we can see that $Q_{a,a+v_j}(0) = Q_{r,r+v_j}(0) = 0$ for every $1 \leq j \leq 10$. As a consequence, both a and r are absorbing states under $Q(0)$. To see that the states in $\mathcal{T}$ are transient under $Q(0)$, consider a state $x = (x_1, x_2, x_3, x_4)^T \in \mathcal{T}$. First, suppose $x_1 + x_3 + x_4 \neq 0$. By having the one-step transitions along $v_2 = (0, -1, 0, 0)^T$ occurring x_2 times where $Q_{z,z+v_2}(0) = \frac{k_E^A}{V} (x_3 + x_4 + 2x_1) z_2 > 0$ for all $z = (x_1, z_2, x_3, x_4)^T$ and $1 \leq z_2 \leq x_2$, then having one-step transitions along $v_3 = (0, 0, 1, 0)^T$ occurring $D_{\text{tot}} - x_1 - x_3 - x_4$ times where $Q_{z,z+v_3}(0) = (D_{\text{tot}} - (x_1 + z_3 + x_4)) \left(k_{W0}^1 + k_W^1 + \frac{k'_M}{V} (x_1 + x_4) \right) > 0$ for all $z = (x_1, 0, z_3, x_4)^T$ and $x_3 \leq z_3 \leq D_{\text{tot}} - x_1 - x_4 - 1$, then having one-step transitions along $v_9 = (1, 0, 0, -1)^T$ occurring x_4 times where $Q_{z,z+v_9}(0) = (x_1 + x_4 - z_1) \left(k_{W0}^1 + \frac{k'_M}{V} \frac{x_1 + x_4 + z_1 - 1}{2} \right) > 0$ for all $z = (z_1, 0, D_{\text{tot}} - x_1 - x_4, x_1 + x_4 - z_1)^T$ and $x_1 \leq z_1 \leq x_1 + x_4 - 1$, and finally having one-step transitions along $v_7 = (1, 0, -1, 0)^T$ occurring $D_{\text{tot}} - x_1 - x_4$ times where $Q_{z,z+v_7}(0) = (D_{\text{tot}} - z_1) \left(k_{W0}^2 + \frac{k_M}{V} z_1 + \frac{\bar{k}_M}{V} \frac{D_{\text{tot}} + z_1 - 1}{2} \right) > 0$ for all $z = (z_1, 0, D_{\text{tot}} - z_1, 0)^T$ and $x_1 + x_4 \leq z_1 \leq D_{\text{tot}} - 1$, we have a positive probability of transition from x to r under $Q(0)$. By SI-Lemma 3.10.6 and the fact that r is an absorbing state, we have that x is a transient state for X^0 . On the other hand, suppose $x_1 + x_3 + x_4 = 0$. Since $x = (0, x_2, 0, 0)^T \in \mathcal{T}$, we have $0 \leq x_2 \leq D_{\text{tot}} - 1$. We can first have a one-step transition along $v_3 = (0, 0, 1, 0)^T$, where $Q_{x,x+v_3}(0) = (D_{\text{tot}} - x_2)(k_{W0}^1 + k_W^1) > 0$, to reach the state $(0, x_2, 1, 0)^T$ and then take the steps in the $x_1 + x_3 + x_4 \neq 0$ case. In

this way, there is a positive probability of transition from x to the absorbing state r , and thus x is transient by SI-Lemma 3.10.6.

Verification of Assumption 3.5.2. To show that Assumption 3.5.2 holds, consider the continuous time Markov chain $\tilde{X}$ with infinitesimal generator $\tilde{Q}$ as described in (3.5.6) and shown in Figure 3.7(d). We will first see that $\{i_{m-1}, i_m, r\}$, with $i_{m-1} = r + v_{10}$ and $i_m = r + v_8$, forms a closed class under $\tilde{Q}$. For this, we see that $Q_{r, r+v_j}(\varepsilon)$ vanishes for every $j = \{1, 2, 3, 4, 5, 6, 7, 9\}$ and $\varepsilon \geq 0$, while $Q_{r, r+v_8}(\varepsilon) = \varepsilon \mu b \frac{k_M^A}{V} D_{\text{tot}}^2$ and $Q_{r, r+v_{10}}(\varepsilon) = \varepsilon \mu' \beta \frac{k_M^A}{V} D_{\text{tot}}^2$. Therefore, the only transitions from r under $\tilde{Q}$ are to i_{m-1} with rate $\tilde{Q}_{r, i_{m-1}} = \mu' \beta \frac{k_M^A}{V} D_{\text{tot}}^2 > 0$ and to i_m with rate $\tilde{Q}_{r, i_m} = \mu b \frac{k_M^A}{V} D_{\text{tot}}^2 > 0$. From (3.8.6), we can see that $Q_{i_{m-1}, i_{m-1}+v_j}(0) = 0$ for $j \in \{1, 2, 3, 4, 5, 6, 7, 8, 10\}$, $Q_{i_m, i_m+v_j}(0) = 0$ for $j \in \{1, 2, 3, 4, 5, 6, 8, 9, 10\}$, $Q_{i_{m-1}, r}(0) = Q_{i_{m-1}, i_{m-1}+v_9}(0) = k_{W0}^1 + \frac{k'_M}{V} (D_{\text{tot}} - 1) > 0$ and $Q_{i_m, r}(0) = Q_{i_m, i_m+v_7}(0) = k_{W0}^2 + \frac{k_M}{V} (D_{\text{tot}} - 1) + \frac{\bar{k}_M}{V} (D_{\text{tot}} - 1) > 0$. Therefore, $\{i_{m-1}, i_m, r\}$ forms a closed class under $\tilde{Q}$. The fact that $\hat{X}_{\mathcal{A}}$ consists of erasing the times from $\tilde{X}$ in which the process is in $\mathcal{T}$, together with Lemma 3.5.2, yields that r is an absorbing state under $Q_{\mathcal{A}}$. From (3.8.6), we can see that $\tilde{Q}_{a, i_1} > 0$ where $i_1 = (0, D_{\text{tot}} - 1, 0, 0)^T$. From the analysis made to prove Assumption 3.5.1, we obtain that i_1 leads to r under $\tilde{Q}$, which is part of the closed class $\{i_{m-1}, i_m, r\}$. By interpreting $\hat{X}_{\mathcal{A}}$ again as a time-change of $\tilde{X}$, by Lemma 3.5.2 we obtain that a is transient under $Q_{\mathcal{A}}$. As a consequence, $Q_{\mathcal{A}}$ has a single recurrent class consisting of the state r . Thus, Assumption 3.5.2 holds, and furthermore, $\alpha = [\alpha_a, \alpha_r]$ with $\alpha_a = 0$ and $\alpha_r = 1$. In addition, the previous arguments show that neither Assumption 3.5.3 nor 3.5.4 holds for this model.

Stationary distribution. Here, we derive an expression for $\pi_x^{(1)}$, $x \in \mathcal{T} = \{i_1, \dots, i_m\}$.

Matrices A_1 , S_1 , and R_0 can be written as

$$A_1 = \begin{bmatrix} -s_1 & 0 \\ 0 & -(s_2 + s_3) \end{bmatrix}, \quad S_1 = \begin{bmatrix} s_1 & 0 & \dots & \dots & 0 & 0 \\ 0 & \dots & \dots & 0 & s_2 & s_3 \end{bmatrix}, \quad R_0 = \begin{bmatrix} r_1 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & r_2 \\ 0 & r_3 \end{bmatrix},$$

with $s_1 = \frac{k_M^A}{V} D_{\text{tot}}^2$, $s_2 = \mu' \beta \varepsilon \frac{k_M^A}{V} D_{\text{tot}}^2$, $s_3 = \mu b \varepsilon \frac{k_M^A}{V} D_{\text{tot}}^2$, $r_1 = k_{W0}^A + k_W^A + \frac{k_M^A}{V} (D_{\text{tot}} - 1)$, $r_2 = k_{W0}^1 + \frac{k_M'}{V} (D_{\text{tot}} - 1)$ and $r_3 = k_{W0}^2 + (\frac{k_M}{V} + \frac{\bar{k}_M}{V}) (D_{\text{tot}} - 1)$. Now, we determine $\beta^{(1)} = [\pi_{i_1}^{(1)}, \dots, \pi_{i_m}^{(1)}] = \alpha S_1 (-T_0)^{-1}$. Given that the only two elements in T_0 that are different from zero are $(T_0)_{i_{m-1}, i_{m-1}} = (k_{W0}^1 + \frac{k_M'}{V} (D_{\text{tot}} - 1))$ and $(T_0)_{i_m, i_m} = (k_{W0}^2 + (\frac{k_M}{V} + \frac{\bar{k}_M}{V}) (D_{\text{tot}} - 1))$, we obtain $\beta^{(1)} = [0, \dots, 0, \pi_{i_{m-1}}^{(1)}, \pi_{i_m}^{(1)}]$, with

$$\pi_{i_{m-1}}^{(1)} = \frac{\mu' \beta \frac{k_M^A}{V} D_{\text{tot}}^2}{k_{W0}^1 + \frac{k_M'}{V} (D_{\text{tot}} - 1)}, \quad \pi_{i_m}^{(1)} = \frac{\mu b \frac{k_M^A}{V} D_{\text{tot}}^2}{k_{W0}^2 + (\frac{k_M}{V} + \frac{\bar{k}_M}{V}) (D_{\text{tot}} - 1)}.$$

Finally, $\alpha^{(1)} = [\pi_a^{(1)}, \pi_r^{(1)}]$ is the unique vector such that $\alpha^{(1)} Q_{\mathcal{A}} = -\beta^{(1)} (R_1 + T_1 (-T_0)^{-1} R_0)$ and $\alpha^{(1)} \mathbf{1} = -\beta^{(1)} \mathbf{1}$.

As an example, suppose $D_{\text{tot}} = 2$, $\beta = b$, $k_W^1 = k_W^2 = k_W^A = 0$, $k_{W0}^1 = k_{W0}^2 = k_{W0}^A = k_{W0}$ and $k_M' = \bar{k}_M = k_M^A = k_M$. Then, we have that $A_0 = 0$, $S_0 = 0$ and matrices $A_1 \in \mathbb{R}^{2 \times 2}$ and $S_1 \in \mathbb{R}^{2 \times 13}$ are equal to

$$A_1 = \begin{bmatrix} -4 \frac{k_M}{V} & 0 \\ 0 & -4 \frac{k_M}{V} (\mu b + \mu' \beta) \end{bmatrix},$$

$$S_1 = \begin{bmatrix} 4\frac{k_M}{V} & 0 & \dots & \dots & 0 & 0 \\ 0 & \dots & \dots & 0 & 4\frac{k_M}{V}\mu'b & 4\frac{k_M}{V}\mu b \end{bmatrix}.$$

Furthermore, $R_1 = 0$ and $R_0 \in \mathbb{R}^{13 \times 2}$ can be written as

$$R_0 = \begin{bmatrix} k_{W0} + \frac{k_M}{V} & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & k_{W0} + \frac{k_M}{V} \\ 0 & k_{W0} + 2\frac{k_M}{V} \end{bmatrix}.$$

Finally, matrices T_0 and T_1 can be written as

$$T_0 = \begin{bmatrix} T_0^1 & T_0^2 \\ T_0^3 & T_0^4 \end{bmatrix}, \quad T_1 = \begin{bmatrix} T_1^1 & T_1^2 \\ T_1^3 & T_1^4 \end{bmatrix},$$

with $T_1^2 = 0^{7 \times 6}$ and

$$T_0^1 = \begin{bmatrix} -3k_{W0} - \frac{k_M}{V} & k_{W0} & 0 & 0 & 0 & k_{W0} & 0 \\ \mu \frac{k_E^A}{V} & -(1 + \mu) \frac{k_E^A}{V} - k_{W0} & 0 & \frac{k_E^A}{V} & 0 & 0 & 0 \\ 2k_{W0} & 0 & -6k_{W0} & 2k_{W0} & 0 & 0 & 2k_{W0} \\ 0 & k_{W0} & 0 & -(4k_{W0} + 2\frac{k_M}{V}) & k_{W0} + \frac{k_M}{V} & 0 & 0 \\ 0 & 0 & 0 & 0 & -(k_{W0} + \frac{k_M}{V}) & 0 & 0 \\ \mu' \frac{k_E^A}{V} & 0 & 0 & 0 & 0 & -(1 + \mu') \frac{k_E^A}{V} - k_{W0} & \frac{k_E^A}{V} \\ 0 & 0 & 0 & 0 & 0 & k_{W0} & -(4k_{W0} + \frac{k_M}{V}) \end{bmatrix},$$

$$T_0^2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & k_{W0} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ k_{W0} + \frac{k_M}{V} & 0 & 0 & k_{W0} & 0 & 0 \\ 0 & 0 & 0 & 0 & k_{W0} + \frac{k_M}{V} & 0 \\ 0 & 0 & k_{W0} & 0 & 0 & 0 \\ k_{W0} + \frac{k_M}{V} & k_{W0} & 0 & k_{W0} & 0 & 0 \end{bmatrix},$$

$$T_0^3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mu' \frac{k_E^A}{V} & 0 & 0 & 0 & \mu \frac{k_E^A}{V} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$T_0^4 = \begin{bmatrix} -(2k_{W0} + \frac{k_M}{V}) & 0 & 0 & 0 & k_{W0} + \frac{k_M}{V} & k_{W0} \\ 0 & -(k_{W0} + \frac{k_M}{V}) & 0 & 0 & 0 & k_{W0} + \frac{k_M}{V} \\ 0 & 0 & -(\mu + \mu' + 2) \frac{k_E^A}{V} & 2 \frac{k_E^A}{V} & 0 & 0 \\ 0 & 0 & k_{W0} & -3(k_{W0} + \frac{k_M}{V}) & k_{W0} + 2 \frac{k_M}{V} & k_{W0} + \frac{k_M}{V} \\ 0 & 0 & 0 & 0 & -(k_{W0} + \frac{k_M}{V}) & 0 \\ 0 & 0 & 0 & 0 & 0 & -(k_{W0} + 2 \frac{k_M}{V}) \end{bmatrix},$$

$$T_1^1 = \begin{bmatrix} -2 \frac{k_M}{V} & 0 & 2 \frac{k_M}{V} & 0 & 0 & 0 & 0 \\ 2\mu b \frac{k_M}{V} & -2(\mu b + 1) \frac{k_M}{V} & 0 & 2 \frac{k_M}{V} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2\mu b \frac{k_M}{V} & -2\mu b \frac{k_M}{V} & 0 & 0 & 0 \\ 2\mu' b \frac{k_M}{V} & 0 & 0 & 0 & 0 & -2(\mu' b + 1) \frac{k_M}{V} & 2 \frac{k_M}{V} \\ 0 & 0 & 2\mu' b \frac{k_M}{V} & 0 & 0 & 0 & -2\mu' b \frac{k_M}{V} \end{bmatrix},$$

$$T_1^3 = \begin{bmatrix} 0 & 0 & 0 & 2\mu'b\frac{k_M}{V} & 0 & 0 & 2\mu b\frac{k_M}{V} \\ 0 & 0 & 0 & 0 & 0 & 0 & 2\mu'b\frac{k_M}{V} \\ 0 & 2\mu'b\frac{k_M}{V} & 0 & 0 & 0 & 2\mu b\frac{k_M}{V} & 0 \\ 0 & 0 & 0 & 2\mu'b\frac{k_M}{V} & 0 & 0 & 2\mu b\frac{k_M}{V} \\ 0 & 0 & 0 & 0 & 2\mu'b\frac{k_M}{V} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$T_1^4 = \begin{bmatrix} -2(\mu + \mu')b\frac{k_M}{V} & 0 & 0 & 0 & 0 & 0 \\ 0 & -2\mu'b\frac{k_M}{V} & 0 & 0 & 0 & 0 \\ 0 & 0 & -2(\mu b + \mu'b + 1)\frac{k_M}{V} & 2\frac{k_M}{V} & 0 & 0 \\ 0 & 0 & 0 & -2(\mu + \mu')b\frac{k_M}{V} & 0 & 0 \\ 2\mu b\frac{k_M}{V} & 0 & 0 & 2\mu b\frac{k_M}{V} & -2(2\mu + \mu')b\frac{k_M}{V} & 0 \\ 2\mu'b\frac{k_M}{V} & 2\mu b\frac{k_M}{V} & 0 & 2\mu'b\frac{k_M}{V} & 0 & -2(\mu + 2\mu')b\frac{k_M}{V} \end{bmatrix},$$

By SI-Theorem 3.10.1, we first obtain that $\pi(0) = \pi^{(0)} = [\alpha, 0] = [\alpha_a, \alpha_r, 0, \dots, 0]$ where α can be obtained by solving $\alpha Q_{\mathcal{A}} = 0$ and

$$Q_{\mathcal{A}} = \frac{K_1((\mu')^2 K_2 + (\mu)^2 K_3 + \mu\mu' K_4 + \mu' K_5 + \mu K_6 + K_7)}{\mu'\mu(\mu' + \mu)K_8 + (\mu')^2 K_9 + (\mu)^2 K_{10} + \mu\mu' K_{11} + \mu' K_{12} + \mu K_{13} + K_{14}} \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix},$$

with

$$\begin{aligned} K_1 &= 4k_{W0}\frac{k_M}{V}, \\ K_2 &= \left(\frac{k_E^A}{V}\right)^2 \left(6\left(\frac{k_M}{V}\right)^3 + 39k_{W0}\left(\frac{k_M}{V}\right)^2 + 68\frac{k_M}{V}(k_{W0})^2 + 32(k_{W0})^3\right), \\ K_3 &= \left(\frac{k_E^A}{V}\right)^2 \left(6\left(\frac{k_M}{V}\right)^3 + 36k_{W0}\left(\frac{k_M}{V}\right)^2 + 64\frac{k_M}{V}(k_{W0})^2 + 32(k_{W0})^3\right), \\ K_4 &= \left(\frac{k_E^A}{V}\right)^2 \left(12\left(\frac{k_M}{V}\right)^3 + 75k_{W0}\left(\frac{k_M}{V}\right)^2 + 132\frac{k_M}{V}(k_{W0})^2 + 64(k_{W0})^3\right), \end{aligned}$$

$$\begin{aligned}
K_5 &= \left(\frac{k_E^A}{V}\right)^2 \left(24 \left(\frac{k_M}{V}\right)^3 + 140k_{W0} \left(\frac{k_M}{V}\right)^2 + 222\frac{k_M}{V}(k_{W0})^2 + 96(k_{W0})^3\right) \\
&\quad + \frac{k_E^A}{V}k_{W0} \left(24 \left(\frac{k_M}{V}\right)^3 + 158k_{W0} \left(\frac{k_M}{V}\right)^2 + 280\frac{k_M}{V}(k_{W0})^2 + 128(k_{W0})^3\right), \\
K_6 &= \left(\frac{k_E^A}{V}\right)^2 \left(24 \left(\frac{k_M}{V}\right)^3 + 134k_{W0} \left(\frac{k_M}{V}\right)^2 + 216\frac{k_M}{V}(k_{W0})^2 + 96(k_{W0})^3\right) \\
&\quad + \frac{k_E^A}{V}k_{W0} \left(24 \left(\frac{k_M}{V}\right)^3 + 152k_{W0} \left(\frac{k_M}{V}\right)^2 + 272\frac{k_M}{V}(k_{W0})^2 + 128(k_{W0})^3\right), \\
K_7 &= \left(\frac{k_E^A}{V}\right)^2 \left(24 \left(\frac{k_M}{V}\right)^3 + 124k_{W0} \left(\frac{k_M}{V}\right)^2 + 180\frac{k_M}{V}(k_{W0})^2 + 72(k_{W0})^3\right) \\
&\quad + \frac{k_E^A}{V}k_{W0} \left(48 \left(\frac{k_M}{V}\right)^3 + 284k_{W0} \left(\frac{k_M}{V}\right)^2 + 456\frac{k_M}{V}(k_{W0})^2 + 192(k_{W0})^3\right) \\
&\quad + (k_{W0})^2 \left(24 \left(\frac{k_M}{V}\right)^3 + 160k_{W0} \left(\frac{k_M}{V}\right)^2 + 288\frac{k_M}{V}(k_{W0})^2 + 128(k_{W0})^3\right), \\
K_8 &= \left(\frac{k_E^A}{V}\right)^2 \left(6 \left(\frac{k_M}{V}\right)^4 + 48k_{W0} \left(\frac{k_M}{V}\right)^3 + 126(k_{W0})^2 \left(\frac{k_M}{V}\right)^2 + 132\frac{k_M}{V}(k_{W0})^3 + 72(k_{W0})^4\right), \\
K_9 &= \left(\frac{k_E^A}{V}\right)^2 \left(6 \left(\frac{k_M}{V}\right)^4 + 51k_{W0} \left(\frac{k_M}{V}\right)^3 + 146(k_{W0})^2 \left(\frac{k_M}{V}\right)^2 + 168\frac{k_M}{V}(k_{W0})^3 + 64(k_{W0})^4\right), \\
K_{10} &= \left(\frac{k_E^A}{V}\right)^2 \left(6 \left(\frac{k_M}{V}\right)^4 + 48k_{W0} \left(\frac{k_M}{V}\right)^3 + 136(k_{W0})^2 \left(\frac{k_M}{V}\right)^2 + 160\frac{k_M}{V}(k_{W0})^3 + 64(k_{W0})^4\right), \\
K_{11} &= \left(\frac{k_E^A}{V}\right)^2 \left(24 \left(\frac{k_M}{V}\right)^4 + 191k_{W0} \left(\frac{k_M}{V}\right)^3 + 509(k_{W0})^2 \left(\frac{k_M}{V}\right)^2 + 547\frac{k_M}{V}(k_{W0})^3 + 200(k_{W0})^4\right), \\
&\quad + \frac{k_E^A}{V}k_{W0} \left(12 \left(\frac{k_M}{V}\right)^4 + 96k_{W0} \left(\frac{k_M}{V}\right)^3 + 252(k_{W0})^2 \left(\frac{k_M}{V}\right)^2 + 264\frac{k_M}{V}(k_{W0})^3 + 96(k_{W0})^4\right), \\
K_{12} &= \left(\frac{k_E^A}{V}\right)^2 \left(19 \left(\frac{k_M}{V}\right)^4 + 149k_{W0} \left(\frac{k_M}{V}\right)^3 + 416(k_{W0})^2 \left(\frac{k_M}{V}\right)^2 + 463\frac{k_M}{V}(k_{W0})^3 + 168(k_{W0})^4\right), \\
&\quad + \frac{k_E^A}{V}k_{W0} \left(18 \left(\frac{k_M}{V}\right)^4 + 161k_{W0} \left(\frac{k_M}{V}\right)^3 + 489(k_{W0})^2 \left(\frac{k_M}{V}\right)^2 + 588\frac{k_M}{V}(k_{W0})^3 + 224(k_{W0})^4\right), \\
K_{13} &= \left(\frac{k_E^A}{V}\right)^2 \left(18 \left(\frac{k_M}{V}\right)^4 + 143k_{W0} \left(\frac{k_M}{V}\right)^3 + 399(k_{W0})^2 \left(\frac{k_M}{V}\right)^2 + 452\frac{k_M}{V}(k_{W0})^3 + 168(k_{W0})^4\right), \\
&\quad + \frac{k_E^A}{V}k_{W0} \left(18 \left(\frac{k_M}{V}\right)^4 + 158k_{W0} \left(\frac{k_M}{V}\right)^3 + 476(k_{W0})^2 \left(\frac{k_M}{V}\right)^2 + 576\frac{k_M}{V}(k_{W0})^3 + 224(k_{W0})^4\right), \\
K_{14} &= \left(\frac{k_E^A}{V}\right)^2 \left(12 \left(\frac{k_M}{V}\right)^4 + 98k_{W0} \left(\frac{k_M}{V}\right)^3 + 276(k_{W0})^2 \left(\frac{k_M}{V}\right)^2 + 306\frac{k_M}{V}(k_{W0})^3 + 108(k_{W0})^4\right), \\
&\quad + \frac{k_E^A}{V}k_{W0} \left(24 \left(\frac{k_M}{V}\right)^4 + 214k_{W0} \left(\frac{k_M}{V}\right)^3 + 654(k_{W0})^2 \left(\frac{k_M}{V}\right)^2 + 780\frac{k_M}{V}(k_{W0})^3 + 288(k_{W0})^4\right), \\
&\quad + (k_{W0})^2 \left(12 \left(\frac{k_M}{V}\right)^4 + 116k_{W0} \left(\frac{k_M}{V}\right)^3 + 384(k_{W0})^2 \left(\frac{k_M}{V}\right)^2 + 496\frac{k_M}{V}(k_{W0})^3 + 192(k_{W0})^4\right).
\end{aligned}$$

Let us now derive an expression for $\pi^{(1)}$. Starting with the transient states $\mathcal{T} = \{i_1, \dots, i_{15}\}$, we obtain that $\beta^{(1)} = [\pi_{i_1}^{(1)}, \dots, \pi_{i_{15}}^{(1)}] = \alpha S_1(-T_0)^{-1}$, and so $\beta^{(1)} = [0, \dots, 0, \pi_{i_{14}}^{(1)}, \pi_{i_{15}}^{(1)}]$, with

$$\pi_{i_{12}}^{(1)} = \frac{4\mu'\beta\frac{k_M^A}{V}}{k_{W0}^1 + \frac{k_M'}{V}}, \quad \pi_{i_{13}}^{(1)} = \frac{4\mu b\frac{k_M^A}{V}}{k_{W0}^2 + (\frac{k_M}{V} + \frac{\bar{k}_M}{V})}.$$

Finally, $\alpha^{(1)} = [\pi_a^{(1)}, \pi_r^{(1)}]$ is the unique vector such that $\alpha^{(1)}Q_{\mathcal{A}} = -\beta^{(1)}[R_1 + T_1(-T_0)^{-1}R_0]$ and $\alpha^{(1)}\mathbf{1} = -\beta^{(1)}\mathbf{1}$. After some calculations, we obtain

$$\begin{aligned} \pi_a^{(1)} &= \frac{(\mu\mu')^2 K_{15}((\mu + \mu')K_{16} + K_{17})}{K_{20}((\mu')^2 K_2 + (\mu)^2 K_3 + (\mu' + \mu)K_4 + \mu'K_5 + \mu K_6 + K_7)}, \\ \pi_r^{(1)} &= -\pi_a^{(1)} - \pi_{i_{14}}^{(1)} - \pi_{i_{15}}^{(1)} \\ &= -\frac{(\mu\mu')^2 K_{15}((\mu + \mu')K_{16} + K_{17})}{K_{20}((\mu')^2 K_2 + (\mu)^2 K_3 + (\mu' + \mu)K_4 + \mu'K_5 + \mu K_6 + K_7)} - \mu'K_{18} - \mu K_{19}, \end{aligned}$$

with

$$\begin{aligned} K_{15} &= 2\frac{k_E^A}{V}\frac{k_M}{V}\left(3\frac{k_M}{V} + 2k_{W0}\right)b^2, \quad K_{16} = \frac{k_E^A}{V}\left(2\left(\frac{k_M}{V}\right)^2 + 16(k_{W0})^2 + 12\frac{k_M}{V}k_{W0}\right), \\ K_{17} &= \frac{k_E^A}{V}\left(21\frac{k_M}{V}k_{W0} + 4\left(\frac{k_M}{V}\right)^2 + 24(k_{W0})^2\right) + k_{W0}\left(24\frac{k_M}{V}k_{W0} + 4\left(\frac{k_M}{V}\right)^2 + 32(k_{W0})^2\right), \\ K_{18} &= \frac{4b\frac{k_M}{V}}{k_W + \frac{k_M}{V}}, \quad K_{19} = \frac{4b\frac{k_M}{V}}{k_W + 2\frac{k_M}{V}}, \quad K_{20} = \left(2\frac{k_M}{V} + k_{W0}\right). \end{aligned}$$

Time to memory loss. As a reminder, we denote the time to memory loss of the active state by $h_{a,r}(\varepsilon)$ and the time to memory loss of the repressed state by $h_{r,a}(\varepsilon)$. Let us start by deriving the order and the leading coefficients of $h_{a,r}(\varepsilon)$ and $h_{r,a}(\varepsilon)$. By (3.6.1), the order of the stationary distribution at a and r are $k_a = -\min\{1 - 2, 0\} = 1$ and $k_r = -\min\{1 - 1, 0\} = 0$, respectively. This is consistent with the results obtained in Section 3.8.2. As obtained for the 3D model, here we obtain $\pi_r^{(0)} = 1$ and $\pi_a^{(1)} > 0$, and

thus

$$h_{a,r}(\varepsilon) = \frac{1}{(Q_{\mathcal{A}})_{a,r}} \frac{1}{\varepsilon} + O(1), \quad \text{and} \quad h_{r,a}(\varepsilon) = \frac{1}{(Q_{\mathcal{A}})_{a,r} \cdot \pi_a^{(1)}} \frac{1}{\varepsilon^2} + O\left(\frac{1}{\varepsilon}\right).$$

Verification of conditions of Theorems S.2 and 3.4 in Campos et al. [14].

Now, in order to exploit Theorems S.2 and 3.4 from Campos et al. [14] and determine how μ' affects $h_{a,r}(\varepsilon)$ and $h_{r,a}(\varepsilon)$, we introduce a small approximation in the transition rates of X^ε , namely, $\frac{x_3-1}{2} \approx x_3$ and $\frac{x_4-1}{2} \approx x_4$ in $f_{R121}(x)$ and $f_{R122}(x)$, respectively. This approximation can be justified by introducing the reasonable assumption that each nucleosome characterized by a repressive modification (D_1^R and D_2^R) has the ability to catalyze the establishment of the opposite repressive mark on itself. Now, let us verify that both conditions (i) and (ii) of Theorem S.2 in Campos et al. [14] hold. These conditions can be written as follows:

(i) For each $1 \leq j \leq n$, the vector Av_j has entries in $\{-1, 0, 1\}$ only.

(ii) For each $x \in \mathcal{X}$, $1 \leq i \leq m$ and $y \in \partial_i(K_A + x) \cap \mathcal{X}$ we have that for each $1 \leq k \leq s$,

$$\sum_{j \in G_i^{k,-}} \check{\Upsilon}_j(y) \leq \sum_{j \in G_i^{k,-}} \Upsilon_j(x), \quad \text{where } G_i^{k,-} = \{j \in G^k \mid \langle A_{i,\bullet}, v_j \rangle = -1\},$$

and

$$\sum_{j \in G_i^{k,+}} \check{\Upsilon}_j(y) \geq \sum_{j \in G_i^{k,+}} \Upsilon_j(x), \quad \text{where } G_i^{k,+} = \{j \in G^k \mid \langle A_{i,\bullet}, v_j \rangle = 1\}.$$

To verify that these conditions hold, let us first note the ten possible transitions vectors for the continuous time Markov chain $X^\varepsilon(t)$: $v_1 = -v_2 = (1, 0, -1, 0)^T$, $v_3 = -v_4 = (1, 0, 0, -1)^T$, $v_5 = -v_6 = (0, 1, 0, 0)^T$, $v_7 = -v_8 = (0, 0, 1, 0)^T$, $v_9 = -v_{10} = (0, 0, 0, 1)^T$, with the associated infinitesimal transition rates $\Upsilon_1(x) = f_{R121}(x)$, $\Upsilon_2(x) = g_{R121}^\varepsilon(x)$,

$\Upsilon_3(x) = f_{R122}(x)$, $\Upsilon_4(x) = g_{R122}^\varepsilon(x)$, $\Upsilon_5(x) = f_A(x)$, $\Upsilon_6(x) = g_A^\varepsilon(x)$, $\Upsilon_7(x) = f_{R1}(x)$,
 $\Upsilon_8(x) = g_{R1}^\varepsilon(x)$, $\Upsilon_9(x) = f_{R2}(x)$, $\Upsilon_{10}(x) = g_{R2}^\varepsilon(x)$. Let

$$A = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}.$$

Then, $(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} : x \preceq_A w\}$. Consider infinitesimal transition rates $\check{\Upsilon}_i(x)$, $i = 1, 2, \dots, 10$, defined as for $\Upsilon_i(x)$, $i = 1, 2, \dots, 10$, with all the parameters having the same values except that μ' is replaced by $\check{\mu}'$, with $\mu' \geq \check{\mu}'$. Now, condition (i) of Theorem S.2 in Campos et al. [14] holds since $Av_1 = -Av_2 = (0, 0, 1, 0)^T$, $Av_3 = -Av_4 = (0, 1, 0, 0)^T$, $Av_5 = -Av_6 = (-1, 0, 0, 0)^T$, $Av_7 = -Av_8 = (0, 1, 0, 1)^T$ and $Av_9 = -Av_{10} = (0, 0, 1, 1)^T$. Assumption S.1 in Campos et al. [14] holds with $G^1 = \{9, 1\}$, $G^2 = \{10, 2\}$, $G^3 = \{7, 3\}$, $G^4 = \{8, 4\}$, $G^5 = \{5\}$, $G^6 = \{6\}$ and $\sigma(1) = 9$, $\sigma(2) = 1$, $\sigma(3) = 10$, $\sigma(4) = 2$, $\sigma(5) = 7$, $\sigma(6) = 3$, $\sigma(7) = 8$, $\sigma(8) = 4$, $\sigma(9) = 5$, $\sigma(10) = 6$.

To verify that also condition (ii) of Theorem S.2 in Campos et al. [14] holds, let us start with considering $x \in \mathcal{X}$ and $y \in \partial_1(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} : x_2 = w_2, x_1 + x_3 \leq w_1 + w_3, x_1 + x_4 \leq w_1 + w_4, x_1 + x_3 + x_4 \leq w_1 + w_3 + w_4\}$. Given that $\langle A_{1\bullet}, v_5 \rangle = -1$ and $\langle A_{1\bullet}, v_6 \rangle = 1$, we must verify that $\Upsilon_5(x) \geq \check{\Upsilon}_5(y)$ and $\Upsilon_6(x) \leq \check{\Upsilon}_6(y)$. Since $x_2 = y_2, x_1 + x_3 \leq y_1 + y_3, x_1 + x_4 \leq y_1 + y_4, x_1 + x_3 + x_4 \leq y_1 + y_3 + y_4$, then $\Upsilon_5(x) = (D_{\text{tot}} - (x_1 + x_2 + x_3 + x_4)) \left(k_{W0}^A + k_W^A + \frac{k_M^A}{V} x_2 \right) \geq (D_{\text{tot}} - (y_1 + y_2 + y_3 + y_4)) \left(k_{W0}^A + k_W^A + \frac{k_M^A}{V} y_2 \right) = \check{\Upsilon}_5(y)$ and $\Upsilon_6(x) = x_2 \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} + \frac{k_E^A}{V} (x_3 + x_4 + 2x_1) \right) \leq y_2 \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} + \frac{k_E^A}{V} (y_3 + y_4 + 2y_1) \right) = \check{\Upsilon}_6(y)$. For the second boundary, let us consider $x \in \mathcal{X}$ and $y \in \partial_2(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} : x_2 \geq w_2, x_1 + x_3 = w_1 + w_3, x_1 + x_4 \leq w_1 + w_4, x_1 + x_3 + x_4 \leq w_1 + w_3 + w_4\}$. Given that $\langle A_{2\bullet}, v_3 \rangle = \langle A_{2\bullet}, v_7 \rangle = 1$ and $\langle A_{2\bullet}, v_4 \rangle = \langle A_{2\bullet}, v_8 \rangle = -1$, we need to verify that $\Upsilon_3(x) + \Upsilon_7(x) \leq \check{\Upsilon}_3(y) + \check{\Upsilon}_7(y)$ and

$\Upsilon_4(x) + \Upsilon_8(x) \geq \check{\Upsilon}_4(y) + \check{\Upsilon}_8(y)$ hold. Since $x_2 \geq y_2, x_1 + x_3 = y_1 + y_3, x_1 + x_4 \leq y_1 + y_4, x_1 + x_3 + x_4 \leq y_1 + y_3 + y_4$, then $\Upsilon_3(x) + \Upsilon_7(x) = (D_{\text{tot}} - (x_1 + x_2 + x_3)) \left(k_{W0}^1 + \frac{k'_M}{V}(x_1 + x_4) \right) \leq (D_{\text{tot}} - (y_1 + y_2 + y_3)) \left(k_{W0}^1 + \frac{k'_M}{V}(y_1 + y_4) \right) = \check{\Upsilon}_3(y) + \check{\Upsilon}_7(y)$ and $\Upsilon_4(x) + \Upsilon_8(x) = (x_3 + x_1)\mu' \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} \beta + x_2 \frac{k_E^A}{V} \right) \geq (y_3 + y_1)\check{\mu}' \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} \beta + y_2 \frac{k_E^A}{V} \right) = \check{\Upsilon}_4(y) + \check{\Upsilon}_8(y)$. For the third boundary, let us consider $x \in \mathcal{X}$ and $y \in \partial_3(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} : x_2 \geq w_2, x_1 + x_3 \leq w_1 + w_3, x_1 + x_4 = w_1 + w_4, x_1 + x_3 + x_4 \leq w_1 + w_3 + w_4\}$. Given that $\langle A_{3\bullet}, v_1 \rangle = \langle A_{3\bullet}, v_9 \rangle = 1$ and $\langle A_{3\bullet}, v_2 \rangle = \langle A_{3\bullet}, v_{10} \rangle = 1$, we need to verify that $\Upsilon_1(x) + \Upsilon_9(x) \leq \check{\Upsilon}_1(y) + \check{\Upsilon}_9(y)$ and $\Upsilon_2(x) + \Upsilon_{10}(x) \geq \check{\Upsilon}_2(y) + \check{\Upsilon}_{10}(y)$ hold. Since $x_2 \geq y_2, x_1 + x_3 \leq y_1 + y_3, x_1 + x_4 = y_1 + y_4, x_1 + x_3 + x_4 \leq y_1 + y_3 + y_4$, then $\Upsilon_1(x) + \Upsilon_9(x) = (D_{\text{tot}} - (x_1 + x_2 + x_4)) \left(k_{W0}^2 + \frac{k_M}{V}(x_1 + x_4) + \frac{\bar{k}_M}{V}(x_1 + x_3) \right) \leq (D_{\text{tot}} - (y_1 + y_2 + y_4)) \left(k_{W0}^2 + \frac{k_M}{V}(y_1 + y_4) + \frac{\bar{k}_M}{V}(y_1 + y_3) \right) = \check{\Upsilon}_1(y) + \check{\Upsilon}_9(y)$ and $\Upsilon_2(x) + \Upsilon_{10}(x) = (x_1 + x_4)\mu \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} b + x_2 \frac{k_E^A}{V} \right) \geq (y_1 + y_4)\mu \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} b + y_2 \frac{k_E^A}{V} \right) = \check{\Upsilon}_2(y) + \check{\Upsilon}_{10}(y)$. For the last boundary, consider $x \in \mathcal{X}$ and $y \in \partial_4(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} : x_2 \geq w_2, x_1 + x_3 \leq w_1 + w_3, x_1 + x_4 \leq w_1 + w_4, x_1 + x_3 + x_4 = w_1 + w_3 + w_4\}$. Given that $\langle A_{4\bullet}, v_7 \rangle = \langle A_{4\bullet}, v_9 \rangle = 1$ and $\langle A_{4\bullet}, v_8 \rangle = \langle A_{4\bullet}, v_{10} \rangle = -1$, we need to verify that $\Upsilon_7(x) \leq \check{\Upsilon}_7(y), \Upsilon_9(x) \leq \check{\Upsilon}_9(y), \Upsilon_8(x) \geq \check{\Upsilon}_8(y)$ and $\Upsilon_{10}(x) \geq \check{\Upsilon}_{10}(y)$. Since $x_2 \geq y_2, x_1 + x_3 \leq y_1 + y_3, x_1 + x_4 \leq y_1 + y_4, x_1 + x_3 + x_4 = y_1 + y_3 + y_4$, which also imply $x_1 \leq y_1, x_3 \geq y_3, x_4 \geq y_4$, then $\Upsilon_7(x) = (D_{\text{tot}} - (x_1 + x_2 + x_3 + x_4)) \left(k_{W0}^1 + \frac{k'_M}{V}(x_1 + x_4) \right) \leq (D_{\text{tot}} - (y_1 + y_2 + y_3 + y_4)) \left(k_{W0}^1 + \frac{k'_M}{V}(y_1 + y_4) \right) = \check{\Upsilon}_7(y)$, $\Upsilon_9(x) = (D_{\text{tot}} - (x_1 + x_2 + x_3 + x_4)) \left(k_{W0}^2 + \frac{k_M}{V}(x_1 + x_4) + \frac{\bar{k}_M}{V}(x_1 + x_3) \right) \leq (D_{\text{tot}} - (y_1 + y_2 + y_3 + y_4)) \left(k_{W0}^2 + \frac{k_M}{V}(y_1 + y_4) + \frac{\bar{k}_M}{V}(y_1 + y_3) \right) = \check{\Upsilon}_9(y)$, $\Upsilon_8(x) = x_3\mu' \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} \beta + x_2 \frac{k_E^A}{V} \right) \geq y_3\check{\mu}' \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} \beta + y_2 \frac{k_E^A}{V} \right) = \check{\Upsilon}_8(y)$ and $\Upsilon_{10}(x) = x_4\mu \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} b + x_2 \frac{k_E^A}{V} \right) \geq y_4\mu \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} b + y_2 \frac{k_E^A}{V} \right) = \check{\Upsilon}_{10}(y)$. Then, condition (ii) of Theorem S.2 in [14] also holds.

We can then conclude that all of the conditions of Theorem S.2 in [14] hold and so do the conclusions of Theorem 3.4 in [14], as per the remarks in SI-Section S.3 in [14].

3.11 Acknowledgement

Chapter 3 is extracted, with suitable enumeration and formatting differences, from the paper “Analysis of singularly perturbed stochastic chemical reaction networks motivated by applications to epigenetic cell memory” by Bruno, S., Campos, F.A., Fu, Y., Del Vecchio, D. and Williams, R.J., which has appeared in the SIAM Journal on Applied Dynamical Systems. The dissertation author developed systematic methodology to determine leading orders and leading coefficients for asymptotic expansions of mean first passage times, assisted in deriving higher order terms for stationary distribution series expansions, provided many details for examples used to illustrate the theory in the paper, and contributed to the writing of the manuscript.

Bibliography for Chapter 3

- [1] C. David Allis, Marie-Laure Caparros, Thomas Jenuwein, and Danny Reinberg. *Epigenetics (Second Edition)*. Cold Spring Harbor Laboratory Press, 2015.
- [2] Eitan Altman, Konstantin E. Avrachenkov, and Rudesindo Núñez-Queija. “Perturbation analysis for denumerable Markov chains with application to queueing models”. In: *Advances in Applied Probability* 36.3 (2004), pp. 839–853. DOI: 10.1017/s0001867800013148.
- [3] David F. Anderson and Thomas G. Kurtz. “Continuous time Markov chain models for chemical reaction networks”. In: *Design and Analysis of Biomolecular Circuits: Engineering Approaches to Systems and Synthetic Biology*. Springer New York, 2011, pp. 3–42.
- [4] David F. Anderson and Thomas G. Kurtz. *Stochastic Analysis of Biochemical Systems*. Mathematical Biosciences Institute Graduate Lecture Series. Springer International Publishing, 2015.

- [5] Søren Asmussen. *Applied Probability and Queues (Second Edition)*. Stochastic Modelling and Applied Probability. Springer New York, 2003.
- [6] Konstantin E. Avrachenkov, Jerzy A. Filar, and Phil G. Howlett. *Analytic Perturbation Theory and Its Applications*. Society for Industrial and Applied Mathematics, 2013.
- [7] Konstantin E. Avrachenkov and Moshe Haviv. “The first laurent series coefficients for singularly perturbed stochastic matrices”. In: *Linear Algebra and Its Applications* 386 (2004), pp. 243–259. DOI: 10.1016/j.laa.2003.12.047.
- [8] Johel Beltrán and Claudio Landim. “Tunneling and metastability of continuous time Markov chains”. In: *Journal of Statistical Physics* 140.6 (2010), pp. 1065–1114. DOI: 10.1007/s10955-010-0030-9.
- [9] Johel Beltrán and Claudio Landim. “Tunneling and metastability of continuous time Markov chains II, the nonreversible case”. In: *Journal of Statistical Physics* 149.4 (2012), pp. 598–618. DOI: 10.1007/s10955-010-0030-9.
- [10] Robert McCallum Blumenthal and Ronald Kay Getoor. *Markov Processes and Potential Theory*. Pure and Applied Mathematics. Academic Press Inc., 1968.
- [11] Simone Bruno, Ruth J. Williams, and Domitilla Del Vecchio. “Analytical and computational study of the stochastic behavior of a chromatin modification circuit”. In: *2022 IEEE 61st Conference on Decision and Control (CDC)* (2022), pp. 3871–3877. DOI: 10.1109/CDC51059.2022.9992654.
- [12] Simone Bruno, Ruth J. Williams, and Domitilla Del Vecchio. “Epigenetic cell memory: the gene’s inner chromatin modification circuit”. In: *PLOS Computational Biology* 18.4 (2022), pp. 1–27. DOI: 10.1371/journal.pcbi.1009961.

- [13] Simone Bruno, Ruth J. Williams, and Domitilla Del Vecchio. “Model reduction and stochastic analysis of the histone modification circuit”. In: *2022 European Control Conference (ECC)* (2022), pp. 264–271. DOI: 10.23919/ECC55457.2022.9838047.
- [14] Felipe A. Campos, Simone Bruno, Yi Fu, Domitilla Del Vecchio, and Ruth J. Williams. “Comparison theorems for stochastic chemical reaction networks”. In: *Bulletin of Mathematical Biology* 85.39 (2023), pp. 1–39. DOI: 10.1007/s11538-023-01136-5.
- [15] Nessa Carey. *The Epigenetic Revolution*. Columbia University Press, 2012.
- [16] Geoffrey M. Cooper. *The Cell: A Molecular Approach (Second Edition)*. Sinauer Associates, 2000.
- [17] Domitilla Del Vecchio and Richard M. Murray. *Biomolecular Feedback Systems*. Princeton University Press, 2014.
- [18] Ian Dodd, Mille A. Micheelsen, Kim Sneppen, and Geneviève Thon. “Theoretical analysis of epigenetic cell memory by nucleosome modification”. In: *Cell* 129 (2007), pp. 813–822. DOI: 10.1016/j.cell.2007.02.053.
- [19] Gary Felsenfeld and Mark Groudine. “Controlling the double helix”. In: *Nature* 421 (2003), pp. 448–453. DOI: 10.1038/nature01411.
- [20] Crispin W. Gardiner. *Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences (Second Edition)*. Springer Series in Synergetics. Springer Berlin, 1994.
- [21] Donald P. Gaver, Patricia A. Jacobs, and Guy Latouche. “Finite birth-and-death models in randomly changing environments”. In: *Advances in Applied Probability* 16.4 (1984), pp. 715–731. DOI: 10.2307/1427338.
- [22] Refael Hassin and Moshe Haviv. “Mean passage times and nearly uncoupled Markov chains”. In: *SIAM Journal on Discrete Mathematics* 5.3 (1992), pp. 386–397. DOI: 10.1137/0405030.

- [23] Edith Heard and Robert A. Martienssen. “Transgenerational epigenetic inheritance: myths and mechanisms”. In: *Cell* 157.1 (2014), pp. 95–109. DOI: 10.1016/j.cell.2014.02.045.
- [24] Roger A. Horn and Charles R. Johnson. *Matrix Analysis (Second Edition)*. Cambridge University Press, 2012.
- [25] Suming Huang, Michael D. Litt, and C. Ann Blakey. *Epigenetic Gene Expression and Regulation*. Elsevier Science, 2015.
- [26] Dean L. Isaacson and Richard W. Madsen. *Markov Chains, Theory and Applications*. R.E. Krieger Publishing Company, 1985.
- [27] M. James. “The generalised inverse”. In: *The Mathematical Gazette* 62.420 (1978), pp. 109–114. DOI: 10.2307/3617665.
- [28] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer New York, 1991.
- [29] John G. Kemeny and J. Laurie Snell. “Finite Continuous Time Markov Chains”. In: *Theory of Probability & Its Applications* 6.1 (1961), pp. 101–105. DOI: 10.1137/1106012.
- [30] Mirang Kim and Joseph Costello. “DNA methylation: an epigenetic mark of cellular memory”. In: *Experimental & Molecular Medicine* 49.4 (2017), pp. 1–8. DOI: 10.1038/emm.2017.10.
- [31] Thomas M. Liggett. *Continuous Time Markov Processes: An Introduction*. Graduate Studies in Mathematics. American Mathematical Society, 2010.
- [32] Carl D. Meyer. “The role of the group generalized inverse in the theory of finite Markov chains”. In: *SIAM Review* 17.3 (1975), pp. 443–464.
- [33] James R. Norris. *Markov chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997. DOI: 10.1017/CBO9780511810633.

- [34] Sebastian Palacios, Simone Bruno, Ron Weiss, Elia Salibi, Andrew Kane, Katherine Ilia, and Domitilla Del Vecchio. “Analog epigenetic cell memory by graded DNA methylation”. In: *bioRxiv* (2024).
- [35] Zigang Pan and Tamer Basar. “H infinity control of large-scale jump linear systems via averaging and aggregation”. In: *International Journal of Control* 72.10 (1999), pp. 866–881. DOI: 10.1080/002071799220605.
- [36] Randolph Gale Phillips and Petar V. Kokotovic. “A singular perturbation approach to modelling and control of Markov chains”. In: *IEEE Transactions on Automatic Control* 26.5 (1981), pp. 1087–1094. DOI: 10.1109/TAC.1981.1102780.
- [37] Mark Ptashne. “Epigenetics: core misconception”. In: *Proceedings of the National Academy of Sciences of the United States of America* 18 (2013), pp. 7101–7013. DOI: 10.1073/pnas.1305399110.
- [38] Wen Xiong and James E. Ferrell Jr. “A positive-feedback-based bistable ‘memory module’ that governs a cell fate decision”. In: *Nature* (2003), pp. 460–465. DOI: 10.1038/nature02089.
- [39] G. George Yin and Qing Zhang. *Continuous-Time Markov Chains and Applications: A Two-Time-Scale Approach*. Stochastic Modelling and Applied Probability. Springer New York, 2012.
- [40] Tianyi Zhang, Sarah Cooper, and Neil Brockdorff. “The interplay of histone modifications – writers that read”. In: *EMBO Reports* 16.11 (2015), pp. 1467–1481. DOI: 10.15252/embr.201540945.

Chapter 4

Coclique Level Structure

4.1 Abstract

Continuous time Markov chains are commonly used as models for the stochastic behavior of chemical reaction networks. More precisely, these Stochastic Chemical Reaction Networks (SCRNs) are frequently used to gain a mechanistic understanding of how chemical reaction rate parameters impact the stochastic behavior of these systems. One property of interest is mean first passage times (MFPTs) between states. However, deriving explicit formulas for MFPTs can be highly complex. In order to address this problem, we first introduce the concept of *coclique level structure* and develop theorems to determine whether certain SCRNs have this feature by studying associated graphs. Additionally, we develop an algorithm to identify, under specific assumptions, all possible coclique level structures associated with a given SCRN. Finally, we demonstrate how the presence of such a structure in a SCRN allows us to derive closed form formulas for both upper and lower bounds for the MFPTs. Our methods can be applied to SCRNs taking values in a generic finite state space and can also be applied to models with non-mass-action kinetics. We illustrate our results with examples from the biological areas of epigenetics, neurobiology and ecology.

4.2 Introduction

4.2.1 Overview

Stochastic Chemical Reaction Networks (SCRNs) are a class of continuous time Markov chain models used to describe the stochastic dynamics of a chemical system undergoing a series of reactions that change the numbers of molecules of a finite set of species over time. These models can be used to conduct theoretical studies in different biological areas such as chromatin regulation (see for example Bruno et al. [5]), enzymatic kinetics (see for example Kang et al. [14]), and intracellular viral kinetics (see for example Srivastava et al. [21] and Haseltine and Rawlings [13]).

Formulas for mean first passage times (MFPTs) between states of the Markov chains are helpful in studying the stochastic behavior. However, while calculating an explicit formula is relatively straightforward for certain one-dimensional models, such as birth-death processes (see SI-Section 4.8.3), this typically becomes significantly more difficult for higher-dimensional SCRNs. A potential approach to compute MFPTs involves matrix inversion techniques (see for example Bruno et al. [4]). While this method can, in principle, lead to closed-form expressions, it becomes algebraically complex when symbolic parameter dependencies are retained, even for moderate system sizes. As the system size grows, the dimension of the matrix to be inverted increases, leading to higher computational costs and even more complex expressions. In this case, fixing parameter values can help reduce the computational burden, but this also reduces information about how parameters influence the MFPT.

The inability to obtain an explicit analytical expression for the MFPT poses a challenge because having one allows for a clear understanding of how reaction rate parameters affect certain stochastic behavior of the system. One possible approach to overcome and study the effect of parameter variations on system dynamics, without calculating explicit formulae, is to exploit comparison theorems for stochastic processes

(see for example Campos et al. [8]). However, this would require applying the theorem, if applicable, to a fixed parameter or combination of parameters of interest, thereby increasing the complexity of the study compared to having an analytical formula that can be directly analyzed.

In this work, we first introduce the concept of *coclique level structure* and develop theorems to determine whether certain SCRN have this feature by studying associated graphs. We also develop an algorithm to identify, under specific assumptions, all possible coclique level structures associated with such SCRN. Then, we demonstrate how the presence of such a structure in a SCRN allows us to derive closed form formulas for both upper and lower bounds for MFPTs. We apply the theoretical tools developed in this work to multiple examples to illustrate how they can be used to determine coclique level structures, to derive formulas for upper and lower bounds for the MFPTs, and then to understand how key biological parameters affect the stochastic behavior of the system. While our focus is on SCRN, our definition of coclique level structure is for associated continuous time Markov chains with state spaces that are finite subsets of the non-negative integer orthant and in which the set of all possible transition vectors is finite as well. Consequently, our approach potentially has broader applications to other models that have similar characteristics to SCRN.

The structure of the work is the following: we first introduce some background on SCRN and mean first passage times (Section 4.3). We then introduce the notion of coclique level structure (Section 4.4) and describe the main theoretical tools developed in this work (Section 4.5). Finally, we apply our results to multiple examples (Section 4.6) and present some concluding remarks (Section 4.7).

4.2.2 Related work

In this work, we introduce the concept of coclique level structure for suitable SCRN, which will be a partition of the finite state space of an associated continuous time Markov

chain, in which the sets in the partition are level sets of a linear function L on the state space and they are cocliques in the sense of graph theory (see for example van Mieghem [18] and Kelsey and Roney-Douga [15]), i.e., there are no direct transitions between states within a set in the partition. To the best of our knowledge, this concept of coclique level structure has not been previously used to derive a reduced stochastic process. Here, we apply the function L to a continuous time Markov chain, and analyze the result to estimate MFPTs for SCRN. Previous theoretical tools developed to evaluate upper and lower bounds for MFPTs are mostly suitable for computational studies (see Gaver et al. [12]) and often focus on specific models, such as imprecise birth-death chains (see Lopatzidis et al. [17]) or population continuous time Markov chains (see Backenköhler et al. [2]). In contrast to these existing works, the theoretical tools that we develop enable the derivation of closed form formulas for MFPT bounds, making them suitable for analytical analysis.

4.2.3 Terminology and Notation

Denote the set of integers by $\mathbb{Z}$. For an integer $d \geq 2$, we denote by $\mathbb{Z}^d$ the set of d -dimensional vectors with entries in $\mathbb{Z}$. Denote by $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$, the set of non-negative integers. For an integer $d \geq 2$, we denote by $\mathbb{Z}_+^d$ the set of d -dimensional vectors with entries in $\mathbb{Z}_+$. We denote by $\mathbf{0}$, respectively, $\mathbf{1}$, a vector of any dimension where all entries are 0's, respectively, 1's. The size of $\mathbf{0}$ or $\mathbf{1}$ will be understood from the context. The set of real numbers will be denoted by $\mathbb{R}$, $\mathbb{R}_+ = [0, \infty)$, and d -dimensional Euclidean space will be denoted by $\mathbb{R}^d$ for $d \geq 2$. For integers $m, n \geq 1$, the set of $m \times n$ matrices with real-valued entries will be denoted by $\mathbb{R}^{m \times n}$. For a matrix $A \in \mathbb{R}^{m \times n}$, we denote the kernel of A by $\ker(A) := \{u \in \mathbb{R}^n : Au = \mathbf{0}\}$. Vectors are column vectors unless indicated otherwise and a superscript of T will denote the transpose of a vector or matrix.

4.3 Stochastic Chemical Reaction Networks (SCRNs)

In this section, we provide basic definitions for a class of continuous time Markov chains called Stochastic Chemical Reaction Networks (SCRNs). The reader is referred to Anderson and Kurtz [1] for a more in depth introduction to this subject. The models considered in our examples will be SCRNs and the state space of all of our models will be finite.

We assume there is a finite non-empty set $\mathcal{S} = \{S_1, \dots, S_d\}$ of d **species**, and a finite non-empty set $\mathcal{R} \subseteq \mathbb{Z}_+^d \times \mathbb{Z}_+^d$ that represents chemical **reactions**. We assume that $(w, w) \notin \mathcal{R}$ for every $w \in \mathbb{Z}_+^d$. The set $\mathcal{S}$ represents d different molecular species in a system subject to reactions $\mathcal{R}$ which change the number of molecules of some species. For each $(v^-, v^+) \in \mathcal{R}$, the d -dimensional vector v^- (the **reactant vector**) counts how many molecules of each species are consumed in the reaction, while v^+ (the **product vector**) counts how many molecules of each species are produced. The associated reaction is usually written as

$$\sum_{i=1}^d (v^-)_i S_i \longrightarrow \sum_{i=1}^d (v^+)_i S_i.$$

To avoid the use of unnecessary species, we will assume that for each $1 \leq i \leq d$, there exists a vector $w = (w_1, \dots, w_d)^T \in \mathbb{Z}_+^d$ with $w_i > 0$ such that (w, v) or (v, w) is in $\mathcal{R}$ for some $v \in \mathbb{Z}_+^d$, i.e., each species is either a reactant or a product in some reaction. The net change in the quantity of molecules of each species due to a reaction $(v^-, v^+) \in \mathcal{R}$ is described by $v^+ - v^-$ and it is called the associated **reaction vector**. We denote the set of reaction vectors by $\mathcal{V} := \{v \in \mathbb{Z}^d : v = v^+ - v^- \text{ for some } (v^-, v^+) \in \mathcal{R}\}$; we let $n := |\mathcal{V}|$, the size of $\mathcal{V}$; and we enumerate the members of $\mathcal{V}$ as $\{v_1, \dots, v_n\}$. Note that $\mathcal{V}$ does not contain the zero vector because $\mathcal{R}$ has no elements of the form (w, w) . Different reactions might have the same reaction vector. For each $v_k \in \mathcal{V}$ we consider the set $\mathcal{R}_{v_k} := \{(v^-, v^+) \in \mathcal{R} : v_k = v^+ - v^-\}$. The matrix $S \in \mathbb{R}^{d \times n}$ whose columns

are the elements in $\mathcal{V}$ will be called the **stoichiometric matrix**¹. In addition, we define a **conservation vector** m (if there is one) as a d -dimensional non-zero vector such that $m^T S = 0$ and we say that the conservation vector is *unique* if m is unique, up to multiplication by a scalar.

Consider sets of species $\mathcal{S}$ and reactions $\mathcal{R}$, a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, and a collection of functions $\Lambda = \{\Lambda_{(v^-, v^+)} : \mathcal{X} \rightarrow \mathbb{R}_+\}_{(v^-, v^+) \in \mathcal{R}}$ such that for each $x \in \mathcal{X}$ and $(v^-, v^+) \in \mathcal{R}$, if $x + v^+ - v^- \notin \mathcal{X}$, then $\Lambda_{(v^-, v^+)}(x) = 0$. Now, for $1 \leq k \leq n$, define

$$\Upsilon_k(x) := \sum_{(v^-, v^+) \in \mathcal{R}_{v_k}} \Lambda_{(v^-, v^+)}(x). \quad (4.3.1)$$

Note that for each $x \in \mathcal{X}$ and $1 \leq k \leq n$, if $x + v_k \notin \mathcal{X}$, then $\Upsilon_k(x) = 0$. The functions $\{\Lambda_{(v^-, v^+)} : \mathcal{X} \rightarrow \mathbb{R}_+\}_{(v^-, v^+) \in \mathcal{R}}$ are called **propensity** or **intensity** functions. A common form for the propensity functions is the following, which is associated with **mass action kinetics**:

$$\Lambda_{(v^-, v^+)}(x) = \kappa_{(v^-, v^+)} \prod_{i=1}^d (x_i)_{(v^-)_i},$$

where $\{\kappa_{(v^-, v^+)}\}_{(v^-, v^+) \in \mathcal{R}}$ are non-negative constants and for $m, \ell \in \mathbb{Z}_+$, the quantity $(m)_\ell$ is the falling factorial, i.e., $(m)_0 := 1$ and $(m)_\ell := m(m-1) \dots (m-\ell+1)$. A **stochastic chemical reaction network (SCRN)** (associated with $(\mathcal{S}, \mathcal{R}, \mathcal{X}, \Lambda)$) is a continuous time Markov chain X with state space $\mathcal{X}$ and infinitesimal generator Q given for $x, y \in \mathcal{X}$ by

$$Q_{x,y} = \begin{cases} \Upsilon_k(x) & \text{if } y - x = v_k \text{ for some } 1 \leq k \leq n, \\ -\sum_{k=1}^n \Upsilon_k(x) & \text{if } y = x, \\ 0 & \text{otherwise.} \end{cases} \quad (4.3.2)$$

If a SCRN associated with $(\mathcal{S}, \mathcal{R}, \mathcal{X}, \Lambda)$ has a conservation vector $m \neq 0$ and

¹In this stoichiometric matrix, there are no duplicate reaction vectors. This parallels combining reaction intensity functions associated with the same reaction vector as in (4.3.1).

$m^T X(0) = x_{\text{tot}}$ for some integer $x_{\text{tot}} \geq 0$, then $m^T X(t) = x_{\text{tot}}$ for every $t \geq 0$. Consequently, we can reduce the dimension of the continuous time Markov chain describing the system by one. In this work, we will initially be considering SCRN for which $m = (1, \dots, 1)^T$ is a conservation vector. Then, the projected process $\check{X} = (X_1, \dots, X_{d-1})^T$ is again a continuous time Markov chain with finite state space $\check{\mathcal{X}} = \{(x_1, \dots, x_{d-1})^T \in \mathbb{Z}_+^{d-1} : (x_1, \dots, x_{d-1}, x_{\text{tot}} - \sum_{i=1}^{d-1} x_i)^T \in \mathcal{X}\}$. We denote its infinitesimal generator by $\check{Q}$. We will assume that $|\check{\mathcal{X}}| > 1$. Let $\mathcal{B}$ be a nonempty subset of $\check{\mathcal{X}}$ such that $\mathcal{B} \neq \check{\mathcal{X}}$, and let

$$\tau_{\mathcal{B}} := \inf\{t \geq 0 : \check{X}(t) \in \mathcal{B}\}. \quad (4.3.3)$$

Then, the **mean first passage time (MFPT)** (for $\check{X}$) from $x \in \check{\mathcal{X}}$ to $\mathcal{B}$ can be defined as

$$h_{x,\mathcal{B}} = \mathbb{E}[\tau_{\mathcal{B}} \mid \check{X}(0) = x]. \quad (4.3.4)$$

If $\mathcal{B} = \{y\}$ for some $y \in \check{\mathcal{X}}$, we use the notation $h_{x,y} := h_{x,\{y\}}$. After considering this simple projection first, later on in Section 4.5.4, we will consider the situation where there are $p > 1$ linearly independent conservation vectors for a SCRN. Then, we can reduce the dimension of the Markov chain X by p , in which case we also denote the projected process by $\check{X}$, its state space by $\check{\mathcal{X}} \subset \mathbb{Z}_+^{d-p}$ and its infinitesimal generator by $\check{Q}$. The mean first passage time for $\check{X}$ is then defined by (4.3.3)–(4.3.4).

4.4 Coclique level structure for SCRN

In this section, we define the notion of *coclique level structure* for SCRN and introduce some useful assumptions, definitions, and lemmas that we use in this work.

First, consider a SCRN, as defined in Section 4.3, and assume the following:

Assumption 4.4.1. Assume that $d \geq 2$. For each $v \in \mathcal{V} = \{v_1, \dots, v_n\}$, the entries of the reaction vector v consist of $d - 2$ zeros, a single one and a single minus one, in any order.

When satisfied, this assumption implies that the net change in the number of species stemming from a reaction consists of consuming one molecule of a given species to produce a molecule of a different species. For example, when $d = 3$, reactions satisfying this assumption include $S_1 \longrightarrow S_2$ and $S_1 + S_3 \longrightarrow S_2 + S_3$, both of which have associated reaction vector $(-1, 1, 0)^T$.

We consider a directed graph $\mathcal{G}$ associated with a SCRNs satisfying Assumption 4.4.1, in which the vertices represent the species of the system (we will label these vertices $1, \dots, d$ or $S_1, \dots, S_d$) and the directed edges between vertices represent the reaction vectors $\{v_1, \dots, v_n\}$. In this graph, an edge $e_k = (i, j)$, with $k \in \{1, 2, \dots, n\}$ and $i \neq j$, exists if and only if there is a reaction vector $v_k \in \mathcal{V}$ where $v_k(i) = -1$, $v_k(j) = 1$, $v_k(\ell) = 0$ for $\ell \notin \{i, j\}$. Note that, if we have two edges $e_k = (i, j)$ and $e_{k'} = (j, i)$, with $i \neq j$, then $v_k = -v_{k'}$.

Before introducing the definition of coclique level structure, we provide some useful definitions and lemmas that we will use in this section. Given the directed graph $\mathcal{G}$, a **coclique** is a non-empty set of vertices such that any two vertices in the set do not have a direct edge between them. The graph $\mathcal{G}$ is **bipartite** if the vertex set can be partitioned into two non-empty cocliques. The directed graph $\mathcal{G}$ is **weakly connected** if the underlying undirected graph is connected. Note that this implies $n \geq d - 1$. In this section, we will focus on SCRNs whose associated graphs $\mathcal{G}$ are weakly connected, and we will prove results for such SCRNs in Sections 4.5.1–4.5.3.

For a SCRNs whose associated graph $\mathcal{G}$ is not weakly connected, we can decompose $\mathcal{G}$ into finitely many disjoint **weakly connected components**. Each of the weakly connected components of $\mathcal{G}$ is a weakly connected graph and there are no edges between different weakly connected components. In Section 4.5.4, we will show how to leverage the results in Sections 4.5.1–4.5.3 for the single weakly connected component case to treat the general case where $\mathcal{G}$ is a finite disjoint union of weakly connected components.

Lemma 4.4.1. *Consider a SCRN satisfying Assumption 4.4.1 with $d \geq 2$ species, $\mathcal{S} = \{S_1, \dots, S_d\}$, and n reaction vectors, $\mathcal{V} = \{v_1, \dots, v_n\}$. Assume that the associated graph $\mathcal{G}$ is weakly connected. Then, the rank of the associated stoichiometric matrix S , whose columns are given by the elements of $\mathcal{V}$, is equal to $d - 1$. Furthermore, $(1, \dots, 1)^T$ is the unique (up to scalar multiplication) conservation vector associated with the stoichiometric matrix S .*

The proof of Lemma 4.4.1 is given in SI-Section 4.8.1.

Remark 4.4.1. If each vertex in $\mathcal{G}$ is regarded as a species and each edge in $\mathcal{G}$ is regarded as a reaction, then $\mathcal{G}$ can be interpreted as a chemical reaction network in the sense of Feinberg [10] (where the complexes in this network are simply the species). For this, one can determine its deficiency $\delta = d - \ell - s$, where d is the number of complexes (which is the number of species in this case), ℓ is the number of weakly connected components in $\mathcal{G}$ and s is the rank of the stoichiometric matrix S associated with $\mathcal{G}$. One can see by Lemma 4.4.1 that, when $\mathcal{G}$ is weakly connected, the chemical reaction network associated with $\mathcal{G}$ has deficiency zero².

Under the assumptions of Lemma 4.4.1, we can introduce a **projected continuous time Markov chain** $\tilde{X} = \{\tilde{X}(t) : t \geq 0\}$ in which the state $\tilde{x}$ represents the number of molecules of the first $d - 1$ species, that is, $\tilde{x} = (x_1, \dots, x_{d-1})^T$. Note that the choice to express x_d as a function of the other x_i with $i \in \{1, \dots, d - 1\}$ is without loss of generality, since the labeling of the species can always be reordered so that the species chosen to be expressed as a function of the others is the last one. The process $\tilde{X}$ is a continuous time

²If one starts with the chemical reactions associated with a SCRN, for example as in (4.6.1), one can form the associated complex-reaction graph, also known as the Feinberg-Horn-Jackson graph, and compute the deficiency associated with that graph. This graph need not be the same as our graph $\mathcal{G}$, and its deficiency need not necessarily be zero, for example as in (4.6.1).

Markov chain defined on the finite state space

$$\begin{aligned}\check{\mathcal{X}} &:= \left\{ \check{x} = (x_1, \dots, x_{d-1})^T \in \mathbb{Z}_+^{d-1} : \left(x_1, \dots, x_{d-1}, x_{\text{tot}} - \sum_{i=1}^{d-1} x_i \right)^T \in \mathcal{X} \right\} \\ &\subset \left\{ \check{x} = (x_1, \dots, x_{d-1})^T \in \mathbb{Z}_+^{d-1} : x_1 + \dots + x_{d-1} \leq x_{\text{tot}} \right\},\end{aligned}$$

where $x_{\text{tot}} = \sum_{i=1}^d X_i(0)$. We will assume that $|\check{\mathcal{X}}| > 1$, and the infinitesimal generator of $\check{X}$ will be denoted by $\check{Q}$.

We now introduce the definitions of coclique level function and coclique level structure. A **coclique level function** for $\check{X}$ is a linear function $L : \mathbb{Z}^{d-1} \rightarrow \mathbb{Z}$ such that for each $k = 1, \dots, n$,

$$L(\check{v}_k) \in \{-1, +1\}, \quad (4.4.1)$$

where $\check{v}_k \in \mathbb{Z}^{d-1}$ is the vector obtained from v_k by removing the last element. If such an L exists, it can be written as

$$L(x) = b^T x \text{ for } x \in \mathbb{Z}^{d-1} \text{ and some } b \in \mathbb{Z}^{d-1}, \quad (4.4.2)$$

where, upon partitioning the set of edges of the associated graph $\mathcal{G}$ into two disjoint subsets $\mathcal{E}_+ = \{e_k : L(\check{v}_k) = 1\}$ and $\mathcal{E}_- = \{e_k : L(\check{v}_k) = -1\}$ (where one of these may be empty), the vector $b = (b_1, \dots, b_{d-1})^T$ solves the system of equations

$$\sum_{i=1}^{d-1} b_i \check{v}_k(i) = \begin{cases} +1 & \text{if } e_k \in \mathcal{E}_+ \\ -1 & \text{if } e_k \in \mathcal{E}_- \end{cases} \quad \text{for } k = 1, \dots, n. \quad (4.4.3)$$

Finally, for a coclique level function L , the (ordered) partition $\{\mathcal{L}_\ell, \dots, \mathcal{L}_u\}$, with

$$\mathcal{L}_z := \{x \in \check{\mathcal{X}} : L(x) = z\} \text{ for } z = \ell, \ell + 1, \dots, u - 1, u, \quad (4.4.4)$$

is called a **coclique level structure** for $\check{X}$, with

$$\ell = \min\{L(x) : x \in \check{\mathcal{X}}\}, \quad u = \max\{L(x) : x \in \check{\mathcal{X}}\}. \quad (4.4.5)$$

The sets $\mathcal{L}_\ell, \dots, \mathcal{L}_u$ are cocliques in the Markov chain graph for $\check{X}$ which consists of states of $\check{\mathcal{X}}$ with edges given by $\left\{(\check{x}, \check{x} + \check{v}_k) : k = 1, \dots, n, \check{x} \in \check{X}, \check{x} + \check{v}_k \in \check{X}\right\}$.

Lemma 4.4.2. *Consider a SCRN satisfying the assumptions in Lemma 4.4.1. we partition the set of the edges of the associated graph $\mathcal{G}$ into two disjoint subsets $\mathcal{E}_+$ and $\mathcal{E}_-$, where one of these may be empty, then, for this partition $\{\mathcal{E}_+, \mathcal{E}_-\}$, the system (4.4.3) admits either zero or one solution, and the solution, if one exists, has integer entries.*

Proof. Let $\check{S} \in \mathbb{Z}^{(d-1) \times n}$ be the first $(d-1)$ rows of the stoichiometric matrix S associated with the SCRN. Then, the system (4.4.3) can be re-written in matrix-vector form as

$$\check{S}^T b = w, \quad (4.4.6)$$

where w_k is $+1$ or -1 depending on whether $e_k \in \mathcal{E}_+$ or $e_k \in \mathcal{E}_-$, respectively, where $k = 1, \dots, n$. By Lemma 4.4.1, we know that $\ker(S^T) = \text{span}\{\mathbb{1}\}$, and thus the last row of S is a linear combination of the first $(d-1)$ rows of S . This means that removing the last row of S does not affect its rank, and thus $\text{rank}(\check{S}^T) = d-1$. Hence, if w is in the range of $\check{S}^T$ the system (4.4.3) admits a unique solution, while if it is not in the range, the system does not admit a solution.

Suppose that (4.4.3) admits a solution $b \in \mathbb{R}^{d-1}$. Since $\mathcal{G}$ is weakly connected, there is an edge connecting the vertex d with another vertex in $\mathcal{G}$, say vertex i . This means that there exists $v_k \in \mathcal{V}$ such that $v_k(i) = -v_k(d) \in \{-1, 1\}$ and $v_k(\ell) = 0$ for $\ell \in \{1, 2, \dots, d-1\} \setminus \{i\}$. Then, since $b \in \mathbb{R}^{d-1}$ solves the k^{th} equation of (4.4.3), we have that $|b_i| = 1$, which means $b_i \in \mathbb{Z}$. Using similar logic, for a vertex $j \notin \{i, d\}$ in $\mathcal{G}$ such that there is an edge between j and $\{i, d\}$, we have that either $|b_i - b_j| = 1$ or $|b_j| = 1$, and

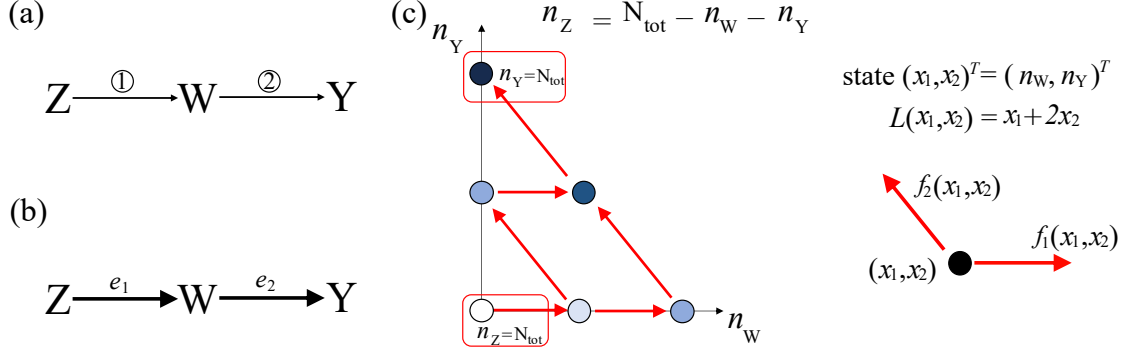
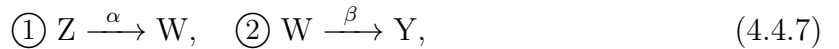


Figure 4.1. Three-species cascade motif: reaction diagram, graph $\mathcal{G}$ and associated Markov chain (Example 4.4.1). (a) Chemical reaction system diagram. The numbers on the arrows correspond to the reactions associated with the arrows as described in (4.4.7) in the main text. (b) Graph $\mathcal{G}$ associated with the chemical reaction system in panel (a). (c) State space and transitions for the projected continuous time Markov chain $\tilde{X} = \{(X_1(t), X_2(t))^T : t \geq 0\}$, which keeps track of (n_W, n_Y) through time. Here, we consider $N_{\text{tot}} = 2$ and we use dots to represent the states, and red double-ended arrows to represent transitions in both directions. Additionally, we use shades of blue to distinguish the level to which each state belongs. The function $L(x_1, x_2)$ associated with the coclique level structure is $L(x_1, x_2) = x_1 + 2x_2$. The rates associated with the one-step transitions for the projected Markov chain $\tilde{X}$ are given in (4.4.9).

thus $b_j \in \mathbb{Z}$. Since $\mathcal{G}$ is weakly-connected and has finitely many vertices, we can iteratively show that all of the entries of b have integer values. $\square$

We use the following example to illustrate our theory in a simple context before giving more complex examples in Section 4.6.

Example 4.4.1. Consider the following SCRN with mass action kinetics, involving three species and two irreversible reactions:



in which $\alpha, \beta > 0$. The diagram of this SCRN is shown in Figure 4.1(a). Let the species vector be $x = (n_W, n_Y, n_Z)$, where n_W, n_Y, n_Z denote the number of molecules of species W, Y, Z, respectively. The total number of molecules in this system is conserved, that is, $n_Z + n_W + n_Y = N_{\text{tot}}$. The reaction vectors associated with this SCRN are $v_1 = (1, 0, -1)^T$ and $v_2 = (-1, 1, 0)^T$. Then, Assumption 4.4.1 is satisfied, and the associated graph

$\mathcal{G}$ is shown in Figure 4.1(b), which is weakly connected. By Lemma 4.4.1, our SCRNs has a unique conservation vector $m = (1, 1, 1)^T$, and then we can introduce a projected continuous time Markov chain $\check{X} = \{(X_1(t), X_2(t))^T : t \geq 0\}$, which keeps track of (n_W, n_Y) through time. Since the total number of molecules N_{tot} is conserved, the state space is $\check{\mathcal{X}} = \{\check{x} = (x_1, x_2)^T \in \mathbb{Z}_+^2 : x_1 + x_2 \leq N_{\text{tot}}\}$. The potential one-step transitions for $\check{X}$ from $x \in \check{\mathcal{X}}$ are illustrated in Figure 4.1(c) (for $N_{\text{tot}} = 2$), where the associated transition vectors are

$$\check{v}_1 = (1, 0)^T, \quad \check{v}_2 = (-1, 1)^T, \quad (4.4.8)$$

and the infinitesimal transition rates are

$$\check{Q}_{\check{x}, \check{x} + \check{v}_1} = f_1(\check{x}) = \alpha(N_{\text{tot}} - (x_1 + x_2)), \quad \check{Q}_{\check{x}, \check{x} + \check{v}_2} = f_2(\check{x}) = \beta x_1. \quad (4.4.9)$$

We now determine a coclique level function and the associated coclique level structure for the projected Markov chain $\check{X}$. To this end, consider the partition of the edge set given by $\mathcal{E}_+ = \{e_1, e_2\}$, $\mathcal{E}_- = \emptyset$. According to Lemma 4.4.2, since $\mathcal{G}$ is weakly connected, and Assumption 4.4.1 is satisfied, the system of equations (4.4.3) admits either zero or one solution. In this case, solving the system of equations

$$b_1 = 1, \quad -b_1 + b_2 = 1$$

yields the unique solution $b = (b_1, b_2)^T = (1, 2)^T$. Therefore, the function $L(x_1, x_2) = x_1 + 2x_2$ is a coclique level function for $\check{X}$. The corresponding coclique level structure is the (ordered) partition $\{\mathcal{L}_\ell, \dots, \mathcal{L}_u\}$ of the state space $\check{\mathcal{X}}$, where each level set is defined as $\mathcal{L}_z := \{x \in \check{\mathcal{X}} : L(x) = z\}$, $z = \ell, \dots, u$, with $\ell = \min \{L(x) : x \in \check{\mathcal{X}}\} = 0$ and $u = \max \{L(x) : x \in \check{\mathcal{X}}\} = 2N_{\text{tot}}$.

It is important to note that this is only one possible coclique level structure for $\check{X}$. In the next section (Section 4.5), we develop theoretical tools to identify all possible

coclique level structures for SCRN s satisfying Assumption 4.4.1.

4.5 Main Results

In Sections 4.5.1–4.5.3, we develop theoretical tools for SCRN s satisfying Assumption 4.4.1 and whose associated graph $\mathcal{G}$ is weakly connected, and in Section 4.5.4, we relax the latter condition and consider the case where $\mathcal{G}$ does not need to be weakly connected. More precisely, in Section 4.5.1, we introduce a theorem to determine all of the possible coclique level functions associated with SCRN s satisfying Assumption 4.4.1 and with $\mathcal{G}$ weakly connected. We then derive theoretical tools to determine when a projected continuous time Markov chain associated with a SCRN admits a coclique level function by studying the structure of the graph $\mathcal{G}$ associated with the SCRN. Additionally, we develop an algorithm to find, under certain assumptions, all of the possible coclique level functions associated with these SCRN s (see Section 4.5.2). We derive analytical formulas for upper and lower bounds on MFPTs for SCRN s having this type of structure (see Section 4.5.3). In Section 4.5.4, we generalize these results to study SCRN s that satisfy Assumption 4.4.1 and whose associated graph $\mathcal{G}$ does not need to be weakly connected. We apply our results to several examples in Section 4.6.

4.5.1 Existence and characterization of coclique level structures

In the first theorem in this section, we show that if (4.4.3) admits a solution, then there is a coclique level function for $\check{X}$. Moreover, all coclique level functions for $\check{X}$ can be obtained in this way.

Theorem 4.5.1. *Consider a SCRN satisfying Assumption 4.4.1. Assume that the associated graph $\mathcal{G}$ is weakly connected. The set of all coclique level functions for $\check{X}$ is the set of all functions of the form (4.4.2), where $b = (b_1, \dots, b_{d-1})^T \in \mathbb{Z}^{d-1}$ is a solution of the system (4.4.3) for some partition $\{\mathcal{E}_+, \mathcal{E}_-\}$ of the edge set of $\mathcal{G}$ (and one of $\mathcal{E}_+$ and $\mathcal{E}_-$*

may be empty).

Proof. Let L be a coclique level function for $\check{X}$. Then, by Section 4.4, L has the form (4.4.2), where $b \in \mathbb{Z}^{d-1}$ satisfies (4.4.3) with $\mathcal{E}_+ = \{e_k : L(\check{v}_k) = 1\}$ and $\mathcal{E}_- = \{e_k : L(\check{v}_k) = -1\}$. Conversely, suppose that b is a solution of the system (4.4.3) for some partition $\{\mathcal{E}_+, \mathcal{E}_-\}$ of the edge set of $\mathcal{G}$ (where $\mathcal{E}_+$ or $\mathcal{E}_-$ may be empty). Then, $b \in \mathbb{Z}^{d-1}$ by Lemma 4.4.2, and with $L(x) = b^T x$ for all $x \in \mathbb{Z}^{d-1}$, L satisfies (4.4.1). Thus, L is a coclique level function for $\check{X}$. $\square$

Remark 4.5.1. For a given partition $\{\mathcal{E}_+, \mathcal{E}_-\}$ of the edge set of $\mathcal{G}$, if the system (4.4.3) has a solution, then by Lemma 4.4.2, it has a unique solution. Furthermore, if we switch the labels of the two subsets, i.e., $\mathcal{E}_+$ becomes $\mathcal{E}_-$ and vice versa, the function L is replaced by $-L$, and the sets in the partition $\{\mathcal{L}_\ell, \dots, \mathcal{L}_u\}$ remain the same but the order is reversed. Therefore, we may consider the coclique level structures associated with L and $-L$ to be the same.

Application to Example 4.4.1: Consider the SCRN introduced in (4.4.7), with associated graph $\mathcal{G}$ shown in Figure 4.1(b). As previously mentioned, by applying Lemma 4.4.1, we know that this SCRN has a unique conservation vector $m = (1, 1, 1)^T$. This allows us to introduce a projected continuous time Markov chain $\check{X} = \{(X_1(t), X_2(t))^T : t \geq 0\}$, which keeps track of (n_W, n_Y) through time, with transition vectors and associated rates given in (4.4.8) and (4.4.9), respectively. Given that Assumption 4.4.1 is satisfied and the associated graph $\mathcal{G}$ is weakly connected, we can use Theorem 4.5.1 to determine all of the coclique level structures for $\check{X}$. To this end, consider all of the possible partitions $\{\mathcal{E}_+, \mathcal{E}_-\}$ of edges of $\mathcal{G}$ that could allow us to determine a coclique level structure. These partitions are the following:

$$\mathcal{E}_+ = \{e_1, e_2\}, \mathcal{E}_- = \emptyset \quad \text{and} \quad \mathcal{E}_+ = \{e_1\}, \mathcal{E}_- = \{e_2\}.$$

We did not consider the partition $\mathcal{E}_+ = \emptyset, \mathcal{E}_- = \{e_1, e_2\}$ or the partition $\mathcal{E}_+ = \{e_2\}, \mathcal{E}_- = \{e_1\}$ because, as explained in Remark 4.5.1, the associated functions L would be the opposite of the ones obtained for the partitions considered above and the resulting coclique level structures are considered to be the same. For each partition, the system of equations in (4.4.3) has a unique solution, these being $(b_1, b_2)^T = (1, 2)^T$ and $(b_1, b_2)^T = (1, 0)^T$, respectively. Then, by applying Theorem 4.5.1, we can conclude that the projected Markov chain $\check{X}$ has two coclique level structures, with associated coclique level functions $L(x_1, x_2) = x_1 + 2x_2$ and $L(x_1, x_2) = x_1$, respectively.

Now, we characterize when there exists a coclique level function for $\check{X}$, assuming Assumption 4.4.1 and the associated graph $\mathcal{G}$ is weakly connected (see Theorem 4.5.3). In the course of this, we show how to derive a simple coclique level function, when one exists, which exploits bipartite structure of $\mathcal{G}$. There can be other coclique level functions and we can use Theorem 4.5.1 above to identify all possible coclique level functions.

Assume the SCRN satisfies Assumption 4.4.1 and its associated graph $\mathcal{G}$ is weakly connected. Consider a weakly directed cycle c in $\mathcal{G}$, i.e., a directed subgraph of $\mathcal{G}$ whose underlying undirected graph is a cycle. We abbreviate weakly directed cycle as wd-cycle³. Choose an orientation for the cycle. Then, define the vector $\vartheta_c \in \mathbb{Z}^n$ associated with the wd-cycle c such that for $k = 1, \dots, n$,

$$\vartheta_c(k) = \begin{cases} +1 & \text{if } e_k \text{ is part of the wd-cycle and } e_k \text{ is in} \\ & \text{the direction of the orientation of the wd-cycle,} \\ -1 & \text{if } e_k \text{ is part of the wd-cycle and } e_k \text{ is in} \\ & \text{the opposite direction of the orientation of the wd-cycle,} \\ 0 & \text{if } e_k \text{ is not part of the wd-cycle.} \end{cases}$$

³Note that we allow two vertices connected by edges e and the reverse of e to be a wd-cycle.

Theorem 4.5.2. *Consider a SCRN satisfying Assumption 4.4.1 and assume its associated graph $\mathcal{G}$ is weakly connected. Let $\mathcal{C}$ be the set of all wd-cycles in $\mathcal{G}$. For each $c \in \mathcal{C}$, we have $\vartheta_c \in \ker(S)$. Furthermore, given a partition $\{\mathcal{E}_+, \mathcal{E}_-\}$ of the edge set of $\mathcal{G}$ (where one of $\mathcal{E}_+, \mathcal{E}_-$ may be empty), we can define a vector w such that $w_k = +1$ if $e_k \in \mathcal{E}_+$ and $w_k = -1$ if $e_k \in \mathcal{E}_-$. Then, if $w^T \vartheta_c \neq 0$ for some $c \in \mathcal{C}$, the system (4.4.3) does not admit a solution $b \in \mathbb{Z}^{d-1}$.*

Proof. The i^{th} entry of the vector $S\vartheta_c$ represents the amount of species S_i consumed/produced after all the reactions associated with the wd-cycle c are triggered, where the signs in ϑ_c ensure that the edges are followed in the direction of the chosen orientation. Since c is a wd-cycle and Assumption 4.4.1 is satisfied, $S\vartheta_c = 0$ and then $\vartheta_c \in \ker(S)$.

Since $\check{S}^T$ is S^T without the last column, we have that $\text{range}(\check{S}^T) \subseteq \text{range}(S^T) = \ker(S)^\perp$. Since $\vartheta_c \in \ker(S)$ for each $c \in \mathcal{C}$, we have that $\text{span}\{\vartheta_c : c \in \mathcal{C}\} \subset \ker(S)$. It follows that if $w^T \vartheta_c \neq 0$ for some $c \in \mathcal{C}$, then $w \notin \ker(S)^\perp \supseteq \text{range}(\check{S}^T)$, and thus (4.4.3) does not admit a solution. $\square$

Corollary 4.5.1. *Consider a SCRN satisfying Assumption 4.4.1 and assume its associated graph $\mathcal{G}$ is weakly connected. Suppose $\{\mathcal{E}_+, \mathcal{E}_-\}$ is a partition of the edge set of $\mathcal{G}$ (where one of $\mathcal{E}_+$ and $\mathcal{E}_-$ can be empty). If there are $v_k, v_{k'} \in \mathcal{V}$ such that $v_k = -v_{k'}$ and the edges associated with them, e_k and $e_{k'}$, belong to the same subset $\mathcal{E}_+$ or $\mathcal{E}_-$, then (4.4.3) does not admit a solution $b \in \mathbb{Z}^{d-1}$, and so there is no coclique level structure for this partition.*

Proof. Suppose that $e_k, e_{k'} \in \mathcal{E}_+$ have the properties described. Consider the wd-cycle c given by e_k and $e_{k'}$, with corresponding vector ϑ_c . Then, $\vartheta_c(\ell) = 0$ for $\ell \notin \{k, k'\}$ and, depending on the orientation of the cycle, $\vartheta_c(k) = \vartheta_c(k') = +1$ or $\vartheta_c(k) = \vartheta_c(k') = -1$. Then, since both the edges belong to $\mathcal{E}_+$, we have w , defined as in (4.4.6), to be such that $w(k) = w(k') = 1$ and $w(\ell) = 0$ for $\ell \notin \{k, k'\}$. This implies that $|w^T \vartheta_c| = 2 \neq 0$. Then, the result follows from Theorem 4.5.2. Similar reasoning can be used for the case where $e_k, e_{k'} \in \mathcal{E}_-$. $\square$

Remark 4.5.2. Based on the results of Corollary 4.5.1, given a partition $\{\mathcal{E}_+, \mathcal{E}_-\}$, we can obtain a coclique level function only if edges e_k and $e_{k'}$ associated with $v_k, v_{k'} \in \mathcal{V}$ such that $v_k = -v_{k'}$ belong to different members of the partition. If this property holds, then the equation (4.4.3) associated with v_k is the opposite of the one associated with $v_{k'}$. This implies that the two equations are providing the same information. Thus, in order to solve system (4.4.3) and reduce the computational cost, we can just write the system for $\hat{\mathcal{V}}$, in which $\hat{\mathcal{V}} \subseteq \mathcal{V}$ is a maximal set of reaction vectors such that, for any $v_k, v_{k'} \in \hat{\mathcal{V}}$, $v_k \neq -v_{k'}$.

Finally, we introduce a theorem demonstrating that for a SCRN satisfying Assumption 4.4.1 and having a weakly connected graph $\mathcal{G}$, there is a coclique level structure for $\check{X}$ if and only if $\mathcal{G}$ is bipartite.

Theorem 4.5.3. *Consider a SCRN satisfying Assumption 4.4.1 and suppose the associated graph $\mathcal{G}$ is weakly connected. If the graph $\mathcal{G}$ is bipartite, then there exists a coclique level function for the projected continuous time Markov chain $\check{X}$. In this case, by relabeling vertices if necessary, there are two disjoint, non-empty cocliques of vertices in $\mathcal{G}$, $B = \{1, \dots, \bar{i}\}$ and $C = \{\bar{i} + 1, \dots, d\}$, and then*

$$L(\check{x}) = L(x_1, \dots, x_{\bar{i}}, x_{\bar{i}+1}, \dots, x_{d-1}) = x_1 + \dots + x_{\bar{i}}$$

defines a coclique level function for $\check{X}$. If the graph $\mathcal{G}$ is not bipartite, there is no coclique level structure for $\check{X}$.

Proof. We start by assuming the graph $\mathcal{G}$ is bipartite, and show there exists a coclique level function for the projected continuous time Markov chain $\check{X}$ as described in the theorem. Suppose $\mathcal{G}$ is bipartite. By relabeling vertices if necessary, there are sets B and C as described in the theorem. Let $\mathcal{E}_+ = \{e_k = (i, j) : v_k \in \mathcal{V}, i \in C \text{ and } j \in B\}$ and $\mathcal{E}_- = \{e_k = (i, j) : v_k \in \mathcal{V}, i \in B \text{ and } j \in C\}$. Note that this $\{\mathcal{E}_+, \mathcal{E}_-\}$ is a partition of the

edge set of $\mathcal{G}$ where one of $\mathcal{E}_+$ or $\mathcal{E}_-$ could be empty. Let $b \in \mathbb{Z}^{d-1}$ be the vector where the first $\bar{i}$ entries are 1 and the remaining $d - 1 - \bar{i}$ entries are zero and define $L(x) = b^T x$ for $x \in \mathbb{Z}^{d-1}$. Since $\mathcal{G}$ is bipartite and B and C are two cocliques, each reaction corresponds either to an edge from a vertex in C to a vertex in B or an edge from a vertex in B to a vertex in C . In particular, when a reaction $v_k \in \mathcal{V}$ is triggered, if $e_k \in \mathcal{E}_+$ then $L(\check{v}_k) = +1$ and if $e_k \in \mathcal{E}_-$ then $L(\check{v}_k) = -1$. Thus, L satisfies (4.4.1) and is a coclique level function for $\check{X}$.

Now suppose the graph $\mathcal{G}$ is not bipartite. Then $\mathcal{G}$ contains at least one odd wd-cycle, c^{odd} (see Theorem 5.1 on page 27 of Wilson [22]). Consider the vector $\vartheta_{c^{odd}}$ associated with this cycle. By definition, $\vartheta_{c^{odd}}$ has all of its entries equal to zero except for an odd number of entries, associated with the edges in the wd-cycle c^{odd} , which take values in $\{-1, +1\}$. For any partition $\{\mathcal{E}_+, \mathcal{E}_-\}$ of the edge set of $\mathcal{G}$ (where one of $\mathcal{E}_+$ or $\mathcal{E}_-$ may be empty), and for w as in (4.4.6), the product $w^T \vartheta_{c^{odd}}$ will result in a sum of an odd number of terms, each of which is $+1$ or -1 , and thus the sum can never be equal to zero. By Theorem 4.5.2, this implies that the system (4.4.3) does not admit a solution. By Theorem 4.5.1, there cannot be a coclique level function for $\check{X}$. $\square$

Application to Example 4.4.1: Consider the SCRN introduced in (4.4.7), with associated graph $\mathcal{G}$ shown in Figure 4.1(b). Earlier, we showed how, by applying Theorem 4.5.1, one can determine all of the coclique level functions for $\check{X}$. On the other hand, by inspecting $\mathcal{G}$, it is also possible to note that the graph is bipartite, and thus we can apply Theorem 4.5.3 to directly identify one of its coclique level functions. Specifically, the two disjoint, non-empty cocliques of vertices in $\mathcal{G}$ are $B = \{1\}$ and $C = \{2, 3\}$, and then $L(x) = x_1$ defines a coclique level function for $\check{X}$.

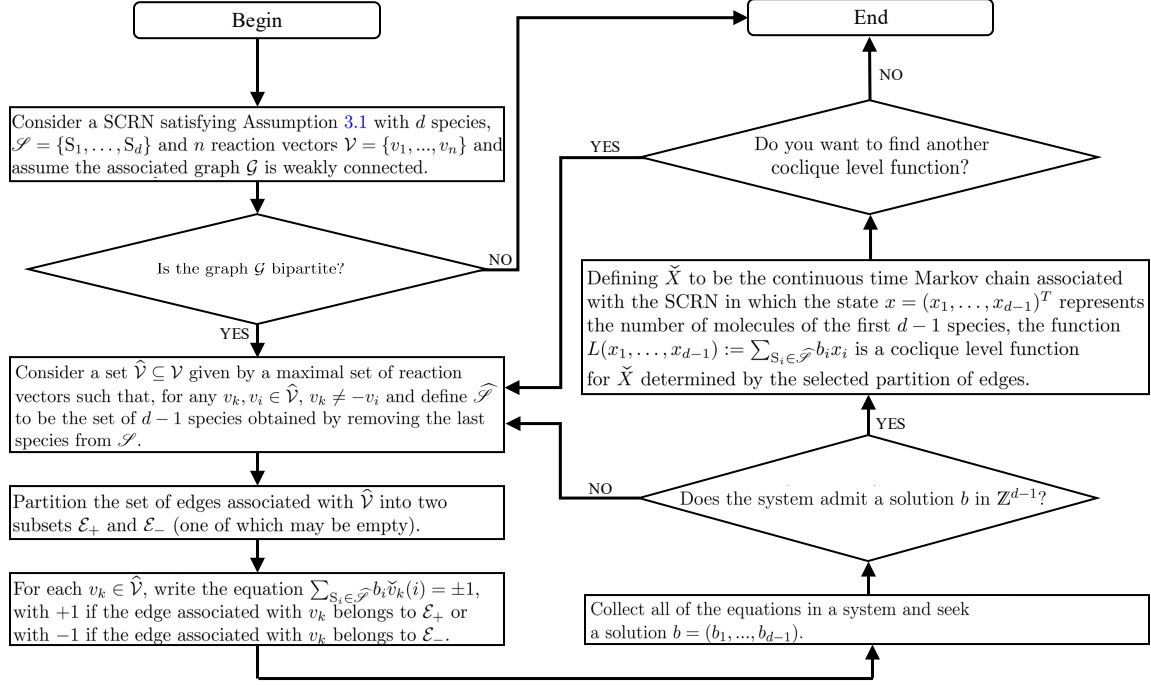


Figure 4.2. Key steps of the algorithm for identifying all coclique level functions for the projected continuous time Markov chain $\tilde{X}$ associated with a SCRN under Assumption 4.4.1 and assuming that $\mathcal{G}$ is weakly connected.

4.5.2 Algorithm to determine all coclique level functions associated with a SCRN satisfying Assumption 4.4.1 and whose associated graph $\mathcal{G}$ is weakly connected

Based on the results given in Theorems 4.5.1 and 4.5.3, given a SCRN satisfying Assumption 4.4.1 and assuming that $\mathcal{G}$ is weakly connected, we can obtain an algorithm that allows us to find all of the coclique level functions for the projected continuous time Markov chain $\tilde{X}$ associated with the SCRN. The steps of the algorithm are described in Figure 4.2. The first step of the algorithm involves checking whether a graph is bipartite. The reader may refer to Section 3.4 in Kleinberg and Tardos [16] for a discussion about a Breadth-First Search implementation to check whether a graph is bipartite, whose run time is linear in terms of the number of vertices in the graph (or, in our case, the number of species in the SCRN). As a second step, the algorithm requires selecting a subset $\hat{\mathcal{V}} \subseteq \mathcal{V}$ defined as a maximal set of reaction vectors such that $v_k \neq -v_i$ for any $v_k, v_i \in \hat{\mathcal{V}}$. Note

that, once such a subset is identified, the specific ordering of the vectors in $\hat{\mathcal{V}}$ does not affect the outcome of the algorithm. Moreover, the choice of $\hat{\mathcal{V}}$ is uniquely determined up to the direction of the vectors and ensures that all directions allowed by the reactions are considered.

Remark 4.5.3. Note that the algorithm described here is for a specific ordering of the species associated with our specification of the SCRNs.

4.5.3 Using coclique level structure to bound MFPTs

As mentioned in the introduction, obtaining an explicit analytical expression for the mean first passage time (MFPT) for one-dimensional birth-death processes is relatively straightforward by using the formula (4.8.3) (see SI-Section 4.8.3). However, it is typically very complicated to obtain an explicit expression for the MFPT for more complex Markov chains, especially in dimension greater than one. In this section, we will show how to obtain closed form formulas for upper and lower bounds for the MFPTs of continuous time Markov chains $\check{X}$ associated with SCRNs having a coclique level structure.

To this end, consider the projected continuous time Markov chain $\check{X}$ with finite state space $\check{\mathcal{X}} \subseteq \mathbb{Z}_+^{d-1}$, associated with a SCRNs satisfying Assumption 4.4.1 and for which $\mathcal{G}$ is weakly connected (as defined in Section 4.4), and having infinitesimal generator $\check{Q}$. Suppose $\check{X}$ has a coclique level function $L : \mathbb{Z}^{d-1} \rightarrow \mathbb{Z}$, defined as in (4.4.1), with (ordered) partition $\{\mathcal{L}_\ell, \mathcal{L}_{\ell+1}, \dots, \mathcal{L}_{u-1}, \mathcal{L}_u\}$ of $\check{\mathcal{X}}$, defined as in (4.4.4) and where $\ell < u$. Given such a coclique level structure for $\check{X}$, we will determine analytical expressions for upper and lower bounds for both the MFPT for $\check{X}$ from $\mathcal{L}_\ell$ to $\mathcal{L}_u$ and the MFPT for $\check{X}$ from $\mathcal{L}_u$ to $\mathcal{L}_\ell$. We first focus on the MFPT from $\mathcal{L}_\ell$ to $\mathcal{L}_u$.

Let

$$G_+ := \{k : L(\check{v}_k) = 1\} \quad \text{and} \quad G_- := \{k : L(\check{v}_k) = -1\}.$$

For $\ell < u$, $\ell \leq z \leq u$ and $\check{x} \in \mathcal{L}_z$, define the “rate of increase” $\lambda_z(\check{x})$ and the “rate of

decrease" $\gamma_z(\tilde{x})$ as follows:

$$\lambda_z(\tilde{x}) = \sum_{k \in G_+} \check{Q}_{\tilde{x}, \tilde{x} + \check{v}_k}, \quad \gamma_z(\tilde{x}) = \sum_{k \in G_-} \check{Q}_{\tilde{x}, \tilde{x} + \check{v}_k}.$$

Then, for each $z \in \{\ell, \ell + 1, \dots, u - 1, u\}$, let

$$\lambda_z^M = \max_{\tilde{x} \in \mathcal{L}_z} \lambda_z(\tilde{x}), \quad \lambda_z^m = \min_{\tilde{x} \in \mathcal{L}_z} \lambda_z(\tilde{x}), \quad \gamma_z^M = \max_{\tilde{x} \in \mathcal{L}_z} \gamma_z(\tilde{x}), \quad \gamma_z^m = \min_{\tilde{x} \in \mathcal{L}_z} \gamma_z(\tilde{x}). \quad (4.5.1)$$

We will define continuous time Markov chains $\check{X}$ and $\hat{X}$ with the same state space as $\check{X}$.

For $\ell \leq z \leq u$ and $\tilde{x} \in \mathcal{L}_z$, let

$$G_+(\tilde{x}) = \{k \in G_+ : \tilde{x} + \check{v}_k \in \check{\mathcal{X}}\} \quad \text{and} \quad G_-(\tilde{x}) = \{k \in G_- : \tilde{x} + \check{v}_k \in \check{\mathcal{X}}\}. \quad (4.5.2)$$

We assume that $G_+(\tilde{x}) : \tilde{x} \in \mathcal{X} \setminus \mathcal{L}_u$ are all non-empty sets when G_+ is non-empty, and $G_-(\tilde{x}) : \tilde{x} \in \mathcal{X} \setminus \mathcal{L}_\ell$ are all non-empty sets when G_- is non-empty. Note that $G_+(\tilde{x})$ is empty when $\tilde{x} \in \mathcal{L}_u$ and $G_-(\tilde{x})$ is empty when $\tilde{x} \in \mathcal{L}_\ell$. Then, we define the infinitesimal generator $\check{Q}$ for $\check{X}$ such that the only positive entries of $\check{Q}$ are given by

$$\check{Q}_{\tilde{x}, \tilde{x} + \check{v}_k} = \frac{\lambda_z^M}{|G_+(\tilde{x})|} \text{ for } k \in G_+(\tilde{x}) \quad \text{and} \quad \check{Q}_{\tilde{x}, \tilde{x} + \check{v}_k} = \frac{\gamma_z^m}{|G_-(\tilde{x})|} \text{ for } k \in G_-(\tilde{x}),$$

for $\ell \leq z \leq u$ and $\tilde{x} \in \mathcal{L}_z$. Similarly, we define the infinitesimal generator $\hat{Q}$ for $\hat{X}$ such that the only positive entries of $\hat{Q}$ are given by

$$\hat{Q}_{\tilde{x}, \tilde{x} + \check{v}_k} = \frac{\lambda_z^m}{|G_+(\tilde{x})|} \text{ for } k \in G_+(\tilde{x}) \quad \text{and} \quad \hat{Q}_{\tilde{x}, \tilde{x} + \check{v}_k} = \frac{\gamma_z^M}{|G_-(\tilde{x})|} \text{ for } k \in G_-(\tilde{x}),$$

for $\ell \leq z \leq u$ and $\tilde{x} \in \mathcal{L}_z$.

Then, by the comparison theorems, i.e., Theorems 3.3 and 3.4 in Campos et al. [8] (see SI-Section 4.8.4), using the matrix A equal to b^T (associated with the coclique

level function L), we have that for $\check{x}, \check{x}$ and $\hat{x}$ in $\mathcal{L}_\ell$, we can realize $\check{X}$ and $\check{X}$ with infinitesimal generators $\check{Q}$ and $\check{Q}$ (resp. $\check{X}$ and $\hat{X}$ with infinitesimal generators $\check{Q}$ and $\hat{Q}$) on the same probability space such that $L(\check{X}) \leq L(\check{X})$, $\check{X}(0) = \check{x}$, $\check{X}(0) = \check{x}$ a.s. (resp. $L(\hat{X}) \leq L(\check{X})$, $\hat{X}(0) = \hat{x}$, $\check{X}(0) = \check{x}$ a.s.). Then, for $\check{\tau}_u = \inf\{t \geq 0 : \check{X}(t) \in \mathcal{L}_u\} = \inf\{t \geq 0 : L(\check{X}(t)) = u\}$, $\check{\tau}_u = \inf\{t \geq 0 : \check{X}(t) \in \mathcal{L}_u\} = \inf\{t \geq 0 : L(\check{X}(t)) = u\}$, and $\hat{\tau}_u = \inf\{t \geq 0 : \hat{X}(t) \in \mathcal{L}_u\} = \inf\{t \geq 0 : L(\hat{X}(t)) = u\}$, we have

$$\mathbb{E}_{\check{x}}[\check{\tau}_u] \leq \mathbb{E}_{\check{x}}[\check{\tau}_u] \leq \mathbb{E}_{\hat{x}}[\hat{\tau}_u]. \quad (4.5.3)$$

By the choice of the rates for $\check{X}$ and $\hat{X}$, we have that $L(\check{X})$ and $L(\hat{X})$ are continuous time Markov chains. In fact, they are simple birth-death processes. It follows that the upper and lower bounds in (4.5.3) can be explicitly evaluated. For the lower bound, suppose that λ_z^M is positive for $\ell \leq z \leq u-1$. We replace λ_i with λ_z^M and γ_i with γ_z^m , respectively, in formula (4.8.7), to obtain

$$\mathbb{E}_{\check{x}}[\check{\tau}_u] = \check{h}_{\ell,u} = \frac{1}{\lambda_{u-1}^M} + \sum_{i=\ell}^{u-2} \frac{1}{\lambda_i^M} \left(1 + \sum_{j=i+1}^{u-1} \frac{\gamma_{i+1}^m \cdots \gamma_j^m}{\lambda_{i+1}^M \cdots \lambda_j^M} \right). \quad (4.5.4)$$

For the upper bound, suppose that λ_z^m is positive for $\ell \leq z \leq u-1$. We replace λ_i with λ_z^m and γ_i with γ_z^M , respectively, in formula (4.8.7), to obtain

$$\mathbb{E}_{\hat{x}}[\hat{\tau}_u] = \hat{h}_{\ell,u} = \frac{1}{\lambda_{u-1}^m} + \sum_{i=\ell}^{u-2} \frac{1}{\lambda_i^m} \left(1 + \sum_{j=i+1}^{u-1} \frac{\gamma_{i+1}^M \cdots \gamma_j^M}{\lambda_{i+1}^m \cdots \lambda_j^m} \right). \quad (4.5.5)$$

With a similar procedure, we can obtain lower and upper bounds for the MFPT for $\check{X}$ from $\mathcal{L}_u$ to $\mathcal{L}_\ell$. In particular, for $\check{\tau}_\ell = \inf\{t \geq 0 : \check{X}(t) \in \mathcal{L}_\ell\} = \inf\{t \geq 0 : L(\check{X}(t)) = \ell\}$ and $\check{x} \in \mathcal{L}_u$, we have

$$\hat{h}_{u,\ell} \leq \mathbb{E}_{\check{x}}[\check{\tau}_\ell] \leq \check{h}_{u,\ell}$$

in which

$$\hat{h}_{u,\ell} = \frac{1}{\gamma_{\ell+1}^M} + \sum_{i=\ell+2}^u \frac{1}{\gamma_i^M} \left(1 + \sum_{j=\ell+1}^{i-1} \frac{\lambda_j^m \cdots \lambda_{i-1}^m}{\gamma_j^M \cdots \gamma_{i-1}^M} \right), \quad (4.5.6)$$

$$\check{h}_{u,\ell} = \frac{1}{\gamma_{\ell+1}^m} + \sum_{i=\ell+2}^u \frac{1}{\gamma_i^m} \left(1 + \sum_{j=\ell+1}^{i-1} \frac{\lambda_j^M \cdots \lambda_{i-1}^M}{\gamma_j^m \cdots \gamma_{i-1}^m} \right), \quad (4.5.7)$$

provided that, for (4.5.6), γ_z^M is positive for $\ell + 1 \leq z \leq u$, and for (4.5.7), γ_z^m is positive for $\ell + 1 \leq z \leq u$.

Application to Example 4.4.1: Consider the SCRN introduced in (4.4.7), with associated graph $\mathcal{G}$ shown in Figure 4.1(b). We seek lower and upper bounds for the MFPT from $n_Z = N_{\text{tot}}$ to $n_Y = N_{\text{tot}}$. To this end, let us consider the coclique level function $L(x_1, x_2) = x_1 + 2x_2$ previously identified. The coclique level structure associated with it can be written as $\mathcal{L}_\ell, \dots, \mathcal{L}_u$, with $\mathcal{L}_z := \{x \in \check{\mathcal{X}} : L(x_1, x_2) = x_1 + 2x_2 = z\}$ for $z = \ell, \dots, u$, and $\ell = 0$, $u = 2N_{\text{tot}}$. This coclique level structure is such that $(0, 0)^T$ (i.e., $n_Z = N_{\text{tot}}$) is the only state belonging to $\mathcal{L}_\ell$ and $(0, N_{\text{tot}})^T$ (i.e., $n_Y = N_{\text{tot}}$) is the only state belonging to $\mathcal{L}_u$ (see Figure 4.1(c)). This feature is critical in order to determine good lower and upper bounds for the MFPT from $n_Z = N_{\text{tot}}$ to $n_Y = N_{\text{tot}}$.

Now,

$$G_+ = \{1, 2\} \quad \text{and} \quad G_- = \emptyset. \quad (4.5.8)$$

The rate of increase $\lambda_z(\check{x})$ and the rate of decrease $\gamma_z(\check{x})$ can be written as

$$\lambda_z(\check{x}) = f_1(\check{x}) + f_2(\check{x}) \quad \text{and} \quad \gamma_z(\check{x}) = 0, \quad (4.5.9)$$

with $f_1(\check{x})$, $f_2(\check{x})$ defined in (4.4.9). The two continuous time Markov chains, $\check{X}$ and $\hat{X}$, are defined on the same state space as $\check{X}$, with infinitesimal generators $\check{Q}$ and $\hat{Q}$, respectively, such that, for $z \in \{\ell, \ell + 1, \dots, u - 1, u\}$ and $\check{x} \in \mathcal{L}_z$, $\check{Q}_{\check{x}, \check{x} + \check{v}_k} = \frac{\lambda_z^M}{|G_+(\check{x})|}$ for $k \in G_+(\check{x})$, $\hat{Q}_{\check{x}, \check{x} + \check{v}_k} = \frac{\lambda_z^m}{|G_+(\check{x})|}$ for $k \in G_+(\check{x})$, with $\lambda_z^M = \max_{\check{x} \in \mathcal{L}_z} \lambda_z(\check{x})$ and $\lambda_z^m = \min_{\check{x} \in \mathcal{L}_z} \lambda_z(\check{x})$, as defined in (4.5.1), and $G_+(\check{x})$ defined as in (4.5.2) where G_+ is given in (4.5.8). Note that

in this example, $G_- = \emptyset$.

Then, as described in Section 4.5.3, we obtain analytical expressions for lower and upper bounds for the MFPT for $\check{X}$ from $n_Z = N_{\text{tot}}$ to $n_Y = N_{\text{tot}}$ as

$$\check{h}_{\ell,u} = \frac{1}{\lambda_\ell^M} + \frac{1}{\lambda_{\ell+1}^M} + \cdots + \frac{1}{\lambda_{u-2}^M} + \frac{1}{\lambda_{u-1}^M}, \quad \hat{h}_{\ell,u} = \frac{1}{\lambda_\ell^m} + \frac{1}{\lambda_{\ell+1}^m} + \cdots + \frac{1}{\lambda_{u-2}^m} + \frac{1}{\lambda_{u-1}^m},$$

respectively. Given that

$$\lambda_z^M = \max_{\check{x} \in \mathcal{L}_z} (\alpha(N_{\text{tot}} - (x_1 + x_2)) + \beta x_1)$$

and

$$\lambda_z^m = \min_{\check{x} \in \mathcal{L}_z} (\alpha(N_{\text{tot}} - (x_1 + x_2)) + \beta x_1),$$

where the functions being maximized and minimized are increasing in α , respectively β , we see that the lower and upper bounds $\check{h}_{\ell,u}$ and $\hat{h}_{\ell,u}$ decrease as either of the rate constants α or β increases.

For this example, due to the simplicity of the SCRN considered, it is also possible to derive a compact analytical expression for the MFPT. Each of the N_{tot} molecules acts independently and must undergo one $Z \rightarrow W$ transition and one $W \rightarrow Y$ transition to be converted from Z to Y . The time for one molecule of Z to be converted to Y is the sum of two independent exponential random variables with parameters α and β , respectively, which has a hypoexponential distribution with distribution function

$$F(x) = (F_1 * F_2)(x) = \int_0^x (1 - e^{-\alpha(x-y)})\beta e^{-\beta y} dy = \int_0^x (1 - e^{-\beta(x-y)})\alpha e^{-\alpha y} dy,$$

for $x > 0$ where F_1 and F_2 are distribution functions for exponential random variables with parameters α and β , respectively. The MFPT we desire is the mean of the maximum of N_{tot} such independent hypoexponential random variables, which is given by $\int_0^\infty (1 - F(x)^{N_{\text{tot}}}) dx$.

Observe that for fixed $\beta > 0$ and $0 < y < x$, $1 - e^{-\alpha(x-y)}$ is increasing with α , and then $F(x)$ is increasing with α . Thus, the MFPT decreases with increasing α . Similarly, for fixed $\alpha > 0$, we have that the MFPT decreases with increasing β .

Overall, this formulation explicitly gives the dependence of the MFPT on the rate constants α and β , and is fully consistent with the monotonic behavior that was observed through our study of its bounds.

4.5.4 Generalization to the non-weakly-connected case

In this section, we will consider SCRN satisfying Assumption 4.4.1 and whose associated graphs do not necessarily have to be weakly connected. In this case, the graph $\mathcal{G}$ associated with a SCRN can be decomposed into finitely many weakly connected components, and we use p to denote the number of weakly connected components in $\mathcal{G}$. For each $q = 1, \dots, p$, we use $\mathcal{G}^q$ to denote the q^{th} weakly connected component of $\mathcal{G}$ and use d_q to denote the number of vertices in $\mathcal{G}^q$.

Example 4.5.1. Consider a SCRN with four species (S_1, S_2, S_3, S_4) and the following three reactions under mass-action kinetics:



where the reaction rate constants $\kappa_1, \kappa_2, \kappa_3$ are positive. The associated graph $\mathcal{G}$ can be represented as follows:



In this case, $\mathcal{G}$ has two weakly connected components. The dynamics of $\{S_1, S_2\}$ evolve independently from those of $\{S_3, S_4\}$, and we can analyze the two subsystems independently. We will show, at the end of this section, how one can study the whole system by studying each subsystem independently.

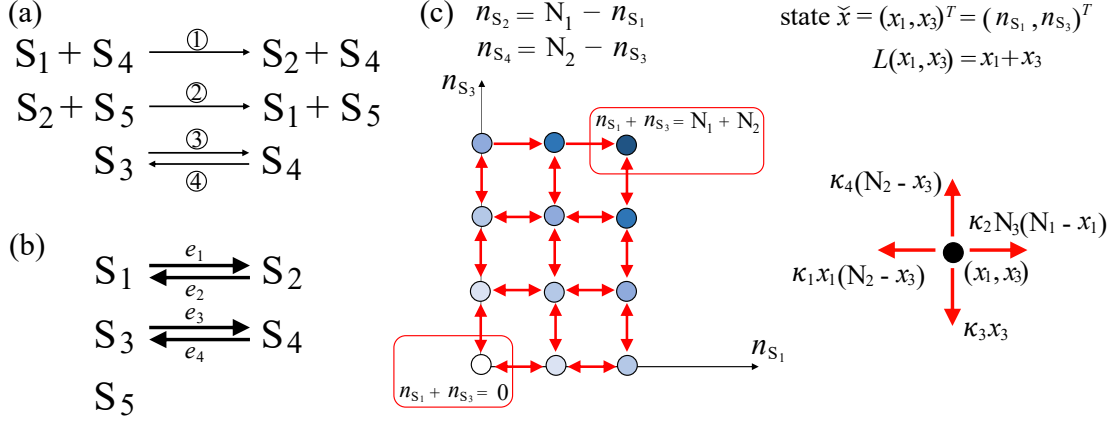
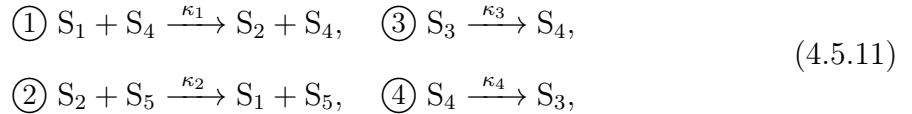


Figure 4.3. Example of a SCRN whose associated graph $\mathcal{G}$ is not weakly connected (**Example 4.5.2**). (a) Chemical reaction system diagram. The numbers on the arrows correspond to the reactions associated with the arrows as described in (4.5.11) in the main text. (b) Graph $\mathcal{G}$ associated with the chemical reaction system in panel (a). (c) State space and transitions of the projected continuous time Markov chain $\tilde{X} = \{(X_1(t), X_3(t))^T : t \geq 0\}$, which keeps track of (n_{S_1}, n_{S_3}) through time. Here, we consider $N_1 = 2, N_2 = 3$ and we use dots to represent the states, and red double-ended (single-ended) arrows to represent transitions in both directions (in a single direction). Additionally, we use shades of blue to distinguish the level to which each state belongs. The function L associated with the coclique level structure is $L(x_1, x_3) = x_1 + x_3$. On the right-hand side of the panel, we show the rates associated with the one-step transitions for the projected Markov chain $\tilde{X}$. Note that, given an initial condition, the quantity of species S_5 does not change over time and we denote this conserved quantity by N_3 .

Example 4.5.2. For a reaction edge e_k lying in some weakly connected component of $\mathcal{G}$, the infinitesimal rate Υ_k may depend on the quantities of some species where the graph vertices for those species may be in other weakly connected components of $\mathcal{G}$. For example, consider the following SCRN under mass-action kinetics:



where the reaction rate constants $\kappa_1, \kappa_2, \kappa_3, \kappa_4$ are positive. The diagram of this SCRN is shown in Figure 4.3(a) and the associated graph $\mathcal{G}$ is shown in Figure 4.3(b). As a convention, we keep species S_5 even though its quantity is conserved over time, and the vertex representing S_5 in $\mathcal{G}$ is in its own weakly connected component. Here, $\mathcal{G}$ has three weakly connected components. The dynamics of $\{S_1, S_2\}$ do depend on the varying

quantity of S_4 , where S_4 is in a different weakly connected component from that of $\{S_1, S_2\}$. In this case, we cannot analyze the first two subsystems independently.

Now, for a SCRN satisfying Assumption 4.4.1 and whose associated graph $\mathcal{G}$ has p weakly connected components, we shall and do relabel the species so that $j \in \{1, \dots, d\}$ belongs to the q^{th} weakly connected component of $\mathcal{G}$ if and only if $\sum_{q'=1}^{q-1} d_{q'} < j \leq \sum_{q'=1}^q d_{q'}$ where a sum over an empty set is considered to be 0. With this relabeling, for a state $x \in \mathcal{X}$, $1 \leq q \leq p$ and $1 \leq i \leq d_q$, we shall use x_i^q to denote the $\left(i + \sum_{q'=1}^{q-1} d_{q'}\right)^{th}$ entry of x . In other words, we have

$$x = (x^1, \dots, x^p)^T \in \mathbb{Z}_+^d \text{ where } x^q = (x_1^q, \dots, x_{d_q}^q)^T \in \mathbb{Z}_+^{d_q} \text{ for } q = 1, \dots, p. \quad (4.5.12)$$

For $q = 1, \dots, p$, each vertex in $\mathcal{G}^q$ corresponds to a species and each edge in $\mathcal{G}^q$ corresponds to a reaction, and in this sense, there is a stoichiometric matrix S^q associated with $\mathcal{G}^q$. Since there is no edge between different weakly connected components of $\mathcal{G}$, the stoichiometric matrix S associated with the SCRN has the form⁴

$$S = \begin{bmatrix} S^1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & S^p \end{bmatrix}. \quad (4.5.13)$$

By applying Lemma 4.4.1 to each weakly connected component of $\mathcal{G}$, we can show that the rank of the stoichiometric matrix S is $d - p$ (which also implies that the chemical reaction network associated with $\mathcal{G}$, in the manner described in Remark 4.4.1, has deficiency zero) and there are p linearly independent conservation vectors for this SCRN. As a result, we shall consider a **projected continuous time Markov chain** $\check{X} = \{\check{X}(t) : t \geq 0\}$ in which the state $\check{x} \in \check{\mathcal{X}} \subset \mathbb{Z}_+^{d-p}$ tracks, for each $q = 1, \dots, p$, the number of molecules of

⁴Here, S^q is a $d_q \times r_q$ matrix, where r_q is the number of edges in $\mathcal{G}^q$. A component $\mathcal{G}^q$ of $\mathcal{G}$ consisting of one vertex has no edges and the associated S^q is degenerate.

each species in $\mathcal{G}^q$, except for the last species in $\mathcal{G}^q$, that is,

$$\tilde{x} = (\tilde{x}^1, \dots, \tilde{x}^p)^T \text{ where } \tilde{x}^q = (x_1^q, \dots, x_{d_q-1}^q)^T \text{ for } q = 1, \dots, p. \quad (4.5.14)$$

Here, we slightly abuse notation, where $\tilde{x}^q$ is a zero-dimensional vector if $|\mathcal{G}^q| = 1$. Note that the choice to express $x_{d_q}^q$ as a function of $x_i^q : 1 \leq i \leq d_q - 1$ is without loss of generality, since the species can always be relabeled so that a species chosen to be expressed as a function of the others is the last one. The process $\tilde{X}$ is a continuous time Markov chain⁵ defined on the finite state space

$$\begin{aligned} \tilde{\mathcal{X}} := & \left\{ \tilde{x} = (x_i^q)_{1 \leq i \leq d_q-1, 1 \leq q \leq p} \in \mathbb{Z}_+^{d-p} : \right. \\ & \left. (x_i^q)_{1 \leq i \leq d_q, 1 \leq q \leq p} \in \mathcal{X} \text{ where } x_{d_q}^q = x_{\text{tot}}^q - \sum_{i=1}^{d_q-1} x_i^q \text{ for } q = 1, \dots, p \right\} \\ \subset & \left\{ \tilde{x} = (x_i^q)_{1 \leq i \leq d_q-1, 1 \leq q \leq p} \in \mathbb{Z}_+^{d-p} : x_1^q + \dots + x_{d_q-1}^q \leq x_{\text{tot}}^q \text{ for } q = 1, \dots, p \right\}, \end{aligned}$$

where $x_{\text{tot}}^q = \sum_{i=1}^{d_q} X_i^q(0)$ for each $q = 1, \dots, p$. We will assume that $|\tilde{\mathcal{X}}| > 1$, and the infinitesimal generator of $\tilde{X}$ will be denoted by $\tilde{Q}$.

A **coclique level function** for $\tilde{X}$ is a linear function $L : \mathbb{Z}^{d-p} \rightarrow \mathbb{Z}$ such that for each $k = 1, \dots, n$,

$$L(\tilde{v}_k) \in \{-1, +1\}, \quad (4.5.15)$$

where $\tilde{v}_k \in \mathbb{Z}^{d-p}$ is the vector obtained from v_k by removing the last element in each of the weakly connected components in $\mathcal{G}$, in a similar manner to how we obtain (4.5.14) from (4.5.12). If such an L exists, it can be written as

$$L(x) = b^T x \text{ for } x \in \mathbb{Z}^{d-p} \text{ and some } b \in \mathbb{Z}^{d-p}, \quad (4.5.16)$$

⁵With this reduction, coordinates corresponding to graph components consisting of a single vertex are removed, although the transition rates of the Markov chain $\tilde{X}$ may depend on the constant amount of any species associated with such an isolated vertex.

where, upon partitioning the set of edges of the associated graph $\mathcal{G}$ into two disjoint subsets $\mathcal{E}_+ = \{e_k : L(\check{v}_k) = 1\}$ and $\mathcal{E}_- = \{e_k : L(\check{v}_k) = -1\}$ (where one of these may be empty), the vector $b = (b_1, \dots, b_{d-p})^T$ solves the system of equations

$$\sum_{i=1}^{d-p} b_i \check{v}_k(i) = \begin{cases} +1 & \text{if } e_k \in \mathcal{E}_+ \\ -1 & \text{if } e_k \in \mathcal{E}_- \end{cases} \quad \text{for } k = 1, \dots, n. \quad (4.5.17)$$

For a coclique level function L , a **coclique level structure** for $\check{X}$ can be defined similar to that in the case where $\mathcal{G}$ is weakly connected.

Now, we are going to characterize all coclique level structures for $\check{X}$. In the following theorem, if the q^{th} weakly connected component $\mathcal{G}^q$ of $\mathcal{G}$ has more than one vertex, i.e., $|\mathcal{G}^q| > 1$, then we will consider a SCRN that consists of all the species and reactions associated with this weakly connected component. If $|\mathcal{G}^q| = 1$, then $\check{x}^q$ is a zero-dimensional vector and we do not need to consider this weakly connected component when constructing coclique level functions as in (4.5.18).

Theorem 4.5.4. *Consider a SCRN satisfying Assumption 4.4.1 and $|\check{X}| > 1$. A coclique level function for the projected continuous time Markov chain $\check{X}$ exists if and only if for every weakly connected component $\mathcal{G}^q$ of the associated graph $\mathcal{G}$ satisfying $|\mathcal{G}^q| > 1$ where $q \in \{1, \dots, p\}$, an associated coclique level function exists. The set of all coclique level functions for $\check{X}$ is the set of all functions of the form*

$$L(\check{x}) = L(\check{x}^1, \dots, \check{x}^p) = \sum_{\substack{q=1, \dots, p: \\ |\mathcal{G}^q| > 1}} L^q(\check{x}^q) \quad (4.5.18)$$

where $L^q : \mathbb{Z}^{d_q-1} \rightarrow \mathbb{Z}$ is a coclique level function associated with $\mathcal{G}^q$ for $q = 1, \dots, p$ and $|\mathcal{G}^q| > 1$. In particular, a coclique level function for $\check{X}$ exists if and only if $\mathcal{G}$ is bipartite.

The proof of Theorem 4.5.4 can be found in SI-Section 4.8.2. For this, we can break

down the problem to look at each weakly connected component separately, which is made possible by the block structure of the stoichiometric matrix shown in (4.5.13).

Suppose the projected continuous time Markov chain $\check{X}$ with finite state space $\check{\mathcal{X}} \subseteq \mathbb{Z}_+^{d-p}$ has a coclique level function $L : \mathbb{Z}^{d-p} \rightarrow \mathbb{Z}$, defined as in (4.5.15), with coclique level structure $\mathcal{L}_\ell, \mathcal{L}_{\ell+1}, \dots, \mathcal{L}_{u-1}, \mathcal{L}_u$ defined as in (4.4.4)–(4.4.5) with $\ell < u$. Similar to the reasoning in Remark 4.5.1, for a coclique level function L as in (4.5.18), the sets $\mathcal{L}_\ell, \dots, \mathcal{L}_u$ in the coclique level structures associated with L and $-L$ are the same, while the orderings of those sets in the two coclique level structure partitions are opposite. As a convention, we consider the coclique level structures associated with L and $-L$ to be the same. Similar to the discussion in Section 4.5.3, we can determine analytical expressions for upper and lower bounds for both the MFPT for $\check{X}$ from $\mathcal{L}_\ell$ to $\mathcal{L}_u$ and the MFPT for $\check{X}$ from $\mathcal{L}_u$ to $\mathcal{L}_\ell$.

Application to Example 4.5.1: In this example, $\mathcal{G}$ has two weakly connected components, and $m_1 = (1, 1, 0, 0)^T$ and $m_2 = (0, 0, 1, 1)^T$ are two linearly independent conservation vectors for this system. Therefore, we can consider a projected continuous time Markov chain $\check{X} = \{(X_1(t), X_3(t))^T : t \geq 0\}$, which tracks the number of S_1 and S_3 through time. The state space for $\check{X}$ is $\{(x_1, x_3)^T \in \mathbb{Z}_+^2 : 0 \leq x_1 \leq N_1, 0 \leq x_3 \leq N_2\}$, where $N_1 = X_1(0) + X_2(0)$ and $N_2 = X_3(0) + X_4(0)$. We shall use Theorem 4.5.4 to identify all of the coclique level structures for $\check{X}$. For the first weakly connected component $\mathcal{G}^1$ of $\mathcal{G}$, the only possible partitions of edges of $\mathcal{G}^1$ are $\mathcal{E}_+^1 = \{e_2\}, \mathcal{E}_-^1 = \{e_1\}$ and $\mathcal{E}_+^1 = \{e_1\}, \mathcal{E}_-^1 = \{e_2\}$, by Remark 4.5.2 and the fact that reactions ① and ② are a pair of reversible reactions. One then can verify that $L^1(\check{x}^1) = x_1^1 = x_1$ and $L^1(\check{x}^1) = -x_1^1 = -x_1$ are the two coclique level functions associated with $\mathcal{G}^1$. Similarly, $L^2(\check{x}^2) = x_1^2 = x_3$ and $L^2(\check{x}^2) = -x_1^2 = -x_3$ are the only coclique level functions associated with $\mathcal{G}^2$. By Theorem 4.5.4, the coclique level functions for $\check{X}$ are $L(\check{x}) = x_1 + x_3$, $L(\check{x}) = x_1 - x_3$, $L(\check{x}) = -x_1 + x_3$ and $L(\check{x}) = -x_1 - x_3$, where the first two give the two distinct coclique level structures for $\check{X}$ by similar reasoning to that in Remark 4.5.1.

Because the dynamics of $\{S_1, S_2\}$ evolve independently from those of $\{S_3, S_4\}$, we can choose to analyze the two subsystems independently. For example, $\check{X}^1 = \{X_1(t) : t \geq 0\}$ is itself a continuous time Markov chain that tracks the number of S_1 and we have shown that $L^1(\check{x}^1) = x_1^1 = x_1$ is a coclique level function associated with $\mathcal{G}^1$. Thus, we may study the MFPT from $\check{X}^1 = 0$ to $\check{X}^1 = N_1$. However, it is not possible to study this MFPT using a coclique function for the whole system, because $L(\check{x}) = x_1$ is not a coclique level function for $\check{X}$. In conclusion, when the dynamics of each subsystem in the SCRNs evolve independently, it is beneficial to apply our theory in Sections 4.5.1–4.5.3 to each subsystem separately.

Application to Example 4.5.2: In this example, $\mathcal{G}$ has three weakly connected components, and $m_1 = (1, 1, 0, 0, 0)^T$, $m_2 = (0, 0, 1, 1, 0)^T$ and $m_3 = (0, 0, 0, 0, 1)^T$ are three linearly independent conservation vectors for this system. Therefore, we can consider a projected continuous time Markov chain $\check{X} = \{(X_1(t), X_3(t))^T : t \geq 0\}$, which tracks the number of S_1 and S_3 through time. The state space for $\check{X}$ is $\{(x_1, x_3)^T \in \mathbb{Z}_+^2 : 0 \leq x_1 \leq N_1, 0 \leq x_3 \leq N_2\}$, where $N_1 = X_1(0) + X_2(0)$ and $N_2 = X_3(0) + X_4(0)$. Note that, given an initial condition, the quantity of species S_5 does not change over time and we shall denote this conserved quantity by N_3 . Accordingly, the trivial dynamics of the system associated with the third weakly connected component of $\mathcal{G}$ is omitted in the projected chain $\check{X}$. Similar to the previous example, we can use Theorem 4.5.4 to identify all of the coclique level structures for $\check{X}$, which are the coclique level structures associated with the coclique level functions $L(\check{x}) = x_1 + x_3$ and $L(\check{x}) = x_1 - x_3$.

As an example, we can use the coclique level structure associated with the coclique level function $L(\check{x}) = x_1 + x_3$ to bound the MFPT from $\check{x} = (x_1, x_3)^T = (N_1, N_2)^T$ to $(0, 0)^T$ above and below by (4.5.6) and (4.5.7), respectively, where the parameters in these

two expressions are given by $\ell = 0$, $u = N_1 + N_2$ and for $\ell \leq z \leq u$,

$$\begin{aligned}\lambda_z^m &= \min_{\substack{0 \leq x_1 \leq N_1, \\ 0 \leq x_3 \leq N_2, \\ x_1 + x_3 = z}} (\kappa_2(N_1 - x_1)N_3 + \kappa_4(N_2 - x_3)), \\ \lambda_z^M &= \max_{\substack{0 \leq x_1 \leq N_1, \\ 0 \leq x_3 \leq N_2, \\ x_1 + x_3 = z}} (\kappa_2(N_1 - x_1)N_3 + \kappa_4(N_2 - x_3)), \\ \gamma_z^m &= \min_{\substack{0 \leq x_1 \leq N_1, \\ 0 \leq x_3 \leq N_2, \\ x_1 + x_3 = z}} (\kappa_1 x_1(N_2 - x_3) + \kappa_3 x_3), \\ \gamma_z^M &= \max_{\substack{0 \leq x_1 \leq N_1, \\ 0 \leq x_3 \leq N_2, \\ x_1 + x_3 = z}} (\kappa_1 x_1(N_2 - x_3) + \kappa_3 x_3),\end{aligned}$$

in which one can verify that $\gamma_z^M \geq \gamma_z^m > 0$ for each $1 \leq z \leq N_1 + N_2$.

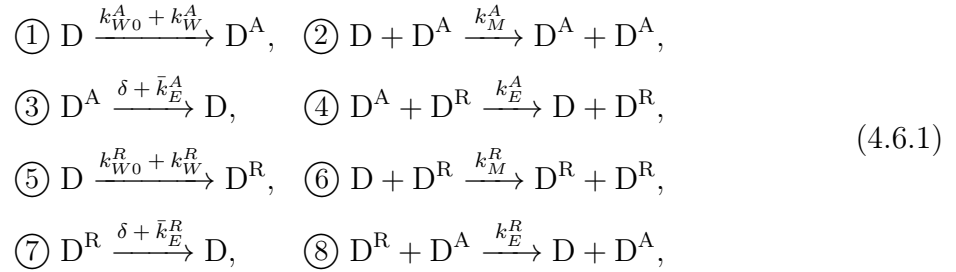
4.6 Examples

In this section, we consider three examples and study their stochastic behavior, in terms of MFPT, by exploiting the theoretical tools developed in this work. The examples come from biological areas of epigenetics, neurobiology and ecology (see Bruno et al. [5], Milo et al. [19] and Brglez et al. [3]). All the continuous time Markov chains associated with the SCRNs considered in our examples have a finite state space.

4.6.1 Chromatin modification circuit including only histone modifications

Epigenetic regulation is the modification of the DNA structure, due to chromatin modifications, that determines if a gene is active or repressed. Various chromatin modifications affect the structure of DNA. In this example, we will consider only histone modifications, while in the next one, we will study a more complex model including also DNA methylation. More precisely, in this example we analyze a well-established model for a histone modification circuit (see Dodd et al. [9] and Bruno et al. [5, 7]), which involves three species: unmodified nucleosome, denoted by D; nucleosome modified with

repressive histone modifications, denoted by D^R ; nucleosome modified with activating histone modifications, denoted by D^A . In this model, each histone modification catalyzes its own establishment on unmodified nucleosomes and catalyzes the erasure of the opposite modification (see Dodd et al. [9] and Bruno et al. [5]). The quantity of each species is denoted by n_D , n_{D^R} and n_{D^A} , respectively. Their sum remains constant, i.e., $n_D + n_{D^R} + n_{D^A} = D_{\text{tot}}$, where D_{tot} denotes the total number of nucleosomes within the gene. Then, the chemical reaction system, whose diagram is shown in Figure 4.4(a), can be written as



in which $k_{W0}^A, k_W^A, k_M^A, \delta, \bar{k}_E^A, k_E^A, k_{W0}^R, k_W^R, k_M^R, \bar{k}_E^R, k_E^R > 0$. The expression of the reaction rate constants is because we combined reactions sharing the same reactants and products. Now, denoting the reaction volume by V , let us introduce $\varepsilon := \frac{\delta + \bar{k}_E^A}{k_M^A(D_{\text{tot}}/V)} = \frac{\delta_A}{k_M^A(D_{\text{tot}}/V)}$, with $\delta_A := \delta + \bar{k}_E^A$, and $\mu := \frac{k_E^R}{k_E^A}$. Furthermore, let us consider the constant $\tilde{b}$ such that $\mu\tilde{b} = \frac{\delta_R}{\delta_A}$, where $\delta_R := \delta + \bar{k}_E^R$. Then, $\delta_R = \delta_A\mu\tilde{b} = \varepsilon \frac{k_M^A D_{\text{tot}}}{V} \mu\tilde{b}$.

Considering $x = (n_{D^R}, n_{D^A}, n_D)$, the reaction vectors associated with (4.6.1) are $v_1 = (0, 1, -1)^T$, $v_2 = (0, -1, 1)^T$, $v_3 = (1, 0, -1)^T$, and $v_4 = (-1, 0, 1)^T$. By examining them, it is possible to verify that Assumption 4.4.1 is satisfied. The graph $\mathcal{G}$ associated with the chemical reaction system (4.6.1) can then be represented as in Figure 4.4(b). By inspecting $\mathcal{G}$, one can verify that the underlying undirected graph is connected and that $\mathcal{G}$ is bipartite. By Lemma 4.4.1, our SCRN has a unique conservation vector $m = (1, 1, 1)^T$ and then we can introduce a projected continuous time Markov chain $\check{X} = \{(X_1(t), X_2(t))^T : t \geq 0\}$, which keeps track of (n_{D^R}, n_{D^A}) through time. Since the total number of nucleosomes D_{tot} is conserved, the state space is $\check{\mathcal{X}} = \{\check{x} = (x_1, x_2)^T \in \mathbb{Z}_+^2 : x_1 + x_2 \leq D_{\text{tot}}\}$. The potential

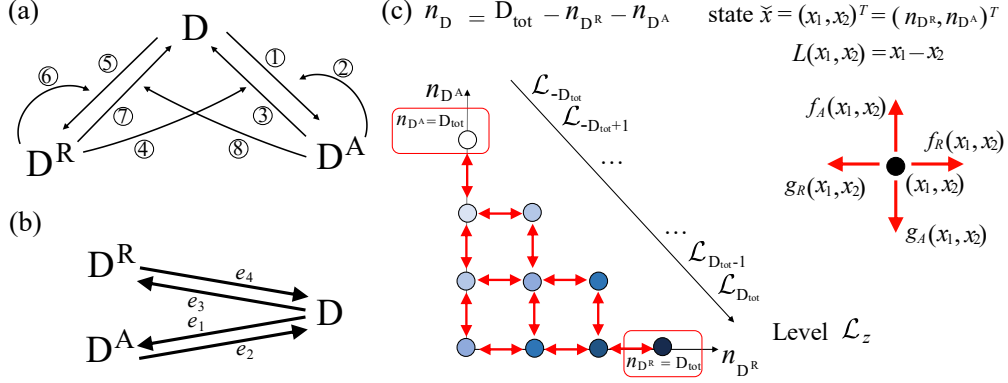


Figure 4.4. Chromatin modification circuit including only histone modifications: reaction diagram, graph $\mathcal{G}$ and associated Markov chain. (a) Chemical reaction system diagram. The numbers on the arrows correspond to the reactions associated with the arrows as described in (4.6.1) in the main text. (b) Graph $\mathcal{G}$ associated with the chemical reaction system in panel (a). (c) State space and transitions of the projected continuous time Markov chain $\tilde{X} = \{(X_1(t), X_2(t))^T : t \geq 0\}$, which keeps track of (n_{D^R}, n_{D^A}) through time. Here, we consider $D_{\text{tot}} = 3$ and we use dots to represent the states, and red double-ended arrows to represent transitions in both directions. Additionally, we use shades of blue to distinguish the level to which each state belongs. The function L associated with the coclique level structure is $L(x_1, x_2) = x_1 - x_2$. The rates associated with the one-step transitions for the projected Markov chain $\tilde{X}$ are given in (4.6.2).

one-step transitions for $\tilde{X}$ from $x \in \tilde{\mathcal{X}}$ are shown in Figure 4.4(c), where the associated transition vectors are $\tilde{v}_1 = -\tilde{v}_2 = (0, 1)^T$ and $\tilde{v}_3 = -\tilde{v}_4 = (1, 0)^T$, and the infinitesimal transition rates (in which we assume mass-action kinetics with the usual rate constant volume scaling) are

$$\begin{aligned}
 \check{Q}_{\tilde{x}, \tilde{x} + \tilde{v}_1} &= f_A(\tilde{x}) = (D_{\text{tot}} - (x_1 + x_2)) \left(k_{W0}^A + k_W^A + \frac{k_M^A}{V} x_2 \right), \\
 \check{Q}_{\tilde{x}, \tilde{x} + \tilde{v}_2} &= g_A(\tilde{x}) = x_2 \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} + x_1 \frac{k_E^A}{V} \right), \\
 \check{Q}_{\tilde{x}, \tilde{x} + \tilde{v}_3} &= f_R(\tilde{x}) = (D_{\text{tot}} - (x_1 + x_2)) \left(k_{W0}^R + k_W^R + \frac{k_M^R}{V} x_1 \right), \\
 \check{Q}_{\tilde{x}, \tilde{x} + \tilde{v}_4} &= g_R(\tilde{x}) = x_1 \mu \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} \tilde{b} + x_2 \frac{k_E^A}{V} \right).
 \end{aligned} \tag{4.6.2}$$

Let us now focus on determining explicit analytical expressions for upper and lower bounds of MFPTs. To this end, let us apply Theorem 4.5.1 to determine the coclique level

structures for $\check{X}$. We can apply the theorem because Assumption 4.4.1 is satisfied and the associated graph $\mathcal{G}$ is weakly connected.

Consider all of the possible partitions $\{\mathcal{E}_+, \mathcal{E}_-\}$ of edges of $\mathcal{G}$ that could allow us to determine a coclique level structure. These partitions are the following:

$$\mathcal{E}_+ = \{e_1, e_3\}, \mathcal{E}_- = \{e_2, e_4\} \quad \text{and} \quad \mathcal{E}_+ = \{e_2, e_3\}, \mathcal{E}_- = \{e_1, e_4\}.$$

We did not consider the partition $\mathcal{E}_+ = \{e_2, e_4\}, \mathcal{E}_- = \{e_1, e_3\}$ or the partition $\mathcal{E}_+ = \{e_1, e_4\}, \mathcal{E}_- = \{e_2, e_3\}$ because, as explained in Remark 4.5.1, the associated functions L would be the opposite of the ones obtained for the partitions considered above and the resulting coclique level structures are considered to be the same. Furthermore, we did not consider the partitions in which two edges in the same element of a partition are associated with two reaction vectors $\check{v}_k, \check{v}_{k'} \in \mathcal{V}$ such that $\check{v}_k = -\check{v}_{k'}$, because, as stated in Corollary 4.5.1, these partitions would not lead to a coclique level structure. For each partition, the system of equations in (4.4.3) has a unique solution, these being $(b_1, b_2)^T = (1, 1)^T$ and $(b_1, b_2)^T = (1, -1)^T$, respectively. Then, by applying Theorem 4.5.1, we can conclude that the projected Markov chain $\check{X}$ has two coclique level structures associated with coclique level functions $L(x_1, x_2) = x_1 + x_2$ and $L(x_1, x_2) = x_1 - x_2$.

Let us consider the coclique level function $L(x_1, x_2) = x_1 - x_2$. The coclique level structure associated with it can be written as $\mathcal{L}_\ell, \dots, \mathcal{L}_u$, with $\mathcal{L}_z := \{x \in \check{X} : L(x_1, x_2) = x_1 - x_2 = z\}$ for $z = \ell, \dots, u$, with $\ell = -D_{\text{tot}}$ and $u = D_{\text{tot}}$. This coclique level structure is such that $a = (0, D_{\text{tot}})^T$ is the only state belonging to $\mathcal{L}_\ell$ and $r = (D_{\text{tot}}, 0)^T$ is the only state belonging to $\mathcal{L}_u$ (see Figure 4.4(c)). As shown below, this feature of the coclique level structure is critical in order to determine lower and upper bounds for $h_{a,r}$, the MFPT from a to r , and $h_{r,a}$, the MFPT from r to a , and this is the reason why we consider the coclique level structure associated with the function $L(x_1, x_2) = x_1 - x_2$ and not the one associated with the function $L(x_1, x_2) = x_1 + x_2$.

Let us now determine the lower and upper bounds for the MFPT from a to r , and vice versa, following the approach described in the previous section. In particular, here we have

$$G_+ = \{2, 3\} \quad \text{and} \quad G_- = \{1, 4\}, \quad (4.6.3)$$

and the rate of increase $\lambda_z(\tilde{x})$ and the rate of decrease $\gamma_z(\tilde{x})$ can be written as

$$\lambda_z(\tilde{x}) = f_R(\tilde{x}) + g_A(\tilde{x}) \quad \text{and} \quad \gamma_z(\tilde{x}) = f_A(\tilde{x}) + g_R(\tilde{x}), \quad (4.6.4)$$

with $f_R(\tilde{x})$, $g_A(\tilde{x})$, $f_A(\tilde{x})$, $g_R(\tilde{x})$ defined in (4.6.2). The two continuous time Markov chains $\check{X}$ and $\hat{X}$, defined on the same state space as $\tilde{X}$ have infinitesimal generators $\check{Q}$ and $\hat{Q}$, respectively, such that, for $z \in \{\ell, \ell + 1, \dots, u - 1, u\}$ and $\tilde{x} \in \mathcal{L}_z$, $\check{Q}_{\tilde{x}, \tilde{x} + \tilde{v}_k} = \frac{\lambda_z^M}{|G_+(\tilde{x})|}$ for $k \in G_+(\tilde{x})$, $\check{Q}_{\tilde{x}, \tilde{x} + \tilde{v}_k} = \frac{\gamma_z^m}{|G_-(\tilde{x})|}$ for $k \in G_-(\tilde{x})$, $\hat{Q}_{\tilde{x}, \tilde{x} + \tilde{v}_k} = \frac{\lambda_z^m}{|G_+(\tilde{x})|}$ for $k \in G_+(\tilde{x})$, and $\hat{Q}_{\tilde{x}, \tilde{x} + \tilde{v}_k} = \frac{\gamma_z^M}{|G_-(\tilde{x})|}$ for $k \in G_-(\tilde{x})$, with λ_z^M , λ_z^m , γ_z^M , and γ_z^m defined as in (4.5.1), and non-empty $G_+(\tilde{x})$ and $G_-(\tilde{x})$ defined as in (4.5.2) where G_+ and G_- are given in (4.6.3).

Then, as described in Section 4.5.3, we can compare the Markov chain $\check{X}$ with $\tilde{X}$ and $\hat{X}$, separately, to conclude that

$$\check{h}_{\ell,u} \leq h_{a,r} \leq \hat{h}_{\ell,u} \quad \text{and} \quad \hat{h}_{u,\ell} \leq h_{r,a} \leq \check{h}_{u,\ell} \quad (4.6.5)$$

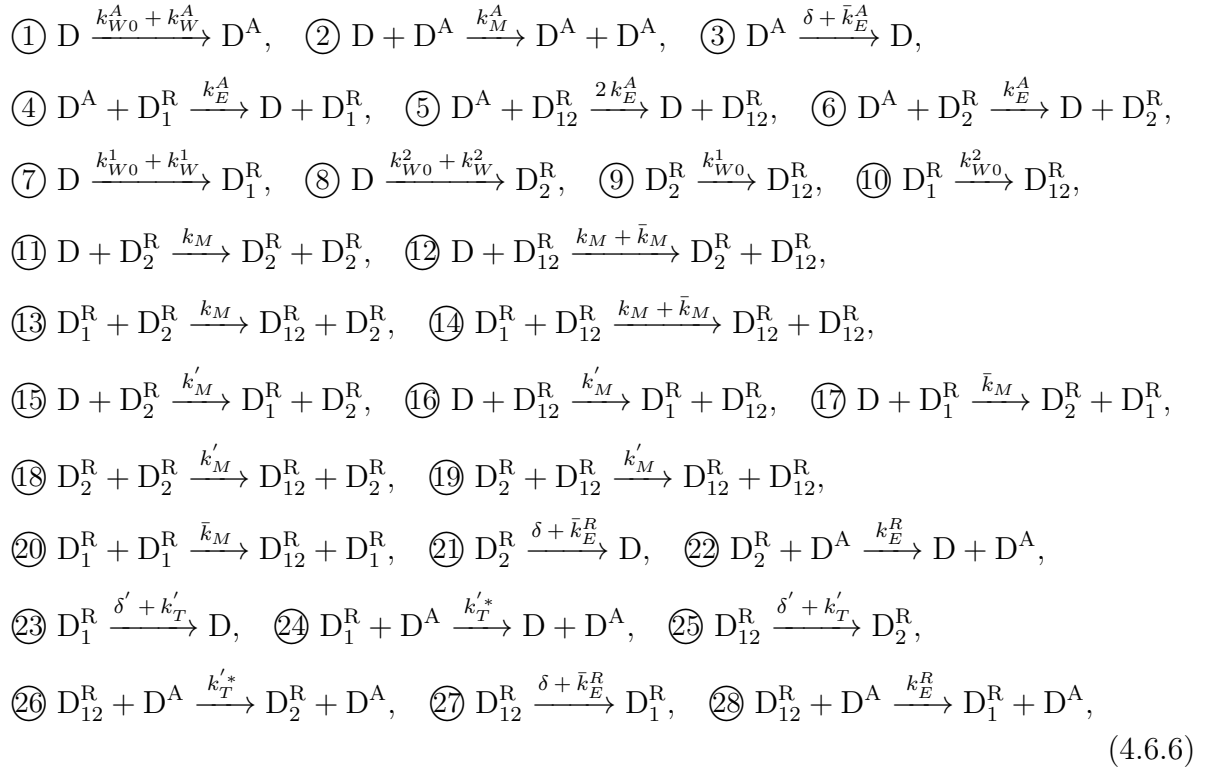
where $\check{h}_{\ell,u}$, $\hat{h}_{\ell,u}$, $\hat{h}_{u,\ell}$, and $\check{h}_{u,\ell}$ can be written as in (4.5.4), (4.5.5), (4.5.6), and (4.5.7), respectively, with the quantities in (4.5.1) replaced by those in (4.6.4).

These analytical expressions allow us to study the effect of parameter ε on the MFPT from a to r (average time to memory loss of the active state) and the MFPT from r to a (average time to memory loss of the repressed state). Specifically, given that the only $O(\varepsilon)$ rates are $\gamma_u^M = \gamma_u^m = g_R(D_{\text{tot}}, 0)$ and $\lambda_\ell^M = \lambda_\ell^m = g_A(0, D_{\text{tot}})$, we can conclude that the upper bounds $\hat{h}_{\ell,u}$, $\check{h}_{u,\ell}$ and the lower bounds $\check{h}_{\ell,u}$, $\hat{h}_{u,\ell}$ in (4.6.5) are all $O(\varepsilon^{-1})$. This implies that the average time to memory loss of both the repressed and active states

are $O(\varepsilon^{-1})$, and as ε approaches 0, the average time to memory loss tend to infinity.

4.6.2 Full chromatin modification circuit

In this example, we consider a chromatin modification circuit model that includes not only histone modifications, but also DNA methylation (see Bruno et al. [5, 6]). The model involves five species: unmodified nucleosome, denoted by D ; nucleosome with CpGme only, denoted by D_1^R ; nucleosome with H3K9me3 only, denoted by D_2^R ; nucleosome with both H3K9me3 and CpGme, denoted by D_{12}^R ; nucleosome with an activating histone modification, denoted by D^A . In terms of molecular interactions, DNA methylation catalyzes the establishment of repressive histone modifications (and vice versa), while enhancing the erasure of activating marks (and vice versa) (see Bruno et al. [5]). The quantity of each species is denoted by n_D , n_{D^A} , $n_{D_1^R}$, $n_{D_2^R}$, and $n_{D_{12}^R}$, respectively. Their sum remains constant, i.e., $n_D + n_{D^A} + n_{D_1^R} + n_{D_2^R} + n_{D_{12}^R} = D_{\text{tot}}$. The chemical reaction system, whose diagram is shown in Figure 4.5(a), can be written as



in which $k_{W0}^A, k_W^A, k_M^A, \delta, \bar{k}_E^A, k_E^A, k_{W0}^1, k_W^1, k_{W0}^2, k_W^2, k_M', \bar{k}_M, k_M, \delta', k_T', k_T'^*, \bar{k}_E^R, k_E^R > 0$ and the expression of the reaction rate constants is because we combined reactions sharing the same reactants and products. As we did for Example 4.6.1, let us introduce parameters $\varepsilon = \frac{\delta + \bar{k}_E^A}{k_M^A D_{\text{tot}}}$ and $\mu = \frac{k_E^R}{k_E^A}$, with a constant $\tilde{b}$ such that $\frac{\delta + \bar{k}_E^R}{\delta + k_E^A} = \tilde{b}\mu$. Furthermore, since in this model we also have DNA methylation, we also introduce $\mu' = \frac{k_T'^*}{k_E^A}$ and a constant β such that $\frac{\delta' + k_T'}{\delta + k_E^A} = \beta\mu'$. The parameter μ' quantifies the relative speed between the erasure rate of DNA methylation and the erasure rate of activating histone modifications.

For $x = (n_{D_{12}^R}, n_{D^A}, n_{D_1^R}, n_{D_2^R}, n_D)^T$, the reaction vectors associated with (4.6.6) are $v_1 = (0, 1, 0, 0, -1)^T$, $v_2 = (0, -1, 0, 0, 1)^T$, $v_3 = (0, 0, 1, 0, -1)^T$, $v_4 = (0, 0, -1, 0, 1)^T$, $v_5 = (0, 0, 0, 1, -1)^T$, $v_6 = (0, 0, 0, -1, 1)^T$, $v_7 = (1, 0, -1, 0, 0)^T$, $v_8 = (-1, 0, 1, 0, 0)^T$, $v_9 = (1, 0, 0, -1, 0)^T$, and $v_{10} = (-1, 0, 0, 1, 0)^T$. By examining them, one can verify that Assumption 4.4.1 is satisfied. The graph $\mathcal{G}$ associated with the chemical reaction system (4.6.6) can then be represented as in Figure 4.5(b). As done for Example 4.6.1, by inspecting $\mathcal{G}$, one can verify that the underlying undirected graph is connected and that $\mathcal{G}$ is bipartite. By Lemma 4.4.1, our SCRN has a unique conservation vector $m = (1, \dots, 1)^T$ and we can introduce the projected continuous time Markov chain $\check{X} = \{(X_1(t), X_2(t), X_3(t), X_4(t))^T : t \geq 0\}$, which keeps track of $(n_{D_{12}^R}, n_{D^A}, n_{D_1^R}, n_{D_2^R})$ through time. Since the total number of nucleosomes D_{tot} is conserved, the state space is $\check{\mathcal{X}} = \{\check{x} = (x_1, x_2, x_3, x_4)^T \in \mathbb{Z}_+^4 : x_1 + x_2 + x_3 + x_4 \leq D_{\text{tot}}\}$. The potential one-step transitions for $\check{X}$ from $x \in \check{\mathcal{X}}$ are shown in Figure 4.5(c), where the associated transition vectors are $\check{v}_1 = -\check{v}_2 = (0, 1, 0, 0)^T$, $\check{v}_3 = -\check{v}_4 = (0, 0, 1, 0)^T$, $\check{v}_5 = -\check{v}_6 = (0, 0, 0, 1)^T$, $\check{v}_7 = -\check{v}_8 = (1, 0, -1, 0)^T$, and $\check{v}_9 = -\check{v}_{10} = (1, 0, 0, -1)^T$, and the infinitesimal transition rates (in which we assume mass-action kinetics with the usual rate constant volume scaling)

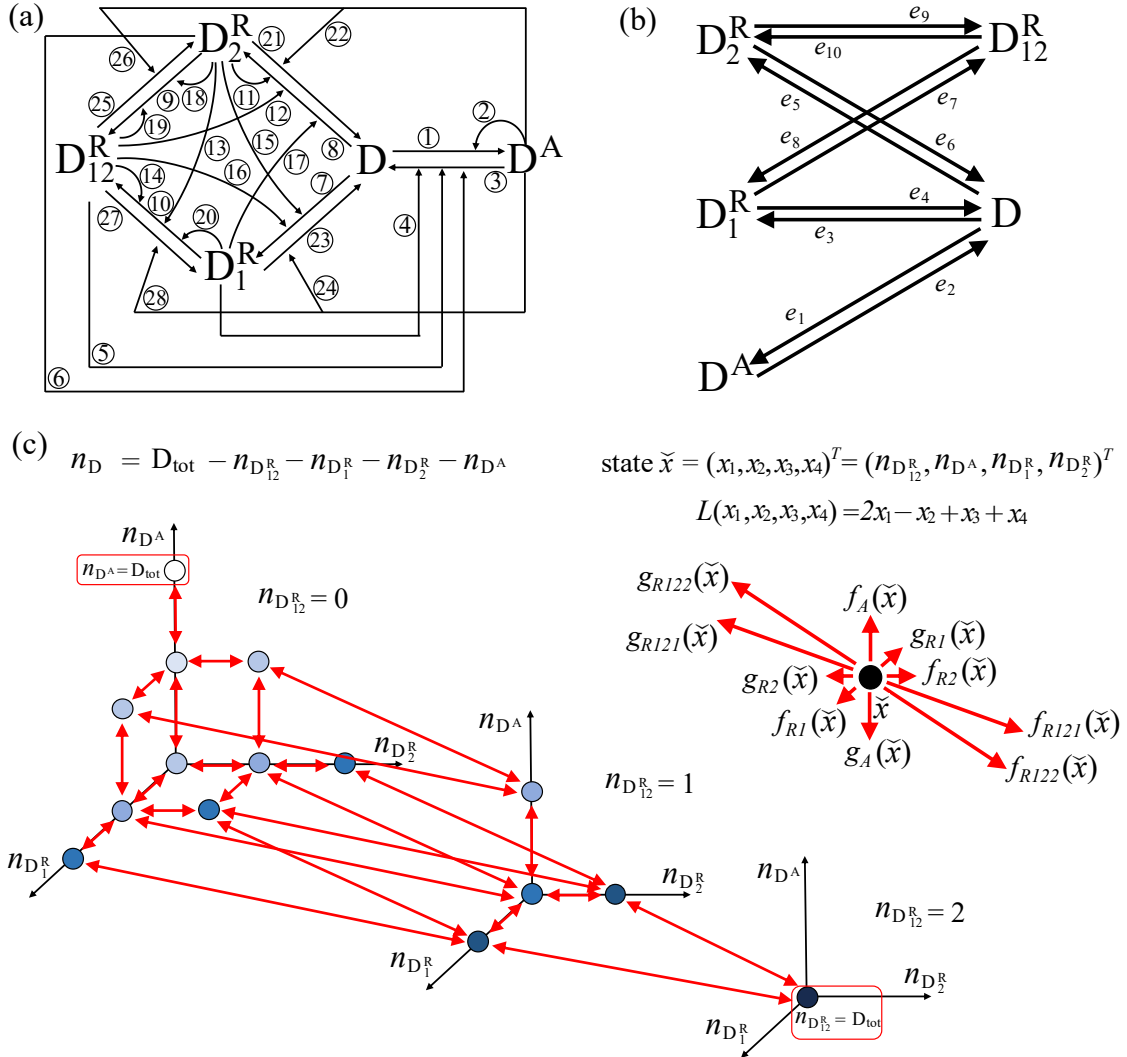


Figure 4.5. Full chromatin modification circuit: reaction diagram, graph $\mathcal{G}$ and associated Markov chain. (a) Chemical reaction system diagram. The numbers on the arrows correspond to the reactions associated with the arrows as described in (4.6.6) in the main text. (b) Graph $\mathcal{G}$ associated with the chemical reaction system in panel (a). (c) State space and transitions of the projected continuous time Markov chain $\tilde{X} = \{(X_1(t), X_2(t), X_3(t), X_4(t))^T : t \geq 0\}$, which keeps track of $(n_{D_{12}^R}, n_{D^A}, n_{D_1^R}, n_{D_2^R})$ through time. Here, we consider $D_{\text{tot}} = 2$ and we use dots to represent the states, and red double-headed arrows to represent transitions in both directions. Additionally, we use shades of blue to distinguish the level to which each state belongs. The function L associated with the coclique level structure is $L(x_1, x_2, x_3, x_4) = 2x_1 - x_2 + x_3 + x_4$. The rates associated with the one-step transitions for the projected Markov chain $\tilde{X}$ are given in (4.6.7).

are

$$\begin{aligned}
Q_{x,x+v_1}(\varepsilon) &= f_A(x) = (D_{\text{tot}} - (x_1 + x_2 + x_3 + x_4)) \left(k_{W0}^A + k_W^A + \frac{k_M^A}{V} x_2 \right), \\
Q_{x,x+v_2}(\varepsilon) &= g_A^\varepsilon(x) = x_2 \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} + \frac{k_E^A}{V} (x_3 + x_4 + 2x_1) \right), \\
Q_{x,x+v_3}(\varepsilon) &= f_{R1}(x) = (D_{\text{tot}} - (x_1 + x_2 + x_3 + x_4)) \left(k_{W0}^1 + k_W^1 + \frac{k_M'}{V} (x_1 + x_4) \right), \\
Q_{x,x+v_4}(\varepsilon) &= g_{R1}^\varepsilon(x) = x_3 \mu' \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} \beta + x_2 \frac{k_E^A}{V} \right), \\
Q_{x,x+v_5}(\varepsilon) &= f_{R2}(x) = (D_{\text{tot}} - (x_1 + x_2 + x_3 + x_4)) \\
&\quad \left(k_{W0}^2 + k_W^2 + \frac{k_M}{V} (x_1 + x_4) + \frac{\bar{k}_M}{V} (x_1 + x_3) \right), \\
Q_{x,x+v_6}(\varepsilon) &= g_{R2}^\varepsilon(x) = x_4 \mu \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} b + x_2 \frac{k_E^A}{V} \right), \\
Q_{x,x+v_7}(\varepsilon) &= f_{R121}(x) = x_3 \left(k_{W0}^2 + \frac{k_M}{V} (x_1 + x_4) + \frac{\bar{k}_M}{V} \left(x_1 + \frac{x_3 - 1}{2} \right) \right), \\
Q_{x,x+v_8}(\varepsilon) &= g_{R121}^\varepsilon(x) = x_1 \mu \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} b + x_2 \frac{k_E^A}{V} \right), \\
Q_{x,x+v_9}(\varepsilon) &= f_{R122}(x) = x_4 \left(k_{W0}^1 + \frac{k_M'}{V} \left(x_1 + \frac{x_4 - 1}{2} \right) \right), \\
Q_{x,x+v_{10}}(\varepsilon) &= g_{R122}^\varepsilon(x) = x_1 \mu' \left(\varepsilon \frac{k_M^A}{V} D_{\text{tot}} \beta + x_2 \frac{k_E^A}{V} \right).
\end{aligned} \tag{4.6.7}$$

A representation of the Markov chain graph for $D_{\text{tot}} = 2$ is given in Figure 4.5(c).

We now focus on determining explicit analytical expressions for upper and lower bounds of the MFPT from the active state $a = (0, D_{\text{tot}}, 0, 0)^T$ to the repressed state $r = (D_{\text{tot}}, 0, 0, 0)^T$, i.e., $h_{a,r}$, and vice versa, i.e., $h_{r,a}$, in order to understand how the parameters ε and μ' affect them.

To calculate the MFPT upper and lower bounds, let us first apply Theorem 4.5.1 in order to determine the coclique level structures for $\check{X}$. We can apply the theorem because Assumption 4.4.1 is satisfied and the associated graph $\mathcal{G}$ is weakly connected. The rank of the stoichiometric matrix S is $4 = d - 1$. Now, for this example, there are several possible

partitions $\{\mathcal{E}_+, \mathcal{E}_-\}$ of edges of $\mathcal{G}$ that could allow us to determine a coclique level structure. Let us consider the following one: $\mathcal{E}_+ = \{e_2, e_3, e_5, e_7, e_9\}$, $\mathcal{E}_- = \{e_1, e_4, e_6, e_8, e_{10}\}$. The reason for this choice is that it is the only one that, as we will see later, allows us to determine a coclique level structure in which the active state $a = (0, D_{\text{tot}}, 0, 0)^T$ and the repressed state $r = (D_{\text{tot}}, 0, 0, 0)^T$ are the two extremum levels. For this partition, the system of equations in (4.4.3) admits a unique solution $(b_1, b_2, b_3, b_4)^T = (2, -1, 1, 1)^T$.

Then, by applying Theorem 4.5.1, we can conclude that the projected Markov chain $\check{X}$ has a coclique level structure associated with the coclique level function $L(x_1, x_2, x_3, x_4) = 2x_1 - x_2 + x_3 + x_4$. This coclique level structure can be written as $\mathcal{L}_\ell, \dots, \mathcal{L}_u$, with $\mathcal{L}_z := \{x \in \check{\mathcal{X}} : L(x) = 2x_1 - x_2 + x_3 + x_4 = z\}$ for $z = \ell, \dots, u$, with $\ell = -D_{\text{tot}}$ and $u = 2D_{\text{tot}}$ (see Figure 4.5(c)). This coclique level structure is such that $a = (0, D_{\text{tot}}, 0, 0)^T$ is the only state belonging to $\mathcal{L}_\ell$ and $r = (D_{\text{tot}}, 0, 0, 0)^T$ is the only state belonging to $\mathcal{L}_u$. Here,

$$G_+ = \{2, 3, 5, 7, 9\} \quad \text{and} \quad G_- = \{1, 4, 6, 8, 10\}, \quad (4.6.8)$$

and the rate of increase $\lambda_z(\check{x})$ and the rate of decrease $\gamma_z(\check{x})$ can then be written as

$$\begin{aligned} \lambda_z(\check{x}) &= f_{R121}(\check{x}) + f_{R122}(\check{x}) + g_A(\check{x}) + f_{R1}(\check{x}) + f_{R2}(\check{x}), \\ \gamma_z(\check{x}) &= f_A(\check{x}) + g_{R121}(\check{x}) + g_{R122}(\check{x}) + g_{R1}(\check{x}) + g_{R2}(\check{x}), \end{aligned} \quad (4.6.9)$$

respectively, with $f_{R121}(\check{x})$, $f_{R122}(\check{x})$, $g_A(\check{x})$, $f_{R1}(\check{x})$, $f_{R2}(\check{x})$, $f_A(\check{x})$, $g_{R121}(\check{x})$, $g_{R122}(\check{x})$, $g_{R1}(\check{x})$, $g_{R2}(\check{x})$ defined in (4.6.7). The two continuous time Markov chains $\check{X}$ and $\hat{X}$ are defined on the same state space as $\check{X}$ and have infinitesimal generators $\check{Q}$ and $\hat{Q}$, respectively, such that, for $z \in \{\ell, \ell + 1, \dots, u - 1, u\}$ and $\check{x} \in \mathcal{L}_z$, $\check{Q}_{\check{x}, \check{x} + \check{v}_k} = \frac{\lambda_z^M}{|G_+(\check{x})|}$ for $k \in G_+(\check{x})$, $\check{Q}_{\check{x}, \check{x} + \check{v}_k} = \frac{\gamma_z^m}{|G_-(\check{x})|}$ for $k \in G_-(\check{x})$, $\hat{Q}_{\check{x}, \check{x} + \check{v}_k} = \frac{\lambda_z^m}{|G_+(\check{x})|}$ for $k \in G_+(\check{x})$, and $\hat{Q}_{\check{x}, \check{x} + \check{v}_k} = \frac{\gamma_z^M}{|G_-(\check{x})|}$ for $k \in G_-(\check{x})$, with $\lambda_z^M = \max_{\check{x} \in \mathcal{L}_z} \lambda_z(\check{x})$, $\lambda_z^m = \min_{\check{x} \in \mathcal{L}_z} \lambda_z(\check{x})$, $\gamma_z^M = \max_{\check{x} \in \mathcal{L}_z} \gamma_z(\check{x})$, and $\gamma_z^m = \min_{\check{x} \in \mathcal{L}_z} \gamma_z(\check{x})$, as defined in (4.5.1), and non-empty $G_+(\check{x})$ and $G_-(\check{x})$ defined as in (4.5.2) where G_+ and G_- are given in (4.6.8).

Then, as described in Section 4.5.3, we can compare the Markov chain $\check{X}$ with $\check{X}$ and $\hat{X}$, separately, to obtain analytical expressions for lower and upper bounds, respectively, for the MFPT $h_{a,r}$ from the fully active state to the fully repressed state:

$$\begin{aligned}\check{h}_{\ell,u} &= \frac{1}{\lambda_{u-1}^M} + \sum_{i=\ell}^{u-2} \frac{1}{\lambda_i^M} \left(1 + \sum_{j=i+1}^{u-1} \frac{\gamma_{i+1}^m \cdots \gamma_j^m}{\lambda_{i+1}^M \cdots \lambda_j^M} \right), \\ \hat{h}_{\ell,u} &= \frac{1}{\lambda_{u-1}^m} + \sum_{i=\ell}^{u-2} \frac{1}{\lambda_i^m} \left(1 + \sum_{j=i+1}^{u-1} \frac{\gamma_{i+1}^M \cdots \gamma_j^M}{\lambda_{i+1}^m \cdots \lambda_j^m} \right),\end{aligned}$$

and expressions for lower and upper bounds for the MFPT $h_{r,a}$ from the repressed state to the active state:

$$\begin{aligned}\hat{h}_{u,\ell} &= \frac{1}{\gamma_{\ell+1}^M} + \sum_{i=\ell+2}^u \frac{1}{\gamma_i^M} \left(1 + \sum_{j=\ell+1}^{i-1} \frac{\lambda_j^m \cdots \lambda_{i-1}^m}{\gamma_j^M \cdots \gamma_{i-1}^M} \right), \\ \check{h}_{u,\ell} &= \frac{1}{\gamma_{\ell+1}^m} + \sum_{i=\ell+2}^u \frac{1}{\gamma_i^m} \left(1 + \sum_{j=\ell+1}^{i-1} \frac{\lambda_j^M \cdots \lambda_{i-1}^M}{\gamma_j^m \cdots \gamma_{i-1}^m} \right).\end{aligned}$$

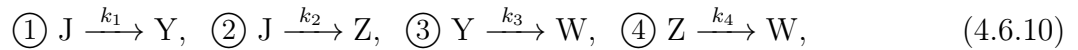
Given that the only $O(\varepsilon)$ rates are λ_ℓ^M , λ_ℓ^m , γ_u^M , γ_{u-1}^M , and $\gamma_z^m : D_{\text{tot}} \leq z \leq u$, with the other rates being $O(1)$, we can conclude that both $\check{h}_{\ell,u}$ and $\hat{h}_{\ell,u}$ are $O(\varepsilon^{-1})$, and thus the MFPT from a to r (average time to memory loss of the active state) is $O(\varepsilon^{-1})$. Furthermore, we have $\hat{h}_{u,\ell}$ is $O(\varepsilon^{-2})$ and $\check{h}_{u,\ell}$ is $O(\varepsilon^{-D_{\text{tot}}})$, and thus the MFPT from r to a (average time to memory loss of the repressed state) is at least $O(\varepsilon^{-2})$. In cases like this, alternative approaches are needed to identify the precise scalings of MFPTs, such as the method we developed in our recent work (see Bruno et al. [4]), which allows us to show that average time to memory loss of the repressed state is indeed $O(\varepsilon^{-2})$. These results suggest that decreasing ε extends the memory of both the active and repressed chromatin states, but with a more pronounced impact on the repressed state. This difference can be attributed to the cooperation of repressive chromatin marks, i.e., DNA methylation and repressive histone modifications, which introduces a structural bias in the chromatin

modification circuit towards a repressed chromatin state.

Let us now determine the effect of μ' , i.e., the parameter quantifying the relative speed between the DNA methylation erasure rate and the activating histone modification erasure rate, on the time to memory loss. Since the rates $g_{R1}(\tilde{x})$ and $g_{R121}(\tilde{x})$ are linear in μ' and are the only transition rates depending on μ' (see (4.6.7)), then, based on the definition in (4.6.9), $\gamma_z(\tilde{x})$ increases for lower values of μ' . This implies that increasing μ' leads to higher $\hat{h}_{\ell,u}$ and $\check{h}_{\ell,u}$ and lower $\check{h}_{u,\ell}$ and $\hat{h}_{u,\ell}$. The opposite happens when μ' decreases.

4.6.3 Bi-parallel network motif

In this example, we analyze a bi-parallel network motif, which is a typical building block found in complex networks (see for example Milo et al. [19]), such as the neuronal connectivity network of the nematode *Caenorhabditis elegans* or food web networks (see Milo et al. [19] and Brglez et al. [3]). The model includes four species, which are J, Y, Z, and W, and the quantity of each species is denoted by n_J , n_Y , n_Z , and n_W , respectively. Their sum remains constant, that is $n_J + n_Y + n_Z + n_W = S_{\text{tot}}$. Then, the chemical reaction system, whose diagram is shown in Figure 4.6(a), can be written as



in which $k_1, k_2, k_3, k_4 > 0$.

Considering $x = (n_Y, n_Z, n_W, n_J)$, the reaction vectors associated with (4.6.10) are $v_1 = (1, 0, 0, -1)^T$, $v_2 = (0, 1, 0, -1)^T$, $v_3 = (-1, 0, 1, 0)^T$, and $v_4 = (0, -1, 1, 0)^T$. By examining them, we see that Assumption 4.4.1 is satisfied. The graph $\mathcal{G}$ associated with the chemical reaction system (4.6.10) can be represented as in Figure 4.6(b). By inspecting $\mathcal{G}$, we see that the underlying undirected graph is connected and that $\mathcal{G}$ is bipartite. By Lemma 4.4.1, our SCRNs has a unique conservation vector $m = (1, 1, 1, 1)^T$ and then we can

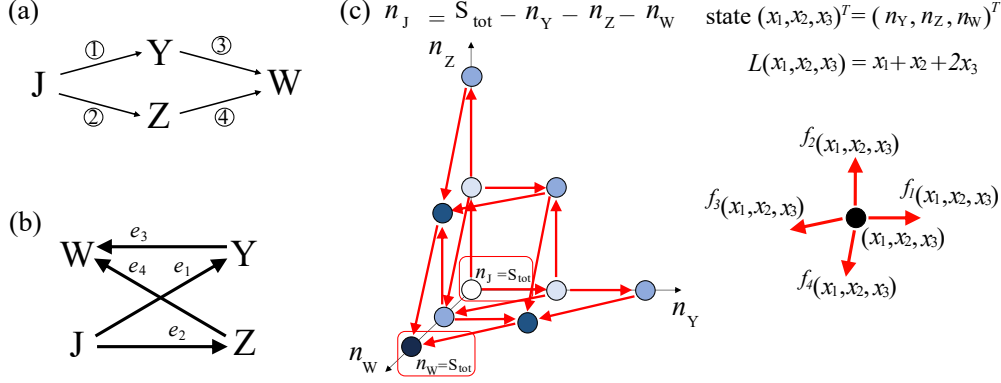


Figure 4.6. Bi-parallel network motif: reaction diagram, graph $\mathcal{G}$ and associated Markov chain. (a) Chemical reaction system diagram. The numbers on the arrows correspond to the reactions associated with the arrows as described in (4.6.1) in the main text. (b) Graph $\mathcal{G}$ associated with the chemical reaction system in panel (a). (c) State space and transitions for the projected continuous time Markov chain $\tilde{X} = \{(X_1(t), X_2(t), X_3(t))^T : t \geq 0\}$, which keeps track of (n_Y, n_Z, n_W) through time. Here, we consider $S_{\text{tot}} = 2$ and we use dots to represent the states, and red double-ended arrows to represent transitions in both directions. Additionally, we use shades of blue to distinguish the level to which each state belongs. The function $L(x_1, x_2, x_3)$ associated with the coclique level structure is $L(x_1, x_2, x_3) = x_1 + x_2 + 2x_3$. The rates associated with the one-step transitions for the projected Markov chain $\tilde{X}$ are given in (4.6.2).

introduce a projected continuous time Markov chain $\tilde{X} = \{(X_1(t), X_2(t), X_3(t))^T : t \geq 0\}$, which keeps track of (n_Y, n_Z, n_W) through time. Since S_{tot} is conserved, the state space is $\tilde{\mathcal{X}} = \{\tilde{x} = (x_1, x_2, x_3)^T \in \mathbb{Z}_+^3 : x_1 + x_2 + x_3 \leq S_{\text{tot}}\}$. The potential one-step transitions for $\tilde{X}$ from $x \in \tilde{\mathcal{X}}$ are shown in Figure 4.6(c), where the associated transition vectors are $\tilde{v}_1 = (1, 0, 0)^T$, $\tilde{v}_2 = (0, 1, 0)^T$, $\tilde{v}_3 = (-1, 0, 1)^T$, and $\tilde{v}_4 = (0, -1, 1)^T$, and the infinitesimal transition rates (in which we assume mass-action kinetics) are

$$\begin{aligned}
 \check{Q}_{\tilde{x}, \tilde{x} + \tilde{v}_1} &= f_1(\tilde{x}) = k_1(S_{\text{tot}} - (x_1 + x_2 + x_3)), \\
 \check{Q}_{\tilde{x}, \tilde{x} + \tilde{v}_2} &= f_2(\tilde{x}) = k_2(S_{\text{tot}} - (x_1 + x_2 + x_3)), \\
 \check{Q}_{\tilde{x}, \tilde{x} + \tilde{v}_3} &= f_3(\tilde{x}) = k_3x_1, \\
 \check{Q}_{\tilde{x}, \tilde{x} + \tilde{v}_4} &= f_4(\tilde{x}) = k_4x_2.
 \end{aligned} \tag{4.6.11}$$

Let us now focus on determining explicit analytical expressions for upper and lower bounds of MFPTs. To this end, we apply Theorem 4.5.1 to determine the coclique level

structures for $\check{X}$. We can apply the theorem because Assumption 4.4.1 is satisfied, and the associated graph $\mathcal{G}$ is weakly connected. The rank of the stoichiometric matrix S is $3 = d - 1$.

Now, consider all of the possible partitions $\{\mathcal{E}_+, \mathcal{E}_-\}$ of edges of $\mathcal{G}$ that could allow us to determine a coclique level structure. These partitions are the following:

$$\begin{aligned} \mathcal{E}_+ &= \{e_1, e_2, e_3, e_4\}, \mathcal{E}_- = \emptyset; & \mathcal{E}_+ &= \{e_1\}, \mathcal{E}_- = \{e_2, e_3, e_4\}; \\ \mathcal{E}_+ &= \{e_2\}, \mathcal{E}_- = \{e_1, e_3, e_4\}; & \mathcal{E}_+ &= \{e_3\}, \mathcal{E}_- = \{e_1, e_2, e_4\}; \\ \mathcal{E}_+ &= \{e_4\}, \mathcal{E}_- = \{e_1, e_2, e_3\}; & \mathcal{E}_+ &= \{e_1, e_2\}, \mathcal{E}_- = \{e_3, e_4\}; \\ \mathcal{E}_+ &= \{e_1, e_3\}, \mathcal{E}_- = \{e_2, e_4\}; & \mathcal{E}_+ &= \{e_1, e_4\}, \mathcal{E}_- = \{e_2, e_3\}. \end{aligned}$$

As was done for Example 4.6.1, we did not consider the partitions obtained by switching the labels of the subsets $\mathcal{E}_+$, $\mathcal{E}_-$ in the partitions listed above because, as explained in Remark 4.5.1, the associated functions L would be the opposite of the ones obtained for the partitions considered above. Therefore, the resulting coclique level structures may be considered to be the same.

For each partition, we can write the system of equations as in (4.4.3). The only systems that admit a solution $b \in \mathbb{Z}^3$ are the ones associated with the first, sixth, and eighth partitions. The solutions are $(b_1, b_2, b_3)^T = (1, 1, 2)^T$, $(b_1, b_2, b_3)^T = (1, 1, 0)^T$, and $(b_1, b_2, b_3)^T = (1, -1, 0)^T$, respectively. Then, by applying Theorem 4.5.1, we can conclude that the projected Markov chain $\check{X}$ has three coclique level structures associated with the coclique level functions $L(x_1, x_2, x_3) = x_1 + x_2 + 2x_3$, $L(x_1, x_2, x_3) = x_1 + x_2$, and $L(x_1, x_2, x_3) = x_1 - x_2$.

We consider the coclique level function $L(x_1, x_2, x_3) = x_1 + x_2 + 2x_3$. The coclique level structure associated with it can be written as $\mathcal{L}_\ell, \dots, \mathcal{L}_u$, with $\mathcal{L}_z := \{x \in \check{\mathcal{X}} : L(x_1, x_2, x_3) = x_1 + x_2 + 2x_3 = z\}$ for $z = \ell, \dots, u$, with $\ell = 0$ and $u = 2S_{\text{tot}}$. This coclique level structure is such that $(0, 0, 0)^T$ (i.e., $n_J = S_{\text{tot}}$) is the only state belonging to $\mathcal{L}_\ell$ and

$(0, 0, S_{\text{tot}})^T$ (i.e., $n_W = S_{\text{tot}}$) is the only state belonging to $\mathcal{L}_u$ (see Figure 4.6(c)). This feature, as shown in the previous examples, is critical in order to determine good lower and upper bounds for the MFPT from $n_J = S_{\text{tot}}$ to $n_W = S_{\text{tot}}$.

Let us now determine the lower and upper bounds for the MFPT from $n_J = S_{\text{tot}}$ to $n_W = S_{\text{tot}}$. In particular, here we have

$$G_+ = \{1, 2, 3, 4\} \quad \text{and} \quad G_- = \emptyset, \quad (4.6.12)$$

and the rate of increase $\lambda_z(\tilde{x})$ and the rate of decrease $\gamma_z(\tilde{x})$ can be written as

$$\lambda_z(\tilde{x}) = f_1(\tilde{x}) + f_2(\tilde{x}) + f_3(\tilde{x}) + f_4(\tilde{x}) \quad \text{and} \quad \gamma_z(\tilde{x}) = 0,$$

with $f_1(\tilde{x})$, $f_2(\tilde{x})$, $f_3(\tilde{x})$, $f_4(\tilde{x})$ defined in (4.6.11). The two continuous time Markov chains $\check{X}$ and $\hat{X}$, are defined on the same state space as $\tilde{X}$ and have infinitesimal generators $\check{Q}$ and $\hat{Q}$, respectively, such that, for $z \in \{\ell, \ell+1, \dots, u-1, u\}$ and $\tilde{x} \in \mathcal{L}_z$, $\check{Q}_{\tilde{x}, \tilde{x}+\check{v}_k} = \frac{\lambda_z^M}{|G_+(\tilde{x})|}$ for $k \in G_+(\tilde{x})$ and $\hat{Q}_{\tilde{x}, \tilde{x}+\check{v}_k} = \frac{\lambda_z^m}{|G_+(\tilde{x})|}$ for $k \in G_+(\tilde{x})$, with $\lambda_z^M = \max_{\tilde{x} \in \mathcal{L}_z} \lambda_z(\tilde{x})$ and $\lambda_z^m = \min_{\tilde{x} \in \mathcal{L}_z} \lambda_z(\tilde{x})$, as defined in (4.5.1), and non-empty $G_+(\tilde{x})$ defined as in (4.5.2) where G_+ is given in (4.6.12).

Then, as described in Section 4.5.3, we can compare the Markov chain $\check{X}$ with $\tilde{X}$ and $\hat{X}$, separately, to obtain analytical expressions for lower and upper bounds for the MFPT from $n_J = S_{\text{tot}}$ to $n_W = S_{\text{tot}}$, which can be written as

$$\check{h}_{\ell,u} = \frac{1}{\lambda_\ell^M} + \frac{1}{\lambda_{\ell+1}^M} + \dots + \frac{1}{\lambda_{u-2}^M} + \frac{1}{\lambda_{u-1}^M}, \quad \hat{h}_{\ell,u} = \frac{1}{\lambda_\ell^m} + \frac{1}{\lambda_{\ell+1}^m} + \dots + \frac{1}{\lambda_{u-2}^m} + \frac{1}{\lambda_{u-1}^m},$$

respectively. Since $\lambda_z^M = \max_{\tilde{x} \in \mathcal{L}_z} \lambda_z(\tilde{x}) = \max_{\tilde{x} \in \mathcal{L}_z} (f_1(\tilde{x}) + f_2(\tilde{x}) + f_3(\tilde{x}) + f_4(\tilde{x}))$, and $\lambda_z^m = \min_{\tilde{x} \in \mathcal{L}_z} \lambda_z(\tilde{x}) = \min_{\tilde{x} \in \mathcal{L}_z} (f_1(\tilde{x}) + f_2(\tilde{x}) + f_3(\tilde{x}) + f_4(\tilde{x}))$, we observe that lower and upper bounds $\check{h}_{\ell,u}$ and $\hat{h}_{\ell,u}$ for the MFPT under consideration decrease as any of the rate constants k_1, k_2, k_3, k_4 increases. This result suggests that, if any reaction associated with

either pathway in the bi-parallel network becomes faster, then the upper and lower bounds for the mean time needed to transform all J into all W (i.e., the MFPT from $n_J = S_{\text{tot}}$ to $n_W = S_{\text{tot}}$) decrease.

We note that, by following the same approach used above, these results can be generalized to any network having two species (J and W) connected by n-parallel pathways (beyond just two).

4.7 Conclusion

In this work, we started by providing a description of Stochastic Chemical Reaction Networks (SCRNs), which are a class of continuous time Markov chain models commonly used to describe the stochastic behavior of chemical reaction systems (see Section 4.3). We then introduced the notion of *coclique level structure* (see Section 4.4) and developed theoretical tools for identifying such coclique level structures for continuous time Markov chains associated with SCRNs, where each reaction involves the consumption of one molecule of a given species and the production of one molecule of another species, and an associated graph $\mathcal{G}$ is weakly connected (see Section 4.5.1). Additionally, we provided conditions for identifying if a SCRN does or does not admit a coclique level structure (see Section 4.5.1). Finally, we derived analytical expressions for upper and lower bounds for MFPTs of SCRNs having a coclique level structure (see Section 4.5.3).

Following this, we provided illustrative examples to demonstrate the utility of our theoretical tools in studying the stochastic behavior of SCRNs (see Section 4.6). More precisely, we focused on models describing the main interactions among histone modifications alone, and in combination with DNA methylation (see Bruno et al. [5]), as well as on a bi-parallel network motif, a typical building block found in complex networks (see Milo et al. [19]). Through these examples, we demonstrated that our algorithm for identifying coclique level structures is easy to apply, and the analytical

expressions for upper and lower bounds of MFPTs obtained with our theoretical tools provide mechanistic insights into how system parameters affect the stochastic behavior of SCRN. This mechanistic insight is particularly valuable for applications where it is crucial to understand which biological parameters must be tuned to modulate system dynamics in a specific manner.

The mathematical results and theoretical tools developed in this work can be applied to all stochastic models that meet the considered assumptions. Future work will be focused on generalizing these results by relaxing some of these assumptions, such as allowing the Markov chain to have countably many states and allowing SCRN to have more than one molecule consumed and produced per reaction. Another valuable direction for future work is the implementation of the algorithm introduced in Section 4.5.2. While in this work we focused on providing theoretical guidance, a practical implementation of the algorithm would broaden the applicability of our approach and facilitate its integration into simulation pipelines.

4.8 Supplementary Information (SI)

4.8.1 Proof of Lemma 4.4.1

Before introducing the proof of Lemma 4.4.1, we provide some definitions used in the proof. The **degree** of a vertex is the number of edges that are incident to the vertex, where we count both incoming and outgoing edges. A **weakly directed tree** is a directed graph whose underlying undirected graph is a tree. Given a directed graph with d vertices, a **weakly directed spanning tree** is a subgraph of the graph with all d vertices and such that it is a weakly directed tree, and it will have $d - 1$ edges. We abbreviate weakly directed spanning tree as wd-spanning tree. Please note that a weakly connected graph $\mathcal{G}$ has a wd-spanning tree.

Proof. Consider a wd-spanning tree of the graph $\mathcal{G}$. Let $\hat{\mathcal{V}}_{st} = \{\hat{v}_1, \hat{v}_2, \dots, \hat{v}_{d-1}\}$ denote

the set of $d - 1$ reaction vectors associated with the wd-spanning tree edges. We first show that the stoichiometric matrix associated with $\widehat{\mathcal{V}}_{st}$, $S_{d,d-1} = [\widehat{v}_1, \widehat{v}_2, \dots, \widehat{v}_{d-1}] \in \mathbb{Z}^{d \times (d-1)}$ has rank $d - 1$. We prove this by induction. First consider $d = 2$. In this case the only possible wd-spanning tree is given by the two vertices and one edge connecting them. This means that $\text{rank}(S_{2,1}) = 1$. Then, let us assume that the result is true for any wd-spanning tree with $d - 1$ vertices (i.e., $\text{rank}(S_{d-1,d-2}) = d - 2$) for some $d \geq 3$, and consider a wd-spanning tree with d vertices. This wd-spanning tree always has a degree-one vertex, that we define as *vertex 1*, and then we can rearrange the rows and columns of $S_{d,d-1}$ so that

$$S_{d,d-1} = \begin{bmatrix} \pm 1 & 0 \\ * & S_{d-1,d-2} \end{bmatrix}.$$

Then, $S_{d-1,d-2}$ is the stoichiometric matrix associated with the wd-spanning tree with the *vertex 1* and associated edge removed. By the induction hypothesis, we can conclude that $S_{d-1,d-2}$ has rank $d - 2$, and then $\text{rank}(S_{d,d-1}) = d - 1$. Since $S_{d,d-1}$ is obtained by removing columns associated with the reaction vectors $v_k \in \mathcal{V} \setminus \widehat{\mathcal{V}}_{st}$ from S , this result implies that $\text{rank}(S) \geq d - 1$.

Under Assumption 4.4.1, $\mathbb{1}^T v_k = 0$ for $k = 1, \dots, n$, and thus $\mathbb{1}^T S = 0$. This means that $\mathbb{1} \in \ker(S^T)$ and then $\dim \ker(S^T) \geq 1$. By the rank-nullity theorem (see for example Friedberg et al. [11]), we have that $\text{rank}(S) = \text{rank}(S^T) = d - \dim \ker(S^T) \leq d - 1$. Putting together the results obtained, we can conclude that $\text{rank}(S) = d - 1$.

Furthermore, given that $\mathbb{1}^T S = 0$ and $\dim \ker(S^T) = 1$, then $\mathbb{1}$ is the only (up to scalar multiplication) conservation vector such that $S^T \mathbb{1} = 0$. $\square$

4.8.2 Proof of Theorem 4.5.4

Let $\check{S}^q \in \mathbb{Z}^{(d_q-1) \times n}$ be the first $(d_q - 1)$ rows of the stoichiometric matrix S^q . Then, similar to (4.4.6) and since the stoichiometric matrix for the SCRNs has the form (4.5.13),

the system (4.5.17) can be re-written in matrix-vector form as

$$\begin{bmatrix} (\check{S}^1)^T & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & (\check{S}^p)^T \end{bmatrix} b = w, \quad (4.8.1)$$

which is equivalent to

$$(\check{S}^q)^T b^q = w^q \quad \text{for every } q = 1, \dots, p \text{ such that } |\mathcal{G}^q| > 1,$$

where for $q = 1, \dots, p$, the vectors $b^q, w^q \in \mathbb{Z}^{d_q-1}$ are the q^{th} entries of $b = (b^1, \dots, b^p)^T$ and $w = (w^1, \dots, w^p)^T$, respectively. Note that for $q = 1, \dots, p$, if $|\mathcal{G}^q| = 1$, then $\check{S}^q \in \mathbb{Z}^{(d_q-1) \times n}$ is a $0 \times n$ matrix, which does not appear in (4.8.1), and $\check{x}^q$ is a zero-dimensional vector, and so we do not need to consider that component in any coclique level function, as in (4.5.18). Consider $q = 1, \dots, p$ such that $|\mathcal{G}^q| > 1$. As noted before Theorem 4.5.4, there is a SCRN associated with each $\mathcal{G}^q$. The stoichiometric matrix for this SCRN is S^q . By Theorem 4.5.1, $(\check{S}^q)^T b^q = w^q$ has a solution $b^q \in \mathbb{Z}^{d_q-1}$ if and only if $L^q : \mathbb{Z}^{d_q-1} \rightarrow \mathbb{Z}$ given by $L^q(\check{x}^q) = (b^q)^T \check{x}^q$ is a coclique level function for the SCRN associated with $\mathcal{G}^q$. Thus, L is a coclique level function for the whole SCRN if and only if

$$L(\check{x}) = b^T \check{x} = (b^1, \dots, b^p)^T (\check{x}^1, \dots, \check{x}^p) = \sum_{\substack{q=1, \dots, p: \\ |\mathcal{G}^q| > 1}} (b^q)^T \check{x}^q = \sum_{\substack{q=1, \dots, p: \\ |\mathcal{G}^q| > 1}} L^q(\check{x}^q),$$

where L^q is a coclique level function for the SCRN associated with $\mathcal{G}^q$ where $|\mathcal{G}^q| > 1$.

For each weakly connected component $\mathcal{G}^q$ such that $|\mathcal{G}^q| > 1$, by Theorem 4.5.3, there exists a coclique level structure for the SCRN associated with $\mathcal{G}^q$ if and only if $\mathcal{G}^q$ is bipartite. Since $\mathcal{G}$ is bipartite if and only if each $\mathcal{G}^q$ with $|\mathcal{G}^q| > 1$ is bipartite, we conclude that a coclique level function for $\check{X}$ exists if and only if $\mathcal{G}$ is bipartite. $\square$

4.8.3 One-dimensional birth-death process: Mean first passage time.

Let us consider a one-dimensional irreducible finite state continuous time Markov chain in which the state space $\mathcal{X} = \{0, 1, \dots, K\}$ and the off-diagonal entries of the infinitesimal generator Q are all zero except for the following positive rates:

$$\begin{aligned} Q_{x,x+1} &= \lambda_x & \text{if } x \in \{0, \dots, K-1\}, \\ Q_{x,x-1} &= \gamma_x & \text{if } x \in \{1, \dots, K\}. \end{aligned} \tag{4.8.2}$$

In other words, Q is the infinitesimal generator for a finite state birth-death process.

We will determine an analytical expression for the MFPT from $x = K$ to $x = 0$ and from $x = 0$ to $x = K$ for this chain. To this end, it is important to note that X can be equivalently characterized using holding times with exponential parameters $\{q_x\}_{x \in \mathcal{X}}$ and a transition probability matrix P for the *embedded discrete time Markov chain*. More precisely, for each $x \in \mathcal{X}$, $q_x = -Q_{x,x} \neq 0$, since X is irreducible, and for all $x, y \in \mathcal{X}$, $P_{x,x} = 0$, $P_{x,y} = \frac{Q_{x,y}}{q_x}$, for $y \neq x$ in $\mathcal{X}$. Note that $Q = \text{diag}(q)(P - I)$. Defining $\mathcal{B}$ as a nonempty subset of $\mathcal{X}$ such that $\mathcal{B} \neq \mathcal{X}$ and using first step analysis (see for example Equation (3.1) in Norris [20]), we obtain that the MFPT from x to $\mathcal{B}$ can be written as

$$h_{x,\mathcal{B}} = \begin{cases} 0 & \text{if } x \in \mathcal{B} \\ \frac{1}{q_x} + \sum_{y \in \mathcal{X}} P_{x,y} h_{y,\mathcal{B}} & \text{if } x \in \mathcal{B}^c. \end{cases} \tag{4.8.3}$$

Now, let us first focus on the MFPT from $x = K$ to $x = 0$. In this case $\mathcal{B} = \{0\}$

and then (4.8.3) can be rewritten as

$$\begin{cases} h_{0,0} = 0, \\ h_{x,0} = \frac{1}{\lambda_x + \gamma_x} + \frac{\lambda_x}{\lambda_x + \gamma_x} h_{x+1,0} + \frac{\gamma_x}{\lambda_x + \gamma_x} h_{x-1,0} & \text{if } x \in \{1, \dots, K-1\}, \\ h_{K,0} = \frac{1}{\gamma_K} + h_{K-1,0}, \end{cases} \quad (4.8.4)$$

where for $x, y \in \mathcal{X}$, $h_{x,y} = \mathbb{E}_x[\tau_y]$, $\tau_y = \inf\{t \geq 0 : X(t) = y\}$, X is the continuous time Markov chain with infinitesimal generator given by (4.8.2). Now, defining $\Delta h_{x,x-1} = h_{x,0} - h_{x-1,0}$ for $x \in \{1, \dots, K\}$, we can rewrite (4.8.4) in the following way:

$$\begin{cases} h_{0,0} = 0, \\ \Delta h_{x,x-1} = \frac{1}{\gamma_x} + \frac{\lambda_x}{\gamma_x} \Delta h_{x+1,x} & \text{if } x \in \{1, \dots, K-1\}, \\ \Delta h_{K,K-1} = \frac{1}{\gamma_K}. \end{cases} \quad (4.8.5)$$

From (4.8.5), we have an explicit formula for $\Delta h_{K,K-1}$ and any $\Delta h_{x,x-1}$ can be expressed as a function of $\Delta h_{x+1,x}$. Furthermore, if we sum the $\Delta h_{x,x-1}$ for $x = 1, \dots, K$, we obtain

$$h_{K,0} = h_{K,0} - h_{0,0} = \sum_{x=1}^K (\Delta h_{x,x-1}) = \Delta h_{1,0} + \Delta h_{2,1} + \dots + \Delta h_{K-1,K-2} + \Delta h_{K,K-1}.$$

Thus, to evaluate the MFPT from $x = K$ to $x = 0$, we can calculate $\Delta h_{x,x-1}$ for $x = K, K-1, \dots, 1$ and then sum all of the terms. We then obtain

$$\begin{aligned} h_{K,0} &= \frac{1}{\gamma_K} \left(1 + \frac{\lambda_{K-1}}{\gamma_{K-1}} + \frac{\lambda_{K-1}\lambda_{K-2}}{\gamma_{K-1}\gamma_{K-2}} + \dots + \frac{\lambda_{K-1}\dots\lambda_1}{\gamma_{K-1}\dots\gamma_1} \right) \\ &\quad + \frac{1}{\gamma_{K-1}} \left(1 + \frac{\lambda_{K-2}}{\gamma_{K-2}} + \frac{\lambda_{K-2}\lambda_{K-3}}{\gamma_{K-2}\gamma_{K-3}} + \dots + \frac{\lambda_{K-2}\dots\lambda_1}{\gamma_{K-2}\dots\gamma_1} \right) + \dots + \frac{1}{\gamma_1} \\ &= \frac{1}{\gamma_1} + \sum_{i=2}^K \frac{1}{\gamma_i} \left(1 + \sum_{j=1}^{i-1} \frac{\lambda_j \dots \lambda_{i-1}}{\gamma_j \dots \gamma_{i-1}} \right). \end{aligned} \quad (4.8.6)$$

With a similar procedure, we can obtain the MFPT from $x = 0$ to $x = K$. More precisely, we have

$$\begin{aligned}
h_{0,K} &= \frac{1}{\lambda_0} \left(1 + \frac{\gamma_1}{\lambda_1} + \frac{\gamma_1\gamma_2}{\lambda_1\lambda_2} + \cdots + \frac{\gamma_1 \cdots \gamma_{K-1}}{\lambda_1 \cdots \lambda_{K-1}} \right) \\
&\quad + \frac{1}{\lambda_1} \left(1 + \frac{\gamma_2}{\lambda_2} + \frac{\gamma_2\gamma_3}{\lambda_2\lambda_3} + \cdots + \frac{\gamma_2 \cdots \gamma_{K-1}}{\lambda_2 \cdots \lambda_{K-1}} \right) + \cdots + \frac{1}{\lambda_{K-1}} \\
&= \frac{1}{\lambda_{K-1}} + \sum_{i=0}^{K-2} \frac{1}{\lambda_i} \left(1 + \sum_{j=i+1}^{K-1} \frac{\gamma_{i+1} \cdots \gamma_j}{\lambda_{i+1} \cdots \lambda_j} \right). \tag{4.8.7}
\end{aligned}$$

A more detailed derivation of the $h_{0,K}$ and $h_{K,0}$ is given in Bruno et al. [5].

4.8.4 Theorems 3.3 and 3.4 from Campos et al. [8]

Let A be an $m \times d$ matrix, with no rows identically zero, and let $K_A = \{y \in \mathbb{R}^d : Ay \geq 0\}$. For $x, y \in \mathbb{R}^d$, we say that $x \preceq_A y$ whenever $A(y - x) \geq 0$. For a non-empty set $\Gamma \subseteq \mathcal{X} \subseteq \mathbb{Z}_+^d$, we say that a set Γ is increasing in $\mathcal{X}$ with respect to $\preceq_A$ if for every $x \in \Gamma$ and $y \in \mathcal{X}$, $x \preceq_A y$ implies that $y \in \Gamma$. Moreover, we say that a set $\Gamma \subseteq \mathcal{X}$ is decreasing in $\mathcal{X}$ with respect to $\preceq_A$ if for every $x \in \Gamma$ and $y \in \mathcal{X}$, $y \preceq_A x$ implies that $y \in \Gamma$. Furthermore, for $x \in \mathbb{R}^d$, let $K_A + x = \{y \in \mathbb{R}^d : x \preceq_A y\}$ and $\partial_i(K_A + x) := \{y \in K_A + x : \langle A_{i\bullet}, y \rangle = \langle A_{i\bullet}, x \rangle\}$ ⁶ for each $1 \leq i \leq m$. We can then characterize the boundary of $K_A + x$ as follows:

$$\partial(K_A + x) = \bigcup_{i=1}^m \partial_i(K_A + x).$$

Finally, we introduce the concept of usual stochastic order $\preceq_{st}$ for two random variables Y, Z : we say that Y is smaller than Z in the usual stochastic order, that is, $Y \preceq_{st} Z$, if $F_Y(t) \geq F_Z(t)$ for every $t \in \mathbb{R}$, where F_Y and F_Z are the cumulative distribution functions

⁶Here, for convenience of notation, let $A_{i\bullet}$ denote the row vector corresponding to the i -th row of A , for $1 \leq i \leq m$. In this article we will adopt the convention of considering the inner product $\langle \cdot, \cdot \rangle$ as a function of a row vector in its first entry and as a function of a column vector in the second entry. In particular, $\langle A_{i\bullet}, x \rangle = \sum_{k=1}^d A_{ik}x_k$.

for Y and Z , respectively.

In the following theorem, we consider the set of distinct vectors $\{\eta^1, \dots, \eta^s\}$ formed by Av_j , for $1 \leq j \leq n$, where s denotes the cardinality of this set, and we consider the subsets of indices

$$G^k := \{j : 1 \leq j \leq n \text{ and } Av_j = \eta^k\}, \quad \text{for } 1 \leq k \leq s.$$

The following theorem applies even if $\mathcal{X}$ is countably infinite, although in this work all of our state spaces are finite.

Theorem S.1 (immediate consequence of Theorem 3.3 in Campos et al. [8]⁷).

Consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, a collection of distinct vectors $v_1, \dots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$ and two collections of non-negative (intensity) functions on $\mathcal{X}$, $\Upsilon = (\Upsilon_1, \dots, \Upsilon_n)$ and $\check{\Upsilon} = (\check{\Upsilon}_1, \dots, \check{\Upsilon}_n)$ such that if $x + v_j \notin \mathcal{X}$, then $\Upsilon_j(x) = \check{\Upsilon}_j(x) = 0$, and assume the associated continuous time Markov chains (with intensity functions given by Υ and $\check{\Upsilon}$, respectively, for the transition directions $v_1, \dots, v_n$) do not explode in finite time. Consider a matrix $A \in \mathbb{Z}^{m \times d}$ with non-zero rows and suppose that both of the following conditions hold:

(i) *For each $1 \leq j \leq n$, the vector Av_j has entries in $\{-1, 0, 1\}$ only.*

(ii) *For each $x \in \mathcal{X}$, $1 \leq i \leq m$ and $y \in \partial_i(K_A + x) \cap \mathcal{X}$ we have that*

$$\sum_{j \in G^k} \check{\Upsilon}_j(y) \leq \sum_{j \in G^k} \Upsilon_j(x), \quad \text{for each } k \text{ such that } \eta_i^k < 0,$$

and

$$\sum_{j \in G^k} \check{\Upsilon}_j(y) \geq \sum_{j \in G^k} \Upsilon_j(x), \quad \text{for each } k \text{ such that } \eta_i^k > 0.$$

⁷Compared to Theorem 3.3 in Campos et al. [8], Theorem S.1 includes some additional clarifications inserted in parentheses to improve clarity and completeness in the context of this work.

Then, for each pair $x^\circ, \check{x}^\circ \in \mathcal{X}$ such that $x^\circ \preceq_A \check{x}^\circ$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with two continuous time Markov chains $X = \{X(t) : t \geq 0\}$ and $\check{X} = \{\check{X}(t) : t \geq 0\}$ defined there, each having state space $\mathcal{X} \subseteq \mathbb{Z}_+^d$, with infinitesimal generators Q and $\check{Q}$, associated with Υ and $\check{\Upsilon}$ respectively (as in (4.3.2)), with initial conditions $X(0) = x^\circ$ and $\check{X}(0) = \check{x}^\circ$ and such that:

$$\mathbb{P} \left[X(t) \preceq_A \check{X}(t) \text{ for every } t \geq 0 \right] = 1.$$

Theorem S.2 (immediate consequence of Theorem 3.4 in Campos et al. [8]⁸).

Consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}_+^d$, a collection of distinct vectors $v_1, \dots, v_n$ in $\mathbb{Z}^d \setminus \{0\}$ and two collections of non-negative (intensity) functions on $\mathcal{X}$, $\Upsilon = (\Upsilon_1, \dots, \Upsilon_n)$ and $\check{\Upsilon} = (\check{\Upsilon}_1, \dots, \check{\Upsilon}_n)$, such that if $x + v_j \notin \mathcal{X}$, then $\Upsilon_j(x) = \check{\Upsilon}_j(x) = 0$, and assume the associated continuous time Markov chains (with intensity functions given by Υ and $\check{\Upsilon}$, respectively, for the transition directions $v_1, \dots, v_n$) do not explode in finite time. Consider a matrix $A \in \mathbb{Z}^{m \times d}$ with non-zero rows and conditions (i) and (ii) in Theorem S.2 are satisfied.

Let $x^\circ, \check{x}^\circ \in \mathcal{X}$ be such that $x^\circ \preceq_A \check{x}^\circ$ and let $X = \{X(t) : t \geq 0\}$ and $\check{X} = \{\check{X}(t) : t \geq 0\}$ be two continuous time Markov chains (possibly defined on different probability spaces), each having state space $\mathcal{X} \subseteq \mathbb{Z}_+^d$, with infinitesimal generators Q and $\check{Q}$, associated with Υ and $\check{\Upsilon}$ respectively, and with initial conditions $X(0) = x^\circ$ and $\check{X}(0) = \check{x}^\circ$. For a non-empty set $\Gamma \subseteq \mathcal{X}$, consider $T_\Gamma := \inf\{t \geq 0 : X(t) \in \Gamma\}$ and $\check{T}_\Gamma := \inf\{t \geq 0 : \check{X}(t) \in \Gamma\}$. If Γ is increasing in $\mathcal{X}$ with respect to the relation $\preceq_A$, then

$$\check{T}_\Gamma \preceq_{st} T_\Gamma,$$

⁸Compared to Theorem 3.4 in Campos et al. [8], Theorem S.2 includes some additional clarifications inserted in parentheses to improve clarity and completeness in the context of this work.

and the mean first passage time of $\check{X}$ from $\check{x}^\circ$ to Γ is dominated by the mean first passage time of X from x° to Γ . If Γ is decreasing in $\mathcal{X}$ with respect to the relation $\preceq_A$, then

$$T_\Gamma \preceq_{st} \check{T}_\Gamma,$$

and the mean first passage time of X from x° to Γ is dominated by the mean first passage time of $\check{X}$ from $\check{x}^\circ$ to Γ .

The proof of these theorems can be found in Sections 5.3 and 3.3 of Campos et al. [8], respectively.

4.9 Acknowledgement

Chapter 4 is reproduced with minor enumeration and formatting differences from the paper “Coclique Level Structure for Stochastic Chemical Reaction Networks.” by Bruno, S., Fu, Y., Campos, F.A., Del Vecchio, D. and Williams, R.J., which has been submitted for publication and is currently under review. The dissertation author is an equal first author with S. Bruno on this paper.

Bibliography for Chapter 4

- [1] David F. Anderson and Thomas G. Kurtz. *Stochastic Analysis of Biochemical Systems*. Mathematical Biosciences Institute Graduate Lecture Series. Springer International Publishing, 2015.
- [2] Michael Backenköhler, Luca Bortolussi, and Verena Wolf. “Bounding mean first passage times in population continuous-time Markov chains”. In: *Quantitative Evaluation of Systems (QEST 2020): Lecture Notes in Computer Science*. Vol. 12289. Springer, Cham, 2020, pp. 155–174.

- [3] Franc Brglez, David Bryan, and Krzysztof Koźmiński. “Combinational profiles of sequential benchmark circuits”. In: *IEEE International Symposium on Circuits and Systems* 3 (1989), pp. 1929–1934. DOI: 10.1109/ISCAS.1989.100747.
- [4] Simone Bruno, Felipe A. Campos, Yi Fu, Domitilla Del Vecchio, and Ruth J. Williams. “Analysis of singularly perturbed stochastic chemical reaction networks motivated by applications to epigenetic cell memory”. In: *SIAM Journal on Applied Dynamical Systems* 23.4 (2024), pp. 2695–2731. DOI: 10.1137/23M1592389.
- [5] Simone Bruno, Ruth J. Williams, and Domitilla Del Vecchio. “Epigenetic cell memory: the gene’s inner chromatin modification circuit”. In: *PLOS Computational Biology* 18.4 (2022), pp. 1–27. DOI: 10.1371/journal.pcbi.1009961.
- [6] Simone Bruno, Ruth J. Williams, and Domitilla Del Vecchio. “Mathematical analysis of the limiting behaviors of a chromatin modification circuit”. In: *Mathematics of Control, Signals, and Systems* 3 (2023), pp. 399–432. DOI: 10.1007/s00498-023-00343-8.
- [7] Simone Bruno, Ruth J. Williams, and Domitilla Del Vecchio. “Model reduction and stochastic analysis of the histone modification circuit”. In: *2022 European Control Conference (ECC)* (2022), pp. 264–271. DOI: 10.23919/ECC55457.2022.9838047.
- [8] Felipe A. Campos, Simone Bruno, Yi Fu, Domitilla Del Vecchio, and Ruth J. Williams. “Comparison theorems for stochastic chemical reaction networks”. In: *Bulletin of Mathematical Biology* 85.39 (2023), pp. 1–39. DOI: 10.1007/s11538-023-01136-5.
- [9] Ian Dodd, Mille A. Micheelsen, Kim Sneppen, and Geneviève Thon. “Theoretical analysis of epigenetic cell memory by nucleosome modification”. In: *Cell* 129 (2007), pp. 813–822. DOI: 10.1016/j.cell.2007.02.053.
- [10] Martin Feinberg. “Chemical reaction network structure and the stability of complex isothermal reactors - I. the deficiency zero and deficiency one theorems”. In: *Chemical*

- Engineering Science* 42.10 (1986), pp. 2229–2268. DOI: 10.1016/0009-2509(87)80099-4.
- [11] Stephen H. Friedberg, Arnold J. Insel, and Lawrence E. Spence. *Linear Algebra (Fourth Edition)*. Pearson Education, 2014.
 - [12] Donald P. Gaver, Patricia A. Jacobs, and Guy Latouche. “Finite birth-and-death models in randomly changing environments”. In: *Advances in Applied Probability* 16.4 (1984), pp. 715–731. DOI: 10.2307/1427338.
 - [13] Eric L. Haseltine and James B. Rawlings. “Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics”. In: *The Journal of Chemical Physics* 117 (2002), pp. 6959–6969. DOI: 10.1063/1.1505860.
 - [14] Hye-Won Kang, Wasiur Rahman KhudaBukhsh, Heinz Koepl, and Grzegorz Rempala. “Quasi-steady-state approximations derived from the stochastic model of enzyme kinetics”. In: *Bulletin of Mathematical Biology* 81 (2019), pp. 1303–1336. DOI: 10.1007/s11538-019-00574-4.
 - [15] Veronica Kelsey and Colva M. Roney-Dougal. “Maximal cocliques in the generating graphs of the alternating and symmetric groups”. In: *Combinatorial Theory* 1 (2022), pp. 1–21. DOI: 10.5070/C62156879.
 - [16] Jon Kleinberg and Eva Tardos. *Algorithm Design (First edition)*. Pearson Education, Inc., 2005.
 - [17] Stavros Lopatzidis, Jasper De Bock, and Gert de Cooman. “Computing lower and upper expected first-passage and return times in imprecise birth-death chains”. In: *International Journal of Approximate Reasoning* 80.2 (2017), pp. 137–173. DOI: 10.1016/j.ijar.2016.08.008.
 - [18] Piet van Mieghem. *Graph Spectra for Complex Networks*. Cambridge University Press, 2010.

- [19] Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. “Network motifs: simple building blocks of complex networks”. In: *Nature* 298.5594 (2002), pp. 824–827. DOI: 10.1126/science.298.5594.824.
- [20] James R. Norris. *Markov chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997. DOI: 10.1017/CBO9780511810633.
- [21] Rishi K. Srivastava, Lingchong You, Jesse W. Summers, and John Yin. “Stochastic vs. deterministic modeling of intracellular viral kinetics”. In: *Journal of Theoretical Biology* 218.3 (2002), pp. 309–321. DOI: 10.1006/jtbi.2002.3078.
- [22] Robin J. Wilson. *Introduction to Graph Theory (Fourth Edition)*. Addison Wesley Longman, 1996.