

Advances in the stochastic and deterministic analysis of multistable biochemical networks



Andreas Petrides

Department of Engineering
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Trinity College

January 2018

To Thekla and my parents

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Andreas Petrides
January 2018

Acknowledgements

This work has been funded by the Engineering and Physical Sciences Research Council, to whom I am grateful. I would also like to thank Trinity College and the A.G. Leventis Foundation for additional financial assistance.

I am hugely indebted to Dr Glenn Vinnicombe for motivating the work in this thesis and providing me with constant inspiration. I have learnt a lot from him, especially on how to approach difficult problems methodologically with a clear mindset. I would also like to thank him for his unlimited patience and for his confidence in me, allowing me the freedom to develop my own path.

I am also grateful to Dr Timothy O'Leary, with whom I spent hours discussing my work, providing me with a different, yet very helpful, perspective, ultimately leading to the development of Chapter 7.

I would also like to thank the members of the control group, past and present, for providing company during my PhD, especially Ilario Cirillo, Luka Ribar, Thiago Burghi and Alessandro Morelli. Ilario and Luka have generously spent many hours listening my work, providing much helpful criticism. With Thiago and Alessandro I shared an office for most of the duration of my studies and I would like to thank them for their friendship.

On a more personal level, I would like to thank my parents, who have always provided me with a lot of love and support. I would also like to thank Fotios Logothetis and Theocharis Amanatides who have been great friends.

Finally I must thank Thekla, who has never lost faith in me and has given me endless patience, love and support.

Abstract

This dissertation is concerned with the potential multistability of protein concentrations in the cell that can arise in biochemical networks. That is, situations where one, or a family of, proteins may sit at one of two or more different steady state concentrations in otherwise identical cells, and in spite of them being in the same environment.

Models of multisite protein phosphorylation have shown that this mechanism is able to exhibit unlimited multistability. Nevertheless, these models have not considered enzyme docking, the binding of the enzymes to one or more substrate docking sites, which are separate from the motif that is chemically modified. Enzyme docking is, however, increasingly being recognised as a method to achieve specificity in protein phosphorylation and dephosphorylation cycles.

Most models in the literature for these systems are deterministic i.e. based on Ordinary Differential Equations, despite the fact that these are accurate only in the limit of large molecule numbers. For small molecule numbers, a discrete probabilistic, *stochastic*, approach is more suitable. However, when compared to the tools available in the deterministic framework, the tools available for stochastic analysis offer inadequate visualisation and intuition.

We firstly try to bridge that gap, by developing three tools: a) a discrete ‘nullclines’ construct applicable to stochastic systems - an analogue to the ODE nullclines, b) a stochastic tool based on a Weakly Chained Diagonally Dominant M-matrix formulation of the Chemical Master Equation and c) an algorithm that is able to construct non-reversible Markov chains with desired stationary probability distributions.

We subsequently prove that, for multisite protein phosphorylation and similar models, in the deterministic domain, enzyme docking and the consequent substrate enzyme-sequestration must inevitably limit the extent of multistability, ultimately to one steady state. In contrast, bimodality can be obtained in the stochastic domain even in situations where bistability is not possible for large molecule numbers.

We finally extend our results to cases where we have an autophosphorylating kinase, as for example is the case with Ca^{2+} /calmodulin-dependent protein kinase II (CaMKII), a key enzyme in synaptic plasticity.

Table of contents

List of figures	xvii
List of tables	xix
Notation	1
1 Dissertation Overview	5
1.1 Motivation and Summary of Contributions	5
1.2 Summary of mathematical tools and concepts	7
1.3 Papers associated with dissertation	8
2 Introduction to biochemical networks	9
2.1 Biochemical networks	9
2.2 Classical deterministic analysis of biochemical networks	10
2.2.1 Generalisation and Definitions	11
2.2.2 Exploitation of timescales: Michaelis-Menten equations	11
2.2.3 Cooperativity, the Hill Equation and Ultrasensitivity	13
2.3 The need for analysis in both deterministic and stochastic domains	14
2.3.1 The reaction rates now represent propensities	14
2.3.2 The Chemical Master Equation	16
2.3.3 The stochastic methods available	17
2.3.3.1 Truncation of state space	17
2.3.3.2 Kinetic Monte Carlo - Gillispie Exact Algorithm	17
2.3.3.3 Analytical approaches - Chemical Langevin Equation	18
2.3.4 Deterministic and stochastic rate constants	19
2.4 Deterministic vs Stochastic tools: A gap that needs to be bridged	19
2.5 Conclusion	21

3	Understanding the genetic toggle switch phenomena - Development of a new discrete ‘nullcline’ construct	23
3.1	Introduction	23
3.1.1	The genetic toggle switch	23
3.1.2	Stochastic nullclines in the literature and our approach	24
3.2	The deterministic analysis provides an invariant to the equilibrium size Jacobian	26
3.3	Revisiting the Chemical Master Equation and its graph theoretic interpretation	27
3.3.1	Chemical Master Equation	27
3.3.2	Directed spanning trees, the Markov chain tree theorem and its consequences	27
3.3.3	Discrete ‘nullcline’ construct proposed	29
3.3.3.1	New type of edge-based discretisation	29
3.3.3.2	Comparison of discretisation procedures on a given example	30
3.3.3.3	‘Nullcline’ construction	31
3.3.4	Applying the discrete nullclines to the genetic toggle switch example	36
3.4	The comparison with Numerical Methods illustrates that the heuristic method can provide upper-bound estimates	39
3.5	The discrete ‘nullcline’ construct can also be used in the asymmetric case .	42
3.6	Conclusion	43
4	A WCDD M-matrix reformulation of the Chemical Master Equation and an associated stochastic tool	45
4.1	Introduction	45
4.2	Development of stochastic tool	46
4.2.1	Reformulating the equation for finding the stationary distribution of the Chemical Master Equation	46
4.2.1.1	Subsection’s proofs	47
4.2.2	Weakly chained diagonally dominant (WCDD) M-matrix	48
4.3	Bounds of the stationary probability of a microstate	49
4.3.1	Section’s proofs	50
4.3.2	Tightness of upper and lower bounds	52
4.3.3	Using the bounds obtained for the development of a tool	53
4.3.4	Contributions with respect to the Markov chain literature	53
4.4	Conclusion	56

5	An algorithm for constructing not necessarily reversible Markov chains with specified discrete-state stationary probability distributions	57
5.1	Introduction	57
5.2	Revisiting the WCDD M-matrix formulation of the CME and the development of the algorithm	59
5.2.1	The first part of the algorithm is concerned with the creation of an irreducible weakly chained diagonally dominant M-matrix	59
5.2.2	The second part of the algorithm is equivalent to controlling the stationary probability distribution formation from a distinguished state	61
5.2.3	The resulting continuous time Markov chain can be converted to a discrete time Markov chain	62
5.3	Example - Graph and Spectral Structure of resulting Markov chain	64
5.4	Comparison with the Reversible Markov chain obtained from Metropolis-Hastings can show a better spectral gap	65
5.5	Advantages of this algorithm	66
5.6	Conclusion	67
6	Enzyme Sequestration by the Substrate: An Analysis in the Deterministic and Stochastic Domains	69
6.1	Introduction	69
6.2	Models	72
6.2.1	A deterministic framework for the excess substrate regime	72
6.2.2	Substrate-Kinase and Substrate-Phosphatase Sequestration	77
6.2.2.1	Model extension with the addition of Substrate Enzyme Sequestration	78
6.2.2.2	Comparison with Legewie sequestration mechanism	78
6.3	Results and Discussion	80
6.3.1	Substrate Enzyme-Sequestration in the Deterministic Domain	80
6.3.1.1	What happens when Substrate Enzyme-Sequestration is considered?	80
6.3.1.2	Sufficient Conditions for Further Limiting the Extent of Multistability	80
6.3.1.3	Does direct decrease of overall substrate/enzyme numbers have the same effect as Substrate Enzyme-Sequestration?	83
6.3.1.4	How does Substrate Enzyme-Sequestration limit multistability?	84

6.3.1.5	Substrate Enzyme-Sequestration effects are not necessarily limited to S_0 and S_n or even to multi-site protein phosphorylation	86
6.3.1.6	Generalisation to arbitrary processivity and sequentiality	87
6.3.2	Substrate Enzyme-Sequestration in the Stochastic Domain	89
6.3.2.1	The behaviour of Substrate Enzyme-Sequestration when the molecule numbers are small require a different analysis	89
6.3.2.2	The same strength of Substrate Enzyme-Sequestration, sufficient for monostability, can lead to both monomodality and bimodality in the stochastic domain, depending on the timescales of the individual sequestration parameters.	91
6.4	Bimodality is feasible even when only one phosphosite is available	93
6.5	Conclusion	96
6.6	Appendix	99
6.6.1	Deterministic Framework	99
6.6.1.1	Theoretical development of the conditions on limiting the extent of multistability	99
6.6.1.2	Representing the derived conditions for limitation of multistability with kinetic parameters	100
6.6.1.3	Results based on the Pratt's tableau	101
6.6.1.4	Algorithm for finding δ^K using the Pratt tableau	104
6.6.1.5	Algorithm for finding δ^P using the Pratt tableau	105
6.6.1.6	Generalisation of results	105
6.6.1.7	Parameters of multisite protein phosphorylation systems under investigation	107
6.6.1.8	Enzyme-sharing scheme parameters	108
7	Substrate Enzyme-Sequestration adapted for CaMKII and a potential role in synaptic plasticity	109
7.1	Introduction	109
7.2	Bistable models of CaMKII in literature	110
7.3	Can Substrate Enzyme-Sequestration models be adapted to incorporate the particular characteristics of CaMKII?	111
7.4	Substrate Enzyme-Sequestration can induce bimodality and consequently activation of CaMKII, which becomes more difficult as system size expands	111
7.5	What can we learn from this model?	116
7.6	Conclusion	117

7.7 Appendix - Parameters used for the simulations of Chapter 7 118

8 Conclusions 119

8.1 Key findings 119

8.2 Further research 120

References 121

List of figures

2.2.1 Ultrasensitive vs Michaelian hyperbolic response	15
2.2.2 Input-output relationships and local sensitivities of Hill functions	15
3.3.1 Illustration of directed spanning tree formation	28
3.3.2 Comparison between standard and proposed discretisation method	32
3.3.3 Inferring the nodes with largest stationary probabilities	33
3.3.4 Illustration of discrete ‘nullcline’ visualisation procedure	34
3.3.5 Stationary probability distributions and discrete ‘nullclines’ for cooperative binding	36
3.3.6 ‘Nullcline’ gap formation in cooperative binding	37
3.3.7 Stationary probability distributions and discrete ‘nullclines’ for non-cooperative binding	38
3.4.1 Comparison with numerical results	40
3.5.1 Discrete ‘nullclines’ and stationary probability distribution in asymmetric case	42
4.3.1 Source/sink analogy of the formulation of the Chemical Master Equation	50
4.3.2 Sensitivity	54
5.2.1 Visualisation of weakly chained and strictly diagonally dominant M-matrices	60
5.2.2 Graphical illustration of the methodology followed.	62
5.3.1 Graph and Spectral Structure of resulting Non-reversible Markov chain	64
5.4.1 Graph and Spectral Structure of Metropolis-Hastings reversible Markov chain	65
5.5.1 Illustration of how pre-specified topology can also be satisfied	68
6.1.1 Inactive complex formation	71
6.2.1 Extension of phosphorylation scheme with enzyme docking considered	72
6.2.2 Tristability presented by either time courses or by plots of total concentrations of kinase and phosphatase as functions of free kinase and phosphatase	76
6.2.3 Illustration of Legewie mechanism	79

6.3.1	Limitation of multistability due to Substrate Enzyme-Sequestration	81
6.3.2	Tightness of derived sufficient conditions	82
6.3.3	Substrate Enzyme-Sequestration is not equivalent to just lowering substrate and enzyme concentrations	85
6.3.4	Robustness of tristability to direct decrease of enzyme concentration numbers	86
6.3.5	Sequestration of kinase by a kinase-substrate intermediate	87
6.3.6	Substrate Enzyme-Sequestration in a different enzyme-sharing mechanism .	88
6.3.7	Stochastic simulation exhibiting bimodality	90
6.3.8	Importance of the timescales of the sequestration parameters in the stochastic domain	92
6.4.1	Development of microstate grid from macroscopic reactions	94
6.4.2	15-substrate molecule single phosphosite system with no sequestration . . .	95
6.4.3	One phosphosite reaction scheme with inactive complexes PS_0 and KS_1 added	96
6.4.4	Design table using characterisation ratios	97
7.3.1	Substrate Enzyme-Sequestration model adapted for CaMKII	112
7.4.1	Bimodality achieved with PP1 Sequestration by CaMKII	114
7.4.2	Extinction of a species and its consequences	115
7.4.3	Monomodality due to expansion of system size	115
7.5.1	Overview of the scheme with negative feedback obtained via changing the system size, moving from the stochastic to the deterministic regime	117

List of tables

1.1	Summary of the mathematical tools and concepts used	8
6.1	The parameters of the original multisite protein phosphorylation system by Thomson and Gunawardena	107
6.2	The stochastic parameters used in the single phosphosite system	107
6.3	The parameters for the enzyme sharing scheme	108
7.1	The stochastic parameters used in the CaMKII system simulations	118

Notation

Graphs and Spanning Trees

$\alpha((x_0, y_0), (x_1, y_1))$	Transition rate between two states represented by Cartesian coordinates. The transition rate from state (x_0, y_0) to state (x_1, y_1) .
E	Set of edges.
G	Graph.
T	Directed spanning tree.
V	Set of vertices.
w	Weight function defined on edges.
$W(T)$	Weight of directed spanning tree.
$w_{i,j}$	The weight of the edge directed from vertex i to vertex j .

Phosphorylation networks

$[S]$	Concentration of species S .
$[S_{tot}]$	Total concentration of species S .
α_i^K	Rate of kinase attachment to substrate S with i phosphorylated sites.
α_i^P	Rate of phosphatase attachment to substrate S with i phosphorylated sites.
β_i^K	Rate of kinase de-attachment from substrate S with i phosphorylated sites.
β_i^P	Rate of phosphatase de-attachment from substrate S with i phosphorylated sites.

γ_i^K	Rate of phosphorylation resulting from the complex KS_i producing S_{i+1} .
γ_i^P	Rate of dephosphorylation resulting from the complex PS_i producing S_{i-1} .
K	Kinase.
KS_i	Complex formed after attachment of the kinase to the substrate S with i phosphorylated sites.
P	Phosphatase.
PS_i	Complex formed after attachment of the phosphatase to the substrate S with i phosphorylated sites.
S	Substrate.
S_i	Substrate S with i phosphorylated sites.

Markov chains, Probability and Linear Algebra

$\lambda_{min}(\mathbf{A})$	The smallest eigenvalue of matrix \mathbf{A} .
\mathbf{A}	Matrix of transition rates (propensities).
$\mathbf{A}_{i,j}^D$	The sub-matrix formed after deleting the i^{th} row and j^{th} column from matrix \mathbf{A} .
\mathbf{P}_s	The stationary probability distribution.
$\mathbf{P}(t)$	The probability landscape at time t .
$\mathbf{x}(t)$	The state (microstate) of the system at time t . The set of molecule numbers of each species at a particular moment.
$\ \cdot \ $	Matrix norm.
$\sigma_{max}(\mathbf{A})$	The largest singular value of matrix \mathbf{A} .
$A(\mathbf{x}_i, \mathbf{x}_j)$	The transition rate from state \mathbf{x}_j to state \mathbf{x}_i .
A_j^i	The j^{th} column of matrix \mathbf{A} with element i deleted.
A_j	The j^{th} column of matrix \mathbf{A} .

a_{ij} or $a_{i,j}$ The element of \mathbf{A} found in row i and column j . $P_s(j)$ or P_s^j The stationary probability of microstate \mathbf{x}_j .

Chapter 1

Dissertation Overview

1.1 Motivation and Summary of Contributions

Multistability, the ability of a cellular component to be found in more than one stable steady state under particular biological conditions, is considered fundamental for understanding cellular decision making. Understanding cellular decision making is required to understand cell signalling, that is, the communication between cells to regulate their cellular activities. This regulation can either be done at the genetic scale, involving the regulation of transcription (of DNA to mRNA) or translation (of mRNA to protein), or at the protein level via post-translational protein modifications. An example of such modification is protein phosphorylation.

Traditionally, the analysis of multistability is based on models of biochemical networks involving Ordinary Differential Equations. This deterministic type of analysis, despite its convenience and its large accompanying mathematical toolbox, is only exactly accurate in the limit of infinite molecule population sizes. When the molecule numbers are large but finite, multistability of the differential equations manifests itself as multimodality. The modes correspond to the stable steady states of the system, and the system undergoes fluctuations within, and random jumps between, the modes. When molecule numbers are small, however, there may be little relationship between the continuous deterministic and discrete stochastic analyses.

A well-studied example is the genetic toggle switch, which, in the absence of cooperativity, is predicted to have only one stable steady state, whereas experimental results and exact stochastic simulations have shown that the system exhibits bimodality. This is usually referred to as ‘noise-induced’ bimodality. Understanding this type of bimodality is hard, yet even harder is the prediction of when this would take place.

Taking these into consideration, it is evident that to fully appreciate the implications of a

biochemical mechanism to the potential multistability of the biochemical system, one should perform analyses in both the deterministic and the stochastic domains.

The biochemical mechanism we are primarily interested in this dissertation is multisite protein phosphorylation, because of its theoretical ability to provide unlimited multistability. In the presence of excess substrate, it has been shown (deterministically) that the achievable number of stable steady states can increase linearly with the number of phosphosites available. The models used in the aforementioned analysis of multisite protein phosphorylation, however, do not consider the effects of enzyme docking, despite it being increasingly recognised as a method of providing specificity in protein phosphorylation and dephosphorylation cycles. Enzyme docking refers to the ability of the kinase (the enzyme catalysing the addition of a phosphate group) or phosphatase (the enzyme catalysing the removal of a phosphate group) to bind with one or more sites on the substrate, which are independent of the catalytic site, making it therefore possible to form non-modifiable inactive complexes. For example, a phosphatase molecule could be attached to a completely unphosphorylated molecule.

To perform the analysis of that system in the stochastic domain, we first need to understand how stationary probability distributions are formed, leading to the development of the necessary stochastic tools.

In particular, in Chapter 3, recognising that ODE nullclines are a poor predictor of the behaviour of discrete state stochastic systems in the low numbers regime, we propose a discrete ‘nullcline’ heuristic construct inspired by the Markov chain tree theorem to understand the discrete genetic toggle switch phenomena. These ‘nullclines’ can, without the need to calculate the steady state distribution, provide an efficient graphical way of predicting the shape of the stationary probability distribution in different parameter regimes, thus allowing for greater insights in the observed behaviours.

In Chapter 4, we focus on the stationary probability of each particular discrete state of the state-space grid that the system can be found in (‘microstate’). Based on a Weakly Chained Diagonally Dominant M-matrix formulation of the Chemical Master Equation, we develop a stochastic tool that separates the effect of the output propensities of the particular microstate from the rest of the parameters of the system (a source/sink analogy).

In Chapter 5, in an effort to better understand how stationary probability distributions are formed, we use the results of Chapter 4 to develop an algorithm for constructing Markov chains on finite discrete spaces, both discrete and continuous-time, with specified discrete-state stationary probability distributions. The Markov chains constructed do not need to be reversible and are not modifications of prior constructions of reversible Markov chains.

In Chapter 6, having developed our stochastic toolbox, we analyse the consequences of adding enzyme docking to models of multisite protein phosphorylation and similar models

with the resultant sequestration of phosphatase and kinase by the fully unphosphorylated and by the fully phosphorylated substrates respectively. In the large molecule numbers limit, where deterministic analysis is applicable, we prove that there are always values for these rates of sequestration which, when exceeded, limit the extent of multistability. For the models considered here, these numbers are much smaller than the affinity of the enzymes to the substrate when it is in a modifiable state. As substrate enzyme-sequestration is increased, we further prove that the number of steady states will inevitably be reduced to one. For smaller molecule numbers where a stochastic analysis is more appropriate, we find that substrate enzyme sequestration can induce bimodality even in systems where only a single steady state can exist at large numbers. To facilitate this analysis, we use the stochastic tool based on the Weakly Chained Diagonally Dominant (WCDD) M-matrix formulation of the Chemical Master Equation developed in Chapter 4, allowing greater insights in the way enzyme sequestration can shape probability distributions and therefore exhibit different behaviour across different regimes.

In Chapter 7, we extend the framework of Chapter 6 to include the case when the substrate can act both as the substrate and as the kinase, autophosphorylating itself, a behaviour exhibited by Ca^{2+} /calmodulin-dependent protein kinase II (CaMKII), a molecule with a dominant role in synaptic plasticity. We speculate that the induced bimodality together with its associated dependence on the system size, could provide an explanation of the experimental findings that the Long Term Potentiation (LTP) induction is preferentially happening in smaller spines and that the timescales of CaMKII activation are similar to the timescales governing spine volume expansion.

1.2 Summary of mathematical tools and concepts

For the development of this dissertation, we used several mathematical tools and concepts. The main ones are summarised below. In the brackets we mention the subsections where they are firstly introduced and/or defined.

Mathematical tools and concepts used in deterministic analysis	Mathematic tools and concepts used in stochastic analysis
Linearisation about fixed points (Ch. 3.2)	Chemical Master Equation (Ch. 2.3.2)
Jacobian matrix (Ch. 3.2)	Gillispie Exact Algorithm (Ch. 2.3.3.2)
Nullclines (Ch. 3.1.2)	Chemical Langevin Equation (Ch. 2.3.3.3)
Thomson and Gunawardena deterministic framework (Ch. 6.6.1)	Markov chain tree theorem and Directed spanning trees (arborescences) (Ch. 3.3.2)
Root locus (Ch. 6.6.1.6)	M- and inverse M-matrices (Ch. 4.2.2)
Vieta formulae (Ch. 6.6.1.1)	Diagonally dominant matrices (Strictly, Weakly, Weakly chained) (Ch. 4.2.2)
Descartes' rule of signs (Ch. 6.6.1.1)	Detailed balance, Reversibility, Irreversibility (Ch. 3.3.3.1)
Pratt tableau (Ch. 6.6.1.3)	Matrix spectral analysis (Ch. 4.3.1)

Table 1.1 Summary of the mathematical tools and concepts used

1.3 Papers associated with dissertation

The work of this dissertation led to the publication of two papers, as summarised below. In the brackets, the relevant chapters in this dissertation are mentioned.

- Petrides A. and Vinnicombe G. Understanding the genetic toggle switch phenomena using a discrete ‘nullcline’ construct inspired by the Markov chain tree theorem. Proceedings of the 56th IEEE Conference on Decision and Control (CDC), 2017 (Chapter 3).
- Petrides, A. and Vinnicombe, G. Enzyme sequestration by the substrate: An analysis in the deterministic and stochastic domains. PLoS Computational Biology, 14(5), p.e1006107, 2018. (Chapters 4 and 6).

Chapter 2

Introduction to biochemical networks

To analyse potentially multistable biochemical networks, it is imperative that one should first understand how biochemical networks are modelled. In this introductory chapter we present some of the methods that are frequently used to analyse biochemical networks. The notion of stochastic analysis, central in this dissertation, is also introduced, contrasted to the classical deterministic analysis.

2.1 Biochemical networks

The discovery of the DNA structure [175] led to great advances in molecular biology, which have significantly enhanced our knowledge for individual cellular components, particularly regarding their biochemistry. Understanding their biochemistry, however, is not the end of the road; we are still far from full understanding and complete appreciation of the biological mechanisms' importance. Individual biochemical components do not live in isolation, hence for understanding *why* certain mechanisms arise in nature, we firstly need to understand what effects these interconnections create on a systemic, network level. These networks are found across all scales: from genes to RNAs, to transcription factors, to proteins, to enzymes [80, 5, 88, 29]. Deciphering the roles of different biochemical mechanisms in system dynamics is key to understanding experimental studies as well as guiding new experimental research.

Of course, this is easier said than done. This is, in fact, the aim of Systems biology [88]; by viewing different cellular components as parts of a greater system, the systemic behaviour is modelled instead. As models are of principal importance in these type of studies, researchers are faced in front of a two-faced problem: a) what models are accurate enough to describe the behaviour of the problem at hand? b) what methods are available to analyse the models selected?

The approach taken in this dissertation, and its emphasis that the biochemical systems must be analysed both in their corresponding deterministic and stochastic domains is a result of an effort to contribute to both of these questions. We understand that to discover potential new functions for a biochemical mechanism, we need, not only to use existing methods, but to extend them, wherever possible, or even to develop new methods.

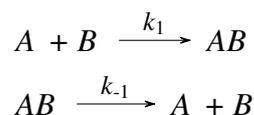
2.2 Classical deterministic analysis of biochemical networks

Biochemical networks are modelled by firstly modelling individual biochemical reactions, which are coupled by one or more shared biochemical species. In a biochemical reaction, one or more biochemical species, called ‘reactants’, react at a particular rate to produce some other biochemical species, called ‘products’.

For example consider the reaction



In this reaction species A is transformed to species B . In the classical *deterministic* formulation, the reaction specifies the *rate* at which the concentration of A , $[A]$, decreases, causing an increase in the concentration of B , $[B]$. This can now be extended to coupled biochemical reactions, forming a biochemical network [68, 29, 31]. The deterministic formulation includes reactant concentrations varying continuously in time and governed by a system of rate equations. A big advantage of this representation is that theoretical methods from linear and nonlinear dynamics and control [157] can be directly utilised. In this formulation, a molecular network is considered to be a series of elementary biochemical reactions, whose kinetics can be described by rate equations according to the *mass action law*. This law was introduced by Guldberg and Waage in the nineteenth century [65] and states that the reaction rate is proportional to the probability of a collision of the reactants. Therefore the reaction rate is not taken to be constant, yet instead it is proportional to the concentrations of the reactants, with the rate constant being the constant of proportionality. Take for example the reactions



This system of reactions can be represented by the following system of differential equations [29]:

$$\begin{aligned}\frac{d[AB]}{dt} &= k_1[A][B] - k_{-1}[AB] \\ \frac{d[A]}{dt} &= \frac{d[B]}{dt} = k_{-1}[AB] - k_1[A][B]\end{aligned}$$

2.2.1 Generalisation and Definitions

Generalising the above to n biochemical components, which can be proteins, mRNAs, chemical complexes, different states of the same protein, or proteins at different locations in a cell [29], the network can be represented with the following equation,

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{f}(\mathbf{x}, t) \quad (2.2.1)$$

where $\mathbf{x}(t) = (x_1(t), \dots, x_n(t)) \in \mathbb{R}^{+n}$ represents the concentrations of all components at time $t \in \mathbb{R}$, where \mathbb{R}^+ is the set of non-negative real numbers. \mathbf{f} represents the set of functions $f_i : \mathbb{R}^{+n} \times \mathbb{R} \rightarrow \mathbb{R}$ for $i = 1, \dots, n$.

From Eq. 2.2.1, one can see that at steady state, $\hat{\mathbf{x}} \in \mathbb{R}^{+n}$, $\mathbf{f}(\hat{\mathbf{x}}) = 0$ [157] is satisfied.

Definition 2.2.1. Steady state [90, 29, 92]: A stationary state or a steady state is a point in the phase plane, where the condition $\frac{d\mathbf{x}(t)}{dt} = 0$ is met.

Definition 2.2.2. Multistationarity [92]: The equation system $\frac{d\mathbf{x}(t)}{dt} = 0$ can have multiple solutions referring to multiple steady states. This phenomenon is called multistationarity.

Definition 2.2.3. Stable steady state [23, 157]: A steady state is (asymptotically) stable, if the system after small perturbations returns to this state for $t \rightarrow \infty$.

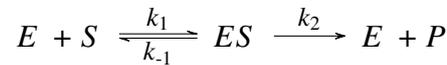
Definition 2.2.4. Bistability [29]: The equation system has two *stable* solutions of $\frac{d\mathbf{x}(t)}{dt} = 0$.

Definition 2.2.5. Multistability [29]: The system has multiple *stable* solutions of $\frac{d\mathbf{x}(t)}{dt} = 0$.

2.2.2 Exploitation of timescales: Michaelis-Menten equations

An old, yet still very effective, idea is the one that tries to approximate the system of differential equations by exploiting different timescales in fast-slow dynamics. For enzymatic reactions, a typical model is the Michaelis-Menten equation [90, 41]. It is firstly assumed that the enzyme is neither consumed or produced; therefore the total concentration of the enzyme

remains constant. An enzyme–substrate complex, ES , is formed from the free enzyme E and the substrate S which results into an irreversible release of the product P . This is presented below.



The corresponding system of ordinary differential equations is

$$\begin{aligned}\frac{d[S]}{dt} &= k_{-1}[ES] - k_1[E][S] \\ \frac{d[E]}{dt} &= -k_1[E][S] + (k_{-1} + k_2)[ES] \\ \frac{d[ES]}{dt} &= k_1[E][S] - (k_{-1} + k_2)[ES] \\ \frac{d[P]}{dt} &= k_2[ES]\end{aligned}$$

The assumption made is that the reversible reactions governing the enzyme-substrate complex ES formation are much faster than the reaction leading to a new product. This leads to the quasi-steady-state assumption that the concentration of the enzyme–substrate complex is unchanging with respect to time and is equal to

$$[ES] = \frac{[E][S]}{K_M}$$

where the Michaelis-Menten constant is

$$K_M = \frac{k_{-1} + k_2}{k_1}$$

As $[E_{tot}] = [E] + [ES]$,

$$[ES] = \frac{[E_{tot}][S]}{K_M + [S]}$$

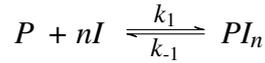
This leads to the well-known Michaelis-Menten equation:

$$\frac{d[P]}{dt} = \frac{V_{max}[S]}{K_M + [S]}$$

where $V_{max} = k_2[E_{tot}]$ is the maximal enzyme velocity. This equation describes the initial velocity of the enzyme as a function of three parameters: the substrate concentration, $[S]$, the Michaelis-Menten constant, K_M , and the maximal velocity, V_{max} . [154].

2.2.3 Cooperativity, the Hill Equation and Ultrasensitivity

Another commonly used function is the Hill equation [90, 50, 51], which is viewed as an extension of the Michaelis-Menten function to account for *cooperativity*. Consider, for example, a protein P which has n active binding sites. When n monomers I bind to P , the protein P can either bind to n molecules of I , forming PI_n , or stay non-bound. The reaction can be represented by:



and the associated ordinary differential equation for $[PI_n]$ is

$$\frac{d[PI_n]}{dt} = k_1[P][I]^n - k_{-1}[PI_n]$$

Assuming quasi-steady-state for the enzyme complex $[PI_n]$ and as $[P_{tot}] = [P] + [PI_n]$, we can derive the so-called Hill equation:

$$[PI_n] = \frac{[P_{tot}][I]^n}{K_{eq}^n + [I]^n}$$

where $K_{eq}^n = \frac{k_{-1}}{k_1}$

Parameter n is the Hill coefficient. In the example above, $[PI_n]$ is a function of $[I]$. More generally, a Hill function is an input-output relationship of the form $Output = \frac{Input^n}{K^n + Input^n}$. Related concepts are those of *sensitivity* and *ultrasensitivity*. Sensitivity refers to how much input is needed to achieve a particular level of output. Assuming that the change in input is infinitesimal, local sensitivity can be defined as $S = \frac{d \ln(Output)}{d \ln(Input)}$ [50]. Ultrasensitivity refers to an output response that is more sensitive to stimulus change than the hyperbolic Michaelis-Menten response. For simple binding interactions, the ligand concentration (input) has to change by 81-fold in order for the protein to go from 10% to 90% bound [93]. According to the ultrasensitivity definition by Goldbeter and Koshland [61], ultrasensitive responses require less than an 81-fold stimulus to drive them from 10% to 90% maximal response. An ultrasensitive response is often sigmoidal and is well-approximated by the Hill equation [50]. The difference between an ultrasensitive and a Michaelian response is illustrated in Fig. 2.2.1.

The larger the Hill coefficient n , the steeper the Hill curve [5], producing ultrasensitive responses, as illustrated by Fig. 2.2.2. Ultrasensitivity coupled with positive feedback is capable of producing switch-like behaviours [167], considered fundamental in cellular decision making. Ultrasensitivity has been mentioned to have a function in several signaling

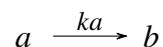
systems, including signaling cascades, bistable switches, and oscillators [19].

2.3 The need for analysis in both deterministic and stochastic domains

Despite being the most usual way of modelling biochemical networks, deterministic analysis is only exact at the limit of infinite molecule numbers [169, 57]. A cellular system is inherently noisy, with finite molecule population numbers. This results to stochasticity due to the random transitions among discrete chemical states [28]. This stochastic noise may not only affect the dynamics of biological systems but may be exploited by living organisms to actively facilitate certain functions. There are several examples in the literature where the behaviour of the system is predicted to behave differently when analysed deterministically and when analysed stochastically. These phenomena and their constructive or destructive influence on the robustness of systems has been a topic of great interest in several areas of biology, from the molecular to the ecological level [60, 94, 84, 141, 146, 161]. Furthermore, taking them into account can reveal fundamental constraints on regulation [99] and even be used in designing synthetic gene circuits [136].

In this section we provide the details of the stochastic framework, which is able to consider the exact numbers of molecules present, which are discrete quantities. In the stochastic framework, the components change discretely. The selection and the timing of the changes are probabilistic.

2.3.1 The reaction rates now represent propensities



In the stochastic framework, the scheme above now specifies the probability that the reaction which would cause a decrease of the number of molecules of A, a , by one, and the increase of the number of molecules of B, b , also by one, will occur. This probability, also provided by the mass action law, is equal to $ka dt$. As it is later to be seen it is usually easier to work with the rate of this probability instead, ka , which is referred to as the *propensity* of a reaction to take place.

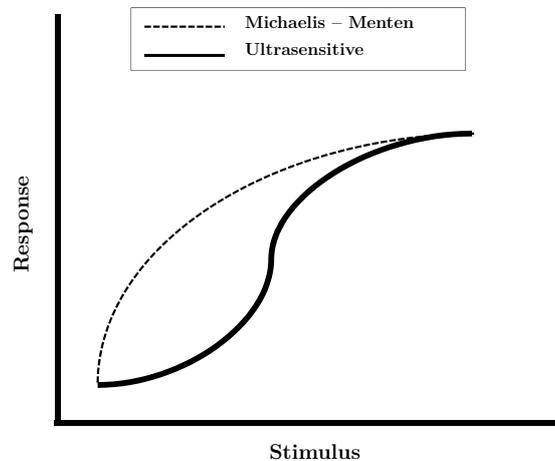


Fig. 2.2.1 Ultrasensitive vs Michaelian hyperbolic response [148]. *Reproduced with permission of Royal Society of Chemistry from 'Engineering and applications of genetic circuits, Sayut, D. et al, Molecular bioSystems by Royal Society of Chemistry (Great Britain), 3(12):835–840, ©2007' in the format Thesis/Dissertation via Copyright Clearance Center.*

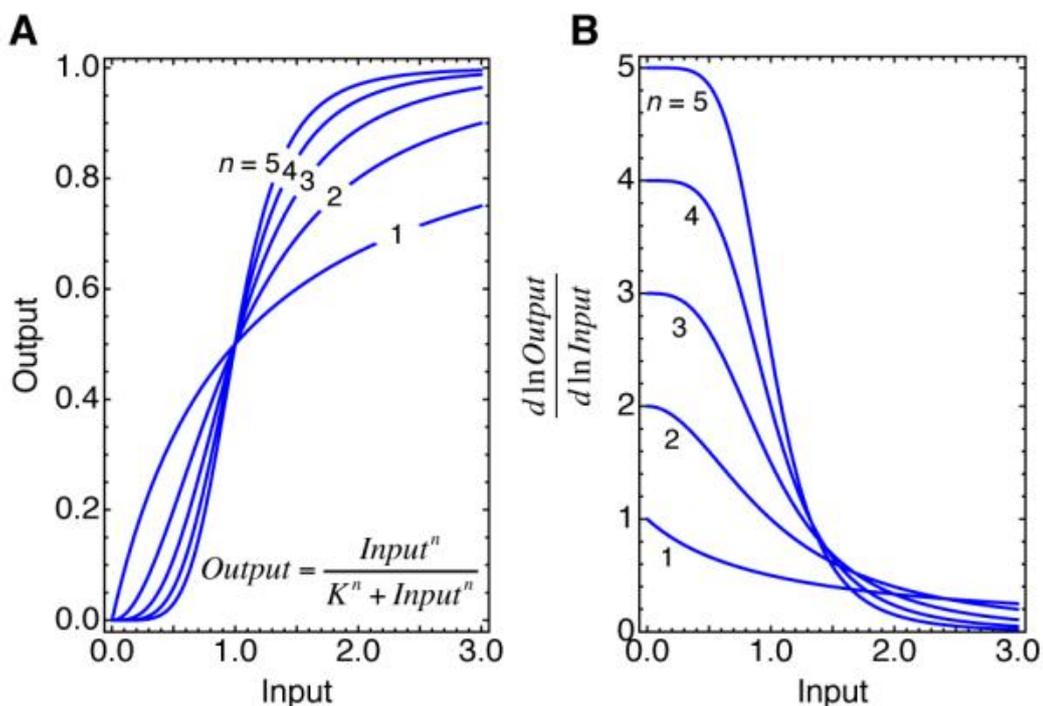


Fig. 2.2.2 (A) Input-output relationships for Hill functions with exponents ranging from one to five. When $n = 1$ the response is Michaelian. When $n > 1$, the response is ultrasensitive. The larger the Hill coefficient n , the more switch-like the response. (B) Local sensitivities for Hill curves with exponents ranging from one to five [50]. *Reproduced with permission of Elsevier Ltd. from 'Ultrasensitivity part I: Michaelian responses and zero-order ultrasensitivity, Ferrell, J. E. and Ha, S. H, Trends in biochemical sciences by International Union of Biochemistry, 39(10):496–503, ©2014' in the format Thesis/Dissertation via Copyright Clearance Center.*

2.3.2 The Chemical Master Equation

An accurate stochastic framework, where the discrete nature of molecule numbers is taken into account is the discrete Chemical Master Equation (CME). This describes the process with a discrete-state continuous-time Markov chain and is, in fact, a Forward Kolmogorov Equation [150]. It is characterised by a probability distribution as a function of time. The state of the system at time t , $\mathbf{x}(t)$, commonly also referred to as a microstate, is the set of molecule numbers of each species at a particular moment.

Formally, the state space of the Chemical Master Equation is represented by a detailed amount of each of the n molecular species in the biochemical reaction network. At time t the microstate of the system is $\mathbf{x}(t) = \{x_1(t), x_2(t), \dots, x_n(t)\} \in \mathbb{N}^n$. The overall state space is the set X of all possible combination numbers $X = \{\mathbf{x}(t) | t \in (0, \infty)\}$. $\mathbf{P}(t) \in [0, 1]^{|X|}$ is the probability landscape [108]. The positive real transition rate (*propensity*) from microstate \mathbf{x}_j to \mathbf{x}_i , r (i.e. $\mathbf{x}_j \xrightarrow{r} \mathbf{x}_i$), is represented by $A(\mathbf{x}_i, \mathbf{x}_j) = r$.

The Chemical Master Equation is illustrated in Eq. 2.3.1, where $P(\mathbf{x}, t)$ is the continuous time probability of each discrete microstate.

$$\frac{dP(\mathbf{x}, t)}{dt} = \sum_{\mathbf{x}' \neq \mathbf{x}} \left[A(\mathbf{x}, \mathbf{x}')P(\mathbf{x}', t) - A(\mathbf{x}', \mathbf{x})P(\mathbf{x}, t) \right] \quad (2.3.1)$$

The probability landscape of the CME, when the state space is finite, can also be written in matrix-vector form, where $\mathbf{A} \in \mathbb{R}^{|X| \times |X|}$ is called the matrix of propensities and is the collection of all $A(\mathbf{x}_i, \mathbf{x}_j)$.

$$\frac{d\mathbf{P}(t)}{dt} = \mathbf{A}\mathbf{P}(t) \quad (2.3.2)$$

Matrix \mathbf{A} is a zero column sum (ZCS) square matrix, as $A(\mathbf{x}, \mathbf{x}) = - \sum_{\mathbf{x}' \neq \mathbf{x}} A(\mathbf{x}', \mathbf{x})$. The zero column sum principle ensures that there is conservation of probability mass in the chain.

Two very important concepts in this dissertation are those of stationary probability distribution and multimodality. These are defined below.

Definition 2.3.1. Stationary probability distribution [67]: A stationary probability distribution \mathbf{P}_s , in the context of this dissertation, is a non-trivial (i.e. $\mathbf{P}_s \neq 0$) solution of the condition $\frac{d\mathbf{P}(t)}{dt} = 0$.

Definition 2.3.2. Mode [1, 24]: A mode is a local maximum of the stationary probability distribution \mathbf{P}_s .

Definition 2.3.3. Bimodality [24]: The phenomenon where the stationary probability distribution \mathbf{P}_s of a system has two local maxima.

Definition 2.3.4. Multimodality [24]: The phenomenon where the stationary probability distribution \mathbf{P}_s of a system has multiple local maxima.

2.3.3 The stochastic methods available

For simple problems like monomolecular reaction systems, analytical solutions can be obtained [81]. However, for most cases, the Chemical Master Equation this is not possible. The state space created by the Chemical Master Equation expands rapidly with increasing number of species and molecular copy numbers [86].

To approximate the solution, there are three approaches found in the literature.

2.3.3.1 Truncation of state space

The first approach approximates the Chemical Master Equation solution by solving a truncated version of the original Markov process [149]. For this method, the truncation must be chosen so that the number of states is large enough to retain most of the probability mass, yet small enough to be computationally efficient. The finite state projection method [123] is a variant of this method which, via using an absorbing state to represent the truncated states, can provide error guarantees. Based on the same principle of state truncation, several other variants have also been proposed in the literature. Examples include the sliding window abstraction method [73], where consecutive windows of the state space are truncated, and the finite buffer method [25], which limits the number of new molecules that can be synthesized in open systems.

2.3.3.2 Kinetic Monte Carlo - Gillispie Exact Algorithm

A second approach is via simulations using kinetic Monte Carlo. The principle behind this approach is to produce realisations, which could be either exact or approximate, of the underlying Markov process. Given enough realizations, the relevant statistics of the process can be obtained. The Gillispie Exact Algorithm [56], the most well-known of this class of algorithms, producing exact realisations, is presented in Algorithm 1 [29].

Following the notation by Gillispie [57], a general molecular network is considered, with N molecular species S_1, \dots, S_N that react through M channels R_1, \dots, R_M . $X = (X_1, \dots, X_N)$ is the state of the molecules at time t . $a_j(X)dt$ is the probability, given state X , that one R_j reaction

occurs in the next infinitesimal time interval $[t, t + dt)$ and v_{ji} is the change in the number of S_i molecules produced by one R_j reaction ($j = 1, \dots, M$ and $i = 1, \dots, N$).

Result: Output simulation trajectory

Step 1. Initialisation: set $t = 0$ and fix the initial numbers of molecules $X = X_0$

Step 2. Calculate the propensity function a_j , $j = 1, \dots, M$

Step 3. Generate two random numbers r_1 and r_2 in $[0, 1)$

Step 4. Calculate $\Delta t = \frac{1}{a_0(X)} \ln(\frac{1}{r_1})$, where $a_0(X) = \sum_{j=1}^M a_j(X)$

Step 5. Calculate $\mu =$ the smallest integer satisfying $\sum_{j'=0}^{\mu} a_{j'} > r_2 a_0(X)$

Step 6. Execute reaction μ and advance time by Δt . If t reaches T_{max} , terminate the computation. Otherwise, go to Step 2.

Algorithm 1: Gillispie Exact Algorithm

2.3.3.3 Analytical approaches - Chemical Langevin Equation

The third approach lies on the principle that it could be beneficial to use asymptotic approximations to trade accuracy for computational or analytical tractability. The Chemical Langevin Equation (CLE) [57] is a method based on this principle. Starting from the Chemical Master Equation, when the number of each species is finite yet large, the Chemical Langevin Equation (CLE) can describe well the dynamics of cellular systems [57, 29]. To illustrate how the Chemical Master Equation can be approximated by the Chemical Langevin equation, a general molecular network is considered as before. For this general network, the Chemical Master Equation is represented by [29]:

$$\frac{dP(X, t)}{dt} = \sum_{j=1}^M a_j(X - v_j)P(X - v_j, t) - a_j(X)P(X, t)$$

where $v_j = (v_{j1}, \dots, v_{jN})$.

The Chemical Langevin equation can then be derived [57, 58] taking the form

$$\frac{dX_i(t)}{dt} = \sum_{j=1}^M v_{ji} a_j(X(t)) + \sum_{j=1}^M v_{ji} \sqrt{a_j(X(t))} \Gamma_j(t)$$

where $\Gamma_j(t)$ are temporally uncorrelated, statistically independent Gaussian white noises, defined by

$$\Gamma_j(t) = \lim_{dt \rightarrow 0^+} N(0, 1/dt)$$

where $N(m, \sigma^2)$ denotes the normal random variable with mean m and variance σ^2 [58].

2.3.4 Deterministic and stochastic rate constants

The rate constants used for stochastic simulations and the rate constants used for deterministic modelling are related, yet are not always identical. For first order reactions (i.e. where there is only one reactant $X \rightarrow$), the deterministic and stochastic rate constants are equal and expressed in sec^{-1} . This is because the rates are independent of the volume the reactants are in. However, for other non-monomolecular reactions (e.g. $X + Y \rightarrow$) the deterministic rate constants are expressed in terms of the concentrations, measured in M (moles per litre) (e.g. $M^{-1}sec^{-1}$), whereas the stochastic rate constants are expressed in terms of molecules (e.g. $molecule^{-1}sec^{-1}$). Therefore, in order to convert the concentrations to number of molecules, the volume of the container, V , needs to be known. For example, a deterministic rate constant of $1M^{-1}sec^{-1}$ that has been measured in a reaction volume of $1 \times 10^{-15}L$ converts to a stochastic rate constant of $1.66 \times 10^{-9}molecule^{-1}sec^{-1}$ [90, 177].

2.4 Deterministic vs Stochastic tools: A gap that needs to be bridged

The stochastic methods reviewed, although they are great computational tools, when compared to the tools available in the deterministic framework, offer inadequate visualisation and qualitative interpretation of the system's behaviour through a systematic way without explicit computation. Such techniques are numerous in the deterministic non-linear dynamics literature and include bifurcations, phase portraits and ODE nullclines [157].

As Hiroaki Kitano mentioned in one of the most cited articles on Systems Biology [88], to obtain a system-level understanding of a biological system one should be able to design based on definite design principles instead of blind trial-and-error. Therefore, before analysing complex potentially multistable systems (which we do in Chapters 6 and 7), we first need to develop new methodological techniques to understand the underlying reasons of the discrepancies observed between deterministic and stochastic analyses.

This thesis is primarily concerned with small sized systems, where the characterisation via the Chemical Master Equation is the most appropriate [87]. For such systems, the literature primarily focuses on two themes: the first is the computational: the ability to calculate the solution of the Chemical Master Equation efficiently, either directly or by Monte Carlo methods (e.g. Gillespie Algorithm) [56]. In recent years there has been extensive progress in this direction [123, 67, 159, 85]. The second theme is concerned with its biological implications:

stochastic simulations are frequently employed either to explain experimental results that are not predictable using classical deterministic methods or to illustrate the non-agreement between deterministic solutions and exact stochastic simulations [136, 107, 64, 39, 100, 16]. In this context, we recognise the need to have new, easily computable as well as easily interpretable methods that would be able to predict when discrepancies between deterministic and stochastic analyses would occur, without the need to fully simulate or fully compute the solution of the Chemical Master Equation.

In Chapter 3 we focus on the genetic toggle switch. We choose to start with this example, as it is very well-studied example in the literature, where in the absence of cooperativity, the system is predicted to have only one stable steady state [53], whereas experimental results and exact stochastic simulations have shown that the system can exhibit bimodality instead [39, 100, 16]. This is usually referred to as ‘noise-induced’ bimodality [39]. Understanding this bimodality is hard, yet even harder is the prediction of when this would take place [16]. Ma et al [107], in their effort to explain their genetic toggle switch surprising experimental results, developed a stochastic nullcline analogue. Their nullclines, however, were meant to be explanatory and not predictive, requiring the stationary probability distribution to be firstly calculated, either directly or via kinetic Monte Carlo. As what we really aim is a *predictive* method, in Chapter 3, utilising the Markov chain tree theorem, we develop our own discrete nullcline analogue. This can be used as a heuristic method to predict, for example, when the stationary probability distribution will turn bimodal from monomodal, *without the need to directly calculate the stationary probability distribution a priori*.

The results of Chapter 3 illustrated to us that to understand the stochastic phenomena, we also need to have a method to understand the behaviour of a *single* microstate, especially when the system under investigation is not a planar system. Despite the abundance of different computational methods of calculating the *entire* stationary probability distribution of the Chemical Master Equation [67, 85], or of a general Continuous-time Markov chain [110, 156], we noted the lack of focus on the stationary probability of a single microstate, as also noted recently in the literature [21, 96, 95]. In Chapter 4 we extend existing computational methods [85, 155], having the single microstate in mind, and by using results from the M-matrix literature we find upper and lower bounds on the stationary probability of a single microstate of the system. As our aim is to develop *easily computable and easily interpretable* methods, our bounds explicitly separate the global from the local interactions of each microstate, whereas we make sure that we present the bounds in such format that there are accurate algorithms which do not depend on the condition number for any computation that needs to be done on the matrices [3, 35].

Having expanded our intuition in how the stationary distribution of an ergodic continuous-

time Markov chain is formed, we also develop a new tool for constructing non-reversible Markov chains in Chapter 5. Other than the need of having general methods that create non-reversible Markov chains [160, 165, 171, 36, 126, 158] due to their better mixing or asymptotic variance properties [17], we also noted that in Systems Biology there is also a need to have a method to create Continuous-Time Markov Chains that respect both a particular topology and a particular stationary probability distribution which are not necessarily reversible [131]. An illustrative example on how this can be done using our algorithm is also shown in Chapter 5.

Having created this toolbox, together with the existing tools found in the literature for deterministic models, we were now able to investigate the effect of having enzyme sequestration by the substrate [112] on models of multisite protein phosphorylation [51, 163], without restricting the analysis to be either only stochastic or only deterministic. Instead, we use the work of the previous chapters to carry out our analysis both in the deterministic (i.e. in the presence of large molecule numbers) as well as in the stochastic regimes (i.e. in the presence of small molecule numbers), thus obtaining fuller insights in the abilities of such biological mechanisms.

2.5 Conclusion

In this introductory chapter, we firstly show how biochemical networks are modelled deterministically using ordinary differential equations, introducing important notions like those of cooperativity and ultrasensitivity. Moreover, appreciating the finite discrete nature of molecular species in cells, the need to also analyse these biochemical networks in a stochastic framework is explained. The Chemical Master Equation is introduced, together with the relevant stochastic tools deriving from it. We also recognised the need to have stochastic tools that offer better visualisation and qualitative interpretation of the system's behaviour systematically without explicit accurate computation.

Chapter 3

Understanding the genetic toggle switch phenomena - Development of a new discrete ‘nullcline’ construct

This chapter has two aims: a) to introduce the discrepancies obtained between deterministic and stochastic analysis and b) to propose a new discrete ‘nullcline’ construct, used as a graphical way of predicting the shape of the stationary probability distribution in different parameter regimes without the need to perform steady state calculations. The system studied is the genetic toggle switch, whose stochastic phenomena are mentioned several times in the literature [107, 100, 16].

3.1 Introduction

3.1.1 The genetic toggle switch

Since its first appearance in 2000 [53], the synthetic genetic toggle switch has inspired a lot of interest among both the biological as well as the engineering community, as it provides a potential explanation of how decisions are made in biological processes. Its design includes two competing proteins, X and Y , each repressing (inhibiting) the transcription of the other. In Eq. 3.1.1 the standard symmetric birth and death reactions for the proteins (transcription factors) X and Y are presented [107, 100, 130, 53], with the difference that their values are normalised by their respective equilibrium values, \bar{x} and \bar{y} . The production of protein X is negatively regulated by protein Y , through binding of m copies of Y to the promoter of X (and vice versa). If m , the Hill coefficient, satisfies $m > 1$, then the transcription factors are said to exhibit cooperative binding, or simply ‘cooperativity’ [100, 15]. k is a constant,

associated with the inverse of the repression strength [107, 100, 16].

The traditional analysis using Ordinary Differential Equations [53] predicts two stable steady states in the cooperative case ($m > 1$) and one stable steady state in the non-cooperative case ($m = 1$). However, experimental results and exact stochastic simulations [107] have shown that when the concentrations are small, then the system with cooperative binding can also exhibit trimodality, something that is not predicted using the Chemical Langevin Equation [57].

Furthermore, by the use of Langevin equations [16], direct use of the Chemical Master Equations [100] or exact stochastic simulations [100] as well as experimentally [164], it has been shown that the non-cooperative case can also exhibit bimodality in certain cases, without this being predicted in the deterministic analysis.

$$\begin{aligned} x \xrightarrow{\frac{k^m \bar{x}}{k^m + \left(\frac{y}{\bar{y}}\right)^m}} x + 1, x \xrightarrow{\beta x} x - 1 \\ y \xrightarrow{\frac{k^m \bar{y}}{k^m + \left(\frac{x}{\bar{x}}\right)^m}} y + 1, y \xrightarrow{\beta y} y - 1 \end{aligned} \quad (3.1.1)$$

3.1.2 Stochastic nullclines in the literature and our approach

A convenient way of characterising and understanding the behaviour of planar nonlinear deterministic systems is the use of nullclines, as they provide an easy way of identifying the equilibrium points of a system [157]. For example, in the following system,

$$\begin{aligned} \frac{dx}{dt} &= f(x, y) \\ \frac{dy}{dt} &= g(x, y) \end{aligned}$$

the x -nullcline is the set of points where $f(x, y) = 0$ and y -nullcline is the set of points where $g(x, y) = 0$. The points of intersection between x -nullcline and y -nullcline are the equilibrium points. Note that along the x -nullcline the velocity vectors are vertical while along the y -nullcline the velocity vectors are horizontal. As long as we are traveling along a nullcline without crossing an equilibrium point, then the direction of the velocity vector must be the same [157].

Noting the practicality of nullclines, Ma et al tried to identify the experimentally found equilibria of the genetic toggle switch problem in the small numbers regime by developing a stochastic 'nullclines' analogue [107]. However, the stochastic 'nullclines' presented by Ma et al [107] were generated by a series of mean values from stationary conditional distributions, which meant that the stationary distribution needed to be firstly calculated (e.g. by a kinetic

Monte Carlo algorithm). As the stationary probability distribution of the system needed prior computation, that ‘nullclines’ method lacked a very important feature of deterministic ‘nullclines’: the ability to predict the behaviour of the system without simulating it.

In this chapter, we propose a new ‘nullclines’ method, based simply on the transition rates (propensities) of the system. As this method is based just on the transition rates already found in the description of the problem, the calculations required are minimal. Most importantly, though, these discrete ‘nullclines’ can provide an efficient graphical way of predicting the shape of the stationary probability distribution in different parameter regimes *without the need to calculate or simulate the steady state distribution*.

Our approach is based on the graphical representation of the Chemical Master Equation and the well-known Markov chain tree theorem and its graph theoretic representation. The Markov chain tree theorem relates the stationary probability of a node in a strongly-connected graph with the sum of the weights of the directed spanning trees that are sinked (rooted) to that node [6]. Looking at the problem directly, however, is combinatorially very challenging. Therefore, our approach is to try to extract as much information as we can from the requirements that must be satisfied for a subgraph to be a directed spanning tree.

The requirements that a subgraph is a rooted directed spanning tree are that it is acyclic, the outdegree of the rooted vertex is equal to zero and the outdegree of each other vertex is equal to one [63].

Selecting a particular directed edge when forming the rooted directed spanning tree (arborescence) places a constraint on the rest of the directed edges that can be selected in order to satisfy the requirements that the directed subgraph created is a rooted directed spanning tree. In fact, it is this constraint that makes the related directed spanning tree problems much more combinatorially difficult than the corresponding undirected spanning tree problems, as for example is the classical problem of finding the spanning tree of minimum weight [52, 62, 83].

Heuristically, for consistently forming large weight arborescences, the location of the root would be expected to allow the consistent selection of the most beneficial (i.e. the largest of the two) direction for each edge. Based on this argument, we suggest that knowing the preferential transition direction of each edge can be a useful tool that could enable inference of the final form of the stationary probability distribution.

3.2 The deterministic analysis provides an invariant to the equilibrium size Jacobian

In this section, we illustrate that the usual deterministic analysis, transforming the discrete problem firstly into a continuous time stochastic differential equation, can provide an invariant Jacobian [157] at equilibrium for our choice of normalised reaction rates. This allows us to discriminate the stochastic effects resulting from the change of equilibrium number from any deterministic phenomena.

It is standard to approximate Eq. 3.1.1 as a pair of coupled Stochastic Differential Equations (SDEs):

$$\begin{aligned} dx &= \left(\frac{k^m \bar{x}}{k^m + \left(\frac{y}{\bar{y}}\right)^m} - \beta x \right) dt + \sqrt{\frac{k^m \bar{x}}{k^m + \left(\frac{y}{\bar{y}}\right)^m} + \beta x} dW_1 \\ dy &= \left(\frac{k^m \bar{y}}{k^m + \left(\frac{x}{\bar{x}}\right)^m} - \beta y \right) dt + \sqrt{\frac{k^m \bar{y}}{k^m + \left(\frac{x}{\bar{x}}\right)^m} + \beta y} dW_2 \end{aligned} \quad (3.2.1)$$

From Eq. 3.2.1, deterministic ODEs can be obtained by setting the noise terms to zero (taking the large numbers limit) [169]. This results in

$$\begin{aligned} \dot{x} &= \frac{k^m \bar{x}}{k^m + \left(\frac{y}{\bar{y}}\right)^m} - \beta x \\ \dot{y} &= \frac{k^m \bar{y}}{k^m + \left(\frac{x}{\bar{x}}\right)^m} - \beta y \end{aligned} \quad (3.2.2)$$

An equilibrium solution of the ODEs of Eq. 6.2.4, when $\frac{k^m}{k^m+1} = \beta$, is $(x, y) = (\bar{x}, \bar{y})$. Linearising and then normalising about this point, letting $\delta x = x - \bar{x}$ and $\delta y = y - \bar{y}$ [157],

$$\begin{pmatrix} \dot{\delta x} \\ \dot{\delta y} \end{pmatrix} = \begin{pmatrix} -\beta & -\frac{\beta m}{k^m+1} \\ -\frac{\beta m}{k^m+1} & -\beta \end{pmatrix} \begin{pmatrix} \delta x \\ \delta y \end{pmatrix} \quad (3.2.3)$$

Thus the equilibrium Jacobian when there is no cooperativity ($m = 1$) is invariant and given by

$$J = \begin{pmatrix} -\beta & -\frac{\beta}{k+1} \\ -\frac{\beta}{k+1} & -\beta \end{pmatrix} \quad (3.2.4)$$

3.3 Revisiting the Chemical Master Equation and its graph theoretic interpretation

3.3.1 Chemical Master Equation

To investigate the stochastic effects due to the change of equilibrium size (e.g. small number effects), it is possible to directly approximate the steady state of the Chemical Master Equation (CME) by truncating the infinite grid into a finite one [149, 123, 159], as it was also mentioned in Chapter 2. The microstate of the system involving, in this case, two species, x and y , is defined as $\mathbf{x}(t) = \{x(t), y(t)\} \in \mathbb{N}^2$. The general discrete Chemical Master Equation (Eq. 2.3.1) can be represented explicitly for this system, where we represent the microstate probability at time t , $P(\mathbf{x}, t)$, using its x and y components, as $P_{x,y}$:

$$\begin{aligned} \frac{dP_{x,y}}{dt} = & \frac{k^m \bar{x}}{k^m + \left(\frac{y}{\bar{y}}\right)^m} P_{x-1,y} + \beta(x+1) P_{x+1,y} \\ & + \frac{k^m \bar{y}}{k^m + \left(\frac{x}{\bar{x}}\right)^m} P_{x,y-1} + \beta(y+1) P_{x,y+1} - \\ & \left(\frac{k^m \bar{x}}{k^m + \left(\frac{y}{\bar{y}}\right)^m} + \beta x + \frac{k^m \bar{y}}{k^m + \left(\frac{x}{\bar{x}}\right)^m} + \beta y \right) P_{x,y}, \end{aligned} \quad (3.3.1)$$

$$\beta = \frac{k^m}{1+k^m}$$

The equation can also be written in matrix vector form, as in Eq. 2.3.2

$$\frac{d\mathbf{P}(t)}{dt} = \mathbf{A}\mathbf{P}(t) \quad (3.3.2)$$

Thus, in order to approximate the stationary probability distribution across all microstates, \mathbf{P}_s , one can just solve Eq. 3.3.2 at equilibrium by finding the null space of \mathbf{A} , which is unique as this process is irreducible (since the associated graph is strongly connected [149]) and then normalising so that the sum of probabilities of the states adds up to one. We perform this calculation for comparison purposes, but our aim is to infer features of the stationary distribution directly from the propensities.

3.3.2 Directed spanning trees, the Markov chain tree theorem and its consequences

Before presenting the discrete nullcline construct, we first recall the notion of rooted directed spanning trees (also described as arborescences in the literature).

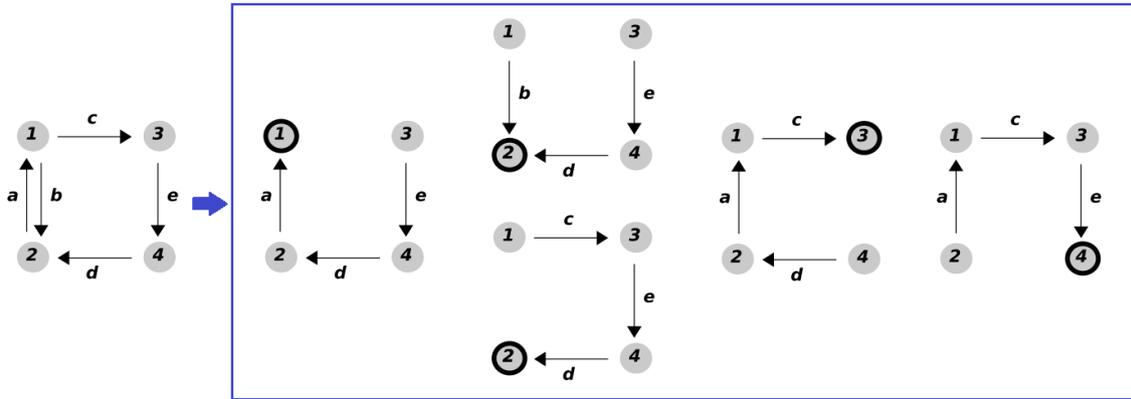


Fig. 3.3.1 Illustration of directed spanning tree formation [121]. Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer, *Bulletin of Mathematical Biology, Laplacian dynamics on general graphs*, Mirzaev, I. and Gunawardena J. ©2013.

Let $G = (V, E, w)$ be a weighted strongly connected directed graph, where $w : E \rightarrow \mathbb{R}$ is a weight function defined on its edges. A directed spanning tree rooted at $r \in V$ with orientation from the leaves to the root (i.e. the root vertex is in fact a sink) is a subgraph Q of G such that the undirected version of Q is a tree, while there is a directed path from all vertices in V to the root r [183]. This means that the outdegree (the number of edges directed away from a vertex) of all vertices in Q is equal to 1, whereas the outdegree of r is equal to 0 [63]. In Fig. 3.3.1, we present the directed spanning trees that can be formed from an example graph, by using each of the four vertices of the graph in turn as the possible root.

From Eq. 3.3.3, one can see that the null space of \mathbf{A} can be calculated by taking any one column of the adjoint matrix, as it is known that if \mathbf{A} is a zero column sum (ZCS) matrix, then $\text{adj}(\mathbf{A})$ has identical columns [44].

$$(\text{adj}(\mathbf{A}))\mathbf{A} = \mathbf{A}(\text{adj}(\mathbf{A})) = (\det \mathbf{A})I = 0 \tag{3.3.3}$$

From here, moving to the well-known Markov Chain Tree Theorem for \mathbf{A} (an $n \times n$ matrix), presented in Theorem 3.3.1, is natural. We can observe that the elements $a_{j,i}$ of \mathbf{A} correspond to the elements $w_{i,j}$ of a strongly connected directed graph, where $w_{i,j}$ represents the weight of the edge directed from vertex i to vertex j . The weight $W(T)$ of a directed spanning tree T is given by $W(T) = \prod_{\text{edge } i \rightarrow j \text{ in } T} w_{i,j} = \prod_{\text{edge } i \rightarrow j \text{ in } T} a_{j,i}$.

Theorem 3.3.1. [44, 6] *The i^{th} diagonal of $\text{adj}(\mathbf{A})$ is $(-1)^{n-1}$ times the sum of the weights over all directed spanning trees (arborescences) with sink i .*

The first consequence of this theorem is that the stationary probability of every microstate is proportional to the sum of all the directed spanning trees rooted (sunked) on the microstate, as mentioned several times in literature dealing with non-equilibrium dynamics [2, 54]. The second consequence comes from the way the directed spanning trees are formed. As the outdegree of all vertices but the root needs to be equal to one, there can be no directed spanning tree which includes both directions of the same edge (i.e. if the edge $i \rightarrow j$ belongs to the directed spanning tree, then the edge $j \rightarrow i$ does not).

The third consequence is indirect, yet is critical for constructing a graphical heuristic tool to infer the formation of the stationary probability distribution. For graphs that are balanced (as per Tutte [166] page 39, we define a *balanced* graph as one where for each vertex the number of inward edges equals the number of outward edges), such as the one obtained in this genetic toggle switch example, the number of directed spanning trees that can be formed given a particular vertex as the root is constant [166]. Therefore the stationary probability of each microstate is also proportional to the mean weight of the rooted directed spanning trees. This is particularly important, as now we can consider the expected weight of a random directed spanning tree T that can be formed given a distinguished vertex $r = j$.

$$P_s(j) \propto \mathbb{E}(W(T)|root = j) \quad (3.3.4)$$

Therefore, if the aim is to infer which roots will have large stationary probabilities, heuristically we need to search for the possible roots that are located in such positions in the graph that would allow the most beneficial directions for each edge to be consistently preferred in their corresponding random directed spanning tree.

3.3.3 Discrete ‘nullcline’ construct proposed

3.3.3.1 New type of edge-based discretisation

It is evident that in the discrete Markov process setting, the classical notion of nullclines is not applicable, as the ODEs do not capture any effects coming from the ‘discreteness’ of the system. In the graph formulation, traditional nullclines calculate the difference between the ‘birth’ and ‘death’ jump given a particular node, i.e. $f_d = w_{i,i+1} - w_{i,i-1}$ [149] when the birth and death jumps considered occur from node i to nodes $i+1$ and $i-1$ respectively. This node-based method, however, does not generalise well with the graph representation coming from the Markov Chain Tree Theorem, especially at the grid boundaries. One can clearly see that at the grid boundaries, f_d in the direction orthogonal to the boundary will always be

positive, suggesting that a 'birth' jump is the most probable operation to take place, no matter how strong the propensity of the 'death' jump towards the grid boundary is. The reason for this problem is that the calculation in discrete space depends on two edges, therefore the calculation is not well-defined at the boundaries, as there is no second edge to perform the calculation.

A more appropriate calculation, given the fact that a directed spanning tree cannot include both directions of the same edge, would be $f_s = \log \left(\frac{w_{i,i+1}}{w_{i+1,i}} \right)$, taking into consideration the birth jump from node i to node $i + 1$ and the death jump from node $i + 1$ to node i instead. That is the sign of f_s provides information about the preferential (largest) transition direction for each edge.

It is important, however, to also note that f_s does not just provide a preferential direction, but it is in itself a quantitative measure of the strength of preference of a particular direction. This is particularly important in asymmetric examples, as we will see later.

Furthermore, note that in the 1-D case as well as in higher dimensional cases where detailed balance is satisfied, knowing the f_s value for each edge is sufficient to calculate accurately the entire stationary probability distribution. This is a direct consequence of the definition of detailed balance, which means that for all nodes (vertices) i, j [54],

$$\frac{w_{i,j}}{w_{j,i}} = \frac{P_s(j)}{P_s(i)} \quad (3.3.5)$$

It is not a coincidence that in thermodynamics literature, where detailed balance is usually assumed, f_s is used to calculate the local energy difference [139].

Furthermore, this calculation now involves only one edge every time, thus it is well-defined near the grid boundaries as well.

From now on, f_s is to be called the 'net propensity' of each edge. This will be calculated by moving either across the x-axis (horizontal net propensity) or across the y-axis (vertical net propensity).

3.3.3.2 Comparison of discretisation procedures on a given example

Fig. 3.3.2 compares the results obtained using the two discretisation methods for an example 3×3 graph where the magnitude of the dominant edges is equal to 2 and the magnitude of the non-dominant edge is equal to 1. The standard node-based method

($f_d = w_{i,i+1} - w_{i,i-1}$ [149]) suggests that node 5 is the steady state of the system, as $f_d = 0$ for both the vertical and horizontal directions. This is, however, inconsistent with the actual stationary probability distribution of the system, where $P_s = [0.0455, 0.0909, 0.1818, 0.0909, 0.1818, 0.0909, 0.1818, 0.0909, 0.0455]$. This means that the standard node-based approach does not correctly capture the fact that nodes 3 and 7 are also modes of the system.

Our proposed edge-based method, on the other hand, is better suited in providing quick insights regarding the formation of directed spanning trees given a particular node as a sink. This is because it retains the information about which direction is dominant for each edge pair. This is very helpful, allowing us, for example, to see that no directed spanning tree sinked at node 5 can be formed by 8 dominant edges. This is because in a directed spanning tree all nodes must always contribute one outward edge. As such, nodes 3 and 7 can only contribute non-dominant edges. Similarly, we see that this also applies to nodes 3 and 7, yet this time the constraining nodes are the nodes 5 and 7 and the nodes 3 and 5 respectively. Furthermore, we can also deduce that if we take any random directed spanning tree sinked at nodes 3, 5 or 7, the worst case scenario is that only 4 dominant edges are selected, which is better than the worst case scenario for the other nodes of the graph. Moreover, if we start creating random directed spanning trees sinked at the aforementioned three nodes, the most likely result would be a spanning tree with 5 dominant edges. This is also better than the random directed spanning trees sinked at the other nodes. Unlike the standard node-based discretisation method, our proposed edge-based method allows us to infer that nodes 3, 5 or 7 are the most likely modes of the system.

For comparison, in Fig. 3.3.3 we enumerate all 192 directed spanning trees that can be formed according to their sink (root) node grouping them by the number of dominant edges selected for each. This confirms that the directed spanning trees sinked at nodes 3, 5 and 7 consistently select the dominant edges of the graph, in turn making them the dominant modes of the system.

3.3.3.3 ‘Nullcline’ construction

The question now is how we can compactly capture the information that some regions primarily contribute dominant or non-dominant edges to the directed spanning trees sinked at other regions of the graph. This is especially important for much larger than 3×3 systems. We note, here, that in the deterministic regime, this is indeed the role of nullclines (with respect to velocity vectors towards or away of the equilibria), yet as expected, standard

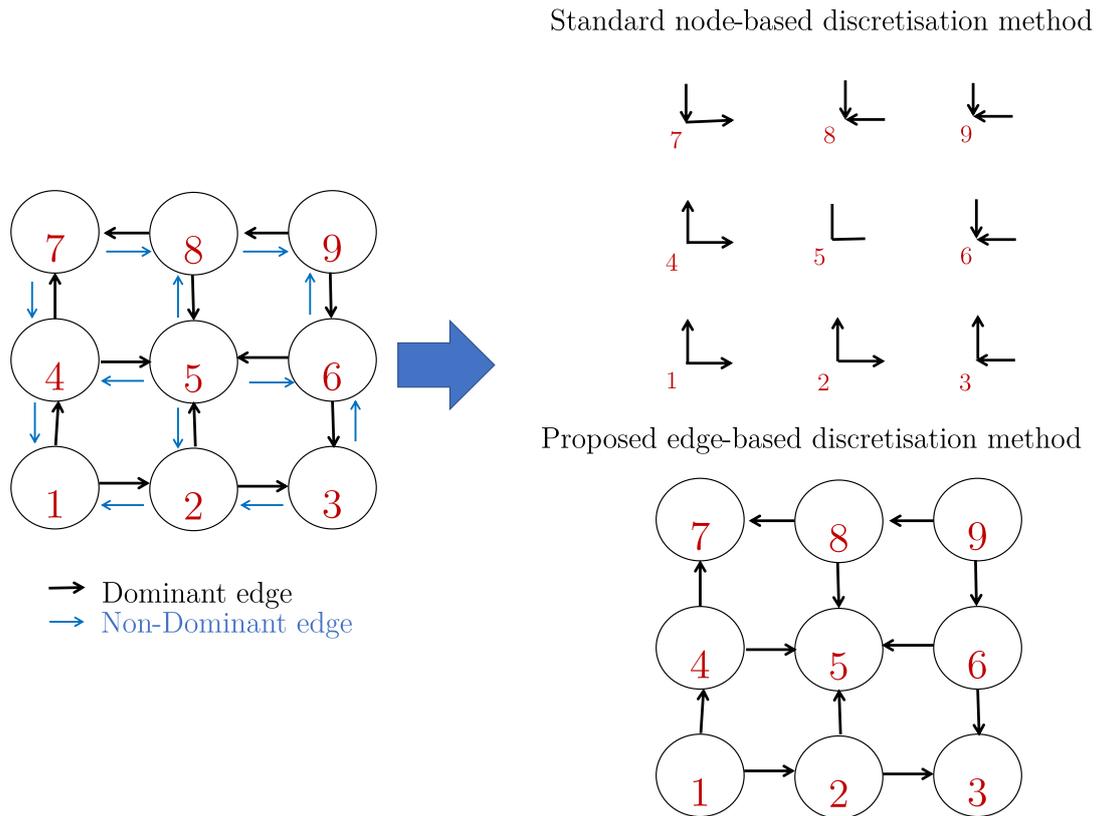


Fig. 3.3.2 Comparison between the standard node-based and the proposed edge-based discretisation methods for creating discrete nullclines. It is seen that in the standard node-based method of calculating the difference between the 'birth' and 'death' jump $f_d = w_{i,i+1} - w_{i,i-1}$, the direction orthogonal to the boundary is always positive, suggesting that a 'birth' jump is the most probable operation to take place. Also note that f_d is zero in both vertical and horizontal directions for node 5, thus suggesting that node 5 is the steady state of the system.

nullclines with standard node-based discretisation are not appropriate. From Fig. 3.3.2, we can see that the problem of node-based discretisation is that it falsely gives the perception that there is a horizontal nullcline passing between nodes 7 and 8 as well as between nodes 2 and 3. It also gives the perception that there is a vertical nullcline passing between nodes 4 and 7 and nodes 3 and 6.

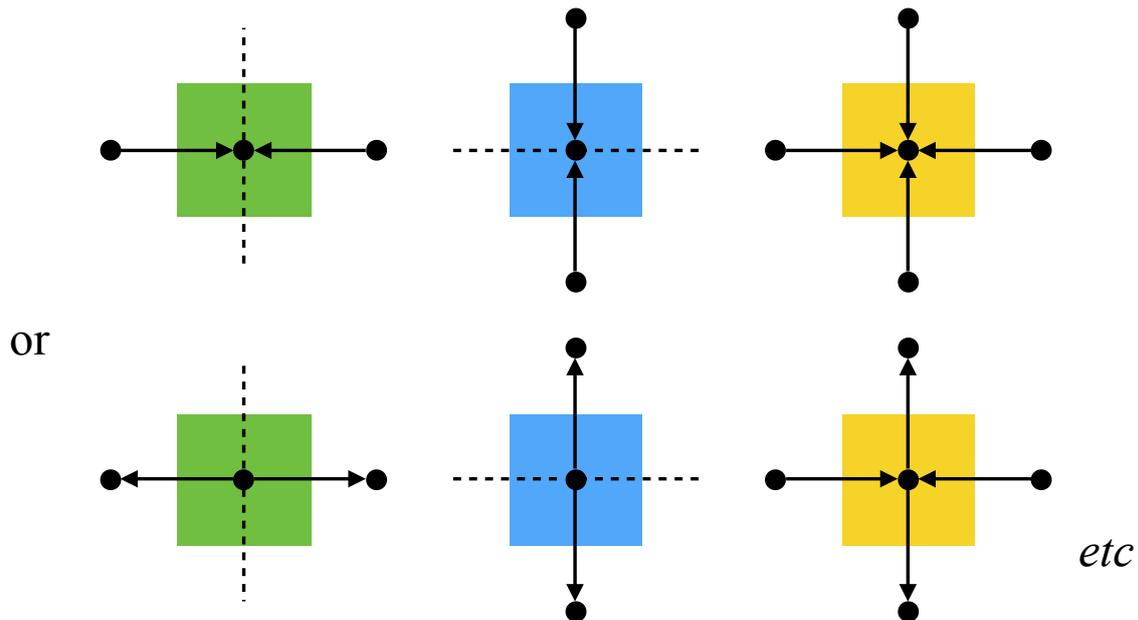


Fig. 3.3.4 Illustration of discrete 'nullcline' visualisation procedure, as that is defined in p.35. Here we present a representative sample of the cases where the conditions will be satisfied. The direction of the arrows represent the direction of the net propensity $(f_s = \log(\frac{w_{i,i+1}}{w_{i+1,i}}))$. The green squares illustrate the net horizontal propensity-reversal nodes, blue squares the net vertical propensity-reversal nodes and yellow squares both net vertical and net horizontal propensity-reversal nodes. Net propensity-reversals take place when the sign of the net propensity changes (i.e. from positive to zero or negative, zero to positive or negative and negative to zero or positive)

Therefore, a new discrete 'nullcline' analogue based on our proposed edge-based discretisation method needs to also allow for the non-existence of 'nuclines' when there is no reversion in the direction of the dominant edges. As we are also going to see later in the genetic toggle switch example, the knowledge of the non-existence of such 'nullclines' could itself be used to identify when modes are expected to be formed on the boundaries in a similar manner to modes 3 and 7 of Fig.3.3.2.

To construct these 'nullclines', we find when the direction of either the net horizontal or the net vertical propensity reverses (i.e. their sign changes from positive to zero or negative, zero

to positive or negative and negative to zero or positive).

The procedure of creating visualised ‘nullclines’ can be done as follows:

- Let $\alpha((p, q), (r, s)), p, q, r, s \in \mathbb{N}_0$ represent the propensity from state (p, q) to state (r, s) i.e. $\frac{dP_{r,s}}{dt} = \alpha((p, q), (r, s))P_{pq} + \dots$
- If $x > 0$ and $\text{sign}\left(\log \frac{\alpha((x-1, y), (x, y))}{\alpha((x, y), (x-1, y))}\right) \neq \text{sign}\left(\log \frac{\alpha((x, y), (x+1, y))}{\alpha((x+1, y), (x, y))}\right)$ then there is a horizontal net propensity reversal. For visualisation purposes, the square associated with state (x, y) is coloured green.
- If $y > 0$ and $\text{sign}\left(\log \frac{\alpha((x, y-1), (x, y))}{\alpha((x, y), (x, y-1))}\right) \neq \text{sign}\left(\log \frac{\alpha((x, y), (x, y+1))}{\alpha((x, y+1), (x, y))}\right)$ then there is a vertical net propensity reversal. For visualisation purposes, the square associated with state (x, y) is coloured blue.
- If both conditions are satisfied then there is both a horizontal and a vertical net propensity reversal. For visualisation purposes, the square associated with state (x, y) is coloured yellow.

Summarising, net horizontal propensity-reversal nodes are colored green, blue squares illustrate net vertical propensity-reversal nodes and yellow squares represent both net vertical and net horizontal propensity-reversal. Note that it is not possible for a square at the vertical boundary ($x = 0$) to be colored green (or yellow therefore). Similarly it is not possible for a square at the horizontal boundary ($y = 0$) to be colored blue (or yellow). This visualisation procedure is illustrated in Fig. 3.3.4.

As already mentioned, this ‘nullcline-like’ construct should not be confused with explanatory stochastic ‘nullclines’ explanations utilising already calculated stationary probability distributions either directly or by Monte Carlo simulations, as in [107], as the construct presented here just utilises the transition rates (propensities) coming directly from the definition of the system. There is no requirement to calculate the steady state solution first. This graphical ‘nullcline-like’ construction is aimed to allow a very quick, without calculations, inference of the stationary probability distribution as well as provide an insight for the appearance of unexpected probability modes in the stationary probability distribution.

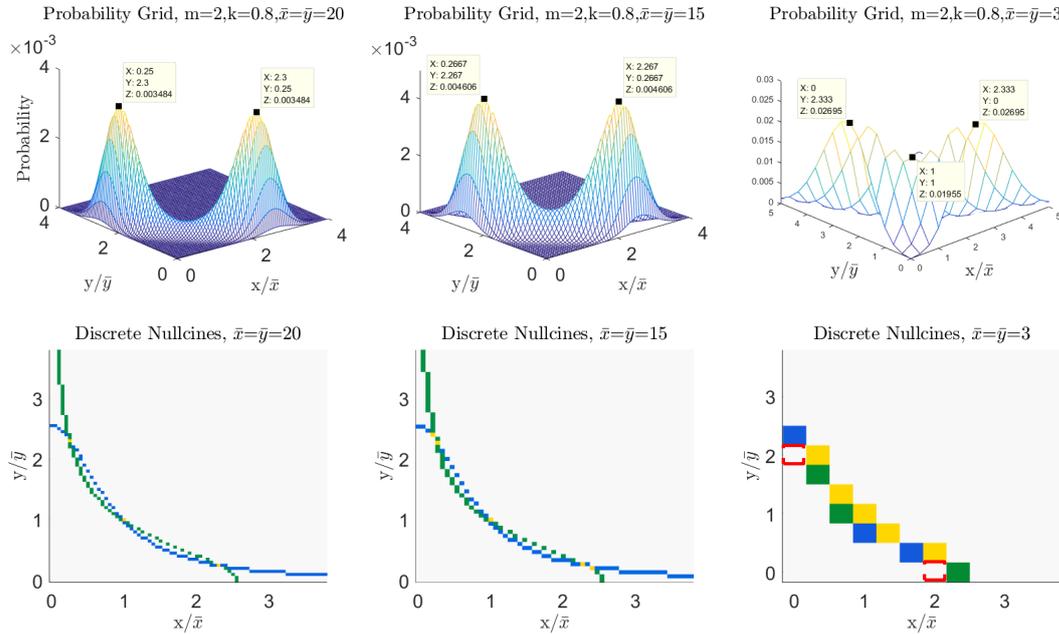


Fig. 3.3.5 The stationary probability distributions (top) with the corresponding discrete ‘nullclines’ (bottom) in the case of cooperative binding ($m=2, k=0.8$). By decreasing $\bar{x} = \bar{y}$ (20 to 15 to 3), the normalised by the equilibrium fixed point (1,1) varies from being the least dominant mode in the stationary probability distribution to a mode of almost equal probability with the initial two dominant modes. Note that there is no boundary ‘nullcline’ gap formation (like the ones dotted in red) in the first two cases ($\bar{x} = \bar{y} = 20$ and $\bar{x} = \bar{y} = 15$), and that the initial two dominant modes are not found exactly on the boundaries of the grid in contrast with the results in Fig. 3.3.6. The stationary probability distribution when the equilibrium size is small (and discretisation is coarse) (right) can be used to explain the trimodality observed experimentally in [107].

3.3.4 Applying the discrete nullclines to the genetic toggle switch example

As shown in Fig. 3.3.5, decreasing the equilibrium size $\bar{x} = \bar{y}$, the normalised by the equilibrium fixed point (1,1) varies from being the least dominant mode in the stationary probability distribution to a mode of almost equal probability with the initial two dominant modes, while other modes are appearing as well. The stationary probability distribution of the coarser discretised system (right) can also be used to explain trimodality as an effect of small numbers as this is experimentally observed in [107].

The discretisation observed, inversely proportional to the equilibrium size, has an immediate effect on the discrete ‘nullcline’ structure of the system, as illustrated in Fig. 3.3.5

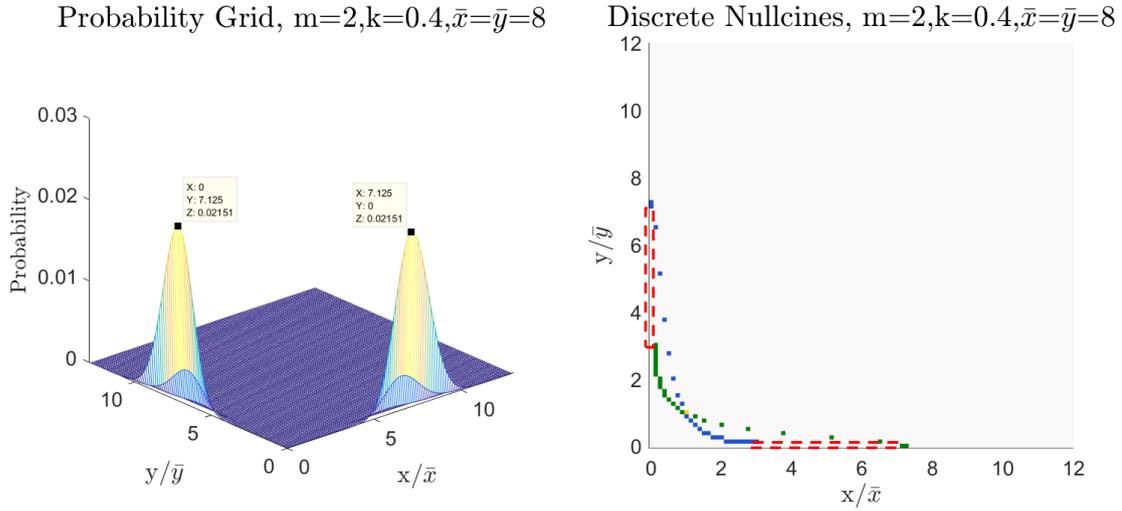


Fig. 3.3.6 $m=2, k=0.4, \bar{x} = \bar{y} = 8$. The net horizontal and net vertical propensity-reversal nodes form a ‘nullcline’ gap on each boundary which is dotted in red, where the orthogonal to the boundary net propensities are pointing towards the boundary.

(bottom). Firstly note that the discrete ‘nullclines’ when the equilibrium size is large (thus the discretisation is very fine) resemble the deterministic nullclines obtained using Ordinary Differential Equations [53]. Secondly and most importantly, note that the first two discrete ‘nullclines’ of Fig. 3.3.5 are different to the third discrete ‘nullclines’ diagram as well as to the ‘nullclines’ of Fig. 3.3.6. In the latter two cases we observe a boundary ‘nullcline’ gap formation, dotted in the figures in red. For the horizontal boundary, for instance, a ‘nullcline’ gap is formed when the first net horizontal propensity reversal node at the horizontal boundary (coloured green) is to the right of all net vertical propensity reversal nodes (coloured blue or yellow). The exact definition is provided below.

Let the x - and y -coordinates of each net horizontal propensity reversal node i be (h_x^i, h_y^i) and belong to set H . Let the x - and y -coordinates of each net vertical propensity reversal node j be (v_x^j, v_y^j) and belong to set V . Let the largest x -coordinate component found in the elements of set V be v_x^{max} and the largest y -coordinate component found in the elements of set H be h_y^{max} . Let $h_{x,0}^{min}$ be the minimum x -coordinate component of the elements $(h_x^i, 0)$ of set H and $v_{0,y}^{min}$ be the minimum y -coordinate component of the elements $(0, v_y^j)$ of set V . Then,

Definition 3.3.2. A horizontal boundary ‘nullcline’ gap is defined to be formed when $h_{x,0}^{min} > v_x^{max}$. If that is true, then the size of the gap is equal to $h_{x,0}^{min} - v_x^{max}$. Similarly, a vertical boundary ‘nullcline’ gap is defined to be formed when $v_{0,y}^{min} > h_y^{max}$. If that is true, then the size of the gap is equal to $v_{0,y}^{min} - h_y^{max}$.

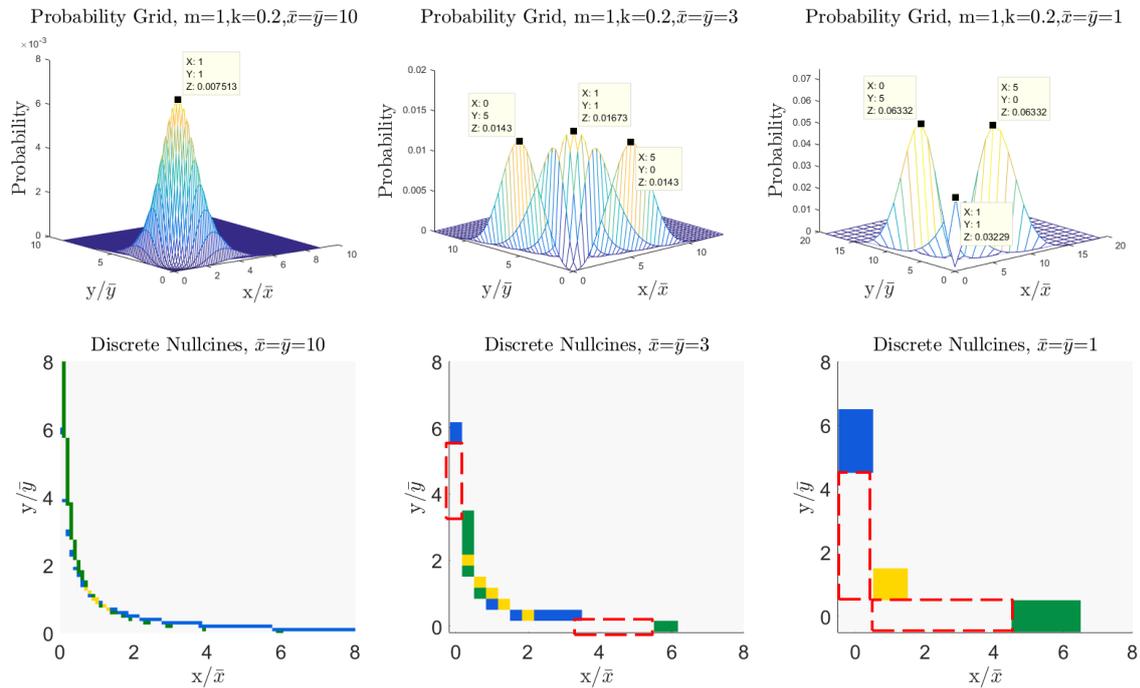


Fig. 3.3.7 The stationary distributions (top) with the corresponding discrete ‘nullclines’ (bottom) for the non-cooperative binding case ($m=1$). By decreasing $\bar{x} = \bar{y}$, the normalised by the equilibrium fixed point (1,1) (which is always coloured yellow as it is both a net vertical and a net horizontal propensity-reversal node) varies from being the dominant mode in the stationary probability distribution to the least dominant mode, while two dominant modes appear on the boundaries. At the same time boundary ‘nullcline’ gaps are formed (dotted in red) in both the horizontal and the vertical boundaries.

This essentially means that the preferential direction of the orthogonal to the boundary edge is towards the boundary (i.e. towards zero), whereas the preferential direction of the parallel to the boundary edge is towards further growth. This can be interpreted by saying that it is preferential for one species to completely vanish, while it is preferential for the other species to continue growing.

Fig. 3.3.6 greatly resembles the examples in literature where the genetic toggle switch with no cooperative binding exhibits bimodality [100].

For that reason, the same analysis was performed for the genetic toggle switch with no cooperative binding ($m=1$), the results of which are shown in Fig.3.3.7. When discretisation is very fine (left), the discrete ‘nullclines’ resemble what we get with deterministic nullclines, accompanied with monomodality in the stationary distribution. Yet, as the equilibrium size becomes smaller and therefore the discretisation becomes coarser, the discrete ‘nullclines’ reveal the gap on the grid boundaries as that observed in Fig. 3.3.6, explaining the movement

from monomodality to an intermediate multimodality and finally to essentially bimodality with the two modes always found on the boundaries, as in Fig. 3.3.6.

These results show that the discrete ‘nullclines’ proposed can capture the effects of the changes in both the equilibrium size and k , associated with the inverse of the repression strength, and provide the insight that the mechanism providing bimodality in the non-cooperative binding case can, in certain cases, be the same mechanism providing bimodality in the cooperative binding case (e.g. as in Fig. 3.3.6).

3.4 The comparison with Numerical Methods illustrates that the heuristic method can provide upper-bound estimates

The discrete ‘nullclines’ constructed can provide a good starting point for inference of potential stochastic effects, prior to any calculations. In this section we aim to compare the value of k , associated with the inverse of the repression strength, obtained through the heuristic nullcline procedure with the minimum numerically-found value of k , which guarantees that the corresponding to the equilibrium point, (\bar{x}, \bar{x}) , $\bar{x} = \bar{y}$, node in the associated graph is the global mode.

Proposition 3.4.1 provides the necessary and sufficient condition relating the equilibrium size $\bar{x}(= \bar{y})$ and the parameter associated with the inverse of the repression strength, k , for the discrete ‘nullcline’ boundary gap to be formed.

Proposition 3.4.1. *Consider the symmetric genetic toggle switch system presented in Eq. 3.1.1 with $\beta = \frac{k^m}{k^m+1}$, $\bar{x} = \bar{y} \geq 1$ and $k > 0$. Horizontal and vertical boundary ‘nullcline’ gaps are formed, as defined in Definition 3.3.2, if and only if $\lceil \frac{k^m+1}{k^m} \bar{x} - 1 \rceil > \lfloor \sqrt[m]{(k^m+1)\bar{x}^{m+1} - k^m \bar{x}^m} \rfloor$*

Proof. At point $(0,0)$ the net vertical and the net horizontal propensities are both equal to $f_s = \log\left(\frac{\bar{x}(k^m+1)}{k^m}\right)$. For $\bar{x} \geq 1$ and $k > 0$, $f_s > 0$. Looking at the horizontal boundary, the horizontal net propensity is given by $f_{sH} = \log\left(\frac{\bar{x}(k^m+1)}{(x+1)k^m}\right)$, whereas the vertical net propensity is given by $f_{sV} = \log\left(\frac{\bar{y}(k^m+1)}{k^m + (\frac{x}{\bar{x}})^m}\right)$. It is easily seen that as we increase x both terms will monotonically decrease and ultimately become negative. For a fixed x , x_0 , the vertical net propensities as y is increased are also monotonically decreasing as can be seen: $f_{sV}^y = \log\left(\frac{\bar{y}(k^m+1)}{(y+1)(k^m + (\frac{x_0}{\bar{x}})^m)}\right)$. Therefore we only need to find the points where the change of sign occurs at the boundary. Balancing the horizontal propensities of the edge on the right of $(x_H, 0)$ for $\bar{x} = \bar{y}$,

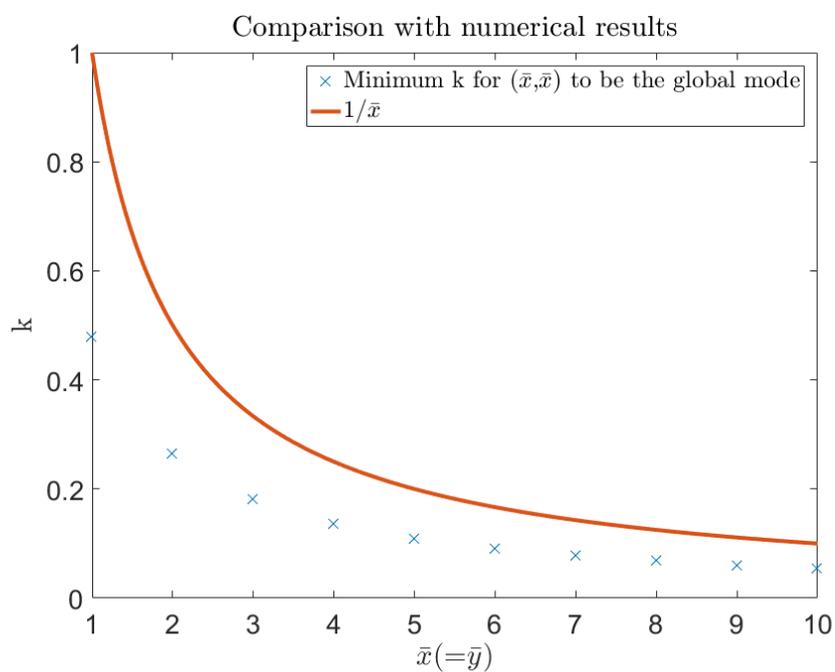


Fig. 3.4.1 In the regime of small numbers and non-cooperativity ($m = 1$), we compare, for different equilibrium solutions, the numerically-found minimum k for (\bar{x}, \bar{x}) , $\bar{x} = \bar{y}$, to be the global mode (i.e. the node with unique maximum stationary probability) against the sufficient ($k < 1/\bar{x}$) condition obtained for the formation of boundary 'nullcline' gaps.

$$\frac{k^m}{k^m + 1} \frac{x_H + 1}{\bar{x}} = 1 \Rightarrow x_H = \frac{k^m + 1}{k^m} \bar{x} - 1$$

Balancing the vertical propensities of the edge on the top of $(x_V, 0)$,

$$\begin{aligned} \frac{k^m}{k^m + 1} \frac{1}{\bar{x}} &= \frac{k^m}{k^m + \left(\frac{x_V}{\bar{x}}\right)^m} \\ \Rightarrow x_V &= \sqrt[m]{(k^m + 1)\bar{x}^{m+1} - k^m \bar{x}^m} \end{aligned}$$

Thus

$$\lceil x_H \rceil > \lfloor x_V \rfloor \Leftrightarrow \left\lceil \frac{k^m + 1}{k^m} \bar{x} - 1 \right\rceil > \left\lfloor \sqrt[m]{(k^m + 1)\bar{x}^{m+1} - k^m \bar{x}^m} \right\rfloor$$

Since $h_{x,0}^{min} = \lceil x_H \rceil$ and $v_x^{max} = \lfloor x_V \rfloor$, it follows that there exists a horizontal boundary ‘nullcline’ gap (see Definition 3.3.2) under the same condition. The same calculation applies to the vertical boundary due to the symmetry of the problem. \square

A sufficient condition for boundary ‘nullcline’ gap formations is the one presented in Proposition 3.4.1, yet without taking the floor and ceiling operations into account. For non-cooperative binding ($m = 1$), $\bar{x} = \bar{y} \geq 1$, and $k > 0$, this sufficient condition simply becomes $k < 1/\bar{x} \leq 1$.

Figure 3.4.1 illustrates that, in the regime of small numbers, the sufficient value of k for boundary ‘nullcline’ gap formations, is an upper-bound estimate of the numerically-found minimum k for the equilibrium point (\bar{x}, \bar{y}) , $\bar{x} = \bar{y}$, to be the global mode when $m = 1$. The reason the latter is calculated for integers only, is that an integer is required for the equilibrium point in the deterministic domain to coincide exactly with a vertex (microstate) in the corresponding graph depicting the Markov process.

Although k in Fig. 3.4.1 can be found numerically directly by calculating the null space of matrix \mathbf{A} for each k , we use a formulation of ours proved in Chapter 4, extending a result by Karim et al [85]. In our formulation $\mathbf{A}'q = b$, where \mathbf{A}' is the principal submatrix of \mathbf{A} after the removal of the column corresponding to the equilibrium microstate j and any row i . b is the j^{th} column of \mathbf{A} with element i deleted. $q = [q_1, q_2, q_3, \dots, q_{j-1}, q_{j+1}, q_{j+2}, \dots, q_n]^T$, $q_k = \left[\frac{P_s(k)}{P_s(j)} \right]$, represents the vector of ratios of stationary probabilities of all the nodes compared to the stationary probability of the microstate j . Then the microstate j is the unique

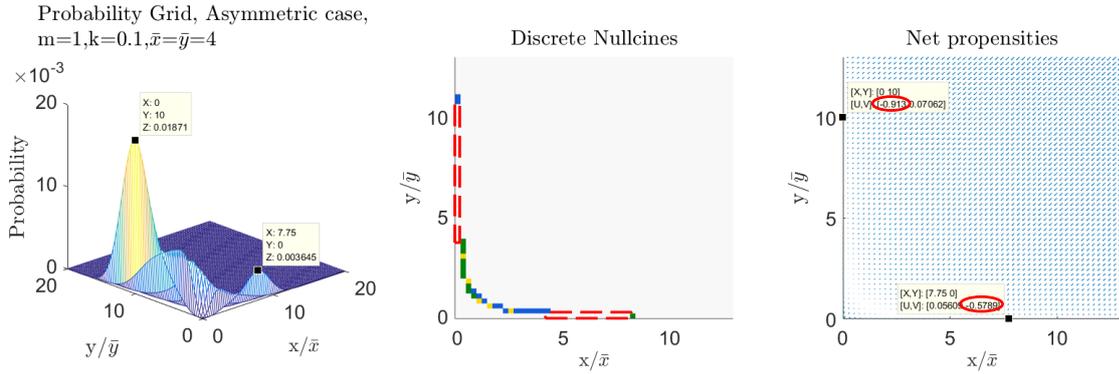


Fig. 3.5.1 The discrete 'nullclines' can be used to provide predictions for the stationary probability distributions in the asymmetric genetic toggle switch example as well. For the asymmetric case presented in Eq. 3.5.1 with $m = 1, k = 0.1, \bar{x} = \bar{y} = 4$, we also need to make a further prediction regarding which of the two boundary 'nullcline' gaps will produce the largest mode. To do so, we can use the value of the net propensities of the edges orthogonal to the boundary gaps. It is observed that in the horizontal boundary, the vertical net propensity is less negative than the horizontal net propensity at the vertical boundary. Therefore, we expect to have the largest mode on the vertical boundary, which is exactly what we observe.

global mode if and only if $\|q\|_\infty < 1$. Therefore another way to investigate this problem numerically is by investigating $\|q\|_\infty$ in the parameter space of \bar{x} and k .

3.5 The discrete 'nullcline' construct can also be used in the asymmetric case

Even though up to now we only considered the symmetric case of the toggle switch, as this is the one most commonly investigated in literature [108, 100], we would like to emphasize that the discrete 'nullclines' we propose can be used effectively in asymmetric cases as well. For example, Eq. 3.5.1 increases both the 'birth' as well as the 'death' rates associated to species x , while leaving the corresponding rates for y equal to the ones shown in Eq. 3.1.1.

$$\begin{aligned}
 x &\xrightarrow{\frac{(1.2k)^{m\bar{x}}}{(1.2k)^m + \left(\frac{y}{\bar{y}}\right)^m}} x + 1, x \xrightarrow{1.3\beta x} x - 1 \\
 y &\xrightarrow{\frac{k^m \bar{y}}{k^m + \left(\frac{x}{\bar{x}}\right)^m}} y + 1, y \xrightarrow{\beta y} y - 1
 \end{aligned} \tag{3.5.1}$$

In Fig. 3.5.1 we can see that the boundary modes in the stationary probability distribution can be predicted from the boundary gaps of the corresponding discrete 'nullclines'. As this case is asymmetric however, we need to make a further prediction regarding which of the two boundary gaps will produce the largest mode. To do so, we can use the value of the net propensities (i.e. the logarithm of the ratio of the weights of the two directions of each edge) of the edges orthogonal to the boundary gaps. It is observed that in the horizontal boundary, the vertical net propensity is less negative than the horizontal net propensity at the vertical boundary, while the size of the 'nullcline' boundary gap is also smaller. Therefore, we expect to have the largest mode on the vertical boundary, which is exactly what we observe.

3.6 Conclusion

We have proposed a new discrete 'nullcline' construct, inspired by the Markov chain tree theorem, to be used as a heuristic graphical tool for investigating stochastic phenomena, requiring minimum calculations. Its effectiveness was investigated through the genetic toggle switch example, where it was illustrated that it is effective in finding parameter regimes where different stochastic phenomena are to be expected as well as providing good inference of the stationary probability distributions to be expected both in the symmetric and asymmetric cases. Unlike other constructs [107], its aim is not to be explanatory but predictive, thus requiring no previous calculation of the steady-state distribution either through direct means [123, 149] or through Monte Carlo Simulations [107].

Chapter 4

A WCDD M-matrix reformulation of the Chemical Master Equation and an associated stochastic tool

In this chapter we develop another tool to be used in stochastic analysis, applicable to larger, non-planar systems, involving more than two chemical species. To do this, we use a Weakly Chained Diagonally Dominant M-matrix formulation of the Chemical Master Equation. The aim of this chapter is introduce mathematical notions like that of Weakly Chained Diagonally Dominant M-matrices as well as lay the mathematical foundation of some of the mathematical results utilised in the following chapters.

4.1 Introduction

The previous chapter focused on the use of the Markov Chain Tree Theorem to develop a discrete ‘nullcline’ analogue of the standard ODE nullclines starting from the Chemical Master Equation. That tool was useful in providing an insight in the way stochastic phenomena can occur in the well-studied problem of the genetic toggle switch. Nullclines, however, are primarily a useful tool for analysing smaller planar systems, where there are only two independent variables [76]. Therefore for larger, non-planar systems, as the ones later considered in Chapter 6, a different approach is required.

Furthermore, as the graph obtained in the genetic toggle switch example was ‘balanced’ (as per Tutte [166] p.39 i.e. the number of edges to and from every vertex are equal), the discrete ‘nullclines’ used the information that could be obtained by considering the local interactions between each microstate and its neighbouring microstates. Now, instead of looking at every

microstate as a local interacting agent, we could think of a microstate as being simultaneously both a sink and a source of probability for the rest of the network. The stationary probability of that microstate is consequently the global result of those interactions.

Based on that intuition, we start again from the accurate stochastic framework provided by the Chemical Master Equation, aiming to reformulate it in such a way to separate the ‘network’ input from the ‘local’ output effects.

4.2 Development of stochastic tool

4.2.1 Reformulating the equation for finding the stationary distribution of the Chemical Master Equation

In the previous chapter we saw that the stationary probability distribution \mathbf{P}_s of a finite state Chemical Master Equation can be calculated by solving the following equation

$$\mathbf{A}\mathbf{P}_s = 0 \tag{4.2.1}$$

Matrix \mathbf{A} (an $n \times n$ matrix) contains all the information required to create the discrete state continuous time Markov process governing the system. As we assume that the Markov chain is irreducible (therefore its associated graph is strongly connected), instead of finding the stationary probability distribution by computing the null space of \mathbf{A} , this can be found by solving a linear system of equations.

Karim et al. [85] suggested the formulation shown in Proposition 4.2.1, which is applicable when the rank of \mathbf{A} is equal to $(n - 1)$, as a way to translate Eq. 4.2.1 from an ill-conditioned problem into a well-conditioned system of linear equations. The stationary probability of microstate j , $P_s(j)$, is here denoted as P_s^j for compactness.

Proposition 4.2.1. [85] Equation Eq. 4.2.1 can be formulated into a well-conditioned system of linear equations in the following way:

$\mathbf{A}'q = -A_j$ where \mathbf{A}' is the matrix \mathbf{A} with column j removed, A_j is the j^{th} column of matrix \mathbf{A} and q is a column vector of size $(n - 1)$, $q = [q_1, q_2, q_3, \dots, q_{j-1}, q_{j+1}, q_{j+2}, \dots, q_n]^T$, $q_k = \begin{bmatrix} P_s^k \\ P_s^j \end{bmatrix}$.

Although this formulation makes the system well-conditioned, the fact that the new \mathbf{A}' matrix is rectangular, means that the available mathematical tools for analysis are limited, as the rectangular representation lacks the mathematical framework that is, for example, available for square matrix representations. For example one cannot compute an inverse or

find eigenvalues for a non-square matrix.

Extending that, we prove that given any $i \in \{1, 2, \dots, n\}$, Eq. 4.2.1 representing an irreducible Markov process with positive transition rates between distinct states can be further formulated into a solvable format in the following way:

$\mathbf{A}_{i,j}^D q = -A_j^i$ where $\mathbf{A}_{i,j}^D$ is the sub-matrix formed after deleting the i^{th} row and j^{th} column from matrix \mathbf{A} and A_j^i is the j^{th} column of matrix A_j as defined in the above proposition of Karim et al [85] with element i deleted. Note that in this operation any row can be deleted, due to the fact that all the simultaneous equations represented by the system are independent due to the irreducibility condition placed on the Markov process. q represents the same column vector as in the aforementioned proposition.

This means that now the $\mathbf{A}_{i,j}^D$ matrix is square. Moreover, the matrix is invertible. A direct corollary is that microstate j is the unique global mode (microstate with maximum stationary probability) of the system represented by Equation 4.2.1 (representing an irreducible Markov process with positive transition rates between distinct states) if and only if $\|(\mathbf{A}_{i,j}^D)^{-1} A_j^i\|_\infty < 1$, for any $i \in \{1, 2, \dots, n\}$.

4.2.1.1 Subsection's proofs

Fact 4.2.2. *Matrix \mathbf{A}' of Proposition 4.2.1 is a rectangular $n \times (n - 1)$ matrix with rank equal to $n - 1$.*

Lemma 4.2.3. [44] *If X is a zero row sum (ZRS) matrix, then $\text{adj}(X)$ has identical rows.*

Corollary 4.2.4. *If X is a zero column sum (ZCS) matrix, then $\text{adj}(X)$ has identical columns.*

Theorem 4.2.5. *Given any $i \in \{1, 2, \dots, n\}$, Equation 4.2.1 and Proposition 4.2.1 representing an irreducible Markov process with positive transition rates between distinct states can be further formulated into a solvable format in the following way:*

$\mathbf{A}_{i,j}^D q = -A_j^i$ where $\mathbf{A}_{i,j}^D$ is the sub-matrix formed after deleting the i^{th} row and j^{th} column from matrix \mathbf{A} and A_j^i is the j^{th} column of matrix A_j as defined in Proposition 4.2.1 with element i deleted. q represents the same column vector as in Proposition 4.2.1.

Proof. From Fact 4.2.2, Theorem 4.2.5 follows if $\det(\mathbf{A}_{i,j}^D) \neq 0$ and A_j^i is not an empty column vector. A_j must have at least two non-zero elements since by definition $A(\mathbf{x}, \mathbf{x}) = -\sum_{\mathbf{x}' \neq \mathbf{x}} A(\mathbf{x}', \mathbf{x})$ and the process is irreducible, thus strongly connected. Therefore A_j^i has at least one non-zero element. \mathbf{A} is a zero column sum (ZCS) matrix, therefore from Corollary

4.2.4, for a given j , $|\det(\mathbf{A}_{i,j}^D)| = \text{constant}$, $i \in \{1, 2, \dots, n\}$. It is also known that a column of the adjugate matrix $\text{adj}(\mathbf{A})$ is proportional to \mathbf{P}_s [11] (from $\mathbf{A} \text{adj}(\mathbf{A}) = \det(\mathbf{A}) \mathbf{I} = 0$ [138]). As the probability of any microstate found in an irreducible Markov process with positive transition rates between distinct states (which can be represented as a strongly connected graph with positive weights on the edges connecting the different states/nodes) is finite and greater than zero (this is obvious from the well-known Markov Chain Tree Theorem [44]), then for a given j , $|\det(\mathbf{A}_{i,j}^D)| = \text{constant} \neq 0$, $i \in \{1, 2, \dots, n\}$. \square

Corollary 4.2.6. *Microstate j is the unique global mode (state with maximum stationary probability) of the system represented by Equation 4.2.1 (representing an irreducible Markov process with positive transition rates between distinct states) if and only if $\|(\mathbf{A}_{i,j}^D)^{-1} \mathbf{A}_j^i\|_\infty < 1$, for any $i \in \{1, 2, \dots, n\}$.*

Proof. It follows from Theorem 4.2.5, $\mathbf{q} = -(\mathbf{A}_{i,j}^D)^{-1} \mathbf{A}_j^i = [q_1, q_2, \dots, q_{j-1}, q_{j+1}, q_{j+2}, \dots, q_n]^T$, $q_k = \left[\frac{p_s^k}{p_s^j} \right]$. For state j to be the unique global mode, $q_k = \left[\frac{p_s^k}{p_s^j} \right] < 1 \forall k \neq j \iff \|\mathbf{q}\|_\infty < 1 \iff \|(\mathbf{A}_{i,j}^D)^{-1} \mathbf{A}_j^i\|_\infty < 1$ \square

4.2.2 Weakly chained diagonally dominant (WCDD) M-matrix

Based on the formulation of the previous section, further results can be obtained with the introduction of the notion of Weakly Chained Diagonally Dominant M-matrices. For completeness, we provide the required definitions in this section.

Definition 4.2.7. [13] A square real matrix is said to be a Z-matrix if all of its off-diagonal entries are nonpositive.

Definition 4.2.8. [43, 13] An $n \times n$ matrix A is called an M-matrix if there exists an $n \times n$ nonnegative matrix B and some nonnegative real number λ such that $A = \lambda I - B$ and $\lambda \geq \rho(B)$, where I is the identity matrix and $\rho(B)$ denotes the spectral radius of matrix B (the largest absolute value of its eigenvalues); if $\lambda > \rho(B)$, A is a nonsingular M-matrix.

There are many equivalent characterisations of nonsingular M-matrices. For example, Plemmons [135] provides forty of such equivalent characterisations. Here we list four of them. If A is a Z-matrix, then the following are equivalent [135, 37]:

- A is a nonsingular M-matrix;
- The real part of each eigenvalue of A is positive;

- A is monotone; that is, $Ax \geq 0$ implies $x \geq 0$;
- A is inverse positive; that is A^{-1} exists and $A^{-1} \geq 0$.

Definition 4.2.9. [7] Row i of a complex matrix $A := (a_{ij})$ is strictly diagonally dominant (SDD) if $|a_{ii}| > \sum_{i \neq j} |a_{ij}|$. Similarly, column i of a complex matrix $A := (a_{ji})$ is strictly diagonally dominant (SDD) if $|a_{ii}| > \sum_{i \neq j} |a_{ji}|$. A is row/column strictly diagonally dominant (SDD) if all its rows/columns are SDD. Weak diagonal dominance (WDD) is defined with weak inequality instead. If not mentioned explicitly, diagonal dominance refers to weak row diagonal dominance.

Definition 4.2.10. [7] A complex square matrix A is said to be a weakly chained diagonally dominant (WCDD) if it satisfies:

1. A is WDD.
2. for each row r , there exists a path in the graph of A from r to an SDD row p .

4.3 Bounds of the stationary probability of a microstate

As we can select any element i in solving $\mathbf{A}_{i,j}^D q = -A_j^i$, we can let $i = j$. This means that we now aim to solve $\mathbf{A}_{j,j}^D q = -A_j^j$ where $\mathbf{A}_{j,j}^D$.

Selecting $i = j$ has several advantages. Firstly, Matrix $\mathbf{C}_j^T = -(\mathbf{A}_{j,j}^D)^T$ is a weakly chained diagonally dominant (WCDD) M-matrix [7]. The proof is found in Section 4.3.1. Intuitively, this means that the negated transpose of the matrix can be represented by a Markov chain, where there is a path from every microstate to reach at least one flux ‘hanging’ out of the grid, which is exactly what we observe in Fig. 4.3.1. Furthermore, this formulation separates, as initially aimed, the output propensities of a particular microstate from the rest of the parameters of the system, as also seen by Fig. 4.3.1.

Secondly, we can now bound the stationary probability of a particular microstate P_s^j using spectral properties of $\mathbf{A}_{j,j}^D$ and the magnitude of a_{jj} (the latter is simply the sum of the output propensities of microstate j). Then, defining $\lambda_i(\mathbf{C}_j)$ and $\lambda_{\min}(\mathbf{C}_j)$ to be the i^{th} and the minimum eigenvalue of matrix \mathbf{C}_j respectively and knowing that matrix \mathbf{C}_j is an $m \times m$ matrix (where $m = n - 1$), $P_s^j \geq \frac{\lambda_{\min}(\mathbf{C}_j)}{\lambda_{\min}(\mathbf{C}_j) + m|a_{jj}|}$. Moreover, defining $\sigma_{\max}(\mathbf{C}_j)$ to be the maximum singular value of matrix \mathbf{C}_j , $P_s^j \leq \frac{\sigma_{\max}(\mathbf{C}_j)}{\sigma_{\max}(\mathbf{C}_j) + \|b\|_2}$. An added benefit of the formulation presented above lies in the fact that accurate algorithms can be developed for this class of matrices [10] in computing the singular values [132, 35, 9], the smallest eigenvalue [3]

and the inverse [4]. The accuracy of these algorithms is independent of any condition number.

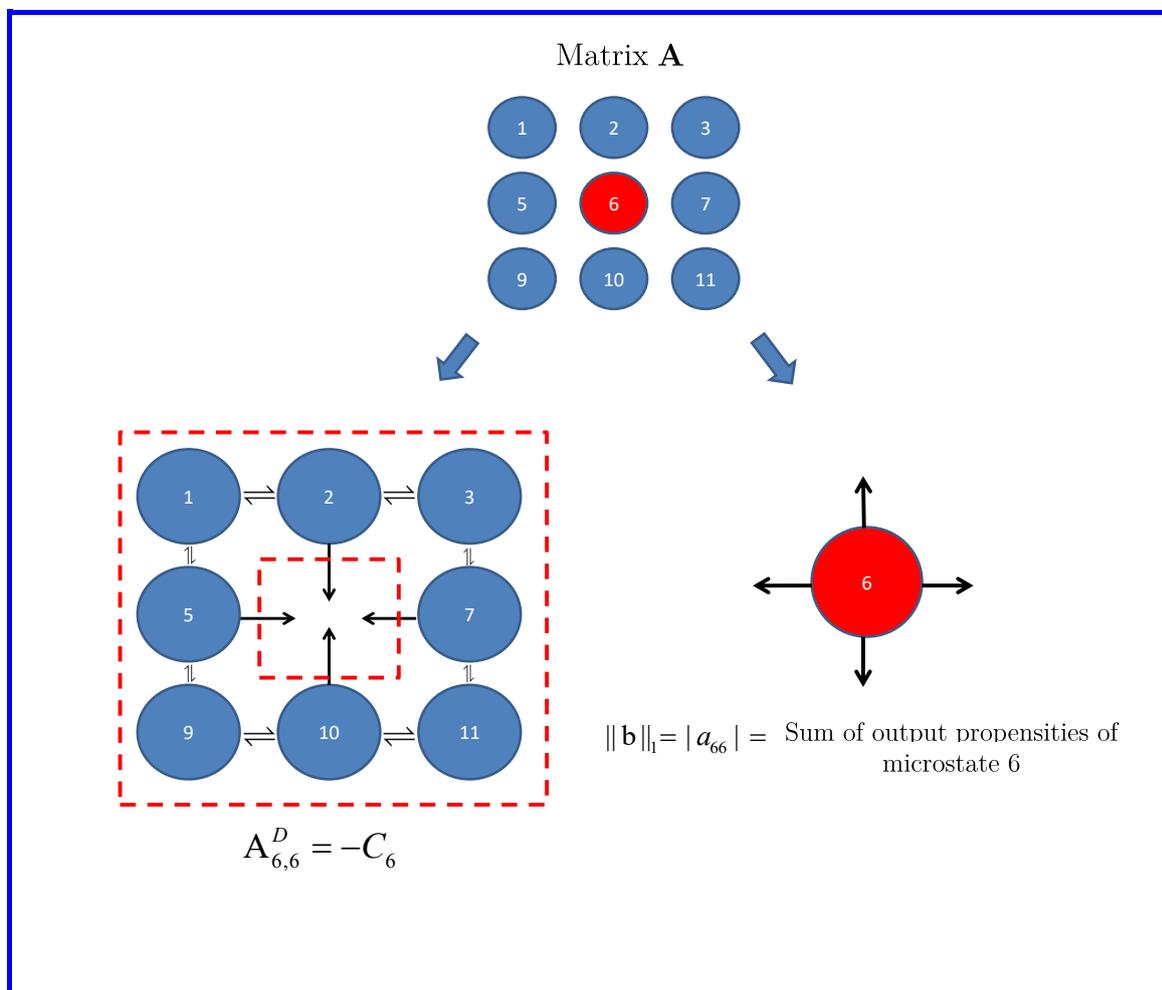


Fig. 4.3.1 The formulation of the Chemical Master Equation allows the separate investigation of the input effects of the rest of the network on the microstate under investigation from its local output propensities (a source/sink analogy)

4.3.1 Section's proofs

Here we provide the proofs of the results mentioned in this section. Note that all the matrix norms hereafter are taken to be induced norms.

Theorem 4.3.1. [7, 135] *A non-singular WDD Z-matrix with positive diagonals is an M-matrix.*

Theorem 4.3.2. [7] *The following are equivalent:*

1. *A is a WCDD Z-matrix with positive diagonals.*
2. *A is a WDD M-matrix.*

Theorem 4.3.3. *Let $\mathbf{C}_j = -\mathbf{A}_{j,j}^D$ and $b = A_j^j$, as defined in Theorem 4.2.5.*

$$\text{Then } P_s^j = \frac{1}{1+\|\mathbf{C}_j^{-1}b\|_1} \geq \frac{1}{1+|a_{jj}|\|\mathbf{C}_j^{-T}\|_\infty}$$

Proof. $\|q\|_1 = \|\mathbf{C}_j^{-1}b\|_1 \leq \|\mathbf{C}_j^{-1}\|_1 \|b\|_1 = \|\mathbf{C}_j^{-T}\|_\infty \|b\|_1$. But $\|q\|_1 = \frac{\sum_{k \neq j} P_s^k}{P_s^j} = \frac{1-P_s^j}{P_s^j}$.
Therefore, $\frac{1-P_s^j}{P_s^j} \leq \|\mathbf{C}_j^{-T}\|_\infty \|b\|_1 \Rightarrow P_s^j \geq \frac{1}{1+\|\mathbf{C}_j^{-T}\|_\infty \|b\|_1} = \frac{1}{1+|a_{jj}|\|\mathbf{C}_j^{-T}\|_\infty}$ \square

Fact 4.3.4. *Matrix $\mathbf{C}_j = -\mathbf{A}_{j,j}^D$ as defined in Theorem 4.2.5 is a square Z-matrix with positive diagonals.*

Fact 4.3.5. *Matrix $\mathbf{C}_j = -\mathbf{A}_{j,j}^D$ as defined in Theorem 4.2.5 is a column weakly diagonally dominant matrix.*

Theorem 4.3.6. *Matrix $\mathbf{C}_j^T = -(\mathbf{A}_{j,j}^D)^T$ as defined in Theorem 4.3.3 is a weakly chained diagonally dominant (WCDD) M-matrix.*

Proof. From Facts 4.3.4 and 4.3.5, \mathbf{C}_j^T is a (row) weakly diagonally dominant (WDD) Z-matrix. From the proof of Theorem 4.2.5, \mathbf{C}_j^T is nonsingular. From Theorem 4.3.1, \mathbf{C}_j^T is a WDD M-matrix. Therefore, from Theorem 4.3.2, \mathbf{C}_j^T is also weakly chained diagonally dominant (WCDD). \square

Definition 4.3.7. [34] A nonnegative matrix U satisfies the Complete Maximum Principle (CMP) if for all $x \in \mathbb{R}^n$, $x_i \geq 0$ implies that $(Ux)_i \leq 1$, then also $Ux \leq \mathbb{1}$.

Definition 4.3.8. [34] A nonnegative matrix U is said to be a *potential* if it is nonsingular and its inverse is a row diagonally dominant M-matrix.

Lemma 4.3.9. [34] *Assume U is a nonnegative nonsingular matrix. Then, U^{-1} is a row diagonally dominant M-matrix, that is U is a potential, if and only if U satisfies the Complete Maximum Principle (CMP).*

Lemma 4.3.10. [34] *Assume U is a nonnegative matrix that satisfies the CMP. Then U is column pointwise diagonally dominant. i.e. $\forall i, j |U_{jj}| \geq |U_{ij}|$.*

Theorem 4.3.11. Let $\mathbf{C}_j = -\mathbf{A}_{j,j}^D$ and $b = A_j^j$, as defined in Theorem 4.2.5. Matrix C_j is an $m \times m$ matrix, where $m = n - 1$. Then $P_s^j \geq \frac{1}{1+|a_{jj}|\text{trace}(\mathbf{C}_j^{-T})} = \frac{1}{1+|a_{jj}|\sum_i \lambda_i(\mathbf{C}_j^{-T})} = \frac{1}{1+|a_{jj}|\sum_i \frac{1}{\lambda_i(\mathbf{C}_j)}} \geq \frac{\lambda_{\min}(\mathbf{C}_j)}{\lambda_{\min}(\mathbf{C}_j)+m|a_{jj}|}$. $\lambda_i(\mathbf{C}_j)$ and $\lambda_{\min}(\mathbf{C}_j)$ denote the i^{th} and the minimum eigenvalue of matrix C_j respectively.

Proof. The inverse of a nonsingular M-matrix is nonnegative [135]. From Lemmas 4.3.9 and 4.3.10 and Theorem 4.3.6, $\|\mathbf{C}_j^{-T}\|_{\infty} \leq \text{trace}(\mathbf{C}_j^{-T}) = \sum_i \lambda_i(\mathbf{C}_j^{-T}) = \sum_i \frac{1}{\lambda_i(\mathbf{C}_j)}$. The real part of each eigenvalue of an M-matrix is positive [135]. The minimum eigenvalue of a nonsingular M-matrix is positive real [43]. From Theorem 4.3.3, $P_s^j \geq \frac{1}{1+|a_{jj}|\text{trace}(\mathbf{C}_j^{-T})} = \frac{1}{1+|a_{jj}|\sum_i \frac{1}{\lambda_i(\mathbf{C}_j)}} \geq \frac{\lambda_{\min}(\mathbf{C}_j)}{\lambda_{\min}(\mathbf{C}_j)+m|a_{jj}|}$. \square

Theorem 4.3.12. Let $\mathbf{C}_j = -\mathbf{A}_{j,j}^D$ and $b = A_j^j$, as defined in Theorem 4.2.5. Then $P_s^j \leq \frac{\sigma_{\max}(\mathbf{C}_j)}{\sigma_{\max}(\mathbf{C}_j)+\|b\|_2}$. $\sigma_{\max}(\mathbf{C}_j)$ denotes the maximum singular value of matrix \mathbf{C}_j .

Proof. $P_s^j = \frac{1}{1+\|\mathbf{C}_j^{-1}b\|_1} \leq \frac{1}{1+\|\mathbf{C}_j^{-1}b\|_2}$. As $\|\mathbf{C}_j^{-1}b\|_2 \geq \sigma_{\min}(\mathbf{C}_j^{-1})\|b\|_2 = \frac{\|b\|_2}{\sigma_{\max}(\mathbf{C}_j)}$, $\Rightarrow P_s^j \leq \frac{1}{1+\|\mathbf{C}_j^{-1}b\|_2} \leq \frac{\sigma_{\max}(\mathbf{C}_j)}{\sigma_{\max}(\mathbf{C}_j)+\|b\|_2}$. \square

4.3.2 Tightness of upper and lower bounds

Although both the upper ($P_s^j \leq \frac{\sigma_{\max}(\mathbf{C}_j)}{\sigma_{\max}(\mathbf{C}_j)+\|b\|_2}$) and lower ($P_s^j \geq \frac{\lambda_{\min}(\mathbf{C}_j)}{\lambda_{\min}(\mathbf{C}_j)+m|a_{jj}|}$) bounds depend only on the spectral properties of the rest of the network and the microstate's output propensities, they are expected to differ in tightness.

The lower bound derivation directly exploits the structure of the strongly connected Markov chain which allows the use of the Complete Maximum Principle (CMP) on $\mathbf{C}_j^{-T} = -(\mathbf{A}_{j,j}^D)^{-T}$. It is the use of CMP that allows the lower bound to be based directly on the eigenvalues of the rest of the network. If we take, for example, the system

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

then we can see that the lower bounds found are $\mathbf{P}_s \geq [1/3, 1/3, 1/3]$, which are tight as these are equal to the actual stationary probabilities of the system. On the other hand, the upper bound derivation is expected to be more conservative, as the first step of the proof ($P_s^j = \frac{1}{1+\|\mathbf{C}_j^{-1}b\|_1} \leq \frac{1}{1+\|\mathbf{C}_j^{-1}b\|_2}$) bounds the exact calculation of the microstate's stationary

probability by a more general result that applies to norms. Indeed, for the system A above, the upper bound is found to be $\mathbf{P}_s \leq [0.61803, 0.61803, 0.61803]$.

As our objective is to use these bounds for design and insights in systems biology problems, beyond checking the tightness of these, we also need to investigate how sensitive they are to changes in the stationary probabilities of the system's microstates. In Fig. 4.3.2 we use the example 3×3 system of Chapter 3 to investigate both the tightness and the sensitivity of the bounds for that system. As we can see, the upper bound is more sensitive in absolute numbers to the changes of stationary probability (top right graph), however, the lower bound is more sensitive in relative terms (bottom graph). As the bottom graph of Fig. 4.3.2 illustrates, when the bounds are normalised by their corresponding estimates of stationary probability for node (microstate) 1 (e.g. for the lower bound, we divide all the lower bounds found with the lower bound of microstate 1), the lower bound tracks more closely the behaviour of the exactly calculated stationary probability distribution.

This result, in conjunction with its tighter nature, make the lower bound a better candidate than the upper bound to be used as the basis of a heuristic tool to investigate the stochastic phenomena arising in our systems biology problems.

4.3.3 Using the bounds obtained for the development of a tool

Observing the lower bound of the individual microstate probability, $\frac{\lambda_{\min}(\mathbf{C}_j^T)}{\lambda_{\min}(\mathbf{C}_j^T) + m|a_{jj}|}$, one can use the minimum eigenvalue of \mathbf{C}_j^T (for which an accurate computational algorithm is presented in [3]) as a means to capture the information about the input effect on the microstate from the entire grid, whereas $|a_{jj}|$ is the sum of the microstate's output propensities. The latter could also provide a measure of the dwell time spent in a microstate. m is a constant, therefore the ratio $\frac{\lambda_{\min}(\mathbf{C}_j^T)}{|a_{jj}|}$ can be used as a heuristic to investigate the effect of the different propensities on the formation of the stationary probability distribution. This is utilised in Chapter 6.

4.3.4 Contributions with respect to the Markov chain literature

As the Chemical Master Equation represents a Continuous-time Markov chain, this chapter has a collection of results that apply not only on problems of Systems Biology but to general Markov chains as well. In particular, our work builds upon the literature on direct methods for numerically calculating the stationary probability distribution of an ergodic Markov chain [155].

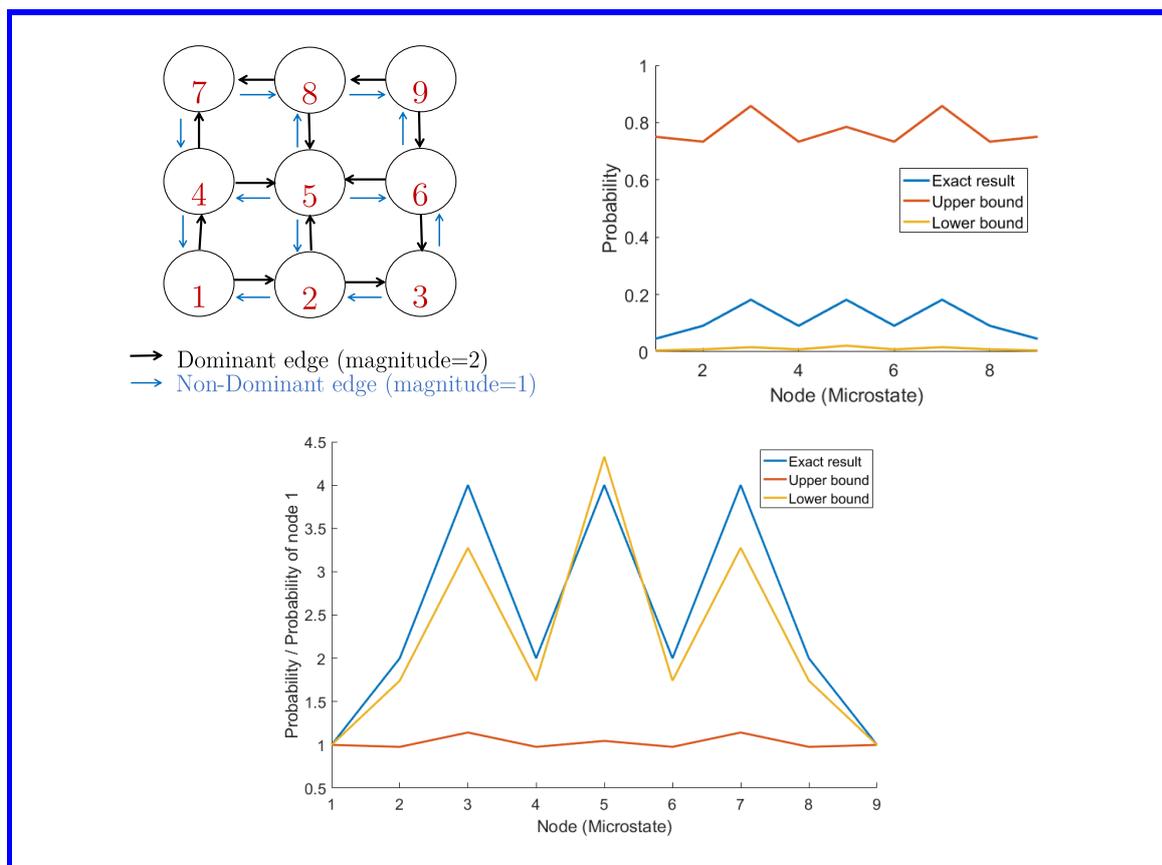


Fig. 4.3.2 Lower and upper bounds (top right) of the 3×3 system of Chapter 3 (top left). When the bounds are normalised by their corresponding estimate of probability of node (microstate) 1 (bottom graph), we can see that the lower bound is tracking more closely the behaviour of the exact stationary probability distribution.

However, unlike the general Markov chain literature [110, 156], the aim of our work is not to find another way of calculating the entire stationary probability distribution. We focus, instead, on the stationary probability of a *single* state, something that only very recently sparked interest in the general Markov chain community, and especially in the fields of Machine Learning and Markov Chain Monte Carlo [21, 96, 95]. Below we explain where our computational framework and our derived bounds from that framework stand in the general literature.

a) Theorem 4.2.5 and Corollary 4.2.6: The theorem describes a method which offers directly as a solution a vector with the ratios of the stationary probabilities of the $(n - 1)$ states of the system to the n^{th} state. The corollary is a necessary and sufficient condition based on the above theorem for a mode to be a global mode.

Theorem 4.2.5 is a slightly more general format of a technique found in the Ergodic Markov chain literature, used to handle the singularity of \mathbf{A} before Gaussian elimination can be applied [110, 91, 155]. As described in p.74 of [155] as well as in [110], the technique partitions the matrix \mathbf{A} into

$$\mathbf{A} = \begin{pmatrix} B & d \\ c^T & f \end{pmatrix}$$

where B is the $(n-1) \times (n-1)$ matrix obtained after removing the last row and last column of matrix \mathbf{A} (or in general the row and column with the same index). d is the last column of matrix \mathbf{A} with the last element removed. c^T is the last row of matrix \mathbf{A} with, f , its last element deleted.

Then, solving the equation

$$\begin{pmatrix} B & d \\ c^T & f \end{pmatrix} \begin{pmatrix} \hat{x} \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

becomes equivalent to solving the $B\hat{x} = -d$. The stationary probability vector is then found by normalising $(\hat{x}, 1)$.

The difference with our format is that unlike this technique, we do not have to constrain the row and the column to be deleted to be of the same index, as we do not have a partitioning setup in mind.

The aim of our results, however, is not to provide small extensions to existing computational frameworks [85, 110, 155, 91] for calculating the entire stationary probability distribution; instead, we aim for new compact ways for numerically identifying when a particular mode becomes a global mode, something that we obtain by our extended framework and especially by its associated corollary, finding use in our work of Chapter 3 and in our corresponding paper [133].

b) The upper $(P_s^j \leq \frac{\sigma_{\max}(\mathbf{C}_j)}{\sigma_{\max}(\mathbf{C}_j) + \|b\|_2})$ and lower $(P_s^j \geq \frac{\lambda_{\min}(\mathbf{C}_j)}{\lambda_{\min}(\mathbf{C}_j) + m|a_{jj}|})$ bounds of the stationary distribution of specific microstates: Unlike the vast literature on techniques of calculating or estimating the entire stationary probability distribution, the literature on *single* states is quite new [21, 96, 95], driven usually by the interest in computing the network centrality of a specific agent, or a subset of agents in a network [95, 21]. However, as that work focuses on Monte Carlo estimation algorithms, their authors' aim was not to find explicit lower or upper bounds on the single state's stationary probability. Bressan et al in a very recent publication [21] note the lack of general lower bounds for the stationary probability of a single state, something that could be useful in estimating the number of repeated sampling required to estimate the stationary probability of a particular state. Our results provide such bounds with the advantage that due to the WCDD M-matrix structure

of the matrices involved, both the lower and upper bound can be calculated by accurate algorithms that are independent of any condition number [3, 35]. Most importantly for this thesis, however, with these bounds we are able to develop new heuristic tools that allow us to decouple the effect of the ‘network inputs’ from the effect that the ‘local outputs’ of a microstate have on the stationary probability distribution. This provides us with insights in problems of Systems Biology, as shown in Chapter 6 and our corresponding paper [134].

4.4 Conclusion

In this chapter, we develop a stochastic tool based on the principle that a microstate can be seen as being simultaneously both as a sink and a source of probability for the rest of the network. In this way, the effects resulting from the ‘network inputs’ and those resulting from the ‘local outputs’ of a microstate are separated. The development of the tool was based on an extension of a result by Karim et al [85] for the Chemical Master Equation, allowing the stationary probability distribution of an irreducible Markov process to be found as a solution of a system of linear equations. By introducing the notion of a Weakly Chained Diagonally Dominant M-matrix, bounds of the stationary probability of a microstate using the spectral properties of the rest of the network (‘network inputs’) and the sum of the output propensities of the microstate could be found.

Chapter 5

An algorithm for constructing not necessarily reversible Markov chains with specified discrete-state stationary probability distributions

The aim of this chapter is to use the theory developed in the previous chapter to enhance our understanding of how stationary probability distributions are formed. This allowed us to develop a new algorithm for constructing Markov chains on finite discrete spaces, both discrete- and continuous-time, with specified discrete-state stationary probability distributions without requiring to satisfy detailed balance.

5.1 Introduction

Since the introduction of the Metropolis algorithm [116, 30] in 1953, Markov Chain Monte Carlo (MCMC) methods have had a profound role in fields such as statistical mechanics [36, 71] and Bayesian inference [18, 124], as they allowed the computation of expectations of high-dimensional probability distributions, especially when direct sampling is infeasible. Markov Chain Monte Carlo methods are based on the idea that an irreducible Markov chain can be constructed, which will have as its unique stationary distribution a desired probability distribution [71]. Then sampling from the probability distribution can be done by simply simulating the constructed Markov chain.

As we saw in previous chapters, understanding how stationary probability distributions are formed is also fundamental in understanding intriguing stochastic biological phenomena

[149, 100].

The most common and familiar algorithms of constructing Markov chains with a desired invariant probability distribution, as for example is the classical Metropolis-Hastings, are based on the principle of *reversibility* [129], which means that there is a strong local condition on the transition rates (also called propensities). The condition is called *detailed balance*, defined in Chapter 3, in Eq. 3.3.5. Being reversible and thus respecting detailed balance is a sufficient, yet not necessary, condition for an irreducible Markov chain to have a unique stationary distribution [172]. On this note, non-reversible Markov chains can have better mixing or asymptotic variance properties [17]. This has been shown both experimentally [160, 165, 171] and theoretically for specific circumstances [36, 126, 158].

In those examples, non-reversible chains are constructed based on reversible chains, either ‘lifting’ them to an expanded state space which allows the addition of a velocity component, aiming to provide momentum through the state space or by introducing ‘vorticity’ [17, 158] directly without the need to expand the state space. In both cases, the construction is based on an underlying reversible chain. In this chapter, we introduce a method, which can create directly finite state Markov chains, both reversible and non-reversible, without the need to consider reversibility at all, as the basis of the algorithm is concerned only with the necessary and sufficient condition of global and not detailed balance. Given the invariant target distribution, this algorithm firstly constructs a continuous-time discrete-space Markov chain which can then be also converted to a discrete-time discrete-space Markov chain.

The algorithm not only provides us with non-reversible Markov chains, without needing to start with a reversible one first unlike other algorithms, but it has the extra advantage that, based on a discrete space of n states, the transition rates and topology relating the $n - 1$ states can be pre-specified. This means that the transitions to and from the last n^{th} state with regards to the rest of the graph can be used to control the stationary probability distribution formation. This decoupled construction of the Markov chain means that our method can also be employed in fields like Systems Biology, where understanding of the possible effects on the stationary probability distribution that can be obtained from a single state and its interaction with the other states is of great interest [2].

We first revisit the Weakly Chained Diagonally Dominant (WCDD) M-matrix formulation of the Chemical Master Equation, presented in Chapter 4. Then, we illustrate that the matrix representing the transitions between the $n - 1$ vertices is required to be a non-singular WCDD M-Matrix.

Having created the transition matrix relating the $n - 1$ states to each other, we then control the formation of the stationary probability distribution by the last left n^{th} state.

Lastly, a standard uniformisation procedure is followed to convert the Continuous-time

Markov chain into a Discrete-time one. Finally, an example is provided to illustrate the method.

5.2 Revisiting the WCDD M-matrix formulation of the CME and the development of the algorithm

The development of the algorithm is based on the WCDD M-matrix formulation of the Chemical Master Equation presented in the previous chapter. For this purpose it is important to remember what the formulation of Theorem 4.2.5 allows us to do when $i = j$. q in a system of n states is the $(n - 1) \times 1$ vector with the ratios of the stationary probability of the $n - 1$ states when compared to the last distinguished n^{th} state with index j . A_j^j represents the $(n - 1) \times 1$ vector of outward transition rates (propensities) from the state of index j to all the other $n - 1$ states of the system. $\mathbf{A}_{j,j}^D$ is now the submatrix created when the row and the column corresponding to the state with index j are deleted.

The importance of this formulation here is that given a desired unnormalised stationary probability distribution \mathbf{P}_s (which only has entries greater than zero and from which q can easily be obtained) and deciding which of the states is going to be the distinguished one j , the problem of creating matrix \mathbf{A} (and so the corresponding continuous-time Markov chain) is distilled down to firstly creating a suitable $\mathbf{A}_{j,j}^D$ and carrying out a matrix-vector multiplication in obtaining A_j^j . From there, the obtained negative propensities out of state j can be converted to positive propensities into j from the corresponding states through the use of their ratios of stationary probabilities.

The first step therefore is to understand the required properties of matrix $\mathbf{A}_{j,j}^D$.

5.2.1 The first part of the algorithm is concerned with the creation of an irreducible weakly chained diagonally dominant M-matrix

We have proved in the previous chapter that Matrix $\mathbf{C}_j^T = -(\mathbf{A}_{j,j}^D)^T$ is a weakly chained row diagonally dominant (WCDD) M-matrix. Consequently, $\mathbf{M} = -(\mathbf{A}_{j,j}^D)$ is a weakly chained column diagonally dominant M-matrix. Therefore what we need is a way to create weakly chained column diagonally dominant M-matrices.

A negated weakly chained diagonally (column) dominant M-matrix in the context of continuous-time Markov chains corresponds to a generator matrix \mathbf{A} of a Markov chain with some flux from at least one of its nodes ‘hanging’ outwards from the chain, not connected to anything. A negated strictly diagonally dominant (column) M-matrix in the same context corresponds to a generator matrix \mathbf{A} of a Markov chain with fluxes coming out of all the

nodes ‘hanging’ outwards from the chain. A graphical representation is provided in Fig. 5.2.1 for better visualisation.

Therefore, one can create the matrix $\mathbf{M} = -(\mathbf{A}'_{j,j})$ by creating a Markov chain on $n - 1$ nodes with generator matrix \mathbf{A}' . \mathbf{A}' now has positive off-diagonals, where the diagonal elements are equal to the negated sum of each column, thus satisfying the Zero Column Sum (ZCS) property. Then in at least one column, a negative value is added to its negative diagonal element, thus making the column strictly diagonally dominant. This corresponds to the addition of flux directed from at least one node, ‘hanging’ out of the chain. Although not strictly required at this stage, ensuring that \mathbf{A}' is also itself irreducible (i.e. strongly connected) is important as it simplifies the algorithm for creating the final irreducible Markov chain without relying on the added distinguished state to have bidirectional transitions to all the nodes of the grid. Note that this construction means that the graph representing the first $n - 1$ nodes can have an arbitrary topology, as long as it is strongly connected. \mathbf{M} can now be obtained simply by $\mathbf{M} = -\mathbf{A}'$.

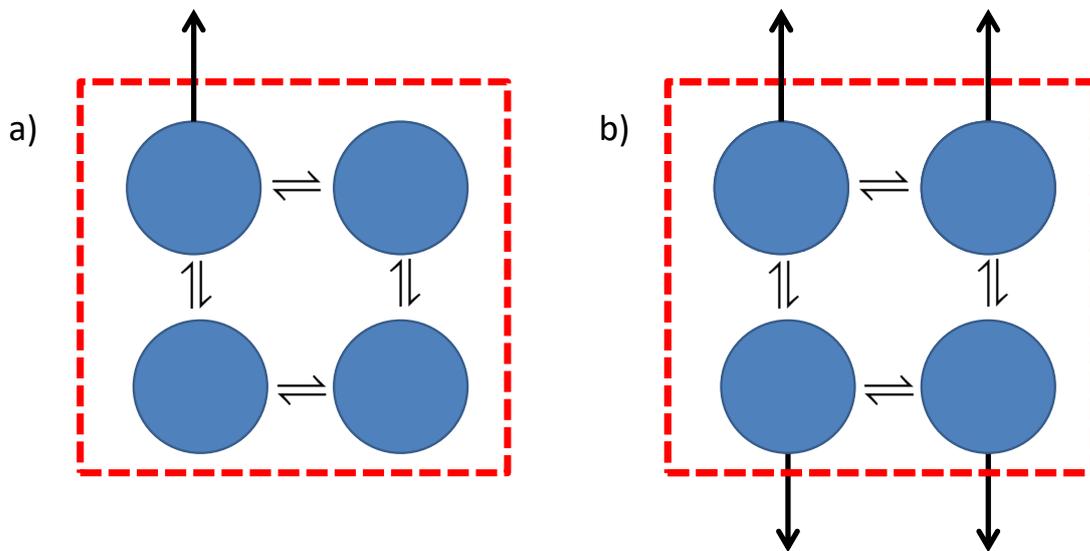


Fig. 5.2.1 a) A negated weakly chained diagonally (column) dominant M-matrix corresponds to a Markov chain with flux directed from at least one node ‘hanging’ outwards from the chain. b) A negated strictly diagonally (column) dominant M-matrix corresponds to a Markov chain with flux directed from all nodes ‘hanging’ outwards from the chain.

Matrix \mathbf{M} can also be generically made by exploiting theory regarding inverse M-matrices. Noting that strictly diagonally dominant matrices are a subset of weakly chained diagonally dominant matrices, an old result on inverse M-matrices [179] provides us with exactly what we want.

Theorem 5.2.1. [179] Assume a positive $n \times n$ matrix $A = (a_{ij})$ with unit diagonal elements and off-diagonal elements which satisfy $0 < y \leq a_{ij} \leq x < 1$. Also define an interpolation parameter s via $x^2 = sy + (1-s)y^2$. $A^{-1} = M$ is a strictly diagonally dominant (both by rows and by columns) M-matrix if $0 < y < x < 1$ and either:

1. $n = 2$ or
2. $n > 3$ and $y < x$ and $s^{-1} \geq n - 2$.

This theorem is particularly useful, since $A > 0$, $A^{-1} = M$ is irreducible as well [34].

5.2.2 The second part of the algorithm is equivalent to controlling the stationary probability distribution formation from a distinguished state

We have so far created matrix $\mathbf{M} = -(\mathbf{A}_{j,j}^D)$, which contains the information about the transition rates between the $n - 1$ states relationship either directly (if for example a particular structure regarding the Markov chain interconnections needs to be enforced) or by using inverse M-matrices. But how can we retrieve matrix \mathbf{A} , given a desired stationary probability distribution \mathbf{P}_s ? \mathbf{M} is a column diagonally dominant matrix; the states corresponding to each column whose sum is not zero are those states which have a transition rate ‘hanging’ outwards, which can now be directed into the distinguished state j . Those transition rates can be retrieved by just using the fact that \mathbf{A} must be a Zero Column Sum matrix.

As the transition rates between the $n - 1$ states are established, the stationary probability distribution needs to be formed by manipulating the input and outward propensities of the distinguished state j . Having created q , as in Proposition 4.2.1, from the vector of the desired (unnormalised) stationary probability distribution, the following multiplication can be done, which gives the outward transition rates (propensities) of state j , $d = \mathbf{M}q$. Vector d , however, obtained in this way, can have both positive and negative entries. A negative entry in d represents a negative outward transition rate from state j to the corresponding state. Without loss of generality, if the distinguished node is selected to be $j = 1$, then the k^{th} entry in d corresponds to the transition rate from state 1 to state $(k + 1)$. As we are interested in the system when found at equilibrium, the negative outward transition rate $d(k)$ from state 1 to state $(k + 1)$ can be converted to positive input transition rate b from state $(k + 1)$ to state 1 by equalising the probability flows: $-d(k)P_s(1) = bP_s(k + 1)$, which simplifies to $b = \frac{-d(k)P_s(1)}{P_s(k+1)} = \frac{-d(k)}{q(k)}$.

These can now be added to the k^{th} diagonal entry in \mathbf{M} . This would mean that the outward

transition rates from the distinguished state to those states would become zero, so if bidirectionality of those connections is aimed to be maintained, a small random value can also optionally be added to the k^{th} diagonal entry in \mathbf{M} . Bidirectionality of all the connections from the distinguished to the other states is not a requirement to obtain irreducibility of the final constructed Markov chain, as the initial Markov chain constructed on the $n - 1$ states is itself irreducible. The product of $\hat{d} = \mathbf{M}q$ now gives the outward transition rates of state 1, which are positive. \mathbf{A} can now be reconstructed from $\mathbf{A}_{1,1}^D = -\mathbf{M}$, $A_1^1 = \hat{d}$ and the fact that \mathbf{A} needs to be a Zero Column Sum matrix.

This procedure is summarised in Fig. 5.2.2 and in Algorithm 2.

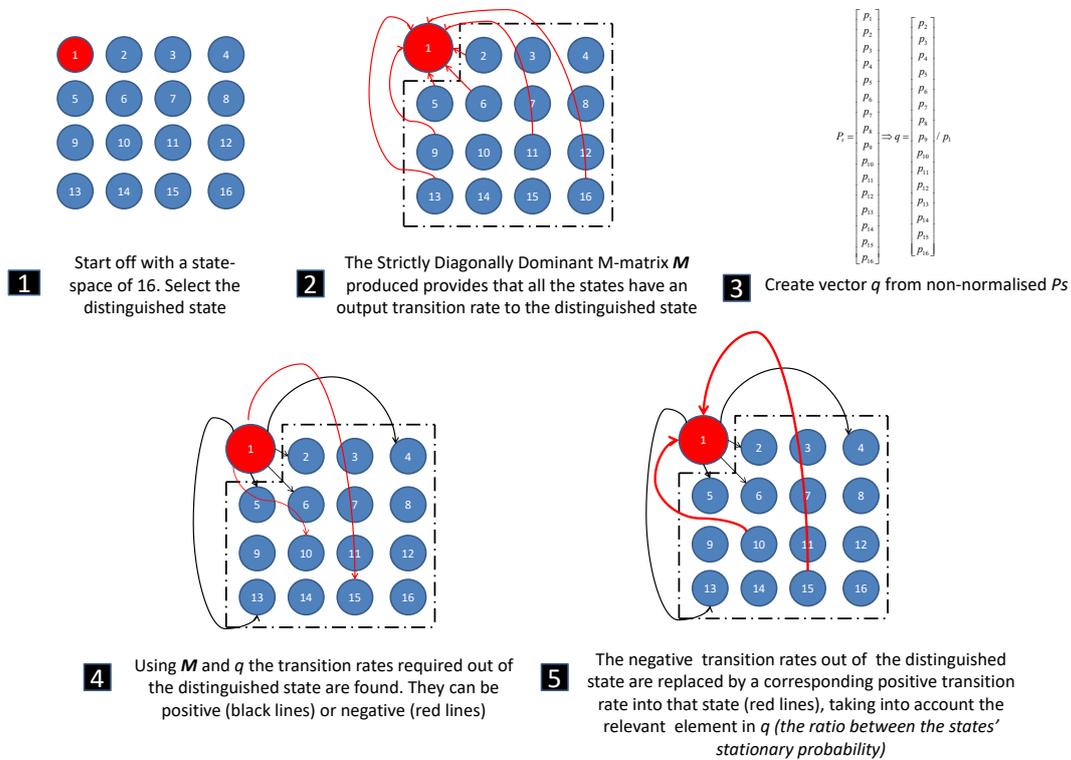


Fig. 5.2.2 Graphical illustration of the methodology followed.

5.2.3 The resulting continuous time Markov chain can be converted to a discrete time Markov chain

The resulting finite space continuous time Markov chain created can now be converted to a discrete time Markov chain by using the technique of uniformisation [82], as illustrated in

Result: Output matrix \mathbf{A} - representing a Continuous time Markov chain with distinguished state 1

1. Input desired unnormalised probability vector \mathbf{P}_s ;
2. Construct the $(n-1) \times 1$ vector $q = [p(2)p(3)\dots p(n)]^T / p(1)$;
3. Construct an irreducible $(n-1) \times (n-1)$ WCDD M-matrix \mathbf{M} either directly or by following Theorem 5.2.1 ;
4. Calculate $d = \mathbf{M}q$
5. **for** $k = 1 : n - 1$ **do**
 - | **if** $d(k) < 0$ **then**
 - | | $\mathbf{M}(k, k) = \mathbf{M}(k, k) - \frac{d(k)}{q(k)} (+rand)$
 - | **end**
- end**
6. $\hat{d} = \mathbf{M}q$;
7. $\mathbf{A}(2 : n, 2 : n) = -\mathbf{M}$;
8. $\mathbf{A}(2 : n, 1) = \hat{d}$;
9. **for** $k = 1 : n$ **do**
 - | $\mathbf{A}(1, k) = -\text{Sum}(\mathbf{A}(:, k))$;
- end**

Algorithm 2: Continuous-time Discrete Space Markov chain

Result: Output matrix \mathbf{R} - representing a Discrete time Markov chain with distinguished state 1

1. Run Algorithm 2 and obtain \mathbf{A} ;
2. Select $q > \max(\text{abs}(\mathbf{A}))$;
3. $\mathbf{R} = \mathbf{A}^T / q + \mathbf{I}$;

Algorithm 3: Discrete-time Discrete Space Markov chain

Algorithm 3. For a discrete time discrete space Markov chain, the stationary distribution \mathbf{P}_s and the matrix of transition rates (where now $R(i, j)$ represents the transition rate from state i to state j) obey the following relation:

$$\mathbf{P}_s \mathbf{R} = \mathbf{P}_s \quad (5.2.1)$$

5.3 Example - Graph and Spectral Structure of resulting Markov chain

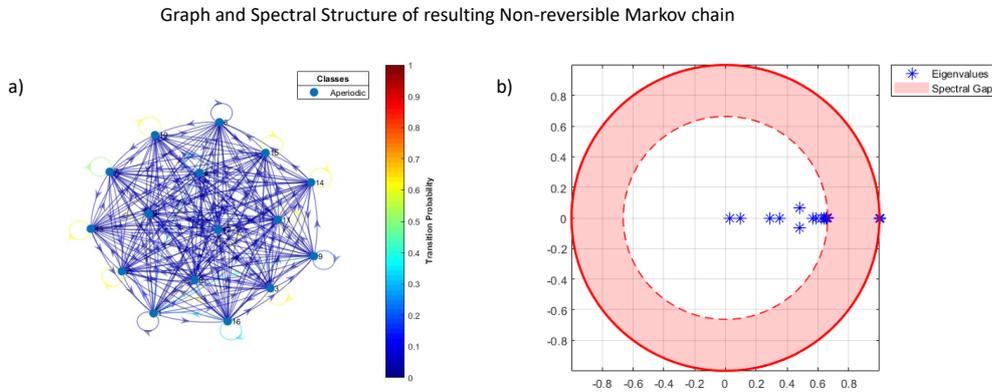


Fig. 5.3.1 a) The resulting Discrete-time Markov chain (DTMC) is represented by a complete digraph with self-loops b) The resulting discrete-time Markov chain transition matrix contains complex eigenvalues, thus it is a non-reversible Markov chain.

Assume that we have a 16-state target stationary distribution $\mathbf{P}_s = [10.5513, 5.8081, 0.7836, 0.3903, 0.7521, 0.2657, 0.4890, 0.7460, 0.2956, 0.7651, 0.9127, 0.6704, 0.7904, 5.7431, 0.6171, 0.3712]^T$. The distribution was formed initially by getting random numbers between 0 and 1 and then for the target probabilities of state 1, 2 and 14 constants 10, 5 and 5 were added to those states respectively. This was done in order to create a strongly multimodal distribution, as it is known that such distributions frequently pose difficulties for Markov Chain Monte Carlo methods [180].

For constructing the inverse M-matrix \mathbf{M}^{-1} , $x = 0.10$ and $y = 0.09$, $s^{-1} = 43.1$, satisfying the condition $s^{-1} \geq n - 2$. The matrix \mathbf{M}^{-1} , apart from the unit diagonal, was filled by parameters randomly selected between x and y .

Fig. 5.3.1 illustrates that the discrete time Markov chain obtained in this case through our algorithm, which has the desired stationary probability distribution, is a complete digraph (i.e. all states are connected to each other in both directions) with self-loops. For this we used

the generic method of creating the initial irreducible Markov chain using inverse M-matrices as well as using the optional version of the step 5 of the algorithm. It is also non-reversible, as illustrated by the fact that it has complex eigenvalues. Note that as the methodology uses random numbers and neither non-reversibility nor reversibility is assumed anywhere, it is possible that in some cases, a reversible Markov chain is obtained. This method aims to be a generic way to obtain Markov chains with a desired stationary probability distribution, without requiring to satisfy detailed balance.

5.4 Comparison with the Reversible Markov chain obtained from Metropolis-Hastings can show a better spectral gap

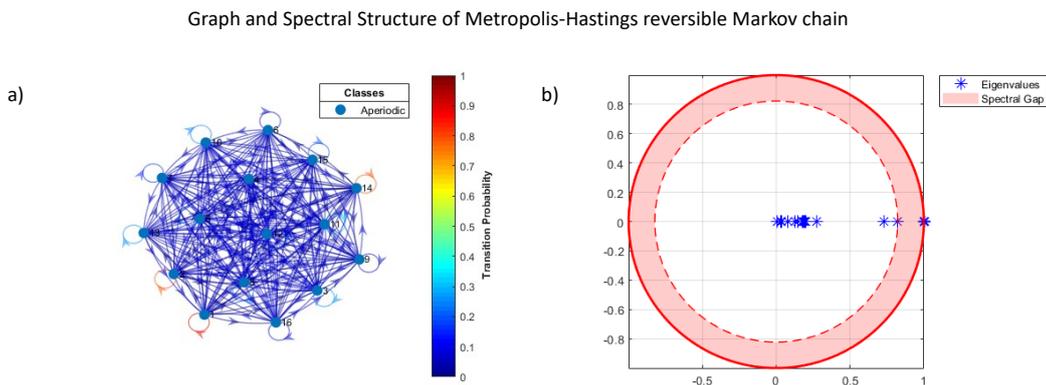


Fig. 5.4.1 a) The reversible Discrete-time Markov chain (DTMC) obtained from Metropolis-Hastings is also represented by a complete digraph with self-loops. Note that the transition probabilities of the self-loops of the modes of the stationary distribution (e.g. state 1) are larger than before. b) This suggests longer mixing times, as also suggested by the thin spectral gap

As mentioned in the introduction, non-reversible Markov chains were shown to perform better than reversible chains [126]. Does this also apply in this case? In order to verify it we used Metropolis-Hastings, which provides us with a reversible Markov chain [20]. In order to be able to do a fair comparison, we set up the proposal distribution of the Metropolis-Hasting to also be represented by a complete digraph, with transition probability $1/16$ for a jump to

take place from any state to any other, including itself.

The result is shown in Fig. 5.4.1. The resulting Markov chain is also represented by a complete digraph with self-loops, yet note that because of the reversibility condition, the transition probabilities of the self-loops of the states with high probability are larger than those of the non-reversible chain. This makes it expected that it will spend longer at those states, thus also making the mixing time longer. This is also predicted by the much thinner spectral gap (the difference between 1 and the second largest eigenvalue modulus μ). The estimated mixing time obtained by $\tau_{mix} = -1/\log(\mu)$ [20], is equal to 2.3872 for the non-reversible Markov chain obtained through our algorithm, whereas it is equal to 5.1202 for the reversible Markov chain obtained through Metropolis-Hastings.

5.5 Advantages of this algorithm

As we have seen, this algorithm can achieve both continuous-time and discrete-time non-reversible Markov chains directly without using previously constructed reversible Markov chains as done in the literature [17, 158].

Beyond, however, the ability to create multiple different generic non-reversible Markov chains with the same target probability distribution (which have in general better mixing properties than reversible ones), this methodology provides us with an extra advantage especially applicable to problems in Systems Biology, where a particular topology of the system is often pre-specified [131].

This is possible due to the algorithm's decomposable nature, as the system returned as an output of the algorithm returns exactly the same topology for the $n - 1$ states with the transition rates connecting the last distinguished n^{th} state to the rest $n - 1$ states, as also shown by the top left table of Fig 5.5.1. These transition rates correspond to the probability flows (to/from the distinguished state) needed to turn the initially provided system to one with the target stationary probability distribution.

This added transparency of how the target stationary probability is achieved allows us to use the output of the algorithm to also ensure that a topology is also maintained.

Take for example the case where the system's topology cannot accept a central distinguished state that can be directly connected to all the other states. It is now possible by using only the output of our algorithm and the target probability distribution, together with classical techniques, to return a system which respects the overall topology of the system and is also

consistent with the target stationary probability distribution.

To illustrate how this can be done, take the 3×3 example system used in Chapters 3 and 4, with the magnitude of dominant edges equal to 2 and the magnitude of non-dominant edges equal to 1. Assume now that we would like to turn this system, by maintaining its topology, to a continuous-time Markov chain with target probability distribution $P_s = [0.1000, 0.1000, 0.2000, 0.1000, 0.1000, 0.1000, 0.1000, 0.1000, 0.1000]$. Without loss of generality choose state 1 as the distinguished state. As Fig. 5.5.1 shows, the output of the algorithm provides 4 probability flows along direct connections not allowed by the topology of the system. These are: a connection from state 5 to state 1 with magnitude 4, a connection from state 7 to state 1 with magnitude 2, a connection from state 1 to state 9 with magnitude 2 and a connection from state 1 to state 8 with magnitude 1. Using classical techniques, such as Breadth-first search [22], we can find a path connecting each pair of states on the graph and then superimpose the probability flows needed using the information from the target probability distribution P_s . Fig. 5.5.1 illustrates diagrammatically how this is done, with the resulting system that satisfies both the topology of the system and the target stationary probability distribution.

5.6 Conclusion

In this chapter, we developed a novel algorithm for constructing continuous- and discrete-time Markov chains with a desired stationary probability distribution on a finite discrete space. This is achieved directly without using previously constructed reversible Markov chains as usually done in literature. Instead, the formation of the stationary probability distribution is controlled from a distinguished state, using the formulation for continuous-time Markov chains and the M-matrix theory, presented in Chapter 4. Finally, for an example with a strongly multimodal stationary probability distribution, we compared the non-reversible Markov chain obtained from our algorithm to the reversible Markov chain obtained from the Metropolis-Hastings algorithm, where we find that it has better spectral and thus estimated mixing properties.

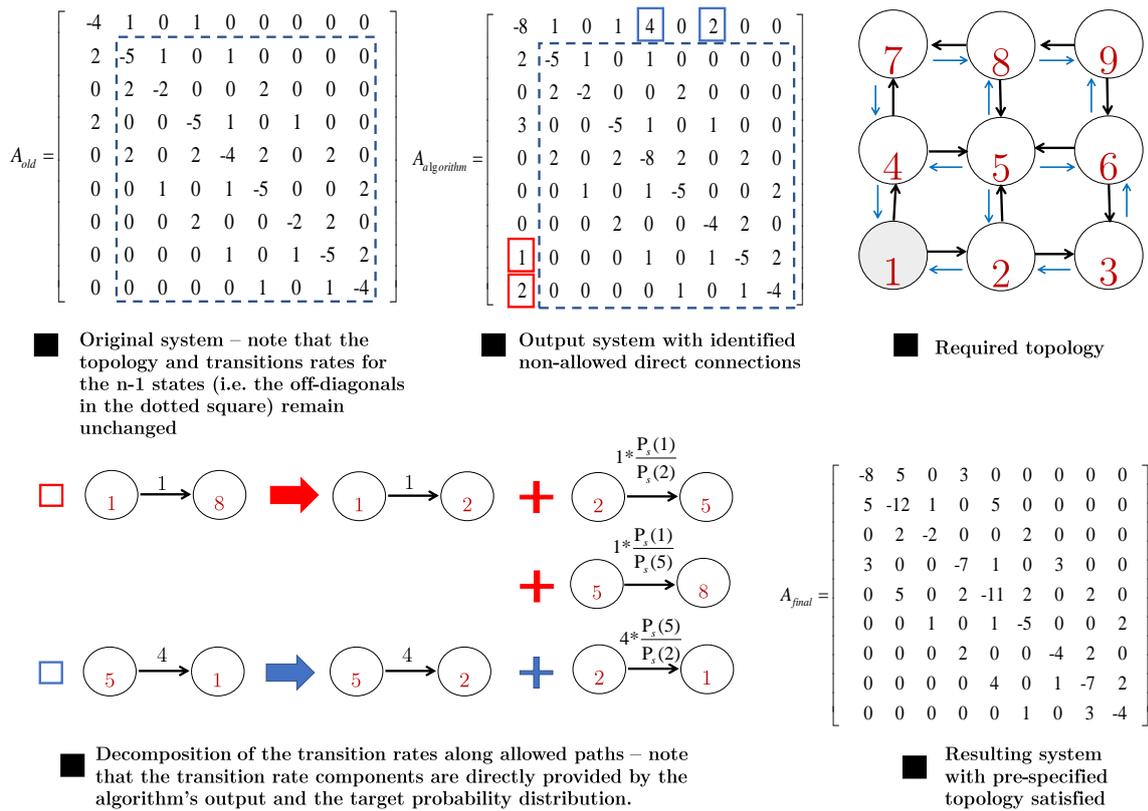


Fig. 5.5.1 Illustration of how the output of the algorithm can be used to return a system which respects the overall topology of the system and is consistent with the target stationary probability distribution

Chapter 6

Enzyme Sequestration by the Substrate: An Analysis in the Deterministic and Stochastic Domains

The previous chapters, along with the development of new tools to facilitate stochastic analysis of biochemical systems, have also enhanced our understanding of how probability distributions in stochastic systems are formed. In this chapter, we use that understanding to analyse more complex potentially multistable systems, focusing on multisite protein phosphorylation both in the deterministic and stochastic domains. In particular, we analyse the consequences of adding enzyme docking to these and similar models, with the resultant Enzyme Sequestration by the Substrate they imply.

6.1 Introduction

Probably the most studied form of protein modification is protein phosphorylation, the binding of a phosphoryl (PO_3^-) group using a kinase enzyme [120]. This, together with dephosphorylation by a phosphatase enzyme, contributes to the regulation of transcription factors, thus regulating the response of a cell to changes in its environment [176]. Goldbeter and Koshland [61] showed that ultrasensitivity can be obtained where a sigmoidal change is observed in output for a linear change in input. This, coupled with positive feedback, can result in bistability. Positive feedback, which can be exhibited implicitly by different mechanisms, is required for bistability and consequently for multistability [46]. Examples of such mechanisms are several [75, 49, 45]. In this chapter we focus on multisite protein phosphorylation, a well studied example of such a mechanism [145, 32, 143, 168, 51, 26, 78,

173, 12], itself belonging to the greater class of enzyme-sharing schemes (i.e. when different substrates or substrate states share the same enzymes). This mechanism is of interest because of its potential unlimited multistable behaviour [51, 66], which could be beneficial for using information from environmental signals to drive internal cell processes.

In the excess substrate regime, Thomson and Gunawardena [163] showed that the number of stable steady states that can be achieved increases linearly with the number of phosphosites available. This is done by introducing enzyme saturation and competition between the unphosphorylated and phosphorylated substrate forms for interaction with the free kinase and with the free phosphatase [51, 163]. The ability of this form of competition to induce bistability in a distributive kinetic mechanism of the two-site MAPK (Mitogen-activated protein kinase) phosphorylation and dephosphorylation was firstly shown by Kholodenko et al [111, 51, 50].

Nevertheless, it is increasingly being recognised that specificity in protein phosphorylation and dephosphorylation cycles can be achieved through enzyme docking: the binding of the interaction domains on the kinase or phosphatase with one or more docking sites on the substrate, where the latter is separate from the motif that is chemically modified [142, 168, 140, 14, 144]. Examples of such docking interactions that have been identified include MAPK and MAPK phosphatases [151, 8, 109], and Glycogen synthase kinase-3 [178], an important kinase for insulin and Wnt signalling [168]. This mechanism implies that a phosphatase molecule can still bind to an unphosphorylated substrate molecule and similarly, a kinase molecule can still bind to a fully-phosphorylated one, forming inactive complexes, as each enzyme can always bind to their docking site [112]. The formation of inactive complexes is graphically illustrated in Fig. 6.1.1. In the excess substrate regime, the formation of such complexes can be thought of as a sequestration mechanism, where the substrate sequesters away the enzymes. This is referred to in the chapter as ‘Substrate Enzyme-Sequestration’. In the complementary regime of excess enzyme, Martins and Swain have already shown that this type of sequestration can provide ultrasensitivity [112].

Note that this mechanism of enzyme sequestration is fundamentally different to that of enzyme sequestration by a different protein (e.g. a scaffold protein) that does not participate in the reaction scheme metabolically [48]. In that type of sequestration, the scaffold-bound population is separated from the rest of the reaction network, creating two compartments. Indeed, compartmentalisation is another mechanism able to provide enhanced ultrasensitivity, bistability and/or multistability [26, 48, 70]. However, in Substrate Enzyme-Sequestration, neither additional proteins nor compartments are sequestering the enzyme; this is done by the substrate itself, as also explained by Martins and Swain [112]. As this sequestration is depen-

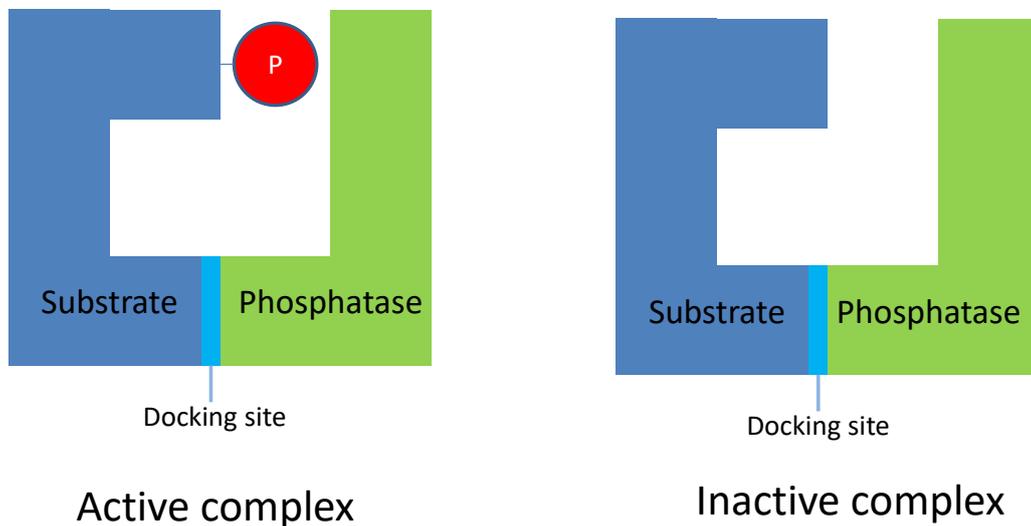


Fig. 6.1.1 An inactive complex can for example be formed when a phosphatase molecule binds with a completely unphosphorylated substrate.

dent on the inherent way the substrate attaches to the enzymes, identifying it experimentally is equivalent to identifying whether the enzyme has any means of avoiding the binding with a substrate found in a phosphorylation state which would create an inactive complex, as for example is the binding of a phosphatase to a completely unphosphorylated substrate.

Here we investigate the effect that this type of sequestration can have on multisite protein phosphorylation in the excess substrate regime in the domains of both large and small numbers of molecules, where a deterministic and a stochastic analysis are respectively more suitable. For the stochastic analysis, we use the weakly diagonally dominant M -matrix formulation of the Chemical Master Equation, and its associated stochastic tool developed in Chapter 4, which allows greater insights on the formation of probability distributions, without the necessity of continuously calculating the exact solution of the steady state distribution or running Monte Carlo simulations.

6.2 Models

6.2.1 A deterministic framework for the excess substrate regime [163]

Our analysis in the large molecule number domain is based on the deterministic framework of Thomson and Gunawardena [163, 162] which was used to show mathematically that unlimited multistability is possible. We first summarise their results, and then extend them to account for Substrate Enzyme-Sequestration by including the reactions outside the red dotted frame of Fig. 6.2.1. We will show that, irrespective of the other parameters of the system, as the strength of sequestration is increased then the number of steady states decreases to one.

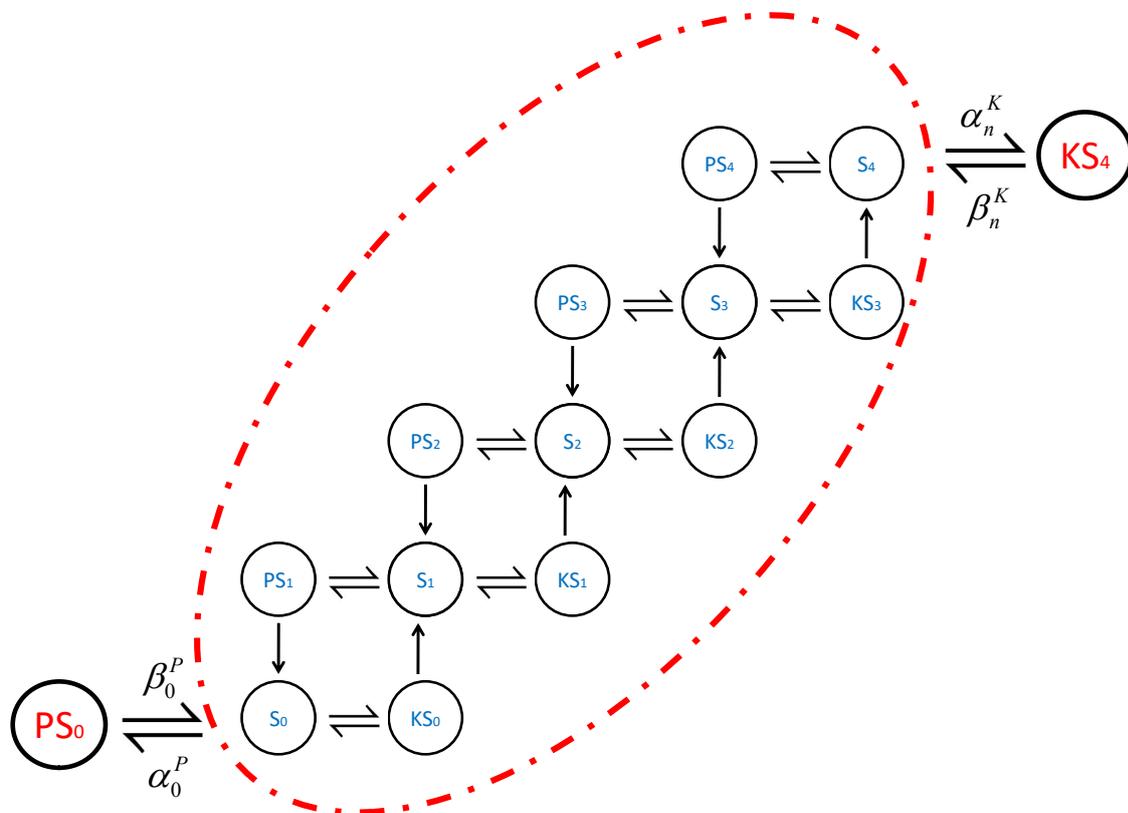
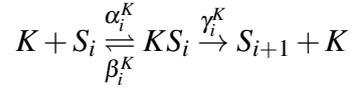


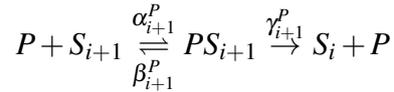
Fig. 6.2.1 The original 4-site protein phosphorylation scheme is extended to include two inactive complexes PS_0 and KS_4 , as implied by enzyme docking.

Fig. 6.2.1 is built from of two kinds of reaction: firstly, a kinase molecule K can attach to a substrate molecule with i phosphorylated phosphosites, S_i . The new complex formed, KS_i , can then either decompose back to K and S_i or phosphorylation can proceed, leading to the

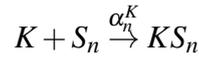
products K and S_{i+1} .



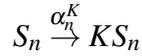
In addition, a phosphatase molecule P can attach to a substrate molecule with $i + 1$ phosphorylated phosphosites, S_{i+1} with the new complex formed PS_{i+1} either decomposing back to P and S_{i+1} or lead to a dephosphorylation reaction with products P and S_i .



Note that in Fig. 6.2.1 and in subsequent schematic figures, we present only the species that account for the total substrate concentrations. The participating enzymes (phosphatase and kinase) are implicitly taken into account as part of the relevant reactions. For example, the reaction



is represented in the figure as



Under the assumption of excess substrate, i.e. that the total concentration of substrate $[S_{\text{tot}}] \gg [K_{\text{tot}}]$ and $[S_{\text{tot}}] \gg [P_{\text{tot}}]$, Thomson and Gunawardena [162] showed that the steady states of a phosphorylation system can be determined by the roots of the following polynomial (where $u = [K]/[P]$ and $w = \frac{[K_{\text{tot}}]}{[P_{\text{tot}}]}$),

$$R(u) = a_{n+1}u^{n+1} + a_nu^n + \dots + a_1u + a_0 \quad (6.2.1)$$

The expressions for the coefficients a_i are given in [162] as

$$\begin{aligned} a_{n+1} &= \lambda_0 \lambda_1 \dots \lambda_{n-2} \lambda_{n-1} \\ a_0 &= -w \\ a_{i+1} &= \lambda_0 \lambda_1 \dots \lambda_{i-2} \lambda_{i-1} \left[(1 - \lambda_i w) + [S_{\text{tot}}] \left(\frac{1}{k_i^K} - \frac{\lambda_i w}{k_{i+1}^P} \right) \right], \\ 0 &\leq i < n, \lambda_{-1} = 1 \end{aligned}$$

where

$$\lambda_i = \left(\frac{\gamma_i^K}{k_i^K} \right) \left(\frac{\gamma_{i+1}^P}{k_{i+1}^P} \right)^{-1}, \quad k_i^K = \frac{\beta_i^K + \gamma_i^K}{\alpha_i^K}, \quad k_i^P = \frac{\beta_i^P + \gamma_i^P}{\alpha_i^P} \quad (6.2.2)$$

For completeness, these can be derived as follows:

Using mass kinetics, the steady state concentration of KS_i can be determined from the concentrations of the free kinase, K , and of the free substrate with i full phosphosites, S_i .

$$[KS_i] = \frac{[K][S_i]}{k_i^K},$$

Similarly for the phosphatase,

$$[PS_i] = \frac{[P][S_i]}{k_i^P},$$

At steady state, in each cycle the net flux into S_{i+1} must be equal to the net flux out of S_{i+1} . Consequently, taking into consideration each cycle in turn, one can express the concentration of any substrate state, $[S_{i+1}]$ as a function of the concentration of the substrate with no phosphorylated phosphosites, $[S_0]$, i.e.

$$[S_{i+1}] = [S_0] \lambda_0 \lambda_1 \dots \lambda_i \left(\frac{[K]}{[P]} \right),$$

Using conservation of mass, we can therefore write the total substrate concentration in terms of the individual species

$$[S_{\text{tot}}] = [S_0] + \dots + [S_n] + [KS_0] + \dots + [KS_{n-1}] + [PS_1] + \dots + [PS_n] \quad (6.2.3)$$

$$= [S_0] (\phi_1 + [K]\phi_2 + [P]\phi_3) \quad (6.2.4)$$

where

$$\begin{aligned}\phi_1 &= \sum_{i=0}^n \left(\prod_{j=0}^{i-1} \lambda_j \right) u^i \\ \phi_2 &= \sum_{i=0}^{n-1} \frac{\left(\prod_{j=0}^{i-1} \lambda_j \right) u^i}{k_i^K} \\ \phi_3 &= \sum_{i=1}^n \frac{\left(\prod_{j=0}^{i-1} \lambda_j \right) u^i}{k_i^P}\end{aligned}$$

Note that the coefficients of the polynomials ϕ_1 , ϕ_2 , and ϕ_3 can be written in terms of the rate constants α , β and γ in their various subscripted and superscripted forms and, importantly, are all positive.

The total concentrations of the kinase and phosphatase can be expressed in terms of the same functions:

$$[K_{\text{tot}}] = [K] (1 + [S_0] \phi_2) = [K] \left(1 + \frac{[S_{\text{tot}}] \phi_2}{\phi_1 + [K] \phi_2 + [P] \phi_3} \right) \quad (6.2.5)$$

(using Eq. 6.2.4)

$$=: K_{\text{tot}}([P], [K]) \quad (6.2.6)$$

That is, the total concentration of the kinase is regarded as a function of free kinase and phosphatase concentrations. Similarly,

$$[P_{\text{tot}}] = [P] (1 + [S_0] \phi_3) = [P] \left(1 + \frac{[S_{\text{tot}}] \phi_3}{\phi_1 + [K] \phi_2 + [P] \phi_3} \right) \quad (6.2.7)$$

$$=: P_{\text{tot}}([P], [K]) \quad (6.2.8)$$

A point $([P], [K])$ that satisfies $K_{\text{tot}}([P], [K]) = [K_{\text{tot}}]$, $P_{\text{tot}}([P], [K]) = [P_{\text{tot}}]$ is a possible steady state for the system and corresponds to an intersection of contours of K_{tot} and P_{tot} in a free phosphatase-free kinase map. This can be shown Fig. 6.2.2, where the stable steady states correspond to the steady states of the time courses when the system is deterministically simulated. Different steady states result according to the starting conditions.

As $\frac{\phi_1}{\phi_1 + [K] \phi_2 + [P] \phi_3} = \frac{[S_{\text{tot}}] - [S_0][K] \phi_2 - [S_0][P] \phi_3}{[S_{\text{tot}}]}$, $[S_0][K] \phi_2 < [K_{\text{tot}}]$ and $[S_0][P] \phi_3 < [P_{\text{tot}}]$, in the regime of excess substrate (i.e. $[K_{\text{tot}}] \ll [S_{\text{tot}}]$ and $[P_{\text{tot}}] \ll [S_{\text{tot}}]$), the expressions of $[K_{\text{tot}}]$ and $[P_{\text{tot}}]$ can be written as follows [162, 163]:

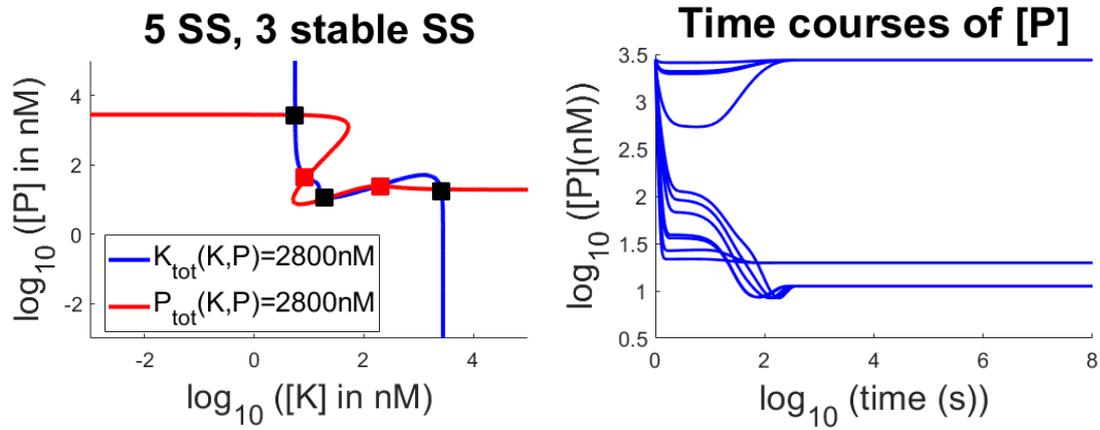


Fig. 6.2.2 Plots of total concentrations of kinase and phosphatase as functions of free kinase and phosphatase (left) of the original tristable system presented in [163]. Intersections correspond to steady states. The stable steady states are coloured in black, whereas the unstable steady states are coloured in red. The stable steady states are the steady states appearing in the time courses (right) when the system is simulated deterministically, starting from different initial conditions.

$$[K_{tot}] = [K] \left(1 + [S_{tot}] \frac{\phi_2}{\phi_1} \right)$$

$$[P_{tot}] = [P] \left(1 + [S_{tot}] \frac{\phi_3}{\phi_1} \right)$$

Dividing the two expressions and rearranging, we obtain:

$$0 = (u - w)\phi_1 + [S_{tot}](u\phi_2 - w\phi_3) =: R(u)$$

where $(w = \frac{[K_{tot}]}{[P_{tot}]})$.

As ϕ_1 , ϕ_2 , and ϕ_3 are functions of u , the above expression can be expressed as

$$R(u) = a_{n+1}u^{n+1} + a_nu^n + \dots + a_1u + a_0$$

$$\begin{aligned}
 a_{n+1} &= \lambda_0 \lambda_1 \dots \lambda_{n-2} \lambda_{n-1} \\
 a_0 &= -w \\
 a_{i+1} &= \lambda_0 \lambda_1 \dots \lambda_{i-2} \lambda_{i-1} \left[(1 - \lambda_i w) + [S_{tot}] \left(\frac{1}{k_i^K} - \frac{\lambda_i w}{k_{i+1}^P} \right) \right], 0 \leq i < n, \lambda_{-1} = 1
 \end{aligned}$$

The roots of $R(u)$ correspond to steady state enzyme ratios. Note that these correspond to non-equilibrium steady states of the underlying biochemical system, since each phosphorylation/dephosphorylation cycle is driven in a counter-clockwise direction by the hydrolysis of ATP. The important point, though, is that the leading coefficient a_{n+1} is positive, as it derives from the leading coefficients of ϕ_1 and ϕ_2 . Conversely, the trailing coefficient a_0 is negative, as it derives from the trailing coefficients of ϕ_1 and ϕ_3 multiplied by $-w$ i.e.

$$a_{n+1} > 0, \quad a_0 < 0 \quad (6.2.9)$$

Thus the problem of finding the steady states of the system is transformed into a problem of finding the roots of a univariate polynomial. This polynomial can have no more than $n + 1$ real positive solutions, and Descartes' rule of signs was used to show that when n is odd there can be no more than n real positive solutions (because the reversal of sign between the first and last coefficients limits the number of changes in sign). Thus the maximum number of stable steady states is equal to $1 + \lfloor \frac{n}{2} \rfloor$ [163, 162]. Gunawardena and Thomson further showed that it was possible to achieve this number by realistic choices of parameter values. Note however that the extent of multistability observed experimentally is much more limited [51], as also mentioned in their seminal paper [163].

6.2.2 Substrate-Kinase and Substrate-Phosphatase Sequestration

Enzyme docking allows the possibility that an enzyme may attach to substrate even when the complex formed will not be active. Fig. 6.2.1 represents the reactions occurring when the multisite protein phosphorylation is distributive and sequential as is often taken to be the case [173, 51, 162].

In sequential (de)phosphorylation, phosphosites are (de)phosphorylated in a strict order where (de)phosphorylation of one site depends on the phosphorylation state of another. This is opposite to a random (de)phosphorylation scheme. Distributive (de)phosphorylation occurs when the enzyme dissociates after each (de)phosphorylation. This is opposite to processive (de)phosphorylation where multiple (de)phosphorylations might take place before the enzyme dissociates from the substrate [152].

Most of our results generalise to the non-distributive, non-sequential case, but we start by describing the simpler case, where the conclusions are sharper. The inactive complexes formed after a kinase binds to a fully phosphorylated substrate and after a phosphatase binds to an unphosphorylated substrate are shown outside the dashed area. These complexes represent an example of substrate enzyme-sequestration. This occurs when a substrate molecule (e.g. a fully phosphorylated substrate molecule) forms an inactive complex with an enzyme molecule (e.g. a kinase), neither allowing the enzyme to bind to other substrate molecules to form active complexes nor any other enzyme (e.g. a phosphatase) to bind to itself. This effect can occur, for example, through competition of the enzymes for the same, or partly the same, docking sites, as illustrated in the literature [77, 8].

6.2.2.1 Model extension with the addition of Substrate Enzyme Sequestration

We now extend the analysis of Gunawardena and Thomson to investigate the potential effects of substrate enzyme-sequestration. First, the polynomials ϕ need to be redefined in order to accommodate the sequestration effects, since now conservation of mass includes two extra species, KS_n and PS_0 :

$$[S_{\text{tot}}] = [S_0] + \dots + [S_n] + [KS_0] + \dots + [KS_{n-1}] \\ + [PS_1] + \dots + [PS_n] + [PS_0] + [KS_n]$$

As shown in Fig. 6.2.1, each of the two new species interacts with just one of the species of the original system (S_n and S_0). Consequently, at steady state, the flow from PS_0 into the original system via S_0 has to be equal to the flow from the original system to PS_0 . Similarly for S_n and KS_n .

In order to take these additional species into account, the polynomials ϕ_2 and ϕ_3 need to be modified to include all n substrate states in the mass conservation equations. We will show that increasing the strength of sequestration changes only the magnitude of the leading and trailing coefficients in the resulting polynomial $R(u)$ which, as a further consequence of the sign change in Eq. 6.2.9, inevitably leads to a reduction in the number of steady states – eventually to one.

6.2.2.2 Comparison with Legewie sequestration mechanism

Before proceeding to the Results section, it is important to note the difference between the Substrate Enzyme-Sequestration we describe in this chapter and the mechanisms where the not fully phosphorylated substrate can inhibit the activation of its kinase, which itself gets activated on a separate phosphorylation cascade. Legewie et al described such mechanism for

the Raf-Mek-Erk mitogen-activated protein kinase cascade [98] which results in a positive circuit, inducing bistability. The mechanism is illustrated in Fig. 6.2.3.

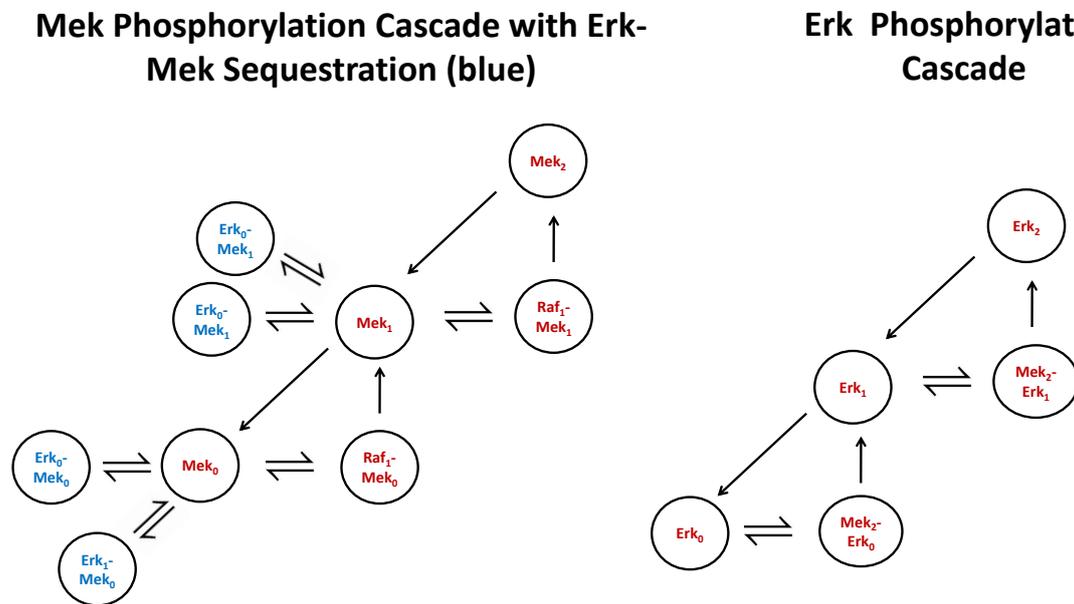


Fig. 6.2.3 Illustration of Legewie mechanism.

In the aforementioned mechanism, the not fully phosphorylated Erk (Erk_0 and Erk_1) species sequester Mek into complexes that cannot be accessed by the Mek's kinase, Raf. This inhibits the formation of the doubly phosphorylated Mek, Mek_2 , which acts as Erk's kinase. Upon weak Raf_1 stimulation, Erk and Mek are mostly unphosphorylated, whereas stronger Raf_1 stimulation increases Mek's double phosphorylation, which in turn phosphorylates the not fully phosphorylated forms of Erk (Erk_0 and Erk_1). This, in turn amplifies Mek phosphorylation, creating the positive circuit which allows for bistability to be observed. Legewie mechanism contrasts with the sequestration mechanism we describe in this chapter in two fundamental ways. Firstly, in our case, sequestration does not inhibit the enzyme activation as the enzymes sequestered by the substrate are already in their active form. Secondly and most importantly, the inactive complexes arising from sequestration are formed by species participating in the same phosphorylation cascade, which in itself is able to portray multistability. This is different to Legewie's mechanism where the inactive complexes formed act as a way to couple two different sequential phosphorylation cascades, introducing a positive feedback mechanism between the fully-phosphorylated substrate of the second phosphorylation cascade and the activation of its kinase, which is the substrate of the first phosphorylation cascade.

6.3 Results and Discussion

6.3.1 Substrate Enzyme-Sequestration in the Deterministic Domain

6.3.1.1 What happens when Substrate Enzyme-Sequestration is considered?

Having extended the deterministic framework to account for substrate enzyme-sequestration in Section 6.2.2.1, we investigated whether this additional competition for enzymes enhances or inhibits the extent of multistability.

The resulting contours for $[K_{\text{tot}}]$ and $[P_{\text{tot}}]$ are shown in Fig. 6.3.1, which shows how the original system in [163, 162], with $[K_{\text{tot}}] = 2.8\mu\text{M}$, $[P_{\text{tot}}] = 2.8\mu\text{M}$, $[S_{\text{tot}}] = 10\mu\text{M}$ and 4 available phosphosites, is affected as the strength of Substrate Enzyme-Sequestration ($\frac{\alpha_n^K}{\beta_n^K}$ and $\frac{\alpha_0^P}{\beta_0^P}$) increases. Furthermore, note that the contours presented in the figures represent the accurate non-approximated $[K_{\text{tot}}]$ and $[P_{\text{tot}}]$. As can be seen, as the strength of sequestration is increased the number of steady states (and stable steady states) decreases continuously from 5 (3 of which are stable) in the original tristable system of [163, 162] to 3 (2 of which are stable) to a single stable steady state. Note that the ratios $\frac{\alpha_i^K}{\beta_i^K}$ for $0 \leq i < n$ and $\frac{\alpha_i^P}{\beta_i^P}$ for $0 < i \leq n$ are approximately equal to $5 \times 10^{-1} \text{ nM}^{-1}$, as seen in Appendix, Section 6.6.1.7.

6.3.1.2 Sufficient Conditions for Further Limiting the Extent of Multistability

Having demonstrated the qualitative effect of Substrate Enzyme-Sequestration we now derive quantitative conditions for the reduction in multistability.

Taking the additional species into account, as in Fig. 6.2.1, ϕ_2 and ϕ_3 need to just be modified in terms of the summation indices to include all n substrate states in the mass conservation equations. Defining $k_n^K = \frac{\beta_n^K}{\alpha_n^K}$ and $k_0^P = \frac{\beta_0^P}{\alpha_0^P}$ for the new species, the ϕ polynomials become:

$$\hat{\phi}_2 = \sum_{i=0}^n \frac{\left(\prod_{j=0}^{i-1} \lambda_j \right) u^i}{k_i^K}$$

$$\hat{\phi}_3 = \sum_{i=0}^n \frac{\left(\prod_{j=0}^{i-1} \lambda_j \right) u^i}{k_i^P}$$

The updated polynomial, applicable for the regime is shown in Eq. 6.3.1. The new polynomial coefficients are presented in terms of the old coefficients in order to allow for comparisons, providing an insight of the quantitative effect of Substrate Enzyme-Sequestration. Note that only the first and the last coefficients are changed.

$$R'(u) = a'_{n+1}u^{n+1} + a_n u^n + \dots + a_1 u + a'_0 \quad (6.3.1)$$

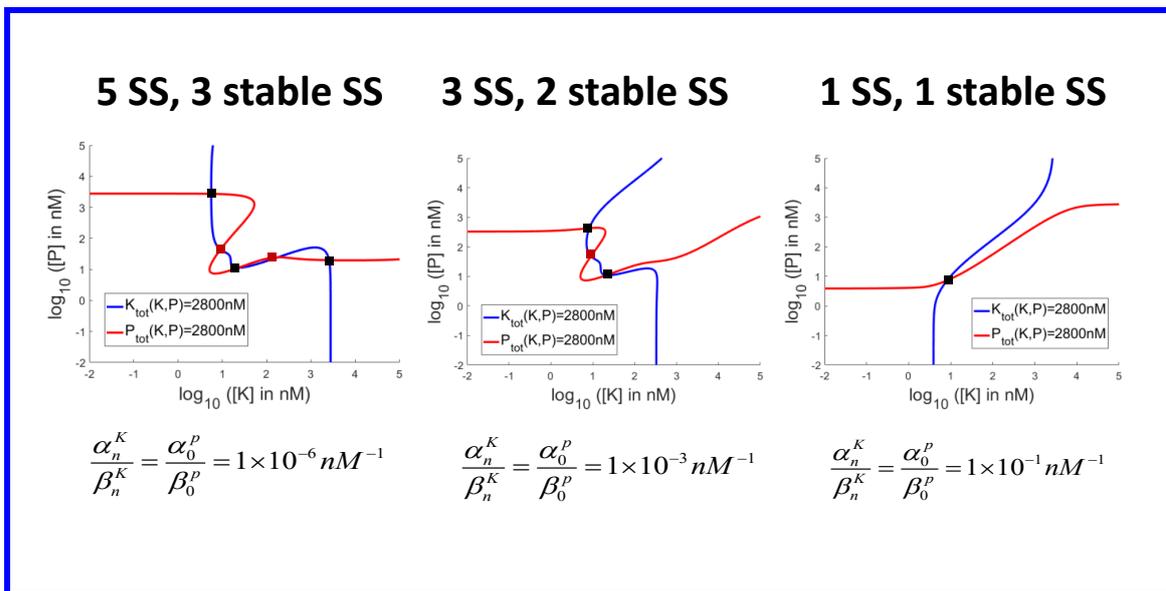


Fig. 6.3.1 Plots of total concentrations of kinase and phosphatase as functions of free kinase and phosphatase. Increasing the strength of both the Substrate Kinase ($\frac{\alpha_n^K}{\beta_n^K}$) and Substrate Phosphatase-Sequestration ($\frac{\alpha_0^P}{\beta_0^P}$), the number of steady states (and stable steady states) falls in a continuous fashion from 5 (3 stable), as in the system without any sequestration, to 3 (2 stable) to 1 (1 stable).

where

$$a'_{n+1} = a_{n+1} \left(1 + [S_{tot}] \frac{\alpha_n^K}{\beta_n^K} \right)$$

$$a'_0 = a_0 \left(1 + [S_{tot}] \frac{\alpha_0^P}{\beta_0^P} \right)$$

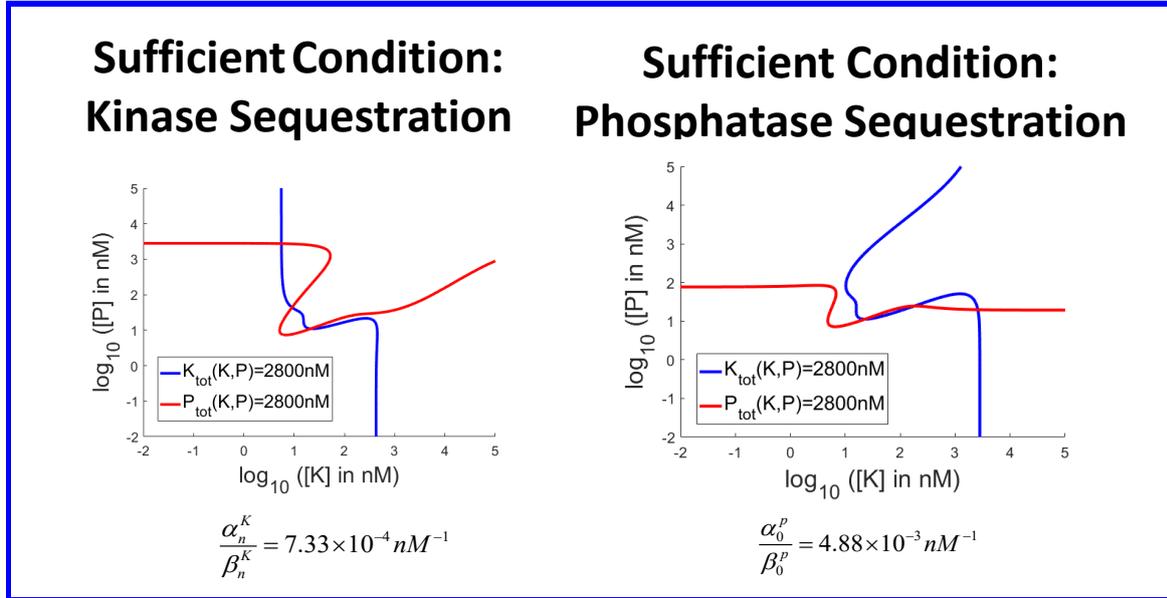


Fig. 6.3.2 The sufficient bounds are reasonably tight for the original tristable system [163, 162] for both Substrate Kinase-Sequestration (left) and Substrate Phosphatase-Sequestration. This tightness is portrayed in the figures by the small separation of the K_{tot} and P_{tot} contours at the regions where there was previously an intersection.

That is, the sign difference between the leading and trailing coefficients is maintained, and they are increased in magnitude. This limits the number of possible positive roots. Using the Vieta formulae, which relate the coefficients of a polynomial to sums and products of its roots [89], in conjunction with the Quadratic Mean - Arithmetic Mean [40] and the Triangle inequalities it is possible to obtain sharp bounds. The following theorem, which is proved in Appendix, shows that if either the leading three terms or trailing three terms fail a discriminant like condition then the polynomial must have a pair of complex roots, in which case the potential number of real positive roots is reduced by two and the number of stable steady states by one:

Theorem 6.3.1. *If any of the following conditions are satisfied, then the number of positive steady states will be no more than $n - 1$ if n is even, or $n - 2$ if n is odd:*

1. $a_{n-1} \leq 0$ and $a_2 \geq 0$
2. $a_{n-1} > 0$ and $\frac{\alpha_n^K}{\beta_n^K} [S_{tot}] > \frac{na_n^2 - 2(n+1)a_{n+1}a_{n-1}}{2(n+1)a_{n+1}a_{n-1}}$ or

$$3. \ a_2 < 0 \text{ and } \frac{\alpha_0^P}{\beta_0^P} [S_{tot}] > \frac{na_1^2 - 2(n+1)a_0a_2}{2(n+1)a_0a_2}$$

Thus it is always possible to choose sequestration rates such that the maximum number of stable steady states is equal to $\lfloor \frac{n}{2} \rfloor$. Fig. 6.3.2 illustrates that the stated bounds are reasonably tight for the original tristable system [163, 162], as the small gaps between the contours $[K_{tot}] = 2.8\mu M$ and $[P_{tot}] = 2.8\mu M$, in the region of their previous intersections, demonstrate. For the original system to be tristable (i.e. with five steady states) both $\frac{\alpha_n^K}{\beta_n^K} \leq 4.4 \times 10^{-4} nM^{-1}$ and $\frac{\alpha_0^P}{\beta_0^P} \leq 3.2 \times 10^{-3} nM^{-1}$ have to be satisfied as found in simulations. If either condition is violated, then the system becomes bistable. If both are violated, then the system becomes monostable. Our derived conditions are quite close to these, as we find that it is sufficient that $\frac{\alpha_n^K}{\beta_n^K} \geq 7.33 \times 10^{-4} nM^{-1}$ or $\frac{\alpha_0^P}{\beta_0^P} \geq 4.88 \times 10^{-3} nM^{-1}$ is satisfied for tristability to be limited to bistability.

Furthermore, based on the Pratt's tableau test [137], which is a method of finding a less conservative than Descartes' rule of signs upper bound for the real positive roots of a real polynomial, the following theorem is proved in Appendix, Section 6.6.1.3

Theorem 6.3.2. *For any $\delta^K \geq 0$ there exists δ^P , directly computable from the rate constants, such that if $\frac{\alpha_n^K}{\beta_n^K} [S_{tot}] = \delta^K$ and $\frac{\alpha_0^P}{\beta_0^P} [S_{tot}] \geq \delta^P$ then the polynomial $R^l(u)$ has precisely one positive root, corresponding to one steady state. Similarly, for any $\delta^P \geq 0$ there exists a δ^K with the same properties.*

As explained in more detail in Appendix, by setting a δ^K , a δ^P can be calculated directly from an algorithm based on the Pratt tableau (which we develop and is found in Appendix, Section 6.6.1.5) and vice-versa. In this way, we iteratively increased δ^K (the input) until δ^P (the output) could not decrease anymore. This algorithm provided us with the following condition: if $\frac{\alpha_n^K}{\beta_n^K} = 6.57 \times 10^{-4} nM^{-1}$ and $\frac{\alpha_0^P}{\beta_0^P} \geq 6.928 \times 10^{-2} nM^{-1}$, then the original tristable system [163, 162] can only have one steady state. Again, these numbers are reasonable when compared to the aforementioned actual values obtained via simulations.

6.3.1.3 Does direct decrease of overall substrate/enzyme numbers have the same effect as Substrate Enzyme-Sequestration?

The inclusion in the model of the inactive complexes PS_0 and KS_n decreases the numbers of both unbound substrate and free enzymes. A natural question then is whether the observed limits on multistability could be attributed simply to these lowered concentrations. In order to investigate this, we found the concentrations of substrate and enzymes that are sequestered away because of the complexes PS_0 and KS_n when $\frac{\alpha_n^K}{\beta_n^K} = \frac{\alpha_0^P}{\beta_0^P} = 3.3 \times 10^{-3} nM^{-1}$ (i.e. when the system is exhibiting a monostable behaviour). Then we checked the behaviour of the

same system without sequestration with the corresponding lower total substrate and enzyme concentrations. We found that substrate enzyme-sequestrations effects cannot be attributed to a simple decrease of the numbers of enzyme and substrate, as the 4-site system still exhibited tristability, as illustrated by Fig. 6.3.3. In fact, tristability persists even when 2799.9nM of the 2800nM total concentration of both the phosphatase and the kinase (together with 5599.8nM of the 10000nM substrate concentration) are removed. This is shown in Fig. 6.3.4.

6.3.1.4 How does Substrate Enzyme-Sequestration limit multistability?

Having established that it is not the decrease in enzyme concentration numbers that limits multistability we search for an intuitive understanding of what does. It is helpful to consider the simpler two-site, bistable system described by Kholodenko et al [111, 51], which corresponds to the bottom left of Fig. 6.2.1. Bistability there occurs because the unphosphorylated substrate S_0 inhibits the production of the fully phosphorylated substrate S_2 by competing with the singly-phosphorylated S_1 for the kinase, while S_2 inhibits the production of S_0 by competing with S_1 for the phosphatase [51]. Thus, allowing S_0 to bind with the phosphatase has the effect that it is now inhibiting its own production as well by competing with S_1 for the phosphatase. As it is inhibiting its own production, it now becomes a worse inhibitor for the production of S_2 . The same applies to S_2 when the binding with the kinase is permitted. Thus, Substrate Enzyme-Sequestration reduces the coupling which caused bistability in the first place.

This explains our finding that no matter what the other kinetic parameters of the system are, we can always calculate a minimum strength of sequestration which limits the extent of multistability (Theorem 6.3.1) or even reduce it to one (Pratt tableau algorithm, Appendix, Section 6.6.1.5). For the parameters of the model (as in Section 6.6.1.7) i.e. for equal concentrations of kinase and phosphatase ($w = 1$), S_{tot} large and $\gamma_0^K \ll \gamma_1^P, \gamma_2^P \ll \gamma_1^K$, we can approximate condition 3) of Theorem 6.3.1 as $\frac{\alpha_0^P}{\beta_0^P} [S_{tot}] \gtrsim \frac{n}{2(n+1)} \frac{k_1^K}{k_0^K k_1^P} \frac{\gamma_1^P \gamma_2^P}{\gamma_0^K \gamma_1^K} \approx 5 \times 10^{-3} nM^{-1}$ (Appendix, Section 6.6.1.2). The exact result, as also shown previously, is $4.88 \times 10^{-3} nM^{-1}$. Here, $k_i^K = \frac{\beta_i^K + \gamma_i^K}{\alpha_i^K}$, $k_i^P = \frac{\beta_i^P + \gamma_i^P}{\alpha_i^P}$ are the Michaelis-Menten constants, which are inversely proportional to the rate constants for the production of enzyme-substrate intermediates from free enzymes and substrates. This approximation is biologically meaningful and consistent with the mechanism described above for the two-site case. The right-hand side of this condition is smaller (making multistability less robust to substrate enzyme-sequestration) when S_1 forms KS_1 intermediates more readily than PS_1 intermediates (i.e. small k_1^K , large k_1^P), when S_0 has a low affinity for forming KS_0 intermediates (i.e. large k_0^K), or when the competition for the phosphatase is higher and phosphatase is less readily made available

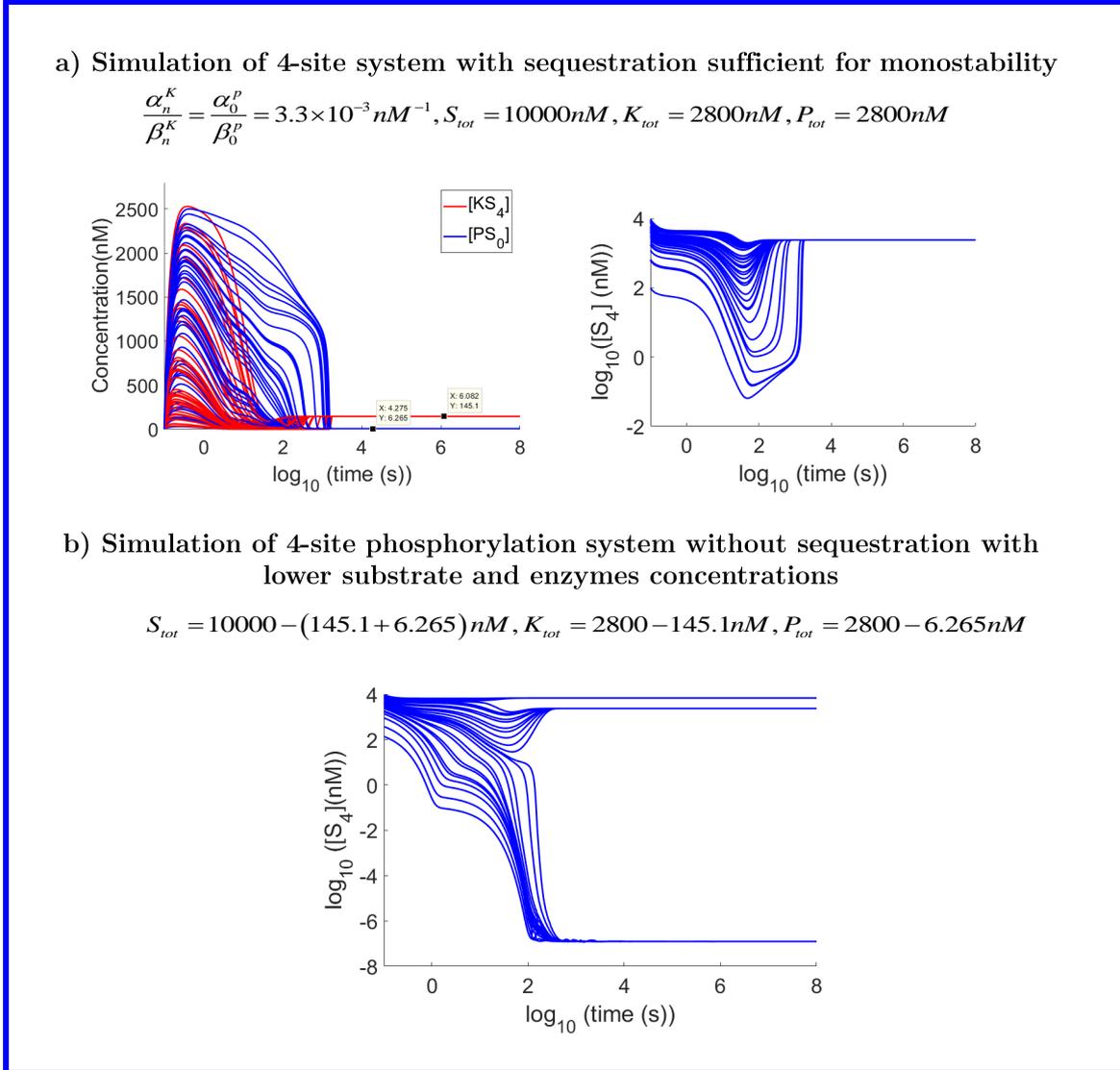


Fig. 6.3.3 Time courses from dynamic simulation of the system with different initial conditions, satisfying total substrate and enzyme concentrations. a) The 4-site original system simulated with sufficient sequestration for monostability to occur (right) ($\frac{\alpha_n^K}{\beta_n^K} = \frac{\alpha_0^P}{\beta_0^P} = 3.3 \times 10^{-3} \text{ nM}^{-1}$). The steady state concentrations of PS_0 and KS_4 due to sequestration are determined (left). b) The same system is simulated with lowering the total substrate and enzyme concentrations by those amounts. Tristability is not affected by that change.

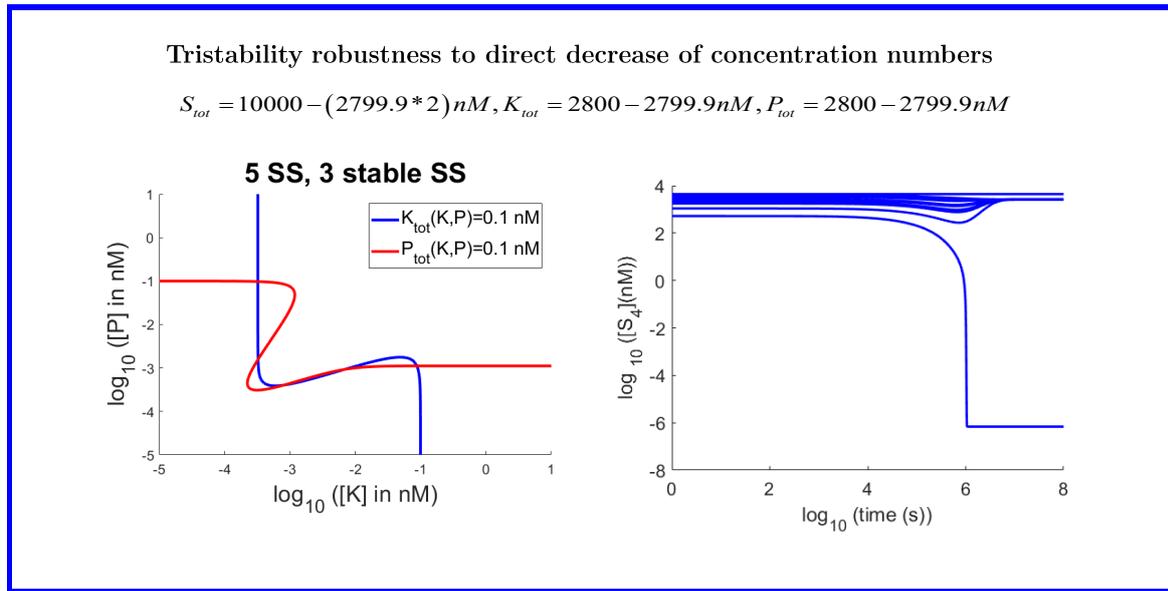


Fig. 6.3.4 Tristability is very robust to direct decrease of enzyme concentration numbers. Tristability is preserved even when 2799.9 nM from the 2800 nM of the total concentration of both the phosphatase and the kinase are removed directly (and therefore 5599.8 nM from the 10000 nM of the substrate concentration) without the substrate enzyme-sequestration mechanism.

from the intermediates than the kinase (i.e. low γ_1^P and γ_2^P , high γ_0^K and γ_1^K).

6.3.1.5 Substrate Enzyme-Sequestration effects are not necessarily limited to S_0 and S_n or even to multi-site protein phosphorylation

Having identified that Substrate Enzyme-Sequestration introduces self-inhibition which disrupts the mechanism that caused multistability, one can see that other inactive complexes might also limit the extent of multistability. For example, if the intermediate complex KS_2 is allowed to form an inactive complex KKS_2 by using an allosteric secondary site perhaps, it is essentially competing with S_2 for its own production. This effectively reduces the coupling provided via KS_2 . A small sequestration strength ($5 \times 10^{-1} \text{ nM}^{-1}$ in Fig. 6.3.5) results in monostability. Note though that not all inactive complexes would have this effect on the system. For example, if KS_2 is allowed to bind with phosphatase P to form PKS_2 , then this does not have the same impact on the coupling via KS_2 . Indeed, in simulations, such an inactive complex formation even with sequestration strengths of the order of 1×10^3 (i.e. 2000 times stronger affinity than before) did not affect the tristability of the original system.

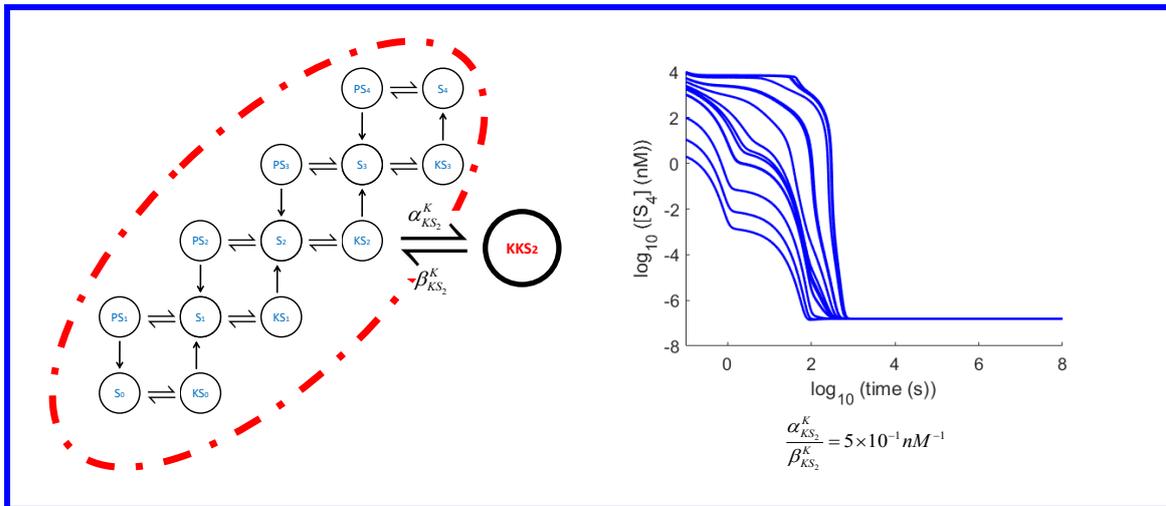


Fig. 6.3.5 Sequestration of kinase by a kinase-substrate intermediate (KKS_2 formation from KS_2) could also cause the system to lose its multistable behaviour, resulting to monostability.

The principle of enzyme competition when substrate is in excess is also prevalent in other enzyme-sharing schemes, for example where two substrates compete for the same kinase and phosphatase. To investigate this we took a one-site substrate model, which has been shown in the literature to exhibit bistability [45]. Following the conditions derived in that paper, we were indeed able to create bistability, which was then turned to monostability on the addition of Substrate Enzyme-Sequestration of strength $1 \times 10^{-1} nM^{-1}$. This is illustrated in Fig. 6.3.6. This can be explained along the same lines as before. For example, S_0 can be thought of as an inhibitor of Z_1 through competition with Z_0 for the kinase, while Z_1 is an inhibitor of S_0 through competition with S_1 for the phosphatase. The same applies to the pair Z_0 and S_1 as well. When S_0 , for example, is allowed to compete for the phosphatase in order to form the inactive complex PS_0 , it is self-inhibiting, gradually weakening the feedback loop with Z_1 .

6.3.1.6 Generalisation to arbitrary processivity and sequentiality

Since the new sequestration species added to the system do not interfere with its internal structure, we can in the same way extend the general framework of Thomson and Gunawardena [163], with arbitrary processivity and sequentiality (i.e. multiple phosphorylations or dephosphorylations can happen per reaction and in any order). The main conclusion, that increasing the strength of either kinase or phosphatase sequestration ultimately reduces the number of steady states to one, remains unchanged.

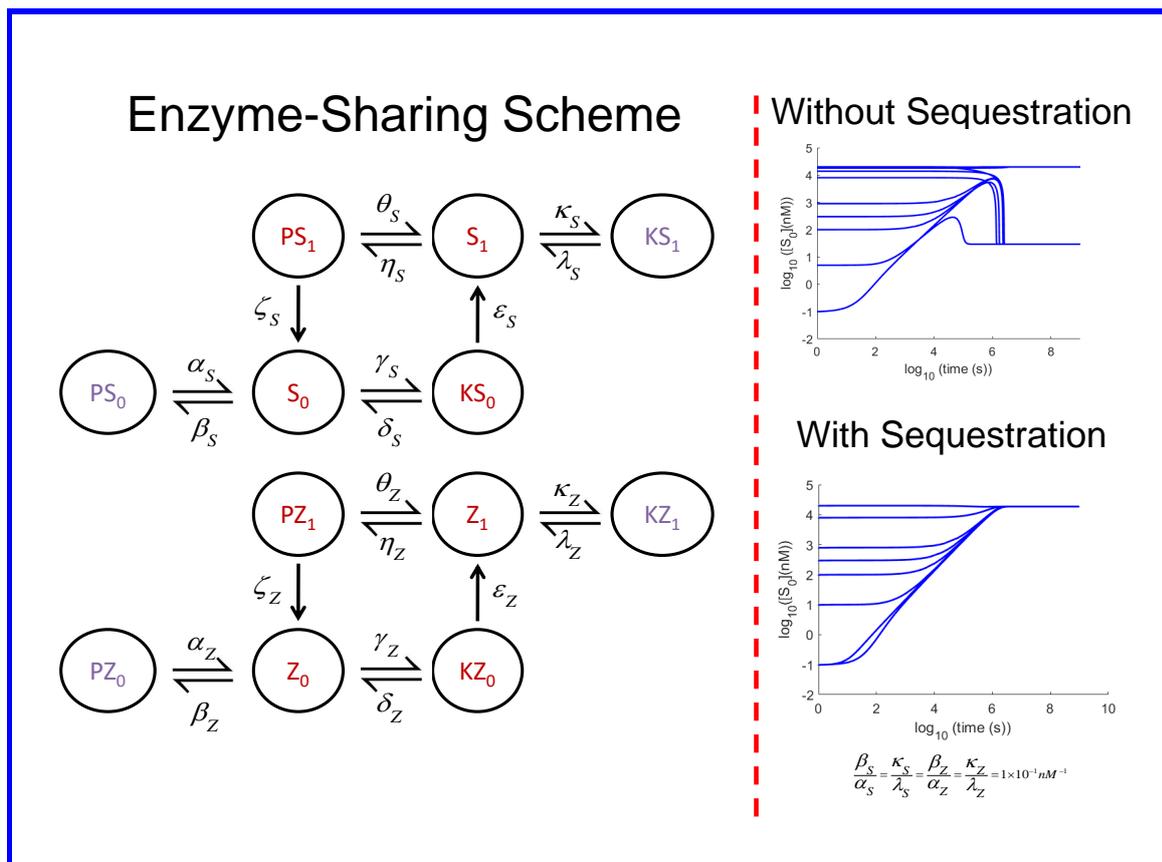


Fig. 6.3.6 Substrate Enzyme-Sequestration can limit bistability in a different enzyme-sharing mechanism as well, where two substrates with one phosphosite compete for the same kinase and phosphatase [45].

In this general framework the new three ϕ functions, are rational positive [163], allowing a rational expression in u , $Z(u)$, to be defined.

$$0 = (u - w)\phi_1 + [S_{\text{tot}}](u\phi_2 - w\phi_3) =: Z(u)$$

where ($w = \frac{[K_{\text{tot}}]}{[P_{\text{tot}}]}$).

It was shown that $Z(u)$ can be expressed as $\frac{R(u)}{Q(u)}$, where $Q(u)$ is an s-positive polynomial (sum of positive monomials). Therefore the steady states of the system can just be found by finding the roots of $R(u)$, with $N + 1$ now lying between $n + 1$ and 2^n depending on the model[163].

$$R(u) = a_{N+1}u^{N+1} + a_Nu^N + \dots + a_1u + a_0 \quad (6.3.2)$$

As before, the leading coefficient a_{N+1} is positive and the trailing coefficient a_0 is negative, (Appendix, Section 6.6.1.6). When phosphatase sequestration is added, the polynomial changes to

$$R'(u) = R(u) - [S_{\text{tot}}]w \frac{\alpha_0^P}{\beta_0^P} Q(u) \quad (6.3.3)$$

Since more than one coefficient is changed it not possible to use the Pratt tableau directly as before. However, since $Q(u)$ is s-positive, and so can't itself have any positive real roots by the Descartes' rule of signs, and its degree is less than that of $R(u)$, it must be the case that $R'(u)$ will have precisely one positive real root for sufficiently large $\frac{\alpha_0^P}{\beta_0^P}$ (Proof using root locus is found in the Appendix).

The same argument applies to kinase sequestration, by relabelling the fully phosphorylated substrate as S_0 and writing the polynomials in terms of $u^{-1} = [P]/[K]$ instead.

6.3.2 Substrate Enzyme-Sequestration in the Stochastic Domain

6.3.2.1 The behaviour of Substrate Enzyme-Sequestration when the molecule numbers are small require a different analysis

So far we have proved that increasing the strength of Substrate Enzyme-Sequestration in multi-site phosphorylation systems leads to the monotonic decrease of whatever multistability would be possible if the inactive complex formation (PS_0 and KS_n) was not considered, no matter what the kinetic parameters. Ultimately, this monotonic decrease leads to one steady state. However, this analysis was done in the deterministic domain, which is technically only valid in the limit of infinite molecule numbers. Therefore, to obtain a full understanding of the effect of the studied Substrate Enzyme-Sequestration is essential that an accompanied

analysis is done for the case when this assumption is not valid.

When molecule numbers are large but finite, bistability of the differential equations manifests itself as bimodality of the stochastic system. The modes correspond to the stable steady states of the system, and the system undergoes fluctuations within, and random jumps between, the modes. To illustrate this we use the original tristable system presented in [163, 162] (Appendix 6.6.1.7). Considering substrate-kinase and substrate-phosphatase sequestration, with respective strengths $\frac{\alpha_n^K}{\beta_n^K} = 1 \times 10^{-3} \text{ nM}^{-1}$ and $\frac{\alpha_0^P}{\beta_0^P} = 1 \times 10^{-3} \text{ nM}^{-1}$, bistability (three steady states, two stable) is obtained, as shown by the intersections of the contours for $[K_{\text{tot}}]$ and $[P_{\text{tot}}]$ in Fig. 6.3.7 (left). The result of the stochastic simulation with the same numerical parameters (including the ratios between enzymes and substrate) but with the parameters converted to units of molecules instead of units of concentration is also shown in Fig. 6.3.7 (right). The system can be seen to jump between the modes.

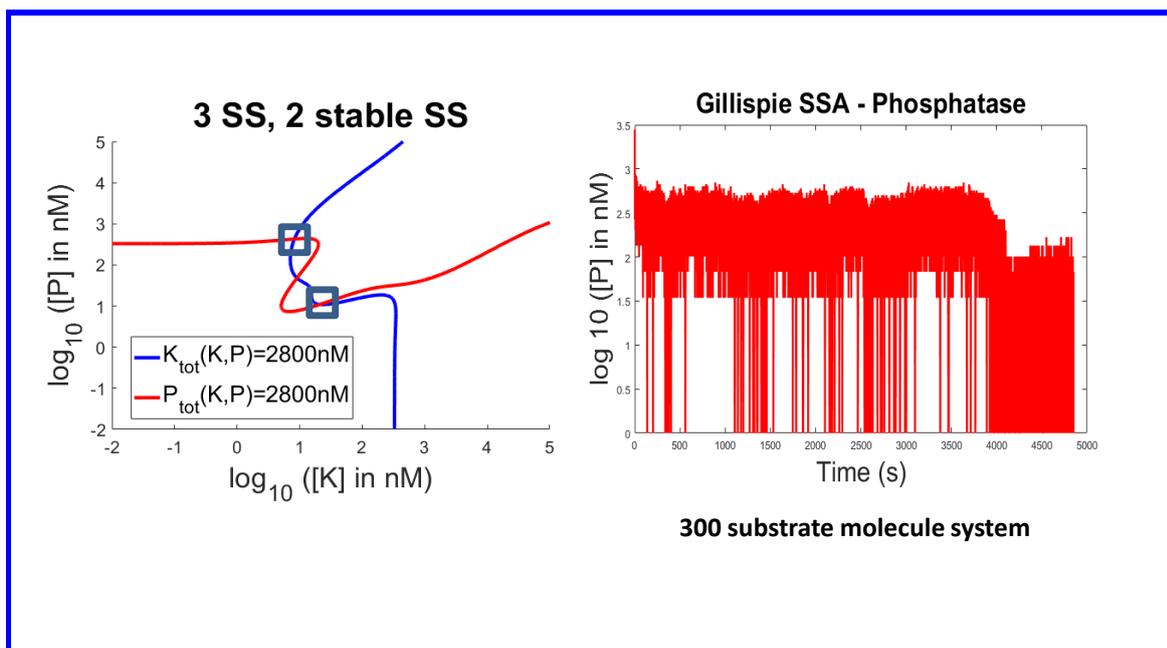


Fig. 6.3.7 Using the original tristable system [163, 162] investigating the effect of Substrate Kinase Sequestration and Substrate Phosphatase Sequestration (with respective strengths $\frac{\alpha_n^K}{\beta_n^K} = 1 \times 10^{-3} \text{ nM}^{-1}$ and $\frac{\alpha_0^P}{\beta_0^P} = 1 \times 10^{-3} \text{ nM}^{-1}$) in a system of 300 substrate molecules, corresponding to the original system in a volume of $4.98 \times 10^{-17} \text{ L}$. Bistability in the deterministic domain (left) manifests itself as bimodality in the stochastic domain (right). The simulation was done using the Gillispie Stochastic Simulation Algorithm (SSA) [56] .

6.3.2.2 The same strength of Substrate Enzyme-Sequestration, sufficient for monostability, can lead to both monomodality and bimodality in the stochastic domain, depending on the timescales of the individual sequestration parameters.

As we have already seen in Chapter 3, when molecule numbers are small, there may be little relationship between the continuous deterministic and discrete stochastic analyses [39, 107]. Understanding this type of ‘noise-induced’ bimodality is hard, yet even harder is the prediction of when this would take place [16]. Therefore, as our results demonstrate that Substrate Enzyme-Sequestration will ultimately lead to one steady state, it is imperative to check whether this mechanism has the same effect in the stochastic domain.

In order to investigate the effect of enzyme sequestration by the substrate in this regime we considered a 15-substrate molecule single phosphosite system, using the same parameters, wherever applicable, as in the original tristable system (in a volume of 2.49×10^{-18} L). Four kinase and four phosphatase molecules (thus having the same substrate/enzyme ratios as before) were selected. To examine whether the predicted monostability is obtained, we use a strength of sequestration found in earlier sections to be sufficient for monostability ($\frac{\alpha_n^K}{\beta_n^K} = \frac{\alpha_0^P}{\beta_0^P} = 5 \times 10^{-3} \text{ nM}^{-1}$). For the same sequestration strength, two different behaviours emerged, depending on the timescale of the kinetic parameters used. This is different to the deterministic case, where the steady states are only dependent on the ratio. For $\beta_n^K = \beta_0^P = 1 \times 10^{-1} \text{ s}^{-1}$, the result was a monomodal probability distribution, agreeing with the prediction from the deterministic analysis. However, when $\beta_n^K = \beta_0^P = 1 \times 10^{-3} \text{ s}^{-1}$, for the same sequestration strength, a bimodal behaviour emerged, as illustrated in Fig. 6.3.8. Note that, as shown in Appendix Section 6.6.1.7, β_i^P and β_i^K are of the order of 10^{-3} to 10^{+0} .

This behaviour shows that the extra mode created is dependent on the time required to get out of that state i.e. the dwell time. This is because the free kinase is completely depleted, as it is trapped in intermediate complexes, leaving only free phosphatase around. In the absence of kinase and in the presence of just phosphatase, the substrate is kept in a mode where it is only unphosphorylated. This mechanism cannot be represented in the deterministic analysis, as the concentration of the enzymes never becomes exactly zero.

Nevertheless, this mechanism could serve a specific function when a system requires different behaviour when its size expands, as we are later going to see in Chapter 7. In smaller sized systems, it could be beneficial that both unphosphorylated and phosphorylated substrate molecules are available, triggering different cascades of reactions. When the system becomes large however, one of the two mechanisms might be more beneficial.

Bimodality is induced when the kinase becomes extinct for a period of time, allowing the

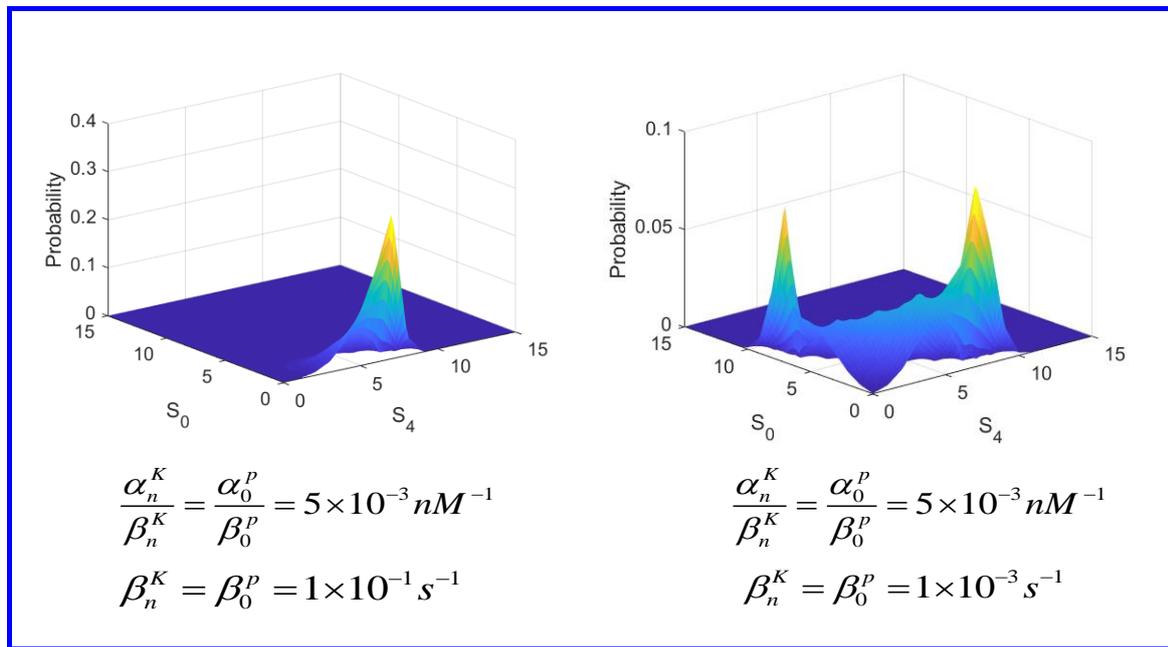


Fig. 6.3.8 In the stochastic domain, the same strength of sequestration can produce both monomodal (left), and bimodal (right) behaviour, depending on the individual timescales of the sequestration parameters. The ratio, unlike in the deterministic case, is not sufficient to deduce behaviour.

phosphatase without competition to completely dephosphorylate the available substrates. This mechanism does not appear to be dependent on the number of available phosphosites, therefore higher orders of multimodality due to this mechanism are not expected (and we were unable to find any). Nevertheless, it is possible to induce it with only one available phosphosite, impossible when the system is analysed deterministically, as we show in the next section using the framework developed in Chapter 4. This provides a methodology to characterise parameter regimes where bimodality can be expected.

6.4 Bimodality is feasible even when only one phosphosite is available

In the previous section we noted that Substrate Enzyme-Sequestration and manipulation of dwell times is enough to create bimodality. This intuition is tied very well with the stochastic framework developed in Chapter 4 which can allow bimodality be investigated in a methodological approach. Here, we use that framework to find the parameter regime where bimodality can be induced in the presence of only one available phosphosite.

To remind ourselves, an accurate stochastic framework that does not depend on simulations is the discrete Chemical Master Equation, represented by a discrete state continuous time Markov process [108]. The microstate of the system involving n species is defined as $\mathbf{x}(t) = \{x_1(t), x_2(t), \dots, x_n(t)\} \in \mathbb{N}^n$. A microstate is therefore describing a possible combination of the different population numbers of each molecular species in the system.

We have already seen that the stationary probability distribution \mathbf{P}_s of a finite state Chemical Master Equation can be calculated by solving the following equation

$$\mathbf{A}\mathbf{P}_s = 0 \quad (6.4.1)$$

The analysis of a stochastic system using the Chemical Master Equation framework requires firstly the conversion of the reaction scheme into a microstate grid. Fig. 6.4.1 shows how this is done for a phosphorylation system in the illustrative example of just two substrate molecules for a system with a single phosphosite. This example assumes excess enzyme, therefore the microstates include all possible permutations, as there is no extra constraint. When substrate is in excess (e.g. there is only one kinase molecule and two substrate molecules), the microstates representing enzyme complexes (e.g. KS_0) greater than the total number of the corresponding enzyme (e.g. K) have to be deleted from the grid, as it is now not possible to have two KS_0 molecules.

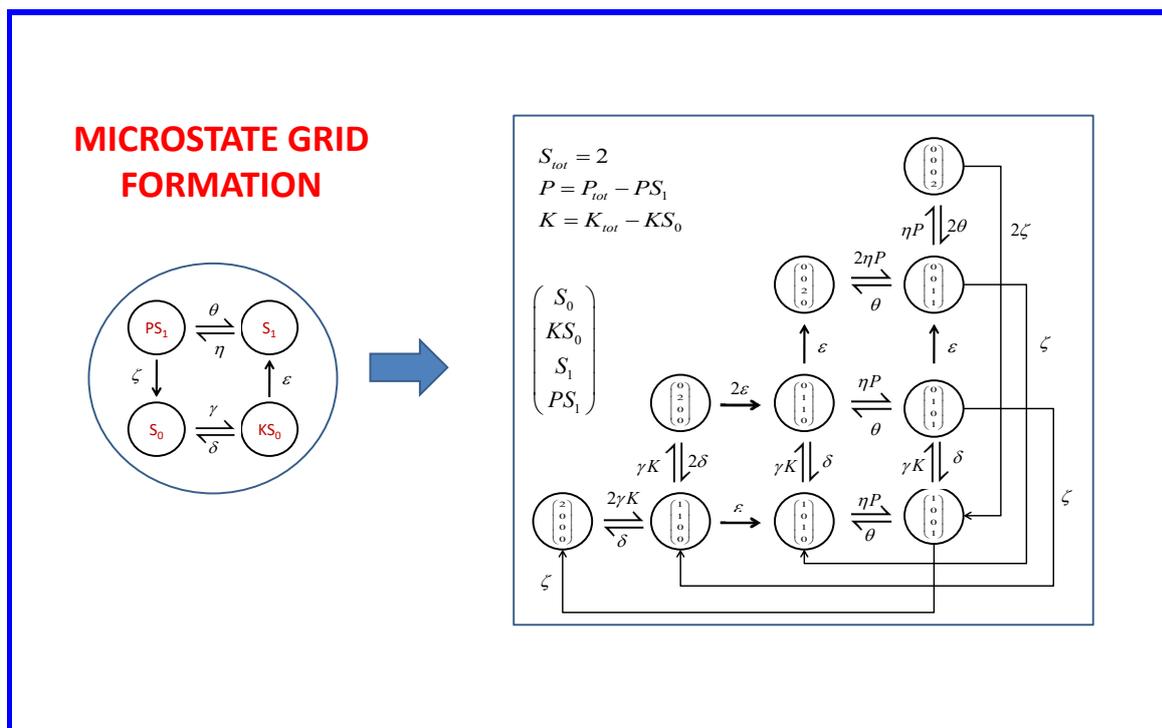


Fig. 6.4.1 This illustrative example with just two substrate molecules and one available phosphorylation site shows how a reaction scheme is converted to a grid of microstates. This example assumes excess enzyme, therefore the microstates include all possible permutations, as there is no extra constraint.

Defining $\mathbf{A}_{j,j}^D$ to be the sub-matrix formed after deleting the j^{th} row and j^{th} column from matrix \mathbf{A} and $\mathbf{C}_j = -\mathbf{A}_{j,j}^D$, in Chapter 4 we saw that we can use the ratio of $\frac{\lambda_{min}(\mathbf{C}_j^T)}{|a_{jj}|}$ to investigate the effect of the different parameters on the formation of the stationary probability distribution. Note that as this was just derived from the lower bound, this is used as an indication/ heuristic for stationary probability distribution design purposes. The parameters found by this heuristic are then tested by explicitly solving the Chemical Master Equation, to verify that the stationary probability distribution is indeed the one desired. The sum of these ratios $\frac{\lambda_{min}(\mathbf{C}_j^T)}{|a_{jj}|}$, to be referred from now on as characterisation ratios, of the microstates corresponding to a particular substrate state can provide a fast and relatively accurate measure of how the stationary probability distribution for the substrate states varies with different macroscopic parameters (reaction rates). The same 15-substrate molecule single phosphosite system was investigated as before (in a volume of 2.49×10^{-18} L), yet with only one phosphosite. Four kinase and four phosphatase molecules were again selected. This provides

us, as expected from the deterministic analysis, with only one mode at $(S_0, S_1) = (11, 0)$. This is illustrated in Fig 6.4.2.

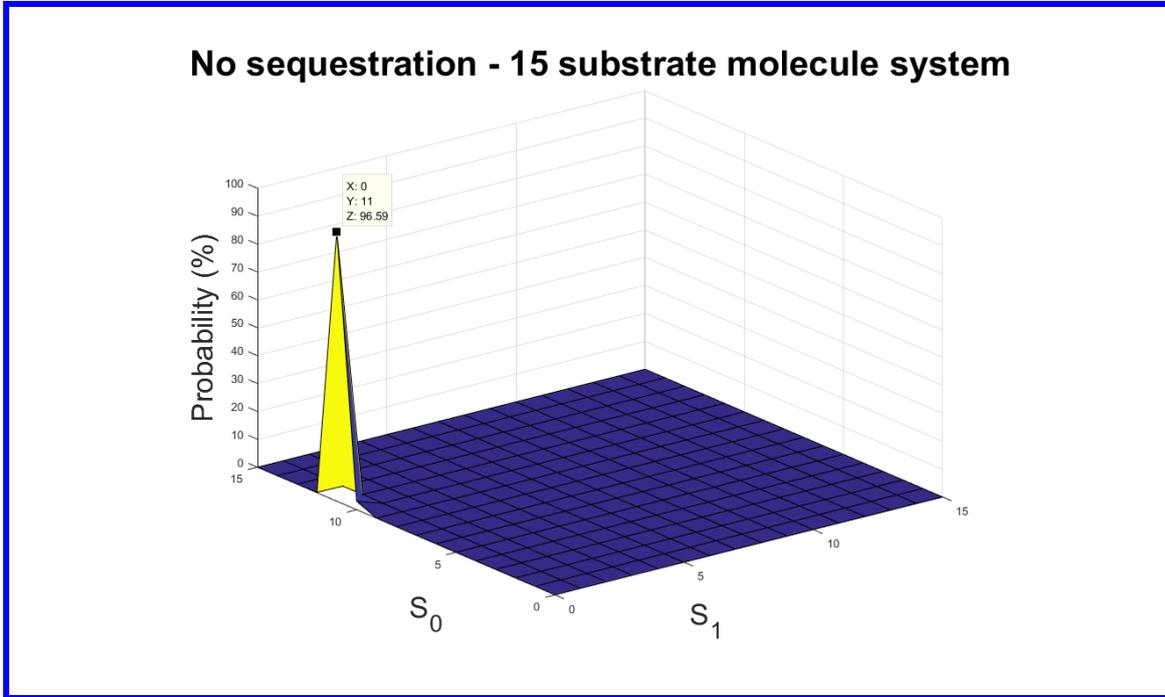


Fig. 6.4.2 No sequestration leads to monomodality for a 15-substrate molecule single phosphosite system, with four kinase and four phosphatase molecules. This is consistent the monostability expected using the deterministic analysis.

The tool developed allows the investigation to take place without considering all the substrate states (S_0, S_1) of the system. Instead, we can initially focus in just some of them. As the original mode $(S_0, S_1) = (11, 0)$ is found at the boundary $S_1 = 0$, we include the boundary states $(S_0, S_1) = (6, 0) - (11, 0)$, in our analysis. As we aim for bimodality, we also include their reciprocal states on the other boundary $S_0 = 0$, $(S_0, S_1) = (0, 6) - (0, 11)$. Finally, we also include some states in between to establish that they do not become more dominant than the ones on the boundaries, e.g. $(S_0, S_1) = (4, 3)$ or $(3, 4)$.

The first step is to set all the parameters of the reaction scheme, as shown in Appendix, Section 6.6.1.7, letting only the sequestration parameter under investigation to be variable. This is α , as shown in the reaction scheme of Fig. 6.4.3. Note that we vary this parameter (which captures the dwell time of the extra mode) instead of the ratio of sequestration strength, following the observation of Fig. 6.3.8. The non-sequestration parameters are the same as the ones in the multisite protein phosphorylation system by Thomson and Gunawardena [163, 162].

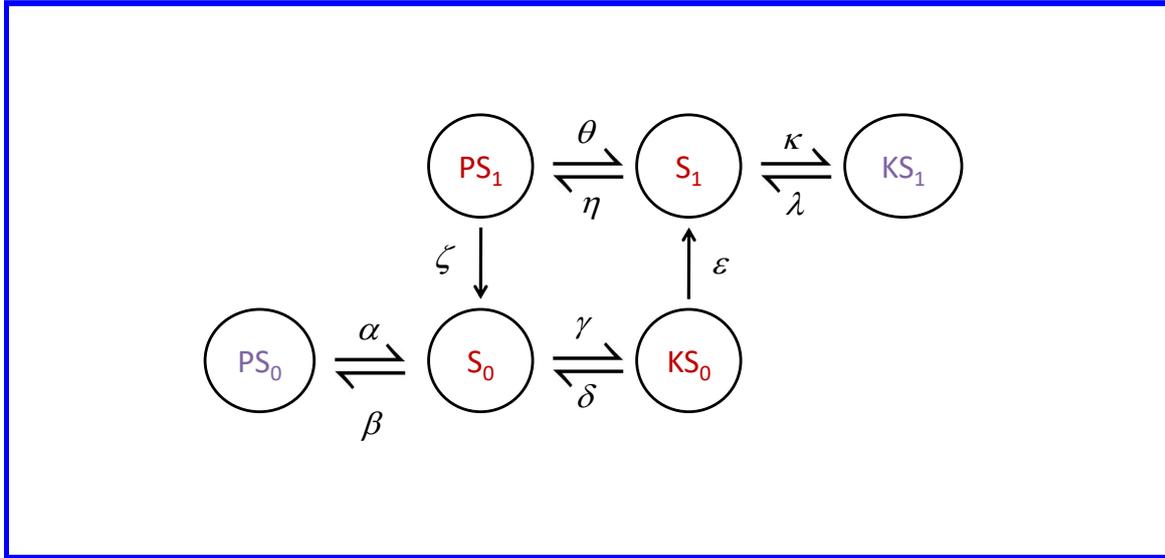


Fig. 6.4.3 One phosphosite reaction scheme with inactive complexes PS_0 and KS_1 added

The next step is to create a design table (bottom left of Fig 6.4.4) using the sum the characterisation ratios $\frac{\lambda_{\min}(\mathbf{C}_J^T)}{|a_{jj}|}$ for the M microstates corresponding to each substrate state as we vary α . The design table allows us to estimate the region of values of α that can allow bimodality. The result is shown in Fig 6.4.4, where it is found that at $\alpha = 10^{-2}$, bimodality can be obtained, as verified by explicitly finding the stationary probability distribution (by solving the Chemical Master Equation), creating modes at $(S_0, S_1) = (8, 0), (0, 7)$ (bottom right). As mentioned before, the numerator of the characterisation ratio, which is the minimum eigenvalue of the developed WCDD M-matrix, provides a network input metric, whereas the denominator, $|a_{jj}|$ represents the local output propensities of the particular microstate. The top left and top right design tables illustrate the sum of the numerators and the sum of the inverse of the denominators of the characterisation ratios respectively. From these we can see that the main driver making $(S_0, S_1) = (8, 0)$ a mode is the effect of the inputs on the network-level, whereas the main driver making $(S_0, S_1) = (0, 7)$ a mode is the low local output propensities of its corresponding microstates.

6.5 Conclusion

In this chapter we first identified the effect of enzyme docking, and the Substrate Enzyme-Sequestration it implies, in the presence of excess substrate and in the regime of large molecule numbers, proving that increasing the strength of sequestration the extent of multista-

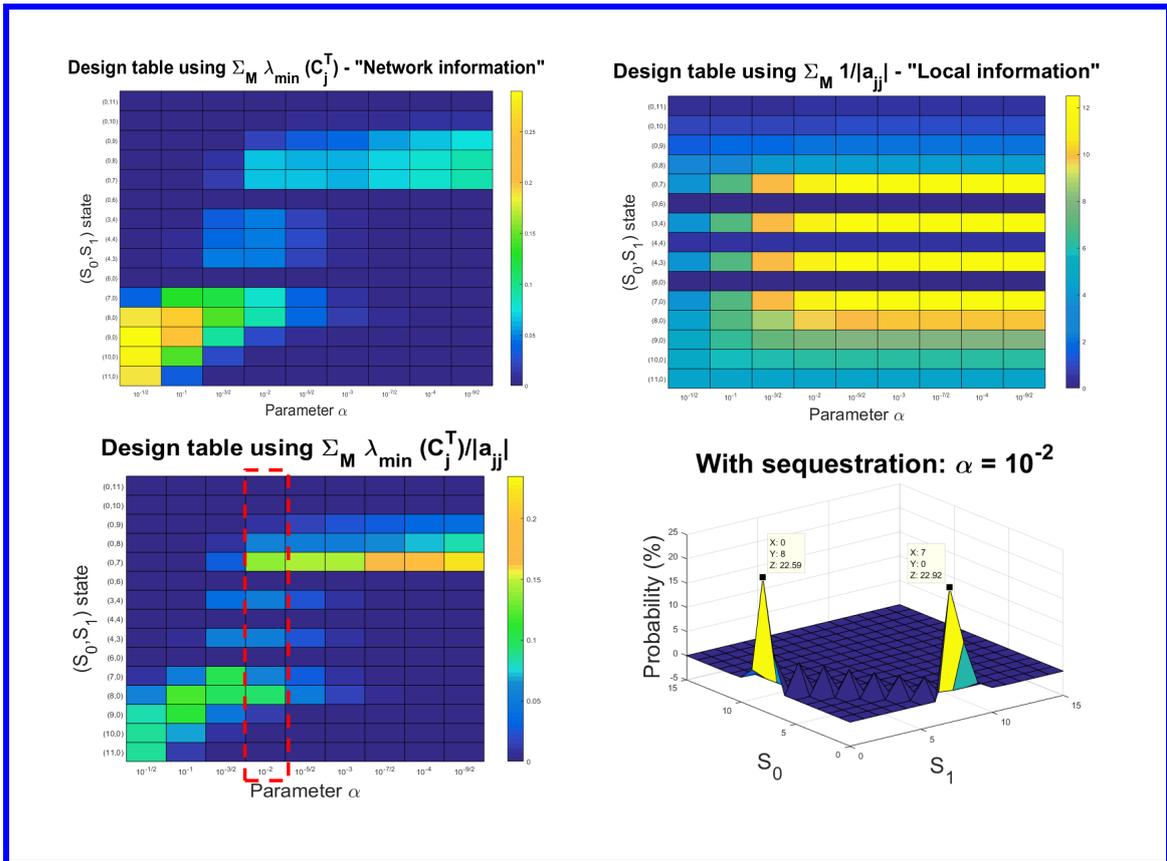


Fig. 6.4.4 The sum of the characterisation ratios corresponding to the microstates of different substrate states under investigation (bottom left) can be used to investigate the parameter regime of α for bimodality to occur. The top left and top right graphs illustrate the network input and local output effects, allowing for greater insights behind the creation of the two modes. The main driver making $(S_0, S_1) = (8, 0)$ a mode is the effect of the inputs on the network-level, whereas the main driver making $(S_0, S_1) = (0, 7)$ a mode is the low local output propensities of its corresponding microstates.

bility is limited and ultimately reduced down to one steady state. Secondly, we explored the mechanism's effect in the presence of small molecule numbers. For the latter, the analysis was naturally placed in the stochastic domain. For that, we note that the sequestration strength, represented as a ratio, cannot provide accurate predictions by itself of the behaviour of the system. We found that the individual dwell times as compared to the spectral properties of the rest of the network need to be considered to identify the behaviour of sequestration. This observation is formalised by the mathematical framework presented in Chapter 4, allowing for a methodology in identifying when bimodality is feasible in the small numbers regime, even when bistability is even deemed as impossible using deterministic analysis.

6.6 Appendix

6.6.1 Deterministic Framework

6.6.1.1 Theoretical development of the conditions on limiting the extent of multistability

Lemma 6.6.1. *If $\frac{1}{1 + \frac{\alpha_n^K}{\beta_n^K} [S_{tot}]} < \frac{2(n+1)a_{n+1}a_{n-1}}{na_n^2}$ or $\frac{1}{1 + \frac{\alpha_0^P}{\beta_0^P} [S_{tot}]} < \frac{2(n+1)a_0a_2}{na_1^2}$ then the number of real steady states is less than or equal to $n-1$.*

Proof. Suppose that $\frac{1}{1 + \frac{\alpha_n^K}{\beta_n^K} [S_{tot}]} < \frac{2(n+1)a_{n+1}a_{n-1}}{na_n^2}$ and that all $n+1$ roots of Eq 6.3.1 are real.

Let x_i represent the i^{th} root of Eq. 6.3.1. Then by the Vieta Formulae [59],

$$\begin{aligned} \left(\frac{a'_n}{a'_{n+1}}\right)^2 &= \left(\sum_i x_i\right)^2 = \sum_i x_i^2 + 2 \sum_{i<j} x_i x_j \\ \Rightarrow \sum_i x_i^2 &= \left(\frac{a'_n}{a'_{n+1}}\right)^2 - 2 \sum_{i<j} x_i x_j \\ \Rightarrow \sum_i x_i^2 &= \left(\frac{a'_n}{a'_{n+1}}\right)^2 - 2 \frac{a'_{n-1}}{a'_{n+1}} \end{aligned}$$

Then by the Quadratic Mean - Arithmetic Mean and the Triangle Inequalities,

$$\begin{aligned} \sqrt{\frac{\sum_i x_i^2}{n+1}} &\geq \frac{\sum_i |x_i|}{n+1} \geq \frac{|\sum_i x_i|}{n+1} \\ \Rightarrow \frac{n}{2} \sum_i x_i^2 &\geq \sum_{i<j} x_i x_j \end{aligned}$$

Using the obtained Vieta Formulae,

$$\begin{aligned} \frac{n}{2} \left[\left(\frac{a'_n}{a'_{n+1}}\right)^2 - 2 \frac{a'_{n-1}}{a'_{n+1}} \right] &\geq \frac{a'_{n-1}}{a'_{n+1}} \\ \Rightarrow a_n^2 &\geq 2a'_{n+1}a_{n-1} \left(\frac{n+1}{n}\right) \end{aligned}$$

Expressing it in terms of the old coefficients,

$$a_n^2 \geq 2a_{n+1}a_{n-1} \left(1 + [S_{tot}] \frac{\alpha_n^K}{\beta_n^K}\right) \left(\frac{n+1}{n}\right)$$

$\Rightarrow \frac{1}{1 + \frac{\alpha_n^K}{\beta_n^K} [S_{tot}]} \geq \frac{2(n+1)a_{n+1}a_{n-1}}{na_n^2}$ which is a contradiction. Therefore the roots x_i cannot all

be real. Thus, due to the expectation of at least one conjugate pair of complex roots, the maximum possible number of positive real roots is less than or equal to $n-1$.

Similarly, $\frac{1}{1 + \frac{\alpha_0^P}{\beta_0^P} [S_{tot}]} < \frac{2(n+1)a_0a_2}{na_1^2}$ is proved by considering the real roots of $P(v)$, $v = \frac{1}{u}$,

$$P(v) = a'_{n+1} + a_n v + \dots + a_1 v^n + a'_0 v^{n+1}.$$

Note that instead of using the Vieta formulae, one could also use a general result proved using differential calculus in Section 4.3 of Hardy's 'Inequalities' book [69].

□

Theorem 6.6.2. *If any of the following conditions are satisfied, then the number of positive steady states will be no more than $n - 1$ if n is even, or $n - 2$ if n is odd:*

1. $a_{n-1} \leq 0$ and $a_2 \geq 0$
2. $a_{n-1} > 0$ and $\frac{\alpha_n^K}{\beta_n^K} [S_{tot}] > \frac{na_n^2 - 2(n+1)a_{n+1}a_{n-1}}{2(n+1)a_{n+1}a_{n-1}}$ or
3. $a_2 < 0$ and $\frac{\alpha_0^P}{\beta_0^P} [S_{tot}] > \frac{na_1^2 - 2(n+1)a_0a_2}{2(n+1)a_0a_2}$

Proof. From Eq. 6.3.1, $a_{n+1} > 0$ and $a_0 < 0$.

For even n , there is a maximum of an even number ($n + 2$) of coefficients, therefore a maximum of $n + 1$ sign changes is possible. As $a_0 < 0$, then if $a_2 \geq 0$ then the last three coefficients, a_0 , a_1 and a_2 , can exhibit a maximum of one sign change. Knowing that there are a maximum of n coefficients from a_{n+1} to a_2 , n being even and $a_{n+1} > 0$, $a_2 \geq 0$, then they can exhibit a maximum of $n - 2$ sign changes. Therefore the total number of sign changes that can be exhibited is equal to $n - 1$. From the Descartes' rule of signs [174] the result follows immediately. Similarly for $a_{n-1} \leq 0$.

For odd n , there is a maximum of an odd number ($n + 2$) of coefficients, and as $a_{n+1} > 0$ and $a_0 < 0$, a maximum of n sign changes is possible. As $a_0 < 0$, then if $a_2 \geq 0$ then the last three coefficients, a_0 , a_1 and a_2 , can exhibit a maximum of one sign change. Similarly for the first three coefficients, $a_{n+1} > 0$, a_n and a_{n-1} , for $a_{n-1} \leq 0$. Knowing that there are a maximum of $n - 2$ coefficients from a_{n-1} to a_2 , n being odd, and $a_{n-1} \leq 0$, $a_2 \geq 0$, then they can exhibit a maximum of $n - 4$ sign changes. Therefore the total number of sign changes that can be exhibited is equal to $n - 2$. From the Descartes' rule of signs the result follows immediately.

The last two parts of the Theorem are directly derived from Lemma 6.6.1. For odd n , as $a_{n+1} > 0$ and $a_0 < 0$, there can be a maximum of n sign changes. The next number of sign changes possible is $n - 2$. Therefore, from the Descartes' rule of signs it is not possible to have $n - 1$ positive roots, as the parity of positive real roots and number of sign changes must be the same. The maximum therefore number of positive roots drops to $n - 2$.

□

6.6.1.2 Representing the derived conditions for limitation of multistability with kinetic parameters

To obtain a better intuition of the conditions derived in the previous section, we expand one of them in its corresponding kinetic parameters. Then we try to approximate the

expression until we obtain a meaningful simple form. The approximation is valid for the parameters of the model (as in Section 6.6.1.7) i.e for equal concentrations of kinase and phosphatase ($w = 1$), S_{tot} being large and $\gamma_0^K \ll \gamma_1^P$, $\gamma_2^P \ll \gamma_1^K$). The condition we use here is the one associated with the inactive complex PS_0 . The condition for the limitation of multistability in that case is that $a_2 < 0$ and $\frac{\alpha_0^P}{\beta_0^P} > \frac{na_1^2 - 2(n+1)a_0a_2}{2(n+1)a_0a_2[S_{tot}]}$.

$$\frac{a_1^2}{a_0a_2} = \frac{\alpha_2^P \gamma_2^P (\beta_1^K + \gamma_1^K) (\alpha_1^P \gamma_1^P (\beta_0^K + \gamma_0^K + a_0^K [S_{tot}]) - w \alpha_0^K \gamma_0^K (\beta_1^P + \gamma_1^P + a_1^P [S_{tot}]))^2}{w \alpha_0^K \alpha_1^P \gamma_0^K \gamma_1^P (\beta_0^K + \gamma_0^K) (\beta_1^K + \gamma_1^K) (-\alpha_2^P \gamma_2^P (\beta_1^K + \gamma_1^K + \alpha_1^K [S_{tot}]) + \alpha_1^K \gamma_1^K w (\beta_2^P + \gamma_2^P + \alpha_2^P [S_{tot}]))}$$

Using that S_{tot} is large,

$$\begin{aligned} \frac{a_1^2}{a_0a_2} &\approx \frac{\beta_1^K + \gamma_1^K}{(\beta_0^K + \gamma_0^K)(\beta_1^P + \gamma_1^P)} \frac{\gamma_2^P}{\gamma_0^K \gamma_1^P} \frac{\alpha_1^P \alpha_0^K}{\alpha_1^K} \frac{(\gamma_1^P - \gamma_0^K w)^2}{\gamma_1^K w - \gamma_2^P} \frac{[S_{tot}]}{w} \\ &= \frac{k_1^K}{k_0^K k_1^P} \frac{\gamma_2^P}{\gamma_0^K \gamma_1^P} \frac{(\gamma_1^P - \gamma_0^K w)^2}{\gamma_1^K w - \gamma_2^P} \frac{[S_{tot}]}{w} \end{aligned}$$

Using that $w = 1$ and $\gamma_0^K \ll \gamma_1^P$, $\gamma_2^P \ll \gamma_1^K$ (from Section 6.6.1.7),

$$\begin{aligned} \frac{na_1^2 - 2(n+1)a_0a_2}{2(n+1)a_0a_2[S_{tot}]} &= \frac{na_1^2}{2(n+1)a_0a_2[S_{tot}]} - \frac{1}{[S_{tot}]} \\ &\approx \frac{na_1^2}{2(n+1)a_0a_2[S_{tot}]} \\ &\approx \frac{n}{2(n+1)} \frac{k_1^K}{k_0^K k_1^P} \frac{\gamma_1^P \gamma_2^P}{\gamma_0^K \gamma_1^K} \approx 5 \times 10^{-3} nM^{-1} \end{aligned}$$

The exact result is $4.88 \times 10^{-3} nM^{-1}$.

6.6.1.3 Results based on the Pratt's tableau

Given any real polynomial, $P(x) = \sum_{j=0}^n c_j x^j$, a family of upper bounds on the number of roots in $(0, \infty)$ can be obtained using a Pratt tableau [137]. This can be created as follows:

Start with

$$\begin{aligned} c_{n+1-j,j} &= c_j, \quad 0 \leq j \leq n \\ c_{i,0} &= c_0 \text{ for all } i \\ \text{and } c_{0,j} &= 0 \text{ for } j > n \end{aligned}$$

Then by successive additions,

$$c_{i,j} = c_{i,j-1} + c_{i-1,j} \text{ for } i \geq 1, j \geq 1, i+j > n+1$$

Definition 6.6.3. [137] An allowed path is any path through the tableau of the $c_{i,j}$ going from any $c_{i,0}$ to any $c_{1,j}$ with $j \geq n$ by horizontal, vertical and diagonal steps.

Theorem 6.6.4. [137] *The number of changes of sign (ignoring 0's) along any allowed path through the tableau generated by the coefficients of $P(x)$ is an upper bound on the number of roots of $P(x)$ in the relevant domain $(0, \infty)$, including multiplicities.*

Fact 6.6.5. *In the Pratt's tableau of any real polynomial, $P(x) = \sum_{j=0}^n c_j x^j$, the coefficient of c_j in the summation required for any entry k in the column j of the tableau is equal to 1.*

Theorem 6.6.6. *Let row m be the first row in which all elements of columns 0 to $n - 1$ of the Pratt tableau of any real polynomial $P(x) = \sum_{j=0}^n c_j x^j$ with $c_0 > 0$ and $c_n < 0$ become positive. Let y be the n^{th} column of the tableau considering the first $(m - 1)$ rows. Then if $\delta \geq \frac{-\max(y)}{c_n}$, then the real polynomial $P(x) + c_n \delta$ has fewer than 2 roots in the $(0, \infty)$ domain.*

Proof. Using the Pratt's tableau, if $c_0 > 0$, all rows will eventually become positive due to the successive additions $c_{i,j} = c_{i,j-1} + c_{i-1,j}$, $i \geq 1, j \geq 1$. Let the m^{th} row denote the row in which all elements of columns 0 to $n - 1$ become positive. If there is no sign change in the n^{th} column up to the $(m - 1)^{\text{th}}$ row, then the upper bound obtained by Theorem 6.6.4, is equal to one, as there will be no sign change across the m^{th} row from $c_{m,0}$ to $c_{m,n-1}$ before having one sign change moving from $c_{m,n-1}$ to $c_{1,n}$. Creating an initial Pratt tableau of $P(x) = \sum_{j=0}^n c_j x^j$ with $c_0 > 0$ and $c_n < 0$ and letting y be the tableau's n^{th} column up to the $(m - 1)^{\text{th}}$ row, then

From Fact 6.6.5,

$$\max(n^{\text{th}} \text{ column}) = \max(y) = c_n + V, \text{ where } V \in \mathbb{R}$$

For the n^{th} column up to the $(m - 1)^{\text{th}}$ row of the Pratt tableau of the real polynomial $P(x) + c_n \delta$ to always be negative or zero,

$$\begin{aligned} c_n(1 + \delta) + V &\leq 0, \\ c_n(1 + \delta) &\leq -V \\ c_n(1 + \delta) &\leq c_n - \max(y) \\ \Rightarrow \delta &\geq \frac{-\max(y)}{c_n}, \text{ as } c_n < 0. \end{aligned}$$

□

Corollary 6.6.7. *Let row m be the first row in which all elements of columns 0 to $n - 1$ of the Pratt tableau of any real polynomial $P(x) = \sum_{j=0}^n c_j x^j$ with $c_0 < 0$ and $c_n > 0$ become negative. Let y be the n^{th} column of the tableau considering the first $(m - 1)$ rows. Then if $\delta \geq \frac{-\min(y)}{c_n}$, then the real polynomial $P(x) + c_n \delta$ has fewer than 2 roots in the $(0, \infty)$ domain.*

Proof. Considering the real roots of $-P(x)$, the result follows directly from Theorem 6.6.6.

□

Corollary 6.6.8. *Let row m be the first row in which all elements of columns 0 to $n - 1$ of the Pratt tableau of any real polynomial $P'(x) = \sum_{j=0}^n c_{n-j}x^j$ with $c_n > 0$ and $c_0 < 0$ become positive. Let y be the n^{th} column of the tableau, considering the first $(m - 1)$ rows. Then if $\delta \geq \frac{-\max(y)}{c_0}$, then the real polynomial $P(x) = \sum_{j=0}^n c_j x^j + c_0 \delta$ has fewer than 2 roots in the $(0, \infty)$ domain.*

Proof. As the relevant domain of roots is $(0, \infty)$, considering the real roots of $P'(v)$, $v = \frac{1}{x}$ the result follows directly from Theorem 6.6.6. \square

Theorem 6.6.9. *For any $\delta^K \geq 0$ there exists δ^P , directly computable from the rate constants, such that if $\frac{\alpha_n^K}{\beta_n^K} [S_{\text{tot}}] = \delta^K$ and $\frac{\alpha_0^P}{\beta_0^P} [S_{\text{tot}}] \geq \delta^P$ then the polynomial $R'(u)$ has precisely one positive root, corresponding to one steady state. Similarly, for any $\delta^P \geq 0$ there exist a δ^K with the same properties.*

Proof. Comparing Corollary 6.6.8 with Eq. 6.3.1, $c_n = a_{n+1} \left(1 + [S_{\text{tot}}] \frac{\alpha_n^K}{\beta_n^K} \right)$. Setting $\delta^K = [S_{\text{tot}}] \frac{\alpha_n^K}{\beta_n^K} \geq 0$, then $c_n > 0$.

Let $c_0 = a_0$. Then $R'(u)$ in Eq. 6.3.1 can be expressed as $R(u) + c_0 \delta^P$, where $c_0 < 0$. Then it follows directly from Corollary 6.6.8 that a $\delta^P = [S_{\text{tot}}] \frac{\alpha_0^P}{\beta_0^P} \geq \frac{-\max(y)}{c_0}$ can be calculated with finite summations from the Pratt tableau, which guarantees that there is at most one real positive root.

In a real polynomial $P(x) = \sum_{j=0}^n c_j x^j$ with $c_n > 0$ and $c_0 < 0$, the number of sign changes has to be odd. From the Descartes' rule of signs the parity of positive real roots and the number of sign changes must be the same. The minimum therefore number of positive roots is 1. Therefore, if the conditions are satisfied there is precisely one positive real root. \square

6.6.1.4 Algorithm for finding δ^K using the Pratt tableau

Result: Output δ^K

Input real polynomial $P(x) = \sum_{j=0}^n c_j x^j$ with $c_0 < 0$ and $c_n > 0$;

Multiply all coefficients c_j by -1;

$c_{n+1-j,j} = c_j, 0 \leq j \leq n$;

$k = 1$;

while (any $c_{k-1,1:n-1} < 0$) OR ($k - 1 < n$) **do**

if ($k > n$) **then**

 | $c_{k,0} = c_0$;

end

for $j = 1 : n$ **do**

if ($k + j > n + 1$) **then**

 | $c_{k,j} = c_{k,j-1} + c_{k-1,j}$

end

end

$k = k + 1$;

end

$m = \max j$ such that $c_{j,n-1} < 0$;

$y = c_{1:m,n}$;

$\delta^K = \frac{-\max(y)}{c_n}$;

6.6.1.5 Algorithm for finding δ^P using the Pratt tableau

Result: Output δ^P

Input real polynomial $P(x) = \sum_{j=0}^n a_j x^j$ with $a_0 < 0$ and $a_n > 0$;

Flip coefficients such that $c_{n-j} = a_j$;

$c_{n+1-j,j} = c_j, 0 \leq j \leq n$;

$k = 1$;

while (any $c_{k-1,1:n-1} < 0$) **OR** ($k - 1 < n$) **do**

if ($k > n$) **then**

$c_{k,0} = c_0$;

end

for $j = 1 : n$ **do**

if ($k + j > n + 1$) **then**

$c_{k,j} = c_{k,j-1} + c_{k-1,j}$

end

end

$k = k + 1$;

end

$m = \max j$ such that $c_{j,n-1} < 0$;

$y = c_{1:m,n}$;

$\delta^P = \frac{-\max(y)}{a_0}$;

6.6.1.6 Generalisation of results

In the general framework of arbitrary processivity and sequentiality, $[S_i] = [S_0]r_i(u)$ [163]. $r_i(u)$ is a rational function of u which was proved to always be well defined and positive in the s-positive sense. A polynomial is said to be sum positive (s-positive) if it is a sum of positive monomials [163]. A rational function is s-positive if it can be represented as a fraction of two s-positive polynomials.

In this general framework, the three $\phi(u)$ functions, all being s-positive, are defined as:

$$\begin{aligned}\phi_1(u) &= \sum_{i=0}^n r_i(u) \\ \phi_2(u) &= \sum_{i=0}^{n-1} \frac{r_i(u)}{k_i^K} \\ \phi_3(u) &= \sum_{i=1}^n \frac{r_i(u)}{k_i^P}\end{aligned}\tag{6.6.1}$$

Therefore, a rational expression in u , $Z(u)$, can be defined, where $Z(u) = \frac{R(u)}{Q(u)}$. $Q(u)$ is also s-positive.

$$0 = (u - w)\phi_1(u) + [S_{\text{tot}}](u\phi_2(u) - w\phi_3(u)) =: Z(u) \quad (6.6.2)$$

where $(w = \frac{[K_{\text{tot}}]}{[P_{\text{tot}}]})$.

From Eq. 6.6.1, if all three $\phi(u)$ functions are expressed with the same least common denominator, which is equal to $Q(u)$, then

$$\begin{aligned} \phi_1(u) &= \frac{R_1(u)}{Q(u)} \\ \phi_2(u) &= \frac{R_2(u)}{Q(u)} \\ \phi_3 &= \frac{R_3(u)}{Q(u)} \end{aligned} \quad (6.6.3)$$

Therefore $R(u)$ can be expressed as:

$$0 = (u - w)R_1(u) + [S_{\text{tot}}](uR_2(u) - wR_3(u)) =: R(u) \quad (6.6.4)$$

Thus, $R(u)$ can be expressed as:

$$R(u) = a_{N+1}u^{N+1} + a_Nu^N + \dots + a_1u + a_0 \quad (6.6.5)$$

It is clear from Eq. 6.6.1 that $\text{order}(R_1(u)) \geq \text{order}(R_2(u))$. Similarly, the degree of $R_1(u)$ is no less than the degree of $R_3(u)$. From Eq. 6.6.4, it is evident that $a_{N+1} > 0$. Conversely, $a_0 < 0$.

Extending Eq. 6.6.5 to include sequestration of the phosphatase by S_0 simply requires adding $\frac{r_0}{k_0^P}$ to ϕ_3 or, equivalently, as $r_0 = 1$ from $[S_i] = [S_0]r_i(u)$, $\frac{1}{k_0^P}Q(u)$ to R_3 in Eq. 6.6.4. Thus, $R'(u) = R(u) - [S_{\text{tot}}]w\frac{\alpha_0^P}{\beta_0^P}Q(u)$ with the degree of Q less than that of R . As $\frac{\alpha_0^P}{\beta_0^P}$ is increased, $\text{deg}(Q)$ roots of $R'(u)$ tend to the roots of $Q(u)$, none of which are real and positive (since Q has all positive coefficients). The remaining m roots (where $m = \text{deg}(R) - \text{deg}(Q)$) tend to infinity at angles which are multiples of $\frac{2\pi}{m}$ by a standard root locus argument [42]. Thus, for sufficiently large $\frac{\alpha_0^P}{\beta_0^P}$, there will be just one root on the positive real axis.

6.6.1.7 Parameters of multisite protein phosphorylation systems under investigation

		$i = 0$	$i = 1$	$i = 2$	$i = 3$
$\alpha_i^K (nM^{-1}sec^{-1})$		8.12×10^{-3}	1.02×10^{-1}	8.12×10^{-3}	1.02×10^{-1}
$\beta_i^K (sec^{-1})$		1.60×10^{-2}	2.04×10^{-1}	1.60×10^{-2}	2.04×10^{-1}
$\gamma_{i,i+1}^K (sec^{-1})$		1.00×10^{-1}	$1.00 \times 10^{+1}$	1.00×10^{-1}	$1.00 \times 10^{+1}$
$\alpha_{i+1}^P (nM^{-1}sec^{-1})$		1.12×10^{-1}	2.64×10^{-3}	6.51×10^{-1}	2.85×10^{-3}
$\beta_{i+1}^P (sec^{-1})$		2.24×10^{-1}	5.00×10^{-3}	$1.30 \times 10^{+0}$	6.00×10^{-3}
$\gamma_{i+1,i}^P (sec^{-1})$		$1.10 \times 10^{+1}$	1.70×10^{-2}	$6.39 \times 10^{+1}$	1.36×10^{-1}
$k_i^K (nM)$		$1.43 \times 10^{+1}$	$1.00 \times 10^{+2}$	$1.43 \times 10^{+1}$	$1.00 \times 10^{+2}$
$k_{i+1}^P (nM)$		$1.00 \times 10^{+2}$	$8.33 \times 10^{+0}$	$1.00 \times 10^{+2}$	$5.00 \times 10^{+1}$
λ_i		6.38×10^{-2}	$5.05 \times 10^{+1}$	1.01×10^{-2}	$3.67 \times 10^{+1}$
$[S_{tot}] (nM)$	$1.00 \times 10^{+4}$				
$[K_{tot}] (nM)$	$2.80 \times 10^{+3}$				
$[P_{tot}] (nM)$	$2.80 \times 10^{+3}$				

Table 6.1 The parameters of the original multisite protein phosphorylation system by Thomson and Gunawardena [163, 162], also used in our analysis

Stochastic Parameter	Stochastic Value	Equivalent Deterministic Parameter	Deterministic Value
Sequestration parameters			
$\alpha (sec^{-1})$	1.0×10^{-2}	$\beta_0^P (sec^{-1})$	1.0×10^{-2}
$\beta (molecule^{-1}sec^{-1})$	1.6×10^{-2}	$\alpha_0^P (nM^{-1}sec^{-1})$	2.4×10^{-5}
$\kappa (molecule^{-1}sec^{-1})$	1.0×10^{-1}	$\alpha_1^K (nM^{-1}sec^{-1})$	1.5×10^{-4}
$\lambda (sec^{-1})$	1.0×10^{-1}	$\beta_1^K (sec^{-1})$	1.0×10^{-1}
Non-sequestration parameters			
$\gamma (molecule^{-1}sec^{-1})$	5.4133×10^0	$\alpha_0^K (nM^{-1}sec^{-1})$	8.12×10^{-3}
$\delta (sec^{-1})$	1.6×10^{-2}	$\beta_0^K (sec^{-1})$	1.6×10^{-2}
$\varepsilon (sec^{-1})$	1.0×10^{-1}	$\gamma_{0,1}^K (sec^{-1})$	1.0×10^{-1}
$\eta (molecule^{-1}sec^{-1})$	$7.47 \times 10^{+1}$	$\alpha_1^P (nM^{-1}sec^{-1})$	1.12×10^{-1}
$\theta (sec^{-1})$	2.24×10^{-1}	$\beta_1^P (sec^{-1})$	2.24×10^{-1}
$\zeta (sec^{-1})$	$1.1 \times 10^{+1}$	$\gamma_{1,0}^P (sec^{-1})$	$1.1 \times 10^{+1}$

Table 6.2 The stochastic parameters used in the single phosphosite system (for a Volume = 2.49×10^{-18} L). The non-sequestration parameters are the same as the ones in the multisite protein phosphorylation system by Thomson and Gunawardena [163, 162]

6.6.1.8 Enzyme-sharing scheme parameters

Parameter	Value	Parameter	Value
$\gamma_s(nM^{-1}sec^{-1})$	1×10^{-5}	$\gamma_z(nM^{-1}sec^{-1})$	1×10^{-3}
$\delta_s(sec^{-1})$	1×10^{-3}	$\delta_z(sec^{-1})$	1×10^{-4}
$\epsilon_s(sec^{-1})$	5×10^{-2}	$\epsilon_z(sec^{-1})$	1×10^{-2}
$\eta_s(nM^{-1}sec^{-1})$	1×10^{-2}	$\eta_z(nM^{-1}sec^{-1})$	1×10^{-1}
$\theta_s(sec^{-1})$	1×10^{-3}	$\theta_z(sec^{-1})$	1×10^{-1}
$\zeta_s(sec^{-1})$	5×10^{-5}	$\zeta_z(sec^{-1})$	2×10^{-4}
$[S_{tot}](nM)$	2×10^4	$[P_{tot}](nM)$	3×10^2
$[Z_{tot}](nM)$	1.8×10^4	$[K_{tot}](nM)$	1×10^2

Table 6.3 The parameters for the enzyme sharing scheme of Fig. 6.3.6

Chapter 7

Substrate Enzyme-Sequestration adapted for CaMKII and a potential role in synaptic plasticity

In Chapter 6, we noted that enzyme docking and the consequent Substrate-Enzyme Sequestration can induce bimodality, depending on the system size. In this chapter, we adapt that framework for the particular characteristics of Ca^{2+} /calmodulin-dependent protein kinase II (CaMKII), a key kinase involved in synaptic plasticity, whose mechanisms remain elusive, speculating that such a mechanism could potentially explain some of the experimental findings.

7.1 Introduction

The mechanisms that underlie how cognitive information is encoded in the brain are still a mystery [113]. In 1949, Hebb suggested that the connections between two neurons might be strengthened if the neurons fire simultaneously [72]. The ability of synapses to strengthen or weaken over time is called synaptic plasticity. It is initiated by changes in dendritic spine Ca^{2+} concentration driven by presynaptic and postsynaptic neuronal activity. The Ca^{2+} signals are detected, triggering a biochemical cascade that leads to potentiation or depression of synaptic efficacy [128].

Dendritic spine size is correlated with the strength of the synapse it hosts [114] and small spines exhibit greater $[Ca^{2+}]$ changes during synaptic activation than large spines [181, 127, 153]. Furthermore, Matsuzaki et al [115] found that spine size is regulated during synaptic plasticity as well as that small spines are preferential sites for long-term potentiation

induction.

In these mechanisms, the presence of Ca^{2+} /calmodulin-dependent protein kinase II (CaMKII) is found to be critical [55].

CaMKII is a serine/threonine kinase holoenzyme, consisting of 12 subunits, being the most abundant protein in the post-synaptic density (PSD), a structure that defines a neuronal subcompartment critical for plasticity. Once activated by calcium/calmodulin, the enzyme can be autophosphorylated, making the CaMKII activity persist even after the calcium concentration falls to baseline levels [101]. In fact, since the introduction of the idea that a kinase with the ability of phosphorylation coupled with a phosphatase can act as a molecular switch [102] and the subsequent discovery of such a kinase in abundance in neurons [119], CaMKII became a likely candidate for being the ‘memory molecule’. In particular, CaMKII is strongly associated with long-term potentiation (LTP) (the long-lasting increase in the efficiency of glutamatergic synaptic transmission) leading to synaptic plasticity [74].

The findings from single-molecule tracking studies illustrate that CaMKII is generally found in subpopulations [105], interacting with several proteins dependent on its particular location. Estimates have shown that in the entire average spine PSD there are only 90 to 240 holoenzymes [47]. Therefore, subpopulations of CaMKII can even be in the orders of tens of holoenzymes, placing any potential analysis of the system into the stochastic regime. This, along with the finding that its associated phosphatase, phosphatase-1 (PP1), can act on several phosphorylation sites of CaMKII and not just the autophosphorylating one (T286) [122], motivates us to investigate whether enzyme docking, as presented in Chapter 6, could provide an explanation of the experimental findings regarding CaMKII activity.

7.2 Bistable models of CaMKII in literature

Initial models of CaMKII suggested that CaMKII autonomous activity could be stable enough to maintain long-term potentiation and memories for a lifetime [103, 182, 104], while a later model suggested that CaMKII could be active as a consequence of T286 phosphorylation for years [118]. The model by Lisman and Zhabotinsky [182, 104] considered many key biochemical properties of CaMKII holoenzymes and the phosphatase-1 (PP1) enzymes that dephosphorylate them. It was shown that an interplay between autophosphorylation of CaMKII holoenzymes and dephosphorylation by PP1 molecules can give rise to two stable states of phosphorylation at basal levels of free $[Ca^{2+}]$. This bistability has recently been criticised, even being described as ‘delicate’ [117]. It has been suggested that the binding of CaMKII to the NR2B subunit of N-methyl-D-aspartate receptors (NMDARs) may help CaMKII maintain their activity due to blocking PP1 from reaching CaMKII’s

autophosphorylated site [122], yet there is more CaMKII than NMDARs in the PSD. In fact, Yasuda et al, using fluorescence resonance energy transfer (FRET) technology, found that LTP induction leads only to transient increase of CaMKII activity, by about 1 min [97, 27], further suggesting that CaMKII activation is required for LTP induction, but not LTP maintenance. This timescale is also consistent with the time found for the volume size of the spine to reach maximum size [97].

7.3 Can Substrate Enzyme-Sequestration models be adapted to incorporate the particular characteristics of CaMKII?

With many models of CaMKII published, others being more detailed [182, 38, 106], than others [117], we can easily see that we are far from consensus on the mechanisms guiding CaMKII bistability and how this could be consistent with the experimental findings. What captured our attention, especially in the light of the results of Chapter 6, are the findings that the timescale of spine volume size is of similar order to that of CaMKII activation and that small spines are preferential sites for long-term potentiation induction.

Therefore a mechanism which would simultaneously account for CaMKII activity and for the spine volume size could be possible. As we have described in the previous chapter, Substrate Enzyme-Sequestration is a mechanism able to induce bistability/bimodality and is indeed dependent on the system size.

CaMKII, however, acts both as the kinase and as the substrate by autophosphorylating itself. This requires a different model than the one presented in Chapter 6. Considering PP1 to be the phosphatase acting on the autophosphorylated active autonomous CaMKII*, we develop a simplified model, presented in Fig 7.3.1. The inactive complex due to enzyme docking is CaMKII.PP1. This inactive complex could be formed while PP1 is acting on other phosphorylated sites of CaMKII, something that has been shown experimentally [79, 33].

7.4 Substrate Enzyme-Sequestration can induce bimodality and consequently activation of CaMKII, which becomes more difficult as system size expands

In order to identify whether bimodality can be induced by this adaptation of Substrate Enzyme-Sequestration, the model was stochastically simulated (by the Gillispie Exact Algorithm), using, wherever possible, parameters of the same order of magnitude as those

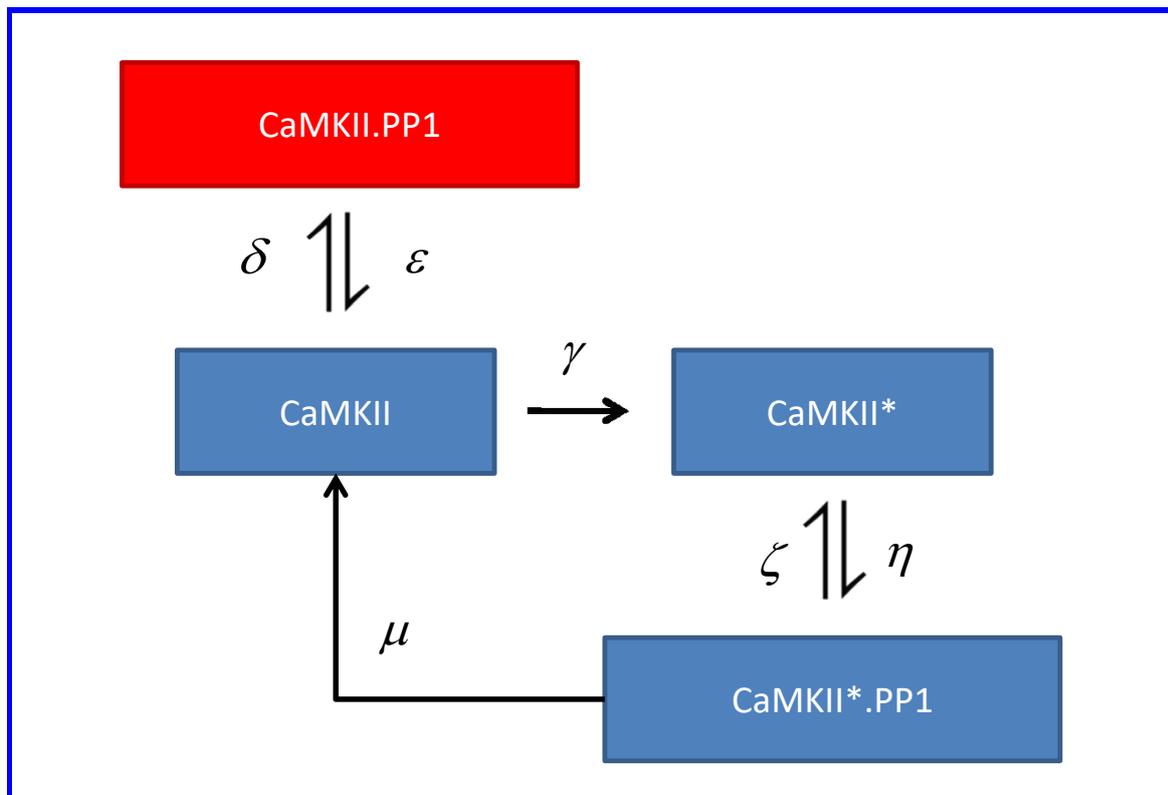


Fig. 7.3.1 Substrate Enzyme-Sequestration model adapted for CaMKII

used in the literature [118, 117]. The parameters used are found in the Appendix. The model assumes that CaMKII is found in excess of the phosphatase, PP1, as in [118].

For a system comprised by 20 CaMKII molecules and 6 PP1 enzymes, we performed exact stochastic simulations using the Gillispie Exact Algorithm. In Fig. 7.4.1, we can see that indeed bimodality can be induced by this adapted version of Substrate Enzyme-Sequestration. The dominant mode is $(\text{CaMKII}, \text{CaMKII}^*) = (0, 14)$. Furthermore, as we can also see from Fig. 7.4.1, the dwell times at this modal state are approximately exponentially distributed with a mean of 55 seconds. This is of the same timescale as the experimentally found activation time of approximately 1 minute [97, 27].

Nevertheless, as we can also see from Fig. 7.4.1, when CaMKII is allowed to autophosphorylate (which happens after a calcium influx) and is found in the low molecule number regime, the system tends to move to the modal active state. This could add robustness to the system, as it would mean that the system would actually get deactivated only when the spine reaches the aimed size (while approaching the deterministic regime) and it would not be so dependent on the actual dwell time spent at a particular state. Therefore, it is possible that the experimentally found activation time is due to the time required for the enlargement of the spine.

In Chapter 6 we saw that bimodality is induced in these systems because of an extinction of a species. Bimodality in this case is induced because all of the phosphatase, PP1, is used in intermediate complexes. When the number of free phosphatase PP1 molecules becomes zero, the transition rates which are directly proportional to the number of phosphatase molecules (as they are based on mass kinetic principles) become zero, as illustrated graphically in Fig. 7.4.2.

As the system size expands, the system approaches monomodality that is expected to be seen via deterministic analysis.

Note that this type of noise-induced temporary extinction of one species due to enzyme sequestration by the substrate, is very different to the boundary bistability that can be described by deterministic analysis [170, 125]. In deterministic models showing boundary bistability [170], a saddle-node bifurcation is present, with population going extinct if certain parameters fall below a particular threshold. Unlike noise-induced extinction phenomena, this extinction will persist. Furthermore, the deterministic prediction, as it is only valid in the regime of large molecule numbers [57], only depends on the kinetic parameters as well as on the concentrations of the relevant species. This means that deterministic analysis will always provide us with the same prediction, regardless of the system size. For example, in the model of CaMKII described in this chapter, deterministic analysis always predicts one stable steady

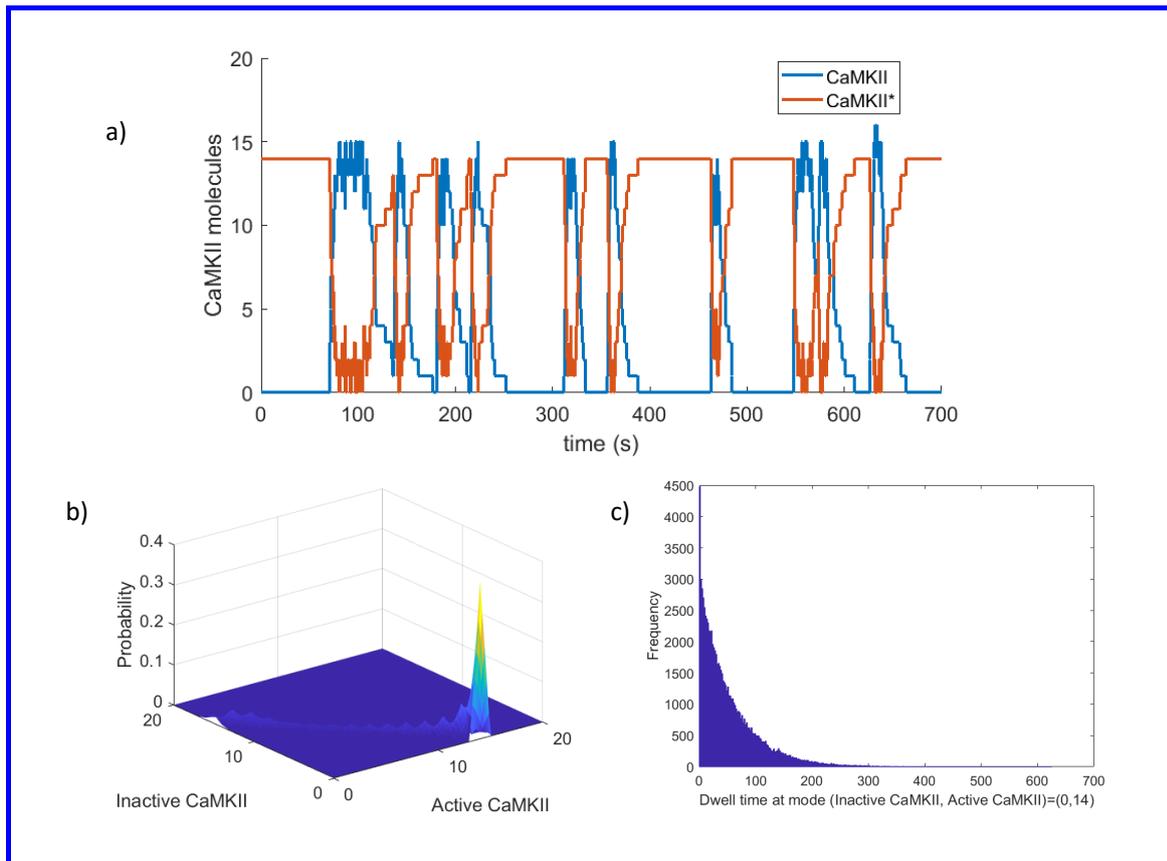


Fig. 7.4.1 a) Time-series illustrating the bimodality achieved due to the PP1 Sequestration by CaMKII. b) The state with 14 active and zero inactive CaMKII molecules is the dominant mode of the corresponding probability distribution. c) The dwell times at the modal state are approximately exponentially distributed with a mean of 55 seconds.

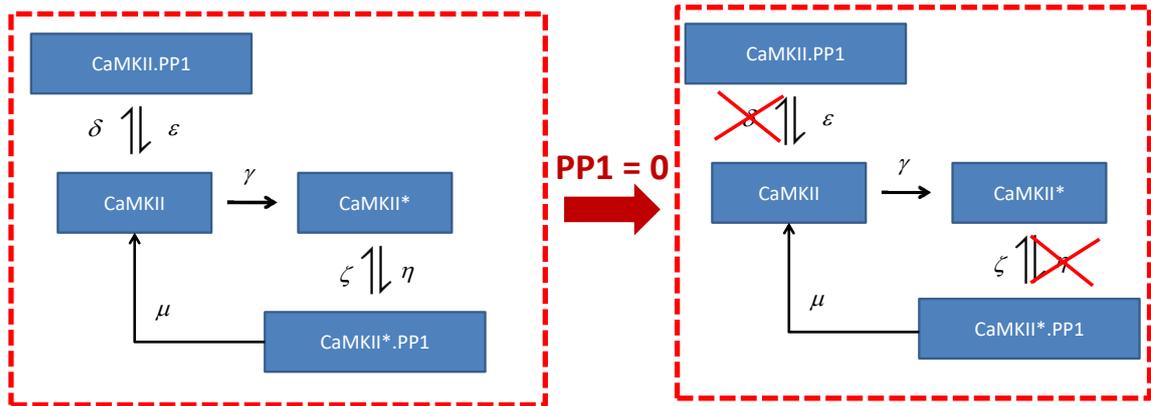


Fig. 7.4.2 As in Chapter 6, the induced bimodality occurs due to the extinction of a species. This species here is PP1.

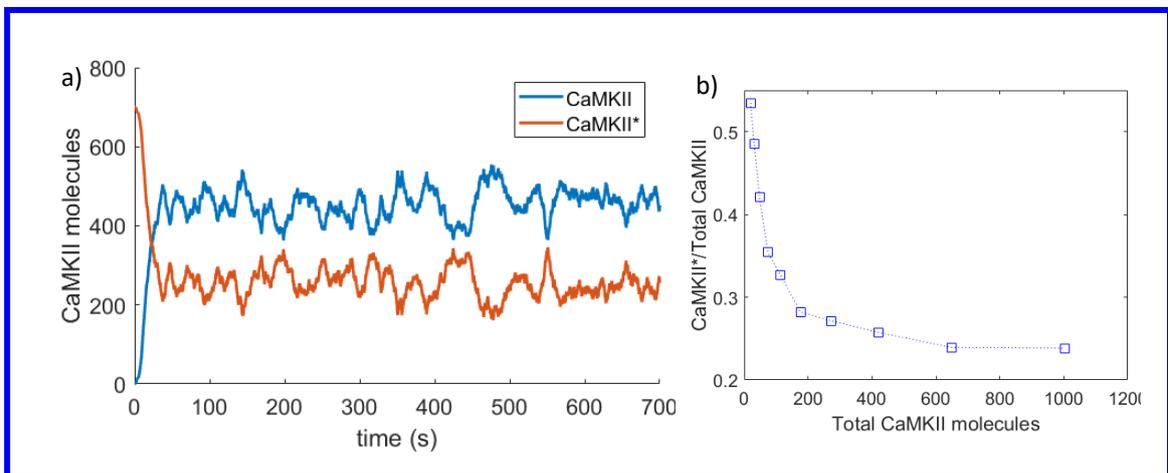


Fig. 7.4.3 a) For an expanded system with 1000 CaMKII and 300 PP1 molecules, monomodal-ity is instead obtained. b) As the system size expands and bimodality is lost, the ratio of CaMKII*/(Total CaMKII) drops, as expected.

state, which we only achieve when we reach the regime of large molecule numbers. This is shown in Fig. 7.4.3, where the system is simulated with 1000 CaMKII and 300 PP1 molecules, keeping the same ratio. The stochastic parameters are also updated, assuming the same concentrations of CaMKII and PP1 are found in a greater volume (i.e. the deterministic rate constant is kept constant). It can be seen that now the system exhibits monomodality, with the amount of inactive CaMKII being greater than that of active CaMKII*. In Fig. 7.4.3 (b), we can indeed see that as the system size expands and bimodality is lost, the ratio of CaMKII*/Total CaMKII drops, as expected. This is consistent with the experimental finding that small spines are more easily LTP induced [181, 127, 153]. In the case that CaMKII indeed functions in this way, this could be an example of what we hinted in Chapter 6, i.e. that Substrate Enzyme-Sequestration could serve a specific function when a system requires different behaviour when its size expands. This is summarised in Fig. 7.5.1, where the suggested mechanism is illustrated. The CaMKII system could be brought to the active mode via an external signal (e.g. a calcium influx), thus promoting spine enlargement, the size of which, Ω , would act as negative feedback on the activation of CaMKII. Interestingly, this type of negative feedback is obtained due to the transition from the stochastic to the deterministic regime as the system size expands.

7.5 What can we learn from this model?

Regardless of whether this model accurately describes the process of CaMKII activation or not, recognising that such simple mechanisms can exhibit the experimentally observed characteristics of CaMKII is important on its own right for two reasons. Firstly, its great simplicity contrasts with the well-known and established deterministic bistable model by Lisman and Zhabotinsky [182, 104, 147], which is essentially a multisite phosphorylation model (the general form of which we investigated in Chapter 6) with 10 available phosphosites [147]. In that model, it is the multisite protein phosphorylation mechanism that creates the deterministically necessary positive feedback for bistability to occur, whereas we show that when the system is analysed stochastically, noise-induced temporary sequestration of the enzyme can induce bimodality even with one available phosphosite.

Secondly and most importantly, however, we illustrate the ability of such simple stochastic systems to leverage the system expansion as a feedback mechanism. We show that the transition of the CaMKII activation regime can occur without complex feedback cascades, but rather because of the transition from the stochastic to the deterministic regime as the system size expands.

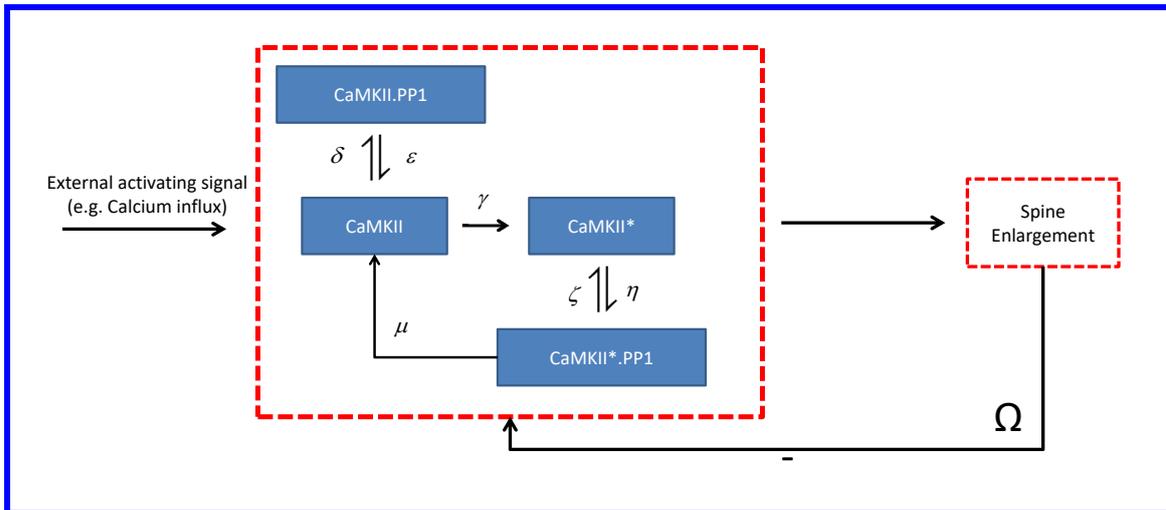


Fig. 7.5.1 Overview of the scheme with negative feedback obtained via changing the system size, moving from the stochastic to the deterministic regime

This becomes even more relevant when recent experimental findings contradict the main premise of the established deterministic models showing bistability of CaMKII. These models are based on the assumption that CaMKII molecules are "activated during LTP induction and remain active" [104]. This is inconsistent with the experimental findings that LTP induction leads only to transient increase of CaMKII activity for only about 1 minute [27] as well as with the finding that this activation time is consistent with the time required for the spine to reach its maximum size [97].

7.6 Conclusion

In this chapter, we show that Substrate Enzyme-Sequestration can induce bimodality in the stochastic domain even when the substrate is both the kinase and the substrate, by autophosphorylating itself. We speculate that this could be applied to CaMKII, an autophosphorylating kinase found to play a dominant role in synaptic plasticity. The dependence of this type of induced bimodality to the system size could provide an explanation of the experimental findings that the LTP induction is preferentially happening in smaller spines and that the timescales of CaMKII activation are similar to the timescales governing spine volume expansion.

7.7 Appendix - Parameters used for the simulations of Chapter 7

Parameter	Value
$\gamma(sec^{-1})$	1×10^{-1}
$\eta(molecule^{-1}sec^{-1})$	1
$\zeta(sec^{-1})$	1×10^{-1}
$\mu(sec^{-1})$	10
$\delta(sec^{-1})$	1×10^{-2}
$\varepsilon(molecule^{-1}sec^{-1})$	3×10^{-3}

Table 7.1 The stochastic parameters used in the CaMKII simulations for a system of 20 CaMKII.CaM and 6 PP1 molecules

Chapter 8

Conclusions

8.1 Key findings

In this dissertation we have provided a quantitative analysis of the effect that enzyme docking and the consequent phosphatase and kinase sequestration by the fully unphosphorylated and the fully phosphorylated substrates respectively have on a system of multisite protein phosphorylation and similar models. The analysis was done both in the deterministic and the stochastic domains for large and small molecule numbers respectively. We have proved, that in the deterministic domain substrate enzyme-sequestration limits the extent of multistability ultimately to one steady state, even for systems with arbitrary processivity or sequentiality. In this context, we also developed algorithms that can calculate, directly from reaction rates, the sufficient strengths of enzyme sequestration that would lead to different orders of multistability. In the stochastic domain, we have shown that substrate enzyme-sequestration can provide bimodality, even when bistability is not possible for large molecule numbers.

We have also extended our results to autophosphorylating kinases, as for example is the Ca^{2+} /calmodulin-dependent protein kinase II (CaMKII), an enzyme with a pivotal role in synaptic plasticity, speculating an explanation of the experimental findings that smaller spines are preferentially chosen for Long-term Potentiation (LTP) induction and that CaMKII activation's timescale is similar to that of spine volume expansion.

These findings were underpinned by the development of several tools. Firstly, we developed a discrete 'nullcline' construct to be used in stochastic analysis, analogue to the ODE nullclines in deterministic analysis, based on the Markov chain tree theorem. This efficient graphical heuristic tool allowed us to obtain a better understanding of the discrepancies reported between deterministic and stochastic moles of the well-studied genetic toggle switch problem. Furthermore, we have developed a weakly chained diagonally dominant M-matrix formula-

tion of the Chemical Master Equation, along with an accompanying tool, allowing greater insights in the way particular mechanisms, like enzyme sequestration, can shape probability distributions and therefore exhibit different behaviour across different regimes.

That formulation was also used to develop an algorithm for constructing not necessarily reversible Markov chains with specified discrete-state stationary probability distributions, enhancing our understanding of how one could control the stationary probability distribution formation, without relying on the establishment of detailed balance.

8.2 Further research

There is much scope for further research, regarding both the biology and the mathematical tools proposed. Firstly, the biological findings of this dissertation could be used to guide further experimental research, especially regarding enzyme docking and the formation of inactive enzyme-substrate substrates. The non-consideration of enzyme docking in many models might have overestimated the potential extent of multistability in particular mechanisms, whereas the non-consideration of the low numbers stochastic effects, might have led the development of complex models to explain an observed bistable behaviour, or even to the conclusion that bistability is not a behaviour exhibited by a system. Particularly in fields where experimental evidence is contradicting, as for example is the case with CaMKII, these considerations should be taken into account.

As far as the mathematical tools proposed are concerned, further research could be done by expanding the discrete ‘nullclines’ concept to higher-dimensional non-planar graphs. Furthermore, more research could be done in improving the bounds obtained for the stationary probability of a microstate, by utilising, for example, the WCDD M-matrix formulation of the Chemical Master Equation.

References

- [1] (1970). *Adaptive, Learning, and Pattern Recognition Systems; theory and applications*. Mathematics in Science and Engineering. Elsevier Science.
- [2] Ahsendorf, T., Wong, F., Eils, R., and Gunawardena, J. (2014). A framework for modelling gene regulation which accommodates non-equilibrium mechanisms. *BMC biology*, 12(1):102.
- [3] Alfa, A., Xue, J., and Ye, Q. (2002a). Accurate computation of the smallest eigenvalue of a diagonally dominant m-matrix. *Mathematics of computation*, 71(237):217–236.
- [4] Alfa, A. S., Xue, J., and Ye, Q. (2002b). Entrywise perturbation theory for diagonally dominant m-matrices with applications. *Numerische Mathematik*, 90(3):401–414.
- [5] Alon, U. (2006). *An introduction to systems biology: design principles of biological circuits*. CRC press.
- [6] Anantharam, V. and Tsoucas, P. (1989). A proof of the markov chain tree theorem. *Statistics & Probability Letters*, 8(2):189–192.
- [7] Azimzadeh, P. and Forsyth, P. A. (2016). Weakly chained matrices, policy iteration, and impulse control. *SIAM Journal on Numerical Analysis*, 54(3):1341–1364.
- [8] Bardwell, A. J., Abdollahi, M., and Bardwell, L. (2003). Docking sites on mitogen-activated protein kinase (mapk) kinases, mapk phosphatases and the elk-1 transcription factor compete for mapk binding and are crucial for enzymic activity. *Biochemical Journal*, 370(3):1077–1085.
- [9] Barreras, A. and Pena, J. (2012). Accurate and efficient ldu decomposition of diagonally dominant m-matrices. *Electronic Journal of Linear Algebra*, 24(1):12.
- [10] Barreras, Á. and Peña, J. M. (2014). Characterisations and tests for almost m-matrices. *Monografias Matemáticas*, 39:57–63.
- [11] Basile, R., Grima, R., and Popović, N. (2013). A graph-based approach for the approximate solution of the chemical master equation. *Bulletin of mathematical biology*, 75(10):1653–1696.
- [12] Basu, K. and Liu, X. (2015). Mathematical modeling for multisite phosphorylation with scaffold binding in cell signaling. *Mathematical Methods in the Applied Sciences*, 38(18):4521–4529.

- [13] Berman, A. and Plemmons, R. J. (1994). *Nonnegative matrices in the mathematical sciences*. SIAM.
- [14] Bhaduri, S., Valk, E., Winters, M. J., Gruessner, B., Loog, M., and Pryciak, P. M. (2015). A docking interface in the cyclin *cln2* promotes multi-site phosphorylation of substrates and timely cell-cycle entry. *Current Biology*, 25(3):316–325.
- [15] Bialek, W. and Setayeshgar, S. (2008). Cooperativity, sensitivity, and noise in biochemical signaling. *Physical Review Letters*, 100(25):258101.
- [16] Biancalani, T. and Assaf, M. (2015). Genetic toggle switch in the absence of cooperative binding: Exact results. *Physical review letters*, 115(20):208101.
- [17] Bierkens, J. (2016). Non-reversible metropolis-hastings. *Statistics and Computing*, 26(6):1213–1228.
- [18] Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer.
- [19] Blüthgen, N., Legewie, S., Herzog, H., and Kholodenko, B. (2007). Mechanisms generating ultrasensitivity, bistability, and oscillations in signal transduction. *Introduction to Systems Biology*, pages 282–299.
- [20] Boyd, S., Diaconis, P., and Xiao, L. (2004). Fastest mixing markov chain on a graph. *SIAM review*, 46(4):667–689.
- [21] Bressan, M., Peserico, E., and Pretto, L. (2018). On approximating the stationary distribution of time-reversible markov chains. In *35th Symposium on Theoretical Aspects of Computer Science*.
- [22] Bundy, A. and Wallen, L. (1984). Breadth-first search. In *Catalogue of Artificial Intelligence Tools*, pages 13–13. Springer.
- [23] Butler, J. and Noble, D. (2014). *Progress in Biophysics and Molecular Biology*. Elsevier Science.
- [24] Cai, G. and Zhu, W. (2016). *Elements of Stochastic Dynamics*. World Scientific Publishing Company.
- [25] Cao, Y. and Liang, J. (2008). Optimal enumeration of state space of finitely buffered stochastic molecular networks and exact computation of steady state landscape probability. *BMC Systems Biology*, 2(1):30.
- [26] Chan, C., Liu, X., Wang, L., Bardwell, L., Nie, Q., and Enciso, G. (2012). Protein scaffolds can enhance the bistability of multisite phosphorylation systems. *PLoS computational biology*, 8(6):e1002551.
- [27] Chang, J.-Y., Parra-Bueno, P., Laviv, T., Szatmari, E. M., Lee, S.-J. R., and Yasuda, R. (2017). Camkii autophosphorylation is necessary for optimal integration of ca²⁺ signals during ltp induction, but not maintenance. *Neuron*, 94(4):800–808.
- [28] Chen, L., Wang, R., and Aihara, K. (2009). Stochastic hybrid system for chemical master equation. In *Proc. 3rd Int. Symp. on Optimization and Systems Biology, Hunan, China*, pages 475–481.

- [29] Chen, L., Wang, R., Li, C., and Aihara, K. (2010). *Modeling biomolecular networks in cells: structures and dynamics*. Springer Science & Business Media.
- [30] Chib, S. and Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The american statistician*, 49(4):327–335.
- [31] Choi, S. (2007). Introduction to systems biology.
- [32] Cohen, P. (2000). The regulation of protein function by multisite phosphorylation—a 25 year update. *Trends in biochemical sciences*, 25(12):596–601.
- [33] Colbran, R. J. (2004). Protein phosphatases and calcium/calmodulin-dependent protein kinase ii-dependent synaptic plasticity. *Journal of Neuroscience*, 24(39):8404–8409.
- [34] Dellacherie, C., Martinez, S., and San Martin, J. (2014). *Inverse M-matrices and ultrametric matrices*, volume 2118. Springer.
- [35] Demmel, J. and Koev, P. (2004). Accurate svds of weakly diagonally dominant m-matrices. *Numerische Mathematik*, 98(1):99–104.
- [36] Diaconis, P. (2009). The markov chain monte carlo revolution. *Bulletin of the American Mathematical Society*, 46(2):179–205.
- [37] Ding, W., Qi, L., and Wei, Y. (2013). M-tensors and nonsingular m-tensors. *Linear Algebra and Its Applications*, 439(10):3264–3278.
- [38] Don, K. and Yao, H. (2017). *Computational Systems Biology Of Synaptic Plasticity: Modelling Of Biochemical Pathways Related To Memory Formation And Impairment*, volume 10. # N/A.
- [39] Duncan, A., Liao, S., Vejchodský, T., Erban, R., and Grima, R. (2015). Noise-induced multistability in chemical systems: Discrete versus continuum modeling. *Physical Review E*, 91(4):042111.
- [40] Dušan, D., Vladimír, J., Ivan, M., and Nikola, P. (2006). *The IMO Compendium: A Collection of Problems Suggested for the International Mathematical Olympiads: 1959-2004*. Springer New York.
- [41] English, B. P., Min, W., Van Oijen, A. M., Lee, K. T., Luo, G., Sun, H., Cherayil, B. J., Kou, S., and Xie, X. S. (2006). Ever-fluctuating single enzyme molecules: Michaelis-menten equation revisited. *Nature chemical biology*, 2(2):87–94.
- [42] Evans, W. R. (1950). Control system synthesis by root locus method. *Transactions of the American Institute of Electrical Engineers*, 69(1):66–69.
- [43] Fang, M. (2007). Bounds on eigenvalues of the hadamard product and the fan product of matrices. *Linear algebra and its applications*, 425(1):7–15.
- [44] Feinsilver, P. (2016). Matrices with zero row sums, tree theorems and a markov chain on trees. *Probability on Algebraic and Geometric Structures*, 668:67.
- [45] Feliu, E. and Wiuf, C. (2012). Enzyme-sharing as a cause of multi-stationarity in signalling systems. *Journal of The Royal Society Interface*, 9(71):1224–1232.

- [46] Feliu, E. and Wiuf, C. (2015). Finding the positive feedback loops underlying multi-stationarity. *BMC systems biology*, 9(1):22.
- [47] Feng, B., Raghavachari, S., and Lisman, J. (2011). Quantitative estimates of the cytoplasmic, psd, and nmdar-bound pools of camkii in dendritic spines. *Brain research*, 1419:46–52.
- [48] Feng, S., Ollivier, J. F., and Soyer, O. S. (2016a). Enzyme sequestration as a tuning point in controlling response dynamics of signalling networks. *PLoS computational biology*, 12(5):e1004918.
- [49] Feng, S., Sáez, M., Wiuf, C., Feliu, E., and Soyer, O. S. (2016b). Core signalling motif displaying multistability through multi-state enzymes. *Journal of The Royal Society Interface*, 13(123):20160524.
- [50] Ferrell, J. E. and Ha, S. H. (2014). Ultrasensitivity part i: Michaelian responses and zero-order ultrasensitivity. *Trends in biochemical sciences*, 39(10):496–503.
- [51] Ferrell, J. E., Ha, S. H., et al. (2014). Ultrasensitivity part ii: multisite phosphorylation, stoichiometric inhibitors, and positive feedback. *Trends in biochemical sciences*, 39(11):556–569.
- [52] Gabow, H. N., Galil, Z., Spencer, T., and Tarjan, R. E. (1986). Efficient algorithms for finding minimum spanning trees in undirected and directed graphs. *Combinatorica*, 6(2):109–122.
- [53] Gardner, T. S., Cantor, C. R., and Collins, J. J. (2000). Construction of a genetic toggle switch in escherichia coli. *Nature*, 403(6767):339–342.
- [54] Gaveau, B. and Schulman, L. (1996). Master equation based formulation of nonequilibrium statistical mechanics. *Journal of Mathematical Physics*, 37(8):3897–3932.
- [55] Giese, K. P. (2015). *Novel mechanisms of memory*. Springer.
- [56] Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361.
- [57] Gillespie, D. T. (2000). The chemical langevin equation. *The Journal of Chemical Physics*, 113(1):297–306.
- [58] Gillespie, D. T. and Seitaridou, E. (2012). *Simple Brownian diffusion: an introduction to the standard theoretical models*. Oxford University Press.
- [59] Girard, A. (1884). *Invention nouvelle en l’algèbre*. Imprimé chez Muré frères.
- [60] Goel, N. S. and Richter-Dyn, N. (2016). *Stochastic models in biology*. Elsevier.
- [61] Goldbeter, A. and Koshland, D. E. (1981). An amplified sensitivity arising from covalent modification in biological systems. *Proceedings of the National Academy of Sciences*, 78(11):6840–6844.

- [62] Gordon, G. and McMahon, E. (1989). A greedoid polynomial which distinguishes rooted arborescences. *Proceedings of the American Mathematical Society*, 107(2):287–298.
- [63] Gould, R. (2012). *Graph theory*. Dover.
- [64] Grima, R., Walter, N. G., and Schnell, S. (2014). Single-molecule enzymology à la michaelis–menten. *The FEBS journal*, 281(2):518–530.
- [65] Guldberg, C. M. and Waage, P. (1867). *Etudes sur les affinités chimiques*. Brøgger & Christie.
- [66] Gunawardena, J. (2005). Multisite protein phosphorylation makes a good threshold but can be a poor switch. *Proceedings of the National Academy of Sciences of the United States of America*, 102(41):14617–14622.
- [67] Gupta, A. and Khammash, M. (2017). Finding the steady-state solution of the chemical master equation. In *Control Technology and Applications (CCTA), 2017 IEEE Conference on*, pages 953–954. IEEE.
- [68] Hannon, B. and Ruth, M. (2014). Modeling dynamic biological systems. In *Modeling dynamic biological systems*, pages 3–28. Springer.
- [69] Hardy, G. H., Littlewood, J. E., and George, P. (1991). *Inequalities*. Cambridge University Press.
- [70] Harrington, H. A., Feliu, E., Wiuf, C., and Stumpf, M. P. (2013). Cellular compartments cause multistability and allow cells to process more information. *Biophysical journal*, 104(8):1824–1831.
- [71] Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.
- [72] Hebb, D. O. (2005). *The organization of behavior: A neuropsychological theory*. Psychology Press.
- [73] Henzinger, T. A., Mateescu, M., and Wolf, V. (2009). Sliding window abstraction for infinite markov chains. In *CAV*, volume 9, pages 337–352. Springer.
- [74] Herring, B. E. and Nicoll, R. A. (2016). Long-term potentiation: from camkii to ampa receptor trafficking. *Annual review of physiology*, 78:351–365.
- [75] Hervagault, J.-F. and Canu, S. (1987). Bistability and irreversible transitions in a simple substrate cycle. *Journal of theoretical biology*, 127(4):439–449.
- [76] Hirsch, M. W., Smale, S., and Devaney, R. L. (2012). *Differential equations, dynamical systems, and an introduction to chaos*. Academic press.
- [77] Hirschi, A., Cecchini, M., Steinhardt, R. C., Schamber, M. R., Dick, F. A., and Rubin, S. M. (2010). An overlapping kinase and phosphatase docking site regulates activity of the retinoblastoma protein. *Nature structural & molecular biology*, 17(99):1051–1057.

- [78] Holstein, K., Flockerzi, D., and Conradi, C. (2013). Multistationarity in sequential distributed multisite phosphorylation networks. *Bulletin of mathematical biology*, 75(11):2028–2058.
- [79] Huang, C.-C., Liang, Y.-C., and Hsu, K.-S. (2001). Characterization of the mechanism underlying the reversal of long term potentiation by low frequency stimulation at hippocampal ca1 synapses. *Journal of Biological Chemistry*, 276(51):48108–48117.
- [80] Ideker, T., Galitski, T., and Hood, L. (2001). A new approach to decoding life: systems biology. *Annual review of genomics and human genetics*, 2(1):343–372.
- [81] Jahnke, T. and Huisinga, W. (2007). Solving the chemical master equation for monomolecular reaction systems analytically. *Journal of mathematical biology*, 54(1):1–26.
- [82] Jensen, A. (1953). Markoff chains as an aid in the study of markoff processes. *Scandinavian Actuarial Journal*, 1953(sup1):87–91.
- [83] Kapoor, S. and Ramesh, H. (1991). Algorithms for generating all spanning trees of undirected, directed and weighted graphs. In *Workshop on Algorithms and Data Structures*, pages 461–472. Springer.
- [84] Karim, M. S., Buzzard, G. T., and Umulis, D. M. (2012a). Secreted, receptor-associated bone morphogenetic protein regulators reduce stochastic noise intrinsic to many extracellular morphogen distributions. *Journal of The Royal Society Interface*, 9(70):1073–1083.
- [85] Karim, S., Buzzard, G. T., and Umulis, D. M. (2012b). Efficient calculation of steady state probability distribution for stochastic biochemical reaction network. *BMC genomics*, 13(Suppl 6):S10.
- [86] Kazeev, V., Khammash, M., Nip, M., and Schwab, C. (2014). Direct solution of the chemical master equation using quantized tensor trains. *PLoS computational biology*, 10(3):e1003359.
- [87] Khanin, R. and Higham, D. J. (2008). Chemical master equation and langevin regimes for a gene transcription model. *Theoretical Computer Science*, 408(1):31–40.
- [88] Kitano, H. (2002). Systems biology: a brief overview. *Science*, 295(5560):1662–1664.
- [89] Klambauer, G. (2012). *Aspects of calculus*. Springer Science & Business Media.
- [90] Klipp, E., Liebermeister, W., Wierling, C., Kowald, A., and Herwig, R. (2016). *Systems biology: a textbook*. John Wiley & Sons.
- [91] Knottenbelt, W. J. (1996). *Generalised Markovian analysis of timed transition systems*. PhD thesis, university of cape town.
- [92] Kulkarni, V., Stan, G., and Raman, K. (2014). *A Systems Theoretic Approach to Systems and Synthetic Biology I: Models and System Characterizations*. A Systems Theoretic Approach to Systems and Synthetic Biology. Springer Netherlands.
- [93] Kuriyan, J., Konforti, B., and Wemmer, D. (2012). *The molecules of life: Physical and chemical principles*. Garland Science.

- [94] Lander, A. D., Lo, W.-C., Nie, Q., and Wan, F. Y. (2009). The measure of success: constraints, objectives, and tradeoffs in morphogen-mediated patterning. *Cold Spring Harbor Perspectives in Biology*, 1(1):a002022.
- [95] Lee, C. E., Ozdaglar, A., and Shah, D. (2013a). Approximating the stationary probability of a single state in a markov chain. *arXiv preprint arXiv:1312.1986*.
- [96] Lee, C. E., Ozdaglar, A., and Shah, D. (2013b). Computing the stationary distribution locally. In *Advances in Neural Information Processing Systems*, pages 1376–1384.
- [97] Lee, S.-J. R., Escobedo-Lozoya, Y., Szatmari, E. M., and Yasuda, R. (2009). Activation of camkii in single dendritic spines during long-term potentiation. *Nature*, 458(7236):299–304.
- [98] Legewie, S., Schoeberl, B., Blüthgen, N., and Herzog, H. (2007). Competing docking interactions can bring about bistability in the mapk cascade. *Biophysical journal*, 93(7):2279–2288.
- [99] Lestas, I., Vinnicombe, G., and Paulsson, J. (2010). Fundamental limits on the suppression of molecular fluctuations. *Nature*, 467(7312):174.
- [100] Lipshtat, A., Loinger, A., Balaban, N. Q., and Biham, O. (2006). Genetic toggle switch without cooperative binding. *Physical review letters*, 96(18):188101.
- [101] Lisman, J., Yasuda, R., and Raghavachari, S. (2012). Mechanisms of camkii action in long-term potentiation. *Nature Reviews Neuroscience*, 13(3):169–182.
- [102] Lisman, J. E. (1985). A mechanism for memory storage insensitive to molecular turnover: a bistable autophosphorylating kinase. *Proceedings of the National Academy of Sciences*, 82(9):3055–3057.
- [103] Lisman, J. E. and Goldring, M. A. (1988). Feasibility of long-term storage of graded information by the ca²⁺/calmodulin-dependent protein kinase molecules of the postsynaptic density. *Proceedings of the National Academy of Sciences*, 85(14):5320–5324.
- [104] Lisman, J. E. and Zhabotinsky, A. M. (2001). A model of synaptic memory: a camkii/pp1 switch that potentiates transmission by organizing an ampa receptor anchoring assembly. *Neuron*, 31(2):191–201.
- [105] Lu, H. E., MacGillavry, H. D., Frost, N. A., and Blanpied, T. A. (2014). Multiple spatial and kinetic subpopulations of camkii in spines and dendrites as resolved by single-molecule tracking palm. *Journal of Neuroscience*, 34(22):7600–7610.
- [106] Lučić, V., Greif, G. J., and Kennedy, M. B. (2008). Detailed state model of camkii activation and autophosphorylation. *European Biophysics Journal*, 38(1):83.
- [107] Ma, R., Wang, J., Hou, Z., and Liu, H. (2012). Small-number effects: a third stable state in a genetic bistable toggle switch. *Physical review letters*, 109(24):248107.
- [108] Maggioni, M., Berger-Wolf, T., and Liang, J. (2013). Gpu-based steady-state solution of the chemical master equation. In *Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW), 2013 IEEE 27th International*, pages 579–588. IEEE.

- [109] Malleshaiah, M. K., Shahrezaei, V., Swain, P. S., and Michnick, S. W. (2010). The scaffold protein ste5 directly controls a switch-like mating decision in yeast. *Nature*, 465(7294):101–105.
- [110] Marek, I. and Szyld, D. B. (2004). Algebraic schwarz methods for the numerical solution of markov chains. *Linear Algebra and its Applications*, 386:67–81.
- [111] Markevich, N. I., Hoek, J. B., and Kholodenko, B. N. (2004). Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. *The Journal of cell biology*, 164(3):353–359.
- [112] Martins, B. M. and Swain, P. S. (2013). Ultrasensitivity in phosphorylation-dephosphorylation cycles with little substrate. *PLoS Computational Biology*, 9(8):e1003175.
- [113] Marx, G. and Gilon, C. (2012). The molecular basis of memory.
- [114] Matsuzaki, M., Ellis-Davies, G. C., Nemoto, T., Miyashita, Y., Iino, M., and Kasai, H. (2001). Dendritic spine geometry is critical for ampa receptor expression in hippocampal ca1 pyramidal neurons. *Nature neuroscience*, 4(11):1086–1092.
- [115] Matsuzaki, M., Honkura, N., Ellis-Davies, G. C., and Kasai, H. (2004). Structural basis of long-term potentiation in single dendritic spines. *Nature*, 429(6993):761–766.
- [116] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092.
- [117] Michalski, P. (2013). The delicate bistability of camkii. *Biophysical journal*, 105(3):794–806.
- [118] Miller, P., Zhabotinsky, A. M., Lisman, J. E., and Wang, X.-J. (2005). The stability of a stochastic camkii switch: dependence on the number of enzyme molecules and protein turnover. *PLoS biology*, 3(4):e107.
- [119] Miller, S. G. and Kennedy, M. B. (1986). Regulation of brain type ii ca²⁺ calmodulin-dependent protein kinase by autophosphorylation: A ca²⁺-triggered molecular switch. *Cell*, 44(6):861–870.
- [120] Milotti, E., Del Fabbro, A., Dalla Pellegrina, C., and Chignola, R. (2007). Dynamics of allosteric action in multisite protein modification. *Physica A: Statistical Mechanics and its Applications*, 379(1):133–150.
- [121] Mirzaev, I. and Gunawardena, J. (2013). Laplacian dynamics on general graphs. *Bulletin of mathematical biology*, 75(11):2118–2149.
- [122] Mullasseril, P., Dosemeci, A., Lisman, J. E., and Griffith, L. C. (2007). A structural mechanism for maintaining the ‘on-state’ of the camkii memory switch in the post-synaptic density. *Journal of neurochemistry*, 103(1):357–364.
- [123] Munsky, B. and Khammash, M. (2006). The finite state projection algorithm for the solution of the chemical master equation. *The Journal of chemical physics*, 124(4):044104.

- [124] Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.
- [125] Nåsell, I. (2011). *Extinction and quasi-stationarity in the stochastic logistic SIS model*. Springer.
- [126] Neal, R. M. (2004). Improving asymptotic variance of mcmc estimators: Non-reversible chains are better. *arXiv preprint math/0407281*.
- [127] Noguchi, J., Matsuzaki, M., Ellis-Davies, G. C., and Kasai, H. (2005). Spine-neck geometry determines nmda receptor-dependent ca²⁺ signaling in dendrites. *Neuron*, 46(4):609–622.
- [128] O’Donnell, C., Nolan, M. F., and van Rossum, M. C. (2011). Dendritic spine dynamics regulate the long-term stability of synaptic plasticity. *Journal of Neuroscience*, 31(45):16142–16156.
- [129] Ottobre, M. (2016). Markov chain monte carlo and irreversibility. *Reports on Mathematical Physics*, 77(3):267–292.
- [130] Pal, R. (2016). Relationships between models of genetic regulatory networks with emphasis on discrete state stochastic models. In *Emerging Research in the Analysis and Modeling of Gene Regulatory Networks*, pages 52–79. IGI Global.
- [131] Parag, K. V. and Vinnicombe, G. (2017). Point process analysis of noise in early invertebrate vision. *PLoS computational biology*, 13(10):e1005687.
- [132] Pena, J. (2004). Ldu decompositions with l and u well conditioned. *Electronic Transactions on Numerical Analysis*, 18:198–208.
- [133] Petrides, A. and Vinnicombe, G. (2017). Enzyme sequestration by the substrate and the consequent drive towards bistability/bimodality: An analysis in the deterministic and stochastic domains. *Manuscript submitted for publication*.
- [134] Petrides, A. and Vinnicombe, G. (2018). Enzyme sequestration by the substrate: An analysis in the deterministic and stochastic domains. *PLoS computational biology*, 14(5):e1006107.
- [135] Plemmons, R. J. (1977). M-matrix characterizations. i-nonsingular m-matrices. *Linear Algebra and its Applications*, 18(2):175–188.
- [136] Potvin-Trottier, L., Lord, N. D., Vinnicombe, G., and Paulsson, J. (2016). Synchronous long-term oscillations in a synthetic gene circuit. *Nature*, 538(7626):514.
- [137] Pratt, J. W. (1979). Finding how many roots a polynomial has in (0, 1) or (0,). *American Mathematical Monthly*, pages 630–637.
- [138] Press, W. H. and Dyson, F. J. (2012). Iterated prisoner’s dilemma contains strategies that dominate any evolutionary opponent. *Proceedings of the National Academy of Sciences*, 109(26):10409–10413.
- [139] Qian, H. (2010). Thermodynamics of markov processes with non-extensive entropy and free energy. *arXiv preprint arXiv:1005.1251*.

- [140] Qvit, N., Disatnik, M.-H., Sho, E., and Mochly-Rosen, D. (2016). Selective phosphorylation inhibitor of delta protein kinase c–pyruvate dehydrogenase kinase protein–protein interactions: application for myocardial injury in vivo. *Journal of the American Chemical Society*, 138(24):7626–7635.
- [141] Raj, A. and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135(2):216–226.
- [142] Remenyi, A., Good, M. C., and Lim, W. A. (2006). Docking interactions in protein kinase and phosphatase networks. *Current opinion in structural biology*, 16(6):676–685.
- [143] Roach, P. J. (1991). Multisite and hierarchal protein phosphorylation. *Journal of Biological Chemistry*, 266(22):14139–14142.
- [144] Sacco, F., Perfetto, L., Castagnoli, L., and Cesareni, G. (2012). The human phosphatase interactome: An intricate family portrait. *FEBS letters*, 586(17):2732–2739.
- [145] Salazar, C. and Höfer, T. (2009). Multisite protein phosphorylation—from molecular mechanisms to kinetic models. *The FEBS journal*, 276(12):3177–3198.
- [146] Samoilov, M. S., Price, G., and Arkin, A. P. (2006). From fluctuations to phenotypes: the physiology of noise. *Sci. STKE*, 2006(366):re17–re17.
- [147] Sandström, M. (2004). *Models of CaMKII activation*.
- [148] Sayut, D., Kambam, P., and Sun, L. (2007). Engineering and applications of genetic circuits. *Molecular bioSystems*, 3(12):835–840.
- [149] Schultz, D., Walczak, A. M., Onuchic, J. N., and Wolynes, P. G. (2008). Extinction and resurrection in gene networks. *Proceedings of the National Academy of Sciences*, 105(49):19165–19170.
- [150] Schuster, P. (2016). *Stochasticity in Processes: Fundamentals and Applications to Chemistry and Biology*. Springer.
- [151] Sharrocks, A. D., Yang, S.-H., and Galanis, A. (2000). Docking domains and substrate-specificity determination for map kinases. *Trends in biochemical sciences*, 25(9):448–453.
- [152] Shenolikar, S. (2012). *Protein phosphorylation in health and disease*. Academic Press.
- [153] Sobczyk, A., Scheuss, V., and Svoboda, K. (2005). Nmda receptor subunit-dependent [ca²⁺] signaling in individual hippocampal dendritic spines. *Journal of Neuroscience*, 25(26):6037–6046.
- [154] Sorensen, R. and Novak, N. (1996). The use of michaelis-menten kinetics in cell biology and physiology teaching laboratories. *Biochemical Education*, 24(1):26–28.
- [155] Stewart, W. J. (1994). *Introduction to the numerical solution of Markov chains*. Princeton University Press.
- [156] Stewart, W. J. (2009). *Probability, Markov chains, queues, and simulation: the mathematical basis of performance modeling*. Princeton University Press.

- [157] Strogatz, S. H. (2014). *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Westview press.
- [158] Sun, Y., Schmidhuber, J., and Gomez, F. J. (2010). Improving the asymptotic performance of markov chain monte-carlo by inserting vortices. In *Advances in Neural Information Processing Systems*, pages 2235–2243.
- [159] Sunkara, V. and Hegland, M. (2010). An optimal finite state projection method. *Procedia Computer Science*, 1(1):1579–1586.
- [160] Suwa, H. and Todo, S. (2010). Markov chain monte carlo method without detailed balance. *Physical review letters*, 105(12):120603.
- [161] Thattai, M. and Van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences*, 98(15):8614–8619.
- [162] Thomson, M. and Gunawardena, J. (2007). Multi-bit information storage by multisite phosphorylation. *arXiv preprint arXiv:0706.3735*.
- [163] Thomson, M. and Gunawardena, J. (2009). Unlimited multistability in multisite phosphorylation systems. *Nature*, 460(7252):274–277.
- [164] To, T.-L. and Maheshri, N. (2010). Noise can induce bimodality in positive transcriptional feedback loops without bistability. *Science*, 327(5969):1142–1145.
- [165] Turitsyn, K. S., Chertkov, M., and Vucelja, M. (2011). Irreversible monte carlo algorithms for efficient sampling. *Physica D: Nonlinear Phenomena*, 240(4):410–414.
- [166] Tutte, W. T. (1998). *Graph theory as I have known it*. OUP Oxford.
- [167] Tyson, J. J., Chen, K. C., and Novak, B. (2003). Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Current opinion in cell biology*, 15(2):221–231.
- [168] Ubersax, J. A. and Ferrell Jr, J. E. (2007). Mechanisms of specificity in protein phosphorylation. *Nature reviews Molecular cell biology*, 8(7):530–541.
- [169] van Kampen, N. G. (1995). *Stochastic processes in physics and chemistry*.
- [170] Vet, S., de Buyl, S., Faust, K., Danckaert, J., Gonze, D., and Gelens, L. (2018). Bistability in a system of two species interacting through mutualism as well as competition: Chemostat vs. lotka-volterra equations. *PloS one*, 13(6):e0197462.
- [171] Vucelja, M. (2014). Non-equilibrium monte carlo: Sampling with irreversible markov chains. In *APS Meeting Abstracts*.
- [172] Vucelja, M. (2016). Lifting—a nonreversible markov chain monte carlo algorithm. *American Journal of Physics*, 84(12):958–968.
- [173] Wang, L. and Sontag, E. D. (2008). On the number of steady states in a multiple futile cycle. *Journal of mathematical biology*, 57(1):29–52.

- [174] Wang, X. (2004). A simple proof of descartes's rule of signs. *The American Mathematical Monthly*, 111(6):525.
- [175] Watson, J. D., Crick, F. H., et al. (1953). Molecular structure of nucleic acids. *Nature*, 171(4356):737–738.
- [176] Whitmarsh, A. and Davis, R. (2000). Regulation of transcription factor function by phosphorylation. *Cellular and Molecular Life Sciences CMLS*, 57(8-9):1172–1183.
- [177] Wilkinson, D. J. (2011). *Stochastic modelling for systems biology*. CRC press.
- [178] Williams, D. D., Marin, O., Pinna, L. A., and Proud, C. G. (1999). Phosphorylated seryl and threonyl, but not tyrosyl, residues are efficient specificity determinants for gsk-3 β and shaggy. *FEBS letters*, 448(1):86–90.
- [179] Willoughby, R. A. (1977). The inverse m-matrix problem. *Linear Algebra and its Applications*, 18(1):75–94.
- [180] Woodard, D. B., Schmidler, S. C., and Huber, M. (2009). Conditions for rapid mixing of parallel and simulated tempering on multimodal distributions. *The Annals of Applied Probability*, pages 617–640.
- [181] Yasuda, R., Nimchinsky, E. A., Scheuss, V., Pologruto, T. A., Oertner, T. G., Sabatini, B. L., and Svoboda, K. (2004). Imaging calcium concentration dynamics in small neuronal compartments. *Sci STKE*, 2004(219):pl5.
- [182] Zhabotinsky, A. M. (2000). Bistability in the ca²⁺/calmodulin-dependent protein kinase-phosphatase system. *Biophysical journal*, 79(5):2211–2221.
- [183] Zwick, L. U. (2013). Lecture notes on “analysis of algorithms”: Directed minimum spanning trees.