Imperial College of Science, Technology and Medicine
Department of Life Sciences

Towards a Non-Equilibrium Thermodynamic Theory of Ecosystem Assembly and Development

Jacob Cook

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Life Sciences of Imperial College London
September 2021

Abstract

Non-equilibrium thermodynamics has had a significant historic influence on the development of theoretical ecology, even informing the very concept of an ecosystem. Much of this influence has manifested as proposed extremal principles. These principles hold that systems will tend to maximise certain thermodynamic quantities, subject to the other constraints they operate under. A particularly notable extremal principle is the maximum entropy production principle (MaxEPP); that systems maximise their rate of entropy production. However, these principles are not robustly based in physical theory, and suffer from treating complex ecosystems in an extremely coarse manner. To address this gap, this thesis derives a limited but physically justified extremal principle, as well as carrying out a detailed investigation of the impact of non-equilibrium thermodynamic constraints on the assembly of microbial communities. The extremal principle we obtain pertains to the switching between states in simple bistable systems, with switching paths that generate more entropy being favoured. Our detailed investigation into microbial communities involved developing a novel thermodynamic microbial community model, using which we found the rate of ecosystem development to be set by the availability of free-energy. Further investigation was carried out using this model, demonstrating the way that trade-offs emerging from fundamental thermodynamic constraints impact the dynamics of assembling microbial communities. Taken together our results demonstrate that theory can be developed from non-equilibrium thermodynamics, that is both ecologically relevant and physically well grounded. We find that broad extremal principles are unlikely to be obtained, absent significant advances in the field of stochastic thermodynamics, limiting their applicability to ecology. However, we find that detailed consideration of the non-equilibrium thermodynamic mechanisms that impact microbial communities can broaden our understanding of their assembly and functioning.

Acknowledgements

Firstly I would like to give thanks to my two supervisors Robert Endres and Samraat Pawar, for their continual dedication to and interest in my project. I would particularly like to thank Robert for his meticulousness and attention to detail, and Samraat for his unceasing enthusiasm and good humour.

Secondly, I wish to thank my fellow researchers, without whom completing this thesis would have been substantively more tricky and less enjoyable. Particular thanks go to Jess McQuade, Harry Cavanagh and Tom Leyshon for making the 3rd floor of SEC a far more pleasant environment to work in. Thanks also go to all members of Pawar lab past and present for contributing to a stimulating discussion environment, particularly Emma Cavan, Dimitrios Kontopoulos, Hsi-Cheng Ho, Tom Smith, Johnathan Zheng, Paul Huxley, Tom Clegg, Alex Christensen, and Olivia Morris. Special thanks go to Tom Smith for always being prepared to answer my (often not terribly well informed) questions about the empirical side of microbial ecology, and to Suhail Islam for answering my (similarly often not terribly well informed) questions about technical computing. In addition, I would like to thank the two master's students whose projects I helped to supervise, Miles Nesbit and Pablo Lechón-Alonso, who I learnt a great amount through working with.

Thirdly, I wish to give thanks for the people whose presence in my life has helped me through the often gruelling process of doing a PhD. Foremost, my partner Josie Taylor for her love, care and exuberance. That is in addition to being the funniest and most intellectually stimulating person I could imagine being confined with in a small flat during a global pandemic. Particular thanks also goes to Liam Croot and Harrison Finch, for being such caring friends and such lovely people to visit. Thanks are also due to the many other people's whose kindness and friendship has helped me throughout the PhD, particularly, Charlotte Alton, Adriana Elgueta, Sybil Grimsey, Jack Hobbs, Eoin Maguire, Tobi Onasanya, and Alex Taylor. I would also like to offer massive thanks to my parents Margaret and Alan for providing me with unwavering support through the decades of my education.

Finally, I wish to thank the NERC Centre for Doctoral Training in Quantitative and Modelling Skills in Ecology and Evolution (QMEE CDT), for providing the funding necessary to support this project.

Copyright Statement

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Statement of Originality

I declare that the work within this thesis was performed and prepared by myself under the supervision of Robert G. Endres and Samraat Pawar. All sources of information used in the preparation of this thesis are appropriately acknowledged and cited. In all cases where work from this thesis has been previously published I was the first author.

Though all work presented in this thesis is my own, the development in my thinking that led to it has inevitably been shaped by interaction and discussion with many other scientists, particularly my supervisors, my fellow PhD students, and the Master's students whose projects I assisted in supervising.

A previous version of **Chapter 2** was made available online on the arXiv pre-print server as "Thermodynamics of switching in multistable non-equilibrium systems" (arXiv:1908.07405), and a previous version of **Chapter 3** was made available online on the bioRxiv pre-print server as "Thermodynamic constraints on the assembly and diversity of microbial ecosystems are different near to and far from equilibrium" (doi: 10.1101/2021.04.19.440392), and on the arXiv pre-print server under the same title (arXiv:2111.08580).

Dedication

Til min morfar Erik Juel Nielsen, der altid støttede mig og opmuntrede mig til at studere verden. Jeg vil altid huske de sjove, spændende og hyggelige timer vi tilbragte sammen på bondegården. Nu skal jeg studere jordfosfor, som du foreslog mig for mange år siden. Tusind tak!

‘A study of humanity and nature is thus a study of systems of energy, materials, money, and information. Therefore, we approach nature and people by studying energy systems networks.’

Howard T. Odum [Odu07]

Contents

Abstract **i**

Acknowledgements **iii**

1 Introduction **1**

1.1 The relevance of thermodynamics to ecology 1

1.1.1 Thermodynamic extremal principles in ecology 4

1.1.2 Stochastic thermodynamics 6

1.1.3 Previous attempts to base extremal principles in stochastic thermodynamics 10

1.1.4 Application of specific thermodynamic mechanisms in ecology 11

1.1.5 Recent developments in microbial community ecology 12

1.2 Motivation and Objectives 15

1.3 Publication Status 16

2 An extremal principle for switching in simple bistable systems **17**

2.1 Introduction 17

2.2 Developments in stochastic thermodynamics 18

2.3 Merging two approaches for state switching 19

2.3.1 Microscopic fluctuation theorems 21

2.3.2 Macroscopic Langevin coarse-graining 26

2.4 Approximating entropy production (EP) and flow (EF) terms 29

2.4.1 Comparison with exact simulation 33

2.5 One-dimensional switching relation 34

2.6 Switching in the Schlögl and toggle switch models 35

2.7 Derivation of a limited form of MaxEPP 39

2.7.1 Multi-dimensional switching relation 43

2.8 Discussion 48

2.9 Outlook 49

3 Development of a novel thermodynamic microbial community model 51

3.1 Introduction 51

3.2 Thermodynamic constraints on microbial ecosystems 52

3.3 The Model 54

3.3.1 Proteome partitioning 55

3.3.2 ATP production 59

3.3.3 Free-energy dissipation and thermodynamic inhibition 61

3.3.4 Emergence of novel interaction types 63

3.4 Simulations 66

3.5 Results 67

3.5.1 Final community states depend on free-energy availability 67

3.5.2 Free-energy availability increases rate of niche generation 69

3.5.3 Thermodynamic interaction types influence community assembly and diversity . . . 73

3.6 Discussion . . . 76

3.7 Outlook . . . 82

4 Thermodynamic and proteomic trade-offs in assembling communities 84

4.1 Introduction . . . 84

4.2 Extension of thermodynamic microbial community model . . . 87

4.3 Key thermodynamic and proteomic trade-offs . . . 89

4.3.1 Reaction number vs rate . . . 92

4.3.2 Translation rate vs efficiency . . . 93

4.3.3 ATP rate vs yield . . . 93

4.4 Assembly simulations . . . 94

4.5 Results . . . 96

4.5.1 Trade-offs in the absence of immigration . . . 96

4.5.2 Dynamics of continuous species immigration . . . 97

4.5.3 Optimal number of reactions changes as the community assembles . . . 100

4.5.4 Outcome of translation rate vs efficiency trade-off depends on maintenance cost . . . 101

4.5.5 The ATP rate vs yield trade-off drives the late stage development of the community . . . 105

4.6 Discussion . . . 107

5 Synthesis **110**

5.1 Viability of extremal principles in ecological modelling 111

5.2 Development of a novel thermodynamic microbial community model 113

5.3 Intervention into broader debates in microbial ecology 114

5.4 Conclusion 116

A Parameter values used in figures and simulations **117**

A.1 Parameters for Figure 2.1 117

A.2 Parameters for Figure 2.6 117

A.3 Parameters used for simulations in Chapter 3 118

B Further details of models used in Chapter 2 **120**

B.1 Concentration-constrained Schlögl model 120

B.2 Flux-constrained Schlögl model 121

B.3 Concentration-constrained toggle switch 123

B.4 Flux-constrained toggle switch 125

B.5 f1 and f2 values for the models 126

B.5.1 Schlögl model 126

B.5.2 Toggle switch model 127

C Algorithms to calculate minimum action paths **128**

C.1 Minimum action path 128

C.2 Geometric minimum action path method 130

C.2.1 Algorithmic procedure 133

D Collection of graphs supporting arguments made in Chapter 2 **135**

D.1 Generality of relation between Langevin and reduced master equation 135

D.2 The source of divergences from the multi-dimensional switching relation 138

E Detailed derivation of the thermodynamic microbial community model **142**

E.1 Irreversible Michaelis–Menten kinetics 142

E.2 Reversible enzyme kinetics 143

E.3 Proteome partitioning model 146

E.3.1 Derivation of simplified proteome partitioning model 147

E.3.2 Validation for single population 152

F Quantification and classification of the interactions found in Chapter 3 **154**

F.1 Determining interaction types and strengths 154

F.2 Interactions between functional groups 158

G Analysis of assumptions underlying the model developed in Chapter 3 **162**

G.1 Choosing energy saturation constants 167

G.2 Assumption of single-reactant single-product reactions 169

H Comparison of thermodynamic microbial model with other models **171**

H.1 Previous work addressing violations of competitive exclusion theory 171

H.2 Previous work addressing effect of energy supply on diversity 174

Bibliography **175**

List of Tables

A.1 **Parameters for Fig. 2.1A** 117

A.2 **Parameters for Fig. 2.1B** 117

A.3 **Parameters for Schlögl paths in Fig. 2.6** 117

A.4 **Parameters for toggle switch paths in Fig. 2.6** 118

A.5 **Fixed parameters for model in Chapter 3** 118

A.6 **Reaction kinetic parameters for model in Chapter 3** 119

G.1 **Assumptions underlying model developed in Chapter 3** 167

List of Figures

2.1 **Overview of models** 20

2.2 **Comparison of methods for obtaining entropy productions** 25

2.3 **Apparent entropy production of steady states and switching paths** . . 30

2.4 **Relationship between force and velocity** 32

2.5 **Comparison between approximate entropy production and actual** . . . 34

2.6 **Dependence of actions on path entropy productions** 38

2.7 **Comparison of states and switching paths** 40

2.8 **Effect of diffusion on action** 41

2.9 **Restriction of feasible paths** 44

2.10 **Comparing action components between paths** 47

3.1 **Overview of the thermodynamic microbial community model** 56

3.2 **Syntrophy leads to favouring more efficient species** 64

3.3 **Pollution can render species non-viable** 65

3.4 **Key dynamical behaviours of the model** 68

3.5 **Free-energy availability increases species diversity** 70

3.6 **Averaged assembly patterns across communities** 71

3.7 **Thermodynamic and conventional interactions can be of comparable strength** . . . 74

3.8 **Interactions in high free-energy case** . . . 75

4.1 **Maximum ribosome fraction varies substantially with free-energy availability** . . . 90

4.2 **Visualisation of trade-offs** . . . 91

4.3 **The continuous assembly simulation procedure** . . . 95

4.4 **The absence of immigration results in straightforward trade-off optima** 97

4.5 **Summary statistics show average community development** . . . 99

4.6 **Optimal number of reactions changes as community assembles** . . . 102

4.7 **Proteome allocation determined by maintenance cost** . . . 104

4.8 **Communities develops towards thermodynamic optima** . . . 106

B.1 **Schematic of flux constrained models** . . . 127

D.1 **Fig. 2.3 alternative parameterisation** . . . 136

D.2 **Switching in opposite direction to Fig. D.1** . . . 136

D.3 **Fig. 2.3 second alternative parameterisation** . . . 137

D.4 **Switching in opposite direction to Fig. 2.3** . . . 137

D.5 **Switching in opposite direction to Fig. D.3** . . . 138

D.6 **Comparison of path differences** . . . 139

D.7 **Low saddle points (SPs) increase divergence** . . . 139

D.8 **Differences in force along paths** . . . 140

D.9 **Differences in diffusion along paths** . . . 141

E.1 **Validation based on growth laws** 153

F.1 **Heat maps of total interaction strengths between functional groups** . . 160

F.2 **Heat maps of interaction frequency between functional groups** 161

H.1 **Test of competitive exclusion** 173

Chapter 1

Introduction

1.1 The relevance of thermodynamics to ecology

A characteristic feature of all living matter is the maintenance of highly ordered structures. As entropy is a measure of the disorder of a system, these structures are inherently low entropy. In a closed system, due to the 2nd law of thermodynamics these structures would be expected to inevitably degrade into more disordered (higher entropy) forms. Because of this, living matter can only (locally) maintain ordered structures through the continuous production of entropy (creation of disorder) elsewhere. For this reason, substantial study has occurred into the role of thermodynamics in many fundamental cellular (and sub-cellular) processes such as metabolism, sensing, and protein folding, to name just a few.

Less obviously, however, thermodynamics has long been of interest to those researching the very highest level of biological organisation, namely ecosystems. This reflects the intuition that energetic and entropic considerations play a fundamental role in structuring interactions at every biological scale. The beginning of the theoretical trajectory that develops this viewpoint is often traced back to Erwin Schrödinger's landmark book "What is Life?" [Sch44]. Here, building upon Boltzmann, Schrödinger makes the argument that all living things are fundamentally competing for negative entropy. That is, living things are in constant competition for sources of energy that are low in entropy, i.e. highly ordered and therefore high in free-

energy. The free-energy obtained by an organism is primarily dissipated as forms of energy with a lower free-energy content (e.g. heat), producing entropy. This continuous entropy production allows living things to decrease their internal entropy, whilst not violating the 2$^{\text{nd}}$ law of thermodynamics. As the synthesis of biomass fundamentally requires the production of low entropy structures, this reduction of internal entropy is integral to the growth of living things. Schrödinger's insight potentially shows the path towards determining how physical (thermodynamic) necessity constrains biological contingency, applicable to both large scale (ecological) and long term (evolutionary) processes.

Schrödinger's insight is predated by the early work of Alfred Lotka (who later developed the Lotka–Volterra equations). This was first formulated in a pair of papers published in 1922 [Lot22a, Lot22b]. Lotka argued that all organisms are fundamentally selected by their ability to capture (free) energy [Lot22a], and also posited natural selection as a 4$^{\text{th}}$ law of thermodynamics [Lot22b]. This thermodynamic approach to evolution was one of a number elements addressed in Lotka's foundational mathematical ecology text "Elements of Physical Biology" [Lot25]. Other elements of the text, particularly about predator-prey dynamics, have been significantly more widely adopted. However, given that Lotka chose to incorporate these elements alongside a further development of his thermodynamic view of evolution, it seems reasonable to conclude that thermodynamics was central to his conception of biological processes. Given the importance of Lotka to the early development of mathematical ecology [Kin95], this highlights the role thermodynamically-informed thinking played in the early development of the science [Kin15].

This type of thermodynamically-informed thinking can be seen to be continued in the work of Howard T. Odum, widely viewed as one of the originators of the ecosystem concept. Building upon the work of Lotka, Odum integrated thermodynamics into his ecological theory. Specifically, through the idea that biological systems, at any level, that obtain and use energy at a greater rate are selected for [OP55], i.e. ecosystems that both generate and use more power are favoured. This concept of power is fundamental to Odum's understanding of the functioning of ecosystems, as evidenced by the central role it is given in his books aimed at a general audience, such as "Environment, Power, and Society for the Twenty-First Century" [Odu07]. Beyond his

work specifically on thermodynamics, it is clear that an appreciation of the constraints placed on ecosystems by thermodynamics informed Odum's concept of an ecosystem which he defined in terms of stores and fluxes of energy and matter. This can be seen throughout Odum's published work, including his famous Silver Springs study [Odu57].

The role of thermodynamics in ecology remains an active area of research. Particularly notable is the proposal that ecosystems should be understood as produce entropy at the maximum possible rate [Dew03]. This has been applied in a wide range of areas such as vegetation modelling [dJFRRI12], the marine biogeochemical cycle [VA16], and plant optimisation theories [Dew10]. It has also been proposed that thermodynamic theory offers the solution to question of why ecosystems generally show far greater stability than predicted by conventional ecological theory [Mic05]. Also worthy of mention is the hypothesis that the origin of life was facilitated by the fact that living systems can act as catalysts to increase the total production of entropy in their environments [Mic11].

As discussed above, the thermodynamic view of ecosystems has historically played a significant role in the development of ecological theory, and the study of the ecological implications of the constraints arising from thermodynamics remains a relatively small but active area of research to this day. In this thesis I seek to evaluate present approaches applying non-equilibrium thermodynamics to ecology, assessing both the robustness of their theoretical foundations and their applicability to ecology. Focusing on microbial communities, I propose extensions to this present body of theory drawing upon more recent developments in thermodynamics, particularly in the areas of non-equilibrium and stochastic thermodynamics. These extensions consider how thermodynamic constraints effect ecosystems not just at steady-state but through the processes of ecosystem assembly and development.

To this end, in the rest of this introduction I will assess the current state of the literature of applying thermodynamic techniques to ecology and introduce the basic elements necessary for the construction of a novel non-equilibrium thermodynamic theory of ecosystem development.

1.1.1 Thermodynamic extremal principles in ecology

When a system under particular constraints tends to either maximise or minimise an objective function, an extremal principle can be said to be operating. The objective function does not need to be thermodynamic; for example, fitness maximisation is an extremal principle. However, thermodynamic objective functions are frequently chosen in order to, at least loosely, provide a physical justification for extremal behaviour. The most famous thermodynamic extremal principle is the second law of thermodynamics, which states that the entropy of an isolated system can never decrease and instead increases towards a maximum value. This law arises from equilibrium thermodynamics, though it retains its validity away from equilibrium. However, ecosystems, and biological systems more generally, operate far-from-equilibrium. This means that non-equilibrium thermodynamic properties such as entropy production can instead be considered. Owing to the difficulty in applying thermodynamic analysis directly to the huge variety of processes making up an ecosystem, and inspired by the success of the second law, a great deal of the ecological thermodynamics literature pertains to non-equilibrium thermodynamic extremal principles.

The most widely used extremal principle in ecology is the maximum entropy theory of ecology [PAS06]. This extremal principle derives from the second law of thermodynamics, and essentially functions as a statistical inference method. Within this theory the assumption is made that the distributions describing the ecological populations are equilibrated. The distributions describing the state would then be expected to have maximal information entropy subject to the constraints on the system [Jay57] (the "Maxent principle"). This approach can be used to predict species distribution patterns, with environmental factors used as constraints on the maximisation. Further incorporation of constraints such as total metabolic rate leads to more accurate prediction of species distributions [HN14]. The theory has been found to successfully predict species abundance and individual size distributions, but poorly predicts intraspecific size distributions and the relation between species size and abundance [XMW15]. Though the use of Maxent, an equilibrium extremal principle, has generated successful results, the intrinsically non-equilibrium nature of ecosystems (i.e. their continual production of entropy) calls for

non-equilibrium principles.

In all ecosystems, large fluxes of energy are occurring. These fluxes travel from highly ordered energy forms (e.g. sunlight) to eventual highly disordered forms (e.g. heat). These fluxes are therefore necessarily accompanied by entropy production (free-energy dissipation). It's therefore unsurprising that one of the first ecological extremal principles was flux maximisation, proposed by Alfred Lotka in 1922 [Lot22a]. Lotka argued that systems that can support larger energy fluxes, and species that draw larger energy fluxes would both be favoured by natural selection. This principle was not based upon any established thermodynamic theory, rather the assumption that evolution would necessarily favour species able to draw upon larger energy fluxes. In addition, although energy flux is a quantity clearly linked to thermodynamics it lacks a clear thermodynamic interpretation, in contrast to the concepts like work, free-energy, temperature, enthalpy, and chemical potential.

As previously mentioned, Howard Odum also advanced an extremal principle, which built upon the principle proposed by Lotka. This was termed the maximum power principle [OP55]: that systems and species maximise the rate of production of work from available energy to drive processes useful for maintenance and reproduction. Though power is a quantity with a clear thermodynamic interpretation, no physical argument for its maximisation was made. Odum's later work moves away from the clearly defined concept of power, to a far less rigorous concept of empower, which can be defined as the total rate of free-energy use required to sustain all the processes required to sustain a higher-level process [Odu88]. For example, in this framework the thermodynamic change involved in a fox eating a rabbit also incorporates the change involved in the rabbit eating grass and in the photosynthesis required to grow that grasses. While approaches like this have a clear utility in understanding the complexity of ecosystem process, they are substantially too complex to be defined in thermodynamic terms. Altered forms of this concept have been proposed, such as exergy, which defined as total free-energy stored by the system [MJ79]. But these concepts are all either too complex to define in simple thermodynamic terms, or the thermodynamic mechanism underlying their maximisation is not given. Therefore, these principles remain thermodynamic-like rather than being truly thermodynamic extremal principles.

In contrast, the (previously mentioned) maximum entropy production principle (MaxEPP), that ecosystems produce entropy at the maximum possible rate, has been argued to arise from non-equilibrium thermodynamics. This entered the ecological literature via the physical sciences where it had notable success in coarse grained models of planetary climates [Pal75, LLWM01]. This principle predicts that steady states with a greater rate of entropy production will be favoured. A derivation of this principle was obtained by Dewar in 2005 [Dew05], which was subsequently challenged [GL07]. In response to this challenge, it was then claimed by Dewar that MaxEPP represents an inference algorithm for non-equilibrium systems similar to Jaynes maximum entropy inference method (MaxEnt) for equilibrium systems [Dew09]. Establishing whether a more solid foundation for this principle can be established using stochastic thermodynamics is one of the principal aims of this thesis.

Despite this challenge to the theoretical basis of the principle, MaxEPP has been applied across ecology to a wide range of problems. This includes prediction of vegetation distributions in river basins [dJFRRI12], providing a unifying basis for different plant optimisation theories [Dew10], simplifying entire ecosystem biogeochemistry models [Val10, VA16], studying phytoplankton temporal strategies [VT20], providing a novel hypothesis for the origin of life [Mic11], and as an explanation for the huge amount of free-energy consumed by transpiration [Mic12]. Notably, in a recent paper, MaxEPP has successfully been used to predict the chemical concentrations of various organic compounds across a small pond over time [VH18]. It is important to note that there has also been work that challenges the predictions of the principle quantitatively by consideration of simple model systems [MB10].

1.1.2 Stochastic thermodynamics

Within statistical physics, the field of stochastic thermodynamics has (relatively) recently emerged. This field aims to develop a thermodynamic framework that pertains to systems where statistical fluctuations are large enough to substantively impact the dynamics. These are often small systems where thermal noise is comparatively large, but larger systems with high levels of noise can also be considered. As a framework, stochastic thermodynamics is

capable of describing systems that are far-from thermodynamic equilibrium, and as such offers real potential of developing a secure theoretical foundation for MaxEPP (or other extremal principles). Due to this relevance, this subsection will now set out the basics of stochastic thermodynamics.

In contrast to classical thermodynamics, which considers transformations of a whole system, stochastic thermodynamics instead considers single fluctuating trajectories. These trajectories can describe a single particle, e.g. a particle trapped in a potential subject to random noise. Alternatively, they can describe the development of a more complex system in an abstract space, e.g. chemical copy number space. As trajectories are considered, the requirement from classical thermodynamics that only systems at equilibrium can be well characterised no longer applies. It is also worth noting that prior non-equilibrium thermodynamic theory applied only to systems with an approximate linear relationship between the size of a thermodynamic force and the flux that it induces [Ons31], which is generally achieved by systems near to a local equilibrium. This is not the case for stochastic thermodynamics, which is capable of characterising trajectories arbitrarily far-from thermodynamic equilibrium.

Essentially any thermodynamic quantity originally defined in classical thermodynamics can be defined either at the level of a single trajectory, or as an average over an ensemble of trajectories. Entropy production can be defined at both the ensemble and single trajectory levels, with its definition at ensemble level being the average over many individual trajectories. The ensemble averaged entropy production is strictly greater than zero, in accordance with the 2^{nd} law of thermodynamics. In order for the entropy production of a trajectory to be calculated, the assumption is made that the system is coupled to a substantially larger medium. The system can then generate entropy in one of two ways, either by increasing the amount of heat in the medium, or by increasing its own degree of internal disorder. However, as all processes are stochastic the system can draw free-energy from the medium without increasing its internal disorder, and can similarly decrease its internal disorder without dissipating free-energy to the medium as heat. Both of these processes result in a net negative entropy production. This negative entropy production is, however, transient, meaning that entropy production is non-negative both for ensembles of trajectories and for longer trajectories.

An important distinction is made in stochastic thermodynamics between entropy production that increases the disorder of the system and entropy production in the medium. While the first of these is commonly used, the concept of medium entropy production is less frequently discussed. This concept was developed by building upon the work of Ilya Prigogine [KP14], who divided the entropy change in a system into two components as

$$\Delta S = \Delta S_e + \Delta S_i, \tag{1.1}$$

where ΔS_e is the flow of entropy into (or out of) the system, and ΔS_i is the production of entropy within the system. From this definition, it can be seen that a system can continuously produce entropy without increasing its internal entropy, provided that it is matched by a greater flow of entropy out of the system, i.e.

$$-\Delta S_e \geq \Delta S_i. \tag{1.2}$$

This entropy flow is often conceived in terms of heat dissipation, but can equally be a transfer of chemical particles between the system and the medium. An expression for total entropy production can be obtained by rearranging Eq. 1.1 as

$$\Delta S_i = \Delta S - \Delta S_e = \Delta S + \Delta S_m, \tag{1.3}$$

where ΔS_m is the entropy production in the medium (i.e. the flow of entropy from the system to the medium). This entropy production in medium accounts for the vast majority of entropy produced by a system over a long trajectory.

The total production of entropy along a trajectory can be calculated using the fluctuation theorem first obtained by Evans and Searles [SE99]. Here, entropy production is calculated based on the ratio of the probability of observing the path with the probability of observing the time-reversal of the path. This can be performed for continuous trajectories by making using of the Fokker–Planck equation [Sei05]. For discrete trajectories it can be defined quite

straightforwardly as

$$\Delta S_\Gamma = \ln\left(\frac{P(\mathbf{X}_0)}{P(\mathbf{X}_N)}\right) + \ln\left(\frac{W_\Gamma}{W_{\overline{\Gamma}}}\right) \tag{1.4}$$

where N is the number of steps in the trajectory, $\mathbf{X}_0$ is the starting point of the trajectory, $\mathbf{X}_N$ is the end point of the trajectory, and $P(\mathbf{X}_i)$ is the steady state probability of finding the system at position $\mathbf{X}_i$. Further, W_Γ and $W_{\overline{\Gamma}}$ are the product of the transition probabilities for the trajectory and its time-reversal, respectively. They are given by

$$W_\Gamma = W_{\rho_1}(\mathbf{X}_0|\mathbf{X}_1)\ldots W_{\rho_N}(\mathbf{X}_{N-1}|\mathbf{X}_N) \tag{1.5}$$

$$W_{\overline{\Gamma}} = W_{\overline{\rho}_N}(\mathbf{X}_N|\mathbf{X}_{N-1})\ldots W_{\overline{\rho}_1}(\mathbf{X}_1|\mathbf{X}_0), \tag{1.6}$$

where $W_{\rho_i}(\mathbf{X}_{i-1}|\mathbf{X}_i)$ is the probability of transitioning from position $\mathbf{X}_{i-1}$ to $\mathbf{X}_i$ due to the reaction (ρ_i) associated with step i of the trajectory, and $\overline{\rho}_i$ denotes the reaction that is the reverse of ρ_i. This framework allows the entropy production of discrete trajectories, such as those obtained from Gillespie simulation, to be calculated.

A major way by which stochastic thermodynamics has advanced our understanding of thermodynamics in general is through obtaining fluctuation theorems (see [Sei12] for a useful summary). These theorems express basic thermodynamic constraints upon either the ensemble of trajectories or individual trajectories, and a significant number of these theorems have been obtained (see [Jar97, SE99, ES02] for important early examples). Fluctuation theorems that apply to ensembles of trajectories are generally termed integral fluctuation theorems. The fact that over an ensemble of trajectories entropy production is strictly non-negative is an example of an integral fluctuation theorem. Detailed fluctuation theorems apply to individual trajectories, and the existence of a detailed fluctuation theorem necessarily implies the existence of a corresponding integral fluctuation theorem. The fluctuation theorem obtained by Searles and Evans that allows calculation of entropy production along a trajectory is an example of a detailed fluctuation theorem [SE99, ES02], and implies the integral fluctuation theorem proving the non-negativity of ensemble level entropy production. In addition to fluctuation theorems, other stochastic thermodynamic relations have been derived. Notably, the thermodynamic uncertainty relation, which links the amount of free-energy dissipated in a stochastic process

with the uncertainty of in its average output [BS15]. The clear link between these results and entropy production, suggests the potential of stochastic thermodynamics as a framework to develop thermodynamic extremal principles.

1.1.3 Previous attempts to base extremal principles in stochastic thermodynamics

Given this potential, it is natural that attempts to apply stochastic thermodynamics to higher levels of biological organisation have been made. An early attempt at this involved application of fluctuation theorems to a model of self-replication of cells [Eng13]. This work was then further developed into the concept, termed dissipative adaptation, that biological systems tend towards states that effectively absorb free-energy from and dissipate free-energy to their environment [Eng15, PME16]. This is effectively a thermodynamic extremal principle with clear similarities to MaxEPP. However, this principle has only been tested for (theoretical) chemical reaction networks [HE17].

There has additionally been work attempting a derivation of MaxEPP by an alternative method [End15, End17]. This derivation involved application of Jaynes' maximum caliber [Jay80] (a non-equilibrium extension of MaxEnt) with the constraint that the stochastic action is minimised. This formulation of MaxEPP has only been analysed analytically for an extremely simple bistable system [End17], and so the question of the relevance of this principle to more complex systems emerges. In all cases where extremal principles have been developed they have either been derived with minimal grounding in thermodynamic theory or they have not been sufficiently tested in complex systems to be realistically argued to apply to ecosystems. Thus, one of the aims of this thesis is to address this by testing and extending the MaxEPP framework developed in [End17], with the aim of establishing the relevance of extremal principles to contemporary ecology.

1.1.4 Application of specific thermodynamic mechanisms in ecology

Another limitation of the application of extremal principles to ecology is their coarseness. That is, the level of abstraction used is such that detailed consideration of the mechanisms underlying ecosystem dynamics becomes impossible. Thus, while extremal principles can be useful for understanding the development of ecosystems in general their specific explanatory potential is limited. Therefore, in this section I review the applications of non-equilibrium thermodynamics in ecology that pertain to specific mechanisms rather than general principles, with the aim of demonstrating that non-equilibrium thermodynamics has a relevance to ecology beyond coarse extremal principles.

A notable example of the direct application of non-equilibrium thermodynamics to ecology can be found in [Mic05]. In this work, the balance of entropy flow (out of each species) and entropy production (by the species) is considered for a model based upon the commonly used Lotka–Volterra equations. The requirement that this entropy balance is satisfied for all species in turn guarantees the stability of an ecosystem, at least in the absence of higher order interactions (i.e. those mediated by additional species). This result thus appears to offer a potential solution to the problem, originally posed by Robert May, of large random ecosystems being unlikely to be stable [May01]. Verifying and extending this work presents substantial difficulties however, as the entropy production and flow considered cannot easily be measured or even estimated. This is because they are determined by the phenomenological parameters of the Lotka–Volterra equations. Though these parameters are linked to actual ecological mechanisms (e.g. growth), it is hard to see how the entropy production caused by an animal's biomass growth would be measured given the number of discrete processes involved. This issue particularly acute when it comes to calculating the entropy produced by the loosely defined 'interference' processes. For this reason direct application of thermodynamics to animal population models is scarcely found in the literature.

Although microbial community dynamics are often modelled using Lotka–Volterra type equations, they have recently begin to be modelled using frameworks based on the MacArthur consumer-resource model [Mac70]. These microbial consumer resource models are capable

of considering the processes through which microbes obtain the energy in sufficient detail to consider them thermodynamically. Detailed consideration of the impact of thermodynamic end-product inhibition has been made in the case of methanogenic microbial communities [HCR96, LM15, DCWS20]. Here, understanding the limitations that thermodynamics places on growth is necessary to understand under which conditions methanogenesis (a process that yields only small amounts of free-energy) can proceed. A similar modelling approach has also been used to suggest that coexistence of multiple microbial strains on a single substrate is be possible provided they utilise different thermodynamic strategies [GS16]. Consideration of thermodynamics has also led to the idea that there is an intrinsic trade-off between the rate and yield of ATP production at the cellular level [PB02].

The limitation that all approaches to the study of non-equilibrium thermodynamics in microbial communities mentioned so far share is a lack of consideration of the multi-species context of microbial communities, with most studies being limited to at most a few competing species. In the next section I survey recent developments in the area of microbial community ecology, to give context to this specific area, that in the latter chapters of this thesis I will aim to develop novel thermodynamic theory for.

1.1.5 Recent developments in microbial community ecology

Microbial communities are of significant interest to ecology for two main reasons. Firstly, the fact that their properties can change radically even on short timescales (i.e. their generation times are measured in minutes). This, coupled with their small size makes them good candidates for experimental manipulations to test general ecological theories. Secondly, microbial communities play a vital role in many key processes in the biosphere. In particular, the role played in human health by microbial communities (often termed 'microbiomes') has become a huge area of study [FP21]. Microbial communities are also one of the most important mediators of global nutrient cycling, particularly of carbon [WFG$^+$15, NSB$^+$20] and nitrogen [KMK18]. Improving our understanding of microbial communities and their dynamics is thus vital for determining the role that biological systems could play in mitigating or exacerbating anthropogenic climate

change.

In recent years, empirical research into microbial communities has made rapid advances. This is large driven by developments in "omnics" methods (genomics, proteomics, metabolomics, etc), e.g. the decreasing cost of next-generation genome sequencing methods. For example, the relative abundance of different taxa can now be tracked throughout a serial dilution experiment, lending insight into the apparent reproducibility (at certain taxonomic levels) of microbial ecosystem development [GLB+18]. Through carefully measurement of the development of communities on micro-particles, the process of succession (commonly observed for macro-ecosystems) has been observed in a microbial system [DSG+16]. Recent work has also suggested that within communities microbes may be organised into cohesive consortia, as a stable level of organisation between the species and community level [PGBB20]. There has also been significant developments in mapping the metabolic networks of entire communities, a particularly nice example of which being the mapping of the microbial community involved in the production of kefir through the entire process of fermentation [BKM+21].

These empirical advances have spurred substantial developments in the mathematical modelling of microbial ecosystems. Of particular interest here, are the modelling approaches that are extensions of the MacArthur consumer-resource model [Mac70]. In some cases, these extensions are minor, for example adding an "enzyme budget" to allow an investigation of the conditions for coexistence in a relatively simple model ecosystem [PTW17]. Other approaches add the production of resources by consumers, allowing more complex questions about the stability and reproducibility of microbial communities to be addressed [GM18]. Particularly notable is the work originally developed in *Marsland et al.* [MCG+19], where the MacArthur consumer-resource model is extended to include both consumption and production of a large number of resources by every consumer. This results in a complex dynamical system, that arguably can capture the complexity of metabolite production and consumption observed in real microbial communities.

Given the fact that microbial community models can now arguably capture metabolite dynamics, a real possibility of applying thermodynamics to microbial community modelling emerges.

The impact of energy supply on diversity has been previously studied [MCG+19], with higher steady state diversity observed when more energy is either supplied or leaked (due to metabolic inefficiency). This occurs as greater leakage of energy allows more secondary metabolites to reach sufficient concentrations to viably support microbial species, that would otherwise be competitively excluded. Though this approach is promising, it falls short of a full thermodynamic consideration in a number of ways. Firstly, through using Michaelis–Menten kinetics, it implicitly assumes that all reactions the species obtain energy from are sufficiently far from thermodynamic equilibrium that product inhibition effects can be neglected, this is far from true in real systems. Secondly, the growth rate of the strains, and thus the rate of development of the system, is determined simply by the rate at which energy is obtained, with no consideration of the factors that would place bounds upon this in real microbes. Finally, in their model, any metabolite can be converted to any other metabolite in a small number of steps, meaning all metabolites rapidly become available. There is thus no interplay between thermodynamics and the ecological processes of ecosystem assembly and development. These are the three key issues that I seek to address in this thesis by developing a novel thermodynamic microbial consumer-resource model.

Despite significant recent progress many open questions remain in microbial community ecology. Here, I will briefly discuss two of them that would be expected to naturally link to thermodynamics. Firstly, there is the question of assembly, or how best to capture the constant exchange of microbes between systems in a model. There are two predominant approaches to this used at present, either allowing the continual addition of new strains to the systems (bottom-up assembly), or introducing a large number of strains at the beginning of the assembly process and forming the final system from the surviving strains (top-down assembly). These two assembly processes have been shown to generate equivalent results in certain limits [SA21], but outside these limits there is clear potential to make links between the assembly process, constraints on the individual species, and the final ecosystem state. Secondly there is the question of trade-offs, this has long been an area of significant interest in microbial ecology and microbiology more broadly. Recent work has suggested that there is little evidence (in microbes) for some of the traditionally invoked trade-offs such cell mass and genome size [WNG+21]. With this

in mind this thesis will consider the possibility that trade-offs emerge from a combination of thermodynamic constraints and the finiteness of the cellular proteome. As real world microbial ecosystems spend significant periods of time away from steady state it is important to assess the role of these trade-offs throughout the process of assembly. I seek to address this by developing a continuous (bottom-up) assembly procedure, and analysing the results of applying this to the previously developed thermodynamic microbial consumer-resource model.

1.2 Motivation and Objectives

The overarching aim of this PhD thesis is establish the importance of a proper consideration of non-equilibrium thermodynamics to theories of ecosystem assembly and development. This will entail both assessing the relevance of existing thermodynamic theories in ecology, and in extending existing modelling approaches to properly incorporate non-equilibrium thermodynamics. I will use this combined approach to suggest the next necessary steps towards developing a theory of ecosystems solidly grounded in non-equilibrium thermodynamics.

This thesis focuses on a few distinct objectives, each addressed in a separate chapter:

(i) In **Chapter 2** an extremal principle is derived for simple systems using stochastic thermodynamics. The implications of this extremal principle are then investigated for two simple bistable systems. The extensibility of this approach to more complex systems is then discussed. The objective of this chapter is to critically assess the viability of (thermodynamic) extremal principles in ecological modelling.

(ii) In **Chapter 3** a novel non-equilibrium thermodynamic consumer-resource model is developed. This model is then studied for a top-down assembly process, in order to establish the constraints that thermodynamics place upon assembly and thus final diversity of microbial ecosystems. The objective of this chapter is to demonstrate that thermodynamics can usefully be incorporated into models of microbial communities, leading to novel insights.

(iii) In **Chapter 4** the model developed in the prior chapter is utilised for a more complex continuous (bottom-up) assembly scenario. In this assembly scenario the community continues to develop through a process of species sorting (or ecological succession) for a long period of time. This continual development allows consideration of how optimal microbial strategies shift as the overall community develops. The objective of this chapter is to explore how thermodynamic considerations can shed light into broader debates in microbial ecology, such as those around the relevance of trade-offs.

1.3 Publication Status

Chapter 2 has been published in the *Journal of Chemical Physics* [CE20]. Additionally, **Chapter 3** has been published in *PLOS Computational Biology* [CPE21].

Chapter 2

An extremal principle for switching in simple bistable systems

2.1 Introduction

This chapter takes up the task of deriving novel extremal principles. To be of value to ecology, these principles have to be both solidly grounded in thermodynamic theory and ecologically relevant. In order to provide a solid grounding in thermodynamics, the attempts to derive a novel extremal principle shall be based in the recently developed field of stochastic thermodynamics. This field considers the thermodynamics of small fluctuating systems and of statistical fluctuations within larger systems. There have been many recent advances in this field, particularly in the area fluctuation theorems (well summarised in [Sei12]). Thus, stochastic thermodynamics appears to offer a novel basis for attempts to derive extremal principles. This chapter aims to take the first step towards a genuinely ecologically relevant principle by deriving and testing a principle for simple bistable systems.

The work presented in this chapter was previously published in the *Journal of Chemical Physics* [CE20]. The rest of this chapter consists of the main text of the previous paper with a number of alterations. Firstly, content previously contained in the supplementary material necessary to understand the results of the main text has been integrated in. The more technical content

of the supplementary material has been added to the thesis as appendices, which are referred to at the relevant points in this chapter. Secondly, a number of changes in wording have been made both for reasons of readability and self-consistency. Finally, an Outlook section has been added in order to situate the results of this chapter within the broader aims of this thesis.

2.2 Developments in stochastic thermodynamics

When Niels Bohr and Erwin Schrödinger asked decades ago whether new physical principles are needed to explain living systems, the answer seemed "no" [Boh33, Sch44]. More recently, however, the field of stochastic thermodynamics with its temporal violations of macroscopic thermodynamic laws at the microscopic scale have provided a new physical perspective on life. Most remarkable corner stones of far-from-equilibrium thermodynamics are the fluctuation theorems and Seifert's thermodynamic uncertainty relation, stressing the important role of entropy production [SE99, ES02, Sei12, BS15]. At equilibrium, detailed balance prohibits any entropy production on average, but far from equilibrium such entropy production is a characteristic feature [BBF$^+$16] and reflects the flow of time [RND$^+$15].

The fluctuation theorem developed by Evans and Searles allows the exact calculation of the entropy production along a trajectory from the time-forward and time-reversed path (corresponding to a movie played backwards), where paths can be calculated from e.g. Gillespie simulations of the underlying chemical master equation [ES02]. The form of this theorem (for a discretized trajectory) is can be found in Section 1.1.2 (along with an extended discussion of fluctuation theorems). However, due to the theorem's intrinsic connection with the time-reversed path, the theorem cannot be used to calculate the probability of a path simply from its entropy production. The situation is different when using the least-action principle, which allows the prediction of the most likely path between two points in a stochastic system from minimising the action (integral over the Lagrangian) [Arn00, AM17]. This is often done with a Langevin approximation of the master equation, such as using stochastic differential equations incorporating noise terms [PCGBP16, dlCPCG$^+$18]. However, now the link to thermodynamics

is less clear as the role of the entropy production is obscured by the action functional.

In this chapter, we combine the fluctuation theorem and least-action principle to address the stability of steady states in non-equilibrium systems. In particular, we will elucidate the roles of steady-state entropy and fluctuations, as well as steady-state and path entropy production in state switching. For this purpose, we use two different low-dimensional minimal models shown in Fig. 2.1. Firstly, the Schlögl model [Sch72, VQ09, End15], which is the simplest model known to exhibit bistability. Here, bistability emerges due to a non-linearity in the decay of chemical X into chemical B, which (subject to parameter choice) can lead the coexistence of a low and high X steady-state. Secondly, the toggle switch model [CA00], a commonly used model in developmental and synthetic biology. Here, bistability emerges as both types of protein repress the genes responsible for producing protein of the opposite type. For both models, external chemical species concentrations are fixed to ensure non-equilibrium behaviour.

2.3 Merging two approaches for state switching

To investigate the thermodynamics of state switching we shall study non-equilibrium bistable systems with macrostates denoted A and B, where both macrostates correspond to sets of microstates in the discrete space of molecule numbers $\mathbf{X}$, which is a vector for multiple chemical species. The assumption is made that no significant amount of time is spent outside these macrostates. In the large volume limit, the process of switching between states can be assumed to be a Poisson process (with exponentially distributed waiting times, see Fig. 2.1B inset). Thus, $\langle \tau_A \rangle = k_{A\to B} \int_0^\infty t \exp(-k_{A\to B} t) dt = k_{A\to B}^{-1}$ where $k_{A\to B}$ is the switching rate from A to B, and similarly for the B state. These switching rates are coarse-grained and will in general have contributions from a number of microscopic paths. The occupation probability of the A state is then given by $p_A = \langle \tau_A \rangle / (\langle \tau_A \rangle + \langle \tau_B \rangle) = 1/(1 + k_{A\to B}/k_{B\to A})$. Hence, such a two-state system is completely described by the ratio of the switching rates. How do we calculate these for actual molecular systems? While switching in equilibrium systems is often determined by Kramer's formula, where the height of the energy barrier matters, the treatment

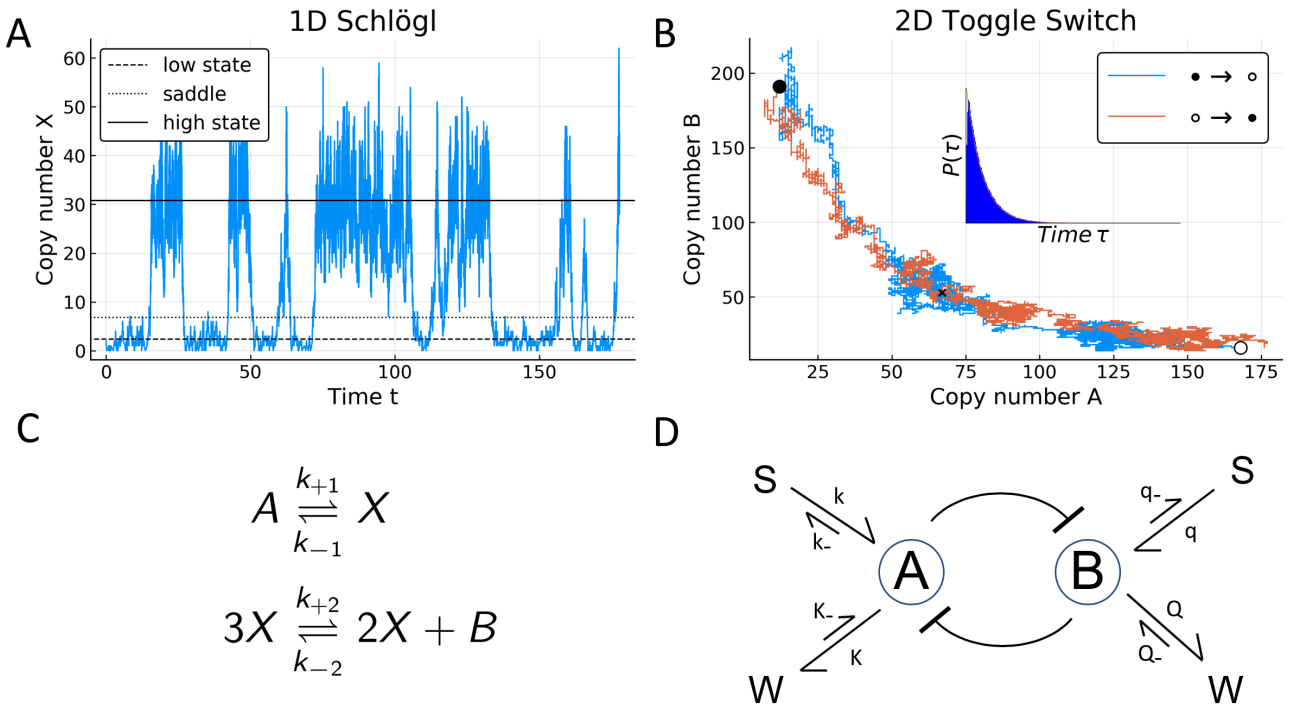


Figure 2.1: **Overview of models.** (a) Example simulation of Schlögl model, displaying switching between high and low X (copy number) states with A and B reservoir species held constant. (b) Example switching paths between A and B states for toggle switch model. (Inset) Waiting time distribution for A to B switching, with a clear exponential distribution beyond a small initial time. (c) Symbolic chemical equations describing the Schlögl model. (d) Schematic illustration of 4 species toggle switch model, consisting of two mutually repressing chemical species (A,B) in addition to substrate (S) and waste (W) species. As the mutual repression between the two chemical species occurs via dimer binding the two rates relevant to this (r and f) are not included in the schematic. The parameter values used to generate the example paths can be found in Appendix A.

of non-equilibrium systems requires entire paths [FZW14, Wan15].

2.3.1 Microscopic fluctuation theorems

Using a microscopic description for non-equilibrium systems, the dynamics can be described by a path Γ, e.g. $\mathbf{X}_0$, $\mathbf{X}_1$, $\mathbf{X}_2$,..., $\mathbf{X}_N$, obtainable from simulations of the chemical master equation [Gas04]. The master equation describes the probability evolution of the microscopic stochastic process and can be formulated as

$$\frac{d}{dt}P(\mathbf{X};t) = \sum_{i=1}^{N_\rho}[W_{\rho_i}(\mathbf{X}-\boldsymbol{\nu}_i|\mathbf{X})P(\mathbf{X}-\boldsymbol{\nu}_i;t) - W_{\overline{\rho}_i}(\mathbf{X}|\mathbf{X}-\boldsymbol{\nu}_i)P(\mathbf{X};t)], \tag{2.1}$$

where $P(\mathbf{X};t)$ is the probability of being in microstate $\mathbf{X}$ at time t and N_ρ is the total number of reactions (including reverse reactions) with $\overline{\rho}_i$ the corresponding reverse reaction to reaction ρ_i. In particular, $W_{\rho_c}(\mathbf{X}_a|\mathbf{X}_b)$ is the transition rate from state a to state b by reaction ρ_c, and $\boldsymbol{\nu}_i$ is the change in molecular copy numbers caused by reaction ρ_i. For every time-forward path, there also exists a time-reversed path $\overline{\Gamma}$, $\mathbf{X}_N$, $\mathbf{X}_{N-1}$,..., $\mathbf{X}_0$. The probability of observing a particular path, e.g. the above time-forward path, is given by the path probability $P_\Gamma = \mathcal{N}P(\mathbf{X}_0)\prod_{i=1}^{N}P(\tau_{i-1,i})W_{\rho_i}(\mathbf{X}_{i-1}|\mathbf{X}_i)$ assuming a continuous-time memory-less Markov process with transition rates $W_{\rho_{i+1}}(\mathbf{X}_i|\mathbf{X}_{i+1})$. Specifically, ρ_i is the reaction that causes step i, $P(\tau_{i,i+1})$ is the probability that there is no reaction in the time interval between τ_i and τ_{i+1}, and $\mathcal{N}$ is a normalisation factor, ensuring $\sum_\Gamma P_\Gamma = 1$.

As we are considering a non-equilibrium steady state (NESS) probability distribution, the total entropy production along the path is given by the steady-state fluctuation theorem (FT) [Sei05, MMB17] as

$$\Delta S_\Gamma = \ln\left(\frac{P(\Gamma)}{P(\overline{\Gamma})}\right), \tag{2.2}$$

where ΔS_Γ is the entropy produced along a trajectory Γ, $P(\Gamma)$ is the probability of the time-forward path Γ, and $P(\overline{\Gamma})$ is the probability of the corresponding time-reversed path. The probability of any path can be decomposed into three components:

1. The steady-state probability of being found at the starting point of the path.

2. The probability of the observed sequence of waiting times, $P_\tau(\Gamma)$.

3. The probability of the transitions that occurred, $P_\mu(\Gamma)$.

The first component is simply the steady-state probability of the initial position $P(\mathbf{X}_0)$ for the time-forward path or the steady-state probability for the final position $P(\mathbf{X}_N)$ for the time-reversed path.

The two remaining components are of more complex form. A compact expression for the second component can be obtained as

$$P_\tau(\Gamma) = e^{-W_{\mathbf{X}_N,\mathrm{tot}}(t_{N+1}-t_N)} \prod_{i=0}^{N-1} W_{\mathbf{X}_i,\mathrm{tot}} e^{-W_{\mathbf{X}_i,\mathrm{tot}}(t_{i+1}-t_i)} \delta t, \tag{2.3}$$

where $W_{\mathbf{X}_i,\mathrm{tot}}$ is the total transition rate out of state $\mathbf{X}_i$ and t_i is the time of transition i (from $\mathbf{X}_{i-1}$ to $\mathbf{X}_i$), with t_0 and t_{N+1} being the start and end times of the path, respectively. Eq. 2.3 also contains a small time interval δt as the probability considered is the probability of waiting for a duration $t_i - t_{i-1}$ and then transitioning in the interval $[t_i, t_i + \delta t]$. This expression is only valid when transition rates have no explicit dependence on time, which is the case in all models we consider.

We can then obtain an expression for the third component as a product of transition probabilities

$$P_\mu(\Gamma) = \prod_{i=0}^{N-1} W_{\mathbf{X}_i,\mathbf{X}_{i+1},\rho_{i+1}} / W_{\mathbf{X}_i,\mathrm{tot}}, \tag{2.4}$$

where $W_{\mathbf{X}_i,\mathbf{X}_{i+1},\rho_{i+1}}$ is the transition rate from state $\mathbf{X}_i$ to state $\mathbf{X}_{i+1}$ by reaction ρ_{i+1}. A combined probability for the time-forward path can now be obtained as

$$\begin{aligned} P(\Gamma) &= P(\mathbf{X}_0) P_\tau(\Gamma) P_\mu(\Gamma) \\ &= P(\mathbf{X}_0) e^{-W_{\mathbf{X}_N,\mathrm{tot}}(t_{N+1}-t_N)} \prod_{i=0}^{N-1} W_{\mathbf{X}_i,\mathbf{X}_{i+1},\rho_{i+1}} e^{-W_{\mathbf{X}_i,\mathrm{tot}}(t_{i+1}-t_i)} \delta t. \end{aligned} \tag{2.5}$$

We now move our consideration onto the time-reversed path, which requires the following

changes. Firstly, the initial state of the path becomes $\mathbf{X}_N$. Secondly, the final state is now $\mathbf{X}_0$ instead of $\mathbf{X}_N$. Finally, transitions are now occurring from state $\mathbf{X}_i$ to $\mathbf{X}_{i-1}$ by reverse reaction $\overline{\rho}_i$, rather than from $\mathbf{X}_i$ to $\mathbf{X}_{i+1}$ by reaction ρ_{i+1}. This leads to an compact expression for the probability of the time-reversed path as

$$\begin{aligned} P(\overline{\Gamma}) &= P(\mathbf{X}_N) P_\tau(\overline{\Gamma}) P_\mu(\overline{\Gamma}) \\ &= P(\mathbf{X}_N) e^{-W_{\mathbf{X}_0,\mathrm{tot}}(t_1-t_0)} \prod_{i=1}^{N} W_{\mathbf{X}_i,\mathbf{X}_{i-1},\overline{\rho}_i} e^{-W_{\mathbf{X}_i,\mathrm{tot}}(t_{i+1}-t_i)} \delta t. \end{aligned} \tag{2.6}$$

Substituting Eqs. 2.5 and 2.6 into Eq. 2.2 leads to an expression for the entropy production as

$$\Delta S_\Gamma = \ln\left(\frac{P(\mathbf{X}_0) e^{-W_{\mathbf{X}_N,\mathrm{tot}}(t_{N+1}-t_N)} \prod_{i=0}^{N-1} W_{\mathbf{X}_i,\mathbf{X}_{i+1},\rho_{i+1}} e^{-W_{\mathbf{X}_i,\mathrm{tot}}(t_{i+1}-t_i)} \delta t}{P(\mathbf{X}_N) e^{-W_{\mathbf{X}_0,\mathrm{tot}}(t_1-t_0)} \prod_{i=1}^{N} W_{\mathbf{X}_i,\mathbf{X}_{i-1},\overline{\rho}_i} e^{-W_{\mathbf{X}_i,\mathrm{tot}}(t_{i+1}-t_i)} \delta t} \right). \tag{2.7}$$

This expression can be significantly simplified by noting that the factors of δt cancel along with the exponential contributions. The overall expression then becomes

$$\Delta S_\Gamma = \ln\left(\frac{P(\mathbf{X}_0)}{P(\mathbf{X}_N)} \right) + \ln\left(\frac{\prod_{i=1}^{N} W_{\mathbf{X}_{i-1},\mathbf{X}_i,\rho_i}}{\prod_{i=1}^{N} W_{\mathbf{X}_i,\mathbf{X}_{i-1},\overline{\rho}_i}} \right) = \ln\left(\frac{P(\mathbf{X}_0)}{P(\mathbf{X}_N)} \right) + \ln\left(\frac{W_\Gamma}{W_{\overline{\Gamma}}} \right), \tag{2.8}$$

with $W_\Gamma = W_{\rho_1}(\mathbf{X}_0|\mathbf{X}_1) \ldots W_{\rho_N}(\mathbf{X}_{N-1}|\mathbf{X}_N)$ and $W_{\overline{\Gamma}} = W_{\overline{\rho}_N}(\mathbf{X}_N|\mathbf{X}_{N-1}) \ldots W_{\overline{\rho}_1}(\mathbf{X}_1|\mathbf{X}_0)$. The entropy considered here is defined as the Shannon entropy of the probability distribution of states $\mathbf{X}$, i.e.

$$S(t) = -\sum_{\mathbf{X}} P(\mathbf{X};t) \ln P(\mathbf{X};t) \tag{2.9}$$

in units where Boltzmann's constant $k_B = 1$. For rare switching, valid for large volume Ω, the steady-state value of S has contributions from S_A with $\mathbf{X} \in A$ and S_B with $\mathbf{X} \in B$.

In order to simplify this calculation for trajectories generated using the Gillespie algorithm [Gil77], a couple of assumptions must be made. Firstly, that for long trajectories differences in probability between the start and end point were small compared to the difference in path probabilities, allowing the first term to be neglected. Secondly, that transition probabilities can be used instead of transition rates with minimal loss of accuracy, as this slightly reduces computational time. Transition probabilities can be defined as $P_{\rho_i}(\mathbf{X}_{i-1}|\mathbf{X}_i) = W_{\mathbf{X}_{i-1},\mathbf{X}_i,\rho_i} / W_{\mathbf{X}_{i-1},\mathrm{tot}}$.

Substituting this into Eq. 2.8 and neglecting the first term leads to a new expression

$$\Delta S_\Gamma = \ln\left(\frac{\prod_{i=1}^{N} P_{\rho_i}(\mathbf{X}_{i-1}|\mathbf{X}_i)}{\prod_{i=1}^{N} P_{\overline{\rho}_i}(\mathbf{X}_i|\mathbf{X}_{i-1})}\right) + \ln\left(\frac{W_{\mathbf{X}_0,\mathrm{tot}}}{W_{\mathbf{X}_N,\mathrm{tot}}}\right), \tag{2.10}$$

where the second term represents the difference between the entropy production we calculate from our Gillespie trajectories and the actual trajectory entropy production. This term becomes vanishingly small in the limit of long trajectories. This expression is equivalent (bar notation) to the expression for the path entropy production found from the Evans and Serales' transient fluctuation theorem [SE99] (see Eq. 1.4). This is because the transient and steady-state fluctuation theorems are equivalent when considering individual paths.

In order to verify that this entropy production (calculated using Eq. 2.10) is the true entropy production, it must be compared with another method. The method we choose to use is the Schnakenberg method [Sch76]. This is a method for calculating the entropy production of a non-equilibrium steady-state. It involves multiplying the chemical fluxes through the steady-state by their corresponding thermodynamic forces. This results in a reasonable approximation for the steady-state entropy production. In order to generate data that could be compared with this method we ran Gillespie simulations of trajectories around each steady state [Gil77]. The modified form of the fluctuation theorem (Eq. 2.10) could then be used to obtain entropy production rates for fluctuations about the steady-states. These were then plotted against the steady-state entropy production rates obtained from the Schnakenberg method in Fig. 2.2, and can be seen to be almost exactly identical.

By restricting our consideration to paths that start within macrostate A and end within B, we can calculate the ensemble-averaged total entropy production $\Delta S_{A\to B} = \sum_{\Gamma|A\to B} P_\Gamma \Delta S_\Gamma = \langle \ln(P(\mathbf{X}_0)/P(\mathbf{X}_N))\rangle_{A\to B} + \langle \ln(W_\Gamma/W_{\overline{\Gamma}})\rangle_{A\to B}$, where the first term on the right-hand side (RHS) corresponds to the average change in entropy of the system between start $\mathbf{X}_0 \in A$ and end $\mathbf{X}_N \in B$ (note N, $\mathbf{X}_0$ and $\mathbf{X}_N$ can vary for different paths). Further, the second term on the RHS corresponds to the flow of entropy from system to medium (here chemical reservoirs), which is termed entropy production in the medium [Sei12]. In the limit of small fluctuations (large Ω, or small noise approximation), the first term becomes negligibly small and the entropy produced

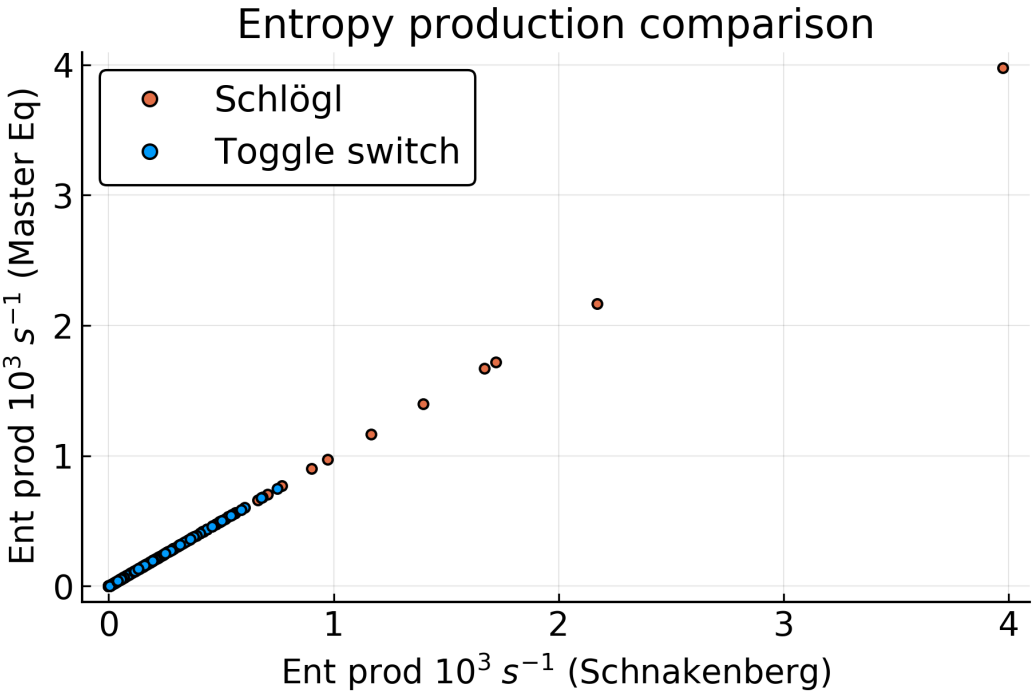


Figure 2.2: **Comparison of methods for obtaining entropy productions.** Entropy productions obtained by applying the fluctuation theorem to master equation trajectories are plotted against entropy productions found by use of the Schnakenberg (flux times affinity) method at steady state. Shown are 100 parameterizations of both the Schlögl model and the toggle switch model colored red and blue, respectively. For each steady-state for each model parameterization Gillespie simulations were run for large volumes ($\Omega = 5000$) to minimize the chance of switching. These simulations were advanced for 10^8 steps and if they had not crossed the saddle point were used. The fluctuation theorem was then applied along the entire trajectory to find the trajectory entropy production, division by the time of the trajectory then gave the entropy production rate.

is given by the second term only, e.g. $\Delta S_{A\to B} \approx \langle \ln (W_\Gamma / W_{\overline{\Gamma}}) \rangle_{A\to B} \approx \ln (k_{A\to B}/k_{\overline{A\to B}})$. This expression is of limited use in determining state switching rates as $k_{A\to B}$ is intrinsically linked to $k_{\overline{A\to B}}$, i.e. the switching rate associated with time-reversal of the paths from A to B. Also note that in general due to the dissipative dynamics the ensemble of time-reversed paths does not correspond to the ensemble of reverse-switching paths, and hence $k_{\overline{A\to B}} \neq k_{B\to A}$ [FZW14]. In order to uniquely determine switching rates we need to utilize a macroscopic formalism.

2.3.2 Macroscopic Langevin coarse-graining

In the macroscopic limit, we can make a continuum approximation of the master equation with the chemical Fokker-Planck equation (FPE), i.e. $X_i = x_i\Omega$ for $i = 1, ..., K$ a K-dimensional chemical system, with concentrations x_i and volume Ω. The chemical Fokker-Planck equation can be derived following Gillespie [Gil00] as

$$\frac{\partial}{\partial t}P(\mathbf{x};t) = -\sum_{i=1}^{N}\frac{\partial}{\partial x_i}\left[f_i(\mathbf{x})P(\mathbf{x};t)\right] + \frac{1}{2}\sum_{i,j=1}^{N}\frac{\partial^2}{\partial x_i \partial x_j}\left[D_{ij}(\mathbf{x})P(\mathbf{x};t)\right], \tag{2.11}$$

where the deterministic force $\boldsymbol{f}(\mathbf{x})$ is given by

$$f_i(\mathbf{x}) = \sum_{j=1}^{N_\rho}\nu_{ji}w_{\rho_j}(\mathbf{x}) \tag{2.12}$$

and the diffusion matrix $\boldsymbol{D}(\mathbf{x})$ is defined as

$$D_{ij}(\mathbf{x}) = \sum_{k=1}^{N_\rho}\nu_{ki}\nu_{kj}w_{\rho_k}(\mathbf{x}). \tag{2.13}$$

In the above definitions, ν_{ji} is the change in molecular copy number of species i that a single occurrence of reaction ρ_j causes, and $w_{\rho_i}(\mathbf{x})$ is the rate (in concentration units) of reaction ρ_i for concentration $\mathbf{x}$, which is the continuum limit of $W_{\rho_i}(\mathbf{X} - \boldsymbol{\nu}_i|\mathbf{X})/\Omega$. This Fokker-Planck equation is generally a reasonable approximation for large (but finite) values of Ω [Gil00], particularly near thermodynamic equilibrium [HGTT84, GTS11]. However, while the accurate prediction of switching rates is difficult, the characterization of relative stability is easier. The

chemical Langevin equation corresponds to this Fokker-Planck equation and can be expressed as $\dot{x}_i = f_i(\mathbf{x}) + g_{ij}(\mathbf{x})\xi_j(t)$ with $\xi_j(t)$ uncorrelated white Gaussian noises of zero mean and autocorrelation $\langle \xi_i(t)\xi_j(t')\rangle = \Omega^{-1}\delta_{ij}\delta(t-t')$. The deterministic force in direction i is given by f_i, and g_{ij} determines the propagation of noise from direction j to i. Here, and throughout this chapter we adopt the convention that repeated indices (in this case j) are summed over. For the models considered in this chapter g_{ij} is always diagonal, representing multiplicative noise. As Eq. 2.11 is an Itô equation, we treat g_{ij} via Itô integration [MM12]. It is worth noting that our definition of entropy is essentially unchanged as we move to the continuum limit, though it is now the Shannon entropy of a continuous probability distribution.

Paths from this Langevin equation can be treated via path integral methods [CCJ06], which can be done more simply in our case. When the probability of escape from a macrostate is sufficiently low, the stochastic transition will be expected to concentrate along a single path $\mathbf{x}^*$, with paths significantly diverging having probabilities so low (for large Ω) as to have negligible impact on overall escape probability [Tou09]. The Wentzel—Kramers—Brillouin (WKB) approximation, the classical analogue of the quantum mechanical WKB [AM17], can then be used to obtain the conditional probability of this path as $P_{A\to B} \sim \exp(-\Omega\mathcal{A}[\mathbf{x}^*])$, where $P_{A\to B}$ is the conditional probability of reaching macrostate B from the initial macrostate A, and $\mathcal{A}$ is the action, as derived in [VF70]. The path $\mathbf{x}^*$ will thus minimize the action $\mathcal{A}[\mathbf{x}^*] = \min \mathcal{A}[\mathbf{x}] = \mathcal{A}_{A\to B}$. This path has a fixed start and end point at the minima of macrostates A and B, respectively. (These points can be reached from any point in their respective macrostates by a zero action path. Thus, the minimum action path $\mathbf{x}^*$ can represent the general path between macrostates.) The relevant action for a path with duration τ is the Freidlin–Wentzell (FW) action [VF70]

$$\mathcal{A}[\mathbf{x}] = \frac{1}{2}\int_0^\tau (\dot{x} - f)_i D^{-1}_{ij} (\dot{x} - f)_j \, dt, \tag{2.14}$$

with the diffusion matrix given by $D_{ij} = g_{ik}g_{kj}$. Any choice of components g_{ij} is acceptable provided that they result in a diffusion matrix that matches Eq. 2.13. As we are considering Itô integration this action possesses one additional term, which can be neglected in the large Ω limit [Tou09]. (If we instead use Stratonovich calculus, three additional terms would appear, which

also disappear in the large Ω limit [Arn00]). Away from this limit, switching paths travelling in opposite directions between two macrostates no longer have to share a saddle point [FZW14]. All of our results will pertain to the large Ω limit. In this limit, the mean-first passage time (MFPT) is given by $T_{A\to B} \sim P^{-1}_{A\to B} = Q_{A\to B}\exp(\Omega\mathcal{A}_{A\to B})$ or $\ln(T_{A\to B}) = \ln(Q_{A\to B}) + \Omega\mathcal{A}_{A\to B}$, so that as Ω grows the contribution from the prefactor $Q_{A\to B}$ becomes less important and the second term on the RHS describes the MFPT to logarithmic precision [BR16, dlCPCG$^+$18]. This approach is often more accurate than treatment based on the FPE obtained by van-Kampen expansion of the master equation [AM17]. While our expression for the MFPT applies in the large deviations limit, alternative forms can be obtained outside of it [FSB05].

An expression for the entropy production based on the time-reversal of Langevin paths can be obtained by noting that the probability of the most probable switching path $A\to B$ is given by $P_{A\to B} = \exp(-\Omega\mathcal{A}_{A\to B})/Q_{A\to B}$. The probability of the corresponding time-reversed path $\overline{A\to B}$ is then found to be $P_{\overline{A\to B}} = \exp(-\Omega\mathcal{A}_{\overline{A\to B}})/Q_{A\to B}$, where $\mathcal{A}_{\overline{A\to B}}$ is the action of the time-reversed path. As factor $Q_{A\to B}$ has not changed, this probability is not in general equal to $P_{B\to A}$, which contains factor $Q_{B\to A}$. Combining the above two expressions as in Eq. 2.8 generates an expression for the entropy production for Langevin paths as $\Delta S^L_{A\to B} = \Omega\,(\mathcal{A}_{\overline{A\to B}} - \mathcal{A}_{A\to B})$ [End17]. Substituting Eq. 2.14 into this relation leads to

$$\Delta S^L_{A\to B} = 2\Omega \int_0^\tau \dot{x}_i D^{-1}_{ij} f_j \, dt. \tag{2.15}$$

Due to our use of a Langevin equation to describe the paths, the above expression is a coarse-grained (apparent) entropy production [MLB$^+$12], which disappears along with the action at steady steady ($\dot{x}_i = f_i(\mathbf{x}) = 0$ in the large Ω limit). This arises as our consideration of the system via the Langevin formalism masks some of the details contained in the master-equation formalism. Namely, the creation (or degradation) of a chemical copy by a production (decay) event cannot be distinguished from creation (degradation) by a reverse-decay (reverse-production) event. This masking of detailed master-equation processes by a more coarse-grained process means that the entropy production calculated is only an apparent entropy production [MLB$^+$12]. Considered via the Langevin formalism our systems are in detailed

balance at steady-state and hence show no apparent entropy production. This zero rate of apparent entropy production implies that the system is quasi-equilibrium. In Section 2.4, we attempt to decompose this apparent entropy production into entropy production and flow terms. Though doing so has obvious illustrative value, in the final instance only the apparent entropy production along a switching path contributes to the state stability.

In Fig. 2.3, a comparison is made between apparent path entropy production and the entropy production of both a reduced model and the full underlying model for a specific parameterization of the toggle switch (further parameterizations are shown in Appendix D.1). This is done both around the steady state and along switching paths. It can be seen that this apparent entropy production approximates that of a reduced model, where the reactions creating a given species can be combined into a single "forward" reaction and likewise for the "reverse" reactions. The entropy production of the switching path for the underlying model is substantially higher (Fig. 2.3E), and at no point generates negative entropy. Zero entropy production at steady state is an expected result for the reduced model as it is an equilibrium model. In contrast, non-zero steady-state entropy production is observed for the underlying model (see Fig. 2.3F). Finally as noted above, the apparent entropy production that we considered does not quantitatively match the entropy production from the time-reversal of the master equation, either at steady state or along the switching paths. That said, the appearance of this Langevin path (apparent) entropy production in our trajectory MaxEPP (Eq. 2.37) suggests that it possesses explanatory potential and is worthy of further study.

2.4 Approximating entropy production (EP) and flow (EF) terms

The fact we are dealing with an apparent entropy production, rather than the entropy production of the full model, is somewhat unsatisfactory. As such, we wish to make use of a similar approach to that taken by Ilya Prigogine [KP14], and divide our apparent entropy production into approximate underlying entropy production and flow terms. The basis for this is described

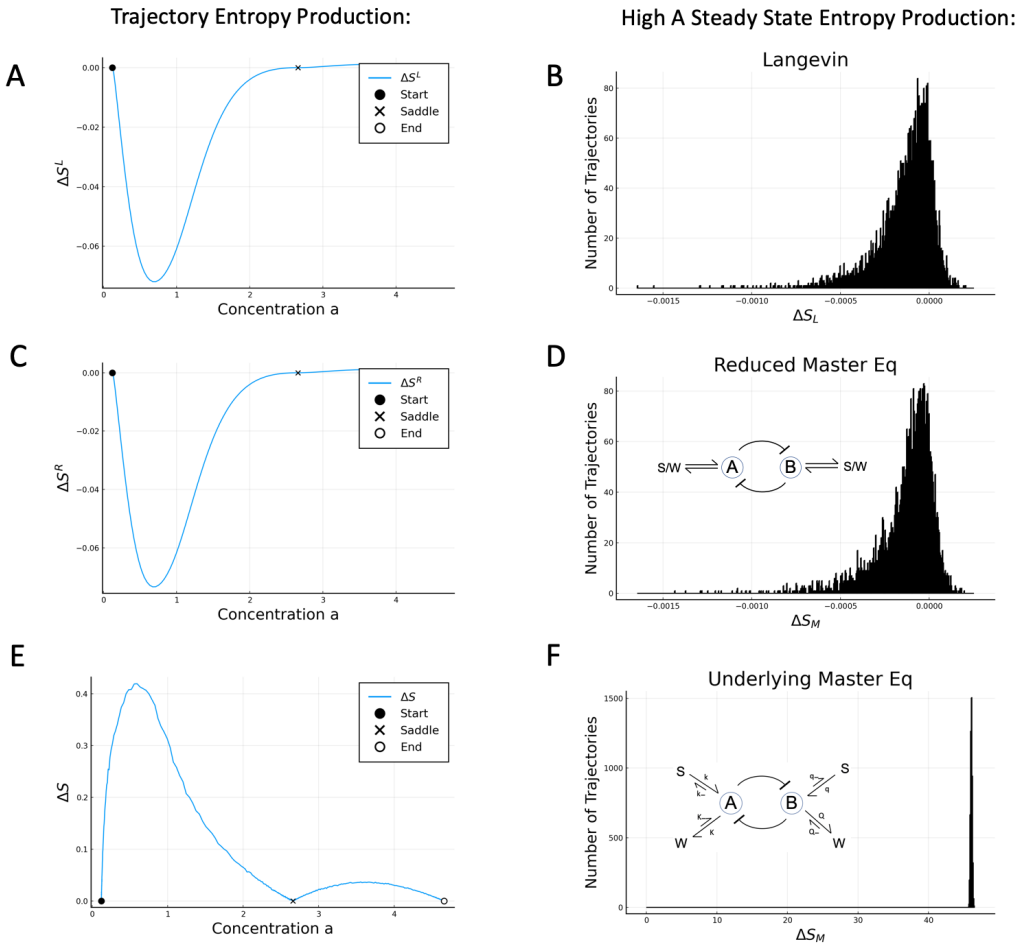


Figure 2.3: **Apparent entropy production of steady states and switching paths.** (a) Apparent entropy production of MAP calculated by reversal of Langevin level path as in [End17]. Entropy production is taken as entropy produced between each pair of points in the $N = 600$ points the MAP was optimized over. (b) Histogram of apparent entropy productions of 500 Gillespie simulations of full bistable switch model around the high A steady state, calculated via the same method. (c) Entropy production of above MAP when projected onto discrete chemical state space. Calculation was performed using transition probabilities from master equation corresponding to reduced model. (d) Histogram of entropy productions of 5000 different Gillespie trajectories calculated via the same method. Inset shows schematic of reduced model. (e) Entropy production of above MAP again projected onto discrete chemical state space, but this time one compatible with the underlying master equation. This projection can only be carried out probabilistically, which results in a (slightly) noisy entropy production curve. (f) Histogram of entropy productions of 5000 different Gillespie trajectories calculated based upon the underlying master equation. Inset shows schematic of the full (underlying) model. All 6 plots are produced using the following parameterization (given to 6 sf): $k = 11.9706$, $k_- = 0.502369$, $q = 5.07189$, $q_- = 0.207451$, $K = 1.74597$, $K_- = 0.0795083$, $Q = 0.439835$, $Q_- = 0.0286310$, $r = 2569.21$, $f = 2574.64$, and $\Omega = 5000$. Appropriate volume (Ω) rescaling of rates was used for the discrete simulations and calculations.

in Section 1.1.2, with the splitting of terms shown in Eq. 1.1. The difficulty in doing this consists in determining how to divide our (Langevin) expression for path entropy production between the underlying entropy production and flow terms. To attempt to do this, we obtain a relation between magnitude of velocity and magnitude of force along the minimum action path. This is done by following the method set out in [JGT18], where the action (Eq. 2.14) is re-expressed in a compact notation as

$$\begin{aligned}\mathcal{A} &= \frac{1}{2}\int_0^\tau \left(|\dot{\boldsymbol{x}}|_d^2 - 2|\dot{\boldsymbol{x}}|_d|\boldsymbol{f}|_d + |\boldsymbol{f}|_d^2\right)\, dt + \int_0^\tau \left(|\dot{\boldsymbol{x}}|_d|\boldsymbol{f}|_d - \dot{\boldsymbol{x}} \bullet \boldsymbol{f}\right)\, dt \\ &= \frac{1}{2}\int_0^\tau \left(|\dot{\boldsymbol{x}}|_d - |\boldsymbol{f}|_d\right)^2\, dt + \int_0^\tau \left(|\dot{\boldsymbol{x}}|_d|\boldsymbol{f}|_d - \dot{\boldsymbol{x}} \bullet \boldsymbol{f}\right)\, dt\end{aligned} \tag{2.16}$$

with $|\boldsymbol{g}|_d^2 = g_i D_{ij}^{-1} g_j$ and $\boldsymbol{g} \bullet \boldsymbol{h} = g_i D_{ij}^{-1} h_j$. The first (magnitude) term in this action reaches its minimum value of zero when $|\dot{\boldsymbol{x}}|_d = |\boldsymbol{f}|_d$. The second (alignment) term depends solely on the alignment of the vector of forces $\boldsymbol{f}$ with the velocity $\dot{\boldsymbol{x}}$. Once the magnitude term has been minimized, changes that decrease the alignment term can either leave the magnitude term unchanged or increase it. Changes that cause an increase will always lead to a decrease in the alignment term that is smaller than the increase of the magnitude term. Therefore for all minimum action paths the following relation would be expected to hold

$$\dot{x}_i D_{ij}^{-1} \dot{x}_j = f_i D_{ij}^{-1} f_j, \tag{2.17}$$

where both sides are summed over i and j. In the case of all non-zero elements D_{ij} being equal or the trivial case of a one-dimensional model (e.g. concentration constrained Schlögl model) this relation reduces to

$$|\dot{\boldsymbol{x}}| = |\boldsymbol{f}|. \tag{2.18}$$

In the concentration constrained toggle switch model the inverse diffusion matrix D^{-1} is no longer uniform i.e. $D_{11}^{-1} \neq D_{22}^{-1}$, but it does not possess cross terms i.e. $D_{12}^{-1} = D_{21}^{-1} = 0$. This means that Eq. 2.18 remains a minimizer of the first part of the action in Eq. 2.16 but is no longer a unique one. We shall proceed by assuming that Eq. 2.18 still approximately holds. We must make an additional assumption in order to obtain an expression linking velocity with

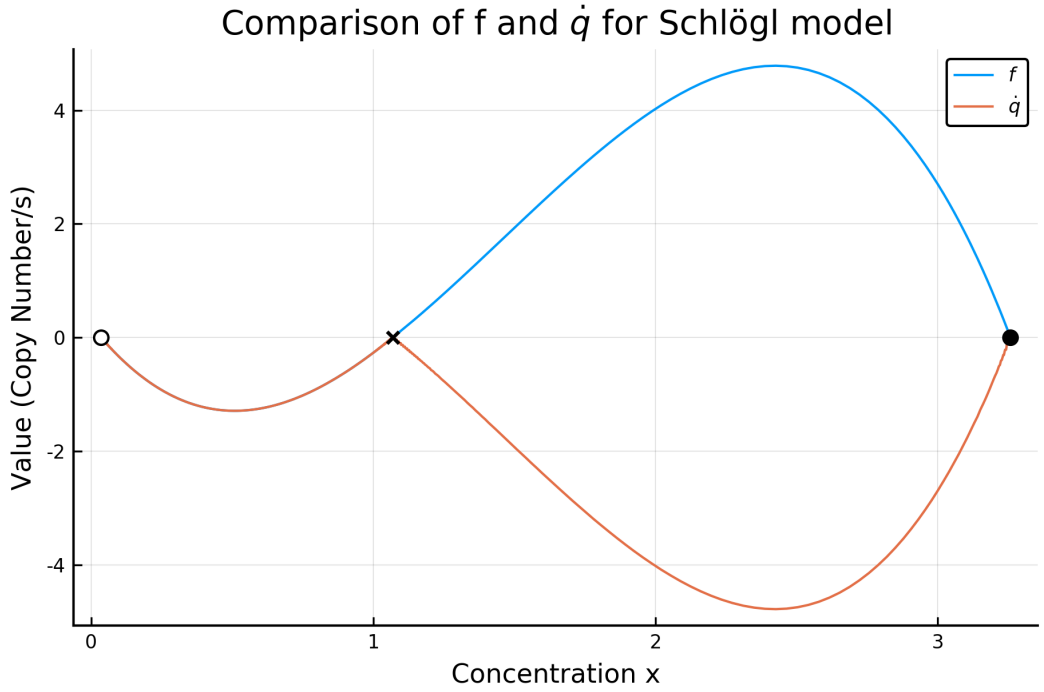


Figure 2.4: **Relationship between force and velocity.** Plot of f and $\dot{x}$ along a minimum action path from high x to low x for the Schlögl model. Moving from the saddle point (black cross) to the final steady state (white dot), f and $\dot{x}$ are identical. From the initial steady state (black dot) to the saddle point, f and $\dot{x}$ can be seen to be equal and of opposite sign. This is in accordance with the relations obtained as Eq. 2.19.

deterministic force. This assumption is that the direction of motion opposes the force when travelling from steady state to saddle point and follows the force when travelling from saddle point to steady state. This can be symbolically expressed as

$$\begin{aligned} \dot{x}_i &\approx -f_i \quad \text{when} \quad t \leq t_s \\ \dot{x}_i &\approx f_i \quad \text{when} \quad t > t_s, \end{aligned} \tag{2.19}$$

where t_s is the time at which the saddle point is reached. Our relation is demonstrated in Fig. 2.4 by plotting force against velocity for a particular Schlögl model minimum action path. Using these approximations alongside a splitting of f into two parts $f = f_1 - f_2$, a new approximate expression for ΔS can be derived as

$$\begin{aligned} \Delta S &= 2\Omega \int_0^t \left(2f_{1,i}D_{ij}^{-1}f_{2,j} - f_{1,i}D_{ij}^{-1}f_{1,j} - f_{2,i}D_{ij}^{-1}f_{2,j}\right) d\tilde{t} \quad \text{when} \quad t \leq t_s \\ \Delta S &= 2\Omega \int_{ts}^t \left(f_{1,i}D_{ij}^{-1}f_{1,j} + f_{2,i}D_{ij}^{-1}f_{2,j} - 2f_{1,i}D_{ij}^{-1}f_{2,j}\right) d\tilde{t} \quad \text{when} \quad t > t_s. \end{aligned} \tag{2.20}$$

These entropy productions correspond to the entropy produced along a single segment. Thus, in order to determine the total entropy produced along the whole path they must be summed. We shall now consider the positive terms in the expression to approximate the entropy production (EP) and the negative terms to approximate the entropy flow (EF). Thus,

$$\Delta S = \Omega \int_0^t (\mathrm{EP} - \mathrm{EF})\, d\tilde{t} \tag{2.21}$$

where

$$\begin{aligned} \mathrm{EP} &= 4 f_{1,i} D_{ij}^{-1} f_{2,j} \quad \text{when} \quad t \leq t_s \\ \mathrm{EP} &= 2\left(f_{1,i} D_{ij}^{-1} f_{1,j} + f_{2,i} D_{ij}^{-1} f_{2,j}\right) \quad \text{when} \quad t > t_s, \end{aligned} \tag{2.22}$$

and

$$\begin{aligned} \mathrm{EF} &= 2\left(f_{1,i} D_{ij}^{-1} f_{1,j} + f_{2,i} D_{ij}^{-1} f_{2,j}\right) \quad \text{when} \quad t \leq t_s \\ \mathrm{EF} &= 4 f_{1,i} D_{ij}^{-1} f_{2,j} \quad \text{when} \quad t > t_s. \end{aligned} \tag{2.23}$$

The values we used for f_1 and f_2 for the Schlögl and toggle switch models are provided in Appendix B.5.

2.4.1 Comparison with exact simulation

In Fig. 2.5 we compare our approximate entropy production (EP) term with exact entropy productions as calculated via the Schnakenberg method or Gillespie simulation. This was done for the parameters sets that will be later used to generate Fig. 2.7. It can be seen from Fig. 2.5A-B that for both models there is a linear relationship between the steady state entropy productions. This relation persists when considering entropy production of switching paths (Fig. 2.5C-D), and in this case many values are approximately identical. It is therefore clear that our approximate entropy production is a reasonable approximation to the true underlying entropy production. This demonstrates that the Langevin entropy production we derived can be decomposed into more conventional thermodynamic variables. Hence, despite the obvious

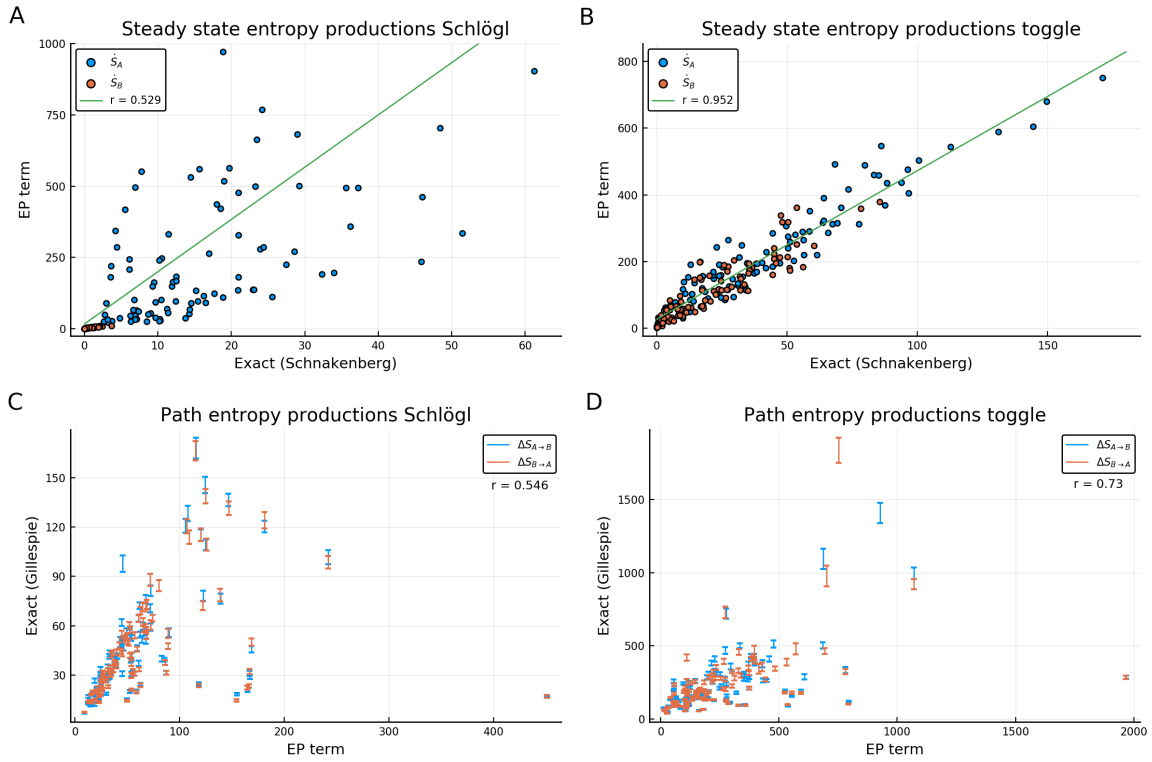


Figure 2.5: **Comparison between approximate (EP) and actual.** (a) A plot of steady state entropy productions calculated using our EP term and by the standard Schnakenberg method. There are some large outliers in this data (outside plotting area), which skew the linear fit and reduce the correlation coefficient. (b) An equivalent plot for the toggle switch model. In this case the correlation is significantly stronger. (c) A plot of the entropy produced by Gillespie simulation for switches between macrostates for the Schlögl model, plotted against our approximate EP term summed along the path length. For a substantial portion of the data our x and y values are approximately equal. (d) An equivalent plot for the toggle switch model. The same approximate correspondence as before can be seen, but due to fewer outliers in this case the weighted correlation (in annotation) is stronger.

difference between the entropy productions from the master equation (Eq. 2.8) and the Langevin paths (Eq. 2.15), the latter has predictive power.

2.5 One-dimensional switching relation

In the special case of one dimensional systems an exact analytic switching relation can be obtained. To do so we shall use a relation that is valid at every point along an action minimising path [JGT18]

$$\sum_{ij} \dot{x}_i D_{ij}^{-1} \dot{x}_j = \sum_{ij} f_i D_{ij}^{-1} f_j, \tag{2.24}$$

which can be interpreted as every point along the minimising path having equal kinetic and potential energy (defined below). In the one-dimensional case this relation simplifies further to

$$\begin{aligned} \dot{x}_1 D_{11}^{-1} \dot{x}_1 &= f_1 D_{11}^{-1} f_1 \\ |\dot{x}_1| &= |f_1|, \end{aligned} \tag{2.25}$$

where || indicates that the magnitude has been taken. This means that both paths have the same speeds along their lengths as they necessarily pass through the same points (in 1D). Thus, for one dimensional systems the path $B \to A$ is identical to the time-reversed path $\overline{A \to B}$. It then follows that for this case $k_{\overline{A \to B}} = k_{B \to A}$, and so the ratio of switching rates are determined solely by the entropy production as

$$k_{A \to B} / k_{B \to A} = \exp(\Delta S^L_{A \to B}). \tag{2.26}$$

When considering the broader class of multi-dimensional systems such simple relations no longer apply and we instead consider numerical approaches to explicit minimal models. This will allow us to investigate how the entropy production varies along the path, and whether diffusion strength and steady-state entropies matter.

2.6 Switching in the Schlögl and toggle switch models

The two models considered are the Schlögl and toggle switch models (Fig. 2.1). For both models concentration-constraints are used to make the models non-equilibrium. The Schlögl model is a simple one-dimensional model that exhibits bistability [Sch72, VQ09]. It involves three chemical species A, B and X, which obey the chemical dynamics shown in Fig. 2.1C. Holding the concentration of external species fixed (i.e. constant a and b), this model can be reexpressed in terms of the dynamics of concentration x as

$$\frac{dx}{dt} = k_{-2} b x^2 - k_{+2} x^3 - k_{-1} x + k_{+1} a + g_x(x) \xi(t), \tag{2.27}$$

where

$$g_x(x) = \sqrt{k_{-2}bx^2 + k_{+2}x^3 + k_{-1}x + k_{+1}a} \tag{2.28}$$

and $(k_{\pm 1}, k_{\pm 2})$ are rate constants. See Appendix B for further details of this model.

The toggle switch model is a two-dimensional model that describes the dynamics of a simple bistable genetic switch [CA00] (see Fig. 2.1D). We consider the adiabatic limit where the fraction of gene copies active is completely determined by the concentration of the corresponding repressing protein (see Appendix B for further details). This means that the model can be expressed in terms of change of protein concentrations a and b as

$$\frac{da}{dt} = \frac{kr}{r + fb^2} - k_- a - Ka + K_- + g_a(a,b)\xi_a(t) \tag{2.29}$$

$$\frac{db}{dt} = \frac{qr}{r + fa^2} - q_- b - Qb + Q_- + g_b(a,b)\xi_b(t), \tag{2.30}$$

where

$$g_a(a,b) = \sqrt{\frac{kr}{r + fb^2} + k_- a + Ka + K_-}, \tag{2.31}$$

and

$$g_b(a,b) = \sqrt{\frac{qr}{r + fa^2} + q_- b + Qb + Q_-}. \tag{2.32}$$

Further, r is the gene activation rate and f is rate constant for gene repression by protein dimers, the other eight rate constants describe the rates of protein production or degradation. Here, the constant external concentrations s and w have been absorbed into the relevant rate constants. Again, see Appendix B for further details of this model.

In order to study the switching between states for these models, we need to find paths that minimise the action (Eq. 2.14). These paths are usually determined by use of a quasi-Newton method (e.g. the L-BFGS algorithm) to find the minimising path for a particular duration τ, with a gradient descent method used to find the value of τ that results in the lowest minimum [PCGBP16]. However, in order to save computational time we made use of the faster (but more complicated) geometric minimum action method [HVE08a, HVE08b, JGT18]. A discussion and derivation of these two alternative algorithms is provided in Appendix C. These paths were then

be used in Eq. 2.14 and Eq. 2.15 in order to generate the corresponding minimum actions and path entropy productions.

Figs. 2.6A&B show exemplar minimum action paths for switching in the Schlögl and the toggle switch model, respectively. As the Schlögl model is one-dimensional its time-reversed switching paths correspond to the switching paths for the contrary direction. Fig. 2.6C shows how this leads to equal and opposite entropy productions (orange lines) along the paths. For the more complicated toggle switch model this simple relation between paths is lost, but Fig. 2.6D shows a linear relation between the difference in minimum action (purple line) and the difference in path entropy productions (gold line). Despite the systems' quasi-equilibrium behavior, non-equilibrium processes still occur. In order to illustrate this fact, we derived approximate entropy production and flow terms (see Section 2.4). Figs. 2.6E&F show plots of the derived entropy production (EP, blue lines) and flow (EF, red lines), demonstrating non-zero contributions at the steady states.

In order to investigate links between occupation probabilities, entropy, and entropy production, 100 random parameters sets were created for each model. Specifically, the procedure for the construction of the toggle switch parameter sets was as follows. First, k was randomly drawn from the (continuous) range $[1, 100]$ and K from the range $[0.1, 10]$. Then random numbers were drawn from the range $[0.001, 0.1]$ to obtain the ratios q/k, k_-/k, q_-/k, Q/K, K_-/K, and Q_-/Q. Finally, r and f were both drawn from the range $[0, 10^4]$. For the Schlögl parameter sets, all four parameters (k_{-1}, $k_{+1}a$, k_{+2}, $k_{-2}b$) were drawn from the range $[0.1, 10]$. In both cases the constructed sets were accepted if they resulted in multiple non-zero steady states. State entropies and occupation probabilities were obtained by direct Gillespie simulation of the chemical master equation [Gil77]. Each of these simulations was run for the largest computationally feasible value of Ω in order to minimise the effect of discretization. Steady-state entropy productions were calculated using the Schnakenberg method (product of flux and reaction affinity) [Sch76].

Fig. 2.7A shows a weak correlation between state occupation probability and steady-state entropy production, which provides some evidence for the maximum entropy production principle

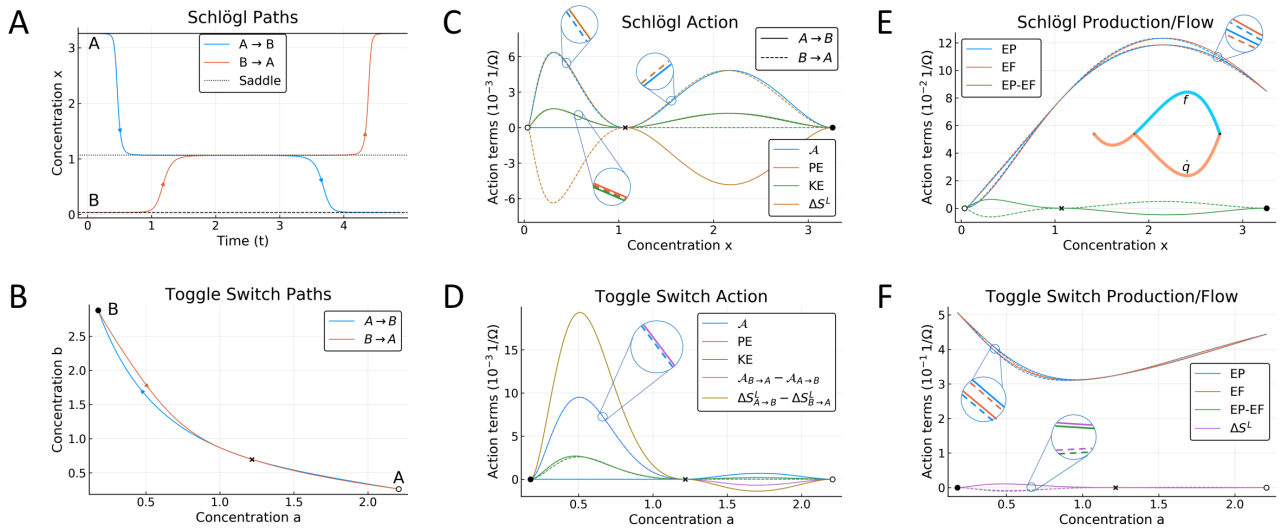


Figure 2.6: **Dependence of actions on path entropy productions.** (a) Minimum action paths for Schlögl model; aside direction, the only difference between paths is the amount of time spent at the fixed points where there is no contribution to the action. (b) Minimum action paths for toggle switch model, now showing clear differences. (c) Action ($\mathcal{A}$), entropy production (ΔS^L), kinetic (KE) and potential energy (PE) of Schlögl paths with KE and PE defined in Eq. 2.36, clearly showing that the entropy production of one path is the opposite of the other. In this and the remaining panels a solid line corresponds to the path from $A \to B$ and a dashed line to $B \to A$. Magnifications are meant to clarify line styles. (d) Action, differences in action and entropy production, KE and PE along the toggle switch paths. The difference in action is proportional to the difference in entropy produced along the path (this linear relationship is further discussed below). (e) Entropy production (EP) and entropy flow (EF) terms along the Schlögl paths. The difference (EP-EF) is equal to the entropy production of time-reversed of Langevin paths. (Inset) Plot showing how f and $\dot{x}$ vary along an exemplar minimum action path. (f) Equivalent plot for the toggle switch paths. The parameter values used to generate the paths can be found in Appendix A.

(MaxEPP) [End17]. This extremal principle proposes that states with higher entropy production are more dynamically stable (subject to other dynamical constraints) [LLWM01, Dew09]. We then approximate the log ratio of state occupation probabilities via the Freidlin–Wentzell theorem as, $\ln(p_A/p_B) \approx \ln(P_{B\to A}/P_{A\to B}) = \ln(Q_{B\to A}/Q_{A\to B}) + \Omega(\mathcal{A}_{A\to B} - \mathcal{A}_{B\to A})$. For large Ω only the second term would be expected to contribute but this limit is difficult to simulate. Simulated occupation probabilities match well with this approximation (see Fig. 2.7B), demonstrating the validity of our use of the FW action. Fig. 2.7C shows a weak correlation between difference in action and difference in state entropy, as expected from equilibrium theory where higher entropy states are more stable. However, state entropy increases sublinearly with Ω so for large Ω it has no effect on the stability. Fig. 2.7D shows a comparison of the difference in action and the difference in path entropy production, showing that the linear relation observed in Fig. 2.6C&D holds generally across parameter sets surveyed. Fig. 2.8 shows that there is no significant relationship between the ratio of diffusion matrix magnitudes and the difference in minimum actions. Ratios and differences are used for this comparison so that significantly different parameter regimes can be compared in a single plot.

2.7 Derivation of a limited form of MaxEPP

Our results suggest a limited form of MaxEPP, which applies to the rate of switching between macrostates. We shall proceed with our derivation by noting that in the large deviations limit (i.e. where transitions between the states are sufficiently improbable), the ensemble of switching paths between two states is predicted to be dominated by paths infinitesimally deviating from a single most probable path. These most probable paths are defined by the fact that they minimize the Freidlin-Wentzell action (Eq. 2.14) [VF70]. The minimum of this action can then be used to determine the probability of a switch between two macrostates, as $P_{A\to B} \sim \exp(-\Omega\mathcal{A}_{A\to B})$. Thus, the ratio of switching rates can be obtained as

$$\frac{P_{B\to A}}{P_{A\to B}} \approx \exp\left[\Omega(\mathcal{A}_{A\to B} - \mathcal{A}_{B\to A})\right]. \tag{2.33}$$

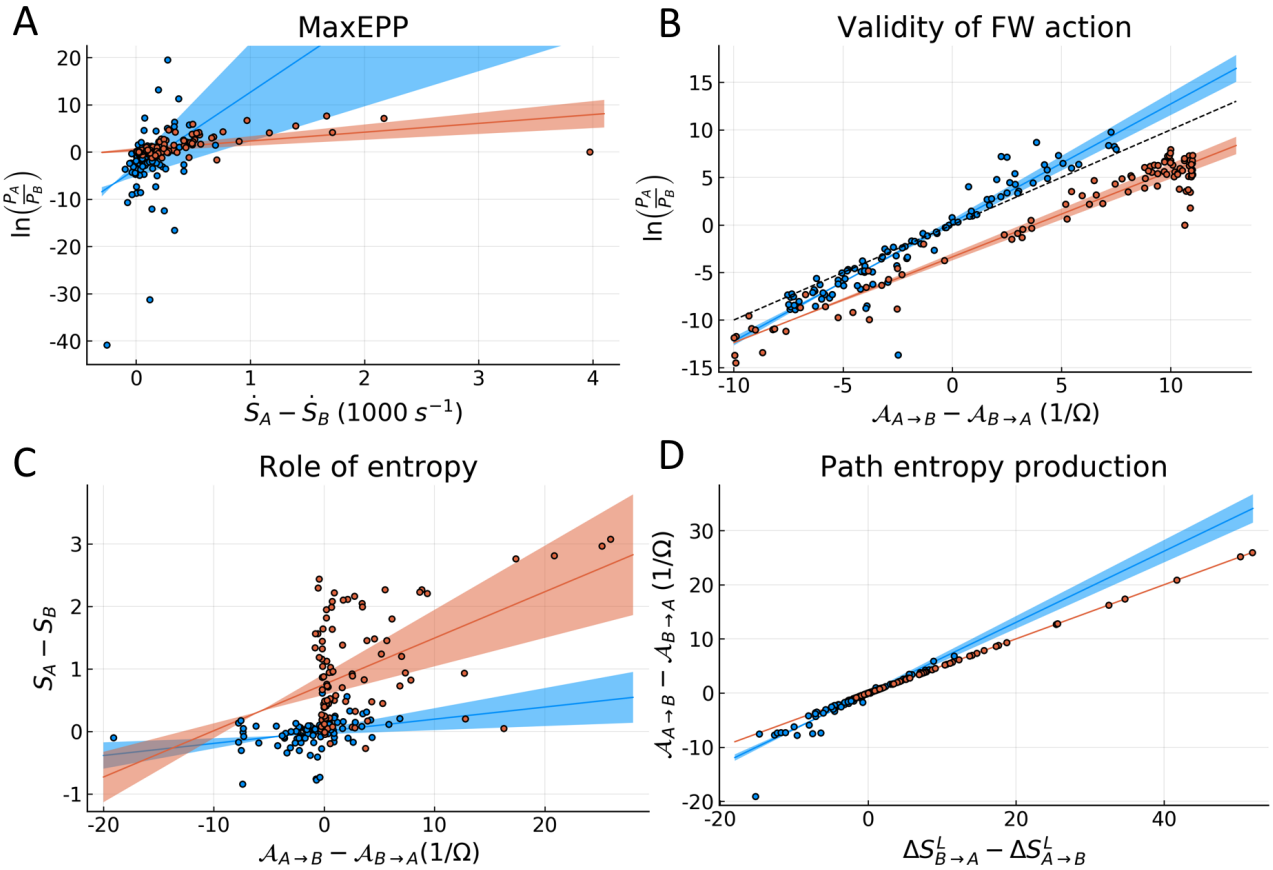


Figure 2.7: **Comparison of states and switching paths.** In all panels, red and blue dots denote specific parameterizations of Schlögl and toggle switch models, respectively. In each plot the lines and shaded regions indicate best fits and 95% confidence intervals for the particular data sets, respectively. (a) Comparison of log ratios of occupation probabilities vs difference in entropy productions at steady states from Gillespie simulations. The Schlögl data has a Pearson correlation of 0.5172 and the toggle switch has 0.3476. (b) Log ratio of occupation probabilities obtained from Gillespie simulation vs the difference in minimum action. The dashed line indicates a perfect correspondence. The toggle switch and Schlögl data have correlations of 0.9515 and 0.9738, respectively. The results shown are coarsely discretized due to the low Ω used to save computational time. The discretization will effect the Schlögl B state disproportionately as it is formed of significantly fewer microstates than A. This represents a potential explanation for the downwards shift of the Schlögl data. (c) Comparison of difference in entropy of steady states vs difference in action. Entropies were found by Gillespie simulation with $\Omega = 1$. Both sets of data show weak correlations of 0.2892 and 0.4622 for the Schlögl and toggle switch models, respectively. (d) Difference in action vs difference in entropy produced along paths. Both models display a strong linear relationship, with correlations of 0.9445 and 1.0000 for the toggle switch and Schlögl models, respectively.

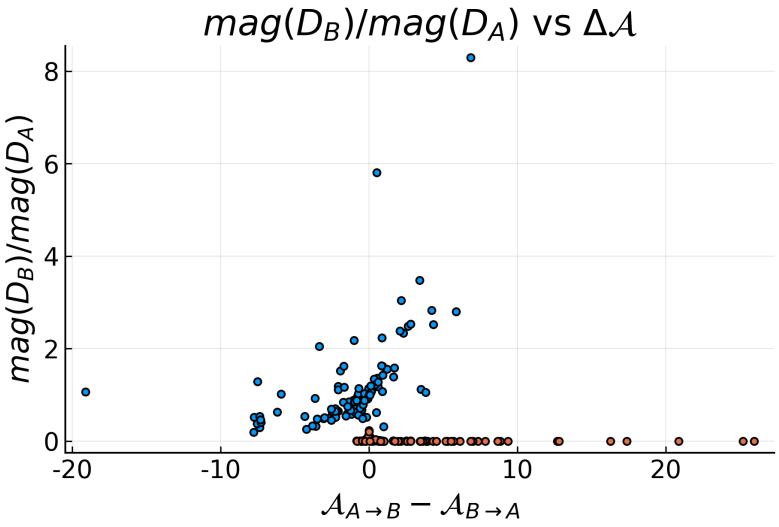


Figure 2.8: **Effect of diffusion on action.** Red and blue points correspond to Schlögl and toggle switch model parameter sets, respectively. The link between the ratio of diffusion matrix magnitudes and difference in action is weak at best. The toggle switch data displays no apparent pattern. While the Schlögl data shows some limited pattern, it is dominated by a few points of extremely large difference in action. Other points of similarly small ratio of diffusion matrix traces do not show such large differences in action.

with the assumption that the multiplicative constants for the forward and backwards processes approximately cancel.

This description pertains to Langevin paths. In order to make a linkage with the fluctuation theorem (Eq. 2.2), a relation to the path probabilities at the level of master equation paths must be obtained. An exact correspondence between the two formalisms would require that the probability P_{Γ^*} of the minimum action path Γ^* (a Langevin path) is equal to the probability W_Γ of the corresponding master equation path (defined in Section 2.3.1). If this assumption holds then the entropy production as calculated from the Langevin expression (Eq. 2.15) should be equivalent to the entropy production calculated using the fluctuation theorem upon master equation paths (Eq. 2.8). Unfortunately, due to our use of a coarse-grained Langevin equation to describe the switching paths the Langevin entropy production is a coarse-grained (apparent) entropy production, and does not match with the master equation entropy production. However, as it appears as a term in our action it is still of interest despite not necessarily corresponding to the entropy production as calculated from the master equation. We can now

express relative switching rates between macrostates as

$$\frac{k_{B\to A}}{k_{A\to B}} \approx \frac{P_{B\to A}}{P_{A\to B}} \approx \exp\left[\Omega(\mathcal{A}_{A\to B} - \mathcal{A}_{B\to A})\right]. \tag{2.34}$$

For one-dimensional models (e.g. the Schlögl model), each minimum action path is the microscopic time-reversal of the other, e.g. $k_{\overline{A\to B}} = k_{B\to A}$. From this relation the previously derived 1D switching relation (Eq. 2.26) can then be recovered.

For the case of more complicated models (e.g. the toggle switch model), the two minimum action paths are no longer expected to be time-reversals of each other. We shall therefore proceed by looking more carefully at the components of the action

$$\Omega\mathcal{A}_{A\to B} = \mathcal{C}_{A\to B} - \frac{1}{2}\Delta S^{L}_{A\to B}, \tag{2.35}$$

where $\mathcal{C}_{A\to B}$ is the conservative action along the path $A\to B$ and $\Delta S^{L}_{A\to B}$ is the Langevin path entropy production (Eq. 2.15) [End17]. From the above decomposition the components of $\mathcal{C}_{A\to B}$ can be written out in a similar form to Eq. 2.15 as

$$\mathcal{C}_{A\to B} = \frac{\Omega}{2}\int_0^{\tau} \left(\dot{x}_i D^{-1}_{ij}\dot{x}_j + f_i D^{-1}_{ij} f_j\right)\, dt, \tag{2.36}$$

where the two terms resemble kinetic (KE) and potential energy (PE) contributions, respectively. By substituting the expanded form of the action into the expression for switching path probability, a reduced form of MaxEPP can be obtained

$$P_{A\to B} = \frac{\exp\left(\frac{1}{2}\Delta S^{L}_{A\to B} - \mathcal{C}_{A\to B}\right)}{Q_{A\to B}}, \tag{2.37}$$

where $P_{A\to B}$ is the probability of the (most probable) switching path along $A\to B$, and $Q_{A\to B}$ is a constant. This equation shows that there is a trade-off between minimization of the conservative action (i.e. fulfilling the equation of motion) and maximization of the path entropy production (i.e. being as dissipative as possible).

2.7.1 Multi-dimensional switching relation

The relation which we shall now derive applies to models with more than one dimension. For one-dimensional models the stronger relation previously derived Eq. 2.26 applies. When this stronger relation holds, the relation we will derive here can be obtained trivially. In order to derive a relation relevant to the multi-dimensional case, we shall consider the restricted space in which feasible minimum action paths can be found. Fig. 2.9 shows how the conservative action of a segment of a minimum action path varies across chemical concentration space. There is a clear region (shown in black) of lower conservative action contribution. Paths outside this region are unlikely to be feasible. Additionally minimal paths are constrained to pass through the saddle point (marked by a white cross). This further constrains the region that paths can feasibly pass through. This means that switching paths in opposite directions should generally not be expected to have large separations between them.

We have previously mentioned the appearance of a potential-energy-like term in our action, defined as

$$\mathrm{PE} = \sum_{ij} f_i D_{ij}^{-1} f_j. \tag{2.38}$$

In order to simplify our terminology from this point onward, we will refer to this as potential energy despite the slightly different form compared to the classical definition of potential energy. If both the force $\mathbf{f}$ and diffusion matrix $\mathbf{D}$ vary slowly across space, then it would be expected that two paths close in space would have approximately the same potential energy along their lengths. The specific model we are considering is the (two-dimensional) concentration-constrained toggle switch. Of the eight components of force (see Eqs. 2.29 & 2.30), two are constant and four are linear. These terms will not vary much between paths that are close together. There are additionally two non-linear terms that take the form $kra/(r + fb^2)$ and $qrb/(r + fa^2)$. In both causes it is the denominator that is causing the non-linearity. This non-linearity will not play a significant effect in two cases. Firstly, if $a, b \ll r/f$ where the r term in the denominator dominates and the force components become effectively linear. Secondly, if $a, b \gg r/f$ where the denominator becomes very large and the force components rapidly approaches zero. Within these regions the potential energies of the paths would be expected

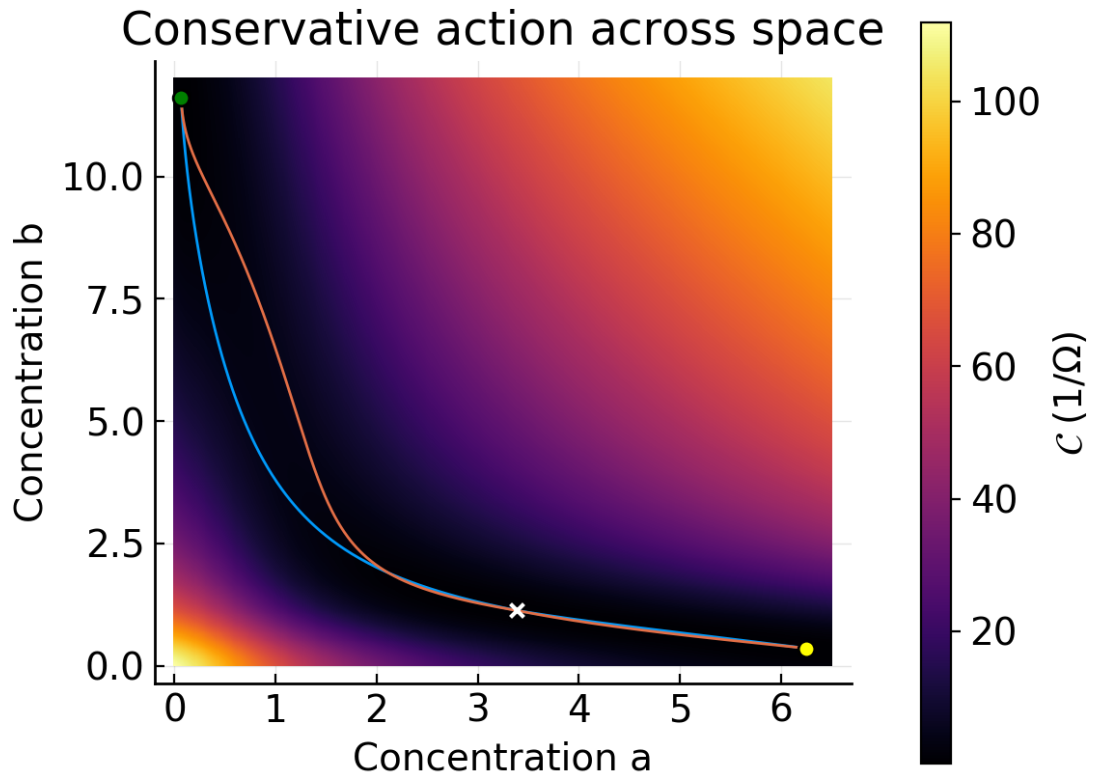


Figure 2.9: **Restriction of feasible paths.** Heatmap showing the value of the conservative action across concentration space. There exists a narrow band where the conservative action contribution is small. The two macrostates A and B are marked by a yellow and a green point, respectively. The saddle point is indicated by a white cross, showing that all paths are constrained to go through here. The minimum action paths $A \to B$ and $B \to A$ are shown as a blue line and a red line, respectively. The parameterization is as follows (given to 6 sf): $k = 58.9655$, $k_- = 2.29765$, $q = 52.9151$, $q_- = 1.31800$, $K = 6.24827$, $K_- = 0.0212992$, $Q = 3.22743$, $Q_- = 0.00844282$, $r = 9003.55$, and $f = 7309.72$.

to be similar.

Previously, when deriving the 1D switching relation (Eq. 2.26) we used the following result (Eq. 2.24)

$$\sum_{ij} \dot{x}_i D_{ij}^{-1} \dot{x}_j = \sum_{ij} f_i D_{ij}^{-1} f_j. \tag{2.39}$$

The RHS of this equation matches our definition of potential energy. It is then natural to make the following definition of kinetic energy

$$\mathrm{KE} = \sum_{ij} \dot{x}_i D_{ij}^{-1} \dot{x}_j. \tag{2.40}$$

Eq. 2.39 can then be written in the form

$$\mathrm{KE} = \mathrm{PE}, \tag{2.41}$$

where KE is the kinetic energy and PE is the potential energy at the same point along a minimum action path. The conservative action can now be expressed as

$$\mathcal{C} = \int_0^\tau \left(\dot{\mathbf{x}} \mathbf{D}^{-1} \dot{\mathbf{x}} + \mathbf{f}\, \mathbf{D}^{-1} \mathbf{f} \right)\, dt = \int_0^\tau (\mathrm{KE} + \mathrm{PE})\; dt = 2 \int_0^\tau \mathrm{PE}\; dt. \tag{2.42}$$

Using the fact that the potential energies along the paths are expected to be approximately equal we can thus conclude that $\mathcal{C}_{A\to B} \approx \mathcal{C}_{B\to A}$. In other words, the conservative component of the action of a switching path and one in the opposite direction are in general approximately equal. This approximation is exact for the 1D Schlögl model as $\overline{A{\to}B} = B{\to}A$, and holds for 90% of the toggle switch parameterizations used (Fig. 2.10).

We can now obtain a relation between switching rates based on path entropy production and conservative actions by substitution of Eq. 2.35 into Eq. 2.34 as

$$\frac{k_{B\to A}}{k_{A\to B}} \approx \frac{P_{B\to A}}{P_{A\to B}} \approx \exp\left(\mathcal{C}_{A\to B} - \frac{1}{2}\Delta S^L_{A\to B} - \mathcal{C}_{B\to A} + \frac{1}{2}\Delta S^L_{B\to A} \right). \tag{2.43}$$

Comparison of Eqs. 2.33 and 2.43, and making use of the approximation $\mathcal{C}_{A\to B} \approx \mathcal{C}_{B\to A}$, leads

to the relation

$$\Omega\left(\mathcal{A}_{A\to B}-\mathcal{A}_{B\to A}\right)\approx\frac{1}{2}\left(\Delta S^{L}_{B\to A}-\Delta S^{L}_{A\to B}\right), \tag{2.44}$$

which describes the empirically observed relationship shown in Fig. 2.7D. This result can now be used to obtain a ratio of switching rates uniquely determined by the difference in path entropy productions

$$\frac{k_{B\to A}}{k_{A\to B}}\approx\frac{P_{B\to A}}{P_{A\to B}}\approx\exp\left[\frac{1}{2}(\Delta S^{L}_{B\to A}-\Delta S^{L}_{A\to B})\right]. \tag{2.45}$$

Note, in general $\Delta S^{L}_{A\to B}\neq-\Delta S^{L}_{B\to A}$ due to the fact that the relation given in Eq. 2.39 only fixes a (diffusion strength corrected) optimal speed. In the multidimensional case there is no strong constraint on the direction of motion. As the path entropy production depends on cross terms of velocity with force, this means that even paths that are close in space can have substantially different entropy productions. A demonstration of this is provided in Fig. 2.10, showing an inequality in panel D.

The relation obtained in Eq. 2.45 can then be extended to link occupation probabilities to difference in entropy produced along the switching paths. This can be done by making the assumption that very little time is spent by the system outside the macrostates A and B. Probability of occupation of state A can then be taken to be $p_A=\langle\tau_A\rangle/(\langle\tau_A\rangle+\langle\tau_B\rangle)$ where $\langle\tau_A\rangle$ is the average time to exit state A. The process of switching between states A and B is then assumed to be an exponentially distributed stochastic process with rate constant $k_{A\to B}$. Thus, $\langle\tau_A\rangle=k_{A\to B}\int_0^\infty t\exp(-k_{A\to B}t)dt=k^{-1}_{A\to B}$, and likewise for the B state. It then follows that $p_A=1/(1+k_{A\to B}/k_{B\to A})$. Taking a ratio of state occupation probabilities and setting this equal to Eq. 2.45 then gives us the result

$$\frac{p_A}{p_B}=\frac{k_{B\to A}}{k_{A\to B}}\approx\exp\left[\frac{1}{2}(\Delta S^{L}_{B\to A}-\Delta S^{L}_{A\to B})\right]. \tag{2.46}$$

Significant divergences from the relations we have derived were generally observed in cases where the saddle point occurred at a low copy number compared to the steady states, due to the substantially faster variation of the force. Further discussion of situations causing divergences

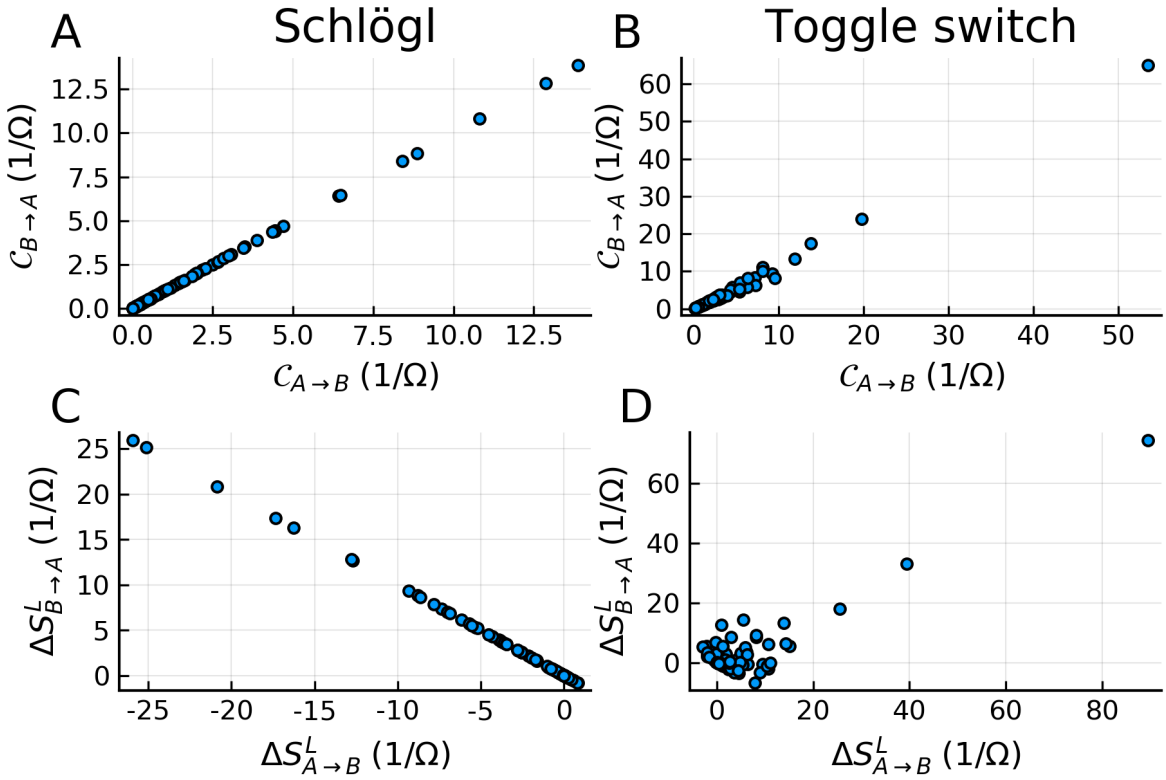


Figure 2.10: **Comparing action components between paths.** (a) Plot of the conservative action for a switching path against the conservative action of the corresponding reversed-switching path, for Schlögl model parameterizations. The two variables can be seen to exactly correspond. (b) Equivalent plot for our parameterizations of the toggle switch model. Though there is no longer an exact correspondence, there is still a clear linear relationship with a slope of approximately one. (c) Plot of the path entropy for a switching path against the path entropy of the corresponding reversed-switching path, for Schlögl model parameterizations. There is an exact negative correspondence between variables as would be expected from our derivation of the !D switching relation (Eq. 2.26). (d) Equivalent plot for our parameterizations of the toggle switch model. There no longer appears to be any relation.

from these relations can be found in Appendix D.2.

2.8 Discussion

Our primary conclusion is that a MaxEPP for switching paths can be obtained within the Langevin approximation (Eq. 2.37), extending the rule that "exergonic reactions occur spontaneously" to switching in multistable systems. In a system with a large number of potential macrostates our relation predicts that for sufficiently large volumes switches that produce more entropy will be favoured. If regions of state space with greater entropy productions also require greater path entropy productions to reach, then this could form a basis for a more extensive maximum entropy production principle. Our secondary conclusion is that there exists a relationship between the difference in action of minimum action paths and difference in entropy produced along these paths (Eq. 2.44), valid for all paths that do not pass through regions of rapidly varying force. We additionally found that steady-state entropies had very little effect on the stability of steady states (see Fig. 2.7C). Finally, we found some evidence to support a broader maximum entropy production principle (see Fig. 2.7A).

There has been significant interest in the thermodynamics of the transition between different steady-state probability distributions when controlled by an external parameter [HS01, BTK13]. This is fundamentally different to our work, which is about the thermodynamics of switching in a bistable system. In our two-state systems, we would naively expect a net zero entropy production through switching as the entropy produced by a switch in one direction would be cancelled by the eventual switch back. Only in cases where the switching path differs from the converse switching path are there net fluxes of probability through the system and thus entropy production. Consistently, for the Schlögl model we find no net entropy production (i.e. $\Delta S^L_{A\to B} = -\Delta S^L_{B\to A}$). In other recent work [Rue16], bounds on the ratio of transition rates between states in a bistable system based on relative entropy ΔH and path entropy production are found. This ratio is determined as $\pi(B \to A)/\pi(A \to B) \geq \exp\left[-\Delta S_{A\to B}\right]$, where $\pi(A \to B)$ is the sum of the rates of switching A to B over all possible switching

channels. With the expectation that a single most probable path will dominate, consideration of the two contrary switching paths will be sufficient. Combining the main result of their paper with our analytic relation (Eq. 2.44) leads to a bound on the path entropy produced as $\Delta S^L_{A\to B} + \Delta S^L_{B\to A} \geq 0$, which becomes an equality in the limit of time-reversed switching paths (e.g. Schlögl). Every parameter set used in Fig. 2.7 was found to satisfy this condition.

Beyond the chemical physics literature, related frameworks to ours have been used in evolutionary science [ML10, dVB11] where the cumulative fitness flux is maximized (like entropy production) subject to the trade-off that speed of allele change and magnitude of selective forces are minimized (like the conservative action). Our results therefore suggest, that states in evolutionary systems that require greater cumulative fitness fluxes to reach should be expected to be more stable. Exploring the application of our theory in ecology, and evolutionary science with systems of multiple stable states could be an interesting way forward [SBB$^+$09].

2.9 Outlook

In the preceding chapter an extremal principle was successfully derived. This principle had an exact form (Eq. 2.37) and an approximate form where (relative) entropy production was the only relevant factor (Eq. 2.45). For both these forms the entropy production along a switching trajectory was the relevant factor, rather than entropy production at steady state. This extremal principle implies that when switches between states in a multi-stable system occur, switches that generate more entropy will be favoured. The fact that an extremal principle could be obtained based upon stochastic thermodynamic theory, suggests that the field has potential to improve and deepen the use of thermodynamic extremal principles in ecology. Either by establishing or undermining the theoretical basis of existing extremal principles, or through the generation of new principles.

That said, this thesis will not continue to investigate extremal principles for two reasons. Firstly, the relevant entropy production in our extremal principle is an apparent entropy production, i.e. one that is associated with a higher level (coarse) master equation. It is not clear how well

an apparent entropy production of this type could be connected to empirical entropy production measurements. This limits the potential applicability of the previously derived principle or any extensions of it. Secondly, the complexity of the methods used to test our principle scales very rapidly with increases in the dimensionality. This makes developing principles for systems including interactions between anything more than four or five species computationally unfeasible, leaving study of realistic ecosystems beyond the scope of our methods. For these reasons this thesis now moves on to the explicit incorporation of non-equilibrium thermodynamics into preexisting ecological models. With the aim of assessing the impact of constraints arising from thermodynamics on broader ecosystem dynamics.

Chapter 3

Development of a microbial community model incorporating non-equilibrium thermodynamic constraints

3.1 Introduction

This chapter aims to move beyond the consideration of coarse extremal principles in order to consider in detail the constraints which non-equilibrium thermodynamics places upon ecosystems. In order to do this, non-equilibrium thermodynamic constraints will be explicitly incorporated into a conventional microbial community model, building upon previous work that has done this for models of simple (few species) microbial communities [HCR96, GS16, LWvB$^+$19, DCWS20]. The major innovation of this chapter is the extension of these methods into the context of complex communities. This will allow the study of not just how thermodynamics effects the growth of single species or interactions between pairs of species, but also how the overall properties of communities are impacted by emergent non-equilibrium thermodynamic constraints.

The work presented in this chapter was previously published in *PLOS Computational Biology*

[CPE21]. A version of this chapter (as it was resubmitted to *PLOS*) is available on bioRxiv (doi: 10.1101/2021.04.19.440392) and arXiv (arXiv:2111.08580). The rest of this chapter consists of the main text of the manuscript as originally submitted to *PLOS Computational Biology* with a number of alterations. Firstly, a number of revisions have been made based on reviewers comments. Secondly, content previously contained in the supplementary material necessary to understand the results of the main text has been integrated in. The more technical content of the supplementary material has been added to the thesis as appendices, which are referred to at the relevant points in this chapter. Thirdly, a number of changes in wording have been made both for reasons of readability and self-consistency. Finally, an Outlook section has been added in order to situate the results of this chapter within the broader aims of this thesis.

3.2 Thermodynamic constraints on microbial ecosystems

The constraints thermodynamics place upon on individual organisms inevitably impact ecosystem dynamics. Thus, attempts to understand ecosystems through thermodynamic principles have been made repeatedly throughout the history of ecology [Lot22a, OP55, Dew03]. These attempts have generally involved the development of coarse, whole-ecosystem level extremal (maximisation or minimisation) principles, such as flux [Lot22a] and power maximisation [OP55]. Most notable of these is the maximum entropy production principle, that ecosystems tend towards states that produce entropy at the maximum achievable rate [Dew03]. This principle has been applied to some degree of success to predicting ecosystem characteristics such as spatial distribution of vegetation [dJFRRI12] and biogeochemical cycling in ponds [VH18]. Though consideration of ecosystem wide entropy production has led to insights in areas such as the conditions for ecosystem stability [Mic05], without detailed consideration of the underlying mechanisms, its explanatory potential remains limited. However, when these mechanisms have indeed been considered, they are implicitly assumed to be dissipating sufficient free-energy that thermodynamic constraints can reasonably be neglected. In contrast, in biophysics, non-equilibrium thermodynamics at the cellular level have been considered in much greater detail [FW20], particularly in the areas of kinetic proofreading [PT20] and sensing accuracy [Tu08, LFMM14].

This opens the possibility of detailed consideration of the impact of thermodynamic constraints on ecosystem dynamics.

Thermodynamic constraints are especially pertinent to microbial community dynamics because microbial growth can be described explicitly in terms of the energetics of carbon substrate processing. Microbes experience a large number of physical constraints on their metabolism (see [AYZP21] for a comprehensive summary). Here, we consider two of the most universally relevant constraints. First, the widely studied constraint of a finite cellular proteome which means that the increased expression of a particular class of metabolic protein must occur at the expense of other metabolic proteins and/or ribosomes [SGM$^+$10, SKMH14, WODS15]. Second, in order to proceed rapidly metabolic reactions must (net) dissipate substantial amounts of free-energy. This both limits the set of possible (catabolic) reactions that microbial cellular metabolic networks can be formed from and leads to a significant reduction in reaction rates close to thermodynamic equilibrium. Both these constraints affect microbial growth rates and ultimately microbial interactions, but the impact of the thermodynamic constraint has rarely been considered previously. Models of the impact of thermodynamic inhibition have been developed to properly capture the growth rate of anaerobic microbial populations [HCR96]. This has been recently extended to study whether thermodynamics leads to distinct ecological strategies [LWvB$^+$19], to explain the coexistence of multiple species on a single substrate [GS16], and to explain empirically observed limitations on the growth of simple methanogenic communities [DCWS20]. However, these studies have considered relatively simple systems of a few interacting species.

The carbon substrates that microbes feed upon are often classified in terms of the difficulty involved in breaking them down, with those easily broken down being termed ‘labile’, and ones that are hard to break down termed ‘recalcitrant’. In energetic terms, a recalcitrant substrate can be thought of as one that requires a significant energy investment by a microbe for a minimal return. This could emerge through multiple mechanisms such as high substrate activation energies, substrates that have to be broken down extracellularly, or reactions that yield small amounts of free-energy. Thus, in a thermodynamic model, recalcitrance can be approximately modelled by considering substrates of low free-energy, potentially shedding light

on the role of substrate lability versus recalcitrance on the dynamics of microbial community assembly.

Here, we develop a mathematical model that explicitly and mechanistically incorporates thermodynamic constraints. We use this model to study the emergent impact of thermodynamic constraints on microbial community diversity by simulating their assembly. Because natural communities may assemble on a diversity of substrates, we focus on the relationship between microbial species efficiency, substrate free-energy, and emergent species diversity. We find free-energy availability to be the single most important driver of community diversity, as greater energy availability allows for a faster rate of niche generation. We also show that consideration of thermodynamic constraints leads to new insights into the fundamental mechanistic and energetic basis of species interactions.

3.3 The Model

Our thermodynamic microbial community model is a thermodynamically-explicit extension of the MacArthur consumer-resource model [Mac70]. Its key features are illustrated in Fig. 3.1 with a detailed derivation given in Appendix E. In this model, catabolic reactions are explicitly modelled, but anabolism is modelled abstractly as protein translation. As the predominant energy cost for microbial cells is protein translation [LM15], the implicit assumption that this is the only energetic cost is a reasonable approximation. Each metabolite (β) is represented by its concentration C_β and each consumer (i) is represented by three variables: its population abundance N_i, ribosome fraction $\phi_{R,i}$, and internal energy (ATP) concentration a_i. The

dynamics of these variables are given by

$$\frac{dC_\beta}{dt} = \kappa\delta_{\beta,1} - \rho C_\beta + \sum_{i=1}^{B} \left(p_{i,\beta}(\mathbf{C}) - c_{i,\beta}(\mathbf{C})\right) N_i \tag{3.1}$$

$$\frac{dN_i}{dt} = (\lambda_i(a_i, \phi_{R,i}) - d_i)N_i \tag{3.2}$$

$$\frac{d\phi_{R,i}}{dt} = \frac{1}{\tau_g}\left(\phi_R^*(a_i) - \phi_{R,i}\right) \tag{3.3}$$

$$\frac{da_i}{dt} = J_i - \chi m \lambda_i - a_i\lambda_i, \tag{3.4}$$

where κ is the substrate supply rate, $\delta_{\beta,\xi}$ is the Kronecker delta, ρ is the metabolite dilution rate, B is the number of species in the community, $p_{i,\beta}$ is the (per cell) rate that the i^{th} species produces metabolite β, and $c_{i,\beta}$ is the (per cell) rate at which it consumes metabolite β. Further, λ_i is the i^{th} species' growth rate, d_i its rate of biomass loss, τ_g is the characteristic time scale for growth, ϕ_R^* is the ideal ribosome fraction, J_i is the ATP production rate for the i^{th} species, χ is ATP use per elongation step, and m is the total mass of the cell (in units of amino acids). The second term in Eq. 3.4 corresponds to the energy use due to protein translation and the third term corresponds to the dilution of energy due to cell growth.

Throughout this chapter, the biomass loss rate for each species d_i is equal to the metabolite dilution rate ρ (see Tab. A.5), which means that our model is effectively a chemostat model. However, we retained biomass loss rate as a separate parameter in order to allow our model to be adapted for ecosystems where the biomass loss rate substantively differs from the washout (dilution) rate. This predominately involves including additional sources of biomass loss (e.g. respiration or mortality), but can also include systems with reduced biomass loss (e.g. due to cellular adhesion).

3.3.1 Proteome partitioning

Due to finite cell size, a higher expression of a given type of protein must always come at the expense of lowered expression of other types of proteins. This means that a cell's ability to obtain energy for growth is constrained not just by thermodynamics, but also by the amount of

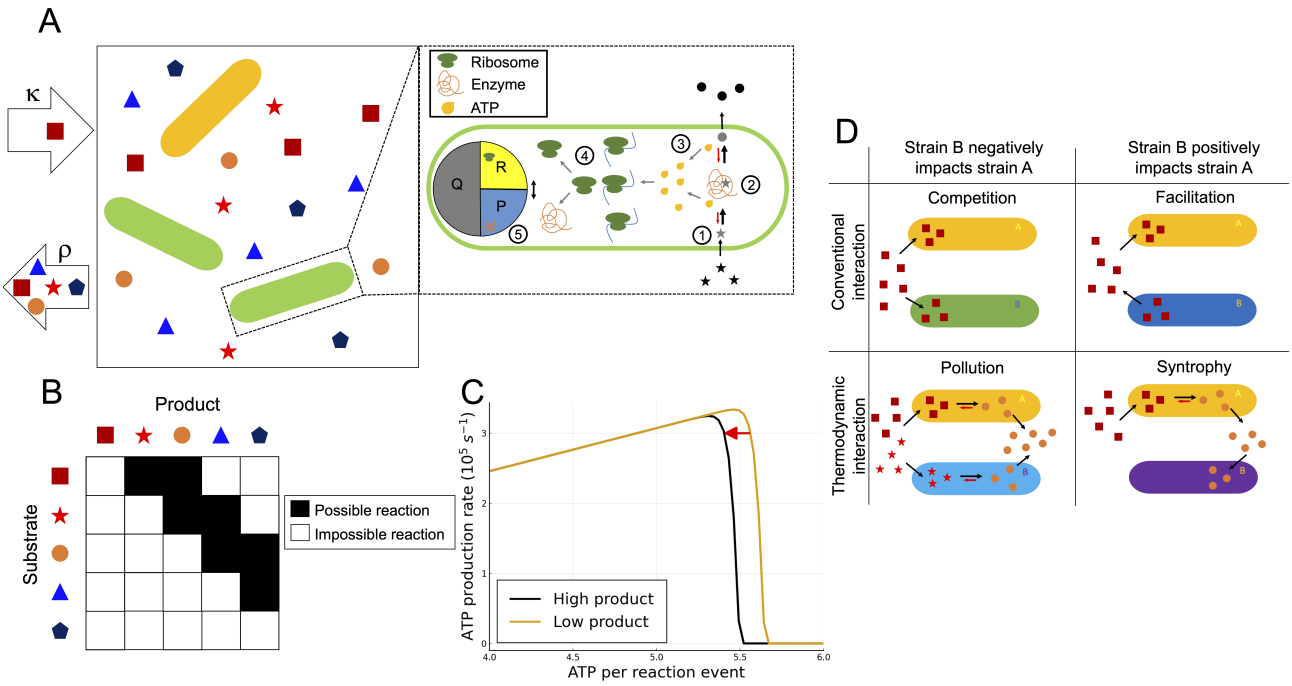


Figure 3.1: **Overview of the thermodynamic microbial community model.** (a) A single substrate is supplied with rate κ, and diversifies into secondary substrates (metabolites) through uptake, metabolism and leakage by the microbial populations. All metabolites are diluted out of the system with constant dilution factor ρ. The magnification shows a schematic of the cellular sub-model, comprising five key processes: (1) uptake of metabolites, (2) their breakdown into other metabolites (Eqs. 3.12–3.13), (3) generation of ATP through this process (Eq. 3.8), and (4) the use of this ATP to drive protein synthesis (Eqs. 3.5–3.6). (5) The cell proteome is partitioned into a constant housekeeping protein compartment (Q), a ribosome compartment (R), and a metabolic protein compartment (P). (b) An example matrix of possible metabolic reactions. Metabolites can only react to form new metabolites that are one or two positions lower on the free-energy hierarchy. (c) The thermodynamic trade-off that emerges from our model; ATP production rate increases linearly with an increase in ATP per reaction event (η), up to the point where thermodynamic inhibition becomes significant. The position of optimal η value shifts as environmental metabolite concentrations change. (d) The four possible species interaction types in our model. Competition and facilitation are present here like in all consumer-resource models that allow for the production of resources. In addition, there are two thermodynamic interaction types: "pollution" where one species produces a metabolite that causes thermodynamic inhibition of the other species, and "syntrophy" where a species consumes a metabolite that is thermodynamically inhibiting another species (and thus benefiting it).

protein it allocates to metabolism. In microbial models, this constraint is typically introduced in the form of a fixed "enzyme budget" that has to be divided between substrates [PTW17]. However, the fact that the size of this "budget" varies with the amount of protein allocated to ribosomes has only recently begun to be considered [PMSMG21], despite the fact that this constraint is known to significantly impact microbial growth [SGM$^+$10, SKMH14]. Therefore, we develop a minimal cell model (Fig. 3.1A), which builds upon an existing mechanistic model of proteome constraints, namely the Weiße–Swain model [WODS15]. Our principal extensions to the model are: reformulating it such that it can be applied to multi-species communities rather than merely homogeneous cell populations, and switching from the assumption that growth is limited by a single (rarest) amino-acid to the assumption that growth is (free-) energy limited.

This sub-model comprises of five key processes: (1) The cell exchanges metabolites with its environment. For simplicity, we assume that the intracellular metabolite concentrations are equal to the environmental concentrations. Even if real cells can use a variety of mechanisms to maintain intracellular metabolite concentrations that make reactions more thermodynamically favourable, this assumption imposes the universally seen limit on efficiency that arises from the 2$^{\text{nd}}$ law of thermodynamics, i.e. maintaining a particular metabolite concentration necessarily consumes more free-energy than it contributes to the reaction free-energy. (2) Within the cell, substrates are broken down into lower free-energy metabolites. This process is thermodynamically-reversible in contrast to the conventionally used irreversible Michaelis–Menten kinetics (for more details see Appendix E). (3) This breakdown of substrates into lower free-energy metabolites allows free-energy to be transduced via the production of ATP. (4) This ATP is subsequently used to fuel protein synthesis [SGM$^+$10, SKMH14, WODS15]. Protein synthesis rate is thus dependent on ATP concentration, but sufficient free-energy is dissipated for it to be considered an irreversible process. (5) The proteome is partitioned into three compartments: a ribosome compartment R, a metabolic protein compartment P, and a compartment for all other proteins required by the cell Q (termed "housekeeping"). As the housekeeping compartment is assumed to be constant, a direct trade-off between the fraction of the proteome dedicated to ribosomes and the fraction dedicated to metabolic proteins arises.

The growth rate of any given microbial species is determined by the total rate at which ribosomes synthesize proteins (process 4 in Fig. 3.1A). This can be modelled by assigning each species (i) two internal variables, the internal energy (ATP) concentration a_i and the ribosome fraction $\phi_{R,i}$. The growth rate depends on both quantities and can be expressed as

$$\lambda_i(a_i, \phi_{R,i}) = \frac{\gamma(a_i) f_b \phi_{R,i}}{n_R}, \tag{3.5}$$

where $\gamma(a_i)$ is the effective translation elongation rate, f_b is the average fraction of ribosomes bound and translating, and n_R is the number of amino acid per ribosome. As we assume that protein synthesis is an irreversible process, the effective translation elongation rate is assumed to be saturating with respect to the energy concentration, taking the form

$$\gamma(a_i) = \frac{\gamma_m a_i}{\gamma_{\frac{1}{2}} + a_i}, \tag{3.6}$$

where γ_m is the maximum elongation rate, and $\gamma_{\frac{1}{2}}$ is its half-maximum constant. With Eqs. 3.5–3.6, we can now define the population dynamics in Eq. 3.2 in terms of the cells' internal state. We therefore now consider the dynamics of a cell's internal state, starting with the ribosome fraction dynamics. We assume that for a specific internal energy concentration (a_i) each cell aims to reach a particular ribosome fraction $\phi_R^*(a_i)$ (the "ideal" ribosome fraction). Similar to Eq. 3.6 it is assumed to be a saturating function of energy concentration,

$$\phi_R^* = \frac{a_i}{\Omega_{\frac{1}{2}} + a_i}(1 - \phi_Q), \tag{3.7}$$

where $\Omega_{\frac{1}{2}}$ is a half-maximum constant and ϕ_Q is the housekeeping protein fraction. By also noting that the characteristic time scale for growth is given by $\tau_g = \log_2(100)/\lambda_i$ (see Appendix E), ribosome dynamics in Eq. 3.3 can now be solved for a given value of the internal energy concentration a_i.

3.3.2 ATP production

To determine the dynamics of the cellular internal energy concentration the rate at which each species produces ATP (process 3 in Fig. 3.1A) must be determined. Fig. 3.1B shows the general pattern of possible reactions in our model for an example case with a small number of metabolites ($M = 5$). Due to the explicit thermodynamic reversibility in our model, only reactions descending in free-energy are allowed. Further to this, the enzyme scheme we used is intended to capture simple enzymatic reactions rather than complex reactions composed of a succession (e.g. glucose respiration). Hence, in our model direct links are only allowed between metabolites with small separations (two positions or less) on the metabolite hierarchy. This network of reactions is fully connected, i.e. every metabolite can be reached from any metabolite higher in the metabolite hierarchy. Each species (i) is assigned a random subset (O_i) of this set of possible reactions. As this subset is small, the reaction networks of individual strains are not complete, though for complex communities the community-level reaction network would tend towards being fully connected. A set amount of ATP ($\eta_{\alpha,i}$) is generated for each species for each reaction α. Though the amount of ATP generated from a specific biochemical reaction is generally invariant, microbial species can still increase the amount of free-energy transduced from a reaction by increasing the number of ions pumped across membranes [LWvB$^+$19].

Based on all this, the ATP production rate can be found by summing the product of this and the reaction rate as

$$J_i = \sum_{\alpha \in O_i} \eta_{\alpha,i} q_{\alpha,i}(E_{\alpha,i}, S_\alpha, W_\alpha), \tag{3.8}$$

where $q_{\alpha,i}$ is the i^{th} species' (mass specific) reaction rate for reaction α, $E_{\alpha,i}$ is the copy number of enzymes for reaction α that the species possesses, S_α is the concentration of reaction α's substrate, and W_α is the concentration of of reaction α's waste product. The enzyme copy number depends on metabolic protein fraction, and so can be expressed in terms of the ribosome fraction as

$$E_{\alpha,i} = \frac{m\nu_{\alpha,i}\phi_{P,i}}{n_P} = \frac{m\nu_{\alpha,i}(1 - \phi_{R,i} - \phi_Q)}{n_P}, \tag{3.9}$$

where $\nu_{\alpha,i}$ is the i^{th} species' proportional expression level for reaction α, and satisfies $\sum_{\alpha \in O_i} \nu_{\alpha,i} =$

1, $\phi_{P,i}$ its metabolic protein fraction, and n_P is the mass (average number of amino acids) of a metabolic protein [WODS15]. The reaction rate in Eq. 3.8 is derived by assuming a reversible kinetic scheme (derivation in Appendix E) and can be expressed as

$$q_{\alpha,i}(E_{\alpha,i}, S_\alpha, W_\alpha) = \frac{k_{\alpha,i} E_{\alpha,i} S_\alpha (1 - \theta_i(S_\alpha, W_\alpha))}{K_{S_\alpha,i} + S_\alpha(1 + r_{\alpha,i}\theta_i(S_\alpha, W_\alpha))}, \tag{3.10}$$

where $k_{\alpha,i}$ is the maximum forward rate for reaction α for the i^{th} species, $K_{S_\alpha,i}$ is its substrate half-saturation constant for reaction α, $r_{\alpha,i}$ is its reversibility factor for reaction α, and $\theta_i(S_\alpha, W_\alpha)$ is a thermodynamic factor given by

$$\theta_i(S_\alpha, W_\alpha) = \frac{Q(S_\alpha, W_\alpha)}{\mathcal{K}_{\alpha,i}}, \tag{3.11}$$

where $Q(S_\alpha, W_\alpha)$ is the reaction quotient, which in the single-reactant/single-product case we consider, is defined as $Q(S_\alpha, W_\alpha) = W_\alpha/S_\alpha$, and $\mathcal{K}_{\alpha,i}$ is the i^{th} species's equilibrium constant for reaction α. The thermodynamic factor $\theta_i(S_\alpha, W_\alpha)$ quantifies how far from equilibrium the reaction is, taking a value of 1 at equilibrium and tending towards 0 far from equilibrium. It is worth noting that for high $\mathcal{K}_{\alpha,i}$ values Eq. 3.10 will not significantly differ from the Michaelis–Menten case. With the rate of energy acquisition (J_i) now sufficiently defined, the dynamics of the internal energy concentration (a_i) can be obtained by Eq. 3.4.

Finally, we wish to consider the impact the species have on the environmental metabolite concentrations through the production and consumption of metabolites (process 2 in Fig. 3.1A). From the expression for the reaction rate, the consumption and production rates for metabolite β that appear in Eq. 3.1 can be defined as

$$p_{i,\beta}(\mathbf{C}) = \sum_{\alpha \in O_i} \delta_{C_\beta, W_\alpha} q_{\alpha,i}(E_{\alpha,i}, S_\alpha, W_\alpha), \tag{3.12}$$

and

$$c_{i,\beta}(\mathbf{C}) = \sum_{\alpha \in O_i} \delta_{C_\beta, S_\alpha} q_{\alpha,i}(E_{\alpha,i}, S_\alpha, W_\alpha). \tag{3.13}$$

In the above, $\delta_{C_\beta,W_\alpha}$ and $\delta_{C_\beta,S_\alpha}$ are Kronecker deltas that are zero unless metabolite β is the

waste product or substrate of reaction α, respectively.

3.3.3 Free-energy dissipation and thermodynamic inhibition

Due to the thermodynamic reversibility in our model, the amount of free-energy dissipated affects reaction rates. We express the amount of free-energy obtained from a given reaction event as the parameter η (units of number of ATP molecules). Cells can also transduce free-energy by pumping ions across membranes. Hence, we assume the minimum value this parameter can take is $\eta = 1/3$, which is approximately equivalent to the amount of free-energy transduced by pumping one ion across a membrane. The maximum possible η value corresponds to all of the free-energy being transduced to ATP and none of it being dissipated. However, in this case the overall reaction will be at equilibrium and there will be no net production of ATP. This impact of free-energy dissipation on the dynamics occurs through the equilibrium constant which is specified as

$$\mathcal{K}_{\alpha,i} = \exp\left(\frac{-\Delta_\alpha G^0 - \eta_{\alpha,i}\Delta G_{\mathrm{ATP}}}{RT}\right), \tag{3.14}$$

where T is temperature, R is the gas constant, ΔG_{ATP} is the Gibbs free-energy per mole of ATP, and $\Delta_\alpha G^0$ is the standard Gibbs free-energy change when one mole of reaction α occurs. As η changes, the reaction rate (Eq. 3.10) changes due to the change in this equilibrium constant. An example of the impact this has on the ATP production rate is visualised in Fig. 3.1C. For low η values, the equilibrium constant is very large so there is a negligible thermodynamic impact on the dynamics, and thus the ATP production rate initially scales linearly with η. However, the exponential form of the equilibrium constant means that there is only a narrow region with anything other than negligible or complete thermodynamic inhibition. This means that ATP production peaks and then rapidly declines to zero as η increases. The narrowness of this rate-yield trade-off means that there exists a clear optimal strategy with alternative thermodynamic strategies being sub-optimal. However, as the position of the peak of this trade-off is determined by the waste product concentration, environmental conditions will determine the optimal strategy.

The inclusion of thermodynamic reversibility in our model has the additional benefit of allowing the entropy production rate (free-energy dissipation rate) to be directly calculated. To do this the free-energy dissipated per reaction event of reaction α must be found for each species (i). This can be expressed as

$$D_{i,\alpha} = \Delta_\alpha G^0 + RT \ln\left(Q(S_\alpha, W_\alpha)\right) + \eta_{\alpha,i}\Delta G_{\mathrm{ATP}}. \tag{3.15}$$

From this, whole community entropy production rate is then found to be

$$\frac{dG_d}{dt} = \sum_{i=1}^{B}\left(\sum_{\alpha\in O_i}\left[\frac{D_{i,\alpha}q_{\alpha,i}(E_{\alpha,i}, S_\alpha, W_\alpha)}{T}\right]N_i\right). \tag{3.16}$$

This entropy production can be used as a state variable allowing the complex dynamics involved in microbial community assembly to be tracked in a simpler but still physically meaningful manner. Without the overall summation the entropy production of individual species can also be found.

We also wish to understand how the overall (thermodynamic) efficiency of the community changes through the assembly process. In order to do this, we need to calculate the efficiency of each reaction for each species. The free energy change of a reaction is not fixed as it depends on both the substrate and waste-product concentrations (see Eq. 3.15). We measure the efficiency against the maximum possible Gibbs free energy transduction, which is determined using the minimum equilibrium product to substrate ratio allowed when choosing η values ($s_r = 10^{-2}$, detailed below in the Simulations section). Species i's efficiency for reaction α is thus given by

$$\zeta_{\alpha,i} = \frac{-\eta_{\alpha,i}\Delta G_{\mathrm{ATP}}}{\Delta_\alpha G^0 + RT\ln\left(s_r\right)}. \tag{3.17}$$

It is worth noting that the free-energy per mole of ATP varies substantially with temperature. However, throughout this work we consider a fixed temperature and so assuming the free-energy of ATP hydrolysis to be constant is a reasonable assumption. For each simulation the average

value of reaction efficiency across all reactions for all species can then be calculated by

$$\bar{\zeta} = \sum_{i=1}^{B} N_i \left(\sum_{\alpha \in O_i} \nu_{\alpha,i} \zeta_{\alpha,i} \right). \tag{3.18}$$

3.3.4 Emergence of novel interaction types

The types of species interactions that emerge from this model and the mechanisms that lead to them are illustrated in Fig. 3.1D. Similarly to other microbial consumer-resource models [GLB$^+$18, TPMW17, MCG$^+$19] species in our model can interact by competing for the same substrate (competition) or by one species producing a metabolite that the other uses as a substrate (facilitation). However, due to the possibility of thermodynamic inhibition interactions via waste products (rather than substrates) are now possible. Species can additionally interact by one species producing a product that thermodynamically inhibits the other species (pollution), or by one species consuming a product that inhibits the other (syntrophy). Example plots of the form syntrophy and pollution interactions take (in a simple system) are shown in Figs. 3.2 & 3.3, respectively. It can be seen that the outcome of competition between two species can be altered due to a syntrophy interaction with a third species (see Fig 3.2). Additionally, the presence of a pollution interaction (with another species) can render a species non-viable that otherwise would be able to survive at steady state (see Fig. 3.3).

In more complex systems these interactions are determined by perturbing each metabolite in turn at steady state, and finding the response of each species. Then the net impact of each species on the concentration of each metabolite is found, which can be combined with the perturbation response to find interactions strength between each species for each metabolite. This method is reminiscent of the calculation of "susceptibilities" in the cavity method [ABM18, CMM20]. An important difference is that our responses are calculated numerically at steady state, as we were unable to obtain the analytic results necessary to find the cavity solution to the dynamics. The interaction types are then classified using the sign of the perturbation response and whether the species makes a net positive or negative impact on the concentration of the perturbed metabolite (for the full process see Appendix F).

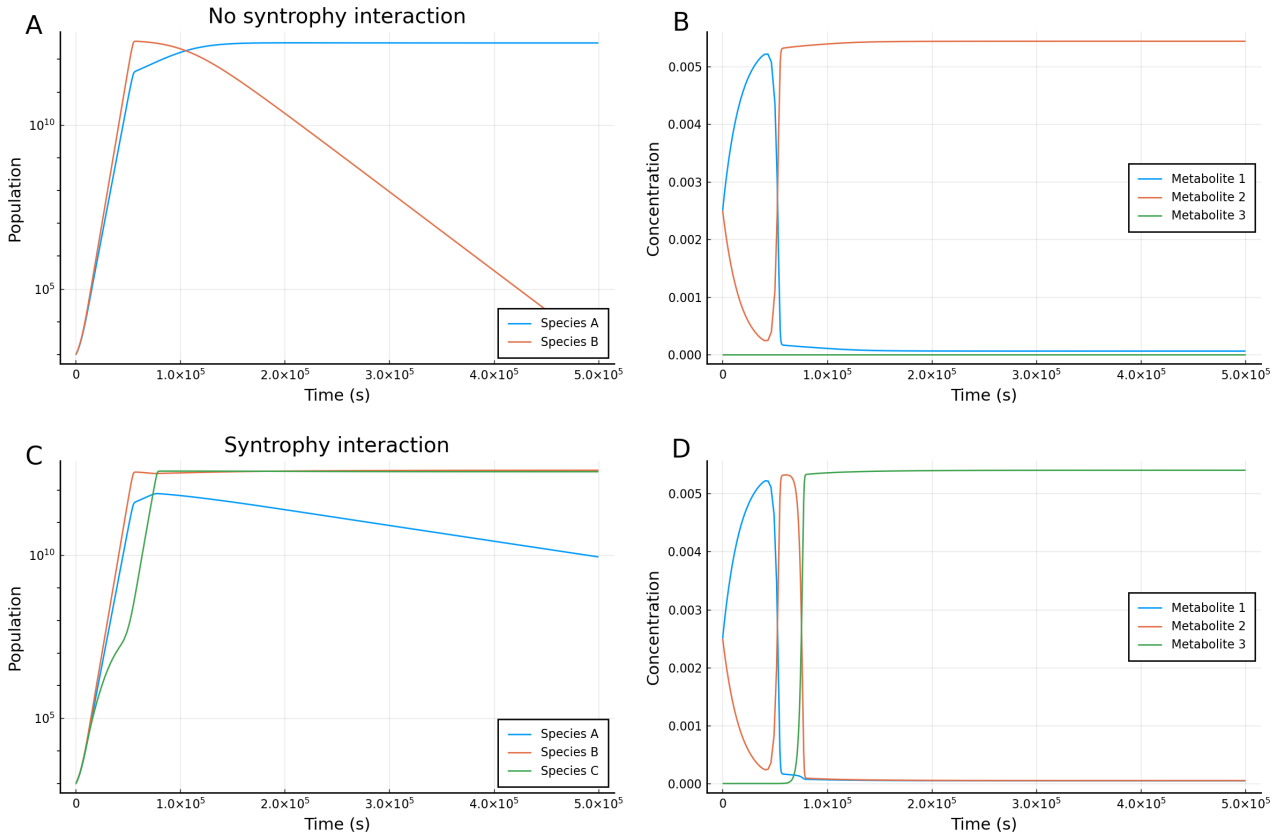


Figure 3.2: **Syntrophy leads to favouring more efficient species.** Our system here consists of three metabolites (only the first of which is supplied) and three species. The species A and B both break down metabolite 1 to produce metabolite 2, and species C breaks down metabolite 2 to produce metabolite 3. Species B generates more ATP per mole of reaction than A does. (a) When species A and species B are grown together species A initially grows faster due to its greater ATP yield. However, as steady state is approached species A dies off to be replaced by species B. (b) Species A dying off occurs due to the build of metabolite 2, which inhibits species A more due to it being closer to thermodynamic equilibrium. (c) When species C is included, species A now survives to steady state (along with species C) instead of species B. (d) The concentration of metabolite 2 is reduced due to consumption by species C, reducing the thermodynamic inhibition of species A. We term this a syntrophy interaction.

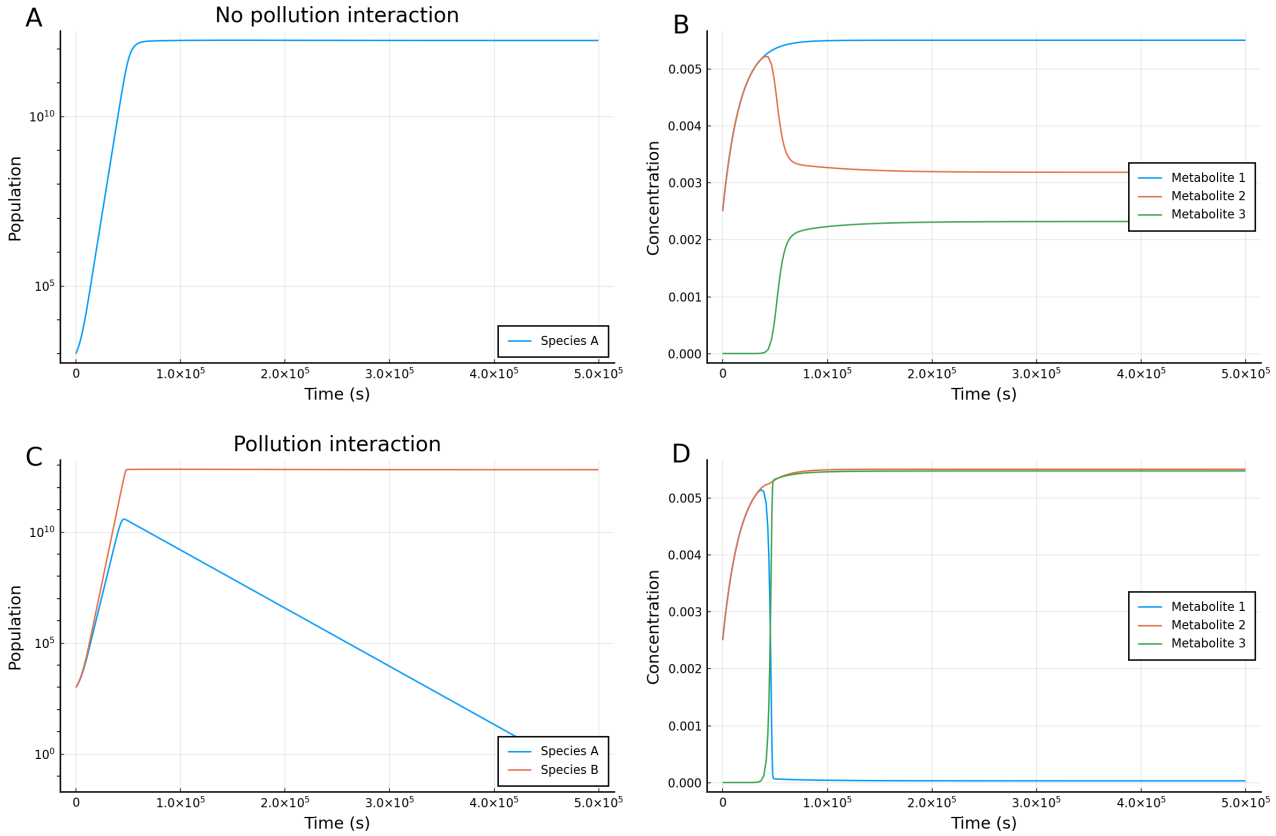


Figure 3.3: **Pollution can render species non-viable.** Our system here consists of three metabolites (the first two of which are supplied) and two species. Species A breaks down metabolite 2 to produce metabolite 3, and species B breaks down metabolite 1 to produce metabolite 3. (a) When species A is grown on its own it reaches a steady state population. (b) As species A only breaks down metabolite 2, both metabolite 1 and 3 accumulate. (c) When species B is added to the system species A is driven to extinction. (d) In this case, greater accumulation of metabolite 3 occurs as species B breaks down metabolite 1 into it. Species A therefore experiences a greater level of thermodynamic inhibition. As species A and B do not share a substrate this competitive exclusion occurs purely via waste products, we therefore term this a pollution interaction.

3.4 Simulations

To simulate community assembly we numerically integrated the system of $M + 3B$ ordinary differential equations (Eqs. 3.1–3.4). We generated the (M metabolite) reaction networks following the pattern shown in Fig. 3.1B, with the free-energy spacing of each step downwards in the metabolite hierarchy being equal. For each community, we then generated a random set of B species. Each species was assigned a number of reactions (N_O) drawn randomly from a uniform distribution. We then generated the reaction set (O) for that species by drawing the N_O reactions at random from the full reaction network. We assigned random kinetic parameters (k, K_S and r) to each reaction of each species, drawn from log-normal distributions as they must be strictly positive. Subsequently, we set the relative reaction expression levels (ν) by assigning each reaction a uniformly distributed random number normalised to the sum of all reactions for the species in question. The next step was to assign the η values to each reaction for a species. When η values are sufficiently low, the dynamics remain far from equilibrium and the enzyme kinetics are effectively irreversible Michaelis–Menten. Thus, we can generate communities with irreversible Michaelis–Menten enzyme kinetics simply by restricting the maximum η value. Henceforth, for the sake of brevity we refer to these communities as having Michaelis–Menten enzyme kinetics. We randomly chose η values from a uniform range between the minimum value ($\eta = 1/3$) and a maximum value which varies based on whether we were considering Michaelis–Menten or reversible kinetics. For Michaelis–Menten kinetics, the maximum possible value of η was chosen such that the reaction will reach equilibrium at a product to substrate ratio of 1×10^5. If instead we considered reversible kinetics, the maximum possible value of η was chosen such that the reaction will reach equilibrium at a product to substrate ratio of 1×10^{-2}. This generation of random initial communities is the only source of stochasticity in our simulations, the model itself is entirely deterministic. After we had generated the random communities, each replicated assembly simulation was initiated with equal abundance across species and the system numerically integrated till the dynamics reached steady state (1×10^8 seconds). A table of parameters and parameter ranges we used for this model can be found in Appendix A. Further, our choice of parametrisation and the

assumptions underlying our model are both justified in Appendix G.

3.5 Results

The dynamical behaviour of our model is demonstrated in Fig. 3.4. The population dynamics (Eq. 3.2) of the species that survive to steady state, along with a small number of species that go extinct, are shown in Fig. 3.4A. The corresponding ribosome fraction dynamics (Eq. 3.3) are shown in Fig. 3.4B. The fractions increase during favourable conditions, leading to an increased growth rate before decreasing to a steady state value (for surviving species). In general, this steady-state growth rate could be matched for a wide range of different ribosome fractions with higher energy concentrations compensating for lower ribosome fractions. However, as the half-maximum constant for the "ideal" ribosome fraction $\Omega_{\frac{1}{2}}$ is fixed across all species in our simulation, each value for the energy concentration a is associated (at steady state) with a specific value for the ribosome fraction ϕ_R (see Eq. 3.7). Hence, all species are observed to converge on the same ribosome fraction at steady state. In Fig. 3.4C the metabolite concentration dynamics (Eq. 3.1) for this community are shown. Initially, the single supplied metabolite accumulates, but as the population of species using this metabolite increases its concentration decreases and the concentration of metabolites one or two steps down the hierarchy increases. This process repeats leading to a sequential diversification of substrates, which leads to clear shifts seen in the ribosome fraction and population dynamics. In Fig. 3.4D the rate of community entropy production (Eq. 3.16) is plotted. The entropy production rises as the substrate diversifies and the total population increases, and shows clear peaks at the time points where accumulated substrates are rapidly depleted. The number of entropy production spikes shows a strong correlation with the final number of substrates diversified (Fig. 3.4D inset).

3.5.1 Final community states depend on free-energy availability

We now assign every microbe to a functional group based on the substrate of its most expressed reaction. The relative abundance of these functional groups with time is shown for a

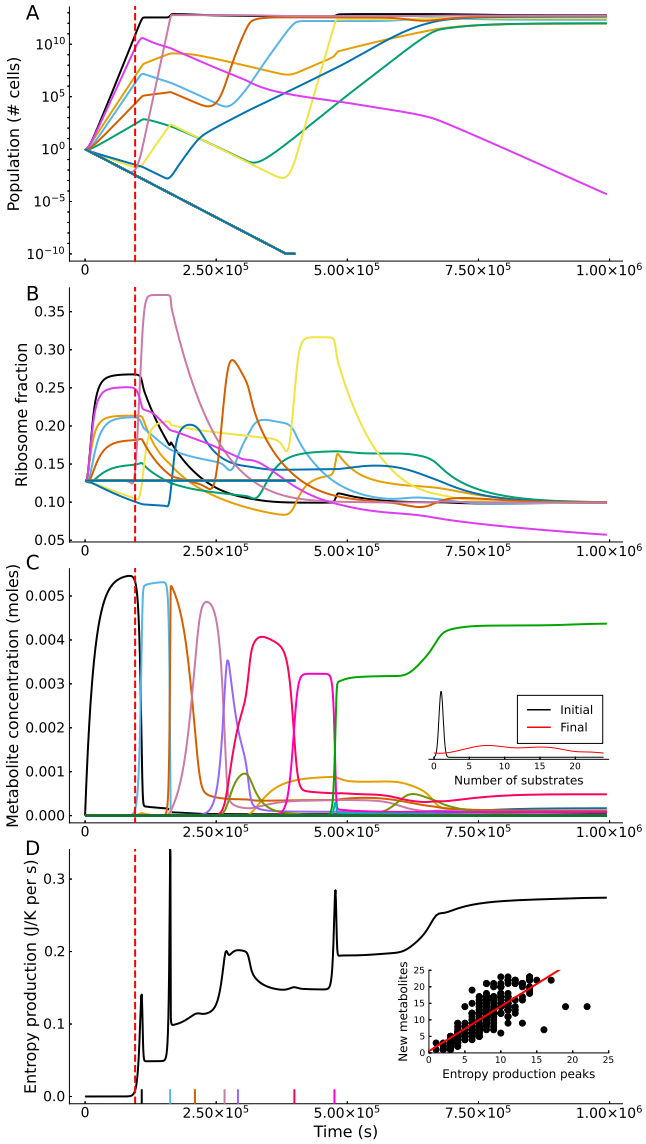


Figure 3.4: **Key dynamical behaviours of the model.** The vertical dashed lines in all plots mark the time point where the third metabolite is available in sufficient quantity (1×10^{-4} moles) to support growth. This specific simulation is started with an initial community of 250 species, each of which is assigned between 1 and 5 reactions from a 25 metabolite (47 reaction) network. The parameter set used here was one of the 250 parameter sets generated using a reversible kinetic scheme and a total free-energy change of $5.0 \times 10^6\,\mathrm{J\,mol^{-1}}$. (a) Microbial population dynamics. (b) Ribosome fraction dynamics. Note that all species that survive to steady state settle to the same ribosome fraction that balances the constant biomass loss rate. (c) Metabolite dynamics. In contrast to (a), these are shown on a linear scale to show the changes in key metabolite concentrations. The inset shows densities of the initial and final number of metabolites across 250 simulations. (d) Entropy production rate of the community with time. The time points where accumulated metabolites drop below an exhaustion threshold (2×10^{-3} moles) are marked on the x-axis (colour-coding corresponds to plot (c)). In the inset the number of entropy production spikes is plotted against the number of new substrates generated (over 250 simulations) with a correlation of 0.769. The best fit line (red) shows a slope of 1.36 and a y intercept of 0.523.

representative parameter set in Fig. 3.5A. The functional diversity initially collapses and then slowly relaxes towards steady state. Using the inverse Simpson index, which corresponds to the effective number of types (here functional groups) in an community we find that functional diversity collapse is common across our simulations. The final community states are now compared across four different regimes in Fig. 3.5B. The regimes compared are low free-energy (recalcitrant) substrates and high free-energy (labile) substrates, and whether the enzyme kinetics are Michaelis–Menten or reversible. For all regimes, species are assigned between 1 and 5 reactions from a 25 metabolite (47 reaction) network, and simulations are started with 250 species. The first property compared is the number of surviving functional groups. Systems supplied with a substrate of higher free-energy see a greater number of surviving functional groups. Using a reversible kinetic scheme (rather than Michaelis–Menten) only increases the number of surviving functional groups in the low free-energy case. A nearly identical pattern is observed for the number of surviving species, suggesting that the decline in functional diversity is predominantly driven by a decline in species diversity. To test whether high free-energy substrates can support more species we then compared the ratio between the number of surviving species and the number of substrates diversified to. We found that the number of survivors per substrate is lower for high free-energy substrates, with no significant difference between kinetic schemes. For the community entropy production rate, higher substrate free-energy corresponds to higher entropy production rates. There is again only a significant difference between kinetic schemes when substrate free-energy is low.

3.5.2 Free-energy availability increases rate of niche generation

The naive explanation for the diversity results in Fig. 3.5B is that higher free-energy availability increases the chance of a species being viable on a particular substrate. However, this is clearly contradicted by the surviving species per substrate results shown in Fig. 3.5B. A more careful investigation of the mechanism that sustains diversity is displayed in Fig. 3.6. The first variable we consider is the number of surviving species, which is shown in Fig. 3.6A. All regimes show a significant loss in species diversity at a specific time. This is the time point where species that

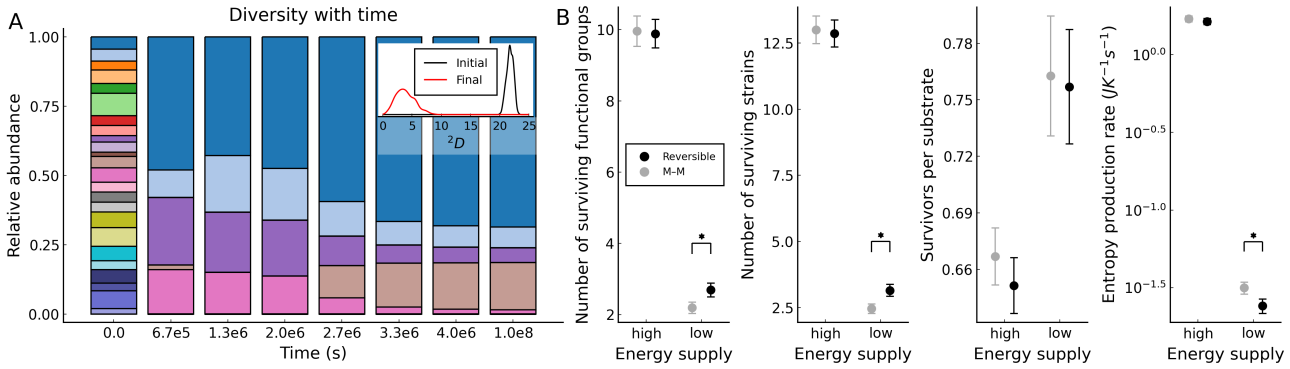


Figure 3.5: **Free-energy availability increases species diversity.** (a) Relative abundances of functional groups over time. The inset density plot shows initial and final distributions of inverse Simpson's diversity indices for functional groups across 250 simulations (using the same simulation parameters as Fig. 3.4). (b) Comparisons between the four free-energy × reversibility regimes. For each regime, the average and 99% confidence intervals obtained from 250 simulations are plotted. Two values of the total free-energy of the supplied substrate are considered: high ($1.5 \times 10^7\,\mathrm{J\,mol^{-1}}$), and low ($1.5 \times 10^6\,\mathrm{J\,mol^{-1}}$). For both energy regimes, we consider both reversible (black symbols) and Michaelis–Menten (M–M grey symbols) kinetics. Cases with a significant difference between these pairs are marked with a star ($P < 0.01$). The greatest differences are seen between high (labile) and low (recalcitrant) free-energy substrates, and only in cases of low free-energy does the kinetic scheme used cause significant differences.

never grow at all reach extinction. The number of surviving species drops to a significantly lower value across all regimes and then levels out, remaining at a substantially higher level for high free-energy (labile) substrates. When low free-energy (recalcitrant) substrates are considered, a small but significant difference in the number of survivors between the reversible and Michaelis–Menten regimes can be seen to emerge after the mass extinction event.

The second variable we consider is the number of substrates diversified, shown in Fig. 3.6B. For all regimes, the number of substrates rapidly plateaus. The plateau occurs after the point where the majority of species can no longer contribute to the substrate dynamics. At high substrate free-energy, the plateau occurs at a far higher number of substrates than at low substrate free-energy, and at low substrate free-energy the reversible case plateaus slightly higher than the corresponding Michaelis–Menten case. This mirrors the species diversity shown in Fig. 3.6A, indicating that diversity is driven by the number of available substrates (niches) being generated during assembly. The average growth rates for the four regimes are shown in Fig. 3.6C. All regimes show an initial period of rapid growth while the energy availability per species is high which settles down to a steady growth rate as the total population increases. The most

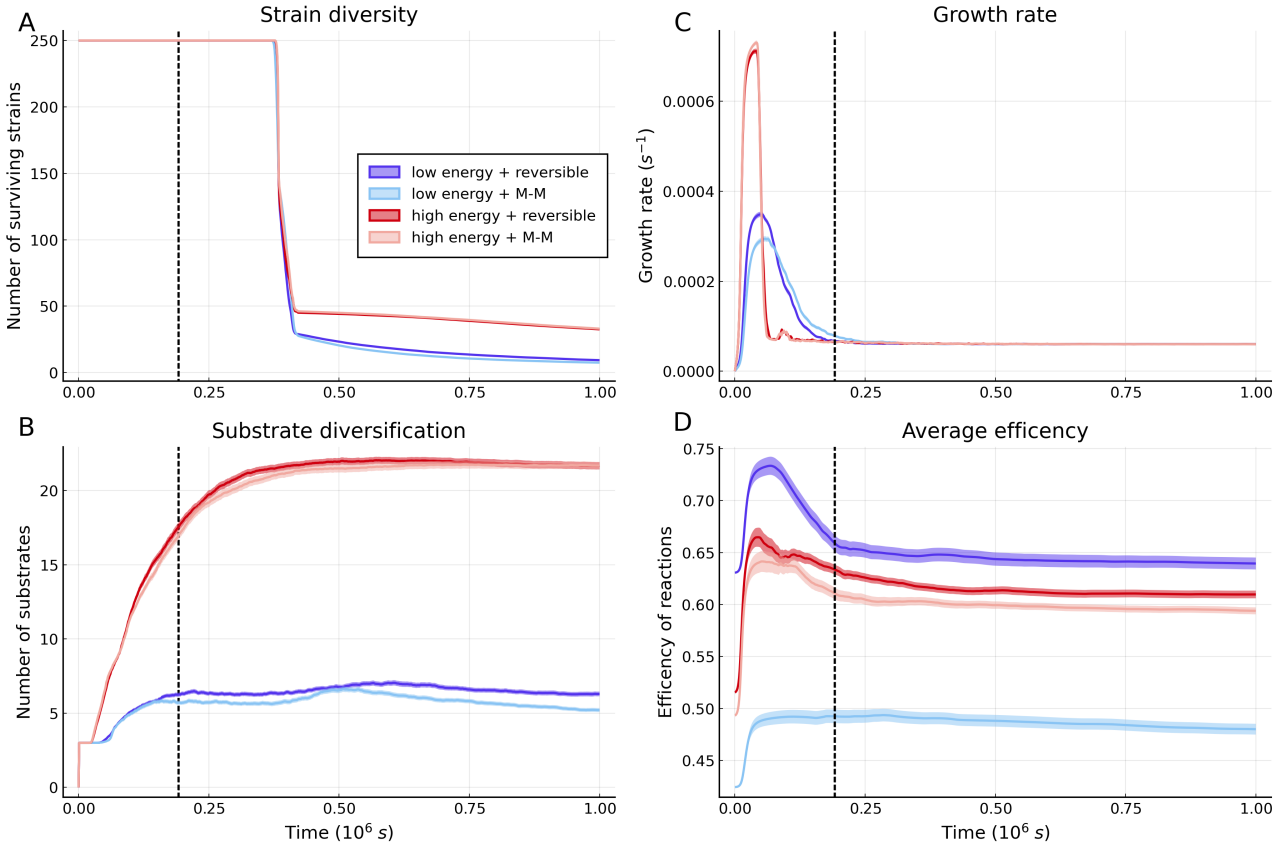


Figure 3.6: **Averaged assembly patterns across communities.** Averages (across 250 simulations) of key quantities over time are shown (with 99% confidence intervals) for the four regimes considered in Fig. 3.5. To better reveal the dynamics only a small initial time window is shown. (a) Average species diversity over time. The dashed black line shown in all four plots marks the time point where 75% of species have dropped below the threshold where they can no longer significantly impact the metabolite dynamics. (b) Substrate diversification plateaus for all regimes at approximately this time. (c) Across all regime growth rate rapidly peaks, the peak is both earlier and higher for high free-energy (labile) substrates. (d) Dynamics of the average reaction efficiency (Eq. 3.18), i.e. the fraction of the free energy change that is used to generate ATP rather than being dissipated. The position of the peaks can be seen to match the position of the corresponding growth rate peaks in plot (c). The four plots use the same parameters for the respective regimes as are used in Fig. 3.5.

pronounced peaks are seen for high free-energy (labile) substrates due to the greatly increased energy availability. There is a small difference in peak height for low free-energy (recalcitrant) substrates as allowing reactions close to equilibrium leads to greater free-energy extraction from the environment. The vastly different growth rates across species' populations account for the differing rates of niche generation (substrate diversification) seen in Fig. 3.6B.

The strategy of allowing near-equilibrium reactions increases free-energy yield, but also means that thermodynamic equilibrium occurs at far lower waste product concentrations. This means that the chance of a species expressing protein for a reaction that cannot proceed becomes far higher. To ascertain the importance of this effect we plot average reaction efficiencies (Eq. 3.18) with time in Fig. 3.6D. The average thermodynamic efficiency of the community changes through a process of species sorting, as the proportional abundance of species with differing average reaction efficiencies changes. Species sorting drives an increase in average reaction efficiency during the initial growth period across all regimes, as species with highly efficient reactions yield more free-energy, and thus grow faster. The two low free-energy (recalcitrant) cases are substantially further apart than the two high free-energy (labile) cases. This arises because low free-energy reactions are inherently closer to equilibrium. Hence, the reversible case has higher efficiency because its closer to equilibrium, and the efficiency of the Michaelis–Menten case is reduced more significantly to ensure reactions are far from equilibrium. In the reversible low free-energy case, after the initial period waste products accumulate and reactions become thermodynamically inhibited, a sharp efficiency peak is created. In the corresponding Michaelis–Menten case, the reactions are always far-from-equilibrium and so this inhibition never occurs and the efficiency simply plateaus. As the higher free-energy (labile) cases are generally further from equilibrium, the pattern here is likely driven by a transition from an initial highly efficient community that breaks down the initial substrate to more diverse community that breaks down a wider range of secondary substrates. This community is less thermodynamically efficient as the broader range of substrate means that competition for metabolites (ultimately, free-energy) is weaker.

3.5.3 Thermodynamic interaction types influence community assembly and diversity

Finally, we examined the preponderance of the different species interaction types (cf. Fig. 3.1D) under the low/high free-energy and reversible/irreversible regimes. As the choice of kinetic scheme only significantly affected previous results for low free-energy (recalcitrant) substrates, we first compare kinetic schemes for this case in Fig. 3.7. We show that under Michaelis–Menten like enzyme kinetics (Fig. 3.7A), there is a clear separation of scales between the two conventional (competition and facilitation) interaction types on the one hand, and the two thermodynamic ones (syntrophy and pollution) on the other. Our results thus imply that when considering the Michaelis–Menten (high dissipation) limit of a reversible kinetic scheme, thermodynamic interactions are observed (Fig. 3.7B), but are so weak as to not affect the dynamics (Fig. 3.7A). In contrast, when considering a reversible kinetic scheme the separation of scales between the strengths of the thermodynamic and non-thermodynamic interaction types disappears (Fig. 3.7C), and thermodynamic interactions become marginally more common (Fig. 3.7D). Taken together, this implies that thermodynamic interactions have a meaningful impact on the overall community dynamics when a reversible kinetic scheme is used and substrate free-energy is low. In Fig. 3.8 the results are shown for the high free-energy (labile) case, where a similar overall pattern is seen. However, the proportion of thermodynamic interactions is significantly lower, as is the mean strength of the thermodynamic interactions relative to the non-thermodynamic interactions. This is expected because the kinetic scheme is largely irrelevant to interaction dynamics when the system is assembled on a high free-energy substrate.

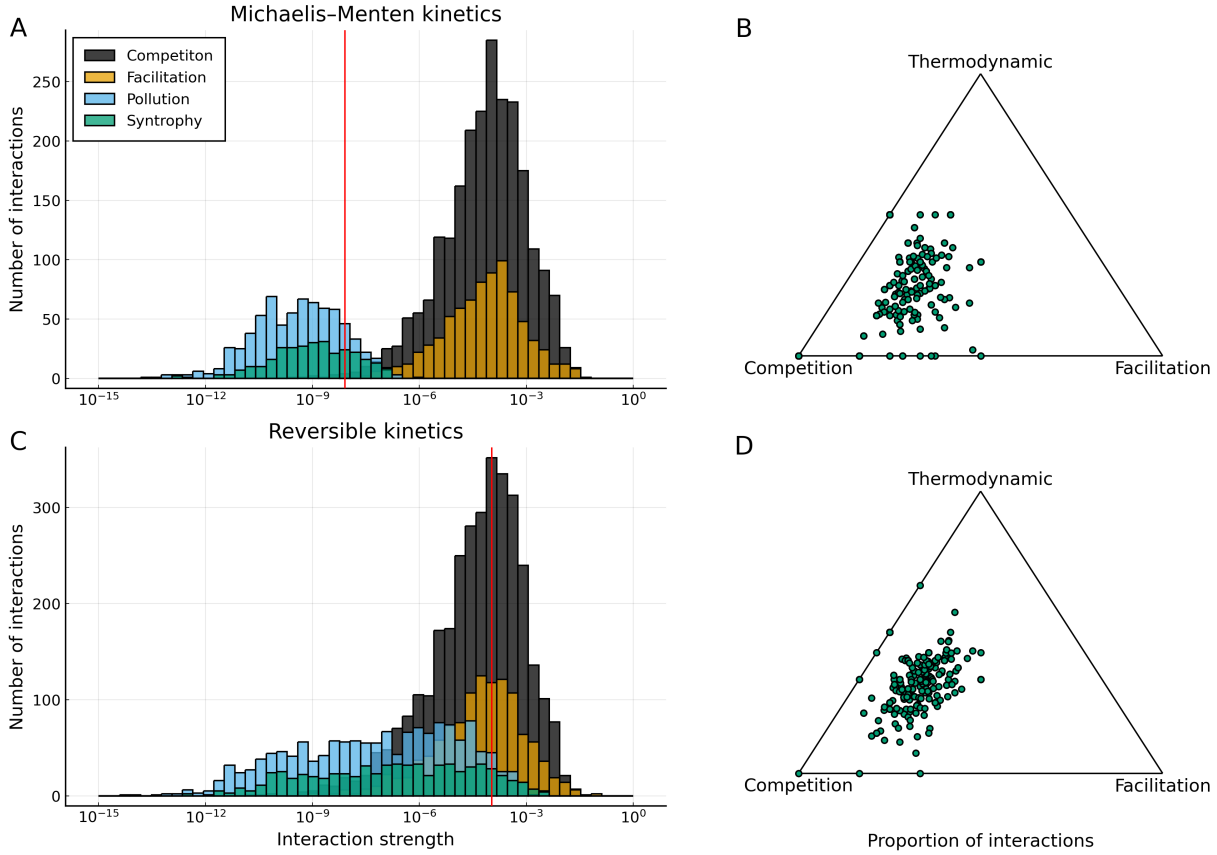


Figure 3.7: **Thermodynamic and conventional interactions can be of comparable strength.** (a) Histogram of strengths of the four interaction types (shown schematically in Fig. 3.1D). The vertical red line marks the point where the distribution of the strengths of the pollution interaction is at half its maximum. A clear separation of strengths between conventional and thermodynamic interactions can be observed. (b) Simplex showing the relative proportions of competition, facilitation and the thermodynamic (pollution and syntrophy) interaction types. Each point represents one out of the 250 communities we simulated for each regime. Competition is by far the most common interaction type, with thermodynamic and facilitation interactions being similarly common. (c) Corresponding plot to Fig. 3.7A for reversible kinetics. Thermodynamic and conventional interactions now reach comparable strengths. (d) Corresponding plot to panel B for reversible kinetics. Competition remains the most common interaction type, but thermodynamic interactions are now marginally more common than facilitation. All cases considered here are for low free-energy (recalcitrant) substrates ($1.5 \times 10^6\,\mathrm{J\,mol^{-1}}$).

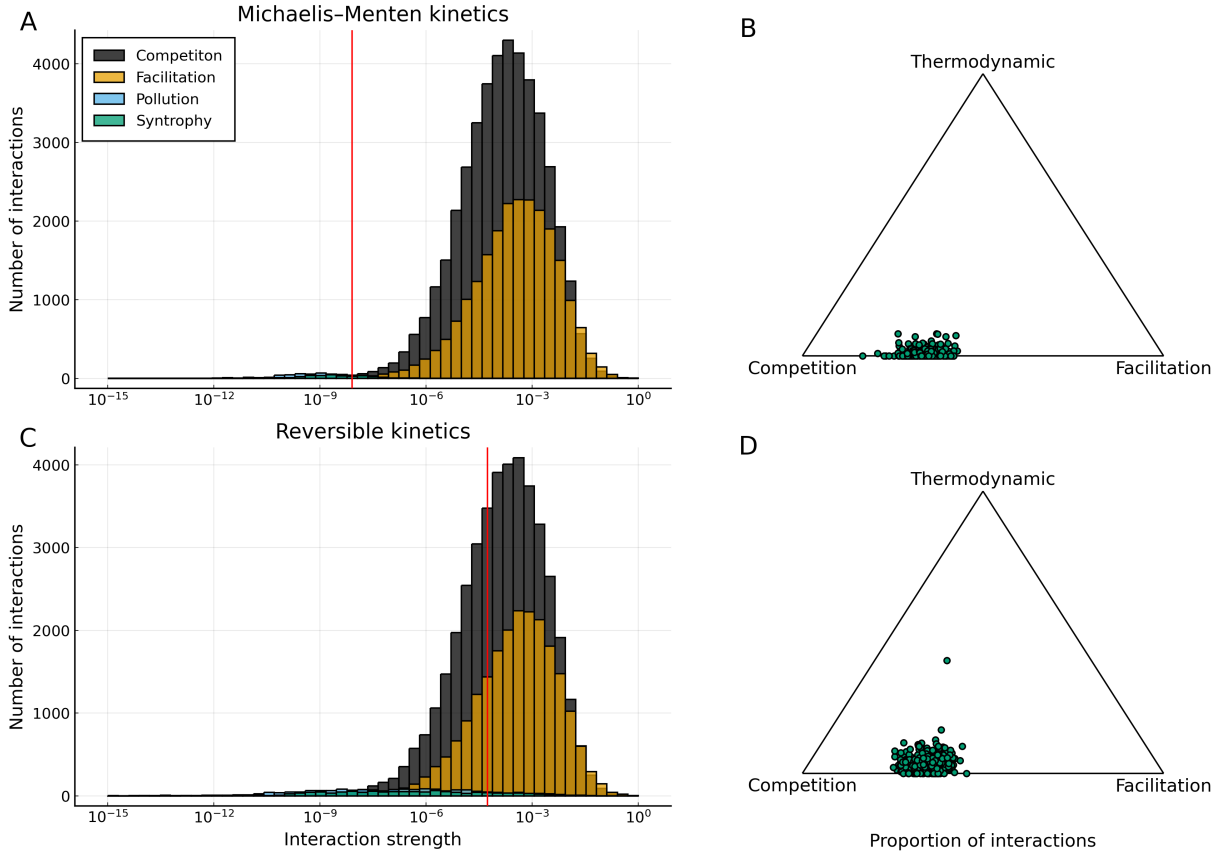


Figure 3.8: **Interactions in high free-energy case. Interactions high energy case.** Identical plot to Fig 3.7 but for the high energy supply case ($1.5 \times 10^7\,\mathrm{J\,mol^{-1}}$).

3.6 Discussion

We have shown that differences in the diversity of microbial communities are strongly influenced by the free-energy level of the substrates that they are assembled on (Fig. 3.5). This is because high free-energy (labile) substrates allow more rapid creation of niches during the initial phase of assembly. At the end of this initial phase of assembly, diversity collapses because species that are not able to grow due to the lack of available substrates (niches) go extinct. As more niches are generated in the high-free energy case than in the low-free energy case, a greater level of species diversity can be sustained at steady state (Fig. 3.6). We also find that for low free-energy (recalcitrant) substrates near-to-equilibrium reactions significantly increase the availability of free-energy, and thus the diversity. However, after the initial phase of community establishment, they are disfavoured as waste products build up. Underlying these dynamics are distributions of the different types of species interactions: thermodynamic interaction types (pollution and syntrophy) are rarer than the conventional ones (competition and facilitation), but they can reach similar magnitudes of strength when the supplied substrate has low free-energy and the kinetics are reversible (allowing non-equilibrium thermodynamics to operate) (Fig. 3.7). Taken together, these results indicate that thermodynamic inhibition can constrain the diversity of microbial communities when free-energy is scarce. A comparison of our models assumptions and results with other recent modelling work in the area can be found in Appendix H. In contrast to previous literature [GS16, PTW17, LLL$^+$20, PMSMG21] we never observe more surviving species than substrates. We also contrast our approach and results with previous work on the impact of energy supply on steady-state diversity [MCG$^+$19].

Thermodynamic constraints are known to be generally more important in anaerobic than aerobic microbial communities due to the lower free-energy availability. This is why previous microbial community models incorporating thermodynamic inhibition were developed specifically for anaerobic systems [HCR96, LWvB$^+$19]. However, aerobic microbial communities frequently assemble on relatively recalcitrant substrates in nature (e.g., high-cellulose or high-chitin resources; [LLMB17]). Therefore our model and results are relevant beyond anaerobic systems. For the relatively labile substrates typically used in laboratory experiments, a non-

thermodynamic model would be adequate. However, when modelling growth on more realistic recalcitrant substrates thermodynamic effects are likely to be important. Most microbes preferentially use labile substrates, but shifts towards using recalcitrant substrates has been observed to occur both as temperature increases [FLMS13], and as communities assemble [RSC+16]. Thus, understanding community-level thermodynamic constraints when the substrate is recalcitrant is necessary for truly understanding microbial community assembly in nature.

The relative importance of competitive versus cooperative interactions to microbial communities is currently an open question. Most of this work has taken a phenomenological empirical approach to characterise and classify such interactions [FB12]. Whole-genome community-scale metabolic modelling now allows a more mechanistic understanding of these interactions by predicting both the substrate overlap and metabolite production (leading to facilitation) between microbial species [MMA+21]. However, these approaches necessarily assume steady state metabolite concentrations, and do not account for the nonlinear feedback loops that develop in dynamic complex real microbial communities. We therefore took the middle road in terms of model complexity by incorporating non-equilibrium thermodynamics into a conventional consumer-resource model with cross feeding [Mac70, MCG+19]. Doing this introduces a dependence of reaction rate on product concentration, which leads to the emergence of novel interaction types between species via products rather than via substrates (see Figs. 3.1D & 3.7). Our results suggest that characterizing communities simply by either the relative proportion of positive and negative interactions, or the relative strength of competition and facilitation is insufficient to capture the true complexity of microbial interactions and their effects on community dynamics.

Because thermodynamic reversibility is explicitly included in our model, we are able to directly calculate the rate at which communities produce entropy (Eq. 3.16). This is a key advance because despite the long-standing interest in ecology about entropy production [OP55, Dew03], calculation of entropy production rates from ecological models and ecosystems has not been done in a explicit and consistent manner. Fig. 3.4 shows that spikes in the rate of entropy production over time can indicate periods where the availability of resources are changing rapidly. This has potential empirical relevance as it suggests that key episodes in the development of

microbial communities could be detected by the rate at which they add heat to their environment. Additionally, as seen in Fig. 3.5B, the rate of entropy production at steady state can be an indicator of free-energy availability, which only varies significantly between kinetic schemes in the case of low free-energy (recalcitrant) substrates. Potentially rates of community entropy production calculated from our model could be compared with experimental measurements of heat production, in order to establish the thermodynamic efficiency of microbial communities.

A fundamental assumption that our model makes is that microbial growth is limited by the availability of (free-) energy. This is one sense in which microbial growth can be considered to be carbon limited. The alternative way in which microbial growth can be carbon limited is through the requirement for organic compounds for anabolism. This is captured in the Weiße–Swain model [WODS15] through the dependence of growth rate on a single limiting amino-acid. Both our approach and the approach of Weiße et al. are clearly limited, as they only capture one factor among the many that limit microbial growth. However, capturing co-limitation in an analytically tractable manner presents significant difficulties, and therefore approaches based on a single factor are necessary. We would argue that there is real value in considering energy limitation, as (free-) energy represents a "currency" that can be expended to obtain other limiting resources (e.g. nitrogen or phosphorus). This is supported by the fact that one of the least thermodynamically efficient microbial taxa is found in conditions of severe nitrogen, iron, and phosphorus limitation [BFC17]. Thus, integrating nutrient limitation into an energetic framework is a challenging but potentially illuminating avenue for future work.

Another major assumption of our model is that the metabolite dynamics of the whole community can be appropriately captured by merely considering well mixed external metabolites that are accessible to all species, i.e. ignoring local metabolite concentrations and internal metabolites. This assumption is frequently made in microbial consumer-resource modelling [GLB$^+$18, MCG$^+$19, MCM20]. If we were to include internal metabolites, species would be able to maintain concentration gradients (through uptake and excretion), this would act to prevent reactions from reaching thermodynamic equilibrium. It is important to note however, that for the catabolic reactions that we consider the free-energy cost of maintaining a particular concentration gradient will always, due to the 2$^{\text{nd}}$ law of thermodynamics, be greater than the

reaction free-energy contributed by the gradient. The limits that thermodynamic inhibition place on microbial growth cannot therefore be evaded through this mechanism. Another limitation of our study is that we only considered metabolite hierarchies with equal free-energy spacing between metabolites. Extending this to allow free-energy gaps of variable sizes in the same hierarchy would mean that thermodynamic bottlenecks develop, which mean that species need to invest substantial amounts of protein in reactions close to thermodynamic equilibrium in order for the system to fully develop [NFBE$^+$16]. This potentially allows the thermodynamics of more realistic scenarios, such as overflow metabolism to be investigated [PB04, BHO$^+$15].

Recently, the existence of an (organism specific) upper limit on the rate of free-energy dissipation (entropy production) was suggested [NLH19]. An "overflow metabolism" was considered, that is when cells incompletely oxidise their growth substrate. By increasing its use of metabolic overflow pathways the cell makes increased use of reactions that dissipate less free-energy relative to the free-energy retained for growth, thus resulting in a maximum free-energy dissipation rate. Cells cannot adapt their metabolisms in such an manner in our model, where changes in the overall metabolism occur solely through species sorting. However, in our model a maximum metabolic flux is implied by the fact that both the metabolic proteome fraction (ϕ_P) and the (randomly chosen) kinetic parameters are bounded. Then for a particular substrate, which has fixed Gibbs free-energy, a maximum possible free-energy dissipation rate is implied. We would not expect this dissipation rate to be reached due to competition between species, and would instead expect the system to settle to a lower dissipation rate (see Fig. 3.4B). Interestingly our maximum dissipation rate arises from proteomic rather than thermodynamic constraints, suggesting a potential link with proteomic constraint based explanations for the existence of overflow metabolism [BHO$^+$15]. It is worth noting that if cells are genuinely constrained by the need to remove the entropy generated by free-energy dissipation, then thermodynamically efficient reactions would be additionally favoured due to producing less entropy. Species would therefore be expected to operate closer to thermodynamic equilibrium, implying that thermodynamic interactions could be more important than our results suggest.

Our results shed further light on the apparent simplicity of community assembly found in recent work [GLB$^+$18]. In Goldford et al., a diverse initial community was inoculated onto a single

carbon source [GLB+18], and the relative abundances of various taxa tracked over successive dilutions. Similar to Fig. 3.5A they observed a rapid collapse in diversity, which then levelled out to a relatively simple community with cross-feeding. Our results suggest this apparent simplicity is likely to emerge only when growth conditions are favourable (e.g. upon labile substrates). When considering assembly upon recalcitrant substrates, we find that the total number of substrates diversified at a given time is far more subject to random variation (see Fig. 3.6B). This suggests a far greater probability of major functional roles not being represented in the final steady-state community. The idea that favourable growth conditions leads to more predictable community assembly is further supported by field assessments of microbial diversity, where in (favourable) high pH conditions more diverse microbial communities were found but with less community variability between sites [GTJ+11]. Even for growth upon labile substrates our results deviate from strict simplicity, with a different set of differentiated metabolites emerging from each simulation run. However, this is probably consistent with Goldford et al. where variations were observed at the strain level but not at the coarser family (functional) level.

In addition, our results provide a new perspective on the modularity of assembly dynamics reported by recent empirical studies [DSG+16, EDS+19]. Datta et al. considered assembly in an open system (one where new species can enter) [DSG+16], finding a highly reproducible assembly process consisting of multiple distinct phases, throughout which species diversity was observed to vary non-monotonically. In contrast, in our closed system diversity declines monotonically, and thus the rate of niche generation is of critical importance (see Fig. 3.6C). Generated niches cannot contribute to diversity after the species that could feasibly occupy them have gone extinct. Furthermore, consistent with both real systems [EDS+19, GLB+18] and conventional microbial consumer-resource models [MCM20], modularity emerges in our model microbial communities (see Fig. 3.5A). We observe that the first functional group predominates across all regimes, but the relative abundances of the secondary functional groups are highly contingent as new species cannot enter the system, so the large number of niches generated by substrate diversification are not guaranteed to be filled. Functional groups also possess definite patterns of interaction types (see Figs. 3.1D & 3.7) with competition occurring primarily within functional groups, facilitation and syntrophy occurring primarily between functional groups,

and pollution occurring both within and between neighbouring functional groups (for further discussion of this point see Appendix F).

In the simulations that we performed the sole source of stochasticity was the choice of initial conditions, i.e. the generation of a random initial community for each simulation run. This is contrary to the highly stochastic growth that has been observed in experimental systems [KL13]. Our reason for ignoring this effect was to reduce the complexity of our simulations so that the impact of changing free-energy supply could be more easily separated from other factors. However, had we decided to incorporate this effect in our model we find it doubtful that it would have greatly affected the overall results. This is due to the substantial degree of functional redundancy that exists in our large initial communities. Thus, though stochastic effects can alter the outcome of competition between two species, any resulting difference in metabolite production would be buffered by increased growth of other community members. However, when the free-energy of the supplied substrate is low, stochastic population growth rate fluctuations could have a greater impact on the function of the final community. This is because in this case species experience greater difficulty establishing themselves (i.e. low growth rates), and thus stochastic fluctuations could prevent the establishment of species able to break down particular substrates. Therefore, we expect that stochastic population fluctuations would act to further disfavour community assembly upon low free-energy (recalcitrant) substrates.

Our model was designed to have sufficient complexity to capture the effects of cellular proteome fractions and non-equilibrium thermodynamics, which are generally ignored in community-scale models. Modelling at the intermediate scale provides novel, empirically-relevant insights, while avoiding the high complexity of models with fully-explicit intra-cellular dynamics [LWMX20]. Including a proteome trade-off generates a direct mechanistic link between free-energy availability and growth rate, because in regimes of greater free-energy availability, less of the proteome needs to be dedicated to metabolism, leading to a higher ribosome fraction, and thus the higher maximum growth rates seen in Fig. 3.6C. A more detailed investigation of the impact of this constraint on the dynamics is a promising avenue for future work, particularly in regards to the observed negative relationship between the number of ribosomal RNA operons a species possesses and its carbon use efficiency [RSS16]. We were unable to explicitly test whether a

specialist vs generalist trade-off exists in our model, because the substrate supply conditions we studied systematically favoured species with more reactions, as they had a greater probability of possessing a reaction for utilizing available substrates early in the assembly process. Changing the assembly process to allow continual immigration of species into the system would allow proper investigation into the existence and strength of the specialist vs generalist trade-off. Our framework could also fruitfully be extended by making links with metabolic ecology, particularly with the suggestion that the recently-reported variation in thermal sensitivity of microbial growth rates [STGC$^+$19, KSBP20] can be linked to proteome allocation [CGM$^+$17].

In summary, our results show that an explicitly thermodynamic model of complex, dynamically assembling ecosystems can provide novel insights into the mechanisms that generate and maintain diversity, how diversity depends on free-energy availability, the role of entropy production, and the nature of underlying interactions between populations

3.7 Outlook

In the preceding chapter, a novel model of microbial communities was developed, combining constraints arising from non-equilibrium thermodynamics with more commonly considered proteome constraints. This allowed the emergent impact of non-equilibrium thermodynamic constraints on the assembly of microbial communities to be studied. It was found that the speed of ecosystem development was set by the availability of free-energy, which then in turn determined the steady state diversity (see Fig. 3.6). Further, we found that in the limit of low substrate free-energy the results of our model significantly diverged from those of conventional models (see Figs. 3.5 & 3.6), i.e. those using irreversible Michaelis–Menten kinetics. This underscores the importance of explicit consideration of thermodynamics, as microbial communities in nature often exist in environments where free-energy is scarce.

Though this chapter has successfully demonstrated that non-equilibrium thermodynamics can be usefully incorporated into microbial community models, our approach possesses a major limitation. This is that the communities we consider assemble from a more diverse initial

community, with no new species ever entering the system. Though this is a reasonable approximation of many lab experiments (e.g. [GLB$^+$18]), the majority of microbial communities in nature experience the continual immigration of new species. Thus, in order to make a full consideration of how non-equilibrium thermodynamics constrains the development of microbial ecosystems (and ecosystems more broadly), a continuous assembly procedure must be considered. The following chapter will attempt to do this, applying a continuous assembly procedure to the model developed in this chapter. There will be a specific focus on the trade-offs that emerge when a combined consideration of proteomic and non-equilibrium thermodynamic constraints is made.

Chapter 4

Thermodynamic and proteomic trade-offs in continuously assembling microbial communities

4.1 Introduction

Understanding ecological assembly is fundamental to understanding the diversity and functioning of natural ecosystems. This is particularly pertinent for microbial communities, due to the fast timescales the operate on, meaning that species turnover can occur rapidly enough to affect the dynamics of key functional processes such as carbon cycling. For microbial communities, species immigration has been found to play a vital role in buffering against metabolic diversity loss due to environmental warming [LBB16], and where even low immigration rates have been found to have a homogenising effect on the community composition [FLVC$^+$19]. Furthermore, the majority of microbial communities found in nature assemble continuously [DSG$^+$16, EDS$^+$19], i.e. with new species perpetually entering the community.

In the quest to understand the dynamics of microbial communities, of similar importance is the need to understand the role of the trade-offs that all microbial species operate under. Trade-

offs generally stem from the fact that limited cellular resources (e.g. protein or free-energy) have to be divided between many different tasks. For example, allocation of energy to sensing must necessarily decrease the energy available to allocate to other energy consuming cellular processes. The microbial cell must, therefore, adopt a particular strategy for the allocation of its limited resources, which results in it being better or worse at critically important tasks (e.g. to grow rapidly or to sense food). These emergent capabilities of the cell determine its competitive ability.

Given the wide variety of cellular processes, many trade-offs have been described in biology. However, only rarely are these linked to fundamental mechanisms. Recent work within the field of microbial ecology has taken the important step of considering how ecologically relevant trade-offs can emerge from fundamental physical constraints. Notably, a key trade-off is between maximum growth rate and carbon use efficiency, which stems from the number of ribosomal RNA (rrn) operons that a species possesses, with more operons being linked to faster but less efficient growth [RSS16]. Another key issue is how a finite cellular proteome leads to a trade-off between thermodynamic efficiency and metabolic protein cost. This manifests both in prokaryotic microbes utilising different glycolysis pathways [FNBE$^+$13], and in microbes switching to use overflow metabolism as substrate availability increases [BHO$^+$15]. However, while trade-offs at the cellular or single population level are widely studied their impact on selection at the community level is rarely considered.

We argue that the problem of continuously-assembling microbial communities cannot be considered separately from the question of the trade-offs that individual microbes are subject to. This is because the fitness effects of any trade-off on a microbial population is determined by both the condition of its environment and its interaction with other species, both of which change significantly as communities assemble and turn-over. Linked to this, species that can successfully establish in the community at a given stage in the assembly will be determined by the dominant trade-offs they are subject to, and therefore, the process of assembly itself is determined by trade-offs on individual species. A few recent attempts to address this issue have been made [LLL$^+$20, NSR20], but these have either considered only simplistic, phenomenological, fixed trade-offs (e.g. fixed enzyme budgets in [LLL$^+$20]), or have had to introduce arbitrary costs

to prevent certain strategies from being universally favourable (e.g. the penalty for possessing more reactions in [NSR20]).

In this study, we consider trade-offs emerging from fundamental mechanisms and the impact they have on selection at the community level. This kind of consideration has been made before for a trade-off between ATP production rate and yield [PSB01]. However, this previous work only considered a single trade-off out of the large number that microbial cells are subject to. To address this limitation, we shall consider three specific trade-offs (shown in Fig. 4.2). Firstly, the trade-off between the number of reactions that a species possesses and the effective rate of each reaction (Fig. 4.2A). Species with more reactions have to split their proteome into more subdivisions, leading to less enzyme for each reaction (on average), hence a trade-off between the number of reactions a species utilises and the (average) effective rate of each reaction emerges. Secondly, the trade-off between the translation rate of ribosomes and their efficiency (in terms translation steps per ATP hydrolysed) (Fig. 4.2B). This trade-off is motivated by the empirical observation that microbial species that possess a greater number of rrn operons have higher maximum growth rates, but at the cost of lower carbon use efficiencies (measured as the ratio of carbon allocated to cellular biomass to the total carbon uptake) [RSS16]. This implies that a higher maximum ribosome fraction (ϕ_R) is achieved, but at the cost of decreased efficiency, likely due to the increased demand for sufficient growth precursors incurred by having more ribosomes translating proteins. Finally, the trade-off between ATP production rate and yield (Fig. 4.2C), which was previously introduced in Chapter 3. The free energy released by catabolic reactions can either be dissipated in order to drive the reaction or transduced as ATP (via a coupled reaction), if the amount of free-energy dissipated becomes too small then the reaction become thermodynamically limited the rate of the reaction (and ATP production rate) decreases. Hence, a trade-off between ATP production rate and yield emerges.

Here we address the limitations of previous work, both, by considering trade-offs arising mechanistically from fundamental thermodynamic and proteomic constraints, and doing so in a more realistic context of continuously-assembling microbial communities. To this end, we use a novel model that extends the model developed in Chapter 3 (see also [CPE21]) such that it better reflects biological reality, allowing key trade-offs to be captured.

4.2 Extension of thermodynamic microbial community model

The model we make use of in this chapter is the model developed in Chapter 3, with a small number of alterations. In this model, each metabolite (β) is represented by its concentration C_β and each consumer (i) is represented by three variables: its population abundance N_i, ribosome fraction $\phi_{R,i}$, and internal energy (ATP) concentration a_i. The dynamics of these variables are given by

$$\frac{dC_\beta}{dt} = \kappa\delta_{\beta,1} - \rho C_\beta + \sum_{i=1}^{B} \left(p_{i,\beta}(\mathbf{C}) - c_{i,\beta}(\mathbf{C})\right) N_i \tag{4.1}$$

$$\frac{dN_i}{dt} = (\lambda_i(a_i, \phi_{R,i}) - d_i)N_i \tag{4.2}$$

$$\frac{d\phi_{R,i}}{dt} = \frac{1}{\tau_g}\left(\phi_R^*(a_i) - \phi_{R,i}\right) \tag{4.3}$$

$$\frac{da_i}{dt} = J_i - \chi(\phi_{R,i})m\lambda_i - a_i\lambda_i, \tag{4.4}$$

where κ is the substrate supply rate, $\delta_{\beta,\xi}$ is the Kronecker delta, ρ is the metabolite dilution rate, B is the number of species in the community, $p_{i,\beta}$ is the (per cell) rate that the i^{th} species produces metabolite β, and $c_{i,\beta}$ is the (per cell) rate at which it consumes metabolite β. Further, λ_i is the i^{th} species' growth rate, d_i its rate of biomass loss, τ_g is the characteristic time scale for growth, ϕ_R^* is the ideal ribosome fraction, J_i is the ATP production rate for the i^{th} species, χ is ATP use per elongation step, and m is the total mass of the cell (in units of amino acids). At the highest level the model is unchanged from the model derived in Chapter 3 (see Eqs. 3.1–3.4). However, several alterations have been made to lower levels of the model, which we shall now detail.

The first alteration we make is to the parameterisation of the function for the 'ideal' ribosome fraction. This function (Eq. 3.7 in the previous chapter) is given by

$$\phi_R^*(a_i) = \frac{a_i}{\Omega_{\frac{1}{2}} + a_i}(1 - \phi_R^{\max}), \tag{4.5}$$

where $\Omega_{\frac{1}{2}}$ is a half-maximum constant, and $\phi_R^{\max}$ is the maximum ribosome fraction. Previously we made the assumption that all species have the same maximum ribosome fraction given by

$$\phi_R^{\max} = 1 - \phi_Q, \tag{4.6}$$

where ϕ_Q is the fixed housekeeping proteome fraction. This is inconsistent with the previously-mentioned differences in the number of ribosomal RNA operons between species [RSS16]. Having more operons allows a greater rate of transcription of genes coding for ribosomes, allowing a higher maximum number of ribosomes in a cell. To address this, we modify Eq. 4.6 in the following way:

$$\phi_R^{\max} = \omega_i \left(1 - \phi_Q\right). \tag{4.7}$$

Here, ω_i is a factor expressing how close species i's maximum ribosome fraction is to the maximum biologically possible (Eq. 4.6); it can take values between 0 and 1.

The second alteration we make is again to the parameterisation of the function for the 'ideal' ribosome fraction (Eq. 4.5). In the previous model the ribosome half-saturation constant ($\Omega_{\frac{1}{2}}$) was treated as constant, but it is now taken to vary with ω_i as

$$\Omega_{\frac{1}{2}}(\omega_i) = \Omega_m + f_\Omega(\omega_i - \omega_m)\Omega_m, \tag{4.8}$$

where Ω_m is the lowest possible values for the saturation constant, ω_m is the minimum possible values of the maximum ribosome factor, and f_Ω is a factor how rapidly the saturation constant increases with increasing maximum ribosome factor. This alteration ensures that as the maximum ribosome factor ω_i increases, the basal ribosome expression (i.e. the ribosome fraction for low a) doesn't increase too rapidly. In real cells, basal expression does rise with increased maximum expression. However, if this effect is too strong a pathological situation emerges, where species with a high maximum ribosome factor (ω_i) express ribosome fractions that are too large to be viable, except when free-energy availability is extremely high. This situation is displayed in Fig. 4.1A. If instead of assuming that $\Omega_{\frac{1}{2}}$ is fixed between species Eq. 4.8 is used, the difference in ribosome expression with changing ω_i can be reduced to a more plausible level,

whilst also ensuring that differences in basal expression persist (see Fig. 4.1B).

The final alteration we make is to the ATP use per translation step (see Eq. 4.4). The previous form of the model assumed that that all modes of growth are equally efficient, i.e. a translation elongation step always costs the same amount of ATP (χ) regardless of the state of the cell. This is inconsistent with the empirical observation that growth efficiencies vary both between species [RSS16], and environmental (e.g. temperature) conditions [SCBP21]. To account for this we define ATP use per translation step as a function of ribosome fraction as

$$\chi(\phi_{R,i}) = \chi_l + \chi_u \frac{\phi_{R,i}}{1 - \phi_Q}, \tag{4.9}$$

where χ_l is the minimum cost of a translation elongation step, and χ_u is maximum additional cost per translation step. The second term in the above expression ensures that the maximum ATP use per protein translation step occurs when the maximum biologically possible fraction of the proteome is allocated to ribosomes (Eq 4.6). The form of the expression arises from the fact that as the number of ribosomes a cell possesses increases, the demand for precursors for protein synthesis must necessarily increase. This additional requirement for precursors can be met by additional energy expenditure. However, the cell will most likely be faced with diminishing returns, with each additional increase in the rate of energy investment resulting in a smaller increase in the rate at which precursors are obtained. Hence, the increase in ATP use per translation step (decrease in efficiency) with increasing ribosome fraction in Eq. 4.9.

4.3 Key thermodynamic and proteomic trade-offs

With the model suitably altered, we can now consider three key trade-offs that are universally experienced by microbial cells (Fig. 4.2). Detailed discussion of each of the three trade-offs is provided below, with attention paid to how they are encoded in our (altered) model.

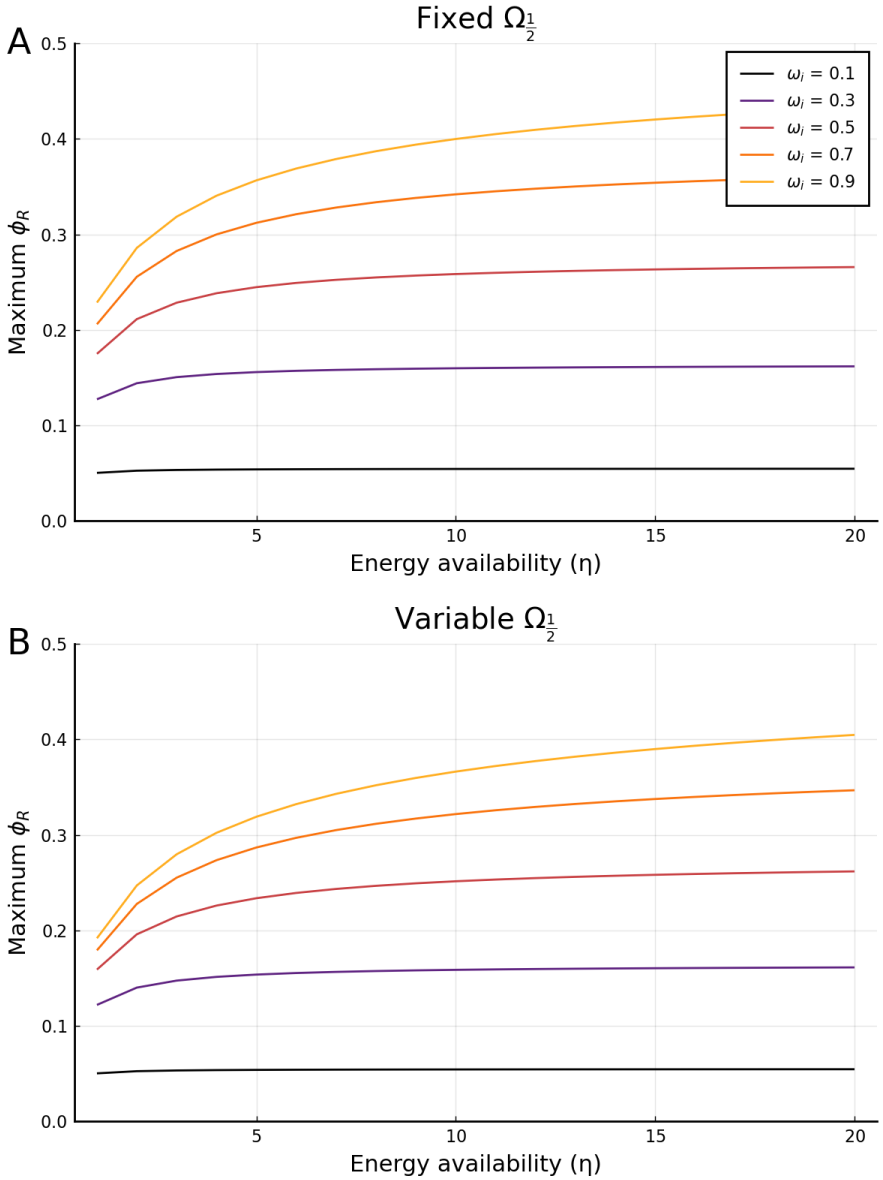


Figure 4.1: **Maximum ribosome fraction varies substantially with free-energy availability.** Simulations were run of a single species with a single reaction and a large initial quantity of substrate, leading to exponential population growth. The maximum ribosome fraction achieved in the exponential growth phase was then recorded. This was then performed for a range of different free-energy availabilities (which we used η as a proxy for) and maximum ribosome factors (ω_i). (a) Comparison the responses to changing energy availability for different ω_i values, when $\Omega_{\frac{1}{2}}$ is held fixed. (b) Comparison the responses to changing energy availability for different ω_i values, when $\Omega_{\frac{1}{2}}$ varies following Eq. 4.8.

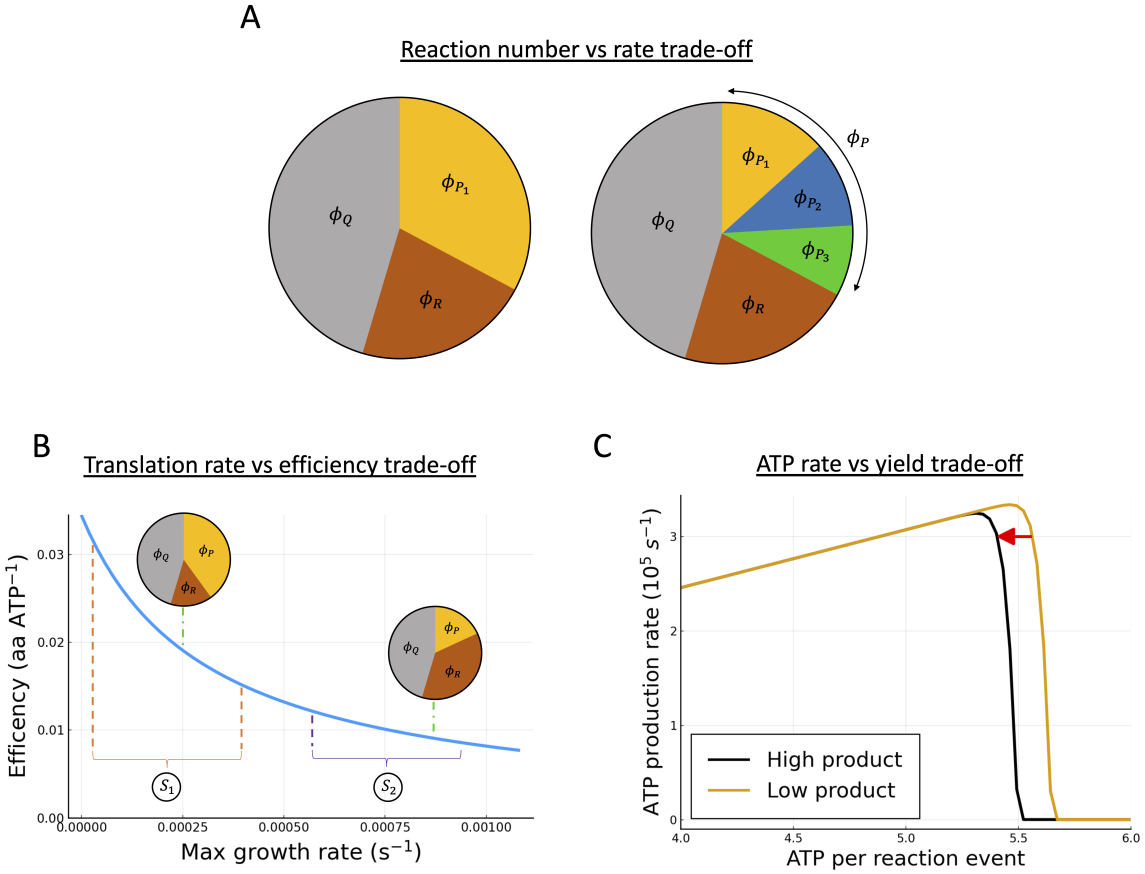


Figure 4.2: **Visualisation of trade-offs.** (a) Reaction number vs rate trade-off: Assuming equal sized ribosome fractions (ϕ_R), when a species possesses more reactions its metabolic proteome fraction (ϕ_P) must be further subdivided. A trade-off therefore emerges between the species being able to utilise more substrates (due to possessing more reactions) and the investment it can make in each reaction. This decreased investment (per reaction) results in lower effective reaction rates. (b) Translation rate vs efficiency trade-off: Increased ribosome fraction allows an increased maximum growth rate, provided internal energy concentration is sufficiently high (see Eq. 3.5). However, this comes at the cost of decreased efficiency, i.e. fewer translation steps (measured in amino acids added) per ATP hydrolysed (see Eq. 4.9). As a reduced maximal ribosome fraction leads to a lower basal expression (see Eq. 4.8), this allows distinct strategies to emerge. A strategy of low basal ribosome expression could be adopted (S_1), this leads to more efficient growth but a lower maximum growth rate. Alternatively, a strategy of high basal ribosome expression could be adopted (S_2). In the latter case, growth is far less efficient, but far higher growth rates can be achieved. (c) ATP rate vs yield trade-off: This thermodynamic trade-off emerges in any model that incorporates thermodynamically reversible enzyme kinetics (mentioned previously in Fig. 3.1). ATP production rate increases linearly with an increase in ATP per reaction event (η), up to the point where thermodynamic inhibition becomes significant. The position of optimal η value shifts as metabolite concentrations in the environment change.

4.3.1 Reaction number vs rate

The reaction number vs rate trade-off (Fig. 4.2A) emerges from the constraint of a finite proteome; the number of proteins a cell can possess is limited, meaning that the synthesis of proteins for a particular task must come at the cost of reduced expression of protein for other tasks. The original application of this constraint was to explain two apparently contradictory relationships between bacterial growth rates and the fraction of their proteome that they dedicate to ribosomes. This apparent contradiction was explained as resulting from a trade-off between the number of ribosomes (required to synthesis protein) and metabolic proteins (required to obtain energy and precursors for growth) [SGM$^+$10, SKMH14]. Recently, this constraint has been applied more broadly in order to consider other classes of protein such as those required for chemotaxis and motility [NCL$^+$20], and chaperone proteins that prevent protein misfolding [CGM$^+$17]. Here, we consider a trade-off between the number of reactions a species possesses and the portion of the metabolic proteome fraction that can be allocated to each reaction. Though a model considering proteome trade-offs between reactions has previously been developed [PMSMG21], no consideration was made of the number of reactions each species possessed.

The model that was previously developed in Chapter 3 (derived in detail in Appendix E) can capture this trade-off, because it mechanistically links population growth rate (Eq. 3.5) to the trade-off between ribosome and metabolic proteome fractions. Further, the amount of enzyme that a species can dedicate to a particular reaction (α) is dependent both on the relative reaction expression ν_α and the metabolic proteome fraction ϕ_P (see Eq. 3.9). As the relative reaction expressions for a given species have to sum to 1, this leads naturally to a trade-off between the number of reactions a species possesses and the effective reaction rate of each reaction. This is because, assuming a fixed ribosome fraction, the amount of enzyme a species can allocate to each reaction is inversely proportional to the number of reactions it possesses. By incorporating this trade-off into our modelling framework, the question of what the optimal number of reactions for a microbial species is and how this changes with changing conditions can be addressed.

4.3.2 Translation rate vs efficiency

The translation rate vs efficiency trade-off (Fig. 4.2B) is captured in our model by allowing the maximum ribosome fraction to vary between species (see Eq. 4.7), with different species possessing different maximum ribosome factors (ω_i). This variation means that different species can now possess different proteome allocation strategies, leading to differences in ribosome fraction (ϕ_R) at steady-state. As the efficiency of species now depends on their ribosome fraction (Eq. 4.9), this leads to a variation between species of the amount of ATP used per translation step. This allows both high efficiency (low maximum growth rate) and low efficiency (high maximum growth rate) growth strategies to exist in our model. Properly capturing this trade-off is important, as it allows the inefficiency inherent in microbial growth to be separated from the inefficiency inherent in the breaking down of substrates in order to obtain energy. Additionally, the incorporation of this trade-off into our model allows us to avoid the peculiar behaviour seen in Fig. 3.4B, where all species show the same growth efficiency and so at steady-state all species possess equal ribosome fractions.

4.3.3 ATP rate vs yield

The ATP rate vs yield trade-off (Fig. 4.2C) emerges from fundamental non-equilibrium thermodynamics. It is encoded in our model through our choice to use thermodynamically reversible enzyme kinetics. Our formulation of it shows a linear increase in the rate of ATP production as ATP per reaction event (η) increases, up to a peak where it then drops very sharply to zero. This peak emerges when the free-energy dissipated by the reaction is small enough that the reaction becomes close to equilibrium, at this point any additional free-energy that is transduced (i.e. by increasing η) rather than dissipated moves the reaction meaningfully closer to equilibrium, thus slowing the reaction rate substantially. The position of maximum ATP production rate depends on the waste-product concentration. This arises as the concentrations of both the substrates and the waste-products determine how close a reaction is to thermodynamic equilibrium. Thus, reducing the build up of waste-products can increase the amount of free-energy that can be transduced without the reaction being slowed significantly by waste-product

inhibition.

4.4 Assembly simulations

A schematic of the community assembly simulation procedure is shown in Fig. 4.3. The procedure starts with a small random initial community. This community is not stable and generally the majority of its members go extinct. Immigration events then occur at random intervals. The reason for these intervals being random is to reduce the possibility of pathological interactions between the timescale of immigration and the timescales of other processes. The number of species entering the system at each immigration event is also random, though with one being the most common number of species. This was done so that our procedure was as generic as possible, accounting for the fact that multi-species invasions frequently occur for real microbial communities. Each simulation is run for a set number of immigration events N_I, taken to be $N_I = 500$ across the simulations performed for this chapter. After the final immigration event the system is integrated for an extended time period (5×10^5 s), so that the system can settle to an (at least approximate) steady state.

Our new simulation procedure is applied to a model that is essentially the model used in Chapter 3, but with alterations made to it as discussed in the previous section. The system was described by the same system of ordinary differential equations (Eqs 4.1–4.4). However, there were a number of changes to at lower levels of the model. The 'ideal' ribosome fraction (Eq. 4.5) now makes use of the maximum ribosome fraction defined in Eq. 4.7 rather than Eq. 4.6. Additionally, the cost per translation step (χ) and the ribosome half-saturation 'constant' ($\Omega_{\frac{1}{2}}$) are no longer treated as constants and are instead defined by Eqs. 4.9 & 4.8, respectively. The parameters used for the simulations in this chapter are identical those provided in Tabs. A.5 & A.6, with the exception of the newly introduced parameters. These take the following values, $\omega_m = 0.1$, $\Omega_m = 1 \times 10^9$, $f_\Omega = 1.5$, $\chi_l = 29.0$ [LM15], and $\chi_u = 101.5$. Values for the the maximum ribosome factor ω for each species were drawn randomly from a uniform distribution between ω_m and 1. Again, the (M metabolite) reaction networks were generated

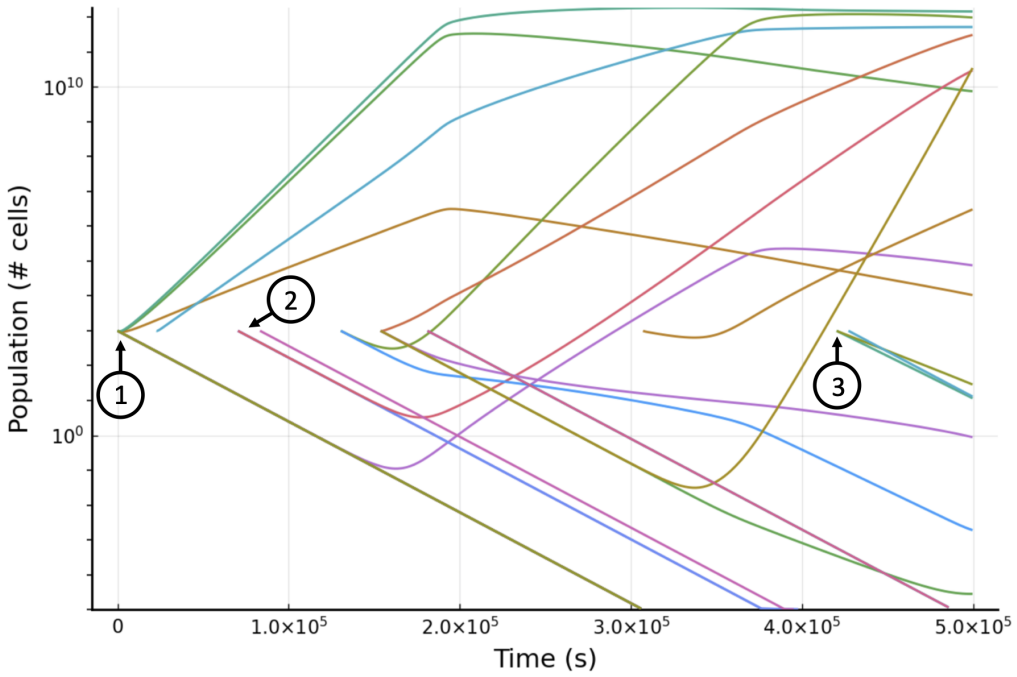


Figure 4.3: **The continuous assembly simulation procedure.** (1) The simulation is started with a small ($N = 10$) initial community. As this community is randomly generated, it is not at steady state and species from it either grow or die off. (2) Immigration events occur at exponentially distributed random time intervals with a mean time of 1×10^5 s. (3) The number of species that immigrate into the system at a given immigration event is Poisson distributed with mean 1.5.

following the pattern shown in Fig 3.1B, with the free-energy spacing of each step downwards in the metabolite hierarchy being equal. The procedure for assigning each species' reaction set (O) and for generating η values, kinetic parameters and proportional reaction expressions for each species for each reaction was the same as the one detailed in Section 3.4. The number of reactions per species was allowed to be an odd number between 1 and 7 (inclusive), and the maximum η value was set such that the reaction will reach equilibrium at a product to substrate ratio of 1×10^{-2}.

4.5 Results

4.5.1 Trade-offs in the absence of immigration

We first consider how the trade-offs manifest for communities that assemble without continuous species immigration. In order to do this, we choose initial random communities with 250 members which assemble upon a high free-energy substrate (5×10^6 J mol^{-1}). Their assembly dynamics are then simulated up to a time of 2.5×10^6 seconds. The results of these simulations are summarised in Fig. 4.4. Without continuous species immigration it is always advantageous for a species to possess more reactions (see Fig. 4.4A). This is driven by the fact that species with more reactions have an increased probability of encountering a substrate they can make use of to grow (see Fig. 4.4B), and thus have an increased probability of surviving in the community. At the very beginning of community assembly high maximum ribosome fractions are selected for (see Fig. 4.4C), in order to sustain a high ribosome fraction (see inset of Fig. 4.4C). After this period the average maximum ribosome fraction collapses, recovers, and then begins a gradual decline. This decline is driven by a decreased need for ribosomes as the community approaches steady state, resulting in species with lower basal ribosome expression being favoured. In Fig. 4.4D, it can be seen that ATP yield initially spikes, and then settles down to a (lower) stable value. This is likely driven by high ATP yield being favourable for growth, but high yields becoming thermodynamically unfeasible as waste products build up. Across all these trade-offs a clear optimal strategy for long term survival exists. However, this

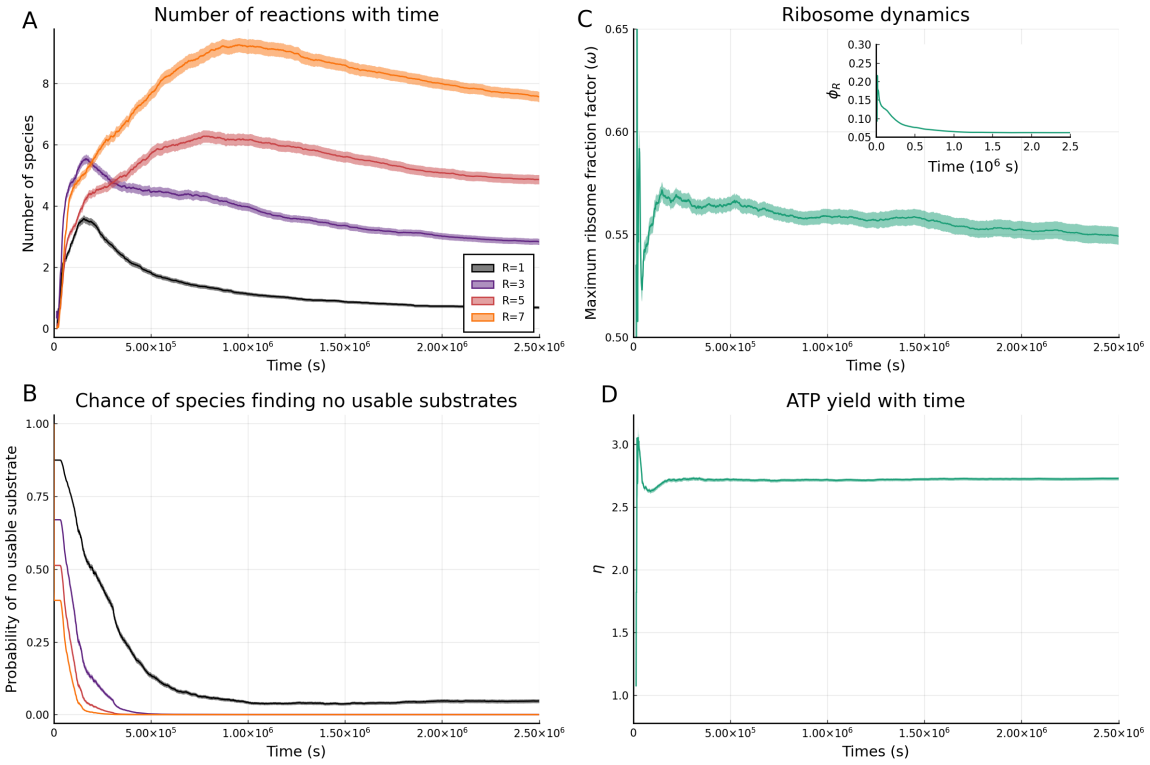


Figure 4.4: **The absence of immigration results in straightforward trade-off optima.** All quantities plotted in this figure are averaged across 250 simulations, and have their corresponding standard errors shown. (a) Plot of the number of species in the community that have specific number of reactions. In the initial phase of community assembly each class increases in frequency irrespective of its number of reactions, but as the system develops the class with the maximum allowed number of reactions ($R = 7$) comes to predominate. (b) Corresponding plot of the probability that a species with a specific number of reactions does not find a usable substrate in the system. The probability for species with multiple reactions drops fairly quickly to (almost) zero. (c) Plot showing how the maximum ribosome factor (ω) varies through the community assembly. The inset plot shows that on average the ribosome fraction (ϕ_R) settles down to a steady value after the initial growth phase. (d) Average ATP yield with time. After an initial spike ATP yield settles to a steady value.

result is likely complicated when continuous immigration is included.

4.5.2 Dynamics of continuous species immigration

The key features of the microbial community assembly dynamics are shown in Fig. 4.5. Both the total population size and the species diversity rise rapidly during the initial phase (see Fig. 4.5A). This is driven by the increasing number of secondary substrates that become available as the initially-supplied substrate is broken down, which leads to more of the free-energy of the substrate being released and more niches becoming available to support species. After

this initial phase the total population continues to grow, driven by the immigration of new species that can more effectively survive upon the substrates in the system. In contrast, the diversity saturates, indicating that substantial changes to the species of comparatively high population occur only rarely. This suggests that a stable subsystem has emerged, that is a community of species with populations several orders of magnitude higher than those of invading species, which remains at an approximate steady state as invasions occur. Generally, invaders die out without causing any significant change to this subsystem, only rarely do successful invaders cause extinctions events which lead to significant changes in the subsystem. It is also worth making note of the steady-state distribution (across all simulations) of ribosome fractions (shown as an inset in Fig. 4.5A). Here, in contrast to the results of the previous model (see Fig. 3.4B), meaningful variation in steady-state values for the ribosome fractions are now observed. This suggests that multiple viable growth strategies now co-exist.

In Fig. 4.5B, the total number of survivors can be observed to peak early and then decline. The fact this decline has no observable impact on the diversity is further evidence for the existence of a relatively stable but still invasible subsystem. The observed decline is likely due to the decreased probability of successful invasion as the system develops, which leads to less frequent extinctions of species at high population. As the competition between invaders and high population residents can take a long time to resolve (i.e. by the resident being driven to extinction), this elevates the total number of species in the system. The invasibility of the community is defined as the proportion of species that enter the community at low population which successfully establish themselves (at higher population) rather than going extinct. This can be approximately measured by calculating the fraction of species that immigrate within a particular time window that are still growing by the end of that time window. We use this measure instead of instantaneous growth (at time of invasion) to control for cases where species demonstrate different growth behaviours over the intermediate term. To capture this we therefore chose a time window of intermediate length ($t_w \sim 6 \times 10^4$ s). When this approximate invasibility measure is plotted it can be seen to peak at roughly the same time as the peak in the total number of species in the system. Finally, the distribution of surviving species with time of immigration (inset to Fig. 4.5B) shows that despite the pattern seen for invasibility,

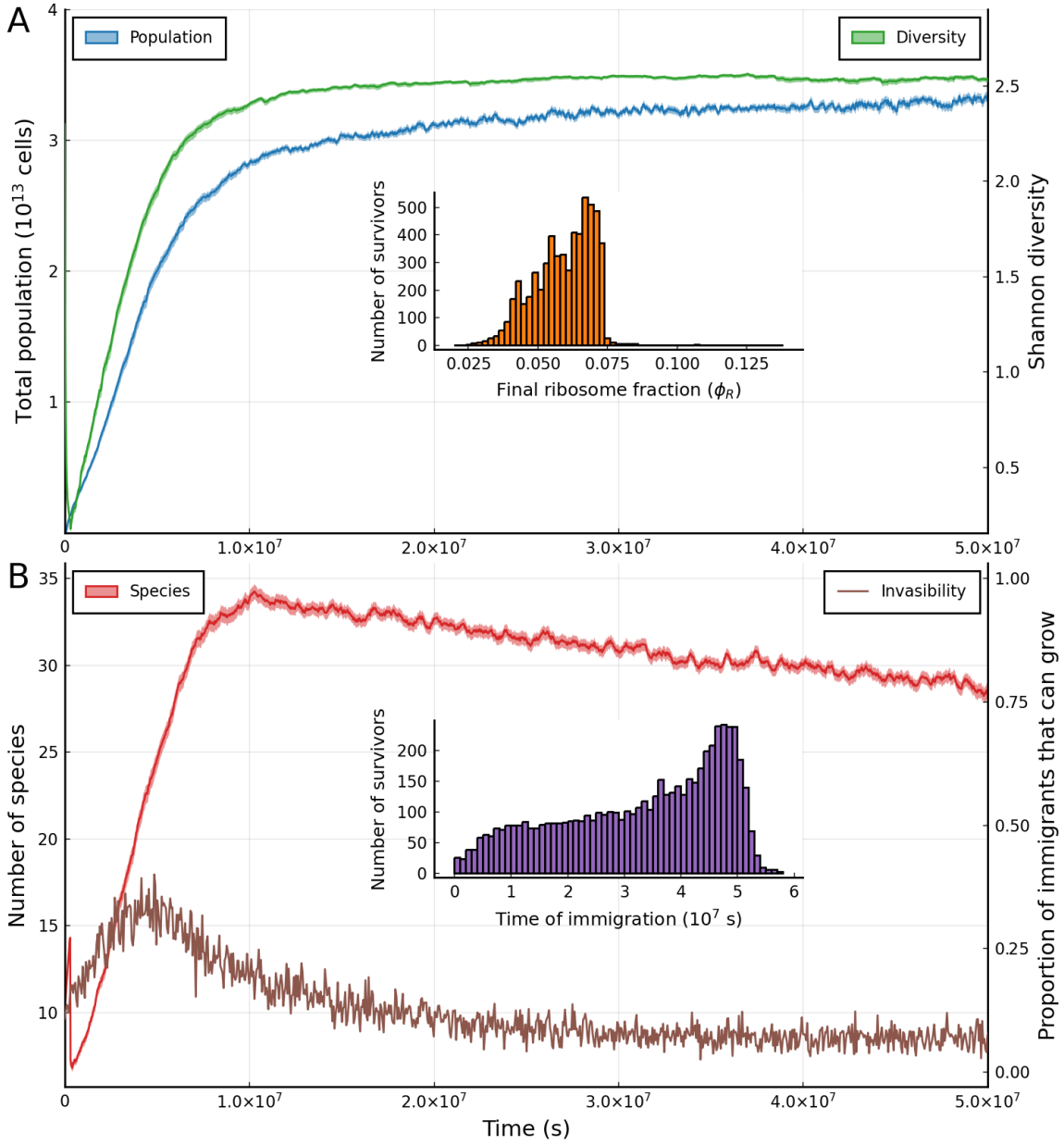


Figure 4.5: **Summary statistics show average community development.** (a) Dynamics of the total population and diversity over time. Total population continues to increase up to the end of the considered time-period, but in contrast the diversity saturates. The inset histogram shows the distribution of final ribosome fractions of surviving species across all simulations. (b) The total number of species in the system reaches a peak at an early time and then declines. This peak corresponds with the maximum probability of an immigrating species being able to grow (invasibility). The inset histogram shows how many of the species that survive to the end of their respective simulations entered the system at different time points. Results presented in this and the following plots are averaged over 250 simulations each involving 500 immigration events. Late time dynamics are excluded as the decreased number of simulations that reach late times leads to a significantly increased statistical uncertainty. For these simulations a high-free energy substrate was considered (5×10^6 J mol^{-1}).

species that enter the system later are significantly more likely to survive to the end of the simulated period. This suggests that the system is continuing to develop through the entire period of the assembly process considered by our simulations.

We now examine the community-level selection for different metabolic strategies and the role that each of the three trade-offs play during the assembly process. To this end, we only consider the properties of species which have populations at least a 1000 times larger than the initial population of an invading species, in order to avoid our results being skewed by the properties of unsuccessful invaders.

4.5.3 Optimal number of reactions changes as the community assembles

We first study how the dynamics of community assembly are impacted by the trade-off between number of reactions and effective reaction rate, and how the optimal strategy with respect to this trade-off changes as the community assembles. To this end, we restrict invading species to randomly have an odd number of reactions between 1 and 7. Only considering odd numbers allows a broader range of reaction numbers to be included, while keeping the total number of different strategies to be tracked small (i.e. four strategies). Fig. 4.6A shows that species with more reactions are initially favoured, but that as the system assembles, species with a single reaction come to dominate. The initial dominance of species with multiple reactions arises due to the fact that they have a significantly decreased chance of finding no substrates that they can use when invading a system (see Fig. 4.6B). This effect is especially pronounced in the initial period of the simulation. In order to understand the eventual dominance of species with one reaction, we now consider how the average ATP per reaction event (η) varies across the different classes over time (Fig. 4.6C). The average value of η does not significantly vary across the classes initially, but as the assembly proceeds species with one reaction come to have the highest average η value. The inset to Fig. 4.6C shows a similar pattern for average substrate saturation constant (K_S), with species with one reaction having a significantly lower average value by the end of the simulation. This effect emerges as reaction parameters (η, K_S,

etc) are selected randomly for each reaction for each species. There is thus a substantially greater probability of generating a species with one reaction with highly optimal parameters than generating a species with multiple reactions all of which have highly optimal parameters. Thus, in the latter stage where more optimal species are selected for, species with one reaction come to predominate (as seen in Fig. 4.6A).

We then test for the existence of this trade-off in conditions of low free-energy availability (see Fig. 4.6D). The first thing to note is that in this case the rate of community assembly is significantly reduced. This arises as decreased availability of free-energy makes it substantially harder for species to establish themselves in the system. Two factors now act to disfavour species with more reactions in the initial stage of assembly. The first of these is that, as previously discussed, it is easier to select for species that are optimal for one reaction than for multiple reactions. This matters as the increased difficultly species experience in establishing themselves means that more optimal reactions are now required. The second factor is that when free energy is scare the cost of allocating proteome to reactions for which the substrate is not available becomes prohibitive. So, it is no longer sufficient for species to find a substrate for just one of their reactions, substrates for most (if not all) of their reactions must already be in the system. Once the initial stage has passed, species with one reaction come to predominate for the same reason as in Fig. 4.6A. Taken together our results suggest that a trade-off in the number of reactions can exist, with the optimal strategy shifting from many reactions to one reaction as the system assembles. However, this shift in optimal strategy only occurs when there is sufficient free-energy to support opportunistic (many reaction) strategies early on.

4.5.4 Outcome of translation rate vs efficiency trade-off depends on maintenance cost

The second trade-off we study is the translation rate vs efficiency trade-off. Again we consider both how this trade-off develops over time and how it varies between conditions. In Fig. 4.7A averages values of the maximum ribosome factor (ω) are plotted for cases of high and low substrate free-energy. There is an initial period where this varies quite rapidly, which then

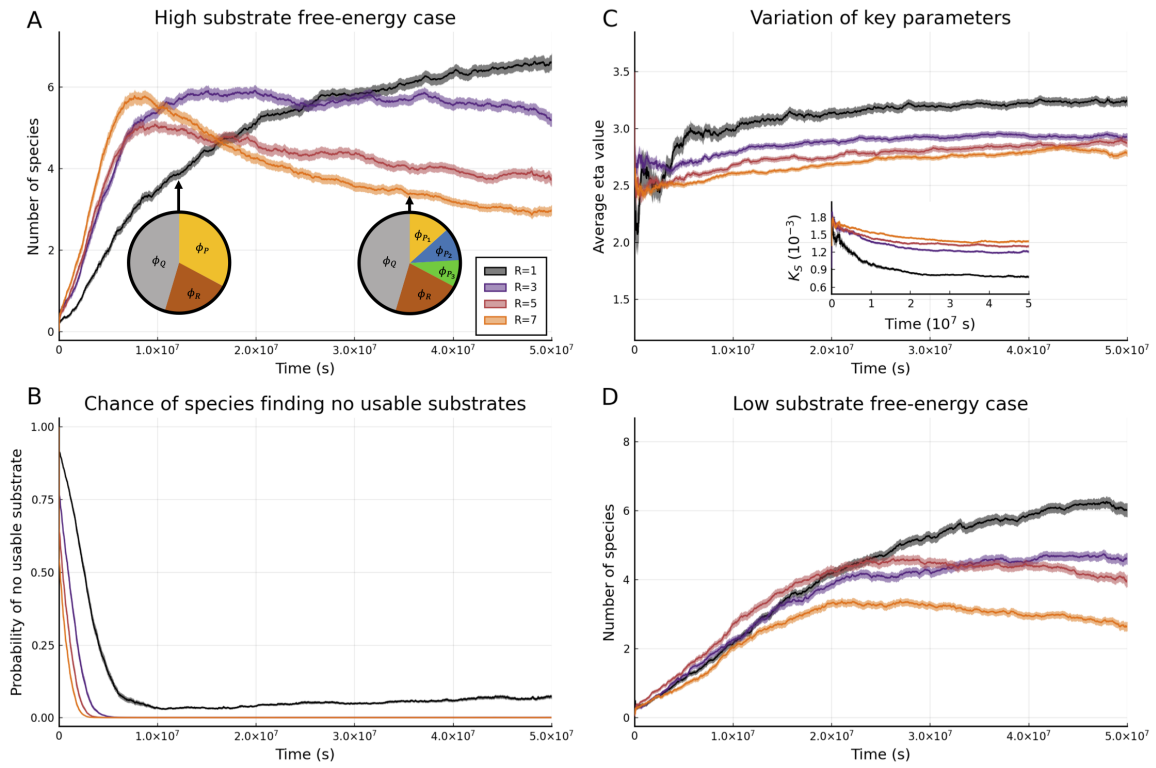


Figure 4.6: **Optimal number of reactions changes as community assembles.** All quantities plotted in this figure (and the following two figures) are averaged across 250 simulations, and have their corresponding standard errors shown. (a) Plot of the number of species in the community that have specific number of reactions, for assembly on a high free-energy substrate (5×10^6 J mol^{-1}). Initially species with many reactions dominate the community, but as the assembly continues species with only one reaction become the most common. Inset schematics of proteome allocation are included to make clear the difference between low and high reaction number strategies. (b) Corresponding plot of the probability that a species with a specific number of reactions does not find a usable substrate in the system. The probability for species with multiple reactions rapidly drops to (almost) zero. (c) The average ATP generated per reaction event (η) is initially approximately equal, regardless of the number of reactions a species possesses. As the simulation proceeds species with fewer reactions come to have higher η values. The inset shows that a comparable relationship is observed for average substrate saturation constant (K_S), though in case the averages decline rather than increase. (d) Plot of the number of species in the community that have specific number of reactions, this time for assembly on a low free-energy substrate (1.5×10^6 J mol^{-1}). In this case, no class of species dominates in the initial period, though species with one reaction come to dominate in the final stages of the simulation.

settles down in the latter stage of community assembly. Notably for the low free-energy case there is a sharp initial spike, this increased ω means species switch to a higher ribosome fraction for the same internal energy concentration (see Eqs. 4.7 & 4.8), providing an advantage in the initial stage when high growth rate is desirable. However, this a costly strategy at steady-state, particularly when free energy is scare, and so this peak declines to a stable lower value. For a similar reason the low free-energy case shows consistently higher ω values, as this allows the maintenance growth rate to be matched with a lower internal energy concentration (a), which provides a competitive advantage when free-energy is scarce. It is important to note that (as shown in the inset for Fig. 4.7A) variations in ω do not result in significant differences in the average ribosome fractions (ϕ_R) between conditions. Thus, at steady state the average efficiency is the same irrespective of free-energy supply (see Eq. 4.9).

As substrate free-energy has essentially no effect on the translation efficiency, we instead compare different maintenance costs in Fig. 4.7B. This is done by comparing low and high biomass loss rates (d), though these rates are generally taken to vary between species rather than ecosystems, they can in certain circumstances vary between systems, e.g. with changes in temperature or when a substrate has to broken down extracellularly. The high maintenance cost condition shows a higher ω value, with the difference between conditions being larger than in the previous comparison. From the inset to the figure in can further be seen that in this case the difference in ω value between conditions results in significantly different average ribosome fractions. This arises due to an increased maintenance cost necessitating a higher rate of biomass synthesis for the species to survive at steady-state. Our results suggest that a trade-off between protein translation rate and efficiency is unlikely to exist for stable steady-state communities, even if these communities are regularly invaded. Instead a clear unvarying optimal strategy exists for a given condition, which members of the steady state community all (approximately) utilise.

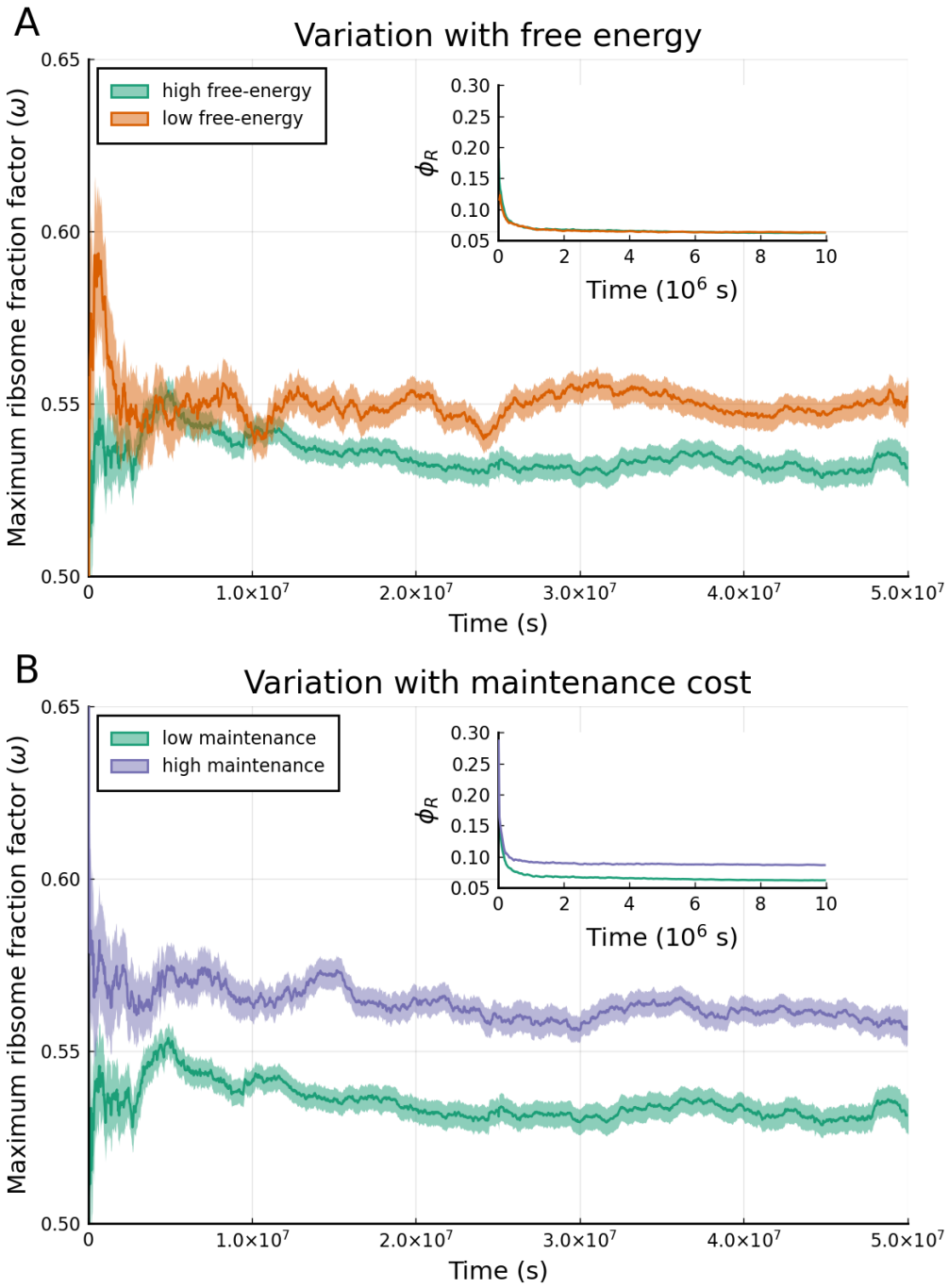


Figure 4.7: **Proteome allocation determined by maintenance cost.** (a) Plot showing how the maximum ribosome factor (ω) varies through the community assembly. Two different conditions are compared, they are high (5×10^6 J mol^{-1}) and low (1.5×10^6 J mol^{-1}) substrate free-energy. The inset plot shows that on average equal ribosome fractions (ϕ_R) are observed between conditions. (b) Identical plot with the conditions compared now being high (1×10^{-4} s^{-1}) and low (6×10^{-5} s^{-1}) biomass loss rate. A high substrate free-energy was used for both cases, and the low maintenance condition corresponds to the high free-energy condition in the previous plot. The insert plot shows that in this case the observed differences in ω result in significantly different average ϕ_R values.

4.5.5 The ATP rate vs yield trade-off drives the late stage development of the community

The final trade-off we consider is the one between ATP rate and yield. This trade-off is addressed in Fig. 4.8, both across time and between low and high free-energy substrates. When assembly occurs upon a high free-energy substrate the average ATP produced per reaction event (η) increases steadily (see Fig. 4.8A). In contrast, when assembling on a low-free energy substrate only small variations in η are observed. The average reaction efficiency reached varies between conditions, but in the opposite manner, with final efficiency being higher upon low free-energy substrates (see Fig. 4.8B). This increased efficiency is necessitated by the decreased availability of free energy. In the high free-energy case, reactions are generally far-from-equilibrium, meaning that increased efficiency does not significantly increase the strength of the thermodynamic end-product inhibition. Thus, in this case efficiency is seen to increase throughout the simulated period. In contrast, the low free-energy case sees a decline in the average efficiency. To understand this effect, Fig. 4.8C considers the efficiency of reactions that descend one step in the metabolite hierarchy (see Fig. 3.1B) separately from those that descend two. The efficiency of reactions is higher when assembling on low free-energy substrates, regardless of whether one or two-step reactions are considered. Additionally, only in the low-free energy case is the difference between the efficiency of one and two-step reactions substantial. This arises because for one-step reactions using low-free substrates free-energy is particularly scarce, so only very efficient species can obtain enough free energy to grow.

It is also important to note that in Fig. 4.8C the efficiency of all reaction types increases with time (a few fluctuations aside). Therefore, in order to explain the decreasing efficiency observed in Fig. 4.8B, the relative frequency of the different types of reactions must be considered (see Fig. 4.8D). In both cases the proportion of reactions that are two-step reactions increases as the community assembles. Making use of two-step reactions increases the free-energy that species can obtain from a single reaction, which in the low-free energy case allows species to support themselves using less efficient reactions. This explains the observed decrease in the reaction efficiency, as species that make use of less efficient but more energetically yielding reactions

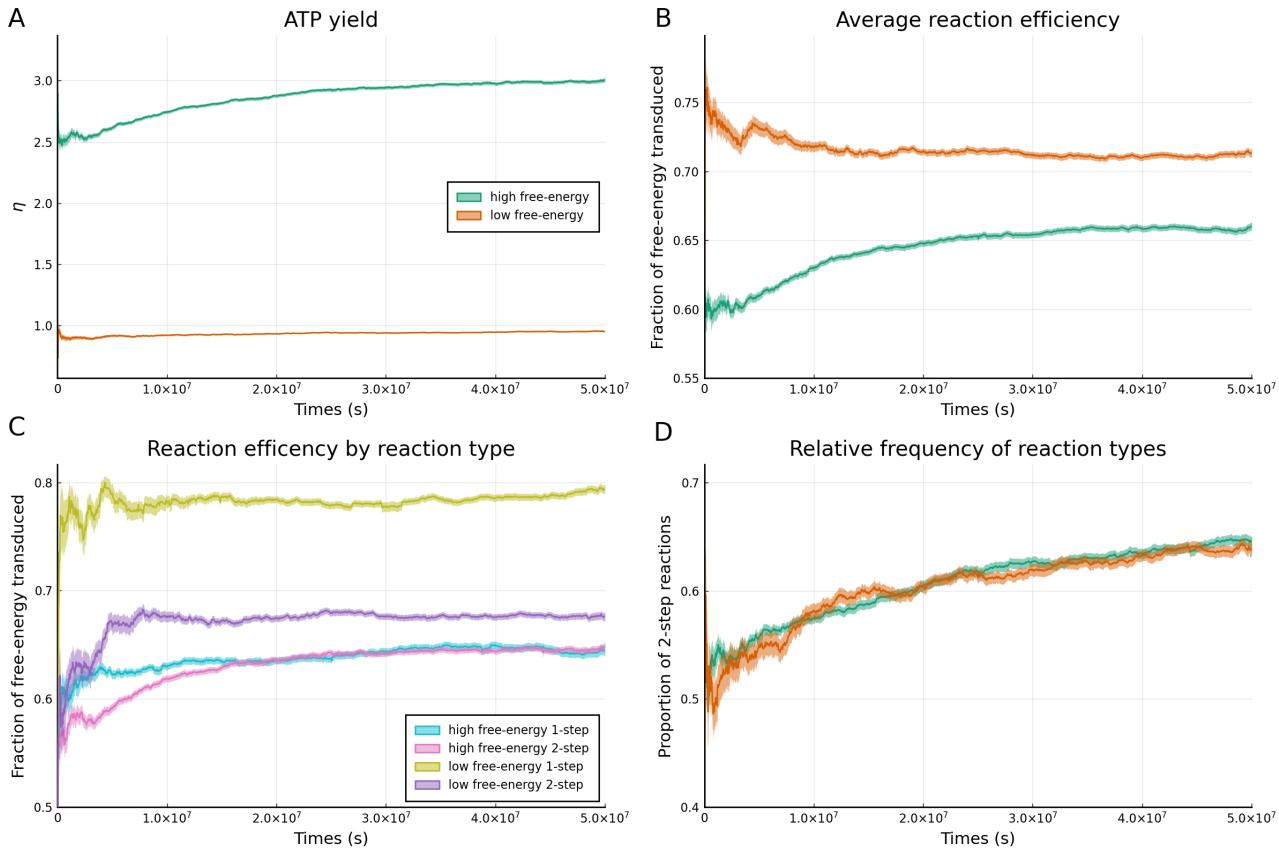


Figure 4.8: **Communities develops towards thermodynamic optima.** (a) Average ATP yield with time for community assembly upon high and low free-energy substrates. (b) Average reaction efficiency with time for the same conditions. (c) Average reaction efficiency for both conditions, but now subdivided into reactions that take one or two steps down the metabolite hierarchy. (d) Average proportion of reactions that take two rather than one steps down the metabolite hierarchy.

successfully invade the community. The shift towards an increased proportion of two-step reactions is still occurring at the end of the simulated period, suggesting that the community is yet to reach an optimal thermodynamic configuration. This suggests that the trade-off between ATP rate and yield drives the continuing development of the community, which is developing towards but has not yet achieved a community level trade-off optimum.

4.6 Discussion

We have shown that in microbial communities assembling continuously (through repeated immigration events) on diversifying substrates, species diversity eventually rises to be comparable to the number of possible substrates (Fig. 4.5). Further, we found that a trade-off between single and many reaction strategies emerges from fundamental proteome constraints (Fig. 4.6). The optimum of this trade-off shifts as the system develops, provided that sufficient free-energy is available to render opportunistic (many reaction) strategies successful at the early stage in the assembly. We also find that a meaningful trade-off between ribosome translation rate and translational efficiency only exists when comparing communities with different cellular maintenance rates (Fig. 4.7). Finally, we find the trade-off between ATP rate and yield drives community development, which continually progresses towards a state with every member satisfying the narrow trade-off optimum (Fig. 4.8).

The final species diversities that we observe in Fig. 4.5 are far higher than those observed in the previous chapter (Chapter 3), suggesting that thermodynamic constraints on diversity are inextricably bound up with the specific process by which the community assembles. Upon high free-energy substrates, the trade-off we observe in the number of reactions species possess (Fig. 4.6A) is consistent with the general predictions of r/K selection theory [FLR$^+$12], where initially fast growing opportunists colonise a community but as the system develops these species are displaced by more efficient specialists. Conversely, upon low free-energy substrates the trade-off we observe (Fig. 4.6B) is consistent with empirical work which has found that specialist species are the first to colonise substrate particles suspended in seawater, with a more diverse

community succeeding these initial specialists [DSG+16, EDS+19]. We observed no ribosome translation rate vs efficiency trade-off, only a difference based on varying maintenance costs (Fig. 4.7). This suggests that further study on this trade-off needs to pay greater attention to the causes of variation of maintenance rate between species. Factors that can effect this include the production of extracellular proteins (notably enzymes), temperature [PS04], energy availability [KvBW+17], and cell mass [DOM+10, KvBW+17].

The form of the ATP rate vs yield trade-off that we found in Fig. 4.2C, a linear increase followed by a sharp drop, differs substantially from the form put forward by historic work on this trade-off [PB02]. Here, ATP production rate was taken to follow a parabolic form with ATP yield, peaking at an intermediate value. This form was obtained by making the assumption that reaction rate scales linearly with the rate of free-energy dissipation, a valid assumption near thermodynamic equilibrium. However, this assumption cannot be universally made, as metabolic reactions frequently operate far from equilibrium. There may very well exist a general relationship between free-energy dissipation and reaction rate that holds arbitrarily far-from equilibrium, but to the best of our knowledge no such relation has been obtained either theoretically or empirically. Therefore, rather than positing an unsupported relationship, our trade-off was based solely upon the well understood mechanism of thermodynamic end-product inhibition [HCR96, LWvB+19]. Future developments in the field of stochastic thermodynamics could however result in a general relationship between free-energy dissipation and reaction rate, in which case a stricter ATP rate yield trade-off could be implemented, i.e. one were the kinetic variables (k, K_S, r) are constrained by the free-energy dissipation.

Recent improvements in the quality of data on microbial communities has cast doubt upon some of the conventionally invoked trade-offs. For example, the growth and metabolic rates of prokaryotic microbes are generally taken to scale with genome size, which was itself thought to scale with cell size [DGL+17]. However, a recent large scale data-synthesis has found that maximum growth rate, cell size and genome size are near-independent properties of microbes [WNG+21]. Additionally, recent work has suggested that commonly invoked rate-yield trade-offs are best thought of as being two-dimensional, accounting for different parameters that microbes would be selected to optimise (in this case growth rate and substrate uptake rate)

[COM+19]. Our work underlines the importance of considering microbial trade-offs that are linked to specific mechanisms, rather than the generic trade-offs (e.g. specialist vs generalist) that are often proposed in the literature.

In this study, we made use of a "bottom-up", continuous assembly procedure. In this mode of assembly, the initial community contains few species and the assembly process is driven by the continual immigration of new species. In the previous chapter, we considered a "top-down" assembly procedure, where an initially diverse community was implanted into an environment and a feasible, less diverse community emerged by non-random species extinctions. These two modes of assembly have been found to return equivalent results provided that invasions are rare (relative to the time required for the system to reach steady state), invaders enter the system at low abundance, and the community possesses an equilibrium which is both feasible and stable [SA21]. These assumptions often hold for macroscopic systems. However, for microbial systems the assumption of rare immigration events only holds for systems that are comparatively isolated (e.g. the human gut). Therefore, in order to capture the assembly dynamics of non-isolated communities, we considered a scenario where the rate of species immigration was relatively high (i.e. mean immigration time of 1×10^5 s). This means that the results arising from our work are fundamentally linked to continuous community assembly, and could not be recovered from a model without continuous immigration of new species.

In summary, in study has demonstrated how trade-offs can emerge from thermodynamic and proteomic constraints in the context of a complex community. It has further demonstrated that these emergent trade-offs fundamentally determine the development of continuously assembling microbial communities. This underscores the importance of considering the role non-equilibrium thermodynamics plays in the assembly and development of microbial communities.

Chapter 5

Synthesis

Thermodynamics has historically played a key role in the development of theoretical ecology. This ranges from early developments in the science made by Lotka [Lot25, Kin15], through the development of the ecosystem concept by Odum [OP55, Odu57], to more recent developments such as the metabolic theory of ecology [BGA+04]. Whilst the areas ecological theory based upon equilibrium thermodynamics (e.g. trophic levels, metabolic theory, etc) tend to to be solidly grounded theoretically, those which draw upon non-equilibrium thermodynamics generally lack any robust basis in physical theory. Though the success of a biological theory does not necessarily hinge upon its physical foundations, if biological insights are to be gained from taking a physical perspective the theories considered must be physically well founded. In order to address this, I proposed three clear objectives for this thesis. They were as follows:

1. To critically assess the viability of (thermodynamic) extremal principles in ecological modelling.

2. To develop a novel modelling framework to understand the effect of non-equilibrium thermodynamic constraints on microbial communities.

3. To utilise this framework to intervene into broader debates in microbial ecology.

5.1 Viability of extremal principles in ecological modelling

In **Chapter 2**, a new extremal principle relevant to switching between states in a multistable system was obtained (see Eq. 2.37). This extremal principles suggests that the entropy produced by a switching trajectory will (on average) be maximised, subject to the simultaneous minimisation of the classical action. This dependence on the classical action means that our theory cannot describe the behaviour of the system solely in terms of the entropy production. In order to address this, we derived an approximate switching relation (Eq. 2.45), where the relative switching between the two states is governed by the difference in entropy produced along the (two) switching paths between the states. We successfully verified the prediction of this switching relation for two simple bistable models the Schlögl model and the toggle switch model (see Fig. 2.7D). It is notable that the relation we obtained here was relevant for switches between states, rather than pertaining to steady-state thermodynamic properties as is generally the case in previous work on extremal principles.

The principle we derived possesses a number of limitations. The first major limitation is that both the extremal principle and the switching relation consider an apparent entropy production. This is the entropy production of a coarse-grained system, which is never larger than the entropy production of the fine-grained system underlying it [MLB$^+$12]. Though we tried to address this by deriving approximate expressions for the underlying entropy production and flows (see Fig. 2.5), both our extremal principle and our switching relation depended on this apparent entropy production. It is worth noting that the question of entropy production coarse-graining remains an active one in stochastic thermodynamics [CSP21], and so a solution to this limitation may emerge with time. The second major limitation arises from the difficulty in establishing the possible states of a system. This is also a problem for previous extremal principles, which it is often hard to establish the informativeness of, as the question of whether alternative configurations could have been adopted gets left unasked. We sidestepped this difficulty with our simple models, by finding analytic solutions for the steady-states and using

a numerical method to find the most probable switching paths between those states. However, as systems become increasingly complex (i.e. as they approach the real-ecosystem case) analytic solutions become significantly harder to achieve and numerical methods become increasingly computationally demanding. This presents significant difficulties for the verification of our switching relation for realistic ecosystems limiting the potential application.

We could have attempted to use the model developed in **Chapter 3** to test either our extremal principle or the conventional maximum entropy production principle. However, we would run into the issue of triviality. That is, we would generally observe increases in the rate of entropy production as microbial communities assemble, but it is not clear that the model offers alternative paths for community assembly that do not increase entropy production. This makes it essentially impossible to determine of whether an observed increase in entropy production is due to an underlying physical principle, or merely the structure of the model. When continuous species immigration is included in the model (see **Chapter 4**) the variation of entropy production throughout community assembly becomes more complex, with it increasing as new substrates become available but (slightly) decreasing as new more efficient species establish themselves in the community. However, despite this, alternative community assembly paths still do not appear to exist, so the issue of triviality remains. This presents a further significant difficulty in testing our switching relation (or other extremal principles), that of setting up a complex system where a non-trivial choice between different development paths exists.

From all this, we conclude that at present the applicability of extremal principles to ecology is limited. Either due to principles being based on weak theoretical foundations, or by better grounded principles being hard to scale beyond simple systems. There may exist cases where this kind of principle could usefully be applied, e.g. to measures of fitness flux in low dimensional systems [ML10], but application to the complex systems generally used to model ecosystems is not feasible. With that said, in this work we successfully demonstrated the potential of applying stochastic thermodynamics to the study of possible extremal principles. It is important to note, that stochastic thermodynamics is significantly growing field of study. Therefore, though immediate progress seems unlikely, there is potential that stochastic thermodynamics will be able to further advance the study of extremal principles at some point in the future. However,

the unlikelihood of immediate progress motivated the rest of this thesis's focus on detailed thermodynamics interactions in microbial community assembly.

5.2 Development of a novel thermodynamic microbial community model

In **Chapter 3**, we developed a novel thermodynamic microbial community model. This model was developed as an extension of a microbial consumer-resource model with metabolite production originally proposed in [MCG+19]. The novelty of this model consisted of it incorporating both thermodynamic end-product inhibition (incorporated before in simpler models e.g. [HCR96, GS16]) and proteomic constraints (incorporated in another recent community model [PMSMG21]). Including all this in our framework allows consideration of how non-equilibrium thermodynamics impacts realistic (many member) microbial communities, as well as addressing the impact of other reasonable constraints (specifically proteomic) on microbial growth. Models incorporating far more detailed thermodynamic and proteomic constraints already exist in the form whole-genome community-scale metabolic models [MATP18]. What our modelling framework adds is consideration of the intermediate scale [LWMX20], being complex enough to capture thermodynamic effects but not so complex that it is unfeasible to simulate diverse communities.

In **Chapter 4**, this model is extended to capture the variation of maximum ribosome fractions between species [RSS16]. In order to avoid high ribosome fraction strategies being universally favoured, the translation efficiency of a species' ribosomes is now taken to vary with the fraction of the proteome allocated to ribosomes. This was motivated by recent empirical work [RSS16], that found an increase in maximum growth rate with the number of ribosomal RNA operons across species, coupled with a decrease in carbon use efficiency. Making this extension allowed our model to encode multiple viable protoeme allocation strategies, in contrast to the unaltered model where all species adopt the same ribosome fraction ϕ_R at steady-state (Fig. 3.4B). This extended model is thus better suited to study fundamental proteomic (and thermodynamic)

trade-offs, as it allows for the coexistence of differing proteome allocation strategies.

There are a couple of natural ways in which our modelling framework could be further expanded. The first of these is altering the model such that the proportional expression of different reactions (see Eq. 3.9) can alter over time. This could consist of an extension to the optimisation procedure developed in [PMSMG21], where species seek to maximise the return (in terms of total substrate uptake) on their protein investment. This would enable our model to capture the shifting between reactions commonly observed for real microbes [Mon49]. The second alteration that could be made to our model is extending the thermodynamic consideration to include anabolic (free-energy consuming) processes rather than merely catabolic (free-energy yielding) processes. In this extended model, the (anabolic) synthesis of certain precursors would become necessary for biomass growth, rather than energy being the only requirement. This would potentially allow thermodynamic bottlenecks to develop, that is reactions that are very close to equilibrium that are a necessary for biomass synthesis. These bottlenecks frequently occur in real-world microbial metabolisms, and when they occur necessitate a large allocation of protein to catalyse the specific bottleneck reaction [NFBE$^+$16]. Thus, including anabolism is important to fully understand the interlinked thermodynamic and proteomic constraints that microbial community members are subject to.

5.3 Intervention into broader debates in microbial ecology

In **Chapter 3**, we consider the impact of free-energy supply on final community diversity. We find that species diversity is higher for communities that grow upon high free-energy substrates. This is driven by the increased rate of growth that greater free-energy availability facilitates. Increased growth rate leads to an increased rate of ecosystem development, meaning that more substrates become available before members of the initially inoculated community begin going extinct. This increased number of niches allows for a greater final diversity. There has been previous theoretical work has addressing the same question, similarly finding that final diversity

increases along with the supplied energy [MCG+19]. In this case, however, the mechanism is far simpler, namely that a greater energy supply leads to more of the byproducts of the initial substrate being sufficiently available to support further species. We believe our result to be far more developed, as it has a clear mechanism and thus timescale, allowing links with other processes to be made. The major drawback of our result is that it pertains to a specific ecosystem assembly procedure, and thus cannot be considered a general result.

Beyond this in **Chapter 4**, we considered the issue of trade-offs, and how they emerge from fundamental thermodynamic (and proteomic) constraints. This consideration focused on the emergent impact of these trade-offs in the context of a community that is, both, complex (many species) and continuously assembling. Our consideration differs from previous approaches in microbial ecology, most of which considered phenomenological rather than mechanistic trade-offs. Where mechanistic trade-offs have been considered previously, this has generally only been for single populations, either at steady-state or exponentially growing [FNBE+13, BHO+15, RSS16]. The importance of understanding community assembly to the study of trade-offs was underlined by the fact that the optimal strategy to satisfy the reaction number vs effective rate trade-off shifted as the assembly proceeded (Fig. 4.6), with many-reaction strategies found to be favoured early on, but at the later stages the single reaction strategy came to dominate. Conversely, understanding trade-offs was also found to be necessary to understanding the process of assembly (Fig. 4.8), with successful invaders generally possessing more optimal strategies for the ATP rate vs yield trade-off. We also found interplay between the trade-offs, with the fact that thermodynamically favourable strategies were more often generated for one-reaction strategists partly driving their selection (Fig. 4.6C).

Our results could be usefully extended by incorporating further trade-offs into our consideration. An obvious candidate is the widely proposed trade-off between genome size and biomass maintenance cost [DGL+17]. Recent work has cast doubt upon the experimental evidence for this trade-off [WNG+21]. It could, therefore, be a useful intervention to consider the costs associated with maintaining a gene vs the costs of maintaining protein copies (following [LM15]), and explicitly incorporating a mechanistic formulation of this trade-off in a model. We could also extend our consideration to communities formed by coalescence rather than assembly, that is

by the combination of previously existing communities. In this context the importance of competitive and cooperative interactions has become a major area of study [PGBB20, LACC$^+$21]. It would be interesting to consider the impact of the thermodynamic interaction types we proposed in Chapter 3 on the success of communities (and community members) in surviving coalescence events.

5.4 Conclusion

In this thesis we have demonstrated the continuing value of non-equilibrium thermodynamics to theoretical ecology. Though physically well grounded extremal principles do not seem likely to be obtained in the near future, non-equilibrium thermodynamics proves useful in other ways. Primarily, through the incorporation of more detailed non-equilibrium thermodynamic mechanisms into ecological models. This is of particular value to the study of microbial ecosystems due to the wealth of detailed metabolic data available. Thus, the rapid experimental advances being made in understanding microbial ecosystems offers the clear potential to broaden our understanding of the non-equilibrium thermodynamic basis of all ecosystems.

Appendix A

Parameter values used in figures and simulations

A.1 Parameters for Figure 2.1

$k_{+1}a$	k_{-1}	k_{+2}	k_{-2}	b	Ω
0.5	3.0	1.0	1.0	4.0	10

Table A.1: Parameter table for Schlögl model simulation shown in Fig. 2.1A.

k	k_-	K	K_-	q	q_-	Q	Q_-	r	f	Ω
20.315	0.19687	7.7926	0.70940	12.294	0.37021	4.6696	0.32111	1516.2	7513.7	80

Table A.2: Parameter table for toggle switch model simulation shown in Fig. 2.1B.

A.2 Parameters for Figure 2.6

$k_{+1}a$	k_{-1}	k_{+2}	$k_{-2}b$
0.21856	6.4414	1.7723	7.7299

Table A.3: Parameter table for Schlögl model paths shown in Fig. 2.6.

k	k_-	K	K_-	q	q_-	Q	Q_-	r	f
26.098	2.1794	8.2401	0.60048	25.772	1.4149	6.9638	0.16574	3875.6	9285.8

Table A.4: Parameter table for toggle switch model paths shown in Fig. 2.6.

A.3 Parameters used for simulations in Chapter 3

	Symbol	Value	Units	Source
ATP free energy	ΔG_{ATP}	75.0	kJ mol^{-1}	[BT13]
Temperature	T	293.15	K	*
Cell mass	m	1×10^8	# amino acids	[BD96]
Maximum elongation rate	γ_m	1260.0	# steps min^{-1}	[BD96]
Average ribosome mass	n_R	7459	# amino acids	[KCVSZ$^+$11]
Average metabolic protein mass	n_P	300	# amino acids	[BEO$^+$09]
Rate of biomass loss	d	6×10^{-5}	s^{-1}	*
Elongation step cost	χ	29	# ATP	[LM15]
Elongation half saturation constant	$\gamma_{\frac{1}{2}}$	5×10^8	# ATP cell^{-1}	*
Average fraction of ribosomes bound	f_b	0.7	N/A	[USP05]
Housekeeping proteome fraction	ϕ_Q	0.45	N/A	[SGM$^+$10]
Ribosome fraction half saturation constant	$\Omega_{\frac{1}{2}}$	1×10^9	# ATP cell^{-1}	*
Metabolite dilution rate	ρ	6×10^{-5}	s^{-1}	[PTW17]
Metabolite supply rate	κ	3.3×10^{-7}	M s^{-1}	*
Initial population	N_0	1.0	cell	*
Initial metabolite concentration	C_0	0.0	M	*
Initial intracellular energy concentration	a_0	1×10^5	# ATP cell^{-1}	*
Initial ribosome fraction	$\phi_{R,0}$	12.8	%	*

Table A.5: Table of fixed parameters for the model developed in Chapter 3, kept constant across all simulated conditions. Parameters not taken from the literature are marked with an asterisk.

	Symbol	Distribution	μ	σ	Magnitude	Units
Maximum forward reaction rate	k	Log-normal	0	$1/\ln 2$	10.0	s^{-1}
Substrate half saturation constant	K_S	Log-normal	0	$1/\ln 2$	1.375×10^{-3}	M
Reversibility factor	r	Log-normal	0	$1/\ln 2$	10.0	N/A

Table A.6: Table of reaction kinetic parameters for the model developed in Chapter 3. These are drawn randomly for each reaction for each species, so the parameters of the random distributions used are given. The random numbers generated from these distributions are then multiplied by their respective magnitude in order to obtain final parameter values.

Appendix B

Further details of models used in Chapter 2

B.1 Concentration-constrained Schlögl model

In the Schlögl model [Sch72] a chemical species A is converted to chemical species B via an intermediary species X. These chemical dynamics are described by the following equations

$$A \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} X \tag{B.1a}$$

$$3X \underset{k_{-2}}{\overset{k_{+2}}{\rightleftharpoons}} 2X + B. \tag{B.1b}$$

The deterministic dynamics of the copy number x in the external species constrained Schlögl model are therefore described by

$$\frac{dx}{dt} = k_{-2}bx^2 - k_{+2}x^3 - k_{-1}x + k_{+1}a, \tag{B.2}$$

where a and b are externally clamped copy numbers and $(k_{\pm 1}, k_{\pm 2})$ are rate constants. Under certain parametrizations Eq. B.2 will have 3 real positive solutions and thus the system will display bistability [End15, Gas04, VQ09]. The external clamping of the copy numbers makes

these states non-equilibrium steady states rather than equilibria.

B.2 Flux-constrained Schlögl model

When moving from the concentration-constrained to the flux-constrained model, a system of three deterministic equations needs to be considered [End17]. The equations for describing the rate of change of real-valued species numbers a, b, and x are given by

$$\frac{da}{dt} = k_{-1}x - k_{+1}a + F \tag{B.3}$$

$$\frac{db}{dt} = k_{+2}x^3 - k_{-2}bx^2 - F \tag{B.4}$$

$$\frac{dx}{dt} = k_{-2}bx^2 - k_{+2}x^3 - k_{-1}x + k_{+1}a, \tag{B.5}$$

where F is the flux constraint on the system. In order to find the steady states of the system we must look for solutions to $\frac{da}{dt} = \frac{db}{dt} = \frac{dx}{dt} = 0$. It can be seen by addition of Eqs. B.3 and B.4 that $\frac{da}{dt} + \frac{db}{dt} = -\frac{dx}{dt}$. This means that $\frac{da}{dt} = \frac{db}{dt} = 0$ ensures $\frac{dx}{dt} = 0$, we therefore cannot extract any additional information about the steady state from Eq. B.5.

The equations we are thus interested in solving are

$$k_{-1}x - k_{+1}a + F = 0 \tag{B.6}$$

$$k_{+2}x^3 - k_{-2}bx^2 - F = 0. \tag{B.7}$$

There are three unknowns in this pair of equations (a, b, and x), therefore a constraint must be introduced in order to find the solution to this system. We make the assumption that as every process conserves the number of molecules in the system that total molecule number is

conserved, i.e

$$a + b + x = n = const. \tag{B.8}$$

With this constraint we can now solve the system for a particular total molecular species number n. Eq. B.6 can be rearranged to,

$$a = \frac{k_{-1}x + F}{k_{+1}}, \tag{B.9}$$

so that Eq. B.8 becomes

$$b = n - x - a = n - x - \frac{k_{-1}x + F}{k_{+1}}. \tag{B.10}$$

Substituting for b in Eq. B.7 and gathering terms leads to

$$\left(k_{+2} + k_{-2} + \frac{k_{-2}k_{-1}}{k_{+1}}\right)x^3 + \left(\frac{k_{-2}F}{k_{+1}} - Nk_{-2}\right)x^2 - F = 0. \tag{B.11}$$

There is only one real positive solution to this equation regardless of parametrization. This can be demonstrated by considering a cubic equation with three real positive solutions x_1, x_2 and x_2.

$$(x - x_1)(x - x_2)(x - x_3) = 0 \tag{B.12}$$

which can be expanded to,

$$x^3 - (x_1 + x_2 + x_3)x^2 + (x_1x_2 + x_2x_3 + x_3x_1)x - x_1x_2x_3 = 0. \tag{B.13}$$

The lack of a first order term in x in Eq. B.11 thus implies $(x_1x_2 + x_2x_3 + x_3x_1) = 0$. This cannot be realized with x_1, x_2 and x_3 being distinct, real and non-negative. Hence under the constraint of conservation of total molecular concentration the flux constrained Schlögl model will have at most one physically sensible (real, +ve) steady state.

If the assumption of Eq. B.8 is not made, then multi-stability is preserved. However, this is not an interesting multi-stability in that if one of (x, a, b) is provided the other two can be determined uniquely by Eqs. B.6 and B.7. The apparent multi-stability here is merely the fact

that there is a wide range of total concentrations n for which the system possesses a stable state.

B.3 Concentration-constrained toggle switch

The chemical equations below describe the dynamics of a simple bistable switch [CA00], with promoters α and β allowing the synthesis of "proteins" A and B. Protein dimers can then repress the promoter of the opposite type:

$$\alpha \underset{k_-}{\overset{k}{\rightleftharpoons}} \alpha + A \tag{B.14a}$$

$$A \underset{K_-}{\overset{K}{\rightleftharpoons}} \varnothing \tag{B.14b}$$

$$2B + \alpha \underset{r}{\overset{f}{\rightleftharpoons}} \varnothing \tag{B.14c}$$

$$\beta \underset{q_-}{\overset{q}{\rightleftharpoons}} \beta + B \tag{B.14d}$$

$$B \underset{Q_-}{\overset{Q}{\rightleftharpoons}} \varnothing \tag{B.14e}$$

$$2A + \beta \underset{r}{\overset{f}{\rightleftharpoons}} \varnothing. \tag{B.14f}$$

These equations contain reverse reactions for thermodynamic consistency. The dynamics can be summarized as a system of deterministic differential equations,

$$\frac{da}{dt} = k\alpha - k_- a - Ka + K_- - 2\beta f a^2 + 2(1-\beta)r \tag{B.15}$$

$$\frac{db}{dt} = q\beta - q_- b - Qb + Q_- - 2\alpha f b^2 + 2(1-\alpha)r \tag{B.16}$$

$$\frac{d\alpha}{dt} = (1-\alpha)r - \alpha f b^2 \tag{B.17}$$

$$\frac{d\beta}{dt} = (1-\beta)r - \beta f a^2. \tag{B.18}$$

We consider the fast switching limit $r, fa^2, fb^2 >> k, q, K, Q$, so that promoters α and β can be treated as continuous variables that run between 0 and 1. This separation of time scales means that Eqs. B.17 and B.18 can be solved for the average promoter values, which can then be inserted into Eqs. B.15 and B.16 (quasi-steady state approximation). The average promoter values are given by

$$\alpha^* = \frac{r}{r + fb^2} \tag{B.19}$$

and

$$\beta^* = \frac{r}{r + fa^2}. \tag{B.20}$$

Substitution then gives,

$$\frac{da}{dt} = \frac{kr}{r + fb^2} - k_- a - Ka + K_- \tag{B.21}$$

$$\frac{db}{dt} = \frac{qr}{r + fa^2} - q_- b - Qb + Q_- \tag{B.22}$$

These two equations for the approximate dynamics can be solved to find the multiple stationary points of the system. These stable states are non-equilibrium steady states as the spontaneous generation and degradation of chemical copies ensures a constant flux through the system, preventing equilibration. These spontaneous reactions can be thought of as equivalent to reactions with concentration-clamped chemical reservoirs.

B.4 Flux-constrained toggle switch

In order to apply a flux constraint on our system, we augment Eq. B.14 by two additional chemical species, substrate S and waste W

$$a + S \underset{k_-}{\overset{k}{\rightleftharpoons}} a + A \tag{B.23a}$$

$$A \underset{K_-}{\overset{K}{\rightleftharpoons}} W \tag{B.23b}$$

$$2B + a \underset{r}{\overset{f}{\rightleftharpoons}} \varnothing \tag{B.23c}$$

$$b + S \underset{q_-}{\overset{q}{\rightleftharpoons}} b + B \tag{B.23d}$$

$$B \underset{Q_-}{\overset{Q}{\rightleftharpoons}} W \tag{B.23e}$$

$$2A + b \underset{r}{\overset{f}{\rightleftharpoons}} \varnothing \tag{B.23f}$$

$$\varnothing \xrightarrow{F} S \tag{B.23g}$$

$$W \xrightarrow{F} \varnothing. \tag{B.23h}$$

Making the same assumption of time scale separation as above, a new set of deterministic differential equations can be obtained as

$$\frac{da}{dt} = \frac{ksr}{r + fb^2} - k_-a - Ka + K_-w \tag{B.24a}$$

$$\frac{db}{dt} = \frac{qsr}{r + fa^2} - q_-b - Qb + Q_-w \tag{B.24b}$$

$$\frac{ds}{dt} = -\frac{ksr}{r + fb^2} - \frac{qsr}{r + fa^2} + k_-a + q_-b + F \tag{B.24c}$$

$$\frac{dw}{dt} = Ka + Qb - (Q_- + K_-)w - F. \tag{B.24d}$$

Once again as in the flux constrained Schlögl model this system of equations is not fully independent, e.g. $\frac{da}{dt} + \frac{db}{dt} = -\frac{ds}{dt} - \frac{dw}{dt}$. At steady state, finding a solution to $\frac{da}{dt} = \frac{db}{dt} = \frac{ds}{dt} = 0$ guarantees that $\frac{dw}{dt} = 0$. To find steady states an additional constraint must be introduced,

$$a + b + s + w = n = const. \tag{B.25}$$

Here, meaningful bistability can be obtained in contrast to the flux constrained Schlögl model. This can be explained through the fact that generally there are three mechanisms through which multi-stability can be maintained:

1. Net flux through the system can be changed.
2. Number of molecules in the system could change, leading to different reaction affinities for the same fluxes.
3. The way the flux is partitioned between reaction pathways could be changed.

In the flux-constrained Schlögl model there exists only one pathway through which flux can pass. Additionally, the assumption is made of conservation of particle number. This means that none of the three mechanisms are available to generate multi-stability. For the flux-constrained toggle switch model two distinct pathways for flux exist so multi-stability can be achieved by changing how flux is partitioned (Fig. B.1). The continual flux of chemical copies through this model ensures that the multiple state states found here are non-equilibrium steady states. This model is not used in Chapter 2 as calculation of minimum action paths became too computationally intensive due to the high dimensionality.

B.5 f_1 and f_2 values for the models

In order for approximate entropy production and flow terms to be calculated (see Section 2.4), the force has to be divided into two parts f_1 and f_2. The definition of these two components of the force are provided below for both the Schlögl and toggle switch models.

B.5.1 Schlögl model

$$\begin{aligned} f_1 &= k_{+1}a - k_{-1}x \\ f_2 &= k_{+2}x^3 - k_{-2}bx^2 \end{aligned} \tag{B.26}$$

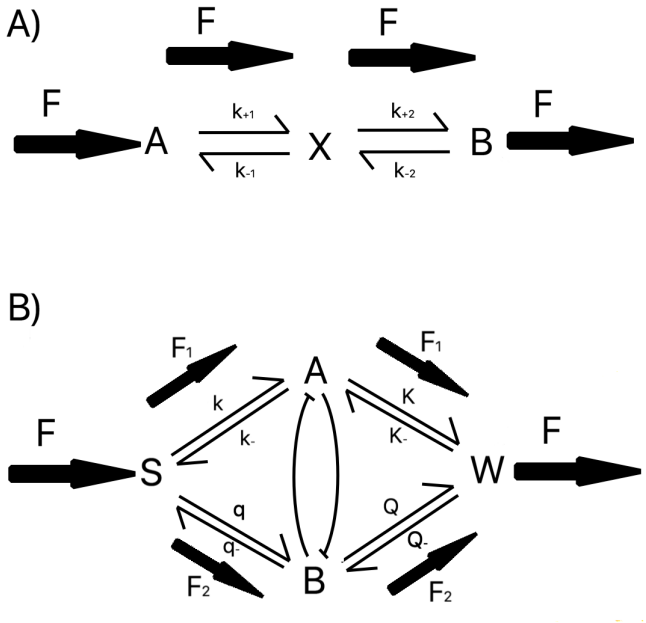


Figure B.1: **Schematic of flux constrained models.** In the Schlögl model (A), there is no ability to partition flux, so if net flux F and total number $n = a + x + b$ are constrained, then there exists on only one stable configuration for a given F and n. In the toggle switch model (B), the flux F can be partitioned down two separate pathways such that $F = F_1 + F_2$. This partitioning means that multiple stable states can exist for a given F and n.

B.5.2 Toggle switch model

$$f_1 = \begin{pmatrix} \frac{kr}{r+fb^2} - k_- a \\ \frac{qr}{r+fa^2} - q_- b \end{pmatrix}$$
$$f_2 = \begin{pmatrix} Ka - K_- \\ Qb - Q_- \end{pmatrix} \tag{B.27}$$

Appendix C

Algorithms to calculate minimum action paths

C.1 Minimum action path

Consideration of the models used in Chapter 2 via deterministic approximation can give information about the steady state behavior but will provide only limited information about state switching. In order to investigate state stability stochastic behavior must therefore be considered. In the large volume (small noise) limit the stochastic evolution of the system can be considered via the over-damped chemical Langevin equation [PCGBP16]

$$\dot{x_i} = f_i(\boldsymbol{x}) + g_{ia}(\boldsymbol{x})\xi_a(t). \tag{C.1}$$

In this equation $\dot{x_i}$ is the rate of change of species i, f_i is the net deterministic force on species i, Ω is the volume of the system, ξ_a is the noise from species a, and g_{ia} determines the propagation of the noise from species a to species i. Here and throughout this appendix we adopt the convention that repeated indices are summed over. This is a continuous stochastic approximation to the underlying master equation, thus allowing a greater range of analytic techniques.

One such technique is the minimization of the Freidlin–Wentzell action [VF70] in order to determine the most probable path between two points in parameter space. According to the large deviations theory this path will be exponentially more probable than paths substantially deviating from it [Tou09]. This exponential fall off means that the mean first-passage time for state switching T_s can be found in an exponential form as

$$T_s = Qe^{\Omega\mathcal{A}}, \tag{C.2}$$

where, due to $\ln T_s = \ln Q + \Omega\mathcal{A}$, T_s is determined by $\Omega\mathcal{A}$ to logarithmic precision for large Ω. Here, $\mathcal{A}$ is the action associated with the minimum action path (MAP) for this stochastic transition and Q is a pre-factor to the exponential behavior. Through this formula the mean rate of transition between two points in parameter space can be found via minimization of the Freidlin–Wentzell action which is given as

$$\mathcal{A} = \frac{1}{2}\int_0^t d\tilde{t}(\dot{x}_i - f_i)D_{ij}^{-1}(\dot{x}_j - f_j) = \int_0^t d\tilde{t}L(\dot{\mathbf{x}}, \mathbf{x}), \tag{C.3}$$

where

$$D_{ij}^{-1} \equiv (g_{ia}^{-1})(g_{ja}^{-1}) \tag{C.4}$$

is the diffusion matrix and parametrizes how easily noise in species i can propagate to species j. $L(\dot{\mathbf{q}}, \mathbf{q})$ is the Lagrangian for a stochastic differential equation with multiplicative white noise to leading order in Ω [Arn00]. The action in Eq. C.3 can be minimized between two steady states in order to determine the relative state stability as described in [PCGBP16]. Briefly, the path is split into N segments such that the points are separated by an equal amount of time Δt. The action can be rewritten into an approximate form as,

$$\mathcal{A}(x_\tau) = \frac{1}{2}\sum_{k=1}^{N}\sum_{i=1}^{N_s}\frac{(\dot{x}_k^i - f_k^i)^2}{D_k^i}\Delta t, \tag{C.5}$$

valid for a system where the noise in one species does not propagate into other species (diagonal D^{-1}). This action corresponds to a path of length τ and $\Delta t = \frac{\tau}{N}$, where f_k^i is evaluated at the centre of each segment $\bar{x}_k = \frac{x_{k-1}+x_k}{2}$, path velocity is taken to be $\dot{x}_k = \frac{(x_k - x_{k-1})}{\Delta t}$, and

N_s is the number of species considered. The action for this path can be minimized using a limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm. A gradient descent method is then used to find the time τ that generates the minimum action. The gradients to be used with this algorithm are given by,

$$\frac{\partial \mathcal{A}}{\partial x_s^l} = \frac{\dot{x}_s^l - f_s^l}{D_s^l} - \frac{\dot{x}_{s+1}^l - f_{s+1}^l}{D_{s+1}^l} - \frac{\Delta t}{2}\sum_{j=1}^{N_s}\left[\frac{\dot{x}_{s+1}^j - f_{s+1}^j}{D_{s+1}^j} f_{l,s+1}'^j + \frac{\dot{x}_s^j - f_s^j}{D_s^j} f_{l,s}'^j \right. \\ + \frac{(D')_{l,s+1}^j}{2(D^2)_{l,s+1}^j}\left(\dot{x}_{s+1}^j - f_{s+1}^j\right)^2 \\ \left. + \frac{(D')_{l,s}^j}{2(D^2)_{l,s}^j}\left(\dot{x}_s^j - f_s^j\right)^2\right]. \tag{C.6}$$

These expressions are significantly more complicated when D^{-1} ceases to be diagonal. Additionally, as the dimensionality (number of chemical species) increases the L-BFGS algorithm becomes increasingly numerically intensive. For these reasons an alternative method of computing the minimum action, termed the geometric minimum action path (gMAP) method, was also used.

C.2 Geometric minimum action path method

The geometric minimum action path (gMAP) method involves discretising the curve by equal units of arc length α rather than by equal time spacing [HVE08a]. This eliminates the need for a gradient descent method to find optimal time of path, saving computational time. The new algorithm is more complicated than the previous algorithm, but through sensible consideration of geometry, avoids the creation of unphysical (non-smooth) paths, thus speeding up the algorithm considerably for higher dimensions.

Contrary to the prior method we now consider the Hamiltonian rather than the Lagrangian. This allows determination of the minima of the action via solution of Hamilton's equations. Solution of these equations speeds the numerical minimization compared to e.g. hill-climbing optimization algorithms, as knowledge of the underlying physics has been integrated into the

procedure. The Hamiltonian for a stochastic differential equation

$$\dot{x} = f(x) + g(x(t))\xi(t) \tag{C.7}$$

is given as

$$H(x,\theta) = \langle f(x), \theta\rangle + \frac{1}{2}\langle \theta, D(x)\theta\rangle, \tag{C.8}$$

where $D = gg^T$, $\xi(t)$ is the vector of noises, x is the vector of copy numbers, f is the vector of deterministic forces, θ is the vector of conjugate momenta, and $\langle a, b\rangle$ means the dot product of vector a and vector b [HVE08a]. Through the rest of this text, $H_x \equiv \frac{\partial H}{\partial x}$, $H_\theta \equiv \frac{\partial H}{\partial \theta}$, $H_{\theta x} \equiv \frac{\partial}{\partial x}(\frac{\partial H}{\partial \theta})$ and $H_{\theta\theta} \equiv \frac{\partial}{\partial \theta}(\frac{\partial H}{\partial \theta})$. The conjugate momenta cannot simply be taken to be $m\dot{x}$ as would be the case when considering the motion of a particle in Cartesian coordinates with no noise. Here, the noise intensity will effect how readily the path can pass through a region, thus the conjugate momenta must be a function of the noise in the vicinity. Although conjugate momenta ϑ can be obtained via $p_i = \frac{\partial L}{\partial \dot{x}_i}$, this expression cannot be solved just with knowledge of position x and arc-length "velocity" $y = \frac{\partial x}{\partial \alpha}$. There will also be a dependence on the undefined function $\lambda(x, y)$, which relates "velocity" in arc-length parametrization to the velocity in species space.

The only information available to the algorithm is the path in q discretized in α. Thus all functions must be definable in terms of x and $y = \frac{\partial x}{\partial \alpha}$. In order to determine ϑ and λ in form dependant only x and y, a number of assumptions are made. The first is that λ is always positive, i.e. the assumption that the movement along the arc is occurring forwards in time. The second is that Hamilton's equations, specifically $\frac{\partial H}{\partial \theta_i} = \dot{x}_i$, are obeyed. Obeying Hamilton's equations is a sufficient and necessary condition for action $\mathcal{A}$ to be a minimum, so this will only give the "correct" momenta when x is the path minimizing $\mathcal{A}$, but as the algorithm proceeds this will be an increasingly good assumption. The final assumption is that the minimizing solution will be a zero energy path $E = H = 0$. This is a reasonable assumption as for all Hamiltonians without time dependence, $H = const$ along the minimizing path. All paths are started with zero velocity in a steady state, therefore the value of this constant E should be

zero. The three assumptions are summarized as

$$\lambda \geq 0, \qquad H_\theta(x, \vartheta) = \lambda y, \qquad H(x, \vartheta) = 0. \tag{C.9}$$

For Eq. C.9, a unique solution $(\vartheta(x, y), \lambda(x, y))$ exists. In the case of a stochastic drift process, equations for both λ and ϑ can be obtained as

$$\lambda(x, y) = \frac{|f(x)|_D}{|y|_D} \tag{C.10}$$

and

$$\vartheta(q, y) = D^{-1}(x)\left(\frac{|f(x)|_D}{|y|_D} y - f(x)\right) \tag{C.11}$$

where $|X|_D = \sqrt{XD^{-1}(x)X}$. The gMAP algorithm selects points that are separated by an equal arc-length. This implies a rediscretization as

$$\varphi(\alpha) = \psi(t), \tag{C.12}$$

where $\alpha \in [0, 1]$, and $t \in [0, T_f]$ where T_f is the total time of the path. The two paths φ and ψ are identical, but the differing discretizations mean that the points that are selected to be used by the optimization algorithm differ. The relation of t and α is done via the function,

$$t = G(\alpha) = \int_0^\alpha \frac{1}{\lambda(a)} da, \tag{C.13}$$

leading to

$$\varphi(\alpha) = \psi(G(\alpha)) \tag{C.14}$$

and

$$\frac{\partial}{\partial \alpha}(\varphi(\alpha)) = \varphi'(\alpha) = \dot{\psi}(t) G'(\alpha) = \frac{\dot{\psi}(t)}{\lambda(\alpha)}. \tag{C.15}$$

Further differentiation of Eq. C.15 by α leads to

$$\lambda^2 \varphi'' + \lambda \lambda' \varphi' = \ddot{\psi}. \tag{C.16}$$

The necessary and sufficient condition for the path ψ to minimize the action $\mathcal{A}$ is that Hamilton's equations are satisfied at every point along the path, i.e. $\dot{\psi} = H_\theta(\psi, \theta)$ and $\dot{\theta} = -H_x(\psi, \theta)$. Together with Eq. C.16 this can be used to derive a condition for the solution φ being stationary dependent only on φ and α

$$\begin{aligned} \ddot{\psi} &= \frac{\partial}{\partial t}(H_\theta(\psi,\theta)) = \frac{\partial}{\partial \psi}(H_\theta)\dot{\psi} + \frac{\partial}{\partial \theta}(H_\theta)\dot{\theta} \\ \lambda^2\varphi'' + \lambda\lambda'\varphi' &= H_{\theta x}\dot{\psi} - H_{\theta\theta}H_x \\ \lambda^2\varphi'' + \lambda\lambda'\varphi' &= H_{\theta x}\lambda\varphi' - H_{\theta\theta}H_x \\ \lambda^2\varphi'' + \lambda\lambda'\varphi' - H_{\theta x}\lambda\varphi' + H_{\theta\theta}H_x &= 0. \end{aligned} \tag{C.17}$$

This provides a condition that is satisfied when the curve minimizes $\mathcal{A}$. At any stage in the algorithm, the deviation of the present curve from this condition can be used to determine the appropriate update for the curve. This deviation δ can be used iteratively in a simple relaxation method towards minima. A relaxation time step τ, is artificially introduced in order to determine the step size the algorithm will use as it relaxes towards this minima. Use of $\dot{\varphi} = \frac{\partial \varphi}{\partial \tau} = \delta$ can then determine φ for the next step. Iterative use of this procedure represents an algorithm to minimize the action.

C.2.1 Algorithmic procedure

Here we explain how the above gMAP algorithm can be implemented as a numerical optimization algorithm to find the gMAP for the case of a stochastic drift process. We then layout how the action and the time of the path can be found through this optimization. The steps are:

1) Discretize $\varphi(\tau, \alpha)$ in both τ and α and thus define $\varphi_i^k = \varphi(k\Delta\tau, i\Delta\alpha)$, k being a natural number, $i = 0, ..., N$, $\Delta\tau$ is the "time" step, and $\Delta\alpha = \frac{1}{N}$. Here, the curve is being discretized into $N + 1$ points.

2) From φ_i^k calculate $\varphi_i'^k = \frac{\varphi_{i+1}^k - \varphi_{i-1}^k}{2/N}$, then use Eqs. C.10 and C.11 to calculate λ_i^k and ϑ_i^k, the edge cases are then taken to be $\lambda_0^k = 3\lambda_1^k - 3\lambda_2^k + \lambda_3^k$ and $\lambda_N^k = 3\lambda_{N-1}^k - 3\lambda_{N-2}^k + \lambda_{N-3}^k$. Now the calculation of $\lambda_i'^k = \frac{\lambda_{i+1}^k - \lambda_{i-1}^k}{2/N}$ can be done.

3) Find $\{\tilde{\varphi}_i\}_{i=0,...,N}$ the solution to the following system of equations

$$\begin{aligned} \frac{\tilde{\varphi}_i - \varphi_i^k}{\Delta\tau} &= (\lambda_i^k)^2 \frac{\tilde{\varphi}_{i+1} - 2\tilde{\varphi}_i + \tilde{\varphi}_{i-1}}{1/N^2} - \lambda_i^k H_{\theta x} \varphi_i'^k + H_{\theta\theta} H_x + \lambda_i^k \lambda_i'^k \varphi_i'^k \\ \tilde{\varphi}_0 &= x_1 \\ \tilde{\varphi}_N &= x_2, \end{aligned} \tag{C.18}$$

where $H_{\theta x}$, $H_{\theta\theta}$ and H_x are evaluated at $(\varphi_i^k, \vartheta_i^k)$.

4) Now interpolate a curve across $\tilde{\varphi}$ and then rediscretize in order to obtain φ^{k+1}. This rediscretization ensures that points remain at equal spacing along the arc length.

5) Repeat steps 2-4 until change in path between steps becomes sufficiently small.

6) The final path can then be used to calculate the action by

$$\mathcal{A} = \frac{1}{N}\left(\frac{3}{2}\langle\varphi_1'^k, \vartheta_1^k\rangle + \sum_{i=2}^{N-2}\langle\varphi_i'^k, \vartheta_i^k\rangle + \frac{3}{2}\langle\varphi_{N-1}'^k, \vartheta_{N-1}^k\rangle\right). \tag{C.19}$$

7) The time to point i can now be found by

$$t_i = \frac{1}{2\lambda_0^k} + \frac{1}{\lambda_1^k} + ... + \frac{1}{\lambda_{i-1}^k} + \frac{1}{2\lambda_i^k}. \tag{C.20}$$

8) Finally, these times can be used to reverse the procedure followed in step 1) in order to generate a path that is discretized in time rather than along arc length.

Appendix D

Collection of graphs supporting arguments made in Chapter 2

D.1 Generality of relation between Langevin and reduced master equation

The generality of the relationship shown in Fig. 2.3 could be questioned as it arises for a specific parameter set. Additionally, plots were only provided for one state and one switching direction. This subsection will provide additional figures in order to address these potential criticisms. Thus, Fig. D.4 shows plots for the same parameter set as Fig. 2.3 but with plots that consider trajectories about the low A steady state and switching from high A to low A. Figs. D.1 & D.2 show the same 8 plots for one alternative parameter set, and Figs. D.3 & D.5 show the same plots for another alternative parameter set. It can been seen that in all cases the entropy production from the Langevin formalism approximates that of consideration of the master equation of the reduced model. Hence, it seems reasonable to conclude that this is generally a good first approximation.

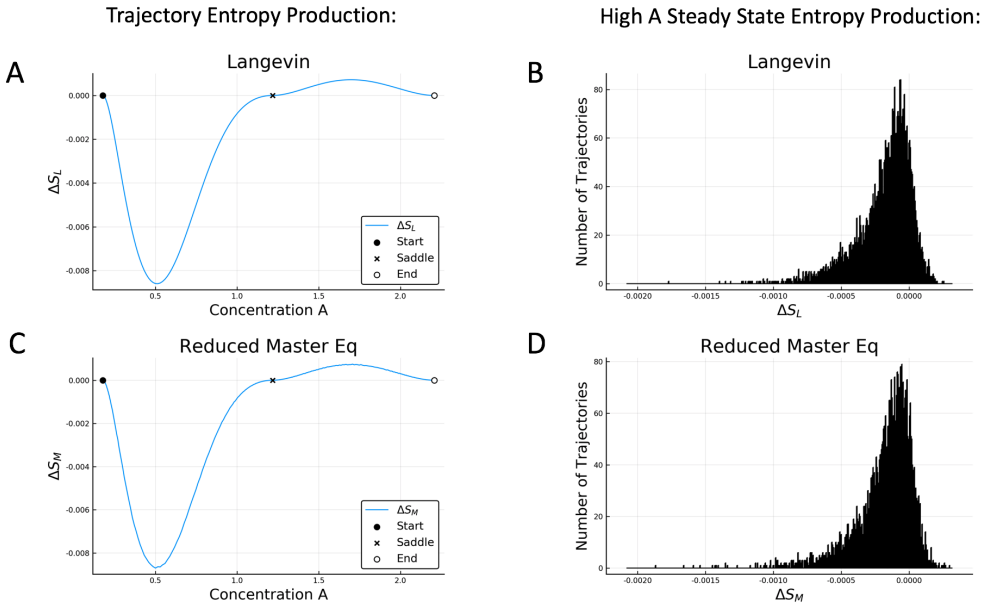


Figure D.1: Same plots as in Fig. 2.3 but all 4 plots are now produced using the following alternative parameterization: $k = 26.0978$, $k_- = 2.17942$, $q = 25.7722$, $q_- = 1.41486$, $K = 8.24010$, $K_- = 0.600485$, $Q = 6.96377$, $Q_- = 0.165740$, $r = 3875.55$, $f = 9285.77$, and $\Omega = 5000$. Appropriate volume (Ω) rescaling of rates was used for the discrete simulations and calculations.

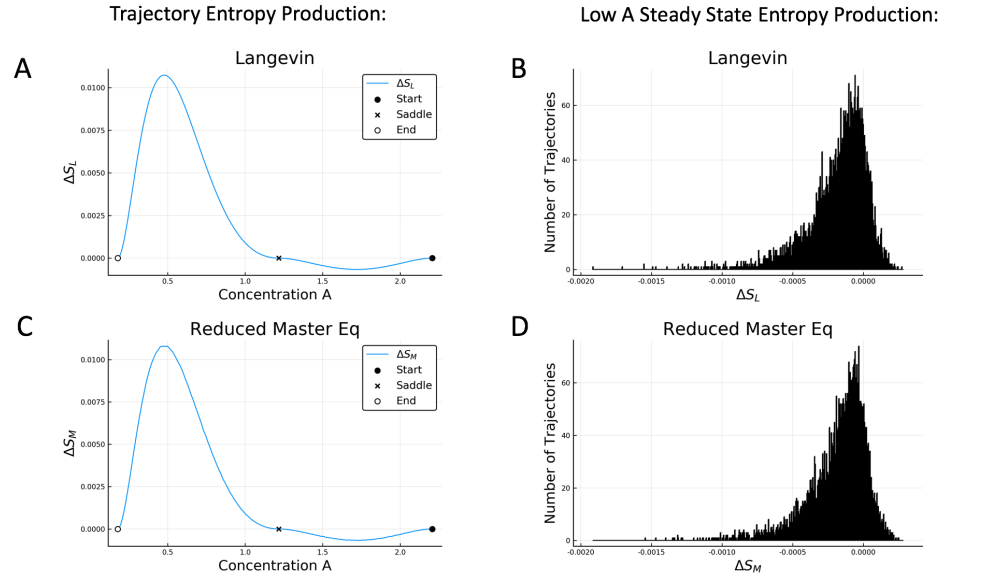


Figure D.2: Same parameterization as Fig. D.1, but now (a) & (c) are switches from high A to low A, and (b) & (d) are simulated trajectories about the low A steady state.

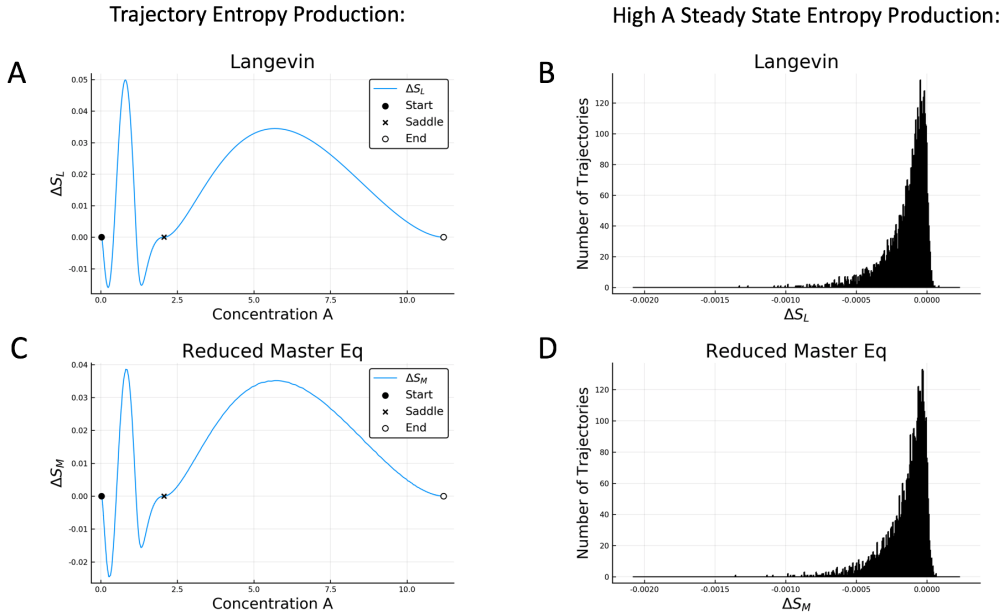


Figure D.3: Same plots as in Fig. 2.3 but all 4 plots are now produced using the following alternative parameterization: $k = 98.7494$, $k_- = 6.53603$, $q = 78.8964$, $q_- = 3.00145$, $K = 2.16553$, $K_- = 0.0566242$, $Q = 2.04786$, $Q_- = 0.135649$, $r = 3832.23$, $f = 9994.70$, and $\Omega = 5000$. Appropriate volume (Ω) rescaling of rates was used for the discrete simulations and calculations.

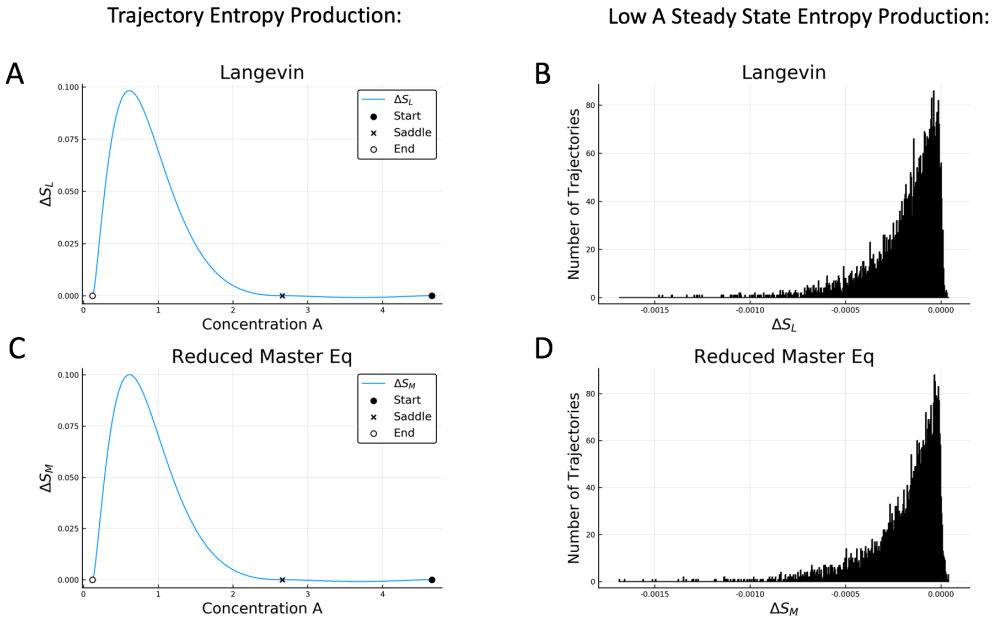


Figure D.4: Same parameterization as Fig. 2.3, but now (a) & (c) are switches from high A to low A, and (b) & (d) are simulated trajectories about the low A steady state.

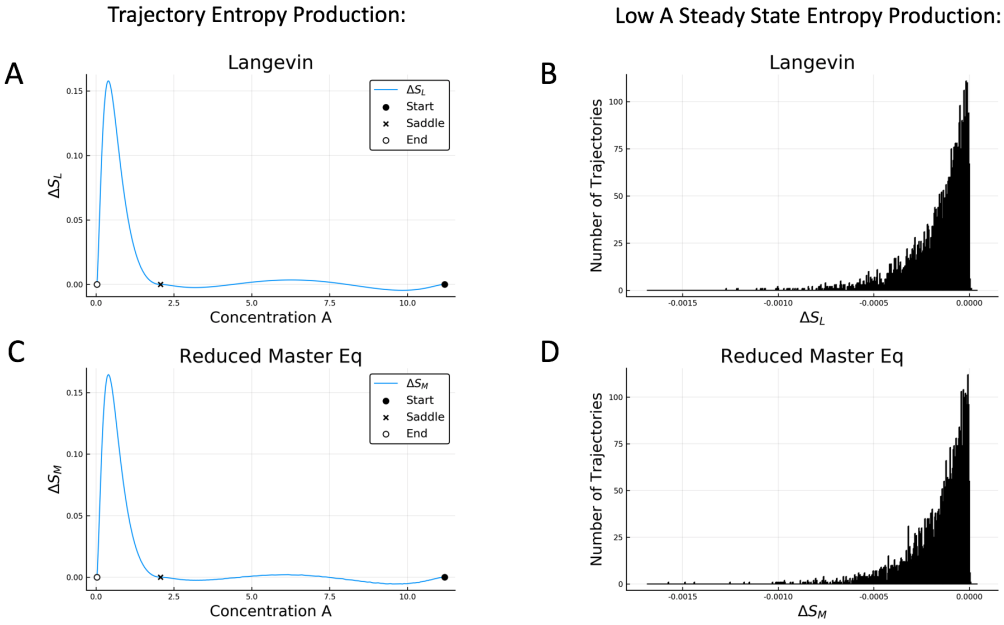


Figure D.5: Same parameterization as Fig. D.3, but now (a) & (c) are switches from high A to low A, and (b) & (d) are simulated trajectories about the low A steady state.

D.2 The source of divergences from the multi-dimensional switching relation

In order to determine what caused the significant divergence of some parameterizations from our analytic relation Eq. 2.44, we initially investigated the difference between switching paths in one direction versus the contrary direction. We defined divergence as

$$\delta = \frac{|\Omega(\mathcal{A}_{A\to B} - \mathcal{A}_{B\to A}) - \frac{1}{2}(\Delta S^L_{B\to A} - \Delta S^L_{A\to B})|}{|\frac{1}{2}(\Delta S^L_{B\to A} - \Delta S^L_{A\to B})|}, \tag{D.1}$$

i.e. the difference of the observed difference in action from the difference predicted by our relation. This is divided by the difference in path entropy production in order to ensure that significant differences between small values are not neglected. In Figs. D.6A and D.6B switching paths are plotted that fit our relation well and poorly, respectively. It can be observed that paths in the poorly fitting case are not substantially more different than in the well fitting case. However, it can be noted that the saddle point lies at a significantly lower total concentration in the poorly fitting case. This is a generally observed feature that is illustrated by Fig. D.7. Here, the divergence δ is plotted against the relative saddle-point concentration $R_s = h_s/(h_1 + h_2)$,

i.e the total chemical concentration at the saddle point (h_s) divided by the total chemical concentration at the two steady-states (h_1 and h_2). R_s can be thought of as the relative height of the saddle point compared to the two steady states. It can be seen from Fig. D.7 that parameter sets with low relative saddle point will always show significant divergences from our relation.

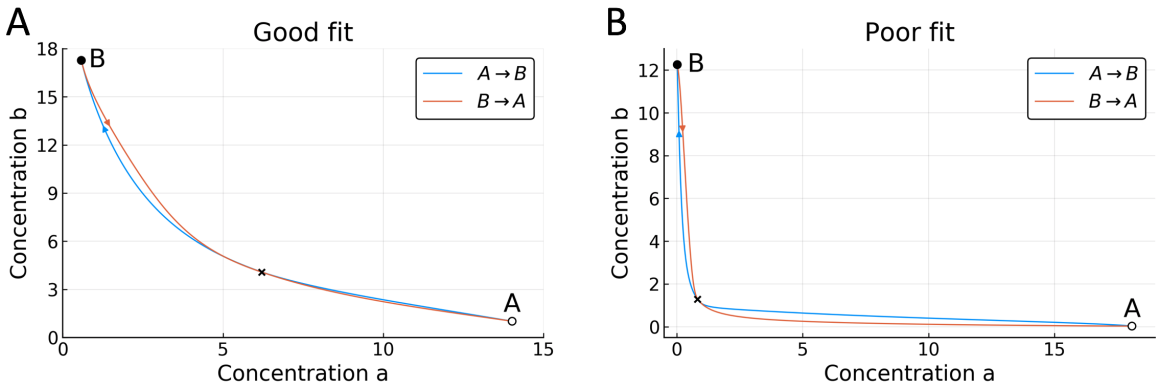


Figure D.6: **Comparison of path differences.** (a) Switching paths for a parameterization of the toggle switch model that fits well ($\delta = 0.0041$) with our analytic relation Eq. 2.44. The parameterization (given to 6 sf) is as follows: $k = 85.7824$, $k_- = 1.31870$, $q = 58.5517$, $q_- = 0.490925$, $K = 4.30008$, $K_- = 0.114606$, $Q = 2.80976$, $Q_- = 0.162321$, $r = 8525.81$, $f = 753.812$. (b) Similar plot but for a parameterization that fits the analytic relation poorly ($\delta = 0.9383$). The parameterization is as follows (given to 6 sf): $k = 99.2988$, $k_- = 2.22327$, $q = 84.0313$, $q_- = 4.45835$, $K = 3.20045$, $K_- = 0.0286723$, $Q = 2.39470$, $Q_- = 0.198633$, $r = 618.219$, and $f = 7949.87$.

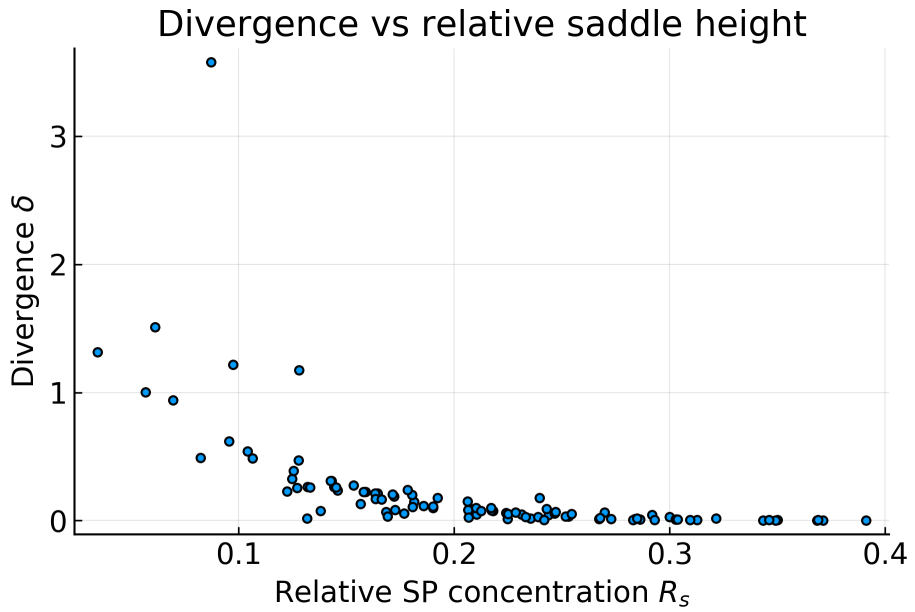


Figure D.7: **Low saddle points (SPs) increase divergence.** Plot of the divergence δ from our relation against relative height of saddle point R_s. Each point here represents a specific parameterization of the toggle switch model.

In order to explain this result we must again note that all terms in the force are constant or vary linearly save the two production terms ($kr/(r + fb^2)$ and $qr/(r + fa^2)$). These two non-linear terms are approximately linear when either $a, b \ll r/f$ or $a, b \gg r/f$. In the intermediate zone between these two regions the value of these terms can vary rapidly in space. In this case two paths that are close in space can possess substantially different forces along their lengths, which can be seen in Fig. D.8. Additionally, the diffusion matrix will be affected as the non-linear terms also play a role here, which is shown in Fig. D.9. This significant variation in force (and to a lesser extent diffusion matrix components) means that the conservative actions of the two paths are no longer approximately equal. Thus, parameterizations that result in low saddle points constrain the switching paths to pass through regions where the approximation of equal conservative action breaks down. This explains the pattern seen in Fig. D.7.

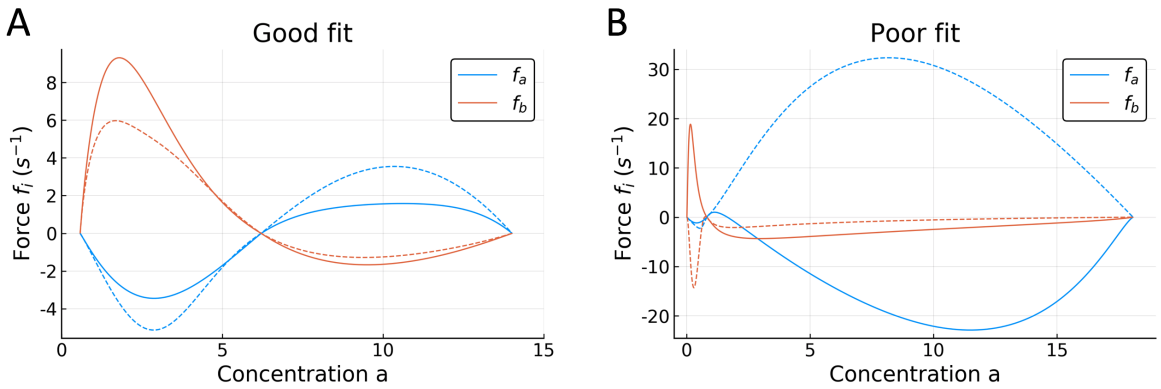


Figure D.8: **Differences in force along paths.** Parameterizations used for (a) and (b) are the same as in Fig. D.6A and B, respectively. As in Fig. 2.6, solid and dashed lines denote a $A \rightarrow B$ and $B \rightarrow A$ paths, respectively. (a) Both components of the force remain similar for the well fitting case. (b) Huge difference in both components of the force along the path for the poorly fitting case.

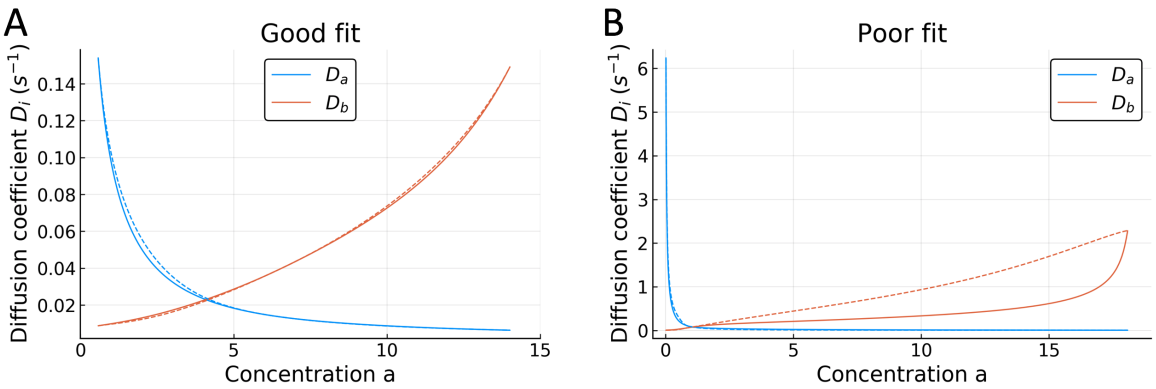


Figure D.9: **Differences in diffusion along paths.** Parameterizations used for (a) and (b) are the same as in Fig. D.6A and B, respectively. As in Fig. 2.6, solid and dashed lines denote a $A \rightarrow B$ and $B \rightarrow A$ paths, respectively. (a) Both components of the diffusion matrix are nearly identical along the two paths. (b) While one component remains similar for the poorly fitting case the other component now differs substantially.

Appendix E

Detailed derivation of the novel thermodynamic microbial community model

In this appendix, the model used in Chapters 3 & 4 is derived in detail. In order to do this, two preliminary sections set out the difference between reversible and irreversible enzyme kinetics, and justify our choice of considering irreversible kinetics. The third section then proceeds to derive the model in detail.

E.1 Irreversible Michaelis–Menten kinetics

Most models of microbial growth assume irreversible Michaelis–Menten kinetics, which can be represented as

$$E + S \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} ES \xrightarrow{k_{\text{cat}}} E + W, \tag{E.1}$$

where k_{+1}, k_{-1} and k_{cat} are rate constants, E is an enzyme, ES is an enzyme substrate complex, S is a substrate and W is a waste product. This scheme leads to an expression for the rate of

substrate consumption (at steady state) as

$$\frac{d[S]}{dt} = -\frac{d[W]}{dt} = -V_{\max}\frac{[S]}{K_M + [S]}, \tag{E.2}$$

where $V_{\max}$ is the maximal rate of substrate consumption (or waste product production), and K_M is the Michaelis (saturation) constant. These constants are defined in terms of rate constants by

$$V_{\max} = k_{\text{cat}}[E]_0 \tag{E.3}$$

and

$$K_M = \frac{k_{-1} + k_{\text{cat}}}{k_{+1}}, \tag{E.4}$$

where $[E]_0$ is the initial enzyme concentration. This provides a mechanistic justification for the phenomenological Monod microbial growth equation, which is frequently used in ecological modelling [SKMH14]. However, the enzyme scheme in Eq. E.1 assumes that the second reaction is irreversible which is only possible (as an approximation) if this reaction is dissipating a very large amount of free energy. Though this is a reasonable assumption for cases of growth on high free energy substrate, it is not when considering the low free-energy change reactions commonly observed in microbial ecosystems. We thus wish to consider a fully reversible scheme of enzyme kinetics.

E.2 Reversible enzyme kinetics

To model non-equilibrium thermodynamics, we can instead consider the scheme

$$E + S \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} EX \underset{k_{-2}}{\overset{k_{+2}}{\rightleftharpoons}} E + W, \tag{E.5}$$

where $k_{\pm 1,2}$ are rate constants [BA53]. This scheme is fully reversible and is thus suitable to describe low free-energy change reactions. In this case the rate of change of substrate/waste-

product concentration (at steady state) can be found to be

$$\frac{d[W]}{dt} = -\frac{d[S]}{dt} = \frac{\frac{V_S}{K_S}[S] - \frac{V_W}{K_W}[W]}{1 + \frac{[S]}{K_S} + \frac{[W]}{K_W}}, \tag{E.6}$$

where V_W and V_S are maximal rates for waste product and substrate production, respectively. These rates are defined for the case where all enzymes are bound as complexes ($[EX] = [E]_0$) and can be written as

$$V_S = k_{+2}[E]_0, \qquad V_W = k_{-1}[E]_0. \tag{E.7}$$

Additionally, K_S and K_W are the dissociation constants for substrate and waste product, respectively. These constants can be defined in terms of rate constants as

$$K_S = \frac{k_{-1} + k_{+2}}{k_{+1}}, \qquad K_W = \frac{k_{-1} + k_{+2}}{k_{-2}}. \tag{E.8}$$

In order to further simplify Eq. E.6 we now make use of a quantity termed the equilibrium constant. This is a thermodynamic constant specific to physical conditions (e.g. temperature, particular reaction type), and constrains the relative values of the rate constants (at a given temperature) by [HCR96]

$$\mathcal{K} = \frac{k_{+1}k_{+2}}{k_{-1}k_{-2}}. \tag{E.9}$$

Eq. E.6 can then be rearranged to

$$\frac{d[W]}{dt} = \frac{V_S\left([S] - \frac{[W]}{\mathcal{K}}\right)}{K_S + [S] + \frac{K_S}{K_W}[W]}. \tag{E.10}$$

The value of the equilibrium constant $\mathcal{K}$ will be affected by the total Gibbs free energy (ΔG_T) of the reaction and the endergonic reactions it is coupled to. Coupling one or more endergonic reactions ($\Delta_r G > 0$) to our original exergonic reaction ($\Delta_r G < 0$) allows the acquisition of free energy for later use in other processes. Cells commonly couple metabolic reactions to the production of ATP which is strongly endergonic in order to transduce free energy [Hil89]. The

coupling of our generic reaction to the synthesis of ATP can be expressed as

$$E + S + \eta\mathrm{ADP} + \eta P_i \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} EX \underset{k_{-2}}{\overset{k_{+2}}{\rightleftharpoons}} E + W + \eta\mathrm{ATP}, \tag{E.11}$$

where P_i is organic phosphate, η is the number ATP produced per reaction event. In order to account for alternative free energy transducing processes such as the pumping of ions across membranes, η can to take non-integer values in our full model [LWvB$^+$19]. The total Gibbs free energy of a reaction coupled to ATP production in this manner is

$$\Delta G_T = \Delta_r G^0 + RT \ln Q + \eta \Delta G_{\mathrm{ATP}}, \tag{E.12}$$

where R is the gas constant, T is the temperature, $\Delta_r G^0$ is the Gibbs free energy of one mole of reaction at standard conditions, Q is the reaction quotient and ΔG_{ATP} is the Gibbs free energy per mole of ATP. The value of ΔG_{ATP} is dependant on the cellular concentrations of ATP, ADP and organic phosphate. In this work we will assume that these are constant and will use the experimentally-derived estimate for ΔG_{ATP} of 75 kJ mol^{-1} throughout [BT13]. For the equilibrium case $\Delta G_T = 0$ the reaction quotient Q takes the value

$$Q_0 = \mathcal{K} = \exp\left(\frac{-\Delta_r G^0 - \eta \Delta G_{\mathrm{ATP}}}{RT}\right), \tag{E.13}$$

equal to the previously-discussed equilibrium constant. This means that the constraint on the relative reaction rates (shown in Eq. E.9) is determined by the reaction quotient of the combined reaction at equilibrium. The first term in the numerator is the Gibbs free energy change at standard conditions (i.e. the substrate and waste product have equal concentration). The equilibrium constant thus tells in which direction we would expect the reaction to proceed when waste product and substrate have equal concentrations (i.e. towards increasing waste product if $\mathcal{K} > 1$ and towards increasing substrate if $\mathcal{K} < 1$). The second term of the numerator shifts this equilibrium so as the amount of free energy dissipated decreases the amount of waste product that needs to build up for equilibrium to be reached also decreases. Increasing temperature T decreases the magnitude of this exponent and so the effect varies dependant on the sign of the

numerator. In our model we will neglect the effect of temperature on the dynamics.

Following [HCR96], we can make use of the fact that $V_S = k_{+2}[E]_0$ (which we shall simplify to $k[E]_0$) to further rearrange Eq. E.10 to

$$\frac{d[W]}{dt} = \frac{k[E]_0\left([S] - \frac{[W]}{\mathcal{K}}\right)}{K_S + [S] + \frac{K_S}{K_W}[W]} = \frac{k[E]_0[S]\left(1 - \frac{[W]}{[S]\mathcal{K}}\right)}{K_S + [S]\left(1 + \frac{K_S[W]}{K_W[S]}\right)}. \tag{E.14}$$

Into to the above expression we can substitute $\frac{K_S}{K_W} = \frac{r}{\mathcal{K}}$ and $\theta = \frac{Q}{\mathcal{K}}$, and then obtain the consumption rate q as

$$q([S],[W]) = \frac{k[E]_0[S](1-\theta)}{K_S + [S](1 + r\theta)}, \tag{E.15}$$

where k is the maximal reaction rate, K_S is the enzyme (substrate) saturation constant, and r is a reversibility factor [HCR96]. Further, θ is a thermodynamic factor that tends towards one when reactions are close to equilibrium, which can be expressed as

$$\theta = \theta\left([S],[W]\right) = \frac{Q([S],[W])}{\mathcal{K}(S,W)} \tag{E.16}$$

where $\mathcal{K}(S,W)$ is the equilibrium constant for the reaction from S to W and Q is the reaction quotient given by

$$Q(A,B) = \frac{B}{A}. \tag{E.17}$$

It is useful to note that for reactions dissipating large amounts of free energy $\mathcal{K} \gg Q$, and thus $\theta \approx 0$, meaning that Michaelis–Menten dynamics are recovered from Eq. E.15. This explains why the Michaelis–Menten scheme is appropriate to model many (but not all) ecosystems, particularly lab cultures where free energy supply is generally high.

E.3 Proteome partitioning model

The basic assumption of our proteome partitioning model is that consistent with empirical observations, cellular proteins can be divided into three broad classes: 1) ribosomes (denoted R) that allow further protein synthesis, 2) metabolic enzymes (denoted P) that obtain energy

from the external metabolites, and 3) a fixed proportion of housekeeping proteins (denoted Q) necessary for other vital cellular processes [SKMH14]. The rate of change of these three classes of proteins can be modelled with the following equations:

$$\frac{dR}{dt} = \frac{\gamma(a)}{n_R} f_b^R R - \lambda R \tag{E.18}$$

$$\frac{dP}{dt} = \frac{\gamma(a)}{n_P} f_b^P R - \lambda P \tag{E.19}$$

$$\frac{dQ}{dt} = \frac{\gamma(a)}{n_Q} f_b^Q R - \lambda Q. \tag{E.20}$$

Here, λ is the cellular growth rate, a is the cellular energy concentration, $\gamma(a)$ is the effective translation elongation rate, n_x is the typical number of translation steps needed to produce a protein of type x, and f_b^x is the fraction of ribosomes bound and producing protein of type x. Further, the number of translations steps needed for each protein step can be approximated from the literature as 7459 (amino acids per protein) [KCVSZ$^+$11], 300 and 300 [BEO$^+$09, WODS15] for n_R, n_P and n_Q, respectively. As protein synthesis operates far from thermodynamic equilibrium in order to minimise translation errors, we assume it to be effectively irreversible. The effective translation elongation rate thus has a saturating dependence on energy availability as

$$\gamma(a) = \frac{\gamma_m a}{\gamma_{\frac{1}{2}} + a}, \tag{E.21}$$

where γ_m is the maximum elongation rate and $\gamma_{\frac{1}{2}}$ is the free energy concentration where the rate is half of the maximum. To the best of our knowledge, this half-maximum constant has not been empirically measured, so is a free parameter in our model. However, the maximum elongation rate (γ_m) has been measured to be 21.0 amino acids per second [BD96].

E.3.1 Derivation of simplified proteome partitioning model

The dynamics for the total mass of the cell (in amino acids) are given by

$$\frac{dm}{dt} = \gamma(a) f_b R - \lambda m \tag{E.22}$$

where f_b is the total fraction of ribosomes that are bound. We assume that the cells retain a constant average protein mass, which an order of magnitude estimate of is taken from the literature as $m = 10^8$ [BD96]. As m is taken to be constant at steady state, Eq. E.22 must equal zero and so can be solved to give an expression for growth rate as

$$\lambda = \frac{\gamma(a) f_b R}{m}. \tag{E.23}$$

Although the fraction of ribosomes bound (f_b) would be expected to vary with the cells internal energy concentration, as a simplifying assumption we take it to be fixed at a value of 0.7 [USP05]. However, the fractions of ribosomes bound making specific protein classes (f_b^R etc) can still vary. Previous mechanistic modelling in the literature suggests that the transcription rate of ribosomal mRNAs shows a stronger response to increasing energy concentration than the transcription rate of non-ribosomal mRNAs [WODS15]. This means that as the cell's energy concentration increases the relative proportion of ribosomes bound to ribosomal mRNAs, i.e. f_b^R increases at the expense of f_b^P as f_b^Q is held constant by steep auto-inhibition. Thus, the fraction of the proteome dedicated to each protein type should be expected to be approximately a direct function of the energy concentration. Each protein class can be expressed as a fraction of the total proteome satisfying the following constraint

$$\phi_R + \phi_P + \phi_Q = 1, \tag{E.24}$$

where $\phi_R = f_b^R / f_b$ is the ribosome fraction, $\phi_P = f_b^P / f_b$ is the metabolic enzyme fraction, and $\phi_Q = f_b^Q / f_b$ is the housekeeping fraction. The maximum value that the ribosome fraction can reach is $\phi_R^{\max} = 1 - \phi_Q$, the cells response to increasing energy concentration must therefore be saturating. As we have no specific mechanism to link this response function to, we assume that it is of Michaelis–Menten form. We can express the ideal ribosome fraction ϕ_R^* (Eq. 3.7) as a function of energy concentration a by

$$\phi_R^*(a) = \frac{a}{a + \Omega_{\frac{1}{2}}} (1 - \phi_Q) \tag{E.25}$$

where $\Omega_{\frac{1}{2}}$ is the half-maximum constant with respect to energy concentration and $\phi_Q = 0.45$ [SGM+10]. As the response function does not correspond to a specific mechanism, $\Omega_{\frac{1}{2}}$ is a treated free parameter of the model. The form of Eq. E.25 ensures that as the energy concentration approaches infinity the ideal ribosome fraction approaches its maximum possible value of $1-\phi_Q$ and as the energy concentration approaches zero the ideal metabolic fraction ϕ_P^* approaches this same maximum. This ideal proteome partition cannot be immediately realised by the cell, as the synthesis of large quantities of protein is required in order to shift the relative size of the fractions. The actual ribosome fraction ϕ_R dynamics (Eq. 3.3) are thus described by

$$\frac{d\phi_R}{dt} = \frac{1}{\tau_g}\left(\phi_R^*(a) - \phi_R\right) \tag{E.26}$$

where τ_g is the characteristic time scale of proteome shifting. In order to define the characteristic time scale we consider the extreme (and unrealisable) case where the ideal ribosome fraction is equal to the maximum $(1-\phi_Q)$ but the actual fraction is zero. We assume that in this extreme case only metabolic proteins would be synthesised, so that each biomass doubling would halve metabolic protein fraction. The ideal fraction will only ever be asymptotically approached through this process of halving, so we consider the process to be complete when the metabolic fraction has dropped to 1% of the maximum. The characteristic time scale can therefore be expressed as

$$\tau_g = \frac{1}{\lambda}\frac{\log 100}{\log 2}, \tag{E.27}$$

where λ^{-1} is the time required for biomass doubling and $\frac{\log 100}{\log 2}$ is the number of halvings required to go from 100% to 1%. In cases where the ideal ribosome fraction is less extreme we would expect the protein production to be balanced between both types. This means that the reduction in time due to needing to synthesise less protein to reach the ideal fraction would be at least partially counteracted by a lower rate of synthesis of the deficient protein type. Thus, we expect the characteristic timescale (τ_g) to be relevant in all cases.

Growth rate (Eq. E.23) can now be re-expressed in terms of the protein fraction as

$$\lambda = \frac{\gamma(a) f_b}{m}\frac{m}{n_R}\phi_R = \frac{\gamma(a) f_b \phi_R}{n_R}, \tag{E.28}$$

which is Eq. 3.5 in Chapter 3. The growth rate is thus heavily determined by the internal energy concentration. The full dynamics of the cells energy concentration is given by

$$\frac{da}{dt} = J - \mathcal{T} - \lambda a, \tag{E.29}$$

where J is the rate at which the cell acquires free energy (as ATP) and $\mathcal{T}$ is the rate of free energy consumption due to protein translation. Following [WODS15], we here make the assumption that protein translation is the only significantly free energy consuming process in the cell. The rate of ATP consumption by protein translation can be expressed as

$$\mathcal{T} = \chi m \lambda, \tag{E.30}$$

where χ is the ATP consumption per translation step, and $m\lambda$ gives the number of translation steps per unit time. The value of χ, including the cost of amino acid synthesis, is taken to be 29.0 [LM15]. In order to find an expression for the energy acquisition rate J we must further divide the proteome. The metabolic protein fraction is in turn divided between a multiple different free energy yielding reactions subject to the constraints

$$\sum_{\alpha \in O} \nu_\alpha = 1 \tag{E.31}$$

and

$$\nu_\alpha \geq 0 \quad \forall \alpha, \tag{E.32}$$

where ν_α is the relative expression of reaction α and O is the set of reactions possessed by the cell. The amount of a enzyme available for reaction α can then be expressed as

$$E_\alpha = \frac{m \nu_\alpha \phi_P}{n_P}, \tag{E.33}$$

where n_P is the mass (i.e. number of translation steps) of a metabolic protein. The rate of free

energy acquisition can now be expressed as

$$J = \sum_{\alpha \in O} \eta_\alpha q_\alpha(E_\alpha, S_\alpha, W_\alpha), \tag{E.34}$$

where η_α is the amount of ATP the cell synthesises from reaction α. The reaction rate for reaction α is a slightly altered form of Eq. E.15

$$q_\alpha(E_\alpha, S_\alpha, W_\alpha) = \frac{k_\alpha E_\alpha S_\alpha (1-\theta)}{K_{S_\alpha} + S_\alpha(1 + r_\alpha \theta)}, \tag{E.35}$$

where k_α is the rate limiting reaction rate (k_{+2} in Eq. E.5) for reaction α, K_{S_α} the saturation constant, and r_α is the reversibility factor.

The cell population is denoted by N and the population dynamics can be described by

$$\frac{dN}{dt} = N\left(\lambda - d\right), \tag{E.36}$$

where d is the death rate accounting for both explicit cell death and other biomass losses. Our model incorporates a large number of metabolites, which can perform the function of both substrates and products for the various reactions. The concentration of metabolite β is given by C_β, and its dynamics can be described by

$$\frac{dC_\beta}{dt} = \kappa_\beta - \rho_\beta C_\beta + \left(p_\beta(\mathbf{C}) - c_\beta(\mathbf{C})\right) N, \tag{E.37}$$

where κ_β is the supply rate of metabolite β, ρ_β is the dilution rate of metabolite β, p_β is the rate at which the cells produce metabolite β, and c_β is the rate at which the cells consume metabolite β. These net reaction rates can be expressed as

$$p_\beta(\mathbf{C}) = \sum_{\alpha \in O} \delta_{C_\beta, W_\alpha} q_\alpha(E_\alpha, S_\alpha, W_\alpha), \tag{E.38}$$

and

$$c_\beta(\mathbf{C}) = \sum_{\alpha \in O} \delta_{C_\beta, S_\alpha} q_\alpha(E_\alpha, S_\alpha, W_\alpha). \tag{E.39}$$

where $\delta_{C_\beta,W_\alpha}$ and $\delta_{C_\beta,S_\alpha}$ are Kronecker deltas that are zero unless metabolite β is the waste product of reaction α or metabolite β is the substrate of reaction α, respectively.

E.3.2 Validation for single population

In order to establish the validity of our model we attempted to replicate the linear growth laws found in [SKMH14]. To do this we simulated the strains with the substrate and waste product concentrations held fixed. We then plotted the final ribosome fraction ϕ_R against the final growth rate λ (due to the fixed nutrient environment the strains exponentially grow forever). The first growth law that we consider is found in experiments where drugs decreasing the rate of protein translation are applied to an exponentially growing population. In these experiments as growth rate decreases a linear increase in ribosome fraction is observed. To approximately replicate these experiments in our model we decreased the maximum translation rate per ribosome γ_m (Fig. E.1, top). The other growth law that we consider is found in experiments where microbial strains are grown upon media of successively higher nutrient quality. Here, as growth rate increases ribosome fraction linearly increases along with it. To adapt our framework to this experimental setup we used $\Delta_r G^0$ is the standard Gibbs free energy of the reaction as a proxy for nutrient quality. As the magnitude of this Gibbs free energy is increased there is more free energy available per reaction event to be retained as ATP and thus the value of η can be increased (Fig. E.1, bottom).

We found that across most growth rates linear growth laws could be recovered (see Fig. E.1). However, for both growth laws at low growth rates deviations were observed. In the case of the nutrient quality law at low growth rates the ribosome fraction is lower than would be expected based on the growth rate. This arises as when strains utilise low free energy substrates significantly less ATP accumulation occurs. This acts to further reduce the growth rate. Hence, as nutrient quality is increased the growth rate initially increases super-linearly. In the case of the translational inhibition law at low growth rates the ribosome fraction is higher than would be expected based on the growth rate. In this case the ability of the strain to make use of energy is significantly hampered by the decreased maximum translation rate γ_m,

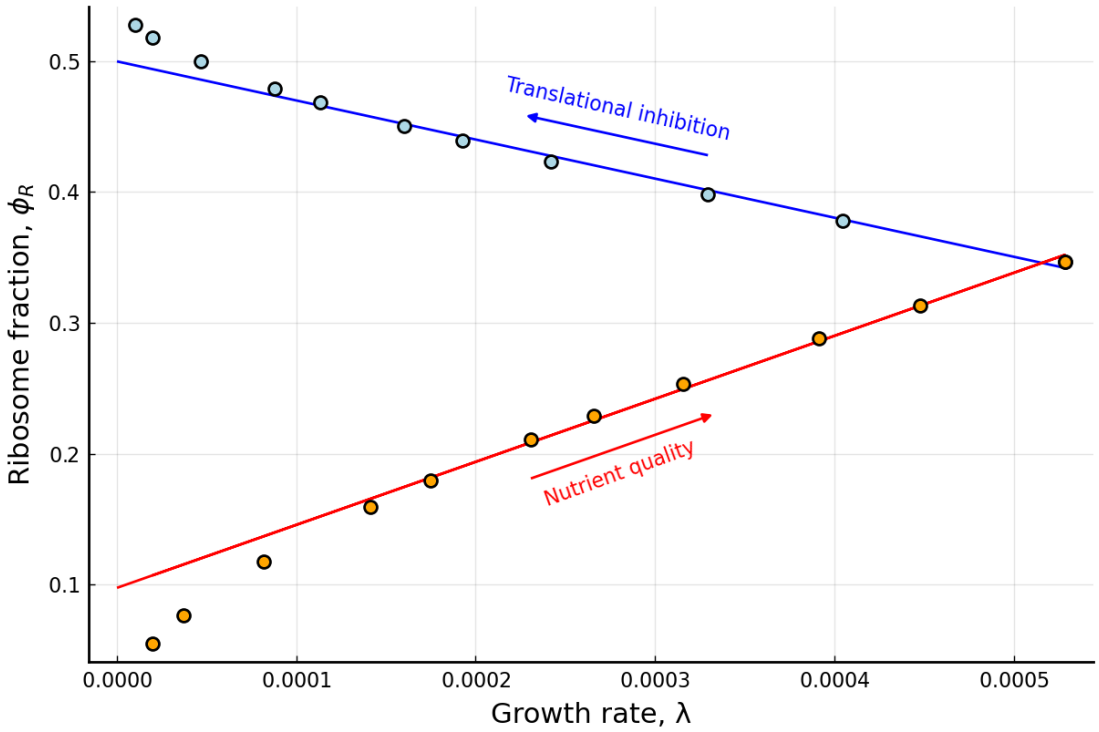


Figure E.1: **Validation based on growth laws.** Final ribosome fraction is plotted against final growth rate for our model simulated using various conditions. The upper (blue) data are simulated for decreasing maximum translation rate γ_m. The lower (orange) data are simulated for increasing standard reaction Gibbs free energy (a proxy for nutrient quality). For both sets of data the linear fit is found excluding the three points with the lowest growth rate.

causing accumulation of energy by the cell. This energy accumulation partially compensates for the decreased maximum translation rate, so growth rate declines more slowly than would be expected.

Appendix F

Details of the quantification and classification of the interactions found in Chapter 3

In this appendix, we detail how the interactions types considered in Chapter 3are calculated and classified. After doing this, we additionally compare the strength and frequency of the different interactions both within and between functional groups.

F.1 Determining interaction types and strengths

All interactions in our model occur via microbes consuming or producing metabolites. Thus, inferring interactions between strains involves calculating both the responses of individual species' populations to changes in metabolite concentrations, and the reciprocal impact on the metabolite concentrations. Calculating these 'susceptibilities' analytically can allow solutions that link interactions to emergent behaviours of such systems. An example is the recent work applying the cavity calculation of statistical physics to model ecosystems [ABM18, CMM20], which allowed susceptibilities to be linked to the ability of new species to invade. However, analytical solutions for the effective resource capacity and the resource susceptibility could not be found

in our model, as reaction rates depend not just on substrate concentrations but also waste-product concentrations. This generates a coupling between resource dynamics (i.e. where the concentration of one metabolite can effect the rate a different metabolite is consumed at), which makes the system far harder to solve analytically. In the absence of the previously mentioned analytic results performing a calculation reminiscent of the cavity calculation does not seem feasible. So rather than making use of this elegant method of determining whether newly introduced species can persist we will make use of our susceptibilities in a relatively crude manner to calculate the instantaneous impact of each species on every other species at steady state, as a measure of interactions between species. These interactions can only be accurately calculated when the ecosystem has reached steady state.

So, in order to calculate species' responses we consider a perturbation of a single metabolite (β) that takes the form

$$\mathbf{C'}_\beta = [C_1, \ldots, (1+\xi) \times C_\beta, \ldots, C_M], \tag{F.1}$$

where C_β is the steady state value of metabolite β, and ξ is the (fractional) perturbation size. We then determine the effect this perturbation has on the rate of change of ATP (Eq. E.29) for strain i which is described by

$$\frac{da_i}{dt}\left(\mathbf{C'}_\beta\right) = J_i\left(\mathbf{C'}_\beta\right) - \chi m \lambda_i - a_i \lambda_i, \tag{F.2}$$

where J_i is the rate of ATP generation (Eq. 3.8), χ is the ATP consumption per translation step, and m is the total mass of the cell (in amino acids). The second term of Eq. F.2 gives the rate at which the strain uses ATP to synthesis protein, and the third term gives the rate at which ATP is diluted by volume expansion. The metabolite perturbation only initially effects the ATP generation rate J_i, as the other two terms have no direct dependence on metabolite concentrations. We cannot find the effect that the perturbation has on the growth rate (Eq. 3.5) directly without finding the new steady state value of a_i, which can only be found numerically. Therefore, we instead assume that the change in the (per cell) rate of ATP generation (J_i)

directly impacts the rate of change of (population) biomass as

$$\frac{dN_i}{dt}\left(\mathbf{C'}_\beta\right) = \frac{N_i}{\epsilon}\frac{dJ_i}{dt}\left(\mathbf{C'}_\beta\right) = \frac{N_i}{\epsilon}\frac{da_i}{dt}\left(\mathbf{C'}_\beta\right),$$

where ϵ is the cost (in ATP units) to synthesis one cell. By assuming that the dilution (3^{rd}) term in Eq. F.2 is sufficiently small as to be neglected this cost simply becomes $\epsilon = \chi m$, and the rate of change of biomass in response to the metabolite perturbation can then be found as

$$\frac{dN_i}{dt}\left(\mathbf{C'}_\beta\right) = \frac{N_i}{\chi m}\frac{da_i}{dt}\left(\mathbf{C'}_\beta\right), \tag{F.3}$$

where N_i is the population of strain i. This allows us to quantify the impact of the perturbation of a specific metabolite β on the strains.

We also need to find the impact the specific strains have on the metabolite concentrations. We do this by finding the steady state nutrient concentrations, which for multiple strains are described by an extension of Eq. E.37 as

$$\frac{dC_\beta}{dt} = \kappa_\beta - \rho_\beta C_\beta + \sum_{i=1}^{B}\left(p_{i,\beta}(\mathbf{C}) - c_{i,\beta}(\mathbf{C})\right)N_i = 0, \tag{F.4}$$

where B is the total number of strains. If consumption (c) and production (p) rates are known the steady state value of metabolite β can be found as

$$C^*_\beta = \frac{\kappa_\beta + \sum_{i=1}^{B}\left(p_{i,\beta}(\mathbf{C}) - c_{i,\beta}(\mathbf{C})\right)N_i}{\rho_\beta}. \tag{F.5}$$

The instantaneous effect of a population perturbation on the concentration will occur due to the changed strain abundance, and the production and consumption rates will only change once the metabolite concentrations have begun to change in response to this. We therefore make the simplifying assumption that the production and consumption rates remain constant. We now calculate the steady-state metabolite concentration (under this assumption) as a proxy for the

strength of the instantaneous response as

$$C^*_\beta + \Delta C_{j,\beta} = \frac{\kappa_\beta + \sum_{i \neq j} \left(p_{i,\beta}(\mathbf{C}) - c_{i,\beta}(\mathbf{C})\right) N_i + \left(p_{j,\beta}(\mathbf{C}) - c_{j,\beta}(\mathbf{C})\right) \left(N_j + \Delta N_j\right)}{\rho_\beta}, \tag{F.6}$$

where ΔN_j is the population perturbation of strain j and $\Delta C_{j,\beta}$ is the resulting change in the steady-state value of metabolite β. By factoring out Eq. F.5 from Eq. F.6 the size of this change is found to be

$$\Delta C_{j,\beta} = \frac{\left(p_{j,\beta}(\mathbf{C}) - c_{j,\beta}(\mathbf{C})\right) \Delta N_j}{\rho_\beta}. \tag{F.7}$$

This concentration change will not be realised (as production and consumption rates adapt), but instead stands as a proxy for the strength of the metabolite's instantaneous response to the population perturbation.

The impact of one strain (j) upon another strain (i) can now be determined using the two metrics derived in Eqs. F.3 & F.7. If the population response of strain i (Eq. F.3) is positive for a positive perturbation ($\xi > 0$) the strain either solely or predominately uses the metabolite as a substrate to grow. Whereas if the response is negative, the strain is predominately inhibited by the metabolite as a waste product. When a positive perturbation of strain j's population ($\Delta N_j > 0$) results in a positive concentration change for metabolite β (Eq. F.7) the strain is a net producer, and if it's negative the strain is a net consumer. This behaviour can be used to classify four interaction types. They are 'competition' ($\Delta C_{j,\beta} < 0$ and $\frac{dN_i}{dt}(\mathbf{C'}_\beta) > 0$), 'facilitation' ($\Delta C_{j,\beta} > 0$ and $\frac{dN_i}{dt}(\mathbf{C'}_\beta) > 0$), 'syntrophy' ($\Delta C_{j,\beta} < 0$ and $\frac{dN_i}{dt}(\mathbf{C'}_\beta) < 0$), and 'pollution' ($\Delta C_{j,\beta} > 0$ and $\frac{dN_i}{dt}(\mathbf{C'}_\beta) < 0$). The strength of the impact of strain j upon strain i via metabolite β is given by

$$\left(\frac{dN_i}{dt}\right)_{j,\beta} = \frac{dN_i}{dt}(\mathbf{C'}_\beta)\frac{1}{\xi C_\beta}\Delta C_{j,\beta} \tag{F.8}$$

where the ξC_β term converts the concentration change ($\Delta C_{j,\beta}$) into units of the metabolite perturbation size ($|\mathbf{C'}_\beta|$). A positive interaction type implies a facilitation or syntrophy interaction, whereas a negative interaction strengths implies a competition or pollution interaction. The interaction strengths obtained from this expression will not be symmetric, i.e.

$|(\frac{dN_i}{dt})_{j,\beta}| \neq |(\frac{dN_j}{dt})_{i,\beta}|$. Additionally, having an interaction type between strain i and strain j does not necessarily determine the type of the converse interaction. Competition interactions are always paired, but pollution interactions only pair if both strains are sufficiently close to thermodynamic equilibrium, and facilitation only pairs with syntrophy when the facilitating strain is close to equilibrium. The total strength of interaction from strain j upon strain i can be calculated by summing over the metabolites as

$$\left(\frac{dN_i}{dt}\right)_j = \sum_{\beta=1}^{M} \left(\frac{dN_i}{dt}\right)_{j,\beta}, \tag{F.9}$$

where M is the total number of metabolites. This total interaction strength is not used in Chapter 3 as it doesn't allow for a full classification of interaction types.

F.2 Interactions between functional groups

In Chapter 3 (see Fig. 3.5), we assign species to functional groups based on the substrate of their most expressed reaction. It logically follows that competition interactions will mostly occur between members of the same functional group. In the case of facilitation a species breaks down a metabolite into a different metabolite which a second species breaks down. From this, we would expect facilitation interactions to (generally) occur between rather than within functional groups. By a similar logic, syntrophy interactions would also be expected to predominately occur within functional groups. Additionally, we would expect pollution interactions to frequently occur within functional groups as reactions in our model that share a substrate often share a waste-product. However, as waste-products can vary we would not expect this to be as exclusively the case as for competition interactions.

To test this hypothesis we produced heatmaps of both the total strength (Fig. F.1) and frequency (Fig. F.2) of the different interaction types across our simulations. Our general hypothesis holds with competition and pollution occurring generally within functional groups, and facilitation and syntrophy generally occurring between functional groups. This is observed regardless of whether the total strength or frequency of interactions is considered, though

the pattern is considerably stronger for interaction strength. For the low-free energy condition we consider, comparatively few surviving species or functional groups are observed (see Figs. 3.5 & 3.6). The functional groups that survive are also likely to be those using substrates higher in the metabolite hierarchy, i.e. those that become available earlier in the process of sequential substrate diversification. For this reason, we only show the first 12 functional groups in our plots as the other 12 occur only extremely rarely. The same effect also leads to functional group 1 being by far the most prevalent functional group. This means that most facilitation and syntrophy interactions involve functional group 1 (see Figs. F.1B-C and Figs. F.2B-C). When considering the frequency of interactions this means that interactions within the first functional group are almost always the most common, regardless of the broader pattern (see Fig. F.2). Finally, it is worth noting the mirroring that emerges between the facilitation and syntrophy. Regardless of whether total interaction strength (Figs. F.1B-C) or frequency (Figs. F.2B-C) is considered, facilitation is more prominent below the diagonal and syntrophy is more prominent above the diagonal, and there is a clear mirroring between the two interactions. Though the interaction strengths differ, each syntrophy interaction is clearly associated with a facilitation interaction.

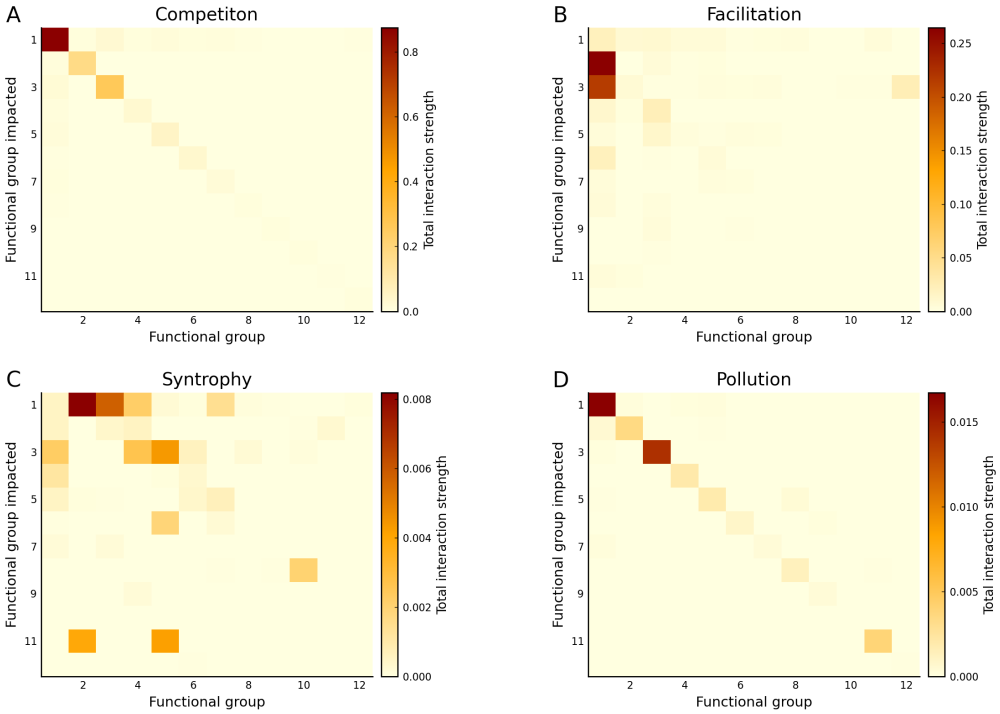


Figure F.1: **Heat maps of total interaction strength within and between functional groups.** The data were generated from our simulations for low-free energy (recalcitrant) substrates (1.5×10^6 J mol^{-1}) with fully reversible kinetics. (a) The total competition interaction strength is highest within functional groups, with the total strengths associated with interactions between different functional groups being negligible in almost every case. (b) Total facilitation interaction strengths are relatively higher between functional groups compared to within functional groups. (c) The relatively highest total interactions strengths are also between functional groups in the syntrophy case. (d) In the case of pollution total interaction strengths are relatively highest within functional groups, with almost negligible total strengths between functional groups.

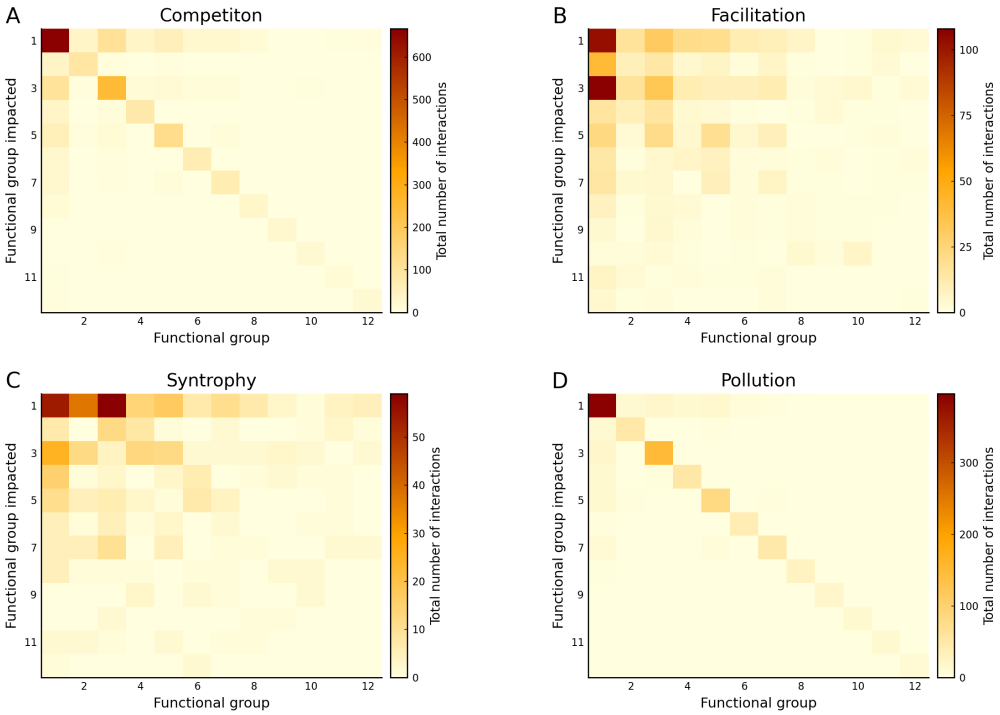


Figure F.2: **Heat maps of interaction frequency within and between functional groups.** The data were generated from our simulations for low-free energy (recalcitrant) substrates (1.5×10^6 J mol^{-1}) with fully reversible kinetics. (a) Almost all competition interactions are within the same functional groups. (b) Though there are significant facilitation interactions within functional group 1, in general there are more interactions between functional groups. (c) A similar pattern is observed for syntrophy with the majority of interactions being between functional groups. (d) In the case of pollution interactions occur almost solely within functional groups.

Appendix G

Analysis of assumptions underlying the model developed in Chapter 3

This appendix summarises all the key assumptions, implicit or explicit, underlying our model. We also address the choice of several key parameter values. Both of these things are contained along with our justifications for them in Tab. G.1. In a couple of cases the justification for our assumptions goes beyond what can be contained in a table as indicated. These are our choice of values for the two saturation constants $\gamma_{\frac{1}{2}}$ and $\Omega_{\frac{1}{2}}$, and the use of single-reactant single-product reactions. Both of these are discussed in further detail in separate subsections.

Assumption	Justification
Proteome can be partitioned into three compartments: ribosomes, metabolic proteins, and a fixed housekeeping fraction.	Assumption has been successfully used to model bacterial growth trade-offs, resulting in novel growth laws [SKMH14].

Only catabolic reactions need to be directly included in our model.	Model aims to investigate the effect of free-energy supply on ecosystem dynamics, catabolism is more impacted by this than anabolism.
All metabolic protein can be treated as contributing to catabolism.	A portion of the metabolic protein fraction could be assigned to anabolism reducing the energy acquisition rate. However, this change could instead be absorbed into a reduction of the maximum enzyme rate (k).
Metabolic protein required to sustain catabolism does not vary between species.	In the model biomass synthesis pathways do not vary between species. This means a species with a catabolism requiring greater protein investment (e.g. one with more thermodynamic bottlenecks), would receive no benefit for the additional cost.
Ideal ribosome fraction (Eq. 3.7) is of saturating functional form.	Required in order to recover the empirically observed growth laws (see Fig. E.1).
Effective elongation rate (Eq. 3.6) is of saturating functional form.	Maximum (per ribosome) translation rate has to be bounded.

Value of housekeeping proteome fraction (ϕ_Q) is fixed across species.	As no compensating mechanism exists (e.g. chemotaxis or motility [NCL+20]), species with a higher housekeeping fraction would only ever be disadvantaged.
Fixed values are used for the two (energy) saturation constants $\gamma_{\frac{1}{2}}$ and $\Omega_{\frac{1}{2}}$.	More detail provided later in this section
Cell use free-energy solely for protein translation.	Translation is one of the most most energy consuming cellular processes [LM15]. Many other growth associated costs can reasonably be expected to scale with protein translation. Other uses of energy (e.g. motility) beyond scope of the model.
The ATP cost per translation step χ does not vary between species or as cellular conditions change.	Simplifying assumption. As no benefit is attached to increased translation cost, allowing it to vary would simply disfavour strains with high χ.
Model considers only single step enzyme reactions.	Simplifying assumption. However, more complex reactions can be captured by species possessing multiple reactions in series.

Reactions can only descend one or two steps downwards in the metabolite hierarchy.	Necessary as the model is intended to capture simple one-step enzyme reactions.
Fraction of ribosomes bound and translating (f_b) does not vary with changing cell conditions.	The value used is taken from the literature [USP05]. Increasing f_b would lead to an increase in the growth rate, but would not change the overall results as this increased growth rate would also have to be met by an increased rate of free-energy acquisition.
Only considered single-reactant single-product reactions.	More detail provided later in this section.
Only external metabolites are modelled, e.g. no internal metabolite concentrations.	Reasonable first approximation. Thermodynamic limits will still apply as discussed in the Discussion section of the Chapter 3.
Constant temperature across all simulations.	The free-energy change caused by plausible temperature changes are very small compared to the variations in substrate free-energy we consider.

The metabolic protein mass (n_p) is the same for all reactions.	Simplifying assumption. Allowing this to vary would make some reactions less beneficial, but beyond reasonable scope of the paper to test this.
High free-energy change processes (e.g. protein translation) treated as irreversible.	Reversible and irreversible Michaelis–Menten kinetics are identical in the high dissipation (i.e. high free-energy change) limit.
Relative reaction expressions (see Eqs. E.31 & E.32) vary between species but don't change over time.	Computationally intensive to implement, reaction expression for the whole community can still shift through species sorting.
The rate at which the proteome fraction can shift (see Eq. 3.3) is set by a characteristic growth time scale τ_g (Eq. E.27).	Captures the fact that synthesised protein is retained for a significant period. τ_g is small enough that steady-state proteome fractions are reached.
Initial conditions (N_0, C_0, a_0, $\phi_{R,0}$) are the same across all simulations.	For all but extreme choices (which lead to all species going extinct) the choice of initial conditions does not effect the final steady states

Kinetic parameters (k, K_S, r) vary by approximately an order of magnitude around a fixed value.	Important to include kinetic variation. However, variation is kept relatively small to avoid kinetic variation dominating competition between strains.
For all simulations the same biomass loss rate (d) is used.	Changing this parameter alters the steady-state populations and the speed of ecosystem assembly. However, the increased population growth would be equal regardless of substrate free-energy and thus our results would still hold.
For all simulations the supply rate of the initial substrate κ is the same.	This is an alternative way of changing the rate of free-energy supply into the system. We kept this fixed in order to more straightforwardly investigate the impact of substrate free-energy.

Table G.1: Table summarising the assumptions made in for the model in Chapter 3, with justifications given in each case. Cases where more justification is required are marked as "More detail provided later in this section".

G.1 Choosing energy saturation constants

Our model contains two saturation constants that govern how cellular processes change as the internal energy concentration changes. They are the half-maximum constant for translational elongation rate $\gamma_{\frac{1}{2}}$ (see Eq. 3.6), and the half-maximum constant for 'ideal' ribosome fraction

$\Omega_{\frac{1}{2}}$ (see Eq. 3.7). Both of these parameters are highly phenomenological, thus we cannot obtain values for them from the existing literature. We therefore had to pick reasonable values, chosen under the following restrictions: The first restriction arises from the fact that the two saturation constants must be of a similar order of magnitude for the proteome model to function. If $\gamma_{\frac{1}{2}}$ is substantially larger, then cells will end up with a large ribosome fraction that only operates at a small fraction of the maximum rate. Whereas if $\Omega_{\frac{1}{2}}$ is substantially larger, ribosomes will always operate at maximum speed but only a vanishingly small fraction of the proteome will be dedicated to them. Neither scenario is consistent with the empirically observed growth laws. The second restriction is that the magnitude of the two saturation constants cannot be too high or too low. In the high value limit, cell populations would never increase due to being unable to achieve sufficient translation rates or ribosome fraction to match the biomass loss rate (d). In the low value limit, cells would rapidly reach maximum translation rate and ribosome fraction, they would show very fast growth but would not be able to sustain the very high energy demand associated with this growth rate, leading to them exhausting their internal energy and dying out. This would of course be biologically unreasonable.

Based on these restrictions the saturation constants must be of intermediate value, and of similar order of magnitude to each other. We specifically chose $\Omega_{\frac{1}{2}}$ to be slightly larger than $\gamma_{\frac{1}{2}}$ to ensure that translation saturates quicker than the ribosome fraction shifts, in order to avoid the situation where species frequently fail to grow due to possessing a large number of inefficient ribosomes. Beyond this the specific values chosen are fairly arbitrary. However, we do not believe that changes in the values of these constants would effect the broad results of the paper, provided pathological cases are avoided by keeping to the restrictions set out in the prior paragraph. The two fundamental constraints on growth in our model are the availability of energy and the number of ribosomes available for protein translation. Increasing energy availability will thus always benefit the cells regardless of the choice of saturation constant values, provided a non-pathological choice is made. Thus, our broad results on the effect of (free) energy availability on the assembly of ecosystems would hold provided all strains are given the same values for the saturation constants.

G.2 Assumption of single-reactant single-product reactions

In our model microbes can only make use of reactions that use a single reactant to produce a single product, these reactions are then assumed to be coupled to ATP production. This represents a significant reduction in complexity from the realistic scenario of microbes generally possessing reactions involving many reactants resulting in many products. We made this simplification in order to reduce the analytic and computational complexity of the model, this made our model both more interpretable and easier to simulate. Additionally, we do not believe that including reactions involving multiple reactants or products would weaken the thermodynamic results that we obtained. In fact we expect that thermodynamic effects would in general be more significant when reactions using multiple reactants or products are included.

To see the thermodynamic effects we should note that in the multiple-reactant multiple-product case the reaction quotient (for a particular reaction α) takes the form

$$Q(\boldsymbol{S}_\alpha, \boldsymbol{W}_\alpha) = \frac{\prod_{j=1}^{G_\alpha} (W_{j,\alpha})^{w_{j,\alpha}}}{\prod_{i=1}^{R_\alpha} (S_{i,\alpha})^{s_{i,\alpha}}}, \tag{G.1}$$

where $W_{j,\alpha}$ is the concentration of the j^{th} product, $S_{i,\alpha}$ is is the concentration of the i^{th} reactant, $w_{j,\alpha}$ is the product coefficient of product j, $s_{i,\alpha}$ is the reactant coefficient of reactant i, R_α is the number of distinct reactants involved in the reaction and G_α is the number of distinct products involved in the reaction. As mentioned in Chapter 3, in the single-reactant single product case this reduces to $Q(S_\alpha, W_\alpha) = W_\alpha / S_\alpha$. The value of the reaction quotient impacts the thermodynamic factor θ (Eq. 3.11), with higher values increasing the thermodynamic factor and lower values decreasing it. How close a reaction is to thermodynamic equilibrium is set by this thermodynamic factor, and therefore this is also (partly) determined by the value of the reaction quotient. When more metabolites are involved in the reaction the multiplicative form of the reaction quotient (see Eq. G.1) means that it varies more quickly with changing metabolite concentrations. So, while thermodynamic effects aren't guaranteed to be more important when considering reactions with multiple reactants and/or products, the variation in

metabolite concentrations required to make these thermodynamic effects significant is smaller.

Appendix H

Comparison of thermodynamic microbial model with other models

In this appendix, we will compare the model developed in Chapter 3 model with previous theoretical work in the area of microbial consumer-resource models, which can be classified into two broad topics which we address separately.

H.1 Previous work addressing violations of competitive exclusion theory

In nature a huge diversity of microbes is observed, despite the expectation from competitive exclusion theory that only one consumer species can survive per limiting resource [Lev70]. This apparent contradiction has been the subject of much historic discussion in the literature (exemplified by Hutchinson's "paradox of the plankton"; [Hut61]). Many potential resolutions to this contradiction have been proposed such as chaotic dynamics [HW99], temporal heterogeneity [DJG05], spatial heterogeneity [HvOW99], limitation by factors other than resource availability (e.g. predation) [RF75], or rapid evolution [YL21]. There are also some approaches that bear strong similarity to ours. The first of these we wish to discuss develops a thermodynamic

consumer-resource model based on irreversible Michaelis–Menten kinetics [GS16]. Using this model they observe that two species can coexist upon the same substrate, provided that they use different reactions (with different products) to do so. This represents a solution to the coexistence constraint predicted by the classical competitive exclusion principle, and suggests thermodynamic limitations can help explain the paradox, more broadly. In another approach that bears similarity to ours, a consumer-resource model was developed with a fixed enzyme budget that can be divided between metabolites, all of which can support cell growth and reproduction [PTW17]. Provided that all species have enzyme budgets of the same size and the resource supply rate are contained within the convex hull of metabolic strategies, an indefinite number of species can survive on just three supplied metabolites. It should however be noted (see [LLL$^+$20]), that this result only holds if both total enzyme budget and biomass loss rates are fixed between species. Additionally, this model only considers metabolite consumption and does not include the commonly observed production of metabolites by microbes.

The proteomic trade-off that our model includes represents a more detailed version of the previously mentioned metabolic trade-off [PTW17, LLL$^+$20]. Our more realistic "enzyme budget" is inversely proportional to the size of the ribosome fraction rather than being fixed. In addition to this, our model includes thermodynamic inhibition (similarly to [GS16]). This makes our framework a natural framework to test the results of the previously discussed approaches. Specifically, whether the increased diversity is retained over the longer run in larger and more realistic stochastically generated communities. We observe that the average number of survivors is lower than the average number of available substrates (see Fig. 3.5B), and that the number of survivors never exceeds the number of diversified substrates (see Fig. H.1). Our results suggest that although metabolic or thermodynamic trade-offs can offer a solution to the mismatch between competitive exclusion theory and ecological reality in special cases, this is not a general effect. This apparent contradiction is far more likely to be resolved by the spatial and temporal heterogeneity of real ecosystems.

Another recent approach to this problem involves incorporating a proteome trade-off into a consumer-resource model [PMSMG21]. In this consumer-resource model, metabolites are consumed but not produced by species, and every species splits its enzyme investment between

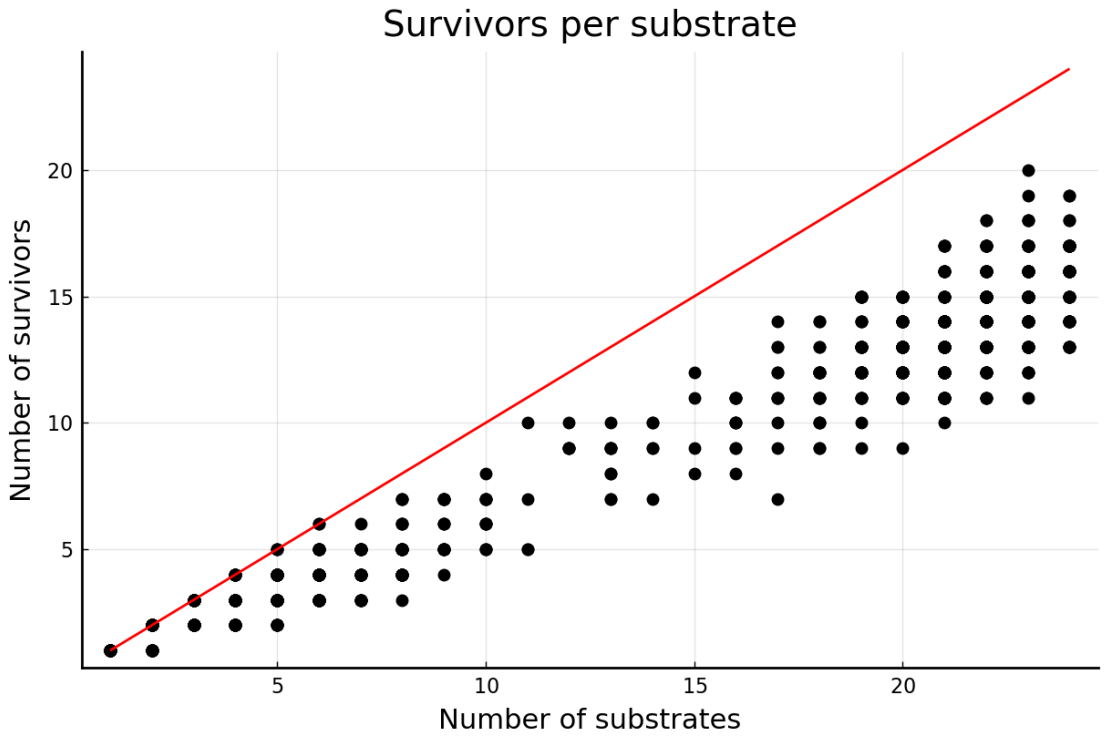


Figure H.1: **Test of competitive exclusion.** Plot of the number of surviving species against the total number of substrates diversified by the time steady state is reached. This shows the full data from across all four regimes considered in Chapter 3, e.g. 1000 simulation runs in total. In no case does the number of survivors exceed the number of substrates (equality indicated by the red line).

the different metabolites. However, this model fundamentally differs from the one presented in Posfai et al. [PTW17], in that the size of the enzyme budget changes as the size of the ribosome fraction changes, and the fraction of a species' enzyme budget dedicated to each metabolite is updated using an optimisation procedure. As previously mentioned our model captures the first of these effects, but in contrast, our model treats the proportional reaction expressions as fixed (see Eq. 3.9), meaning that overall reaction expression of the community can change only through species sorting. The reason we do not make use of an optimisation procedure is that it would be far more complex and computationally intensive to implement for a model including metabolite production.

Using their model, Pacciani-Mori et al. find more species can survive than there are available substrates [PMSMG21], provided that the biomass loss rate for each species is proportional to the proteome fraction allocated to both metabolic protein and ribosomes (i.e. the non-'housekeeping' fractions). This assumption does not require that different species must have the same biomass loss rate when they possess the same 'housekeeping' fraction, and so is a far less restrictive assumption than the one made by Posfai et al. In our model, all species have the same biomass loss rates and 'housekeeping' fractions so this condition is met. However, in contrast to our model it is implicitly assumed that all species are equally good competitors for the various substrates (i.e. equal uptake rates per unit protein investment in a particular substrate). Species are therefore selected for their overall metabolic strategies, rather than being intrinsically better competitors for particular substrates. Our model breaks this neutrality, leading to us never observing more surviving species than substrates at steady-state (see Fig. H.1).

H.2 Previous work addressing effect of energy supply on diversity

Previous theoretical work has addressed the link between energy supply and steady state microbial community diversity [MCG+19]. The model they develop is similar to our model. Their model incorporates microbial metabolite production in addition to consumption, with the level

of metabolite exchange between species being governed by the leakage parameter. A key difference from our model is that they do not introduce a metabolite (free-energy) hierarchy and so the supplied metabolite can generally become any other metabolite through a small number of conversions. This means that full substrate diversification occurs more quickly, leading to a more rapid ecosystem assembly compared to our model (compare Fig. 3.4 with Fig. 1E in [MCG$^+$19]). In this work, energetic limitations on the diversity emerge based on the total amount of energy leaked by species (in the form of metabolites), with this leaked energy allowing more intermediate substrates to become sufficiently available to viably support further species. Maximal diversity is observed when both the energy supply is large and species leak the majority of the energy they uptake back to the environment.

This result is broadly consistent with the results we found (see Fig. 3.5), though the leakage parameter has no obvious interpretation in our framework. However, our results relating to the impact of (free-) energy availability on microbial diversity differ from this previous work as we establish and investigate a mechanism for this effect. We find that the relationship between (free-) energy availability and steady state diversity is intrinsically linked to the process of ecosystem assembly. Low free-energy availability leads to a slower growth rate and thus a slower rate of substrate diversification (see Fig. 3.6). As we considered a top-down assembly procedure (one without additional immigration events), the number of niches (substrates) that are generated before species from the initially inoculated community reach extinction significantly determines the final diversity. From this it can be seen that the availability of free energy sets the timescale of ecosystem development. This is a useful result as this timescale can be compared with those of other processes (e.g. species immigration, disturbance, etc), in order to establish possible inter-linkages between them

Bibliography

[ABM18] Madhu Advani, Guy Bunin, and Pankaj Mehta. Statistical physics of community ecology: a cavity solution to MacArthur's consumer resource model. *Journal of Statistical Mechanics: Theory and Experiment*, 2018(3):033406, 2018.

[AM17] Michael Assaf and Baruch Meerson. WKB theory of large deviations in stochastic populations. *Journal of Physics A: Mathematical and Theoretical*, 50(26):263001, 2017.

[Arn00] Peter Arnold. Symmetric path integrals for stochastic equations with multiplicative noise. *Physical Review E*, 61(6):6099–6102, June 2000.

[AYZP21] Amir Akbari, James T. Yurkovich, Daniel C. Zielinski, and Bernhard O. Palsson. The quantitative metabolome is shaped by abiotic constraints. *Nature Communications*, 12(1):3178, December 2021.

[BA53] Robert M. Bock and Robert A. Alberty. Studies of the Enzyme Fumarase. I. Kinetics and Equilibrium. *Journal of the American Chemical Society*, 75(8):1921–1925, April 1953.

[BBF+16] Christopher Battle, Chase P. Broedersz, Nikta Fakhri, Veikko F. Geyer, Jonathon Howard, Christoph F. Schmidt, and Fred C. MacKintosh. Broken detailed balance at mesoscopic scales in active biological systems. *Science*, 352(6285):604–607, April 2016.

[BD96] Hans Bremer and Patrick P Dennis. Modulation of Chemical Composition and Other Parameters of the Cell by Growth Rate. In Frederick C. Neidhardt and

Roy Curtiss, editors, *Escherichia coli and Salmonella: cellular and molecular biology*, pages 1553–1569. ASM Press, Washington, D.C, 2nd ed edition, 1996.

[BEO+09] Florian Brandt, Stephanie A. Etchells, Julio O. Ortiz, Adrian H. Elcock, F. Ulrich Hartl, and Wolfgang Baumeister. The Native 3D Organization of Bacterial Polysomes. *Cell*, 136(2):261–271, January 2009.

[BFC17] Rogier Braakman, Michael J. Follows, and Sallie W. Chisholm. Metabolic evolution and the self-organization of ecosystems. *Proceedings of the National Academy of Sciences*, 114(15):E3091–E3100, April 2017.

[BGA+04] James H. Brown, James F. Gillooly, Andrew P. Allen, Van M. Savage, and Geoffrey B. West. Toward a Metabolic Theory of Ecology. *Ecology*, 85(7):1771–1789, July 2004.

[BHO+15] Markus Basan, Sheng Hui, Hiroyuki Okano, Zhongge Zhang, Yang Shen, James R. Williamson, and Terence Hwa. Overflow metabolism in Escherichia coli results from efficient proteome allocation. *Nature*, 528(7580):99–104, December 2015.

[BKM+21] Sonja Blasche, Yongkyu Kim, Ruben A. T. Mars, Daniel Machado, Maria Maansson, Eleni Kafkia, Alessio Milanese, Georg Zeller, Bas Teusink, Jens Nielsen, Vladimir Benes, Rute Neves, Uwe Sauer, and Kiran Raosaheb Patil. Metabolic cooperation and spatiotemporal niche partitioning in a kefir microbial community. *Nature Microbiology*, January 2021.

[Boh33] N. Bohr. Light and Life*. *Nature*, 131(3308):421–423, March 1933.

[BR16] Freddy Bouchet and Julien Reygner. Generalisation of the Eyring–Kramers Transition Rate Formula to Irreversible Diffusion Processes. *Annales Henri Poincaré*, 17(12):3499–3532, December 2016.

[BS15] Andre C. Barato and Udo Seifert. Thermodynamic Uncertainty Relation for Biomolecular Processes. *Physical Review Letters*, 114(15):158101, April 2015.

[BT13] Wolfgang Buckel and Rudolf K. Thauer. Energy conservation via electron bifurcating ferredoxin reduction and proton/Na+ translocating ferredoxin oxidation. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1827(2):94–113, February 2013.

[BTK13] G. Baris Bagci, Ugur Tirnakli, and Juergen Kurths. Second law for transitions between nonequilibrium steady states. *Physical Review E*, 87(3), March 2013.

[CA00] JOSHUA L Cherry and FREDERICK R Adler. How to make a biological switch. *Journal of Theoretical Biology*, 203(2):117–133, March 2000.

[CCJ06] Vladimir Y Chernyak, Michael Chertkov, and Christopher Jarzynski. Path-integral analysis of fluctuation theorems for general Langevin processes. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(08):P08001–P08001, August 2006.

[CE20] Jacob Cook and Robert G. Endres. Thermodynamics of switching in multistable non-equilibrium systems. *The Journal of Chemical Physics*, 152(5):054108, February 2020.

[CGM+17] Ke Chen, Ye Gao, Nathan Mih, Edward J. O'Brien, Laurence Yang, and Bernhard O. Palsson. Thermosensitivity of growth is determined by chaperone-mediated proteome reallocation. *Proceedings of the National Academy of Sciences*, 114(43):11548–11553, October 2017.

[CMM20] Wenping Cui, Robert Marsland, and Pankaj Mehta. Effect of Resource Dynamics on Species Packing in Diverse Ecosystems. *Physical Review Letters*, 125(4):048101, July 2020.

[COM+19] Chuankai Cheng, Edward J. O'Brien, Douglas McCloskey, Jose Utrilla, Connor Olson, Ryan A. LaCroix, Troy E. Sandberg, Adam M. Feist, Bernhard O. Palsson, and Zachary A. King. Laboratory evolution reveals a two-dimensional rate-yield tradeoff in microbial metabolism. *PLOS Computational Biology*, 15(6):e1007066, June 2019.

[CPE21] Jacob Cook, Samraat Pawar, and Robert G. Endres. Thermodynamic constraints on the assembly and diversity of microbial ecosystems are different near to and far from equilibrium. *PLOS Computational Biology*, 17(12):e1009643, December 2021.

[CSP21] Luca Cocconi, Guillaume Salbreux, and Gunnar Pruessner. Scaling of Entropy Production under Coarse Graining. *arXiv:2109.05288 [cond-mat, physics:math-ph]*, September 2021. arXiv: 2109.05288.

[DCWS20] Hadrien Delattre, Jing Chen, Matthew J. Wade, and Orkun S. Soyer. Thermodynamic modelling of synthetic communities predicts minimum free energy requirements for sulfate reduction and methanogenesis. *Journal of The Royal Society Interface*, 17(166):20200053, May 2020.

[Dew03] Roderick Dewar. Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states. *Journal of Physics A: Mathematical and General*, 36(3):631, 2003.

[Dew05] R. C. Dewar. Maximum entropy production and the fluctuation theorem. *Journal of Physics A: Mathematical and General*, 38(21):L371, 2005.

[Dew09] Roderick Dewar. Maximum Entropy Production as an Inference Algorithm that Translates Physical Assumptions into Macroscopic Predictions: Don't Shoot the Messenger. *Entropy*, 11(4), November 2009.

[Dew10] Roderick C. Dewar. Maximum entropy production and plant optimization theories. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365(1545):1429–1435, May 2010.

[DGL+17] J. P. DeLong, J. P. Gibert, T. M. Luhring, G. Bachman, B. Reed, A. Neyer, and K. L. Montooth. The combined effects of reactant kinetics and enzyme stability explain the temperature dependence of metabolic rates. *Ecology and Evolution*, 7(11):3940–3950, June 2017.

[dJFRRI12] Manuel del Jesus, Romano Foti, Andrea Rinaldo, and Ignacio Rodriguez-Iturbe. Maximum entropy production, carbon assimilation, and the spatial organization of vegetation in river basins. *Proceedings of the National Academy of Sciences of the United States of America*, 109(51):20837–20841, December 2012.

[DJG05] Blandine Descamps-Julien and Andrew Gonzalez. Stable coexistence in a fluctuating environment: an experimental demonstration. *Ecology*, 86(10):2815–2824, October 2005.

[dlCPCG+18] Roberto de la Cruz, Ruben Perez-Carrasco, Pilar Guerrero, Tomas Alarcon, and Karen M. Page. Minimum Action Path Theory Reveals the Details of Stochastic Transitions Out of Oscillatory States. *Physical Review Letters*, 120(12), March 2018.

[DOM+10] J. P. DeLong, J. G. Okie, M. E. Moses, R. M. Sibly, and J. H. Brown. Shifts in metabolic scaling, production, and efficiency across major evolutionary transitions of life. *Proceedings of the National Academy of Sciences*, 107(29):12941–12945, July 2010.

[DSG+16] Manoshi S. Datta, Elzbieta Sliwerska, Jeff Gore, Martin F. Polz, and Otto X. Cordero. Microbial interactions lead to rapid micro-scale successions on model marine particles. *Nature Communications*, 7(1):11965, September 2016.

[dVB11] Harold P. de Vladar and Nicholas H. Barton. The contribution of statistical physics to evolutionary biology. *Trends in Ecology & Evolution*, 26(8):424–432, August 2011.

[EDS+19] Tim N. Enke, Manoshi S. Datta, Julia Schwartzman, Nathan Cermak, Désirée Schmitz, Julien Barrere, Alberto Pascual-García, and Otto X. Cordero. Modular Assembly of Polysaccharide-Degrading Marine Microbial Communities. *Current biology: CB*, 29(9):1528–1535.e6, May 2019.

[End15] Robert G. Endres. Bistability: requirements on cell-volume, protein diffusion, and thermodynamics. *PLOS ONE*, 10(4):e0121681, April 2015.

[End17] Robert G. Endres. Entropy production selects nonequilibrium states in multistable systems. *Scientific Reports*, 7(1):14437, October 2017.

[Eng13] Jeremy L. England. Statistical physics of self-replication. *The Journal of Chemical Physics*, 139(12):121923, September 2013.

[Eng15] Jeremy L. England. Dissipative adaptation in driven self-assembly. *Nature Nanotechnology*, 10(11):919–923, November 2015.

[ES02] Denis J. Evans and Debra J. Searles. The Fluctuation Theorem. *Advances in Physics*, 51(7):1529–1585, November 2002.

[FB12] Kevin R. Foster and Thomas Bell. Competition, Not Cooperation, Dominates Interactions among Culturable Microbial Species. *Current Biology*, 22(19):1845–1850, October 2012.

[FLMS13] Serita D. Frey, Juhwan Lee, Jerry M. Melillo, and Johan Six. The temperature response of soil microbial efficiency and its feedback to climate. *Nature Climate Change*, 3(4):395–398, April 2013.

[FLR+12] Noah Fierer, Christian L Lauber, Kelly S Ramirez, Jesse Zaneveld, Mark A Bradford, and Rob Knight. Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. *The ISME Journal*, 6(5):1007–1017, May 2012.

[FLVC+19] Stilianos Fodelianakis, Alexander Lorz, Adriana Valenzuela-Cuevas, Alan Barozzi, Jenny Marie Booth, and Daniele Daffonchio. Dispersal homogenizes communities via immigration even at low rates in a simplified synthetic bacterial metacommunity. *Nature Communications*, 10(1):1314, December 2019.

[FNBE+13] A. Flamholz, E. Noor, A. Bar-Even, W. Liebermeister, and R. Milo. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proceedings of the National Academy of Sciences*, 110(24):10039–10044, June 2013.

[FP21] Yong Fan and Oluf Pedersen. Gut microbiota in human metabolic health and disease. *Nature Reviews Microbiology*, 19(1):55–71, January 2021.

[FSB05] A. Fiasconaro, B. Spagnolo, and S. Boccaletti. Signatures of noise-enhanced stability in metastable states. *Physical Review E*, 72(6):061110, December 2005.

[FW20] Xiaona Fang and Jin Wang. Nonequilibrium Thermodynamics in Cell Biology: Extending Equilibrium Formalism to Cover Living Systems. *Annual Review of Biophysics*, 49(1):227–246, May 2020.

[FZW14] Haidong Feng, Kun Zhang, and Jin Wang. Non-equilibrium transition state rate theory. *Chem. Sci.*, 5(10):3761–3769, 2014.

[Gas04] Pierre Gaspard. Fluctuation theorem for nonequilibrium reactions. *The Journal of Chemical Physics*, 120(19):8898–8905, April 2004.

[Gil77] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, December 1977.

[Gil00] Daniel T. Gillespie. The chemical Langevin equation. *The Journal of Chemical Physics*, 113(1):297–306, July 2000.

[GL07] G. Grinstein and R. Linsker. Comments on a derivation and application of the 'maximum entropy production' principle. *Journal of Physics A: Mathematical and Theoretical*, 40(31):9717, 2007.

[GLB+18] Joshua E. Goldford, Nanxi Lu, Djordje Bajić, Sylvie Estrela, Mikhail Tikhonov, Alicia Sanchez-Gorostiaga, Daniel Segrè, Pankaj Mehta, and Alvaro Sanchez. Emergent simplicity in microbial community assembly. *Science*, 361(6401):469–474, August 2018.

[GM18] Akshit Goyal and Sergei Maslov. Diversity, Stability, and Reproducibility in Stochastically Assembled Microbial Ecosystems. *Physical Review Letters*, 120(15):158102, April 2018.

[GS16] Tobias Großkopf and Orkun S. Soyer. Microbial diversity arising from thermodynamic constraints. *The ISME Journal*, 10(11):2725, November 2016.

[GTJ+11] Robert I. Griffiths, Bruce C. Thomson, Phillip James, Thomas Bell, Mark Bailey, and Andrew S. Whiteley. The bacterial biogeography of British soils: Mapping soil bacteria. *Environmental Microbiology*, 13(6):1642–1654, June 2011.

[GTS11] Ramon Grima, Philipp Thomas, and Arthur V. Straube. How accurate are the nonlinear chemical Fokker-Planck and chemical Langevin equations? *The Journal of Chemical Physics*, 135(8):084103, August 2011.

[HCR96] C. Y. Hoh and R. Cord-Ruwisch. A practical kinetic model that considers end-product inhibition in anaerobic digestion processes by including the equilibrium constant. *Biotechnology and Bioengineering*, 51(5):597–604, September 1996.

[HE17] Jordan M. Horowitz and Jeremy L. England. Spontaneous fine-tuning to environment in many-species chemical reaction networks. *Proceedings of the National Academy of Sciences*, 114(29):7565–7570, July 2017.

[HGTT84] Peter Hanggi, Hermann Grabert, Peter Talkner, and Harry Thomas. Bistable systems: Master equation versus Fokker-Planck modeling. *Physical Review A*, 29(1):371–378, January 1984.

[Hil89] Terrell L Hill. *Free energy transduction and biochemical cycle kinetics*. Springer-Verlag, New York, 1989. OCLC: 903749690.

[HN14] John Harte and Erica A. Newman. Maximum information entropy: a foundation for ecological theory. *Trends in Ecology & Evolution*, 29(7):384–389, July 2014.

[HS01] Takahiro Hatano and Shin-ichi Sasa. Steady-State Thermodynamics of Langevin Systems. *Physical Review Letters*, 86(16):3463–3466, April 2001.

[Hut61] G. E. Hutchinson. The Paradox of the Plankton. *The American Naturalist*, 95(882):137–145, May 1961.

[HVE08a] Matthias Heymann and Eric Vanden-Eijnden. The geometric minimum action method: A least action principle on the space of curves. *Communications on Pure and Applied Mathematics*, 61(8):1052–1117, August 2008.

[HVE08b] Matthias Heymann and Eric Vanden-Eijnden. Pathways of maximum likelihood for rare events in nonequilibrium systems: application to nucleation in the presence of shear. *Physical Review Letters*, 100(14), April 2008.

[HvOW99] Jef Huisman, Paul van Oostveen, and Franz J. Weissing. Species Dynamics in Phytoplankton Blooms: Incomplete Mixing and Competition for Light. *The American Naturalist*, 154(1):46–68, July 1999.

[HW99] Jef Huisman and Franz J. Weissing. Biodiversity of plankton by species oscillations and chaos. *Nature*, 402(6760):407–410, November 1999.

[Jar97] C. Jarzynski. Nonequilibrium Equality for Free Energy Differences. *Physical Review Letters*, 78(14):2690–2693, April 1997.

[Jay57] E. T. Jaynes. Information Theory and Statistical Mechanics. *Physical Review*, 106(4):620–630, May 1957.

[Jay80] E. T. Jaynes. The Minimum Entropy Production Principle. *Annual Review of Physical Chemistry*, 31:579–601, October 1980.

[JGT18] John C. Neu, Akhil Ghanta, and Stephen Teitsworth. The geometry of most probable trajectories in noise-driven dynamical systems. In *Coupled mathematical models for physical and biological nanoscale systems and their applications*, volume 232 of *Springer Proceedings in Mathematics & Statistics*. Springer International Publishing, Cham, 2018.

[KCVSZ+11] I. M. Keseler, J. Collado-Vides, A. Santos-Zavaleta, M. Peralta-Gil, S. Gama-Castro, L. Muniz-Rascado, C. Bonavides-Martinez, S. Paley, M. Krummenacker, T. Altman, P. Kaipa, A. Spaulding, J. Pacheco, M. Latendresse, C. Fulcher, M. Sarker, A. G. Shearer, A. Mackie, I. Paulsen, R. P. Gunsalus, and P. D.

Karp. EcoCyc: a comprehensive database of Escherichia coli biology. *Nucleic Acids Research*, 39(Database):D583–D590, January 2011.

[Kin95] Sharon E. Kingsland. *Modeling nature: episodes in the history of population ecology*. Science and its conceptual foundations. University of Chicago Press, Chicago, 2nd ed edition, 1995.

[Kin15] Sharon Kingsland. Alfred J. Lotka and the origins of theoretical population ecology. *Proceedings of the National Academy of Sciences*, 112(31):9493–9495, August 2015.

[KL13] Konstantinos P. Koutsoumanis and Alexandra Lianou. Stochasticity in Colonial Growth Dynamics of Individual Bacterial Cells. *Applied and Environmental Microbiology*, 79(7):2294–2301, April 2013.

[KMK18] Marcel M. M. Kuypers, Hannah K. Marchant, and Boran Kartal. The microbial nitrogen-cycling network. *Nature Reviews Microbiology*, 16(5):263–276, May 2018.

[KP14] Dilip Kondepudi and Ilya Prigogine. *Modern Thermodynamics: From Heat Engines to Dissipative Structures*. Wiley, 1 edition, November 2014.

[KSBP20] Dimitrios—Georgios Kontopoulos, Thomas P. Smith, Timothy G. Barraclough, and Samraat Pawar. Adaptive evolution shapes the present-day distribution of the thermal sensitivity of population growth rate. *PLOS Biology*, 18(10):e3000894, October 2020.

[KvBW+17] Christopher P. Kempes, Peter M. van Bodegom, David Wolpert, Eric Libby, Jan Amend, and Tori Hoehler. Drivers of Bacterial Maintenance and Minimal Energy Requirements. *Frontiers in Microbiology*, 8, January 2017.

[LACC+21] Pablo Lechón-Alonso, Tom Clegg, Jacob Cook, Thomas P. Smith, and Samraat Pawar. The role of competition versus cooperation in microbial community coalescence. *PLOS Computational Biology*, 17(11):e1009584, November 2021.

[LBB16] Diane Lawrence, Thomas Bell, and Timothy G. Barraclough. The Effect of Immigration on the Adaptation of Microbial Communities to Warming. *The American Naturalist*, 187(2):236–248, February 2016.

[Lev70] Simon A. Levin. Community Equilibria and Stability, and an Extension of the Competitive Exclusion Principle. *The American Naturalist*, 104(939):413–423, September 1970.

[LFMM14] Alex H. Lang, Charles K. Fisher, Thierry Mora, and Pankaj Mehta. Thermodynamics of Statistical Inference by Cells. *Physical Review Letters*, 113(14), October 2014.

[LLL+20] Zhiyuan Li, Bo Liu, Sophia Hsin-Jung Li, Christopher G. King, Zemer Gitai, and Ned S. Wingreen. Modeling microbial metabolic trade-offs in a chemostat. *PLOS Computational Biology*, 16(8):e1008156, August 2020.

[LLMB17] Salvador Lladó, Rubén López-Mondéjar, and Petr Baldrian. Forest Soil Bacteria: Diversity, Involvement in Ecosystem Processes, and Response to Global Change. *Microbiology and Molecular Biology Reviews*, 81(2):e00063–16, e00063–16, June 2017.

[LLWM01] Ralph D. Lorenz, Jonathan I. Lunine, Paul G. Withers, and Christopher P. McKay. Titan, Mars and Earth : Entropy production by latitudinal heat transport. *Geophysical Research Letters*, 28(3):415–418, February 2001.

[LM15] Michael Lynch and Georgi K. Marinov. The bioenergetic costs of a gene. *Proceedings of the National Academy of Sciences*, 112(51):15690–15695, December 2015.

[Lot22a] Alfred J. Lotka. Contribution to the Energetics of Evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 8(6):147–151, June 1922.

[Lot22b] Alfred J. Lotka. Natural Selection as a Physical Principle. *Proceedings of the National Academy of Sciences of the United States of America*, 8(6):151–154, June 1922.

[Lot25] Alfred James Lotka. *Elements of Physical Biology*. Williams & Wilkins, 1925.

[LWMX20] Chen Liao, Tong Wang, Sergei Maslov, and Joao B. Xavier. Modeling microbial cross-feeding at intermediate scale portrays community dynamics and species coexistence. *PLOS Computational Biology*, 16(8):e1008135, August 2020.

[LWvB+19] T.A. Lynch, Y. Wang, B. van Brunt, D. Pacheco, and P.H. Janssen. Modelling thermodynamic feedback on the metabolism of hydrogenotrophic methanogens. *Journal of Theoretical Biology*, page S0022519319302164, May 2019.

[Mac70] R. MacArthur. Species packing and competitive equilibrium for many species. *Theoretical Population Biology*, 1(1):1–11, May 1970.

[MATP18] Daniel Machado, Sergej Andrejev, Melanie Tramontano, and Kiran Raosaheb Patil. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Research*, 46(15):7542–7553, September 2018.

[May01] Robert M. May. *Stability and complexity in model ecosystems*. Princeton landmarks in biology. Princeton University Press, Princeton, 1st princeton landmarks in biology ed edition, 2001.

[MB10] Filip J. R. Meysman and Stijn Bruers. Ecosystem functioning and maximum entropy production: a quantitative test of hypotheses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1545):1405–1416, May 2010.

[MCG+19] Robert Marsland, Wenping Cui, Joshua Goldford, Alvaro Sanchez, Kirill Korolev, and Pankaj Mehta. Available energy fluxes drive a transition in the diversity, stability, and functional structure of microbial communities. *PLOS Computational Biology*, 15(2):e1006793, February 2019.

[MCM20] Robert Marsland, Wenping Cui, and Pankaj Mehta. A minimal model for microbial biodiversity can reproduce experimentally observed ecological patterns. *Scientific Reports*, 10(1):3308, December 2020.

[Mic05] K. Michaelian. Thermodynamic stability of ecosystems. *Journal of Theoretical Biology*, 237(3):323–335, December 2005.

[Mic11] K. Michaelian. Thermodynamic dissipation theory for the origin of life. *Earth Syst. Dynam.*, 2(1):37–51, March 2011.

[Mic12] K. Michaelian. Thermodynamic Function of Life. *Hydrology and Earth System Sciences*, 16(8):2629–2645, August 2012. arXiv: 0907.0040.

[MJ79] Henning Mejer and Sven Erik Jørgensen. Exergy and Ecological Buffer Capacity. In *State-of-the-Art in Ecological Modelling*, pages 829–846. Elsevier, 1979.

[ML10] Ville Mustonen and Michael Lässig. Fitness flux and ubiquity of adaptive evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 107(9):4248–4253, March 2010.

[MLB+12] J. Mehl, B. Lander, C. Bechinger, V. Blickle, and U. Seifert. Role of hidden slow degrees of freedom in the fluctuation theorem. *Physical Review Letters*, 108(22):220601, June 2012.

[MM12] Riccardo Mannella and Peter V. E. McClintock. Itô versus Stratonovich: 30 years later. *Fluctuation and Noise Letters*, 11(01):1240010, March 2012.

[MMA+21] Daniel Machado, Oleksandr M. Maistrenko, Sergej Andrejev, Yongkyu Kim, Peer Bork, Kaustubh R. Patil, and Kiran R. Patil. Polarization of microbial communities between competitive and cooperative metabolism. *Nature Ecology & Evolution*, January 2021.

[MMB17] M. Malek Mansour and F. Baras. Fluctuation theorem: A critical review. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(10):104609, October 2017.

[Mon49] J Monod. The Growth of Bacterial Cultures. *Annual Review of Microbiology*, 3(1):371–394, October 1949.

[NCL+20] Bin Ni, Remy Colin, Hannes Link, Robert G. Endres, and Victor Sourjik. Growth-rate dependent resource investment in bacterial motile behavior quantitatively follows potential benefit of chemotaxis. *Proceedings of the National Academy of Sciences*, 117(1):595–601, January 2020.

[NFBE+16] Elad Noor, Avi Flamholz, Arren Bar-Even, Dan Davidi, Ron Milo, and Wolfram Liebermeister. The Protein Cost of Metabolic Fluxes: Prediction from Enzymatic Rate Laws and Cost Minimization. *PLOS Computational Biology*, 12(11):e1005167, November 2016.

[NLH19] Bastian Niebel, Simeon Leupold, and Matthias Heinemann. An upper limit on Gibbs energy dissipation governs cellular metabolism. *Nature Metabolism*, 1(1):125–132, January 2019.

[NSB+20] Dan Naylor, Natalie Sadler, Arunima Bhattacharjee, Emily B. Graham, Christopher R. Anderton, Ryan McClure, Mary Lipton, Kirsten S. Hofmockel, and Janet K. Jansson. Soil Microbiomes Under Climate Change and Implications for Carbon Cycling. *Annual Review of Environment and Resources*, 45(1):29–59, October 2020.

[NSR20] Naoise Nunan, Hannes Schmidt, and Xavier Raynaud. The ecology of heterogeneity: soil bacterial communities and C dynamics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1798):20190249, May 2020.

[Odu57] Howard T. Odum. Trophic Structure and Productivity of Silver Springs, Florida. *Ecological Monographs*, 27(1):55–112, January 1957.

[Odu88] H. T. Odum. Self-Organization, Transformity, and Information. *Science*, 242(4882):1132–1139, November 1988.

[Odu07] Howard T. Odum. *Environment, power, and society for the twenty-first century: the hierarchy of energy*. Columbia University Press, New York, 2007.

[Ons31] Lars Onsager. Reciprocal Relations in Irreversible Processes. I. *Physical Review*, 37(4):405–426, February 1931.

[OP55] Howard T. Odum and Richard C. Pinkerton. Time's Speed Regulator: The Optimum Efficiency for Maximum Power Output in Physical and Biological Systems. *American Scientist*, 43(2):331–343, 1955.

[Pal75] G. W. Paltridge. Global dynamics and climate - a system of minimum entropy exchange. *Quarterly Journal of the Royal Meteorological Society*, 101(429):475–484, July 1975.

[PAS06] Steven J. Phillips, Robert P. Anderson, and Robert E. Schapire. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3-4):231–259, January 2006.

[PB02] Thomas Pfeiffer and Sebastian Bonhoeffer. Evolutionary Consequences of Tradeoffs between Yield and Rate of ATP Production. *Zeitschrift für Physikalische Chemie*, 216(1):51, 2002.

[PB04] Thomas Pfeiffer and Sebastian Bonhoeffer. Evolution of Cross-Feeding in Microbial Populations. *The American Naturalist*, 163(6):E126–E135, June 2004.

[PCGBP16] Ruben Perez-Carrasco, Pilar Guerrero, James Briscoe, and Karen M. Page. Intrinsic noise profoundly alters the dynamics and steady state of morphogen-controlled bistable genetic switches. *PLOS Computational Biology*, 12(10):e1005154, October 2016.

[PGBB20] Alberto Pascual-García, Sebastian Bonhoeffer, and Thomas Bell. Metabolically cohesive microbial consortia and ecosystem functioning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1798):20190245, May 2020.

[PME16] Nikolai Perunov, Robert Marsland, and Jeremy England. Statistical Physics of Adaptation. *Physical Review X*, 6(2), June 2016. arXiv: 1412.1875.

[PMSMG21] Leonardo Pacciani-Mori, Samir Suweis, Amos Maritan, and Andrea Giometto. Constrained proteome allocation affects coexistence in models of competitive microbial communities. *The ISME Journal*, 15(5):1458–1477, May 2021.

[PS04] P. B. Price and T. Sowers. Temperature dependence of metabolic rates for microbial growth, maintenance, and survival. *Proceedings of the National Academy of Sciences*, 101(13):4631–4636, March 2004.

[PSB01] T. Pfeiffer, S. Schuster, and S. Bonhoeffer. Cooperation and Competition in the Evolution of ATP-Producing Pathways. *Science*, 292(5516):504–507, April 2001.

[PT20] William D. Piñeros and Tsvi Tlusty. Kinetic proofreading and the limits of thermodynamic uncertainty. *Physical Review E*, 101(2):022415, February 2020.

[PTW17] Anna Posfai, Thibaud Taillefumier, and Ned S. Wingreen. Metabolic Trade-Offs Promote Diversity in a Model Ecosystem. *Physical Review Letters*, 118(2):028103, January 2017.

[RF75] Jonathan Roughgarden and Marcus Feldman. Species packing and predation pressure. *Ecology*, 56(2):489–492, March 1975.

[RND+15] Édgar Roldán, Izaak Neri, Meik Dörpinghaus, Heinrich Meyr, and Frank Jülicher. Decision Making in the Arrow of Time. *Physical Review Letters*, 115(25):250602, December 2015.

[RSC+16] Damian W. Rivett, Thomas Scheuerl, Christopher T. Culbert, Shorok B. Mombrikotb, Emma Johnstone, Timothy G. Barraclough, and Thomas Bell. Resource-dependent attenuation of species interactions during bacterial succession. *The ISME journal*, 10(9):2259–2268, September 2016.

[RSS16] Benjamin R. K. Roller, Steven F. Stoddard, and Thomas M. Schmidt. Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. *Nature Microbiology*, 1(11):16160, November 2016.

[Rue16] David Ruelle. A generalized detailed balance relation. *Journal of Statistical Physics*, 164(3):463–471, August 2016.

[SA21] Carlos A. Serván and Stefano Allesina. Tractable models of ecological assembly. *Ecology Letters*, 24(5):1029–1037, May 2021.

[SBB+09] Marten Scheffer, Jordi Bascompte, William A. Brock, Victor Brovkin, Stephen R. Carpenter, Vasilis Dakos, Hermann Held, Egbert H. van Nes, Max Rietkerk, and George Sugihara. Early-warning signals for critical transitions. *Nature*, 461(7260):53–59, September 2009.

[SCBP21] Thomas P. Smith, Tom Clegg, Thomas Bell, and Samrāt Pawar. Systematic variation in the temperature dependence of bacterial carbon use efficiency. *Ecology Letters*, page ele.13840, July 2021.

[Sch44] Erwin Schrödinger. *What is Life? The Physical Aspect of the Living Cell*. Cambridge University Press, 1944.

[Sch72] F. Schlögl. Chemical reaction models for non-equilibrium phase transitions. *Zeitschrift für Physik*, 253(2):147–161, April 1972.

[Sch76] J. Schnakenberg. Network theory of microscopic and macroscopic behavior of master equation systems. *Reviews of Modern Physics*, 48(4):571–585, October 1976.

[SE99] Debra J. Searles and Denis J. Evans. Fluctuation theorem for stochastic systems. *Physical Review E*, 60(1):159–164, July 1999.

[Sei05] Udo Seifert. Entropy Production along a Stochastic Trajectory and an Integral Fluctuation Theorem. *Physical Review Letters*, 95(4), July 2005.

[Sei12] Udo Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on Progress in Physics*, 75(12):126001, 2012.

[SGM+10] M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa. Interdependence of Cell Growth and Gene Expression: Origins and Consequences. *Science*, 330(6007):1099–1102, November 2010.

[SKMH14] M. Scott, S. Klumpp, E. M. Mateescu, and T. Hwa. Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Molecular Systems Biology*, 10(8):747–747, August 2014.

[STGC+19] Thomas P. Smith, Thomas J. H. Thomas, Bernardo García-Carreras, Sofía Sal, Gabriel Yvon-Durocher, Thomas Bell, and Samrāt Pawar. Community-level respiration of prokaryotic microbes may rise with global warming. *Nature Communications*, 10(1):5124, December 2019.

[Tou09] Hugo Touchette. The large deviation approach to statistical mechanics. *Physics Reports*, 478(1):1–69, July 2009.

[TPMW17] Thibaud Taillefumier, Anna Posfai, Yigal Meir, and Ned S. Wingreen. Microbial consortia at steady supply. *eLife*, 6:e22644, May 2017.

[Tu08] Yuhai Tu. The nonequilibrium mechanism for ultrasensitivity in a biological switch: Sensing by Maxwell's demons. *Proceedings of the National Academy of Sciences*, 105(33):11737–11741, August 2008.

[USP05] Kelly A. Underwood, James R. Swartz, and Joseph D. Puglisi. Quantitative polysome analysis identifies limitations in bacterial cell-free protein synthesis. *Biotechnology and Bioengineering*, 91(4):425–435, August 2005.

[VA16] Joseph J. Vallino and Christopher K. Algar. The Thermodynamics of Marine Biogeochemical Cycles: Lotka Revisited. *Annual Review of Marine Science*, 8(1):333–356, January 2016.

[Val10] J. J. Vallino. Ecosystem biogeochemistry considered as a distributed metabolic network ordered by maximum entropy production. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1545):1417–1427, May 2010.

[VF70] A D Ventsel' and M I Freidlin. On small random perturbations of dynamical systems. *Russian Mathematical Surveys*, 25(1):1–55, February 1970.

[VH18] Joseph J. Vallino and Julie A. Huber. Using Maximum Entropy Production to Describe Microbial Biogeochemistry Over Time and Space in a Meromictic Pond. *Frontiers in Environmental Science*, 6:100, October 2018.

[VQ09] Melissa Vellela and Hong Qian. Stochastic dynamics and non-equilibrium thermodynamics of a bistable chemical system: the Schlögl model revisited. *Journal of The Royal Society Interface*, 6(39):925–940, October 2009.

[VT20] Joseph J. Vallino and Ioannis Tsakalakis. Phytoplankton Temporal Strategies Increase Entropy Production in a Marine Food Web Model. *Entropy*, 22(11):1249, November 2020.

[Wan15] Jin Wang. Landscape and flux theory of non-equilibrium dynamical systems with application to biology. *Advances in Physics*, 64(1):1–137, January 2015.

[WFG+15] A. Z. Worden, M. J. Follows, S. J. Giovannoni, S. Wilken, A. E. Zimmerman, and P. J. Keeling. Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science*, 347(6223):1257594–1257594, February 2015.

[WNG+21] Mark Westoby, Daniel Aagren Nielsen, Michael R. Gillings, Elena Litchman, Joshua S. Madin, Ian T. Paulsen, and Sasha G. Tetu. Cell size, genome size, and maximum growth rate are near-independent dimensions of ecological variation across bacteria and archaea. *Ecology and Evolution*, page ece3.7290, March 2021.

[WODS15] Andrea Y. Weiße, Diego A. Oyarzún, Vincent Danos, and Peter S. Swain. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proceedings of the National Academy of Sciences*, 112(9):E1038–E1047, March 2015.

[XMW15] Xiao Xiao, Daniel J. McGlinn, and Ethan P. White. A Strong Test of the Maximum Entropy Theory of Ecology. *The American Naturalist*, 185(3):E70–E80, March 2015.

[YL21] Masato Yamamichi and Andrew D. Letten. Rapid evolution promotes fluctuation-dependent species coexistence. *Ecology Letters*, 24(4):812–818, April 2021.