Nonequilibrium stochastic dynamics of brain networks



Ramón Nartallo-Kaluarachchi
St Catherine's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Trinity 2026

Acknowledgements

This thesis is the culmination of four formative and unrepeatable years. It has been an experience shaped by numerous friends and colleagues to whom I am grateful. The list of those who have touched this work is countless but I will endeavour to name just a few.

First, I would like to acknowledge my colleagues in the Mathematical Institute, particularly those in OCIAM, the Mathematical Brain Modelling group, and the Networks group; with a special thank you to Torin and Nick for a wonderful office environment. Next, those in the Centre for Eudaimonia and Human Flourishing, my second home at Oxford, especially Leonardo, Kenneth, Kat, Elisa, Briana, and Vera, who have shared so much of this journey with me.

I would like to thank the many friends who made my time at Oxford so special. Margaret, Lucas, Adam, Vivian, and Alessandro, thank you for kicking off my first year in Oxford with so much joy. To the dream team of Tom, Sam, Michail, Zach, and (of course) Romain, no one has made me laugh and argue half as hard as you – I hope that our friendship knows no bounds in the years to come.

I would like to thank Professors Morten Kringelbach and Gustavo Deco for countless conversations about physics, neuroscience, and everything in between. Thank you for welcoming me into your labs with so much openness – I will be forever grateful.

This thesis would not have been possible without my supervisors Professors Alain Goriely and Renaud Lambiotte. To Alain, thank you for trusting me to chart my own course. It is only with your confidence and support that I have been brave enough to push boundaries and pursue the new. Your passion for science is contagious and I will carry the lessons you have taught me throughout the rest of my career. To Renaud, thank you for your open door, your patience, and your calm. It has been a steadying force these past four years and something I can only try to emulate in my own approach to research and to leadership. If I become half the scientist that either of you are, then I will consider it a lifetime well spent.

I owe an enormous debt of gratitude to my family. First, to the Cerberean ensemble of Lola, Nacho, and Maya, thank you for always being the happiest to see me, for dragging me outdoors at every opportunity, and for never caring who or what I am

beyond your family. To my siblings, Pedro, Nicole, and Gus, thank you for being my anchor to the world outside Oxford, for always understanding, and for always being there. You never fail to make me laugh, smile, and cry. To my parents, all that I am I owe to you and these are debts that I can never repay. I hope this thesis makes you proud and that it is some recompense for the start to life for which I am so grateful.

Finally, to Pauline, thank you for being my best friend. It is the time spent with you that I will remember most fondly from my doctorate. It is the promise of more time with you that inspires me to hope for the future. For what little it is worth, these pages and their author belong to you.

Funding and resources I am grateful to the EPSRC for funding my DPhil, and the Mathematical Institute and St Catherine's College for hosting me. During my DPhil, I also benefitted from additional funding from St Catherine's College, in the form of a Light Scholarship, and from the The Alan Turing Institute, in the form of a Community Enrichment Award.

I would like to acknowledge the use of the University of Oxford Advanced Research Computing facility for the analysis in Ch. 6. The computational work in this thesis would have been impossible if not for countless software packages and innumerable lines of open-source code, written by others, and made freely available. For this I am grateful.

For Pauline,
If not for you, I would still have written this thesis, but I might not remember why

And I said to me
I would like to be
Someone kind over cavalier
What have I done with my years?
A Boat To An Island, Pt. 2 / Agatha's Song
Ben Howard

Abstract

The human brain is the quintessential example of a complex system. It performs information processing through the coordinated activity of networks of neurons and neural populations that produce emergent spatiotemporal dynamics across scales. Neural recordings show that these dynamics are best approximated as a stochastic process in a nonequilibrium steady-state, characterised by time-irreversibility and persistent probability fluxes. Nevertheless, a fundamental theory of neural dynamics remains outstanding and many of the frameworks used to model and analyse neural systems rely on insufficient equilibrium assumptions. In this thesis, we introduce a number of approaches to investigate the brain and other complex systems from the perspective of nonequilibrium, irreversible dynamics, with a focus on stochastic processes, network theory, and data analysis. In particular, we investigate the structural origins of nonequilibrium steady-states on networks, how irreversibility can be used to infer higher-order interactions, the theoretical consistency and practical use of discrete-state approximations of nonequilibrium diffusions, the evidence for the formation of vortices in brain network oscillations, and nonequilibrium extensions of recurrent neural network models of associative memory and neural computation. This work combines theoretical results, numerical simulations, and analysis of empirical data including structural and functional neuroimaging, as well as an array of other examples. Our results highlight the importance of considering asymmetry and irreversibility when analysing neural systems, from networks of individual neurons to the large-scale dynamics of functional imaging. The frameworks we introduce provide practical methods to probe neural data, develop further theoretical models of the brain, and offer interesting mathematical insights into the areas of network and graph theory, stochastic processes, and a wide range of applied and computational mathematics.

Contents

1 Introduction **1**

1.1 A unifying framework for complex systems . . . 2

1.2 The brain as a complex system . . . 2

1.2.1 Neurons, populations, and dynamics . . . 2

1.2.2 Network neuroscience . . . 4

1.2.3 The brain out of equilibrium . . . 7

1.3 Outline and contributions . . . 9

2 Mathematical preliminaries **11**

2.1 Networks and network dynamics . . . 11

2.2 Stochastic processes and nonequilibrium steady-states . . . 13

2.2.1 Equilibrium, entropy, and entropy production . . . 14

2.2.2 Stochastic differential equations . . . 16

2.2.2.1 The Fokker-Planck equation . . . 18

2.2.2.2 Nonequilibrium stationary diffusions . . . 20

2.2.3 Discrete-state Markov processes . . . 24

2.2.3.1 Continuous-time Markov chains . . . 24

2.2.3.2 Discrete-time Markov chains . . . 26

2.3 Hodge theory on graphs . . . 30

2.3.1 Flows and operators on graphs . . . 30

2.3.2 Hodge decompositions on graphs . . . 35

3 Structural origins of entropy production in network dynamics **38**

3.1 Directed networks . . . 39

3.1.1 Measures of directedness on networks . . . 40

3.1.1.1 Parametrising the directedness of a network . . . 44

3.2 Prototypical dynamical processes on networks . . . 45

3.2.1 The Ornstein-Uhlenbeck process . . . 46

3.2.2 The kinetic Ising model . . . 50
3.3 Directedness drives entropy production on networks . . . 52
3.3.1 Dynamics on interpolated directed networks . . . 54
3.3.2 Erdős-Rényi graphs as a null model . . . 55
3.4 Linear models for multivariate time-series . . . 57
3.4.1 Applications to whole-brain network dynamics . . . 59
3.4.2 Trophic hierarchy in the New York Stock Exchange . . . 62
3.5 Conclusions . . . 63

4 Time-irreversibility and higher-order interactions in the brain 65
4.1 Irreversibility and functional interactions . . . 66
4.2 Directed multiplex visibility graphs . . . 70
4.3 The 'DiMViGI' framework . . . 72
4.3.1 The case of independent variables . . . 74
4.4 Applications to MEG recordings during long-term auditory recognition 75
4.4.1 Experimental paradigm and MEG recordings . . . 76
4.4.2 Results of the DiMViGI analysis . . . 77
4.5 Conclusions . . . 80

5 Discrete-state approximations of diffusion processes 83
5.1 Deriving a discrete-state approximation . . . 84
5.1.1 Approximation with finite volume methods . . . 85
5.1.1.1 The Scharfetter-Gummel discretisation . . . 87
5.1.1.2 Convergence of the EPR . . . 89
5.1.1.3 Diagonalisable and non-diagonal diffusion . . . 90
5.1.2 Examples and numerical experiments . . . 92
5.2 Statistical inference for discrete-state models . . . 97
5.2.1 The embedding problem . . . 99
5.2.2 Numerical experiments . . . 100
5.2.3 Determining if paths are from an NESS . . . 101
5.2.3.1 The shuffling procedure . . . 101
5.2.3.2 Applications to polarisation in schooling fish . . . 103
5.3 Decomposing processes as flows on graphs . . . 106
5.3.1 The many faces of the Helmholtz-Hodge decomposition . . . 106
5.3.2 Data-driven grids and the pipeline . . . 108
5.3.3 Numerical experiments . . . 109
5.3.4 Applications to physiological dynamics . . . 113

5.3.4.1 Pipeline modifications for real-world data 114
5.4 Conclusions 117

6 Vortex-like dynamics in brain network oscillations 120
6.1 Vortices and the XY model 121
6.1.1 Sampling from the XY model 123
6.1.2 Vortices on a lattice: formation and identification 125
6.2 Normalised Hodge Laplacians 127
6.2.1 The normalised Hodge decomposition 128
6.2.2 Eigendecomposition of the normalised 1-Laplacian 130
6.2.2.1 Localisation in $\widetilde{L}_1$ 131
6.3 Detecting vortices with the Hodge decomposition 133
6.3.1 Expansion into eigenmodes of $\widetilde{L}_1$ 137
6.4 Numerical experiments 138
6.4.1 XY simulations on the lattice 138
6.4.2 XY simulations on the connectome 140
6.5 Applications to neuroimaging data 142
6.5.1 Hodge vorticity during cognitive tasks 143
6.5.2 Eigenmode expansion of edge-based neural signals 145
6.6 Conclusions 147

7 Embedding latent dynamics in neural networks 149
7.1 Introduction 149
7.1.1 Hopfield's model of associative memory 150
7.1.2 Neural manifolds 153
7.2 The 'drift-diffusion matching' framework 154
7.2.1 A symmetric-asymmetric decomposition 157
7.2.2 Examples and numerical experiments 160
7.2.3 Reconstructing the subspace from paths 161
7.3 Diffusions and networks that switch and cycle between attractors . . 165
7.3.1 Input-driven, attractor-switching dynamics 166
7.3.2 Attractor-cycling, irreversible dynamics 171
7.3.2.1 Constructing rotational planar diffusions 172
7.3.2.2 Examples and numerical experiments 173
7.4 Neural networks in the NESS 175
7.4.1 A Helmholtz-Hodge decomposition for RNNs 177
7.5 Conclusions 181

8 Discussion **184**

A Appendices for Ch. 3 **189**
A.1 Proof of Lemma 3.1.1 . . . 189
A.2 Algorithm for sampling hierarchically directed networks . . . 190
A.3 Proof of Lemma 3.3.2 . . . 191

B Appendices for Ch. 4 **193**
B.1 Participant-level results with DiMViGI . . . 193
B.2 Surrogate testing with shuffled time-series . . . 193

C Appendices for Ch. 5 **197**
C.1 Deriving the Scharfetter-Gummel discretisation . . . 197
C.2 Non-diagonalisable diffusion . . . 198
C.3 Optimisation-based solution to the embedding problem . . . 199
C.4 Sampling for oscillatory and chaotic processes . . . 200
C.4.1 Splitting method for stochastic oscillators . . . 200
C.5 Inferring the spatial dependence of the diffusion in real-world data . . . 201

D Appendices for Ch. 7 **203**
D.1 Optimal decomposition of a network under constraints . . . 203
D.2 Mapping coded inputs to network inputs with ridge-regression . . . 204
D.3 Designed rotational drift is DFSD . . . 204
D.4 Flux and EPR in a low-rank RNN . . . 205
D.5 Parameters and details for numerical experiments . . . 206
D.5.1 Nonlinear systems and RNN representations . . . 206
D.5.2 Switching and cycling attractors . . . 206

Bibliography **208**

Papers related to this thesis

Parts of this thesis have appeared in various publications.

Parts of Ch. 1, Ch. 2, and Ch. 8 appear in papers:

R. Nartallo-Kaluarachchi, M. Kringelbach, G. Deco, R. Lambiotte, A. Goriely. Nonequilibrium physics of brain dynamics, *Physics Reports*, 1151, 1-43, 2026, https://doi.org/10.1016/j.physrep.2025.10.003.

R. Nartallo-Kaluarachchi, R. Lambiotte, A. Goriely. From reductionism to realism: holistic mathematical modelling for complex biological systems, *Journal of the Royal Society Interface*, 22 (232) 20250468, 2025, https://doi.org/10.1098/rsif.2025.0468

Ch. 3 appears in the paper:

R. Nartallo-Kaluarachchi, M. Asllani, G. Deco, M. Kringelbach, A. Goriely, R. Lambiotte. Broken detailed balance and entropy production in directed networks, *Physical Review E*, 110 (034313), 2024, https://doi.org/10.1103/PhysRevE.110.034313.

Ch. 4 appears in the paper:

R. Nartallo-Kaluarachchi, L. Bonetti, G. Fernández-Rubio, P. Vuust, G. Deco, M. Kringelbach, R. Lambiotte, A. Goriely. Multilevel irreversibility reveals higher-order organization of nonequilibrium interactions in human brain dynamics, *Proceedings of the National Academy of Sciences*, 122 (10) e2408791122, 2025, https://doi.org/10.1073/pnas.2408791122.

Ch. 5 appears in the papers:

R. Nartallo-Kaluarachchi, R. Lambiotte, A. Goriely. Coarse-graining nonequilibrium diffusions with Markov chains, *Journal of Statistical Mechanics: Theory and Experiment*, 033205, 2026, https://doi.org/10.1088/1742-5468/ae4f7d.

R. Nartallo-Kaluarachchi, P. Expert, D. Beers, A. Strang, M. Kringelbach, R. Lambiotte, A. Goriely. Decomposing force fields as flows on graphs reconstructed from stochastic trajectories, *Proceedings of the Third Learning on Graphs conference (LoG 2024)*, PMLR 269:4:1-4:26, 2024, https://proceedings.mlr.press/v269/nartallo-kaluarachchi25a.html

Ch. 6 is in preparation for submission as:

R. Nartallo-Kaluarachchi[†], D. Beers[†], E. Guzmán, G. Deco, M. Kringelbach, R. Lambiotte, A. Goriely, M. Porter. Vortex-like patterns in oscillator networks and neural dynamics.

Ch. 7 has been accepted for publication:

R. Nartallo-Kaluarachchi, R. Lambiotte, A. Goriely. Embedding stochastic dynamics in latent manifolds of recurrent neural networks, 2026, *Physical Review Research* (accepted).

In addition, parts of this thesis motivated the popular articles:

R. Nartallo-Kaluarachchi. Mathematical memory machines, *Mathematics Today*, 175, 2024.

R. Nartallo-Kaluarachchi. Cycles and potentials: From structure to stochastic dynamics, *London Mathematical Society Newsletter*, Sept 2025.

Statement of authorship & originality

Ch. 3 was completed with input from Malbor Asllani, Renaud Lambiotte, and Alain Goriely. Morten Kringelbach and Gustavo Deco provided preprocessed neuroimaging data.

The framework presented in Ch. 4 was conceived with input from Renaud Lambiotte and Alain Goriely. The analysis was completed with input from Leonardo Bonetti. The experimental data were collected and preprocessed by Leonardo Bonetti and Gemma Fernández-Rubio, with supervision from Morten Kringelbach and Peter Vuust.

Ch. 5 was completed with supervision from Alain Goriely and Renaud Lambiotte. The framework in Sec. 5.3 was conceived with input from Paul Expert, David Beers, and Alexander Strang.

The framework in Ch. 6 was designed by equal contribution with David Beers, with supervision from Mason Porter, Alain Goriely, and Renaud Lambiotte. Neuroimaging data were preprocessed by Elvira Guzmán, Gustavo Deco, and Morten Kringelbach.

Ch. 7 was completed under the supervision of Renaud Lambiotte and Alain Goriely.

All writing, results, and presentation of figures not otherwise attributed, are my own.

Declaration of generative AI use

I confirm that I have read the University and MPLS Division guidance on generative AI and have complied with it in preparing this thesis. The original mathematical contributions of the thesis, including the problem formulation, algorithms, numerical analysis, experiments, and conclusions, are my own work except where standard scholarly references and co-authorship statements indicate otherwise.

I used the following generative AI tools for the preparation of this thesis: ChatGPT Edu, OpenAI, [model GPT-5] and Gemini, Google, [model 3.1 Flash/Pro]. These tools were used as follows: I used both ChatGPT Edu and Gemini for debugging code, drafting plotting scripts, and setting up the training of the deep learning models in Ch. 7. Suggestions from these tools were never accepted without my own scrutiny, all numerical simulations were run by me locally, and all final code was reviewed.

---

† denotes equal contribution.

Chapter 1

Introduction

Mathematical modelling has become the dominant paradigm for abstracting physical systems and describing their behaviour over time and space. Nowhere has this relationship been more successful than in fundamental physics. From the quantum to the cosmological scale, physical laws can be expressed in governing equations that are often simple yet incredibly rich. Nevertheless, there exists a class of systems for which no ubiquitous set of governing equations has been derived. Often found in biology, these systems are characterised by a myriad of time- and length-scales, heterogeneous interactions, and a zoo of different elements [1]. For lack of a better word, such systems are often called *complex*. A central feature of complex systems is that they lack the symmetries assumed by the mean-field framework of statistical mechanics or continuous description through partial differential equations (PDEs). Instead, they have non-trivial topologies that are neither regular nor random, restricting one's ability to eliminate degrees of freedom, rendering descriptions that are convoluted and unwieldy.

The human brain is an archetypal example of a complex system. Part of a living organism, it has an intricate topology, with billions of neurons organised into an irregular, yet purposeful, circuit. This architecture supports communication between neurons via electrical signals, which, in turn, encodes representations of the world, performs computations, and is the seat of our conscious mind. Yet, despite great and continual advancements in neuroscience, its mechanisms remain noticeably opaque. The modern consensus eschews the historical notion that neurons and brain areas are *only* specialised units, responsible for isolated functions, instead embracing the notion that brain function is realised in a distributed manner, through shared representations across the system [2].

In this thesis, we will focus on the modelling and analysis of so-called *neural dynamics*, the time-varying activity of the brain, at a number of different scales. We

will take a 'complex-systems perspective' focusing on the structure and stochastic dynamics of brain networks. In particular, a central theme will be the recent discovery that the brain may be operating in a *nonequilibrium steady-state (NESS)* [3], which allows us to integrate notions from stochastic thermodynamics into the study of neural systems.

1.1 A unifying framework for complex systems

> *"I think the next century will be the century of complexity"*
>
> Stephen Hawking (2000)

Despite their diversity, a wide array of complex systems have been found to display a number of universal and common features. As a result, the science of complex systems has often sought a single unifying framework with which to describe them. Haken's *synergetics* [4], Wiener's *cybernetics* [5], and Prigogine's *irreversible and dissipative structures* [6], are all attempts to describe the emergence of self-organisation in systems far from equilibrium using the language and techniques of statistical mechanics and nonlinear dynamics. In more recent years, the science of complex systems has focused on more modern areas including: the study of *networks*, as an abstraction bridging the structure and dynamics of a system [7,8]; *information theory*, as a general framework for understanding computations [9,10]; and data-driven approaches to nonlinear dynamics [11]. Nevertheless, the goal, for many, remains an integration of paradigms into a single framework with which we may describe the organisation, evolution, and behaviour of systems in a range of domains [12]. It is from this perspective that we will approach the brain and its dynamics.

1.2 The brain as a complex system

1.2.1 Neurons, populations, and dynamics

It has long been known that the brain is the seat of cognition in humans and other animals. It was not until the work of Golgi and Ramón y Cajal, that *neurons* were identified as the principal computational units of the brain [15,16]. Since then, despite the array of other cell types in the brain, the study of cognition has focused on neurons, their connectivity, and their dynamics.

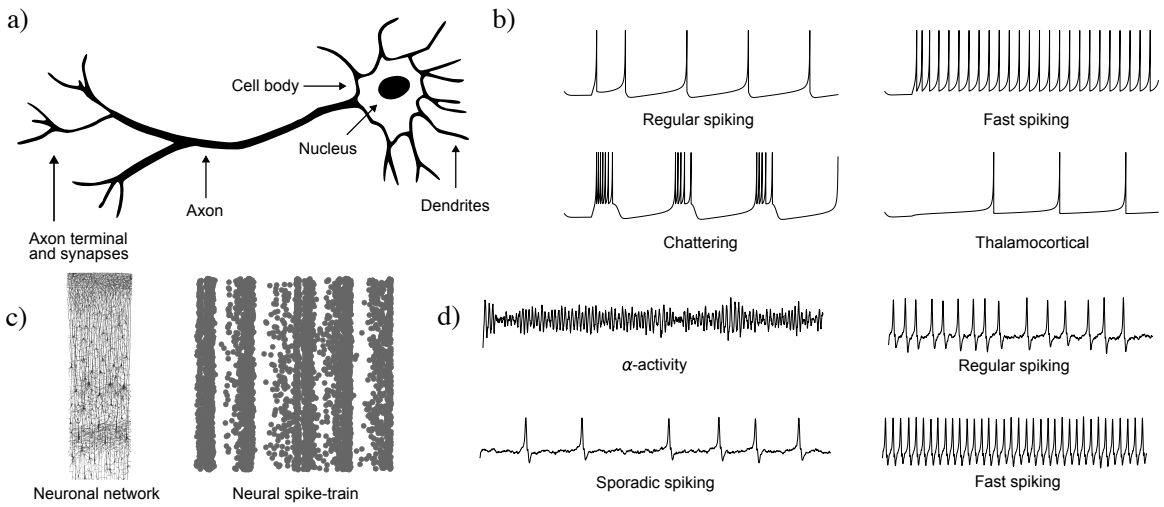

Figure 1.1: **Neurons, populations, and their dynamics.** *a*) A schematic diagram of a neuron.* *b*) Action potentials from different dynamical regimes of the *Izhikevich* model [13]. *c*) Neurons are organised into networks of connected units, which produce collective activity.† This can be visualised as a point-process, a spike-train, in a *raster* plot. *d*) The average activity of a population or network of neurons can be approximated with a mean-field model. These models also have an array of dynamic regimes. We plot trajectories from the *Jansen–Rit* neural mass model [14].

*Adapted from Notjim by CC BY-SA 3.0 (https://commons.wikimedia.org/w/index.php?curid=4824168). †Drawing by Santiago Ramón y Cajal (Public Domain).

Briefly, neurons communicate via electrical signals which propagate along their cell-body, known as the *axon* [Panel *a*), Fig. 1.1]. Neurons are characterised by *excitable* dynamics, which cause them to behave as threshold units, requiring a sufficient input of current to prompt a dynamical excursion known as an *action potential* or *spike* [13]. In their seminal work, Hodgkin and Huxley were able to derive equations for neuronal spiking by modelling the electrical circuit coupling the membrane potential of a neuron to ionic channels in the 'giant axon' of the squid [17]. Through time-scale separation, bifurcation methods, or asymptotic approximations, the Hodgkin-Huxley model can be simplified to define a number of reduced systems. Neglecting to model the propagation of the potential along the axon, therefore abstracting each neuron as a point, these models take the form of ordinary differential equations (ODEs) and have a range of dynamic regimes with different spiking behaviours [Panel *b*), Fig. 1.1].

Through *synapses*, neurons are connected, allowing action potentials from a *presynaptic* neuron to serve as input current for the *postsynaptic* neuron, triggering a cascade of activity through the neural network, and leading to emergent collective

behaviour [Panel *c*), Fig. 1.1]. This observation has since led to the dominance of the so-called 'connectionist' perspective, which posits that computation in the brain is performed in a distributed manner through a network of relatively simple units. It is from this paradigm that modern deep learning emerges, coopting the term 'neural network' to describe their statistical models [18].

Recording from individual neurons in the brain is challenging and relies on invasive electrophysiology. As a result, this modality primarily focuses on small animals, such as rats or mice. Moreover, it focuses on specific, smaller areas of the brain. On the other hand, neuroimaging and encephalography methods, such as functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), and electroencephalography (EEG), non-invasively capture large-scale activity in the brain. These modalities record proxies of neural activity such as blood-flow into brain regions (e.g. fMRI), magnetic fields measurable from outside the head (e.g. MEG), and electrical potentials on the scalp (e.g. EEG). As a result, these data can be considered to be the average activity in a region of interest (ROI) of the brain, known as a neural mass. This neural mass is composed of a large number of interacting neurons with heterogeneous connectivity and cell types. Nevertheless, with a mean-field approach, a single model can be used to describe the average activity of the neural population. A diverse range of neural mass models exist, some able to reproduce an array of different activity types, in alignment with neuroimaging data [Panel *d*), Fig. 1.1] [14].

1.2.2 Network neuroscience

The mathematical study of neural systems can be roughly separated into three subareas, with significant overlap and blurred boundaries. *Network neuroscience* is the study of large-scale brain networks and their dynamics [19, 20]; *theoretical* and *computational neuroscience* focuses on analysis of electrophysiological data and the modelling of learning and computation in neural circuits [21]; and *mathematical neuroscience* studies the properties of neuronal and neural mass models with a focus on statistical and dynamical systems theory [22]. In order to study human brain network dynamics, we find ourselves focusing on the macroscopic scale for the majority of this thesis.

Large-scale brain activity is typically modelled as a network dynamical system, where the nodes correspond to macroscopic brain areas and the edges correspond to connections between them. This demands that we select a *parcellation* of the brain into a number of predetermined, discrete regions [Panel *a*), Fig. 1.2]. Parcellations vary in size from $10^0 - 10^3$ nodes. *Functional* imaging data, which measures activity,

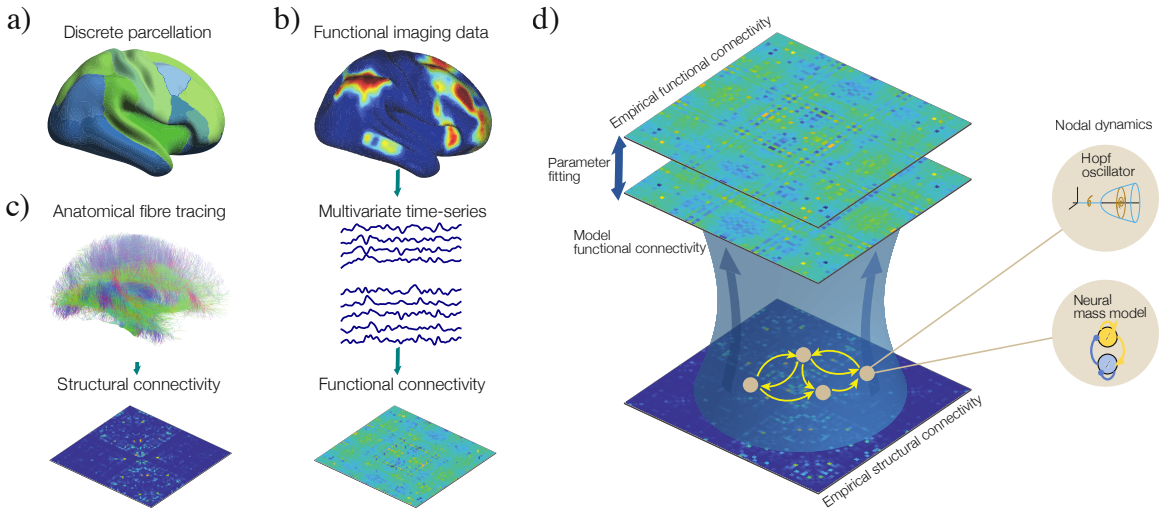

Figure 1.2: **Structural and functional brain networks.** *a*) To model the brain as a network, we first choose a parcellation into a number of predetermined discrete regions. *b*) Given a parcellation, a time-series of activity can be extracted for each brain region. This can be used to compute the functional connectivity (FC). *c*) Anatomical fibre tracing can be used to extract the structural connectivity (SC). *d*) A whole-brain network model consists of a SC network, local dynamics for each region, and a target FC network. The parameters are optimised to align the simulated model FC and the target.

can then be co-located with this parcellation to produce a time-series of activity for each region [Panel *b*), Fig. 1.2].[1]

There are three typical notions of connectivity in network neuroscience:[2]

- **Structural connectivity (SC)**. This refers to physical connections between regions through white-matter tracts. This data can be obtained through non-invasive diffusion tensor imaging (DTI), or invasive tract-tracing [Panel *c*), Fig. 1.2]. The direction of edges cannot be determined from DTI and thus in humans we do not have access to the direction of connections. Networks are typically averaged over a large number of people and are assumed to be constant between most experimental conditions.[3]

- **Functional connectivity (FC)**. This is a statistical correlation between signals. Two regions are considered 'connected' by the degree to which their ac-

[1]For fMRI this time-series is known as blood-oxygenation-level–dependent (BOLD) activity. For MEG and EEG, is it normalised into the unit-less *neural activity index*.

[2]For a detailed discussion of these notions of connectivity, we recommend Ref. [2].

[3]With the exception of comparing vastly different age-groups or those with neurodegenerative disease.

tivity is correlated [Panel *b*), Fig. 1.2]. Whilst Pearson's correlation is standard, other measures, including directed metrics, can be used.

- **Effective connectivity (EC)**. This is model-based connectivity inferred from data. It refers to the minimal network that can explain activity data under an assumed model of neural dynamics.

The human brain network is often referred to as the *connectome*. The structural features of the SC network are of profound interest in their own right, including its modularity, small-worldedness, and node-centrality [20]. However, from a neural dynamics perspective, its role in constraining activity, so called 'structure-function coupling', is of primary interest.

A *whole-brain network model* consists of a SC network, local dynamics for each neural mass, a coupling function, and a statistical target [23]. Examples of local dynamics include linear systems, neural mass models, or limit-cycle oscillators. A typical statistical target is to align the model FC with that of the empirical data [Panel *d*), Fig. 1.2]. The model is then fit via the challenging procedure of optimising the parameters to align the model statistics with that of the target. One can choose to take homogeneous or heterogeneous regions, or to modify the structural network as a parameter. Additionally, we can introduce time-delays in the coupling terms to model the transmission delays between regions, which leads to a delay differential equations (DDEs) [24]. When comparing two distinct experimental conditions, we can contrast the inferred parameters.

Remark 1.2.1 (Neural field theory)**.** In this thesis, we will focus heavily on network approaches to neural systems, as detailed above. In contrast with this approach, the mathematical framework of neural field theory (NFT) models brain activity as a field over a continuous geometry, typically the cortex, using PDEs [25]. These methods lend themselves to the study of travelling waves, patterns, and other nonlinear phenomena [26]. Such an approach is not purely mathematical, and some experimental results indicate that a geometric picture of the brain could, in fact, be more informative [27]. Nevertheless, it will not be the focus of our enquiry.

1.2.3 The brain out of equilibrium

> How would we express in terms of the statistical theory the marvellous faculty of a living organism, by which it delays the decay into thermodynamical equilibrium (death)?
>
> *What is Life? The Physical Aspect of the Living Cell (1944)*
> Erwin Schrödinger

Like many complex systems, the brain produces an array of emergent dynamic phenomena including synchronisation, travelling and standing waves, criticality, chaos, and turbulence [28]. As a result, it has often been modelled and analysed with the tools and language of statistical mechanics. However, one feature of neural dynamics that appears to have been neglected, is the degree to which it is better described by equilibrium statistical mechanics (ESM) or nonequilibrium statistical mechanics (NESM). The similarities and differences between these paradigms are vast[4] so we will focus on a few key distinctions.

In a pair of seminal papers [30, 31], Jaynes introduced the notion of maximum entropy (ME) in an effort to bridge statistical inference with mechanics. In simple terms, this principle argues that given a set of constraints, such as those derived from empirical data, one should choose the probability distribution to represent the system that is the least biased i.e. the one with the largest entropy. Constraints are often related to the *energy* of the system, which neglects dynamics and firmly entrenches this approach in the Boltzmann-Gibbs paradigm of ESM [32].[5] The ME approach has been particularly useful in neuroscience where it has linked the Ising model to the spiking of neuronal networks [33] and has inspired learning algorithms for biological and artificial circuits [34, 35]. Moreover, as we will discuss in greater detail in Ch. 2 and 3, symmetric interactions between elements are a hallmark of an equilibrium system. Whilst we know that interactions between neurons and neural masses are directed and asymmetric, we typically do not have access to this data, especially in the case of human structural brain networks, thus we are implicitly restricted to the ESM paradigm.

Nevertheless, an ESM perspective predicts that neural dynamics will be *time-reversible*, another canonical feature of systems at thermodynamic equilibrium. This can be easily investigated, and we find that they are, in fact, markedly *irreversible* [3].

[4] We point the interested reader to Ref. [29].
[5] This will be defined more formally in Sec. 2.2.1.

Beyond this, Lynn et al. found that large-scale human brain dynamics could be approximated as a stochastic process in a NESS, and that the level of irreversibility, which is defined by the entropy production rate (EPR), was higher in task-based neural dynamics when compared to resting-state.[6] Since this discovery, a plethora of techniques and datasets have been developed to investigate the variation of the 'level of nonequilibrium' in neural data across a range of experimental conditions, recording modalities, and spatial scales. Results show that neural dynamics have lower irreversibility during sleep compared with wakefulness [36], under the influence of anaesthetic drugs [37], psychedelic drugs [38], in Alzheimer's disease [39], and in disorders of consciousness like coma [40]. These observations are at odds with the traditional modelling and analysis paradigms in neuroscience and call for the development of mathematical theory and data-analytic methods to address this problem. This requires a transition to NESM, with a focus on irreversibility and entropy production in trajectory data, as well as asymmetry in network dynamics.

The brain is not alone in its nonequilibrium behaviour. Beginning as early as Schrödinger, NESM has been central to understanding the ability of biological systems to self-organise and maintain themselves away from thermodynamic equilibrium [41]. As a result, irreversibility and entropy production are ubiquitous in biological dynamics across scales [42, 43]. For microscopic (bio)physical systems, such as cells, the EPR directly measures the heat dissipated by the system to its environment in order to maintain itself away from thermodynamic equilibrium [44]. Despite the fact that the brain itself is an energetic biological organ, it is worth clarifying that the EPR of neural activity is not necessarily related to its metabolic cost as an organ. This is due to the coarse-grained and indirect nature of neural recordings, which are far removed from the energetic behaviours of cells in the brain. Coarse-graining, in either space or time, has a complex effect on stochastic thermodynamics. In particular, it can obscure irreversible currents and thus leads to a lower-bound on the EPR of the system [45]. Throughout this thesis, the irreversibility and EPR that we describe in the brain refers to violations of detailed balance[7] in phase-space dynamics rather than physical dissipated heat.

In a different direction, the brain's position as an 'information-processing machine' operating out of equilibrium, draws an interesting relationship with the stochastic thermodynamics of information and its storage, transformation, and erasure [46, 47].

[6] We will define a NESS, irreversibility, and the EPR in mathematical detail in Sec. 2.2.

[7] To be defined in detail in Ch. 2.

As a result, this is a research direction filled with opportunities to reshape perspectives on the functional dynamics of the brain and other biological complex systems.

1.3 Outline and contributions

This thesis is organised as follows.

In Ch. 2, we introduce the mathematical foundations on which this thesis will develop, specifically: networks and network dynamics; stochastic thermodynamics; stochastic processes and NESSs in discrete and continuous spaces; and Hodge theory on graphs. Overall, the focus is placed on applied probability and dynamics, with the use of some methods from deep learning and numerical analysis. Whilst the terms *networks* and *graphs* are often used interchangeably, we will reserve the term *networks* to refer to a system composed of a number of interacting elements, typically physically instantiated in the world, such as a brain network. On the other hand, *graphs* will be mathematical objects that we construct, typically as part of a data analysis pipeline.

In Ch. 3, we will focus on the structural origins of NESSs on networks. In particular, we will illustrate how network directedness is related to the EPR of a nonequilibrium system. More specifically, we will introduce three measures of network directedness and a scheme to generate networks of increasing directedness. Next, we will use three prototypical network stochastic processes, namely the Ornstein-Uhlenbeck, Ising model, and random walk, and show that their EPR increases with directedness using both analysis and simulations. Finally, we will introduce an approach to fit an Ornstein-Uhlenbeck process to a multivariate time-series (MVTS), from which we can infer both the asymmetry of the structural network, and the EPR of the data. We will apply this approach to fMRI from resting and task-evoked brain activity, as well as financial time-series. Our results show that EPR is higher in task-activated dynamics compared to the resting state, and that this can be interpreted as a reorganisation of brain networks into more hierarchical and asymmetric structures.

In Ch. 4, we will focus on the problem of inferring higher-order interactions from neural recordings. We propose time-irreversibility as an effective measure of significance for a higher-order interaction. Next, we introduce the 'DiMViGI' framework in order to measure this irreversibility using visibility graphs. Applying this framework to MEG data recorded during a music-listening task, we show that our approach highlights neuroscientifically insightful groups of regions, illustrating the effectiveness of methods inspired by NESM.

In Ch. 5, we shift our focus from neural dynamics and focus on the more general problem of approximating diffusion processes with discrete states. Using finite volume methods, we derive a master equation from a diffusion process and show that the EPR of this approximation converges in the limit. However, by considering the inference of a discrete-state model from path data, we show that, in practice, we obtain a poor lower-bound on the true EPR. Next, we consider the discrete versus continuous forms of the Helmholtz-Hodge decomposition and provide a heuristic method to decompose a stochastic process into reversible and irreversible components from path data. We illustrate this on simulated paths as well as experimental data from flickering red-blood cells and healthy and arrhythmic heartbeats.

In Ch. 6, we consider the emergence of vortices in brain network oscillations. More specifically, we focus on models of phase-oscillators on a network that produce vortices on a lattice i.e. the XY model. Whilst lattice-based approaches have been used to analyse the vorticity of neural data, this has not been extended to heterogenous (complex) networks and thus cannot take into account the brain's SC. Using the Hodge decomposition, we present an approach to measure the vorticity of phase-oscillators on an arbitrary network, which we illustrate on simulations of the XY model and on fMRI data from tasks and resting-state. We also introduce a measure of complexity for vortex-like dynamics, by considering an expansion of neural signals in a Hodge-Laplace eigenbasis, which extends node-based signal processing to the edge. We find that the vorticity and complexity of task-activated dynamics is higher than in the resting state.

In Ch. 7, we focus on recurrent neural networks and the theory of neural manifolds. We introduce the 'drift-diffusion matching' framework for training a network to encode an arbitrary stochastic process in an affine linear subspace, which we illustrate using examples of nonequilibrium and chaotic systems. Next, by constructing appropriate target processes, we show that our framework can be used to extend models of associative and episodic memory. In particular, we encode dynamics that switch between attractors, which represent distinct memories, based on an input, and autonomous dynamics that cycle between attractors. Using decompositions of the learned connectivity networks, we analyse the role of asymmetry in encoding nonequilibrium dynamics.

In Ch. 8, we offer a discussion of the work, its limitations, and consider future directions.

Chapter 2

Mathematical preliminaries

We begin by introducing the requisite mathematical preliminaries which will recur throughout the chapters of the thesis. In particular, we will focus on networks and applied stochastic processes, without delving into measure-theoretic details or probability spaces.

2.1 Networks and network dynamics

Networks A *network* or *graph*, $G = (V, E)$, is a mathematical object which represents pairwise relationships between the elements of a system. It is comprised of *nodes*, the elements of the system, which we index with natural numbers $i \in V \subseteq \mathbb{N}$, and *edges*, $(i, j) \in E \subseteq V \times V$. Nodes $i \in V$ and $j \in V$ are said to be *connected* if $(i, j) \in E$. We represent a network with its *adjacency matrix*, $A \in \mathbb{R}^{|V|\times|V|}$, whose entries are given by,

$$A_{ji} = \begin{cases} 1 & \text{if } (i,j) \in E \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

If the pairs (i, j) are considered to be *ordered* i.e. $(i, j) \neq (j, i)$, then the network is said to be *directed*, which implies that $A \neq A^\top$, otherwise it is *undirected* and $A = A^\top$ [Panel a), Fig. 2.1]. The connection from *source* node j to *target* node i is given by the edge (i, j), which we will also write as $j \to i$ i.e. $A_{ij} = 1$.

We may extend networks by introducing *weights* to the edges, where we replace the adjacency matrix by the *weight matrix*, where $A_{ij} \in \mathbb{R}$ is the weight, positive or negative, along the edge from j to i, or zero if the nodes are not connected.

Whilst regular structures, such as the regular lattice, are an example of a network, the majority of this thesis will focus on what are known as *complex* networks. These are characterised by an irregular and heterogenous topology [48]. As mentioned in the

introduction, these systems lack the symmetries required for a mean-field approach and place the structure of the system front and center.

Network dynamics When considering dynamical processes, it is important to differentiate between dynamics *of* and *on* networks. Dynamics of networks refers to the evolution of a network's structure over time, such as that seen in the growth of a network or in the study of temporal networks.

We will focus on dynamics on networks, where the system state, $x(t)$, evolves over time with dynamics that are influenced by the structure of the network. One example is the random walk, where $x(t) \in V$, and a walker jumps between connected nodes along edges. Another example is the situation where we consider $x(t) \in \mathbb{R}^N$, where $N = |V|$, and each node i corresponds to one entry, $x_i(t)$, in the *state-vector*, $x(t)$. For example, we might have

$$\frac{dx_i}{dt} = f_i(x_i) + \sum_{j=1}^{N} A_{ij} g(x_i, x_j). \tag{2.2}$$

This is an ODE whose coupling terms are constrained by the structure of the network. Here $f_i : \mathbb{R} \to \mathbb{R}$ represents local dynamics at node i, and $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a *coupling function*, which defines the nature of the interaction between nodes. As a result, a node i receives a contribution from all its neighbours in the network.[1]

Higher-order networks Networks are limited to representing *pairwise* relationships i.e. edges are relationships between sets of precisely two nodes. *Higher-order networks* is a general term used to describe alternative forms of networks that include three-way, or more general, interactions [49]. One example is a *hypergraph* where $E \subseteq \mathcal{P}(V)$, where $\mathcal{P}(S)$ is the set of all subsets of S. As a result, a *hyperedge*, $e \in E$, is a collection of an arbitrary number of nodes, $e = (i_1, \ldots, i_k)$, which can be extended to include direction or weights [Panel b), Fig. 2.1]. A hyperedge is said to be of *order* k if it includes k nodes.

A special case of the hypergraph is the *simplicial complex* [49]. A simplicial complex is a hypergraph with the implicit restriction that if $(i_1, \ldots, i_k) \in E$, then $\{(j_1, \ldots, j_l) : j_m \in \{i_1, \ldots, i_k\}, 1 \leq l \leq k\} \subseteq E$ i.e. for any order k hyperedge to be included in the complex, then all lower order 'subedges' must be included also [Panel

[1] One can also consider $x(t) \in \mathbb{R}^{dN}$, where there are d variables at each node i.e. $x_i(t) \in \mathbb{R}^d$. This is common in neuronal and neural mass models.

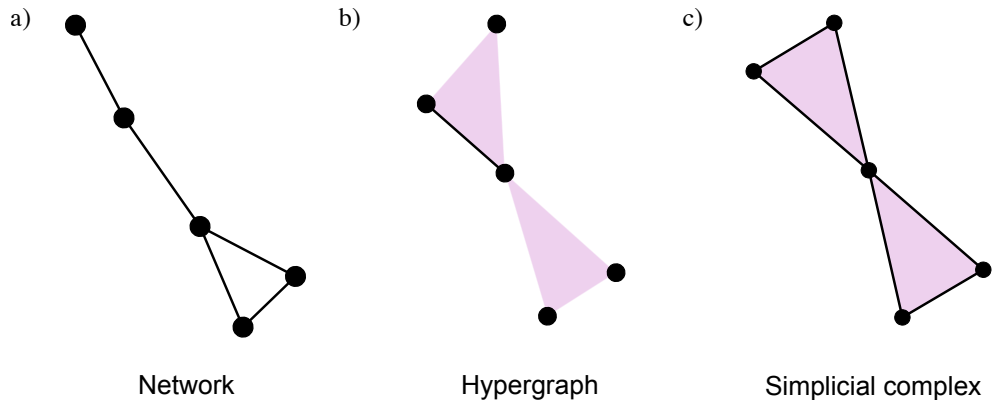

Figure 2.1: **Networks and extensions**. *a*) A simple network represents interactions between pairs of nodes. These can be weighted or directed. *b*) A hypergraph allows for more general interactions between sets of nodes. *c*) A simplicial complex is a special case of a hypergraph, where lower-order structures must be present for higher-order structures to be defined.

c), Fig. 2.1]. A natural example is the *clique complex*, where an ordinary network, $G = (V, E)$, defines a simplicial complex, $S = (V, \widetilde{E})$, with hyperedges,

$$(i_1, \ldots, i_k) \in \widetilde{E} \text{ if } (a, b) \in E \text{ for all } a, b \in \{i_1, \ldots, i_k\}.$$

This implies that the order k edges are simply the *cliques*, sets of nodes where each pair is connected, of size k in the original network.

2.2 Stochastic processes and nonequilibrium steady-states

The laws of classical physics are deterministic, which implies that the time-evolution of a system is precisely specified by a detailed accounting of the forces exerted on it and its initial state. However, at coarse-grained levels of abstraction, such an accounting is unfeasible, and initial states that are practically indistinguishable result in a variety of different outcomes. A canonical example is the behaviour of a macroscopic particle colliding with a multitude of microscopic particles. This is precisely the phenomenon recorded by Brown, when he observed the behaviour of pollen on the surface of water – so-called *Brownian motion* [50]. In this scenario, it is practical to model dynamics as being *stochastic processes*, and thus describe them probabilistically. These processes can be defined over both continuous and discrete state-spaces. Neural recordings, such as those obtained with neuroimaging and encephalography

techniques, are noisy, continuous-valued data, and are thus best described by stochastic differential equations (SDEs), whilst spike-train data is naturally described by discrete states. We will place a particular focus on the NESS, which bridges the areas of stochastic processes and of thermodynamics, where we will begin.

2.2.1 Equilibrium, entropy, and entropy production

Thermodynamics is a central pillar of physical theory, and describes the relationship between heat and all other forms of energy. It is also deeply intertwined with both ESM and NESM. A central quantity in thermodynamics is *entropy*. If we consider a system composed of a large number of microscopic elements, such as molecules, we find that there are a number of different microscopic arrangements (*microstates*) which project the same macroscopic observables (*macrostates*), such as energy or volume. This leads to an uncertainty in the microstates of a system given knowledge of its macrostates. Assuming the system is isolated and at a fixed energy i.e. at thermodynamic equilibrium, then the fundamental assumption of ESM is that all accessible microstates are equally probable [51]. It was Boltzmann who first introduced *entropy* as a measure of the uncertainty in the microstates of such an isolated system. Explicitly, the Boltzmann entropy is given by $S_{\mathrm{B}} = k_B \log \Omega$, where $k_B \approx 1.380 \times 10^{-23}$ J/K is *Boltzmann's constant*, and Ω is the number of microstates corresponding to the macrostate.[2] Gibbs generalised this to the case of non-isolated systems through the *canonical ensemble*, where energy can fluctuate and where some microstates are more probable than others. The probability of a microstate i that has energy E_i is given by the *Boltzmann distribution*,

$$p_i = \frac{e^{-\beta E_i}}{Z}, \tag{2.3}$$

where $\beta = 1/(k_B T)$ is known as the *inverse thermodynamic temperature*, and $Z = \sum_i e^{-\beta E_i}$ is a normalising constant known as the *partition function* [51]. We often describe such equilibrium systems as 'static' as the distribution is stationary and the average macroscopic observables are constant. As a result, we are not interested in their time-dependent dynamics, which are time-reversible, and instead focus on sampling from the Boltzmann distribution using Monte Carlo methods (MCMs) [52]. This is in contrast with the dynamics of nonequilibrium systems.

The Gibbs entropy for the canonical ensemble is given by,

$$S_{\mathrm{G}} = -k_B \sum_i p_i \log p_i, \tag{2.4}$$

[2] Throughout the thesis, we will use $\log(\cdot)$ to be the natural logarithm.

where p_i is the probability that the system is in microstate i. This was further generalised by Shannon to quantify the uncertainty associated with an arbitrary random variable described by a probability distribution or density [9],[3]

$$S = -\sum_i p(x_i) \log p(x_i), \qquad S = -\int_{\mathcal{X}} p(x) \log p(x)\, dx. \tag{2.5}$$

The relationship between Shannon and Gibbs entropy has profound implications for the thermodynamic properties of computation [47]. In neuroscience, entropy has also become an important quantity, used to investigate neural coding [53], the mechanisms by which neurons encode information in their activity, and to construct measures of brain complexity for differentiating between experimental conditions and states of consciousness [54].

Whilst entropy, thanks to Shannon's information theory, has found applications in diverse fields, the idea of entropy production and dissipation is mostly used in the particular case of thermodynamics at the microscopic scale. Unlike the large systems typically treated by thermodynamics, these microscopic systems, such as those found in biology or quantum mechanics, have thermal fluctuations which are no longer insignificant. As a result, the energy and entropy of the system fluctuate significantly, thus must be described probabilistically. From this emerges the paradigm of *stochastic thermodynamics* [44].

The entropy production rate Considering a stochastic process, we can describe the probability density that the process takes a specific path. This is known as a *path-measure*.[4] Given the measure of both a specific path and its time-reversal, we can define the EPR of a stochastic process at time t to be,

$$\Phi(t) = \lim_{\epsilon \to 0} H(\mu_{x_{[t,t+\epsilon]}}, \mu_{\widetilde{x}_{[t,t+\epsilon]}}), \tag{2.6}$$

where H is the relative entropy of the forward, μ_x, and time-reversed, $\mu_{\widetilde{x}}$, path-measures of a stochastic process, $x(t)$, on the interval $[t, t+\epsilon]$. Taking the limit gives the EPR at the instant t [55, 57]. We will rarely refer to this more abstract and measure-theoretic definition of the EPR, and will instead focus on its the specific form for SDEs and Markov chains in the stationary state, when the EPR is no longer time-dependent, but a positive constant that measures the distance of the process from thermodynamic equilibrium [55, 57].

[3] We usually do not include Boltzmann's constant when using the abstract definition of entropy.
[4] See Ref. [55–57] for measure-theoretic details using probability spaces.

The *relative entropy*, also known as the *Kullback-Leibler divergence (KLD)*, measures the divergence between two probability distributions [10]. It can be defined for both continuous and discrete distributions,

$$H(p|q) = \sum_{x\in\mathcal{X}} p(x) \log \frac{p(x)}{q(x)}, \qquad H(p|q) = \int_{\mathcal{X}} p(x) \log \frac{p(x)}{q(x)}\, dx, \tag{2.7}$$

where p and q are distribution or density functions. It is non-negative, and zero if and only if $p = q$.

Eq. (2.6) shows that the EPR is defined as the divergence between the density of a particular path forward in time and its time-reversal. This highlights that the EPR is defined as the irreversibility of a stochastic process. This violation of reversibility implies that the process is out of equilibrium. Irreversibility and the EPR of stochastic processes have important implications for the thermodynamics of fluctuating systems [58]. Physically, the EPR quantifies the rate of irreversible dissipation required to sustain nonequilibrium currents in a steady-state, and therefore quantifies the heat dissipated by a system to its environment.[5] This has made it an important measure in biophysics for studying the metabolic cost of biological processes, such as molecular motors [59]. The EPR can also be thought of as the information-theoretic evidence for the 'arrow of time', as it quantifies the degree to which a process can be distinguished from its time-reversal [60].

2.2.2 Stochastic differential equations

An SDE is the prototypical example of a stochastic process taking values in $\mathbb{R}^N$, and thus for modelling continuous-valued neural systems, such as those found in neuroimaging and encephalography.

Let $x(t) \in \mathbb{R}^N$ be a stochastic process. It is governed by an SDE if it is a realisation of the stochastic integral equation,[6]

$$x(t) - x(0) = \int_0^t f(x(s))\, ds + \Sigma(x)\, dw(s), \tag{2.8}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ defines a vector field, known as the *drift*, $\Sigma(x) \in \mathbb{R}^{N\times M}$, and where $w(t)$ is a Wiener process with M independent components [56].

The Wiener process is a real-valued, continuous-time process that satisfies,

[5] There are a few exceptions to this relationship, such as Hamiltonian dynamics, where the physical EPR differs from the mathematical. Whilst we do not confront this case in this thesis, we point the reader to Remark 2.7 in Ref. [55].

[6] This integral is only defined if we adopt a convention for stochastic integration. Throughout this thesis, we will use the Itô convention, as opposed to the Stratonovich or Klimontovich conventions. For further details see Ref. [56].

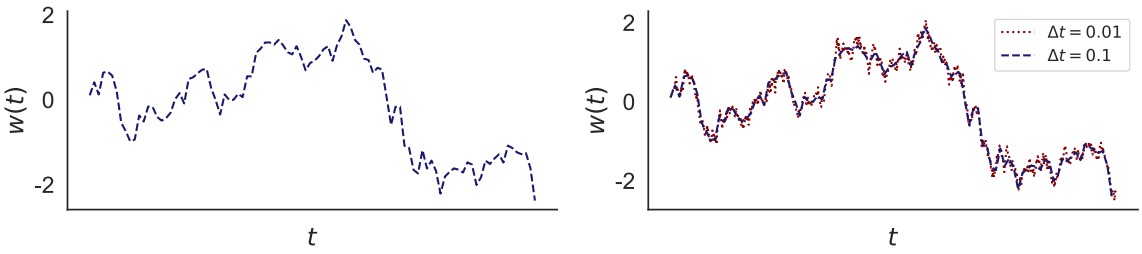

Figure 2.2: **Examples of the Wiener process.** We plot a single path sampled from the Wiener process with $\Delta t = 0.1$ and $\Delta t = 0.01$.

- $w(0) = 0$.
- $w(s_1) - w(t_1)$ is independent of $w(s_2) - w(t_2)$ if $t_1 \leq s_1 < t_2 \leq s_2$.
- $w(t + s) - w(t) \sim \mathcal{N}(0, s)$ where $\mathcal{N}$ is the normal distribution.
- w has almost surely continuous sample paths.[7]

Fig. 2.2 shows an example path from the Wiener process for $\Delta t = 0.1$ and 0.01, respectively.

We typically write Eq. (2.8) in the form,

$$dx(t) = f(x)\, dt + \Sigma(x)\, dw(t). \tag{2.9}$$

Additionally, we define $D(x) = \frac{1}{2}\Sigma(x)\Sigma^\top(x) \in \mathbb{R}^{N\times N}$ to be the *diffusion matrix*. We have already assumed that f and D do not depend on t, and will additionally assume that they are Lipschitz continuous and satisfy the linear growth conditions needed for a strong solution to the SDE (see Ref. [56] for details). Unless otherwise stated, one can assume $N = M$, and that D is full-rank everywhere that the process can reach with non-zero probability.[8]

To sample paths from an SDE we often use a numerical approximation. The standard approximation is the *Euler-Maruyama* discretisation, a stochastic extension of the forward Euler approach for ODEs,

$$x_{t+\Delta t} = x_t + \Delta t[f(x_t)] + \Delta w_t[\Sigma(x_t)], \tag{2.10}$$

[7] Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, where Ω is the sample space, $\mathcal{F}$ is a σ-algebra of subsets of Ω (the events), and $\mathbb{P}$ is a probability measure on $\mathcal{F}$. There exists an event $\Omega_0 \in \mathcal{F}$ with $\mathbb{P}(\Omega_0) = 1$ such that, for every $\omega \in \Omega_0$, the sample path $t \mapsto w_t(\omega)$ is continuous on $[0, \infty)$. For a more detailed introduction using probability spaces and a proof that such a process exists, we point the reader to Ref. [56].

[8] This is common in network dynamical systems where noise is added to each node individually i.e. D is diagonal or a multiple of the identity.

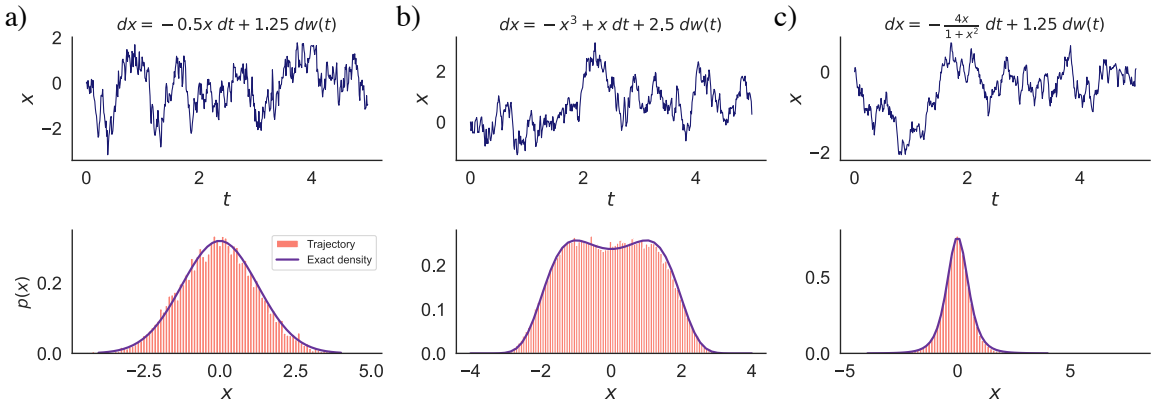

Figure 2.3: **Example SDEs in one dimension.** $a-c$) Example paths from stationary SDEs over $\mathbb{R}$. The paths are sampled from the Ornstein-Uhlenbeck, double-well potential, and logarithmic potential processes respectively. Their stationary densities are Boltzmann distributions.

where Δt is the time-step and Δw_t are independent and identically distributed (i.i.d) multivariate normal random variables with mean zero and covariance $\Delta t I_M$ [61]. Unless otherwise stated, paths will be sampled with this method. Fig. 2.3 shows example paths from 1-dimensional SDEs alongside their empirical and stationary densities.

2.2.2.1 The Fokker-Planck equation

Due to the stochasticity of the process, each path sampled from an SDE is different. However, we can describe the evolution of the *probability density function (PDF)* of the process – so-called 'density dynamics'. We denote the probability density that the process $x(t)$ takes the specific value x at time t by $p(x,t)$, whose dynamics are then given by a PDE called the *Fokker-Planck (FP)* equation [56],

$$\partial_t p(x,t) = -\sum_{i=1}^{N} \partial_{x_i}[f_i(x)p(x,t)] + \sum_{i,j=1}^{N} \partial^2_{x_i,x_j}[D_{ij}(x)p(x,t)]. \tag{2.11}$$

Remark 2.2.1. Whilst we opt to use lowercase letter $x(t)$ for a solution to the SDE in Eq. (2.9),[9] it is worth noting that this is a random variable. On the other hand, the argument x of the probability density $p(x,t)$ is not.

[9]We opt for lowercase letters for both x, the solution to the SDE and w, the Wiener process, to avoid conflicting with X, the edge-flow, introduced in Sec. 2.3, and W which is often the weight matrix for a network.

Next, we can define the *probability flux* to be,

$$J(x,t) = f(x)p(x,t) + \nabla \cdot (D(x)p(x,t)), \tag{2.12}$$

where $\nabla\cdot$ is the divergence operator.[10] As a result, the FP equation can be written as a conservation law,

$$\partial_t p(x,t) = -\nabla \cdot J(x,t). \tag{2.13}$$

If the flux is defined on an infinite domain, we additionally enforce that it decays at infinity. The density is said to be *stationary* if it no longer evolves in time i.e. $\partial_t p(x,t) = 0$, in which case the density is called the *stationary density*, which we denote by $\pi(x)$, and satisfies,

$$\nabla \cdot J_{\text{ss}}(x) = 0, \tag{2.14}$$

$$J_{\text{ss}} = f(x)\pi(x) - \nabla \cdot (D(x)\pi(x). \tag{2.15}$$

The *stationary probability flux*, $J_{\text{ss}}(x)$, is a divergence-free vector field. As stationarity does not preclude that paths are still evolving in time, the process is said to be in a *steady-state*.

Example 2.2.1. As an example, consider a smooth and confining potential, $V : \mathbb{R}^N \to \mathbb{R}$.[11] We define an SDE where the drift is the negative gradient of the potential and the diffusion is a multiple of the identity, known as *isotropic diffusion*, with constant intensity $\sigma > 0$,[12]

$$dx(t) = -\nabla V(x)\, dt + \sigma\, dw(t). \tag{2.16}$$

If the potential has local minima, in the absense of noise this system follows the direction of steepest descent of the potential and arrives at a stable local minimum. However, in the presence of noise, there is a possibility that a sufficiently large fluctuation pushes the process out of the local minimum, causing it to converge to some other minimum point. It is simple to verify that the stationary density is given by,

$$\pi(x) = \frac{1}{Z} \exp\left(-\frac{2V(x)}{\sigma^2}\right), \tag{2.17}$$

[10] The divergence acts row-wise on matrix-valued functions to give a vector field whose $i-$th entry is given by $(\nabla\cdot D(x))_i = \sum_j \partial_{x_j} D_{ij}(x)$. It acts on vector fields to give a scalar $\nabla\cdot f(x) = \sum_j \partial_{x_j} f_j(x)$.

[11] We assume $V \in C^2(\mathbb{R}^d)$ is confining, i.e. $\lim_{|x|\to\infty} V(x) = +\infty$ and $Z := \int_{\mathbb{R}^d} e^{-2V(x)/\sigma^2}\, dx < \infty$. For further details see Ref. [56].

[12] This is sometimes known as a *Smoluchowski dynamic*, but we will refer to it as a *gradient flow*.

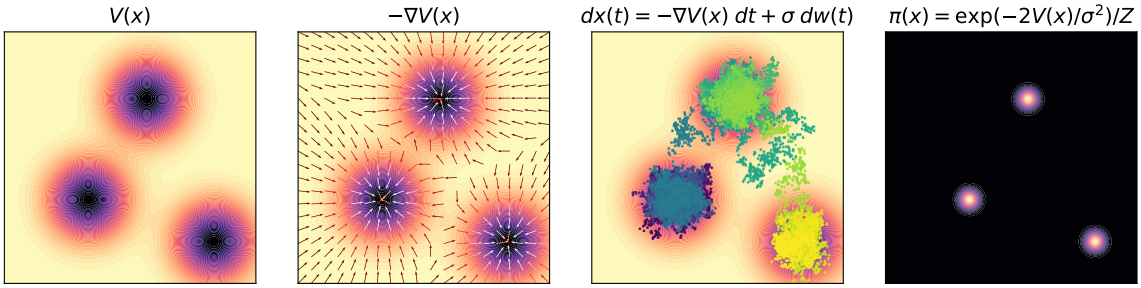

Figure 2.4: **A potential-gradient process**. An illustrative example of a potential on $\mathbb{R}^2$ with three minima, the drift and SDE that its negative gradient defines, and its stationary Boltzmann density.

where Z is a normalising constant, and the continuous analogue of the partition function defined in Sec. 2.2.1 and is also known as a *Boltzmann density*. The conditions on V are chosen to guarantee that Z is finite [56]. This density implies that the process spends most of its time near the minimum points. Fig. 2.4 shows an example of such a potential, its negative gradient, a sampled path from the process, and its stationary Boltzmann density.

2.2.2.2 Nonequilibrium stationary diffusions

A stationary SDE is said to obey the *detailed balance* condition if $J_{\text{ss}} = 0$ [57]. Note that this is a stronger condition than that of stationarity, $\nabla \cdot J_{\text{ss}} = 0$. The detailed balance condition implies that the process is in an *equilibrium steady-state (ESS)*, and thus has zero entropy production. On the other hand, if $J_{\text{ss}} \neq 0$ then the process is in an NESS and detailed balance is violated.

In the case that the diffusion matrix, $D(x)$, is uniformly positive definite, Jiang et al. derive the EPR to be,

$$\Phi = \int_{\mathbb{R}^N} \frac{J_{\text{ss}}^\top(x) D^{-1}(x) J_{\text{ss}}(x)}{\pi(x)}\, dx, \tag{2.18}$$

where the assumptions guarantee that D has a unique inverse [57]. It is clear that this is non-negative as it is the integral of a (positive-definite) quadratic form, and is zero if and only if $J_{\text{ss}} = 0$ i.e the process is in detailed balance.

The Helmholtz-Hodge decomposition Stationary SDEs admit the *Helmholtz-Hodge decomposition (HHD)*, which splits the drift into two components. The HHD

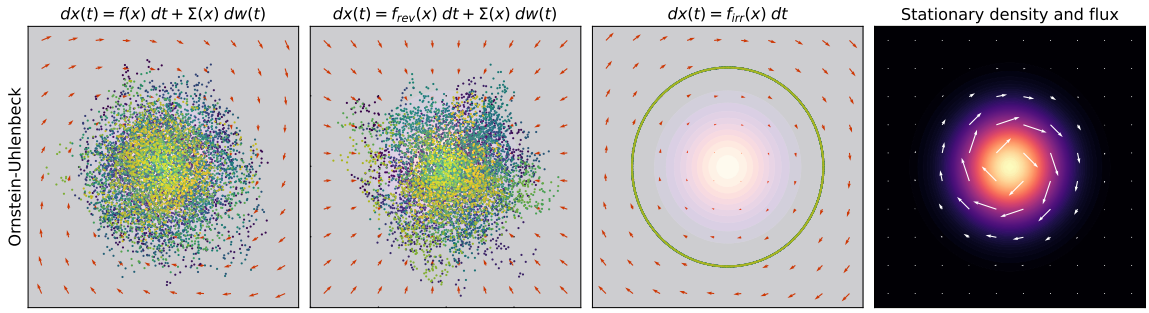

Figure 2.5: **The Helmholtz-Hodge decomposition**. We plot the drift and a sample path from the each component of the HHD of the Ornstein-Uhlenbeck process, as well as its stationary density and flux.

is given by,

$$f(x) = f_{\text{rev}}(x) + f_{\text{irr}}(x), \tag{2.19}$$

$$f_{\text{rev}}(x) = D(x)\nabla \log \pi(x) + \nabla \cdot D(x), \tag{2.20}$$

$$f_{\text{irr}}(x) = J_{\text{ss}}(x)/\pi(x). \tag{2.21}$$

where f_{rev} represents the time-reversible drift of the process, whilst f_{irr} represents the time-irreversible drift [55]. More specifically, f_{rev} is even under time-reversal whilst f_{irr} is odd. As $\nabla \cdot (f_{\text{irr}}\pi) = 0$, we say that f_{irr} is *divergence free with respect to the stationary density (DFSD)*. This decomposition provides a parsimonious and interpretable splitting of the forces at play in the NESS, and we will return to it a number of times.[13]

As illustrated in Fig. 2.5, the reversible component balances the diffusion to maintain the process on the stationary distribution, whilst the irreversible component contours the stationary density in a rotational manner. We can also rewrite the EPR in Eq. (2.18) in the following equivalent form,

$$\Phi = \int_{\mathbb{R}^N} f_{\text{irr}}^{\top}(x) D^{-1}(x) f_{\text{irr}}(x) \pi(x)\, dx, \tag{2.22}$$

where it is clear that it is zero if and only if $f_{\text{irr}} = 0$ i.e. the process is reversible.

Finally, we can define the *thermodynamic force* to be the field,

$$h(x) = D^{-1}(x)(f(x) - \nabla \cdot D(x)). \tag{2.23}$$

This gives a neat criterion for detailed balance and an ESS.

[13] It is related, but not identical to, the traditional *Helmholtz decomposition* of a vector field into a gradient and divergence-free field [62].

Corollary 2.2.1. *Given a stationary SDE with thermodynamic force $h(x)$, the process obeys detailed balance if and only if h is conservative i.e. there exists a potential $V(x)$ such that $h = \nabla V$.*[14]

Remark 2.2.2. In the case that $D(x)$ is not positive definite, Da Costa and Pavliotis derive the EPR in two distinct cases [55]. If $f_{\text{irr}}(x) \in \text{Range}(\Sigma(x))$ for all $x \in \mathbb{R}^N$, then we can replace $D^{-1}(x)$ with its Moore-Penrose (MP) pseudo-inverse, $D^\dagger$,[15]

$$\Phi = \int_{\mathbb{R}^N} f_{\text{irr}}^\top(x) D^\dagger(x) f_{\text{irr}}(x) \pi(x)\, dx. \tag{2.24}$$

If not, then $\Phi = \infty$.

We now consider some simple, immediate results and examples.

Corollary 2.2.2. *A stationary SDE over $\mathbb{R}$ must obey detailed balance. If it has constant diffusion, then it is a gradient flow and converges to a Boltzmann density.*

This follows immediately by noticing that the divergence in one dimension is simply the derivative. The stationarity condition, $\nabla \cdot J_{\text{ss}} = \partial_x(J_{\text{ss}}) = 0$, implies the flux is constant and, as it must vanish at infinity, it is zero everywhere. If we then consider constant diffusion with intensity σ, from the HHD we obtain that the drift must be $f(x) = \nabla\left(\frac{\sigma^2}{2}\log(\pi(x))\right)$.

Example 2.2.2. We return to the negative potential-gradient example of Ex. 2.2.1 i.e. $f = -\nabla V$, $D = \frac{\sigma^2}{2}$. Either by computing the flux directly or via the HHD, we can see that this process obeys detailed balance, thus is reversible, and has an EPR of zero.

Example 2.2.3. One-dimensional SDEs with periodic boundary conditions need not obey detailed balance. For example, we can consider an isolated phase oscillator,

$$d\theta(t) = \omega\, dt + \sigma\, dw(t), \tag{2.25}$$

where $\theta(t) \in [0, 2\boldsymbol{\pi})$ and $\omega > 0$ is the natural frequency [66].[16] We can show that this SDE has a uniform stationary density around the unit circle. This process has

[14] Whilst this result is immediate from the HHD, it is first proved with Stratonovich convention in Ref. [63], and with Itô convention in Ref. [64].

[15] The MP inverse is a generalisation of the matrix inverse for singular or rectangular matrices. The MP inverse of A, $A^\dagger$, is the unique matrix that satisfies $AA^\dagger A = A$, $A^\dagger AA^\dagger = A^\dagger$, $(AA^\dagger)^\top = AA^\dagger$, and $(A^\dagger A)^\top = A^\dagger A$ [65].

[16] We use $\boldsymbol{\pi}$ to denote the ratio of a circle's circumference and diameter, as opposed to $\pi(x)$ which is the stationary density.

$f_{\text{irr}}(x) = \omega$, $J_{\text{ss}} = \omega/2\boldsymbol{\pi}$, and $\Phi = 2\omega^2/\sigma^2$. This implies that a process driven around the circle in a particular direction is irreversible, and that the irreversibility depends quadratically on its speed of rotation.[17]

We will consider a number of further examples in Ch 3, 5, and 7.

Remark 2.2.3 (Even and odd variables)**.** The definitions of detailed balance and the EPR given in this section apply to SDEs composed of even variables e.g. spatial positions. When the system contains odd variables, such as velocity or momentum, there is an alternative formulation of detailed balance [67, 68]. This arises due to the fact that odd variables change sign under time-reversal.

Consider an SDE where $\varepsilon_i = \pm 1$ represents the time-reversal parity of variable x_i. Gardiner offers a generalisation of the detailed balance condition for this case [67],

$$\pi(x)p_t(x', x) = \pi(\varepsilon x')p_t(\varepsilon x, \varepsilon x'), \tag{2.26}$$

where $p_t(x, y)$ is the transition probability i.e. the probability the process is at x at time t given it started in y at time zero. One can go further to write the conditions for detailed balance in terms of a decomposition of the flux into irreversible and reversible components (see Sec. 5.3 of Ref. [67]).

The consequences of this are that equilibrium systems with odd variables, such as Hamiltonian underdamped Langevin dynamics, do not satisfy $J_{\text{ss}} \equiv 0$, yet have zero *physical* entropy production. In light of this contradiction, Da Costa and Pavliotis differentiate between what they call time-reversibility ($J_{\text{ss}} \equiv 0$) and 'generalised reversibility' where the parity of odd variables can be taken into account when computing the EPR [55].

In this thesis, we often consider variables which have no obvious parity, such as neural recordings. Following the standard classification of variables by their behaviour under time reversal [69], we consider these variables to be functions of the system's configuration and therefore treat them as even. This is consistent with previous studies on NESSs in neural systems [3].

Remark 2.2.4 (Dissipation and reversibility in deterministic dynamics)**.** We introduced the irreversibility of an SDE in an NESS. This definition has a well-established deterministic counterpart. Given an ODE, $\dot{x} = f(x)$, the flow φ_t is called *reversible*

[17] Quadratic relationships between asymmetry parameters and the EPR will emerge a number of times. This is a consequence of the EPR being a quadratic form of the irreversibility (Eq. (2.22)).

if there is an involution ϖ with $\varpi \circ \varphi_t = \varphi_{-t} \circ \varpi$ [70].[18] This definition of reversibility is most similar to the definition of detailed balance with odd variables given in Eq. (2.26). In thermostatted deterministic systems, those that are kept at a fixed temperature despite driving, the EPR has an analogous quantity in the mean phase-space contraction rate,

$$-\langle \nabla \cdot f \rangle = -\sum_i \lambda_i, \tag{2.27}$$

where λ_i are the Lyapunov exponents [71].

The flow of the ODE is called dissipative if there is a bounded set that every trajectory eventually enters and remains in [72]. This can be certified by a Lyapunov function i.e. a function V with compact sublevel sets that does not increase along trajectories,

$$\dot{V} = \nabla V \cdot f \leq 0, \tag{2.28}$$

and decreases strictly outside a compact set. All trajectories then converge to the largest invariant subset of $\{\dot{V} = 0\}$, and the flow possesses a compact global attractor. The quantity $-\dot{V} \geq 0$ is the rate which V decreases along trajectories as they approach the attractor and is known as the dissipation rate.

We introduced the EPR at stationarity. However, it can also be defined along the relaxation to stationarity. In this case, it decomposes into two non-negative components [73].[19] The first, the *non-adiabatic* EPR, equals the dissipation rate $-\dot{V}$ of the Fokker–Planck flow on densities, with Lyapunov function given by the relative entropy to the stationary density.[20] The second, called the "housekeeping" EPR, is the cost of maintaining the stationary probability current. At stationarity the non-adiabatic term vanishes and the housekeeping EPR reduces to the quantity defined in Eq. (2.18).

2.2.3 Discrete-state Markov processes

2.2.3.1 Continuous-time Markov chains

Some stochastic processes take values in a countable and discrete state-space $x(t) \in \Omega \subseteq \mathbb{N}$. In continuous time, such processes appear to 'jump' between states at ran-

[18] An involution is a map that is its own inverse. Explicitly, an involution $\varpi : \Omega \to \Omega$ satisfies $\varpi(\varpi(\omega)) = \omega$ for all $\omega \in \Omega$.

[19] A similar decomposition exists for the EPR of a discrete Markov process as it relaxes [74]. In the next section, we will focus only on its steady-state.

[20] That is, $V(t) = H\big(p(\cdot, t) \,|\, \pi\big)$, where $p(x, t)$ is the density of the process at time t and π is the stationary density.

domly distributed intervals. Describing the process typically requires both a *waiting-time distribution (WTD)*, $\varphi_i(\Delta t)$, which defines a density over positive intervals, $\Delta t > 0$, where the process remains in state i, and transition probabilities between all pairs of states [75]. As with the FP equation, it is natural to describe the density dynamics of this process, where $p_i(t)$ is the probability that the process is in state i at time t, given that it was distributed initially with probability vector $p(0) \in \mathbb{R}^{|\Omega|}$.[21] The (possibly infinite) state probability vector $p(t) = (p_1(t), p_2(t), \ldots)$ describes the density dynamics of the entire process. In general, for state-dependent waiting-times, the evolution, $p(t)$, cannot be written in a closed form.

When considering SDEs, we are naturally restricted to processes that are *Markovian* i.e. the dynamics at time t depend entirely on the state $x(t)$ and are independent of $x(s)$ for $s < t$ [76]. Markovianity is a common and extremely important assumption that is implicit in the definition of the Itô SDE [56]. For a general WTD, discrete-state dynamics are not Markovian as the duration that the process has been waiting may influence when a jump will occur i.e. the future of the process at time t does depend on the history prior to t. Markovianity arises from the common and simplifying assumption that the WTD is an exponential distribution as it has the following 'memoryless property'.

Lemma 2.2.1. *Let τ be a random variable with distribution Exp(λ), for all $a, b \geq 0$, the cumulative distribution satisfies,*

$$P(\tau > a + b | \tau > a) = P(\tau > b). \tag{2.29}$$

This implies that the probability of an event occurring within a given future window is independent of the past i.e. how long the process has already been in that state. Under the assumption that all waiting times are exponential, $\varphi_i \sim \text{Exp}(\lambda_i)$, the dynamics become Markovian. The process is known as a continuous-time Markov chain (CTMC), and its dynamics are given by the *master equation (MEq)*,

$$\frac{dp}{dt} = Lp, \tag{2.30}$$

which is an ODE for the evolution of $p(t)$ [76]. Here, L is the *Laplacian* matrix which contains the *transition rates* between states. More specifically, for $i \neq j$, $L_{ij} > 0$ is the transition rate from state j to state i. Additionally, diagonal entries are set to $L_{jj} = -\sum_i L_{ij}$ which enforces that columns of L sum to zero. This ensures that probability is conserved over time i.e. $\sum_i p_i(t) = \sum_i p_i(0) = 1$. The master equation is derived by taking exponential WTDs with rates $\lambda_i = -L_{ii} > 0$.

[21] Alternatively, this can be seen as the proportion of paths that will be in state i at time t given an infinite number of paths distributed initially with probability vector $p(0)$.

Nonequilibrium steady-states As in the case of the FP equation, the MEq admits a stationary distribution, $\pi \in \mathbb{R}^{|\Omega|}$, when $dp/dt = 0$ which occurs when $L\pi = 0$ i.e. π is in the kernel of the Laplacian.

A CTMC is said to be *irreducible* if each discrete state can be reached from each other state in a non-zero time-interval. It is said to be *periodic* if it can only return to a state in a multiple of $d > 1$ steps, and said to be *aperiodic* otherwise. We have the following canonical result for ergodicity [76].

Lemma 2.2.2. *Given an irreducible CTMC with a finite state-space with Laplacian $L \in \mathbb{R}^{N\times N}$, there exists a unique $\pi \in \mathbb{R}^N$ such that $L\pi = 0$ and $\sum_{i=1}^N \pi_i = 1$. If additionally the chain is aperiodic, then $p(t) \to \pi$ as $t \to \infty$.*

For a countably infinite state-space, we additionally need that every state has finite expected return time, known as *positive recurrence*. The Perron-Frobenius theorem guarantees that all entries of π are non-negative.

Time-reversibility of a Markov chain is given by the discrete form of the detailed balance condition [76].[22]

Lemma 2.2.3. *Given a stationary CTMC with Laplacian L and stationary distribution π. The process is time-reversible if and only if,*

$$L_{ij}\pi_j = L_{ji}\pi_i, \tag{2.31}$$

for all pairs (i, j). This is known as the detailed balance condition.

For the stationary CTMC, Jiang et al. derive the EPR to be,

$$\Phi = \frac{1}{2}\sum_{i,j}(L_{ij}\pi_j - L_{ji}\pi_i)\log\frac{L_{ij}\pi_j}{L_{ji}\pi_i}, \tag{2.32}$$

which is also known as the *Schnakenberg formula* [57,77]. This vanishes if the CTMC satisfies detailed balance, and is positive if it does not.

2.2.3.2 Discrete-time Markov chains

Another common formulation of a Markov chain is in discrete-time. A *discrete-time Markov chain (DTMC)* is a stochastic process, $x(t) \in \Omega$, taking values at times $t = 1, 2, \ldots$. As jumps can only occur at specified time-steps, the process is defined entirely by the transition probabilities,

$$T_{ij} = P(x(t+1) = i | x(t) = j), \tag{2.33}$$

[22] For proofs of Lemmas 2.2.1, 2.2.2, 2.2.3 see Ref. [76].

which form the *transition probability matrix (TPM)*, T, whose columns sum to 1.[23] The density dynamics are given by $p(t) \in \mathbb{R}^{|\Omega|}$ which satisfies,

$$p(t) = Tp(t-1) = T^t p(0), \tag{2.34}$$

for some initial $p(0)$ with $\sum_i p_i(0) = 1$. As in the case of the CTMC, if a DTMC is irreducible and aperiodic, it converges to a unique stationary distribution which satisfies,

$$\pi = T\pi, \tag{2.35}$$

i.e. π is the Perron right-eigenvector of T such that $\sum_i \pi_i = 1$. We also have a detailed balance condition for the DTMC [76].

Lemma 2.2.4. *Given a stationary DTMC with transition probability matrix T and stationary distribution π. The process is time-reversible if and only if,*

$$T_{ij}\pi_j = T_{ji}\pi_i, \tag{2.36}$$

for all pairs (i, j). This is known as the detailed balance condition.

It is worth noting that $P_{ij} = T_{ij}\pi_j$ is the *joint* transition probability[24] in the steady-state i.e. $P_{ij} = P(x(t+1) = i \text{ and } x(t) = j)$. The detailed balance condition is then $P_{ij} = P_{ji}$ for all (i, j). The EPR for the DTMC then simplifies to,

$$\Phi = \sum_{i,j} P_{ij} \log \frac{P_{ij}}{P_{ji}}, \tag{2.37}$$

which is specifically the relative entropy between the forward and backward joint transition probabilities [3].

Remark 2.2.5. In both a C- and DTMC, it is evident that if the transition $i \to j$ occurs with non-zero probability, but $j \to i$ has zero probability, then the EPR diverges. A Markov chain satisfies *microscopic reversibility* if for any transition that occurs with non-zero probability, the reverse transition occurs with non-zero probability – an assumption that is typical when considering any physical system [64]. Some attempts have been made to define a finite EPR in chains with unidirectional transitions [78], and in Ch. 3, we will consider one method to quantify irreversibility in this case.

[23] Often T is defined with an opposite convention i.e. T_{ij} is the transition from i to j. Our choice to be consistent with the convention for directed networks [7].

[24] As opposed to the *conditional* transition probability, T_{ij}.

Remark 2.2.6 (Random walks)**.** In this thesis, we will reserve the term random walk (RW) to refer to a discrete-state Markov process (in either discrete or continuous time), modelled on a network [7]. The term is also often used for continuous-space processes such as the Wiener process defined in Sec. 2.2.2, but we will not refer to them as such.

Example 2.2.4 (Random walks on networks)**.** Any Markov chain can be seen as a RW on a network, where the nodes are possible states and the weighted, directed edges represent transitions [79]. The RW on the network, shown in Fig. 2.6, is a canonical system for both studying the structure of a network via dynamics, and for casting properties of the chain in terms of network structure. As a result it has been used to develop techniques for detecting community structure in graphs [80–82].

Given a positive-weighted, directed network with weight-matrix A, the RW in discrete-time is defined by the transition probability matrix,

$$T_{ij} = \frac{A_{ij}}{k_j}, \tag{2.38}$$

where $k_j = \sum_i A_{ij}$.

In continuous-time, we have a choice whether we allow the WTD at each node to depend on the total weight incident to a node (the nodal degree[25] for unweighted networks) or to be the same for each node. In the first case, the MEq is defined by the weighted graph Laplacian,

$$L = A - D, \tag{2.39}$$

where $D = \mathrm{diag}(k_1, \ldots, k_n)$, which means that the WTD at node i is given by $\mathrm{Exp}(k_i)$.

Alternatively, one can define T as in discrete-time, and then define the MEq with the *normalised random walk Laplacian*,

$$L = T - I, \tag{2.40}$$

where I is the identity matrix [83]. Here the WTD is $\mathrm{Exp}(1)$ at each node and it is immediate that the stationary distribution of this process coincides with that of the RW in discrete-time i.e. $T\pi = \pi$ implies $(T - I)\pi = 0$. This is not the case for the weighted graph Laplacian as $T\pi = \pi$ does not imply $(A - D)\pi = 0$.

[25] The *degree* of a node is the number of edges. For directed networks we differentiate between *in-degree*, the number of edges that terminate at a node, and the *out-degree*, the number of edges that begin at a node.

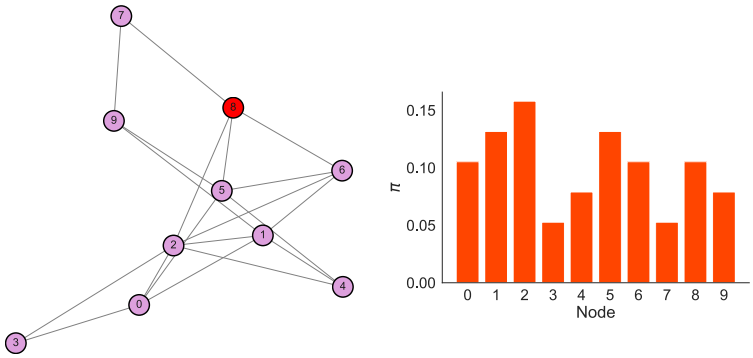

Figure 2.6: **Random walk on a network.** Any Markov chain can be represented as random walk on a weighted, directed network where states become nodes and transitions are edges. The random walk is a prototypical model to investigate network structure via dynamics.

A network is said to be *connected* if there exists a path between any pair of nodes. A directed network is said to be *strongly connected* if a directed path exists between every pair of nodes. The irreducibility condition of the Markov chain becomes equivalent to strong connectedness for the network [7].

We have the following theorem for stationary RWs on networks.

Theorem 2.2.1. *The RW on an undirected network, in discrete or continuous-time, satisfies detailed balance and time-reversibility.*

Proof. First we consider the case of the discrete-time and continuous-time RW with normalised Laplacian, $L = T - I$. The stationary distribution is $\pi_i = k_i / \sum_j k_j$, which can be verified by noticing,

$$[(T - I)k]_i = \sum_j T_{ij} k_j - k_i = \sum_j A_{ij} - k_i = 0, \tag{2.41}$$

where we use that $A_{ij} = A_{ji}$ in the final step. Detailed balance follows immediately as $L_{ij} k_j = L_{ji} k_i \Leftrightarrow T_{ij} k_j = T_{ji} k_i \Leftrightarrow A_{ij} = A_{ji}$.

Next we consider $L = A - D$. In this case the stationary distribution is uniform, $\pi_i = 1/N$ where we have N nodes, as shown below,

$$[(A - D)\pi]_i = \sum_j A_{ij} \pi_j - k_i \pi_i = \frac{1}{N} \left(\sum_j A_{ji} - k_i \right) = 0. \tag{2.42}$$

Detailed balance again follows immediately as $L_{ij} = L_{ji} \Leftrightarrow A_{ij} = A_{ji}$. □

We have shown that directedness is a necessary condition for a RW on a network to be in an NESS. In Ch. 3, we will investigate this further, including showing that it is not a sufficient condition i.e. directed networks may preserve detailed balance.

In this section, we have introduced a number of different notions of stochastic processes, and characterised their out of equilibrium behaviour under diverse and increasingly prescriptive assumptions. Table 2.1 offers a summary of these models, their assumptions, and their NESSs.

2.3 Hodge theory on graphs

We turn now to discrete Hodge theory for graphs and simplicial complexes. At first glance, this may appear disconnected with the study of stochastic processes or neural dynamics, however, as we will show in Ch. 5 and 6, it has important mathematical implications and is of significant practical use [64]. In brief, discrete Hodge theory generalises the Laplacian operator to discrete spaces. This begins with the standard graph Laplacian and then extends to 'higher-order' Hodge Laplacians. In addition, we will introduce the Hodge decomposition which has much in common with the HHD of a stochastic process introduced in Sec. 2.2.2.2. For this exposition, we will closely follow Ref. [84].

2.3.1 Flows and operators on graphs

Let $G = (V, E)$ be an undirected, unweighted graph with N nodes and M edges.[26] Additionally, let T be the set of *triangles*, 3-cliques, in the clique complex defined by the graph (see the definition in Sec. 2.1). More explicitly,

$$\{i, j, k\} \in T \text{ if and only if } \{i, j\}, \{i, k\}, \{j, k\} \in E, \tag{2.43}$$

and where Θ is the cardinality of T.

We can then define real-valued functions on the nodes, edges, and triangles of this graph. A *potential* on the nodes is a real-valued function $f : V \to \mathbb{R}$. A *flow* on the edges is given by $X : V \times V \to \mathbb{R}$, where we enforce that,

$$X(i, j) = -X(j, i), \tag{2.44}$$

for $\{i, j\} \in E$, which is referred to as the *alternating* property, and $X(i, j) = 0$ if $\{i, j\} \notin E$. A function on the triangles is given by $\phi : V \times V \times V \to \mathbb{R}$, such that,

$$\phi(i, j, k) = \phi(j, k, i) = \phi(k, i, j) = -\phi(j, i, k) = -\phi(i, k, j) = -\phi(k, j, i), \tag{2.45}$$

[26] In this section, we will use curly braces $\{\cdot\}$ for edges and triangles and opposed to $(\cdot)$. As will become apparent, it is necessary to choose an arbitrary, but consistent, orientation for structures in this theory, thus (i, j) represents a specific orientation and $\{i, j\}$ is an unoriented set.

Level	**Additional assumption**	**Entropy production rate**	**Implication**
General			
Stochastic process	Path measure μ_x and its time reversal $\mu_{\tilde{x}}$	$\Phi(t)$: rate of $H(\mu_x \vert \mu_{\tilde{x}})$, Eq. (2.6)	$\Phi(t) \geq 0$, with equality if and only if the process is time-reversible
Diffusions			
SDE	Drift split $f = f_{\text{rev}} + f_{\text{irr}}$ and even variables	$\Phi(t)$ in terms of f_{irr}	$\Phi(t) = \infty$ if $f_{\text{irr}} \notin \text{Range}\, \sigma$
Stationary SDE	$p(x,t) = \pi(x)$ and D full-rank	$\Phi = \int f_{\text{irr}}^{\top} D^{-1} f_{\text{irr}}\, \pi \, \mathrm{d}x$, Eq. (2.18) and (2.22)	**NESS** $\Leftrightarrow f_{\text{irr}} \neq 0$
Continuous-time Markov chains			
CTMC	Rates L_{ij}	$\Phi(t) = \frac{1}{2}\sum_{i,j}(p_j L_{ij} - p_i L_{ji}) \log \frac{p_j L_{ij}}{p_i L_{ji}}$, Eq. (2.32)	$\Phi = \infty$ if $L_{ij} > 0$ but $L_{ji} = 0$
Stationary CTMC	$p = \pi$, irreducible and acyclic	Same with $p = \pi$	**NESS iff** $\pi_j L_{ij} \neq \pi_i L_{ji}$ **for some** i, j
Discrete-time Markov chains			
DTMC	Transition matrix T_{ij}	Same as CTMC with $L_{ij} \to T_{ij}$, Eq. (2.37)	$\Phi = \infty$ if $T_{ij} > 0$ but $T_{ji} = 0$
Stationary DTMC	$p = \pi$, irreducible and acyclic	Same with $p = \pi$	**NESS iff** $\pi_j T_{ij} \neq \pi_i T_{ji}$ **for some** i, j

Table 2.1: A summary table for the hierarchy of assumptions for stochastic process introduced in Sec. 2.2. Each row adds an assumption to the one above; the final column gives the resulting characterisation of equilibrium and NESS.

for all $\{i,j,k\} \in T$, which is the alternating property for triangles, and $\phi(i,j,k) = 0$ if $\{i,j,k\} \notin T$. For each edge and triangle, we must choose an (arbitrary) *orientation* for each edge and triangle in the complex, although this is simply for mathematical consistency and does not indicate a structural direction. Nodes, edges, and triangles are referred to as 0-, 1-, and 2-simplices.

The functions f, X, and ϕ are often referred to as 0-, 1-, and 2-cochains respectively [84]. We can equip the set of cochains with corresponding inner products,

$$\langle f, g\rangle_V = \sum_{i=1}^{N} w_i f(i) g(i) \qquad \langle X, Y\rangle_E = \sum_{i<j} w_{ij} X(i,j) Y(i,j)\,, \tag{2.46}$$
$$\langle \phi, \psi\rangle_T = \sum_{i<j<k} w_{ijk} \phi(i,j,k) \psi(i,j,k)\,,$$

where $\{w_i\}$, $\{w_{ij}\}$, and $\{w_{ijk}\}$ are positive weights given to each node, edge, and triangle, respectively. Given an inner product, we define the Hilbert spaces $H(V)$, $H(E)$, $H(T)$, to be the set of (alternating) functions on nodes, edges, and triangles.

The space $H(V)$ has a standard basis consisting of functions f_i, for $i = 1, \ldots, N$, where $f_i(i) = 1$ and $f_i(j) = 0$ for $j \neq i$. After choosing an orientation for each edge and triangle, the Hilbert spaces $H(E)$ and $H(T)$ have analogous standard bases which are defined similarly. As a result, we have that $H(V) \cong \mathbb{R}^N$, $H(E) \cong \mathbb{R}^M$, and $H(T) \cong \mathbb{R}^\Theta$, and thus can be represented by real vectors.[27]

The standard inner products are given by $w_i = w_{ij} = w_{ijk} = 1$ for $i \in V$, $\{i,j\} \in E$, and $\{i,j,k\} \in T$, or zero otherwise, and these define the standard L^2 inner product on the Euclidean spaces. In this case, we denote the function spaces by $L^2(V)$, $L^2(E)$, and $L^2(T)$.

Discrete operators on graphs We now introduce discrete operators, which act on these flows. The operator grad : $H(V) \to H(E)$, is defined as,

$$(\operatorname{grad} f)(i,j) = f(j) - f(i), \tag{2.47}$$

for all $\{i,j\} \in E$ and zero otherwise. This operator acts on a potential on the nodes and returns a flow on the edges, much like the continuous gradient operator acts on a real potential and returns a vector field.

[27] Note that these isomorphisms hold as vector spaces but not as Hilbert spaces. When we consider a Euclidean space, we assume the standard norm. The spaces $H(\cdot)$ do not have the standard inner product structure unless we take the standard weights for nodes, edges, and triangles.

The *divergence* is the linear operator $\operatorname{div} : H(E) \to H(V)$, which is defined by,

$$(\operatorname{div} X)(i) = \sum_{j=1}^{N} \frac{w_{ij}}{w_i} X(i,j), \tag{2.48}$$

for all $i \in V$. Fig. 2.7 shows an example of these operators applied to a potential on the nodes.

Example 2.3.1 (The combinatorial graph Laplacian)**.** Given these definitions, we can define the discrete analogue of the Laplacian operator. We define the 0-Laplacian to be the operator $\Delta_0 : H(V) \to H(V)$,

$$\Delta_0 = -\operatorname{div} \operatorname{grad}. \tag{2.49}$$

First, we assume the standard L^2 inner product and substitute in the definitions of the operators. We have that,

$$(\Delta_0 f)(i) = \sum_{j:(i,j)\in E} f(i) - f(j) = k_i f(i) - \sum_{j:(i,j)\in E} f(j)\,, \tag{2.50}$$

where k_i is the degree of node i. Using the adjacency matrix of the graph, we can write this as

$$(\Delta_0 f)(i) = k_i f(i) - \sum_{j=1}^{N} A_{ij} f(j)\,. \tag{2.51}$$

As functions can be represented by real-valued vectors, operators can be represented by real-valued matrices. The Laplacian can be written as a matrix $L_0 \in \mathbb{R}^{N\times N}$, where we can see that $L_0 = D - A$, where $D = \operatorname{diag}(k_1, \ldots, k_N)$ is the diagonal matrix of degrees. Thus, we have derived the standard 'combinatorial' graph-Laplacian [7].

The eigendecomposition of the graph-Laplacian has important implications for structure-function coupling on networks and has been used to analyse co-activation patterns in the brain, using modes derived from its connectivity [85]. However, it is well-known that this standard form for the Laplacian matrix suffers from *localisation*, meaning that its eigenmodes are concentrated on only a few high-degree nodes, which can obscure structural information [86]. To combat this, it is typical, in both analysis of neural data and other domains like spectral clustering, to use a normalised form of the 0-Laplacian [85, 87]. For a further discussion of localisation, see Sec. 6.2.2.1.

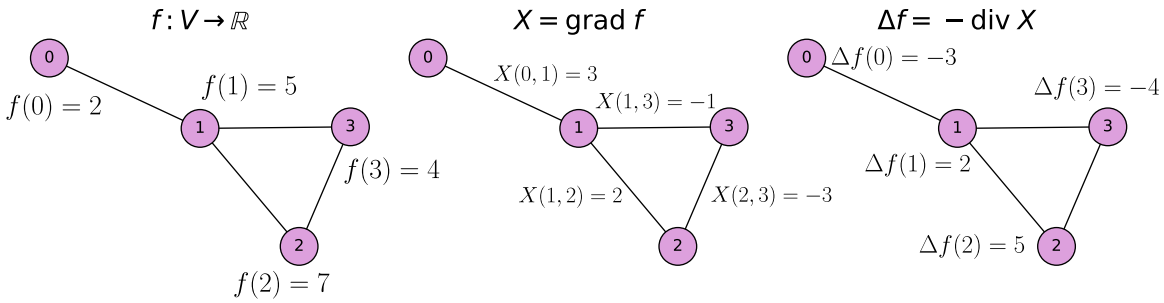

Figure 2.7: **Discrete calculus on graphs.** We illustrate Ex. 2.3.1 with an example. We take a potential on the nodes of a small graph, define a edge-flow by its gradient, then take the (negative) divergence. We obtain the Laplacian of the original potential. This coincides with what we would have obtained if we calculated Lf, where $L = D - A$. For this example, we take the natural orientation (i, j) for $i < j$.

Example 2.3.2 (Normalised Laplacians)**.** In order to prevent this localisation, we modify the inner product on the nodes by assigning the weights $w_i = k_i$ i.e. nodes are weighted according to their degree. It can then be shown that,

$$(\Delta_0 f)(i) = f(i) - \sum_{j=1}^{N} \frac{A_{ij}}{k_i} f(j)\,. \tag{2.52}$$

In matrix form this can be written as $\widehat{L}_0 = I - D^{-1}A$, which is precisely the normalised random-walk Laplacian introduced in Ex. 2.2.4.[28]

Whilst this matrix is not symmetric in general, it is similar to the symmetric, positive semi-definite matrix,

$$\widetilde{L}_0 = D^{1/2}\widehat{L}_0 D^{-1/2} = I - D^{-1/2}AD^{-1/2} = D^{-1/2}L_0 D^{-1/2}, \tag{2.53}$$

which we define as the *normalised 0-Laplacian*. It is this matrix whose eigendecomposition is typically used in both neural dynamics and spectral clustering [85, 87].

Next, we introduce the discrete curl operator, curl : $H(E) \to H(T)$, which is given by,

$$(\mathrm{curl}\, X)(i, j, k) = X(i, j) + X(j, k) + X(k, i), \tag{2.54}$$

for all $\{i, j, k\} \in T$ and zero otherwise. Its adjoint operator, curl* : $H(T) \to H(E)$, can be derived to be,[29]

$$(\mathrm{curl}^*\phi)(i, j) = \sum_{k=1}^{N} \frac{w_{ijk}}{w_{ij}} \phi(i, j, k)\,, \tag{2.55}$$

[28] Up to an arbitrary factor of -1, depending on convention.

[29] Given an inner product space and an operator A, its adjoint operator A^* satisfies $\langle A^*x, y\rangle = \langle x, Ay\rangle$ for all x and y in the space.

for all $\{i, j\} \in E$ and zero otherwise [84].

Finally, we can define the first higher-order Hodge Laplacian, referred to as the *1-Laplacian*, which is the operator $\Delta_1 : H(E) \to H(E)$, given by [84],

$$\Delta_1 = -\text{grad}\,\text{div} + \text{curl}^*\,\text{curl} = \text{grad}\,\text{grad}^* + \text{curl}^*\,\text{curl}. \tag{2.56}$$

Boundary maps and matrix representations The 1-Laplacian, Δ_1, can also be represented by a matrix, $L_1 \in \mathbb{R}^{M\times M}$. In order to derive this matrix representation, we introduce the *boundary maps*, $B_1 \in \mathbb{R}^{N\times M}$ and $B_2 \in \mathbb{R}^{M\times\Theta}$. The matrix B_1 is called the node-edge incidence matrix,[30] and is given by,

$$(B_1)_{v,e} = \begin{cases} -1 & v = s(e) \\ +1 & v = t(e) \\ 0 & \text{otherwise} \end{cases}, \tag{2.57}$$

where $s(e)$ and $t(e)$ denote the source and target of e respectively, as prescribed by our choice of orientation.

The matrix B_2 is the triangle-edge incidence matrix and is given by,

$$(B_2)_{e,t} = \begin{cases} -1 & \text{the edge } e \text{ disagrees with the orientation of the triangle } t \\ +1 & \text{the edge } e \text{ agrees with the orientation of the triangle } t \\ 0 & \text{the edge } e \text{ is not in the triangle } t \end{cases}. \tag{2.58}$$

When using the standard inner products, it can be shown that these matrices encode the discrete operators when applied to the real-valued vectors representing the functions i.e. $\text{grad} \cong B_1^\top$, $\text{div} \cong -B_1$, $\text{curl} \cong B_2^\top$, and $\text{curl}^* \cong B_2$. Therefore, the 1-Laplacian can be written as a $M \times M$ matrix,

$$L_1 = B_1^\top B_1 + B_2 B_2^\top. \tag{2.59}$$

For a general inner product, it is given by,

$$L_1 = B_1^\top (B_1^\top)^* + (B_2^\top)^* B_2^\top, \tag{2.60}$$

where A^* is the adjoint of the matrix A.

2.3.2 Hodge decompositions on graphs

Each 1-Laplacian implies a decomposition of $H(E)$ into orthogonal components [84],

$$H(E) = \text{im}(\text{grad}) \oplus \text{im}(\text{curl}^*) \oplus \ker(\Delta_1), \tag{2.61}$$

[30] In network science, it is typically referred to as just the *incidence matrix* [7].

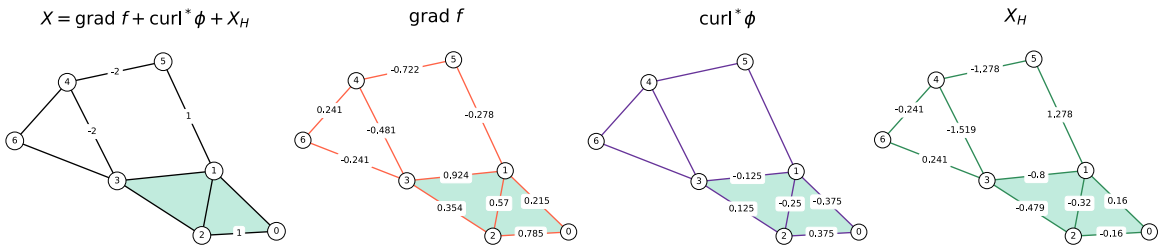

Figure 2.8: **The graph Helmholtz-Hodge decomposition.** We illustrate the GHHD of Eq. (2.62) with an example. We define an edge-flow around a simplicial complex with both edges and triangular faces. Note that this is not a clique complex. We decompose this flow into a gradient, curl, and harmonic component. The curl component is only defined on the triangles.

which we will refer to as the *graph Helmholtz-Hodge decomposition (GHHD)*. This implies that any edge-flow can be written in the form,

$$X(i,j) = (\text{grad}\, f)(i,j) + (\text{curl}^*\, \phi)(i,j) + X_H(i,j), \tag{2.62}$$

where $f \in H(V)$, $\phi \in H(T)$, and $X_H \in \ker(\Delta_1)$. These are called the 'gradient', 'curl', and 'harmonic' parts, respectively. The terms on the right-hand side (RHS) of Eq. (2.62) are unique for a given $X \in H(E)$. Fig. 2.8 illustrates this decomposition with an example. We take an edge-flow on a simplicial complex with both edges and faces. This can be decomposed into its gradient, curl, and harmonic parts, where we notice that the curl component is only defined on the edges forming the triangular faces.[31]

It is worth noting that,

$$\ker(\text{div}) = \text{im}(\text{curl}^*) \oplus \ker(\Delta_1), \tag{2.63}$$

$$\ker(\text{curl}) = \ker(\Delta_1) \oplus \text{im}(\text{grad}), \tag{2.64}$$

thus we can refer to $R = \text{curl}^*\, \phi + X_H$ as a single 'divergence-free' component, in analogy with the Helmholtz decomposition of a vector field [62, 84].

Computing the GHHD The GHHD is easy and efficient to calculate in practice. The orthogonality of the components in the GHHD imply that the decomposition can be realised as a series of least-squares problems, where the flows are vectors and the operators are matrices. We solve

$$f = \arg\, \min_{g \in H(V)} ||\text{grad}\, g - X||, \tag{2.65}$$

$$\phi = \arg\, \min_{\varphi \in H(T)} ||\text{curl}^*\, \varphi - X||, \tag{2.66}$$

[31] Note that this example is not a clique complex. We take the standard orientation for edges and faces with (i,j) and (i,j,k) for $i < j < k$.

where we can then take $X_H = X - \text{grad}\, f - \text{curl}^*\, \phi$. As the matrices representing these operators are sparse, we can solve these problems efficiently using the least-squares QR method (LSQR) [88].

Applications In recent years, the GHHD has found a number of applications for studying both the structure and dynamics of discrete objects. For example, as we will elaborate in Ch. 3, it has been used to uncover hierarchical structure in networks and tournaments, and serves as the basis for many ranking algorithms [89–91]. Additionally, it has been used to analyse dynamic data that live on the edges of a simplicial complex – so-called 'topological signal processing' [92]. In Ch. 5 and 6, we will introduce two novel uses of the HHD for the detection of irreversible and vortex-like dynamics respectively, with applications to biological and neural time-series data.

Chapter 3

Structural origins of entropy production in network dynamics

> (T)he possible role played by network topology in a system's thermodynamics, particularly that of the brain is still poorly understood [...] But does network structure have a role in entropy production and dissipation? If so, what aspects of brain network structure can affect a neural system's thermodynamics?
>
> *'Does the brain behave like a (complex) network? I. Dynamics'*
> *Physics of Life Reviews, 48, 47-98, (2024)*
> *D. Papo & J.M. Buldú*

At the heart of the study of complex networks, is the interplay between structure and emergent dynamics. Quintessential structural features, like modularity or small-worldedness, can determine the character and properties of network dynamics [8]. In this chapter, we focus on the structural origins of irreversible, entropy-producing dynamics on networks. In particular, we will highlight hierarchical network asymmetry as the principal feature determining the EPR of an NESS on a network, which we will illustrate with solvable examples and numerical simulations from prototypical network dynamical processes. Finally, we will introduce an approach for inferring both the EPR and a directed network from MVTS data, which we will apply to both neuroimaging and financial time-series. Our neuroscientific results highlight that the increase in the EPR seen in task compared to rest [3] is driven by a reorganisation of brain regions into a more hierarchical structure. This allows us to interpret results pertaining to entropy production and irreversibility in the brain from a perspective that focuses on interactions between regions.

3.1 Directed networks

As highlighted by Papo & Buldú [28], the role of network structure in the emergence of entropy production and irreversibility in complex systems, and therefore the brain, is unclear. However, from the study of NESM in active matter and collective dynamics, we know that nonequilibrium dynamics emerge in systems with *non-reciprocal* interactions [93, 94]. Such systems are characterised by a natural asymmetry between elements, which often leads to phase-transitions and an NESS. The analogue of non-reciprocal interactions in the network setting is directedness.

Recall that a network is directed if its weight or adjacency matrix satisfies that $A \neq A^\top$ i.e. it is asymmetric. This implies that the edge-weight from node j to node i may differ from the reciprocal edge from i to j, which may not be present. Despite the ubiquity of this setting, results in network dynamics overwhelmingly favour the case of undirected networks, as this simplifies mathematical analysis and data is more readily available. For example, as mentioned in Sec. 1.2.2, the human SC network, due to the limitations of DTI, is undirected, despite the fact that neuronal connectivity is inherently directed. This is because determining the direction of white-matter connectivity relies on invasive techniques which cannot be applied in humans. Nevertheless, analyses of cat and monkey connectomes show that directedness is not only prevalent, but has significant consequences for structural network features [95]. Directed networks can produce vastly different dynamics to their undirected counterparts as it pertains to their linear dynamics [96], pattern formation [97], synchronisation [98], and beyond [99].

An interesting feature of directed networks is that they can lead to the emergence of a *hierarchy*, where some nodes emerge as *leaders*, with many outgoing connections, and others emerge as *followers*, with mostly incoming connections. For example, a food-web can be represented by a network where nodes are animals and directed edges are drawn from predator to prey. In such a network, there is an implicit hierarchy, where each node can be assigned a 'trophic level' consistent with its position in the food-chain. A range of methods for unravelling this structure have been proposed, many of which are inspired by the GHHD from Sec. 2.3.2 [89, 91, 100, 101].

As we will show for the handful of solvable networked systems, directedness implies irreversible dynamics with a non-zero EPR. This presents the following question: does a *more* directed structure imply *more* entropy production? We will focus on this question for the remainder of this chapter.

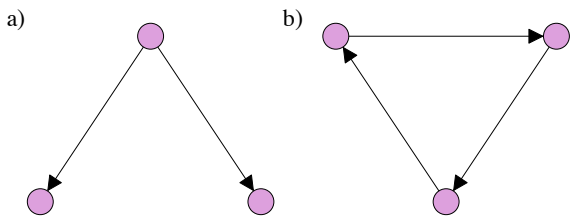

Figure 3.1: **Global and local directed networks.** Networks can display local and global directedness. *a*) This is a DAG, which has perfect local directedness i.e. each edge runs in only one direction, and global directedness, as it contains no directed cycles. *b*) This is a cycle graph, which has perfect local directedness, but a flat hierarchy as nodes cannot be ranked.

3.1.1 Measures of directedness on networks

We must first define what it means to be *more* directed. A naive approach would be to consider $A - A^\top$, which would be zero if and only if the network was undirected. However, this fails to address a number of concerns. For example, how do we derive a scalar measurement from this matrix? Should the measure be normalised allowing us to compare networks of different sizes or with different weights? Moreover, it also neglects the following important caveat about network directedness.

As mentioned previously, directed networks can give rise to a hierarchy amongst nodes, however this is not always the case. Consider for example a *cycle* graph with directed edges [Panel *b*), Fig. 3.1]. No edge in this network has a reciprocal, thus it is perfectly directed according to $A - A^\top$, yet it has no *global* direction or hierarchy. In many cases, it is this global direction that is of primary concern, which must then be differentiated from *local* directedness, which is defined by directly comparing reciprocal edges in isolation.

Remark 3.1.1. This prompts the definition of a 'perfectly directed' network, whose edges run in only one direction between nodes, but do not form cycles that violate the nodal hierarchy. These are known as *directed acyclic graphs (DAGs)*. Panel *a*) of Fig. 3.1 shows an example with three nodes. DAGs can be formally defined as directed graphs where every strongly connected component is a single node [7].

Undirected networks are described by symmetric matrices, thus all eigenvalues are real. On the other hand, directed networks are represented by asymmetric matrices and have complex eigenvalues. As a result, we can also investigate *spectral* asymmetry, which is concerned with the imaginary components of the eigenvalues and, by extension, the behaviour of the network as a linear system [96].

For the remainder of this chapter, we will focus on positively weighted networks without self-loops i.e. $A_{ij} \geq 0$, $A_{ii} = 0$.

Irreciprocity Our first measure of network directedness, is *irreciprocity* [102].[1] Given a network with weight matrix A, we define the reciprocated portion of an edge to be,

$$\overleftrightarrow{A}_{ij} = \min(A_{ij}, A_{ji}) = \overleftrightarrow{A}_{ji}. \tag{3.1}$$

The *reciprocity* is then quantified by,

$$r(A) = \frac{\sum_{i,j} \overleftrightarrow{A}_{ij}}{\sum_{i,j} A_{ij}}, \tag{3.2}$$

where $r \in [0, 1]$ with $r = 0$ corresponding to a perfectly unreciprocated network, where edges can only run in a single direction, and $r = 1$ corresponding to an undirected network. We then define the *irreciprocity* to be $1 - r$.

This is a measure of local directedness, which directly compares a single edge with its reciprocal, and produces a normalised measurement. Directedness and positive irreciprocity are equivalent.

Trophic directedness Next, we turn to the notion of global directedness and define a measure of the degree to which a network follows it.

Given a network with N nodes and weight matrix A, we aim to find a vector of *trophic levels*, $h \in \mathbb{R}^N$, where h_i represents the level of node i. We follow MacKay et al. [89] and consider the *trophic confusion*,

$$F(h) = \frac{\sum_{i,j} A_{ij}(h_i - h_j - 1)^2}{\sum_{i,j} A_{ij}}. \tag{3.3}$$

This acts as a loss function that quantifies how well a given trophic level assignment fits the structure of the network and vanishes if the level difference along each edge is 1. We then minimise the confusion to find the *best* level assignment that represents the network.[2]

[1] Our definition of irreciprocity is derived from Garlaschelli et al.'s measure of *reciprocity* [102], and its extension to weighted networks by Squartini et al. [103]. It differs from Newman's reciprocity, which begins at the same place, by whose expression is derived specifically for unweighted, directed graphs [7].

[2] This approach is closely related to the GHHD from Sec. 2.3.2. The level assignment is the potential on the nodes whose gradient most closely aligns with the asymmetry of the network weights. For further details on the relationship between the two, see Ref. [89]. *SpringRank* and *HodgeRank* are other popular ranking algorithm that uses the GHHD [91, 101].

In order to minimise this function, we note that the second derivative of F is positive semi-definite, thus all critical points are minima. By taking the partial derivative with respect to h_j, we find that critical points satisfy,

$$h_j \left[\sum_i A_{ij} + A_{ji} \right] - \sum_i (A_{ij} + A_{ji}) h_i = \sum_i A_{ji} - A_{ij}. \tag{3.4}$$

In order to simplify this expression, we define a symmetric 'pseudo-Laplacian' matrix,

$$\Lambda = \text{diag}\left(\sum_i A_{ij} + A_{ji} \right) - A - A^\top, \tag{3.5}$$

which acts as the following operator on h,

$$(\Lambda h)_j = h_j \left[\sum_i A_{ij} + A_{ji} \right] - \sum_i (A_{ij} + A_{ji}) h_i. \tag{3.6}$$

Finally, by defining $v_j = \sum_i A_{ji} - A_{ij}$, we note that solutions to the minimisation problem are precisely solutions to the equation,

$$\Lambda h = v, \tag{3.7}$$

which gives us a simple way to calculate h. Eq. (3.7) always has a solution, but it is not unique in general, as there is one arbitrary constant per connected component.

In practice, we replace one equation per connected component with $h_i = 0$, and then linearly scale the solution such that the lowest value of h, in each component, is 0. Given the optimal trophic levels, h^*, we evaluate $F_0 = F(h^*)$, which we call the trophic *incoherence* of the network. Finally, we measure the degree to which a network is directed with the *trophic directedness*, $\sqrt{1 - F_0}$.[3]

We define a network as *trophically flat* if $F_0 = 1$ ($\sqrt{1 - F_0} = 0$) and *perfectly trophically directed* if $F_0 = 0$ ($\sqrt{1 - F_0} = 1$).

Non-normality We turn now to spectral asymmetry and, specifically, *non-normality*. A matrix A is said to be *normal* if it satisfies,

$$AA^\top = A^\top A. \tag{3.8}$$

It is, therefore, non-normal if $AA^\top \neq A^\top A$ [96]. Symmetric matrices are automatically normal whilst non-normal matrices are necessarily asymmetric. Correspondingly, undirected networks are normal whilst non-normal networks are directed. The

[3]The inclusion of the square-root is useful as it causes the increase in trophic directedness to be linear under the interpolation described in Eq. (3.12), as stated in Lemma 3.1.1.

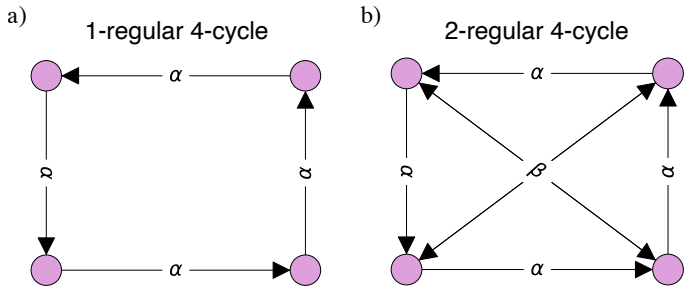

Figure 3.2: **Directed networks from circulant matrices.** Circulant matrices are normal and define cyclically symmetric networks which have no hierarchy.

has been much investigation into the role and uses of non-normal interactions in linear systems [96], neuronal [104], and communication networks [105]. Non-normality can also capture hierarchical asymmetry [89, 96], as we illustrate with the following example.

Example 3.1.1 (Circulant matrices and k-regular cycle graphs)**.** We can consider networks given by *circulant matrices*, which have the form,

$$A_{ij} = c_{(i-j) \bmod N}, \tag{3.9}$$

for $0 \leq i, j \leq N-1$, and where $c = (c_0, \ldots, c_{N-1})$ is the *generating vector* of the matrix $A \in \mathbb{R}^{N \times N}$. Circulant matrices are normal (for a proof see Ref. [106]). The class of networks that correspond to circulant matrices are called *cyclically symmetric*, and include $k-$regular cycle graphs, which extend cycle graphs to those with connections to their k-nearest neighbours on each side [Fig. 3.2]. Even if these networks are directed, they are hierarchically *flat*. The fact that normality groups circulant networks with undirected networks is one example illustrating how non-normality is a useful measure of hierarchical asymmetry.

We quantify the degree to which a network is non-normal using the *Henrici index*,

$$d_H(A) = \sqrt{||A||_F^2 - \sum_{i=1}^{N} |\lambda_i|^2}, \tag{3.10}$$

where $||A||_F = \sqrt{\sum_i \sum_j A_{ij}^2}$ is the Frobenius norm, $\{\lambda_i\}$ is the set of eigenvalues of A, and $|\lambda_i|$ is the modulus of the complex number [96]. The Henrici index is zero when a matrix is normal and positive otherwise. We also introduce the *normalized Henrici index*,

$$\widehat{d}_H(A) = \frac{d_H(A)}{||A||_F}, \tag{3.11}$$

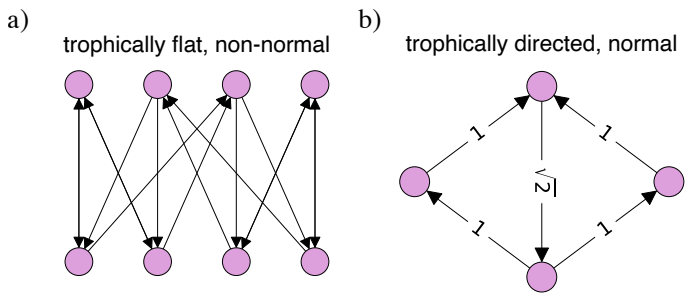

Figure 3.3: **Non-normality is not equivalent to trophic directedness.** *a*) This is an example of a network which is trophically flat ($F_0 = 1$), but non-normal ($d_H = 2.8284$). *b*) This is an example of a network which is trophically directed ($F_0 = 0.9737$), yet it is normal.

which has values between 0 and 1.

Remark 3.1.2. Non-normality and trophic directedness are not equivalent, nor does one imply the other, although they have been found to be strongly correlated in empirical networks [89]. Panel *a*) of Fig. 3.3 shows an example of a network which is trophically flat, yet non-normal, whilst Panel *b*) shows a network which is trophically directed, yet normal. MacKay et al. show that normal unweighted networks must be trophically flat [89]. Additionally, they show that perfect trophic directedness implies that all eigenvalues are zero and thus the network is 'perfectly non-normal' (like a DAG). Finally, we note that perfect non-normality does not imply perfect trophic directedness, where Panel *a*) of Fig. 3.1 serves as a counterexample ($F_0 = 0.1111$).

3.1.1.1 Parametrising the directedness of a network

In order to investigate the effect of directedness on the EPR, we require the ability to vary the directedness of a network whilst preserving as much of the remaining structure as possible. To do this, we begin with a directed network and consider the following parameterisation,

$$\widehat{A}(\epsilon) = (1-\epsilon)\widetilde{A} + \epsilon A, \tag{3.12}$$

where $\widetilde{A} = \frac{1}{2}\left(A + A^\top\right)$ and $\epsilon \in [0, 1]$. In other words, by increasing ϵ, the asymmetry parameter, we are smoothly interpolating between a symmetrised version of the network, and the original network itself. Fig. 3.4 shows the four measures of directedness as a function of ϵ for a single directed network, normalised by the value at $\epsilon = 1$. We find that ϵ increases the directedness almost linearly for each of the four measures,

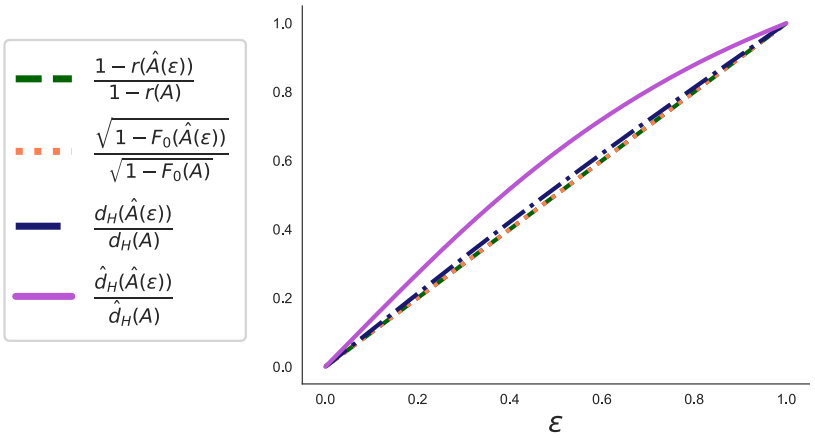

Figure 3.4: **Parametrising the directedness of a network.** We plot the four measures of directedness as a function of ϵ for a single directed network, normalised by the value at $\epsilon = 1$. We find that the measures increase approximately linearly in ϵ, a result that is exact for irreciprocity and trophic directedness.

with some nonlinearity for the non-normality.[4] Additionally, we have the following result.

Lemma 3.1.1. *Irreciprocity and trophic directedness increase linearly in ϵ under the interpolation in Eq. (3.12), i.e. we have that*

$$1 - r(\widehat{A}(\epsilon)) = \epsilon(1 - r(A)), \qquad \sqrt{1 - F_0(\widehat{A}(\epsilon))} = \epsilon\sqrt{1 - F_0(A)}, \tag{3.13}$$

for $\epsilon \in [0, 1]$.

For a proof see App. A.1.

3.2 Prototypical dynamical processes on networks

We now consider three prototypical dynamical processes on networks whose dynamics converge to a stationary distribution, and whose EPR can be calculated or well-approximated. Namely, these are the RW, introduced in Ex. 2.2.4, the (network) *Ornstein-Uhlenbeck process (OU)*, and the *kinetic Ising model*.

Entropy production in the random walk The RW is introduced, in both discrete- and continuous-time, in Ex. 2.2.4, and we recall that the stationary distribution of these formulations coincide. The EPR for the RW can then be computed

[4]The eigenvalues of the matrix $\widehat{A}(\epsilon)$ are not calculable from $\widetilde{A}$ and A so the relationship between ϵ and the spectral measures is non-trivial, but could be computed asymptotically close to $\epsilon = 1$.

using Eq. (2.32) or Eq. (2.37) which are also equivalent. However, computing a finite EPR for the RW requires that every edge in the network is reciprocated, which guarantees that the RW is microscopically reversible (see Remark 2.2.5). This requirement implies that unweighted networks must be undirected, where the EPR is then zero. Additionally, for real-world, weighted, directed networks, this requirement is seldom met.

To avoid this problem, we introduce an alternative quantity, which is a measure of irreversibility for the RW, even though it is not mathematically equivalent to the EPR. We begin by introducing the *Jensen-Shannon divergence (JSD)* between distributions (or densities) p and q with common support. The JSD is defined to be,

$$H_{\mathrm{JS}}(p, q) = H(p|m) + H(q|m), \tag{3.14}$$

where $H(\cdot|\cdot)$ is the KLD defined in Eq. (2.7), and $m = \frac{1}{2}(p + q)$ is a mixture of the two distributions [107]. Unlike the KLD, the JSD is symmetric, and does not diverge if $p(x) > 0$ and $q(x) = 0$. We can use this to define the 'pseudo-EPR' for the RW as,

$$\Phi = \frac{1}{2} \sum_{i,j} \left(P_{ij} \log \frac{P_{ij}}{\widetilde{P}_{ij}} + P_{ji} \log \frac{P_{ji}}{\widetilde{P}_{ij}} \right), \tag{3.15}$$

where $\widetilde{P} = \frac{1}{2}(P + P^\top)$. The quantity Φ is the JSD between the joint transition probabilities.

Remark 3.2.1. Unlike in the subsequent two processes, directedness is not sufficient to guarantee that a RW will converge to an NESS. If we consider a directed cycle graph on N nodes, such as in Panel *b*) of Fig. 3.1, then the uniform distribution $\pi_i = \frac{1}{N}$ is stationary.[5] For $N = 2$, it also obeys detailed balance.[6] These processes are not ergodic, and thus special cases, but we can also construct less trivial, ergodic examples, such as the network in Fig. 3.5. This shows that neither directedness nor non-normality is a sufficient condition for an NESS in the RW.

3.2.1 The Ornstein-Uhlenbeck process

The OU process is one of the most fundamental models in statistical mechanics. In a single dimension, it models a Brownian particle under friction. In the multivariate case, it generalises the linear system to the case of additive noise. We will return to

[5] Stationarity: $(T\pi)_i = \sum_j T_{ij}\pi_j = T_{i,i-1}\pi_{i-1} = \frac{1}{N} = \pi_i$.
[6] Detailed balance: $P_{12} = P_{21} = \frac{1}{2}$.

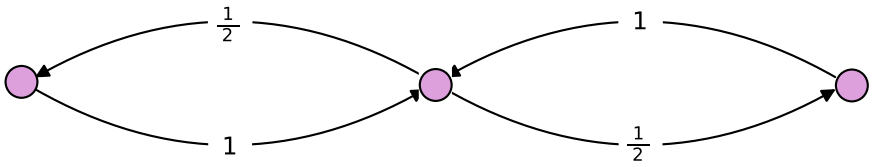

Figure 3.5: **Random walks on directed networks can satisfy detailed balance.** This is a non-normal network where the RW converges to $\pi = (1/4, 1/2, 1/4)$, which satisfies detailed balance. This shows that neither directedness nor normality is sufficient to guarantee an NESS for the RW.

this model a number of times in the thesis, thus we shall now consider a full treatment of its NESS. The OU process is given by

$$dx(t) = -Bx(t)\,dt + \Sigma\,dW(t), \tag{3.16}$$

where $x(t) \in \mathbb{R}^N$, $B \in \mathbb{R}^{N\times N}$ is the *friction matrix*, and $D = \frac{1}{2}\Sigma\Sigma^\top \in \mathbb{R}^{N\times N}$ is the diffusion [56]. Our derivation will closely follow Godrèche and Luck [108]. The process has an exact solution,

$$x(t) = e^{-Bt}x(0) + \int_0^t e^{-B(t-s)}\Sigma\,dw(s), \tag{3.17}$$

thus, when the eigenvalues of B have positive real parts, the process will converge to a mean value of zero with Gaussian fluctuations.[7] The covariance, $S(t)$, and its steady-state limit are given by,

$$S(t) = e^{-Bt}S(0)e^{-B^\top t} + \int_0^t e^{-B(t-s)}\Sigma\Sigma^\top e^{-B^\top(t-s)}\,ds, \tag{3.18}$$

$$S = \int_0^\infty e^{-B(t-s)}\Sigma\Sigma^\top e^{-B^\top(t-s)}\,ds, \tag{3.19}$$

where S, in turn, satisfies the *Lyapunov equation*,

$$BS + SB^\top = \Sigma\Sigma^\top = 2D. \tag{3.20}$$

The stationary density is a mean-zero multivariate Gaussian given by,

$$\pi(x) = (2\boldsymbol{\pi})^{-N/2}(\det S)^{1/2}\exp\left(-\frac{1}{2}x^\top S^{-1}x\right), \tag{3.21}$$

and the stationary probability flux is given by,

$$J_{\text{ss}}(x) = \mu x\pi(x), \tag{3.22}$$

where $\mu = DS^{-1} - B$. Thus we arrive at the following result.

[7] Under suitable conditions [56], the mean of the solution to an SDE is given by $\bar{x}(t) = \bar{x}(0) + \int_0^t \mathbb{E}[f(x(s), s)]\,ds$. The covariance is defined as $S(t) = \mathbb{E}[(x(t) - \bar{x}(t))(x(t) - \bar{x}(t))^\top]$.

Lemma 3.2.1. *Let $x(t)$ be an ergodic OU process with friction matrix B and diffusion matrix D. The Gaussian steady-state satisfies detailed balance if and only if $BD = DB^\top$.*

Proof: $J_{\text{ss}} = 0 \Leftrightarrow \mu = 0 \Leftrightarrow D = BS$. We then combine this with Eq. (3.20) to obtain $BS = SB^\top$. We right multiply $D = BS$ by $B^\top$, $DB^\top = BSB^\top$ and then use that $SB^\top = D$ to obtain $BD = DB^\top$.□

When this is not satisfied, we can calculate the EPR with Eq. (2.18). We have that,

$$\Phi = \langle x^\top \mu^\top D^{-1} \mu x \rangle, \tag{3.23}$$

where $\langle \cdot \rangle$ is expectation. Next, we define $Q = BS - D$, which implies that $\mu = -QS^{-1}$ and that $Q = -Q^\top$ from Eq. (3.20). We can then use that $\langle x^\top A x \rangle = \text{Tr}(SA)$ for Gaussian x with covariance S to obtain,

$$\begin{aligned}
\Phi &= \text{Tr}(S\mu^\top D^{-1}\mu) && (3.24)\\
&= -\text{Tr}(S^{-1}QD^{-1}Q) && (3.25)\\
&= \text{Tr}(B^\top D^{-1}Q) - \underbrace{\text{Tr}(S^{-1}Q)}_{=0} && (3.26)\\
&= \text{Tr}(B^\top D^{-1}Q), && (3.27)
\end{aligned}$$

where $\text{Tr}(S^{-1}Q) = 0$ as the trace of a symmetric matrix times an antisymmetric matrix is zero [55]. Finally, we compute the HHD where we have that $f_{\text{irr}}(x) = \mu x$ i.e.

$$f_{\text{irr}}(x) = -QS^{-1}x, \qquad f_{\text{rev}}(x) = -DS^{-1}x, \tag{3.28}$$

where $B = QS^{-1} + DS^{-1}$ is a decomposition into an asymmetric and symmetric matrix capturing the irreversible and reversible components respectively. Fig. 2.5 shows the HHD of an OU process in two dimensions.

Exact simulation of the Ornstein-Uhlenbeck process Given that we have the exact solution of the OU process in Eq. (3.17), we can derive its transition kernel,

$$\begin{aligned}
x(t) &\sim \mathcal{N}(e^{-Bt}x(0), C(t)), && (3.29)\\
C(t) &= \int_0^t e^{-B(t-s)}\Sigma\Sigma^\top e^{-B^\top(t-s)}\, ds. && (3.30)
\end{aligned}$$

When C can be computed analytically or numerically, we can sample paths directly from this kernel, $x_{t+\Delta t} \sim \mathcal{N}(e^{-B\Delta t}x_t, C(\Delta t))$, thus we do not need to use a numerical approximation.

Given (B, D), we can compute the EPR of the OU process with Eq. (3.27) after solving the Lyapunov equation in Eq. (3.20). An exact solution can be derived with Kronecker algebra [109], however it relies on the inversion of a matrix, thus it is more practical to use standard numerical methods [110]. However, in the special cases of $N = 2$ and circulant matrices, the EPR has a simplified form.

Example 3.2.1 (The OU process for $N = 2$)**.** For $N = 2$, solving the Lyapunov equation is equivalent to solving a three-variable linear system. For the following process,

$$B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad D = \begin{pmatrix} u & w \\ w & v \end{pmatrix}, \tag{3.31}$$

the closed form for the EPR can be obtained,

$$\Phi = \frac{(w(d-a) - bv + cu)^2}{(a+d)\,(uv - w^2)}. \tag{3.32}$$

Example 3.2.2 (The OU process for circulant matrices)**.** If B and D are circulant (where D is also symmetric), the Lyapunov equation (3.20) can also be solved by the discrete Fourier transform. We find that the EPR has closed form,

$$\Phi = \sum_{k=1}^{N} \frac{\Im(\lambda_k)^2}{\Re(\lambda_k)}, \tag{3.33}$$

where $\{\lambda_k\}_{k=1}^N$ are the eigenvalues of B, which can be calculated with the discrete Fourier transform [106], $\Im(\cdot)$ is the imaginary part, and $\Re(\cdot)$ is the real part. Thus the process is reversible if and only if $B = B^\top$ [108].

The network Ornstein-Uhlenbeck model The OU can be considered to be coupled through a positively-weighted network, A,

$$dX(t) = \Theta(\gamma A - I)X(t)\,dt + \sqrt{2\sigma}I\,dW(t), \tag{3.34}$$

where I is the identity matrix, Θ is the reversion rate, γ is the coupling parameter, and $D = \sigma I$ [111]. If the process is stationary, then the condition for detailed balance is $A = A^\top$ i.e. the process violates detailed balance on directed networks and satisfies it on undirected networks.

3.2.2 The kinetic Ising model

Originally developed as a model of magnetism, the Ising model is a model of interacting spins that take binary values $\{\pm 1\}$, where spins correspond to upward and downward dipoles. It has also been used extensively as a model of neuronal network dynamics where $x = \pm 1$ correspond to on and off states for neurons [33].

In its most common form, the Ising model is given by a Hamiltonian energy function,[8]

$$H(x) = -\sum_{i,j} A_{ij} x_i x_j, \tag{3.35}$$

where A_{ij} is the (undirected) coupling strength between nodes j and i [52]. This has an associated Boltzmann distribution,

$$P(x) = \frac{1}{Z} \exp\left(\beta \sum_{i,j} A_{ij} x_i x_j\right), \tag{3.36}$$

where β is the inverse thermodynamic temperature. This formulation as an energy function with a Boltzmann distribution is called the *equilibrium* or *Hamiltonian* Ising model and corresponds to an ESS. Moreover, this is the ME distribution for a binary system with specified mean values and pairwise correlations [33]. This relies on the network being undirected. In this form, the Ising model does not have any 'dynamics', and numerical simulations of this model consist of *sampling* configurations from the Boltzmann distribution with MCMs, but where the sequence of visited states is irrelevant [52].

On the other hand, when the connections are asymmetric, the model exhibits rich time-dependent behaviour. We then define a *dynamic* model, namely a DTMC known as the *kinetic* Ising model. There are a number of choices for the dynamics of the Ising model, including both sequential and parallel updating of spins [112]. We focus on parallel updates, where the dynamics are given by,

$$P(x(t+1)|x(t)) = \prod_{i=1}^{N} \frac{\exp\left(\beta x_i(t) \sum_{j=1}^{N} A_{ij} x_j(t)\right)}{2\cosh\left(\beta \sum_{j=1}^{N} A_{ij} x_j(t)\right)}, \tag{3.37}$$

[8] We have set the external fields to equal zero, which we assume for the rest of this thesis (see Ref. [52]).

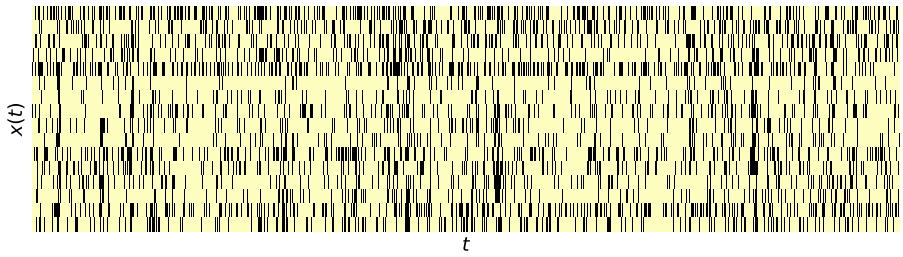

Figure 3.6: **Sample path from the kinetic Ising model**. Here is a path sampled from a kinetic Ising model with a random 16 node network, using Glauber dynamics. Black squares are $+1$ and yellow squares are -1.

where $\beta = 1$ from this point on [113].[9] The stationary distribution does not have an analytical form, and estimating the stationary distribution via MCMs is only possible for small networks, as there are 2^N possible configurations.

Glauber dynamics and small networks To generate the sampled paths from the model, we use synchronous 'Glauber' updates [52].[10] At each time-step we update each spin according to,

$$P(x_i(t+1) = 1|x(t)) = \left[1 + \exp\left(-2\sum_{j=1}^{N} A_{ij}x_j(t)\right)\right]^{-1}. \tag{3.38}$$

Fig. 3.6 shows an example path sampled from a network with 16 nodes.

For large networks, the number of possible states, 2^N, is too large to estimate the stationary distribution from samples. However, we can do so for networks of size $N \leq 10$. After an equilibration period, we estimate the stationary distribution and evaluate the EPR with Eq. (2.37).

Mean field approximation for large networks For large networks where the number of possible states is too large for MCMs, we use the *naive mean-field (NMF)* approximation. First, we define the *activation rates*, $m_i(t)$, and the delayed correla-

[9] ME models in ESM can be generalised to *maximum-caliber* models in NESM [114]. The kinetic Ising model is the maximum-caliber model for a binary system with specified mean activations and pairwise correlations [113].

[10] Note that these Glauber dynamics are not identical to those originally proposed by Glauber in Ref. [115], which update spins sequentially rather than in parallel (which do not satisfy Eq. (3.37)). However, it is a commonly used term for this synchronous update rule [3].

tions, $D_{ij}(t)$,

$$m_i(t) = \sum_{x(t)} x_i(t) P(x(t)), \tag{3.39}$$

$$D_{ij}(t) = \sum_{x(t),x(t-1)} \left(x_i(t) x_j(t-1) P(x(t), x(t-1)) - m_i(t) m_j(t-1) \right), \tag{3.40}$$

which are sufficient statistics to define a particular Ising model. The expression for the EPR of the Ising model reduces to,

$$\Phi = \sum_{i,j} (A_{ij} - A_{ji}) D_{ij}, \tag{3.41}$$

where D_{ij} becomes constant as the process converges to an NESS [113]. The NMF approximation is given by,[11]

$$m_i(t) \approx \tanh \left(\sum_j A_{ij} m_j(t-1) \right), \tag{3.42}$$

$$D_{ij}(t) \approx A_{ij}(1 - m_i^2(t))(1 - m_j^2(t-1)). \tag{3.43}$$

Iteratively computing $m(t)$ and $D(t)$ until they converge to stationary quantities, we then approximate the EPR using Eq. (3.41).

3.3 Directedness drives entropy production on networks

In this section, we use the measures of directedness and the dynamical processes to investigate the relationship between directedness and the EPR. We begin with a simple 2-node network where we can derive analytical results.

Example 3.3.1 (A 2-node network)**.** Consider the 2-node network [Panel *a*), Fig 3.7],

$$A = \begin{pmatrix} 0 & A_{12} \\ A_{21} & 0 \end{pmatrix}. \tag{3.44}$$

The irreciprocity of this network is given by,[12]

$$1 - r = \frac{|A_{12} - A_{21}|}{A_{12} + A_{21}}. \tag{3.45}$$

[11] See Ref. [113] for a derivation and further details.
[12] We use that $\min(a, b) = \frac{1}{2}(a + b - |a - b|)$.

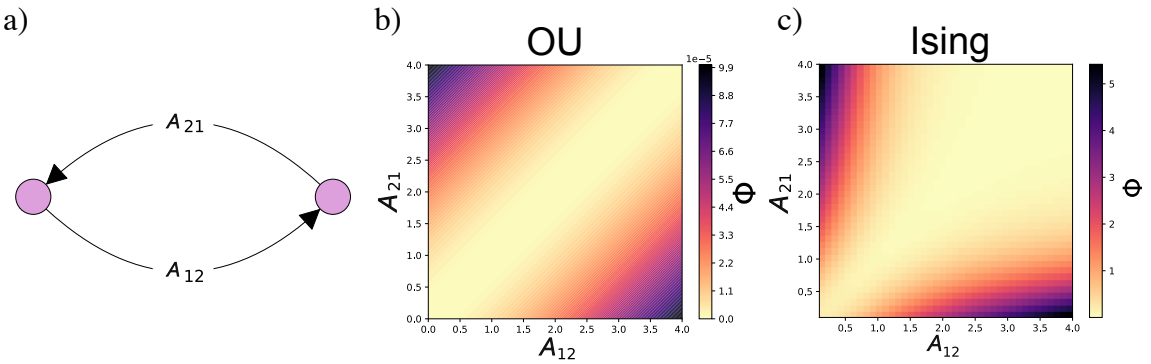

Figure 3.7: **Entropy production in a 2-node network.** *a*) A 2-node directed network. *b*) The EPR of the OU process on a 2-node network increases quadratically with asymmetry (Eq. 3.49) *c*) The EPR of the Ising model on a 2-node network, estimating using MCMs. It increases nonlinearly with asymmetry (Eq. 3.41).

The trophic levels, h, are given by the general solution,

$$h_1 - h_2 = \frac{A_{12} - A_{21}}{A_{12} + A_{21}}, \tag{3.46}$$

which gives trophic directedness,[13]

$$\sqrt{1 - F_0} = \frac{|A_{12} - A_{21}|}{A_{12} + A_{21}}, \tag{3.47}$$

which coincides with the irreciprocity. The Henrici and normalised Henrici indices are given by,

$$d_H = |A_{12} - A_{21}|, \qquad \widehat{d}_H = \frac{|A_{12} - A_{21}|}{\sqrt{A_{12}^2 + A_{21}^2}}. \tag{3.48}$$

Each of the measures captures the asymmetry between the pair of weights, but with unique nuances. The irreciprocity and trophic directedness are normalised by the L_1 norm of the matrix whilst the normalised Henrici index uses the L_2 norm, and the Henrici index is unnormalised.

Now we turn to the dynamics. The RW preserves detailed balance on this network (see Remark 3.2.1). The OU process has friction matrix, $B = \Theta(I - \gamma A)$, which has eigenvalues $\lambda_\pm = \Theta \pm \Theta\gamma\sqrt{A_{12}A_{21}}$, thus stationarity is equivalent to the condition $\gamma < \frac{1}{\sqrt{A_{21}A_{12}}}$. Using the expression for the EPR of the OU with $N = 2$, Eq. (3.32) in Ex. 3.2.1, we have that,

$$\Phi = \frac{\Theta\gamma^2}{2}\left(A_{12} - A_{21}\right)^2, \tag{3.49}$$

[13] Here we use an equivalent formula for trophic directedness proved in Ref. [89], $1 - F_0 = \frac{\sum_{i,j} A_{ij}(h_i - h_j)}{\sum_{i,j} A_{ij}}$.

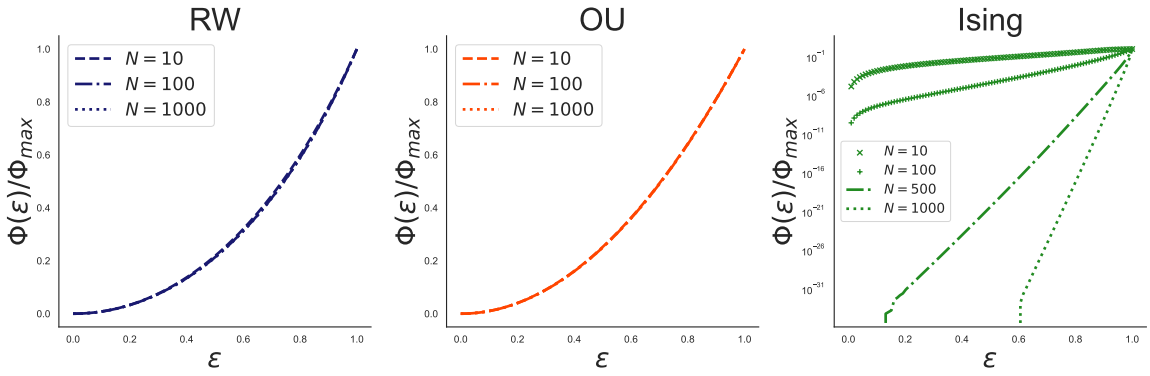

Figure 3.8: **Entropy production in interpolated networks.** We plot the EPR of each of the three systems as a function of the asymmetry ϵ, $\Phi(\epsilon)$, normalised by the maximal value at $\epsilon = 1$. We find that the EPR increases with directedness for each of the processes, and across network size.

i.e. the EPR scales quadratically with the asymmetry [Panel b), Fig. 3.7].

Finally, we consider the Ising model. As there are only four possible spin configurations in this system, we can compute the EPR by MCMs with Glauber dynamics. Panel c) of Fig. 3.7 shows the EPR of the Ising model computed across the plane (A_{12}, A_{21}), which appears to scale nonlinearly with the asymmetry. For both the OU and Ising models, the positive diagonal represents symmetric networks which preserve detailed balance, thus the distance from this line is the level of asymmetry.

3.3.1 Dynamics on interpolated directed networks

We return to the interpolation presented in Sec. 3.1.1.1, which will allow us to vary the directedness of a large network and investigate the effect on the EPR. To begin with, we sample hierarchical, directed, weighted networks with $N = 10$, 100, 500, and 1000, using the algorithm proposed in Ref. [96] (see App. A.2). We then use the interpolation from Eq. (3.12), and measure the EPR as a function of $\epsilon \in [0, 1]$.

Fig. 3.8 shows the EPR, $\Phi(\epsilon)$, along the interpolation, for a single network for each of the sizes $N = 10$, 100, and 1000.[14] This is normalised by the maximal value at $\epsilon = 1$. We see that for the RW and OU process, the EPR scales quadratically with ϵ across network size. For the Ising model, the EPR increases with ϵ, but the scaling relationship changes with the size of the network. This numerical experiment supports our hypothesis that more directed networks produce more entropy.

[14]Whilst Fig. 3.4 shows a single network at each network size, we find this relationship holds for all networks in a large sample.

Whilst we can only perform numerical analysis for large, complex networks, we can derive analytical results for the interpolation in special cases of the OU process.

Lemma 3.3.1. *Consider the interpolation of a 2-node asymmetric network,*

$$\widehat{A}(\epsilon) = \begin{pmatrix} 0 & (1-\epsilon)\bar{A} + \epsilon A_{12} \\ (1-\epsilon)\bar{A} + \epsilon A_{21} & 0 \end{pmatrix}, \tag{3.50}$$

for $\epsilon \in [0,1]$*, where* $\bar{A} = \frac{1}{2}(A_{12} + A_{21})$*. Then a stationary network OU process has EPR,*

$$\Phi(\epsilon) = \frac{\Theta\gamma^2\epsilon^2(A_{12} - A_{21})^2}{2} = \epsilon^2\Phi(1), \tag{3.51}$$

i.e. the EPR scales quadratically in ϵ*.*

This result follows from the formula in Eq. (3.32). We have a similar result for the OU on normal networks.

Lemma 3.3.2. *Let* A *be a positive-weighted, normal network with the interpolation* $\widehat{A}(\epsilon)$*. The EPR along the interpolation is given by* $\Phi(\epsilon) = \epsilon^2\Phi(1)$*.*

For a proof see App. A.3.

Remark 3.3.1. Lemma 3.3.2 challenges the idea that *hierarchical* asymmetry drives the EPR, as the OU on normal networks follows the quadratic scaling relationship despite the absence of a hierarchy. Nevertheless, as we will show in the next section, the EPR is still correlated with hierarchical asymmetry in many cases.

3.3.2 Erdős-Rényi graphs as a null model

The interpolation approach allowed us to vary the directedness of all the measures simultaneously [Fig. 3.4], however this does not allow us to differentiate between them and determine which 'type' of directedness drives the EPR of each process. To this end, we introduce the *Erdős-Rényi (ER)* random graph model. The directed ER model is an ensemble of unweighted, directed graphs, $\mathcal{G}(N,p)$, each with N nodes, where a directed edge is drawn from node j to node i with probability p, independent of all other edges [7]. This process does not assume or enforce a particular structure, however hierarchical directedness can emerge spontaneously. As a result, we use it as a 'null model' where the measures of directedness can behave differently as a function of p.

For a range of values of $p \in [0,1]$, we randomly sample ER graphs with $N = 100$ and measure their directedness. We repeat this for a sample of 100 graphs at each

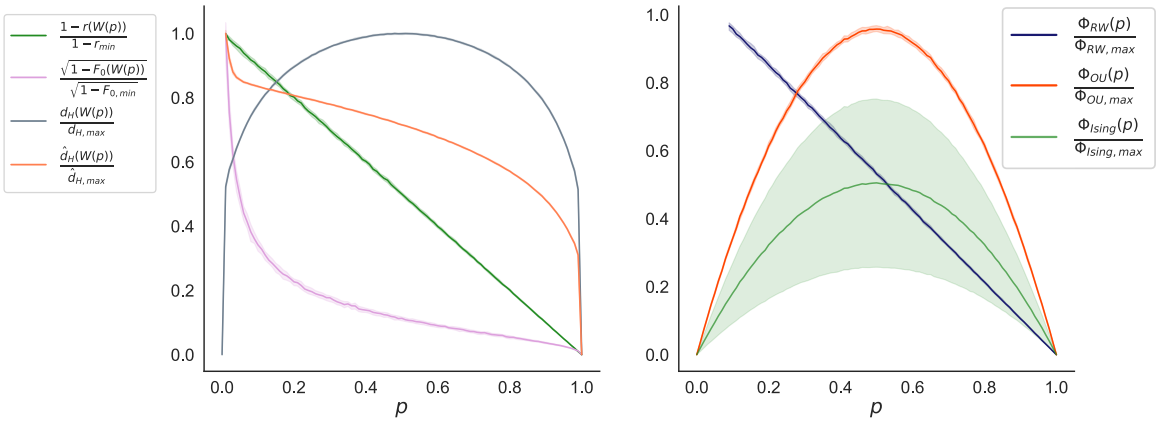

Figure 3.9: **Directedness and entropy production in Erdős-Rényi graphs.** Left. We plot the measures of directedness as a function of p in the ER model, using a ensemble of 100 graphs at each value of p. Each measure shows unique behaviour. Right. We plot the EPR of each process as a function of p, which allows us to match up directedness measures and processes.

value of p and normalise by the maximum of each measure [Left, Fig. 3.9]. We notice that each measure traces a distinct curve as a function of p, with noticeably small variance. For example, irreciprocity decays linearly with p, whilst trophic directedness and the normalised Henrici index decay also. The unnormalised Henrici index behaves parabolically, peaking at $p = 1/2$.

Next, we repeat the same procedure for each of the dynamical processes acting on the sampled graphs [Right, Fig. 3.9].[15] We find that the RW follows linear decay with p, thus is directly correlated with the irreciprocity of the ensemble. On the other hand, the EPR of the OU and Ising processes follows the parabolic behaviour of the unnormalised Henrici index, peaking at $p = 1/2$, suggesting that the EPR is most directly correlated with the non-normality of the underlying network. We offer a tentative explanation for this phenomenon.

The RW is a 'local' process where the dynamics are only affected by the neighbours of the current state-node. As a result the irreversibility of the process is most directly affected by the asymmetry of the individual adjacent edges, as captured by the irreciprocity. Moreover, the RW dynamics are invariant to a global scaling of the network weights, thus the relationship can be captured by a normalised measure. On the other hand, the entire network features in the dynamics of the OU and Ising processes at each instant, and the dynamics are not invariant under scaling. As a

[15] Ergodicity of the RW relies on the weak connectedness of the graph, thus we begin our sampling at $p = 0.1$, and discard any graphs that are disconnected until we have a total sample of 100 graphs.

result, we find that the unnormalised Henrici index, a measure of global asymmetry, is most directly correlated with the EPR.

Remark 3.3.2. The Ising model has very large variance in the EPR, unlike the RW and OU processes. This is most likely because of a greater sensitivity to small structural network changes, or due to 'critical' behaviour, where entropy production is maximised [116].

3.4 Linear models for multivariate time-series

Using a combination of analytical results in simple cases, and numerical analysis for large networks, we have shown a variety of evidence for a link between hierarchical network directedness and the EPR of a stochastic process. Next, we turn to the empirical dynamics of complex networks, in the form of an MVTS of nodal states. Often one attempts to infer information directly from this nodal activity, in the absence of a directed, structural network or model.

In order to investigate the link between the EPR and the directedness of interactions in empirical dynamics, we must simultaneously solve two challenging problems. The first of these is the inference of a directed, structural network from an MVTS. Whilst a number of approaches exist, they are often limited, requiring the ability to perturb the system [117], or are 'functional connections' derived from directed measures of correlation [118]. The second challenge is the inference of the EPR from observed data. Again, such methods have been developed, but are typically designed from low-dimensional data stemming from biophysical experiments and cannot be extended to large networks [119, 120].

To tackle this challenge, we propose a simple but effective approach that relies on the assumption of linear stochastic dynamics. We model the MVTS as observations of an OU process with either general friction or constrained to a network. Discretising these models with the EM approximation, we obtain,

$$x(t_{i+1}) = [I - \Delta t B]\, x(t_i) + \Sigma \xi_i, \tag{3.52}$$

$$x(t_{i+1}) = [I + \Delta t (A - I)]\, x(t_i) + \sqrt{2\sigma}\xi_i, \tag{3.53}$$

where $\Delta t = t_{i+1} - t_i$ and ξ_i are i.i.d multivariate normal vectors from $\mathcal{N}(0, \Delta t I)$. In the network model, we absorb the coupling parameter, γ, into the network A, and the reversion rate, Θ, into the time-scale of the process. The discretised models now coincide with the first-order *multivariate autoregressive process (MAR)*, which is

extremely common in time-series analysis in a range of applied domains [121]. The MAR process is typically written as,

$$x_{t+1} = \Gamma x_t + \chi_t, \tag{3.54}$$

where $\Gamma \in \mathbb{R}^{N\times N}$ is the matrix of regression coefficients and $\{\chi_t\}$ is the time-series of *residuals* i.e. the remainder term after regression of the linear component. To find the matrix Γ, we solve the *least-squares* optimisation problem,

$$\min_{\Gamma} ||X_{1:T} - \Gamma X_{0:T-1}||_F^2, \tag{3.55}$$

where $X_{0:T-1}$ and $X_{1:T-1}$ are matrices whose $T-1$ columns are observations of the process are times t_i. This problem has a closed-form solution [122],

$$\Gamma = X_{1:T} X_{0:T-1}^\top (X_{0:T-1} X_{0:T-1}^\top)^{-1}, \tag{3.56}$$

from which we obtain $B = \frac{1}{\Delta t}(I - \Gamma)$, where $\Delta t \ll 1$ is a time-scale parameter that we set to 0.1.

When fitting the network model, we notice that the problem is *convex*, thus we can add non-negativity constraints on the regression coefficients.[16] In this manner, we can modify the problem to fit the network OU process,

$$\min_{A:A_{ij}\geq 0} ||X_{1:T} - ((1-\Delta t)I + \Delta t A) X_{0:T-1}||^2, \tag{3.57}$$

which can be solved efficiently with convex optimisation [123]. Finally, we calculate the residuals, $\chi_{t+1} = x_{t+1} - \Gamma x_t$, from which we can recover the diffusion matrix with $\mathrm{Cov}(\chi) = 2\Delta t D$, taking the empirical covariance of the residuals over the time-points. If the diffusion is assumed to be diagonal, we estimate the scalar noise intensity with $\sigma = \frac{1}{2\Delta t}\langle \mathrm{diag}\,(\mathrm{Cov}(\chi))\rangle$.

Given the parameters of the OU process, we can examine the eigenvalues and determine if the model is stationary. If so, we can analyse the structure of the inferred connectivity, and compute the EPR with Eq. (3.27).[17]

[16] A problem is *convex* if the function to be optimised, f, satisfies $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$ for $\lambda \in [0,1]$ and $x, y \in \mathcal{C}$ where $\mathcal{C}$ has the property that for every $x, y \in \mathcal{C}$ then $\lambda x + (1-\lambda)y \in \mathcal{C}$ [123].

[17] Similar approaches using the OU process as neural model were developed and published independently by Gilson et al. [36], and by Benozzo et al. [124], although each fitting approach has important differences.

3.4.1 Applications to whole-brain network dynamics

As discussed in Sec. 1.2.2, neuroimaging data takes the form of an MVTS of regional activities. As a result, we can apply our approach to neural data to determine the relationship between the observed irreversibility and the structure of the inferred effective network.[18] We consider fMRI data from the Human Connectome Project (HCP) taken from 100 (unrelated) participants in a resting state, during a motor task, and during a social task [125, 126].[19] The data is parcellated in the Desikan-Killiany brain atlas with 80 regions (DK80), 62 of which are cortical and 18 of which are sub-cortical [127]. The data is preprocessed as described in Ref. [128].

Remark 3.4.1. It is worth noting that models of neural dynamics are highly nonlinear, across scales, in contradiction with our linear assumption [13, 14]. However, fitting nonlinear models to data is extremely challenging and often relies on heuristic and 'ad-hoc' approaches. In recent work, Nozari et al. found that linear models outperformed nonlinear models in the modelling of fMRI data using precisely this dataset, thus it appears to be a valid assumption given our current modelling limitations [129].

Entropy production and directedness in fMRI data First, we fit the unconstrained model and find that the fitted OU processes are stable for each participant in each task. We then compute the EPR, Φ, and find that it is significantly higher in the social and motor tasks, when compared to the resting state [Panel a), Fig. 3.10], supporting the original results of Lynn et al. [3].[20] Next, we fit the network-constrained model and measure the directedness of the inferred effective network. We find that the directedness is higher in both tasks compared to resting state, according to all four measures [Panel a), Fig. 3.10]. Moreover, we find a significant correlation between the EPR and each of the directedness measures [Panel b), Fig. 3.10].

Remark 3.4.2. We choose to use the unconstrained model to infer the EPR and the constrained model to infer the directed network, rather than simply using the constrained model for both. This is because we found that the constrained model was only stable for a partial collection of participants. We considered optimisation

[18] Effective brain networks are defined in Sec. 1.2.2.

[19] The HCP is an open database of neuroimaging data that has been strongly validated, and is often used for method development [125].

[20] We perform significance testing with a paired t-test. For the remainder of the thesis, significance is denoted in the plots by $*$, with (ns, $p \geq 0.05$), (*, $p < 0.05$), (**, $p < 0.01$), (***, $p < 0.001$), and (****, $p < 0.0001$).

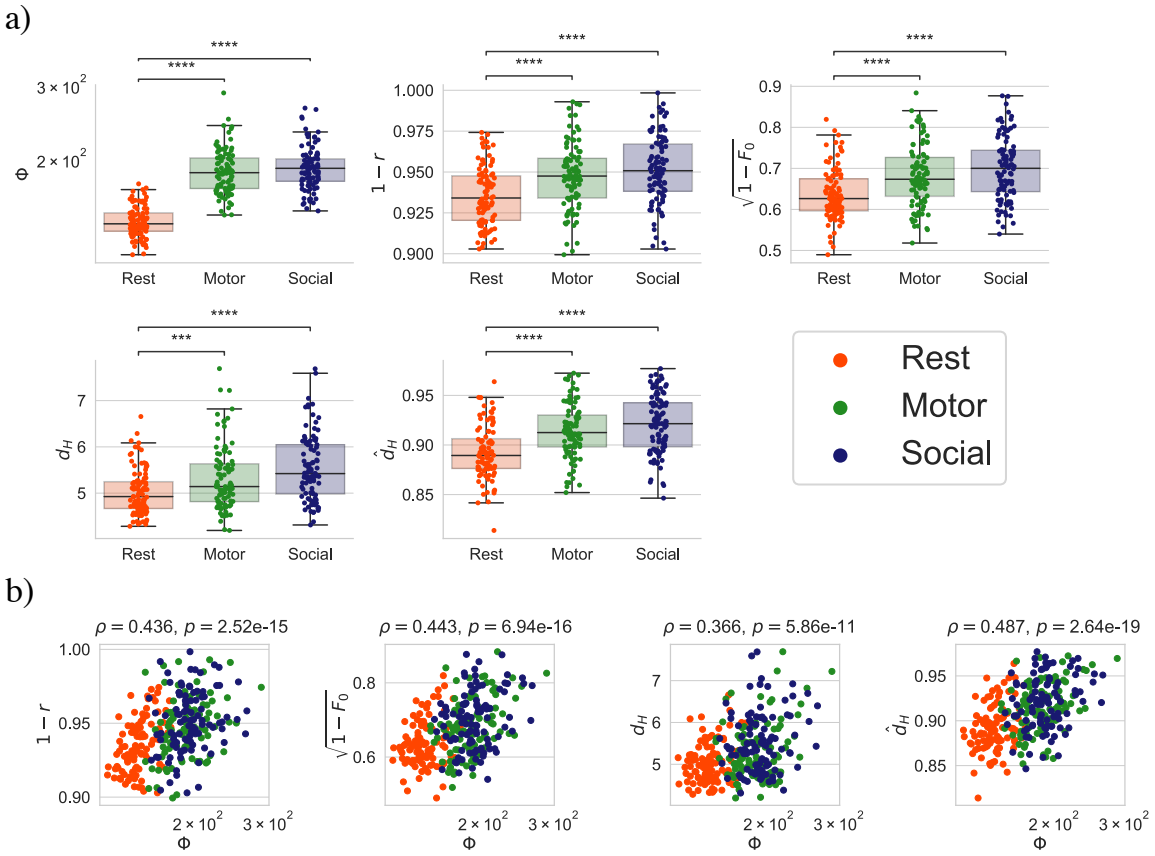

Figure 3.10: **Entropy production and directedness in fMRI data.** *a*) We fit the linear model in Eq. (3.34) with $N = 80$ to fMRI data from 100 participants using least-squares. We measure both the EPR and the four measures of asymmetry using the inferred matrix W. We find that the EPR of the neural dynamics during tasks is higher than at rest. The directedness of the inferred effective network is also higher according to all four measures of directedness. *b*) The EPR and the directedness are also significantly postively correlated for each of the measures.

approaches where the feasible set was constrained to only include stable matrices, but this set is non-convex, thus optimisation is particularly challenging. Moreover, constraining to a convex subset is a significant restriction on the connectivity. Instead, we compromise by using the unconstrained model, which is stable, indicating that the data is stationary, to infer the EPR, and then use the constrained model to infer an effective network.

Trophic directedness of effective brain networks In order to calculate the trophic directedness of a network, we must calculate the trophic level of each node in the network, as in Eq. (3.7), which we can use to investigate the structure of the inferred effective networks. In order to get a better intuition on the role of each

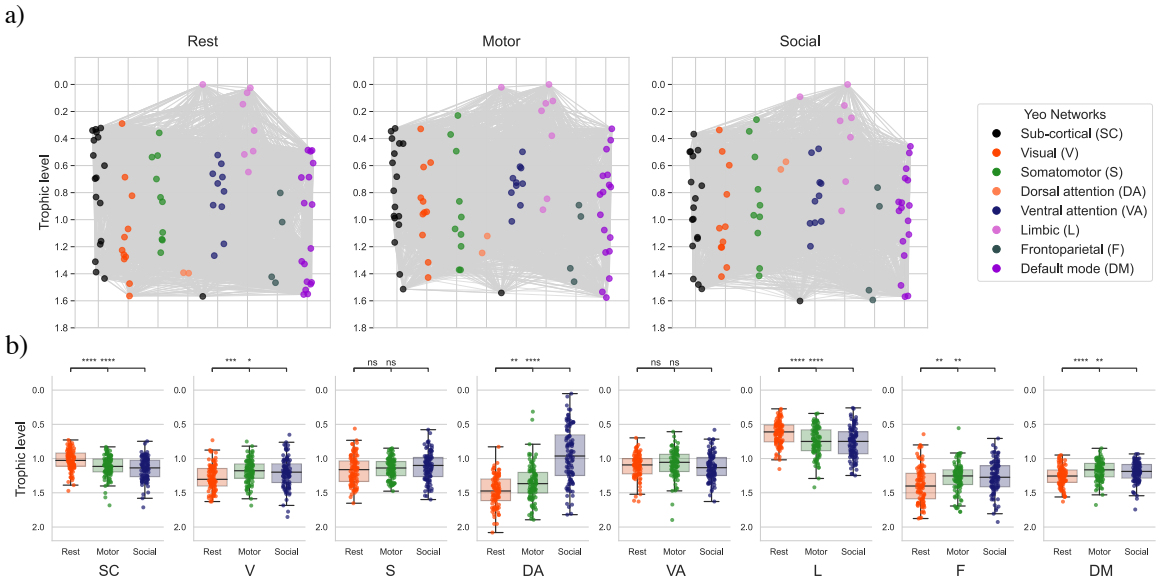

Figure 3.11: **Trophic directedness of effective brain networks.** *a*) Using the inferred network, we can use Eq. (3.7) to solve for the trophic level of each node in the inferred effective networks, averaged over participants. We can then plot the level for each node and for each task. We distinguish nodes using a colour for each of the Yeo functional subnetworks. *b*) Here we look at the average trophic level in each Yeo subnetwork, distributed over participants, for each task. This allows us to see the position of a subnetwork within the overall hierarchy, and how this moves across tasks. We find that the limbic network sits atop the hierarchy, the dorsal attention network climbs to the top during tasks, and the sub-cortical regions descend the hierarchy during tasks.

node in the brain network, we use the canonical 'Yeo functional subnetworks', a partition of the cortex into seven disjoint subnetworks and then the remaining sub-cortical regions, each associated with a particular cognitive function [130].[21] We can visualise the hierarchy in the inferred network, averaged over participants, by looking at the trophic level of each node, coloured by its sub-network, in each task [Panel *a*), Fig. 3.11]. Moreover, we can look at the average trophic level in each subnetwork [Panel *b*), Fig. 3.11]. We find that the limbic network sits atop the trophic hierarchy, a result that is in-keeping with the outsized role of the limbic subnetwork in cortical processing [131]. Moreover, we see that the role of subnetworks can change depending on the task stimulus. For example, we see that the dorsal attention network climbs to the top of the hierarchy during social tasks, whilst the sub-cortical regions descend the hierarchy.

[21]Named for B.T. Thomas Yeo, these functional subnetworks are another well-known tool in the analysis of functional neuroimaging data [130].

Our results show a close relationship between the hierarchical directedness of effective brain network interactions and the resulting irreversibility of their dynamics. In particular, we have shown that the observed increase in the EPR in task-based dynamics compared to resting state can be explained by a reorganisation of effective interactions into a more hierarchical structure. This casts a somewhat esoteric result about the stochastic thermodynamics of large-scale brain networks in a more familiar and interpretable form that is concerned with the structure of interactions. As a result, this presents a solution to the challenge proposed by Papo and Buldú, with which we began this chapter. The directedness, asymmetry, and hierarchy of brain network interactions plays a crucial role in the emergence of NESSs in large-scale neural dynamics.

3.4.2 Trophic hierarchy in the New York Stock Exchange

Beyond neural data, we can apply the same linear modelling approach to other complex systems which may be operating in an NESS, and whose dynamics are approximately linear. We consider stock-price data from 119 U.S. companies from the New York Stock Exchange (NYSE) over the period 2000-2021, as analysed and published in Ref. [132]. Each company represents a node whose state is the fluctuating stock-price, thus we consider this dataset as a single MVTS. We fit the constrained model to the time-series to infer a directed network of interactions between companies.[22] We measure the overall directedness of the inferred stock network to obtain $\sqrt{1-F_0} = 0.405$, $1-r = 0.975$, $d_H = 1.53$, and $\widehat{d}_H = 0.160$, each to three significant figures (3 s.f.). Thus the system is highly locally directed, with some overall hierarchy.

As an analogue of the functional subnetworks, this network can be partitioned into nine disjoint industries.[23] Fig. 3.12 shows the trophic levels of the inferred network [Panel *a*)] and its breakdown by industry [Panel *b*)]. We offer only a brief explanation of these results, as complexity economics is not the focus of this thesis. Our results offer some evidence of a consumer-driven hierarchy in the economy with Consumer Goods sitting near the top of the hierarchy, whilst more speculative industries, such as Financials, sitting lower in the overall hierarchy. This suggests that a hierarchical leader-follower structure is an emergent property of economic forces in the stock-market, in-keeping with previous analyses of economic systems with Granger's causality measure [133–135].

[22] We note that linear autoregressive models, such as ours, are extremely common in economics and financial analysis [121]. Typically these would allow negative correlations, however our model and measures of directedness are restricted to positive weights.

[23] We use the partition into industries from Ref. [132].

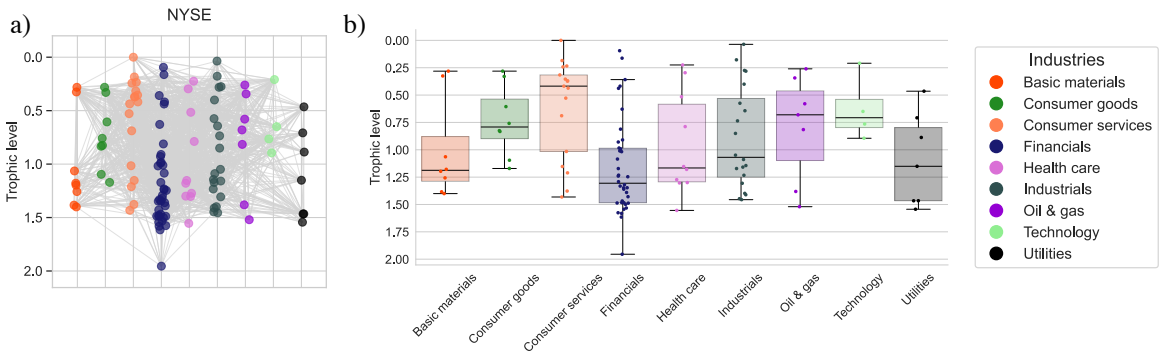

Figure 3.12: **Trophic hierarchy in the New York Stock Exchange.** *a*) We can rank companies in the NYSE by the trophic level of each node in the inferred network. *b*) We can then look at the distribution over trophic levels of each industry. We find that Consumer Goods sit near the top of the hierarchy, suggesting they are leaders, whilst Financials is at the bottom, suggesting this is a speculative follower industry.

3.5 Conclusions

In this chapter, we have focused on the role of (hierarchical) directed network structure in the emergence of NESSs and entropy production in network dynamics. Our results show that directedness is the principal feature driving network dynamics out of equilibrium and that it is highly correlated with the EPR in both synthetic and real-world dynamics. This is another result linking the structure and dynamics of complex networks, that highlights the profound effect of directedness – an area that is often overlooked – on emergent behaviours. Moreover, our linear modelling approach allows us to measure both the EPR and directedness of brain network activity using human neuroimaging data. Our findings support, and present an explanation for, a result which is central to this thesis – that neural dynamics operate in an NESS and that the EPR varies across experimental conditions, such as tasks. We have shown that this variation is driven by a reorganisation of brain regions into a more hierarchical structure. This result ties together disparate results in neural dynamics by bridging the EPR of brain activity [3], with the study of effective network structure [128], and could be further linked to theories of hierarchical processing during tasks, such as that posited by the Friston [136] or Dehaene et al. [137].

Limitations and future directions Whilst we provide new evidence on the link between directedness and the EPR, there remain important questions and edge-cases. For example, whilst we showed that hierarchical structure and non-normality is strongly correlated with the EPR in some network ensembles [Fig. 3.8, 3.9], we

also showed that normal networks, such as cycle graphs, are an important edge-case that do not follow this relationship [Lemma 3.3.2]. This disagreement could extend to complex networks with cyclic patterns i.e. those that possess a large number of loops, which have low trophic-directedness [138].[24] Moreover, our analysis is constrained to a number of simple systems where the EPR is estimable. We have no evidence that this relationship holds in other nonlinear systems, including models of neuronal dynamics, where individual nodes can produce entropy through oscillations. Similarly, our framework for analysing neural data relies on a linear model, which is only a first approximation. Developing a framework for a nonlinear whole-brain model, where the EPR and the effective network can be inferred, remains an open and extremely challenging problem. We note that our analysis focuses only on positive weighted networks, as this is where our measures of directedness are well-defined.[25] However, when considering neuronal dynamics, as we will in Ch. 7, negative interactions are crucial to represent inhibitory connections. Both fitting algorithms and analytical results concerning the EPR of large asymmetric Ising networks admit general signed connectivity, making it challenging to apply our framework in this important case [116,139]. Finally, we note that our approach, like the rest of this thesis, only considers the brain as a network and fails to consider the approach of neural fields (see Remark 1.2.1).

[24] We note that these loop-like structures have been shown to be particularly uncommon in real-world networks [138].

[25] It is non-trivial to generalise notions of hierarchical directedness to negative edge-weights. For example, if $A_{ij} = 5$ and $A_{ji} = -5$, which node is leading and which is following? Should we consider a net effect along the edge of 0 or of +10? Additionally, we note that such skew-symmetric matrices are normal i.e. $A = -A^\top \Rightarrow AA^\top = -A^2 = A^\top A$.

Chapter 4

Time-irreversibility and higher-order interactions in the brain

Networks are the canonical mathematical representation for complex systems, including the brain. However, as discussed in Sec. 2.1, they represent static, unimodal, and pairwise relationships between the elements of a system. In acknowledgment of this limitation, in recent years we have seen an explosion of interest into so-called 'higher-order' networks [49, 140]. By the most general definition, higher-order networks can include temporal, multilayer, and annotated networks [Fig. 4.1], but the term has more recently become associated with systems consisting of 'many-body' interactions e.g. the hypergraphs and simplicial complexes, introduced in Sec. 2.1. In theory, this allows for richer and more accurate representations of complex systems, yet inevitably complicates mathematical analysis. So far, this research direction has predominantly focused on the structural implications of higher-order interactions (HOIs) on emergent dynamics, such as synchronisation or contagion [141], with only recent interest in the problem of inferring these interactions from empirical data using time-series analysis [132, 142, 143] and information theory [144–147].

The data-driven perspective has particular implications for neuroscience, where the study of HOIs focuses on functional interactions i.e. the inference of synergistic relationships between the activity in sets of regions. Whilst there exists a plethora of different correlation measures for capturing pairwise interactions in time-series data [118], extending these to triadic or more general interactions is non-trivial. Existing approaches have focused on co-fluctuations [132] and topological methods [148]. Nevertheless, these approaches often identify relationships in an ad-hoc manner, as they do not directly generalise a functional correlation, and are usually limited to

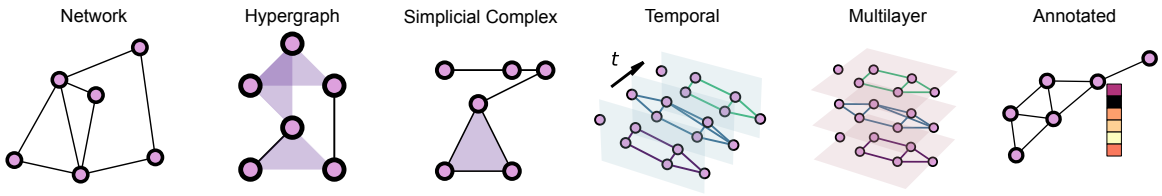

Figure 4.1: **Higher-order representations for complex systems.** Networks are limited to static, unimodal, and pairwise relationships. They can be extended to 'higher-order' systems which offer richer representations of complex systems, including many-body interactions, as a hypergraph or simplicial complex, as well as temporal, multilayer, and annotated networks.

triadic interactions for computational reasons.

In this chapter, we introduce a novel approach for identifying HOIs in neural recordings which, inspired by NESM, uses time-irreversibility as a proxy measure of an HOI. As we will illustrate in Sec. 4.1, the irreversibility of an interaction can be used a proxy measure of the degree to which the relationship between the activity of a set of nodes is significant and causal. As a result, we will use it to determine which sets of brain regions are engaged in a more cognitively significant interaction. Under this premise, we introduce a framework, which we refer to as the DiMViGI framework, that uses techniques from the emerging field of network analysis of time-series (NATS) [149], specifically visibility graphs [150], to measure the irreversibility associated with each individual HOI in an MVTS, and identify those which are significant. We illustrate our framework on sampled paths from irreversible stationary diffusions and then apply it to MEG recordings from human participants engaged in a long-term auditory recognition task [151]. Our framework is able to highlight which key groups of regions are engaging in a significant HOI. In particular, our results display the expected hemispheric symmetry. Moreover they indicate that there are more significant HOIs across sensory and processing regions than within sensory and processing blocks. Finally, we are able to identify the crucial role played by medial regions in facilitating HOIs across hemispheres.

4.1 Irreversibility and functional interactions

First, we consider a brief illustration that underscores irreversibility as a measure of interaction between variables.

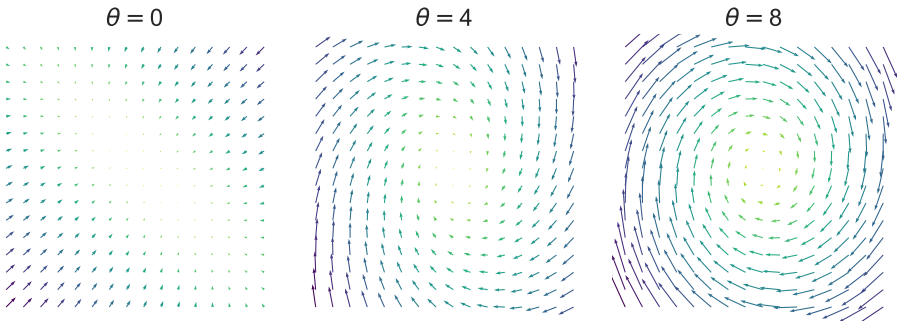

Figure 4.2: **Irreversibility as causal interaction.** We plot the drift field from a process with increasing irreversibility as a function of θ. This illustrates how greater irreversibility can be an indicator of a more significant functional interaction between variables.

Example 4.1.1. Fig. 4.2 shows the drift of a two-dimensional OU process of the form,

$$dx(t) = -\begin{pmatrix} 2 & 1-\theta \\ 1+\theta & 2 \end{pmatrix} x(t)\, dt + \sigma\, dw(t), \tag{4.1}$$

which is reversible if and only if $\theta = 0$, and has EPR $\Phi = \theta^2$ (see Eq. (3.32)). Whilst irreversibility is distinct from correlation, covariance, or information-theoretic measures of causality, we use it as evidence of a synergistic, causal, and hierarchical interaction between variables. In this example, we see that the asymmetric cross-terms drives the irreversibility of the process, which helps us to motivate the use of irreversibility in this context.

Remark 4.1.1 (Cross-correlations)**.** In an effort to detect causality between variables, instantaneous and time-delayed cross-correlations have been studied in depth in complex systems science. Canonical methods include Granger causality [133], transfer entropy [152], and convergent cross mapping [153,154]. They are also a fundamental feature of irreversible systems [155,156]. Frameworks such as 'INSIDEOUT' [37] have been developed to measure irreversibility of neural signals directly from time-delayed cross-correlations.

Remark 4.1.2. In Ex. 4.1.1, we see that this two variable system must be coupled in order to produce irreversibility. This motivates the question: can we observe irreversibility from two uncorrelated variables? This is only possible if each of the two variables are coupled to other, unseen degrees of freedom, as Corr. 2.2.2 indicates that one-dimensional processes over $\mathbb{R}$ must be reversible (in the absense of additional

boundary conditions).[1] Assuming that we have observe all variables in our system, cross-correlations between variables are necessary for irreversibility. As a result, a sensible requirement of any method measuring the irreversibility of a multivariate process, is that the irreversibility of a combination of two mutually independent processes decomposes into the irreversibility of each independent process, which we show for our method in Sec. 4.3.1.

Irreversibility We consider a path from a continuous-valued stochastic process sampled in discrete time, $\{x(t_i)\}$ for $t_i \in \{t_0, \ldots, t_{T-1}\}$ with $x(t) \in \mathbb{R}^N$. We define the *irreversibility* of the path to be,

$$I\left(\{x(t_i)\}_{t_i=t_0}^{t_{T-1}}\right) = H\left(p\left(\{x(t_i)\}_{t_i=t_0}^{t_{T-1}}\right), p\left(\{x(t_i)\}_{t_i=t_{T-1}}^{t_0}\right)\right), \tag{4.2}$$

i.e. the KLD between the probability density of observing a path and the probability of seeing its time-reversal [58]. Assuming that the path comes from a stationary diffusive process, then this path constitutes a 'partial observation' as we only observe the process at discrete time-points and have access to a potentially incomplete set of phase-space variables. As a result we have the following inequality,

$$\Phi \geq I\left(\{x(t_i)\}_{t_i=t_0}^{t_{T-1}}\right), \tag{4.3}$$

where Φ is the EPR of the underlying stationary process [58].

Remark 4.1.3. This result is part of a more fundamental law of stochastic thermodynamics. When observing a stochastic process, any coarse-graining, in space or in time, constitutes a loss in information about the arrow of time i.e. the direction of the process. This reduction in information can only obscure irreversible currents, and thus the evidence for the arrow of time in a partial observation can only be a lower-bound on its true value [45]. We note that this is an inherent property of a stochastic path and does not apply to any practical scheme for inferring the EPR, which can have an error in either direction.

[1] Partial observations of stochastic processes are often non-Markovian, even if the underlying process is Markovian. For SDEs, this means that projections of a Markovian SDE are non-Markovian [157]. For discrete-state processes, this usually takes the form of 'lumped' states [158], which we will discuss in Ch. 5.

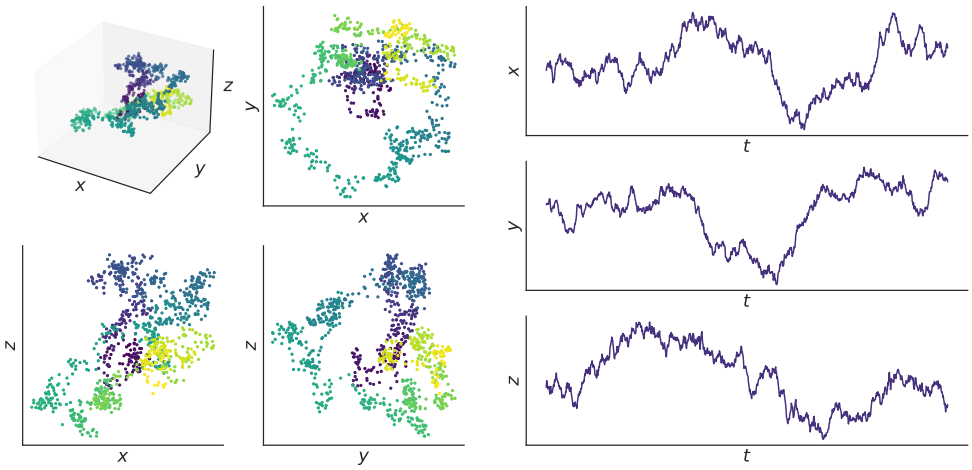

Figure 4.3: **Sub-interactions in multivariate systems.** We can consider multivariate systems to be composed of sub-interactions. We will attempt to identify a significant sub-interaction using the irreversibility in this reduced phase-space. Here we consider a path, $(x(t), y(t), z(t)) \in \mathbb{R}^3$, sampled from an OU process, which has seven sub-interactions, which each define a (reduced) phase-space.

Irreversibility of sub-interactions Within a multivariate system, $(x_1, \ldots, x_N)$, we consider the irreversibility of an individual *sub-interaction* defined by a subset of k variables, $(x_{n_1}, \ldots, x_{n_k})$, where $1 \leq k \leq N$, which we refer to as a k-tuple.[2] Our goal is to approximate the irreversibility of this particular tuple, denoted by $\varsigma^{(n_1,\ldots,n_k)}$.

Example 4.1.2. Consider a system with three variables, (x, y, z). This system has seven possible sub-interactions: the triplet (x, y, z); three pairs (x, y), (x, z), and (y, z); and three singletons (x), (y), (z). Fig. 4.3 shows an example path from such a system, as well as the projection of the path onto each of these seven sub-interactions.

The framework we present in this chapter aims to quantify the irreversibility of each of these interactions and identify, at each level defined by k, which interactions are most significant. As a result, we may move beyond a single aggregate measure and express irreversibility in more interpretable components.[3][4]

[2]The term k-tuple is a common generalisation of a pair, triplet, quadruplet etc. It is worth noting that in computer programming a tuple is an ordered set, whilst here it is unordered.

[3]This approach was inspired by Ref. [159] which presents a more theoretical decomposition of irreversibility but in the case of binary, multipartite dynamics, with applications to neural spike-trains.

[4]A more general approach would be to consider k functions of the N variables as 'observables'.

4.2 Directed multiplex visibility graphs

Quantifying the irreversibility of a time-series, particularly an MVTS, is a challenging problem as it requires computing the relative entropy between two (inaccessible) high-dimensional probability densities. When computing information-theoretic quantities in time-series, one typically has to 'bin' or 'symbolise' the data, often in an unprincipled manner [160]. On the other hand, NATS is a collection of methods for mapping time-series to graphs – a naturally discrete structure – from which such quantities can be estimated more easily [149]. In particular, we will focus on Lacasa et al.'s *visibility graph (VG)*, its generalisation to MVTS, and its applications to irreversibility [150, 161, 162]. We begin by introducing the *visibility algorithm* in the case of a univariate time-series.

The visibility algorithm Given a univariate time-series $\{x_i\}_{i\in I}$ with time indices $\{t_i\}_{i\in I}$, where $x_i \in \mathbb{R}$ and I is the index set, the VG has one node for each $i \in I$.

Nodes $i, j \in I$ are connected by an edge if the corresponding data-points (t_i, x_i) and (t_j, x_j) are *mutually visible* i.e. if they satisfy, for any intermediate data-point (t_k, x_k) with $t_i < t_k < t_j$,

$$x_k < x_j + (x_i - x_j)\frac{t_j - t_k}{t_j - t_i}. \tag{4.4}$$

In geometric terms, this condition is met if the line connecting (t_i, x_i) and (t_j, x_j) does not cross any intermediate data-points [Panel a), Fig. 4.4]. Trivially, each node is connected to its neighbours whilst large positive fluctuations become hubs with many connections. Panels a) and b) of Fig. 4.4 show a time-series and its associated VG.

In order to quantify irreversibility, we first direct the edges 'forward in time' i.e. $i \to j$ if $t_i < t_j$, which we call the directed visibility graph (DVG) [Panels $c - d$), Fig. 4.4]. We can then decompose the degree d of a node into the sum of the in-going and out-going degree,

$$d = d_{\text{in}} + d_{\text{out}}. \tag{4.5}$$

Applying the visibility algorithm to the time-reversal of a stochastic process, we would note that each node would swap its in-degree and out-degree. If the stochastic process was, in fact, reversible, then the in- and out-degree distributions of the associated DVG would coincide, as the time-series would be statistically indistinguishable from

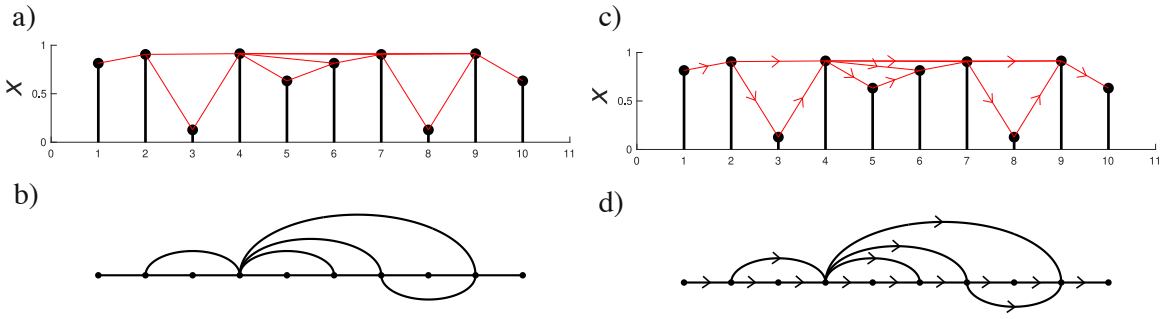

Figure 4.4: **The visibility algorithm for time-series**. *a*) We take a time-series and apply the condition in Eq. (4.4) to determine which points are mutually visible. *b*) These visible points become connected nodes in the VG. *c* – *d*) We can direct edges forward in time to define the DVG, from which irreversibility can be measured.

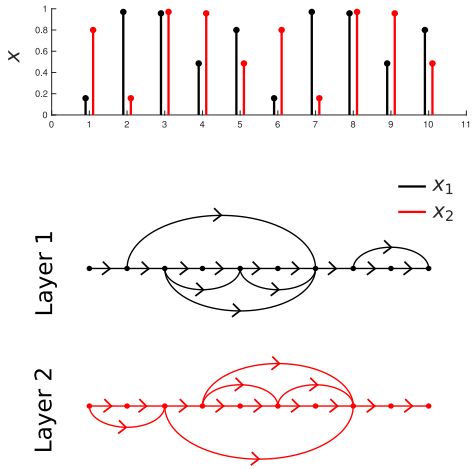

Figure 4.5: **The directed multiplex visibility graph.** We can extend the VG to MVTS by constructing a multi-layer graph, where each layer is defined by applying the visibility algorithm to a single variable, as shown here with a 2-variable system. The result is a multi-layer graph where each has the same nodes, as each variable has the same time points, but different connections.

its reversal. It follows that the divergence between the in- and out-going degree distributions is a simple measure of irreversibility [162].

This construction can be naturally extended to an MVTS by considering the *directed multiplex visibility graph (DMVG)* [161]. Given an MVTS with N variables, we now construct a multi-layer directed graph, also known as a *multiplex* network. Each of the N variables corresponds to a distinct layer in the DMVG, which each contain the same nodes but different edges. The edges are obtained by applying the visibility algorithm to each variable in turn. Fig. 4.5 illustrates this on a 2-variable MVTS.

4.3 The 'DiMViGI' framework

We now arrive at the directed multiplex visibility graph irreversibility (DiMViGI) framework. We combine the multilayer generalisation of the VG with the approach for quantifying irreversibility, to arrive at a method able to quantify the irreversibility of each sub-interaction in a multivariate system. More specifically, we consider an MVTS with N variables and T time points, given by $\{x(t_1), \ldots, x(t_T)\}$, where $x(t_i) = (x_1(t_i), \ldots, x_N(t_i)) \in \mathbb{R}^N$ and construct its associated DMVG.

For a given k-tuple of variables, $(n_1, \ldots, n_k)$, we calculate the multivariate *marginal* in-going and out-going degree distributions:

$$P_{\text{in}}^{(n_1,\ldots,n_k)}(d_1, \ldots, d_k), \qquad P_{\text{out}}^{(n_1,\ldots,n_k)}(d_1, \ldots, d_k), \tag{4.6}$$

where $P^{(n_1,\ldots,n_k)}(d_1, \ldots, d_k)$ is the probability of a node (any node) having degree d_i in layer n_i for all i simultaneously.[5]

Next, we compute the JSD (see Eq. (3.14)) between these particular marginal in- and out-degree distributions to obtain a measure, the *k-order irreversibility*,[6]

$$\widetilde{\varsigma}^{(n_1,\ldots,n_k)} = H_{\text{JS}}(P_{\text{in}}^{(n_1,\ldots,n_k)}, P_{\text{out}}^{(n_1,\ldots,n_k)}). \tag{4.7}$$

[5]In practice, these distributions are estimated from either a single, long time-series or a number of samples of graphs from the process. Empirical distributions can be smoothed via interpolation or with a Bayesian prior [163]. We chose to use a uniform prior – an early and somewhat regrettable choice – but which prevents singularities of the form $P_{\text{in}}(x) = 0 < P_{\text{out}}(x)$. As the method compares the divergence between distributions *relative* to other tuples, the choice of smoothing method is inconsequential, as long as it is consistent over all k-tuples. Stronger priors or more aggressive smoothing will simply reduce the size of all the divergence values, but their relative sizes will remain the same.

[6]We use the notation $\widetilde{\varsigma}^{(n_1,\ldots,n_k)}$ to highlight that we are computing an approximation of the irreversibility, $\varsigma^{(n_1,\ldots,n_k)}$.

We are quantifying the irreversibility of the path after projection into the subspace defined by a particular k-tuple of variables. Repeating this for all possible k-tuples in the system, we quantify the relative irreversibility of each interaction at a given level k. We repeat this process for all values of $k \in \{1, \ldots, N\}$.

We will illustrate this approach on two examples using the OU process.

Example 4.3.1 (A two-dimensional OU process)**.** Consider a 2-dimensional OU process with drift and diffusion,

$$B = \begin{pmatrix} 4 & 1 \\ 2 & 1 \end{pmatrix}, \qquad D = \begin{pmatrix} 1 + \frac{2}{2a+1} & 1 \\ 1 & 1 + \frac{2}{2a+1} \end{pmatrix}, \tag{4.8}$$

which converges to a NESS with EPR given by $\Phi = \frac{4a^2}{5(2a+2)} \propto a$. We sample paths for values $a \in [0, 10]$ to obtain an MVTS with $T = 2000$. Fig. 4.6 shows $\widetilde{\varsigma}$ for both the individual variables x_1 and x_2, alongside the irreversibility of the pair, (x_1, x_2). On one hand, there is no significant correlation between the EPR and the 1-order irreversibility, indicating that the irreversibility measured at this level is the *noise-floor*, the intrinsic error in the irreversibility that comes from using finite data and approximate sampling methods ($r = -0.256$ and $r = 0.263$, respectively, both n.s.). This is expected given that single-variable paths from the OU process are reversible.[7] On the other hand, the irreversibility of the pair is strongly correlated with the EPR ($r = 0.965$, ****).

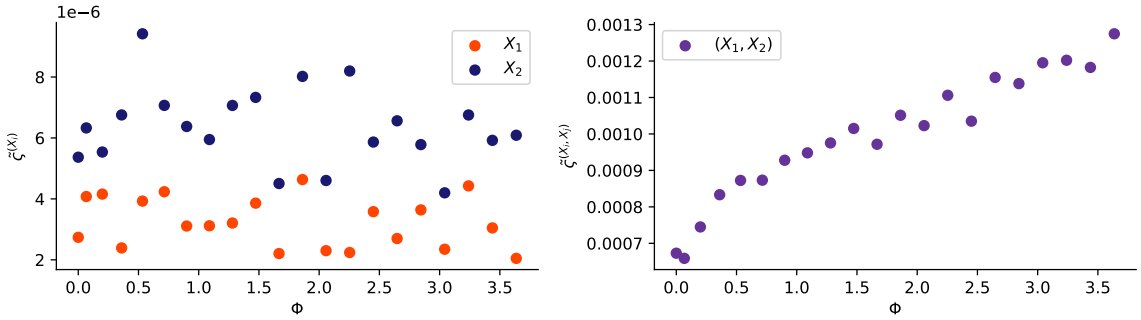

Figure 4.6: **Ex. 4.3.1 Irreversibility of the OU with $N = 2$.** Applying the DiMViGI framework to a 2D OU process, we see that the irreversibility of individual variables is not correlated with the EPR (Left), Φ, whereas the irreversibility of the pair is (Right).

[7] Even if an OU process is in a NESS, the dynamics of individual variables are reversible. This can be shown by noting that the time-course of individual variables form a Gaussian process, and unidimensional Gaussian processes are reversible [164].

Example 4.3.2 (A three-dimensional circulant OU process)**.** We consider a 3-dimensional circulant OU process given by,

$$B = \begin{pmatrix} 1 & a & -a \\ -a & 1 & a \\ a & -a & 1 \end{pmatrix}, \qquad D = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}, \tag{4.9}$$

where $\Phi = 6a^2$ (see Ex. 3.2.2). We sample paths for $a \in [0.5, 8.5]$ and apply DiMViGI. Fig. 4.7 shows that the irreversibility of the individual variables is uncorrelated (or weakly correlated) with the EPR, whereas the irreversibilities of the pairs and triplet are.[8]

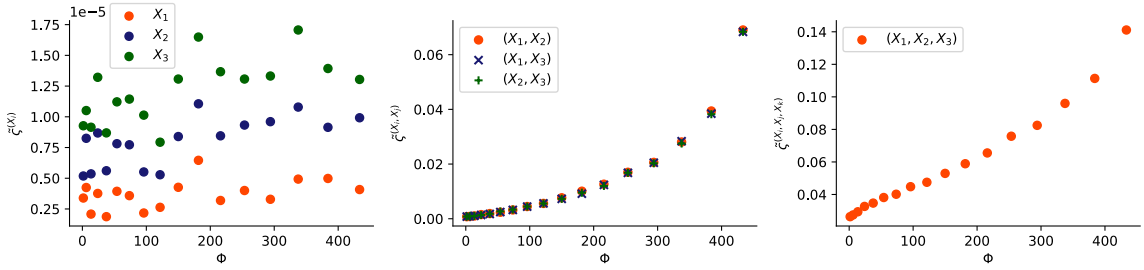

Figure 4.7: **Ex. 4.3.2 Irreversibility of the OU with** $N = 3$. Applying the DiMViGI framework to a 3D OU process, we see, again, that the irreversibility of individual variables is not correlated with the EPR, whilst the irreversibility of the pairs and triplet are.

4.3.1 The case of independent variables

To improve our understanding of how the DiMViGI framework operates on multivariate paths, we consider the special case of independent variables. Consider $(x_1, \ldots, x_k)$ such that x_1 is independent of the other variables. As x_1 is independent, the edges in the associated layer of the multiplex network are also independent. As a result, the joint in- and out-degree distributions of the multiplex factorise as follows,

$$P(d_1, \ldots, d_k) = P(d_1)P(d_2, \ldots, d_k). \tag{4.10}$$

[8]The 1-order interactions have correlations ($r = 0.440$, n.s.), ($r = 0.682$, **), ($r = 0.650$, **), respectively. The 2-order interactions have ($r = 0.904$, ****), ($r = 0.903$, ****), ($r = 0.900$, ****), respectively. The 3-order interaction has ($r = 0.981$, ****).

Under the DiMViGI framework, we quantify the irreversibility of the k-tuple as

$$\widetilde{\varsigma}^{(x_1,\ldots,x_k)} = H_{\mathrm{JS}}(P_{\mathrm{in}}(d_1,\ldots,d_k), P_{\mathrm{out}}(d_1,\ldots,d_k)), \tag{4.11}$$

$$= \sum_{d_1,\ldots,d_k} P_{\mathrm{in}}(d_1,\ldots,d_k) \log \frac{P_{\mathrm{in}}(d_1,\ldots,d_k)}{P^*(d_1,\ldots,d_k)} \tag{4.12}$$

$$+ \sum_{d_1,\ldots,d_k} P_{\mathrm{out}}(d_1,\ldots,d_k) \log \frac{P_{\mathrm{out}}(d_1,\ldots,d_k)}{P^*(d_1,\ldots,d_k)},$$

where $P^* := \frac{1}{2}(P_{\mathrm{in}} + P_{\mathrm{out}})$. Focusing first on the term related to the in-degree distribution, we use the independence of x_1 and the properties of logarithms to factorise and simplify this expression,

$$\begin{aligned}
&\sum_{d_1,\ldots,d_k} P_{\mathrm{in}}(d_1,\ldots,d_k) \log \frac{P_{\mathrm{in}}(d_1,\ldots,d_k)}{P^*(d_1,\ldots,d_k)} \\
&= \sum_{d_1,\ldots,d_k} P_{\mathrm{in}}(d_1) P_{\mathrm{in}}(d_2,\ldots,d_k) \log \frac{P_{\mathrm{in}}(d_1)P_{\mathrm{in}}(d_2,\ldots,d_k)}{P^*(d_1)P^*(d_2,\ldots,d_k)} \\
&= \sum_{d_1,\ldots,d_k} P_{\mathrm{in}}(d_1) P_{\mathrm{in}}(d_2,\ldots,d_k) \left(\log \frac{P_{\mathrm{in}}(d_2,\ldots,d_k)}{P^*(d_2,\ldots,d_k)} + \log \frac{P_{\mathrm{in}}(d_1)}{P^*(d_1)} \right) \\
&= \sum_{d_2,\ldots,d_k} P_{\mathrm{in}}(d_2,\ldots,d_k) \log \frac{P_{\mathrm{in}}(d_2,\ldots,d_k)}{P^*(d_2,\ldots,d_k)} + \sum_{d_1} P_{\mathrm{in}}(d_1) \log \frac{P_{\mathrm{in}}(d_1)}{P^*(d_1)}.
\end{aligned} \tag{4.13}$$

By symmetry the same is true for the term concerning the out-degree distribution. Substituting in the simplified expression, we get

$$\begin{aligned}
\widetilde{\varsigma}^{(x_1,\ldots,x_k)} &= H_{\mathrm{JS}}(P_{\mathrm{in}}(d_2,\ldots,d_k), P_{\mathrm{out}}(d_2,\ldots,d_k)) + H_{\mathrm{JS}}(P_{\mathrm{in}}(d_1), P_{\mathrm{out}}(d_1)), \\
&= \widetilde{\varsigma}^{(x_2,\ldots,x_k)} + \widetilde{\varsigma}^{(x_1)}.
\end{aligned} \tag{4.14}$$

By induction, we have the following result.

Lemma 4.3.1. *For a k-tuple, $(x_1,\ldots,x_k)$, where all variables are independent, the DiMViGI irreversibility fully decomposes into the sum of the 1-order irreversibilities,*

$$\widetilde{\varsigma}^{(x_1,\ldots,x_k)} = \sum_{i=1}^{k} \widetilde{\varsigma}^{(x_i)}. \tag{4.15}$$

This is a particularly desirable property for such a measure of irreversibility in multivariate interactions.

4.4 Applications to MEG recordings during long-term auditory recognition

Next, we apply the DiMViGI framework to MEG recordings taken from human participants during a long-term memory recognition task involving auditory processing.

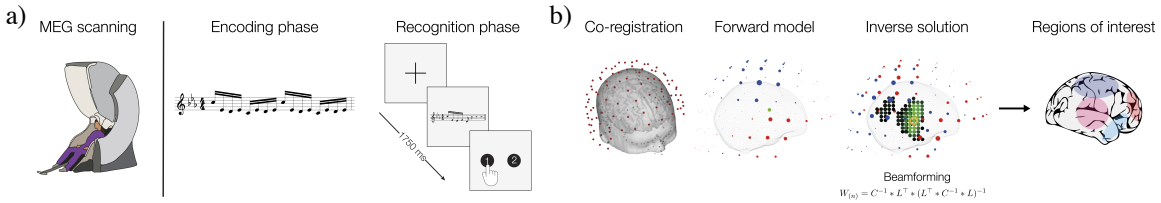

Figure 4.8: **Experimental paradigm and MEG recordings**. *a*) The brain activity in 51 participants was collected using MEG while they performed a long-term auditory recognition task where participants memorised a 5 tone musical sequence. *b*) The MEG data was co-registered with the individual anatomical MRI data, and source reconstructed using a beamforming algorithm. This procedure returned one time-series for each of the 3559 reconstructed brain sources. Six main functional ROIs were derived and neural activity for each ROI was extracted yielding an MVTS.

4.4.1 Experimental paradigm and MEG recordings

In this particular experiment, described at length in Ref. [151], participants performed an auditory recognition task whilst undergoing MEG recording. First, they memorised a five-tone musical sequence. They were then played five further sequences of tones that were either the original sequence or a modified version. They were then asked to state whether the sequence belonged to the original music or was a varied version of the original sequences. We will focus on the first experimental condition, using trials where participants recognised the original, previously memorised sequence.

We consider MEG recordings from 51 participants with 15 trials per participant source-localised into 6 ROIs. Fig. 4.8 shows a diagram of the experimental paradigm. The ROIs include: auditory cortex (left) (ACL), auditory cortex (right) (ACR), hippocampal and inferior temporal cortex (left) (HITL), hippocampal and inferior temporal cortex (right) (HITR), and two medial regions, the bilateral medial cingulate gyrus (MC) and the bilateral ventromedial prefrontal cortex (VMPFC). The ACL/R regions are *sensory* regions, whilst others can be considered *processing* regions. For details on preprocessing see Ref. [151]. The dataset, available in Ref. [165], is comprised of 51×15 samples of a 6 variable MVTS, each with 1026 time-points.

Pipeline For each participant and trial, we construct the DMVG. Next we estimate every marginal in- and out- degree distribution using each DMVG as a sample and calculate the k-order irreversibility. Alternatively, for each participant in isolation, the degree distributions can be calculated using only their associated trials to get an estimate of the k-order irreversibility for each participant and each tuple. However, due to the larger number of samples used to construct the distribution, the cohort-level

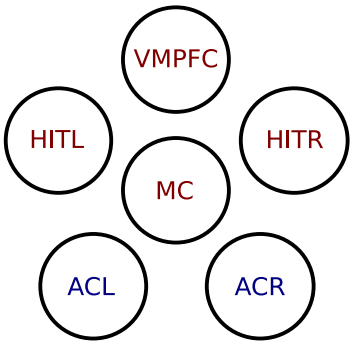

Figure 4.9: **Schematic representation of the ROIs**. This schematic serves as a legend for the icons that denote the tuples in the subsequent figures. The red regions are processing regions, whilst the blue regions are sensory. Regions that end in 'L' or 'R' are on left/right side of the brain respectively.

analysis is more robust and hence is our focus here. Nevertheless, participant-level analysis is available in App. B.1.

4.4.2 Results of the DiMViGI analysis

We present the results of the DiMViGI analysis for $k = 1$ to 5.[9] We denote each k-tuple with a small icon of 6 circles, where a circle represents each region, and blacked-in circles imply that a region is included in the tuple. The circles are in reference to the schematic in Fig. 4.9. Darker colours indicate more strongly irreversible interactions. Tuples that are marked with a $*$ are more than one standard deviation *above* the mean irreversibility at level k, with multiple $*$ indicating as many standard deviations. Those indicated with a † are *below* the mean irreversibility. As we will see, the raw divergence numbers increase significantly as we increase k. This is because we are computing an entropy over an exponentially growing state space. As a result, comparisons can only be made between k-tuples at a single value of k.

1-order irreversibility Fig. 4.10 shows the results at $k = 1$ i.e. individual regions. We find that the hippocampal regions are more irreversible than the medial regions, and that the sensory regions are the least irreversible, suggesting that the dynamics of processing regions are more complex than those of sensory regions. This hemispheric symmetry and consistency is encouraging, suggesting we are capturing genuine effects. Moreover, there is a skew towards the right side of the brain being more irreversible than the left. As we will see, this is consistent across k.

[9] We do not consider $k = 6$ as there is only a single tuple, the entire system, thus there is nothing to compare its irreversibility to.

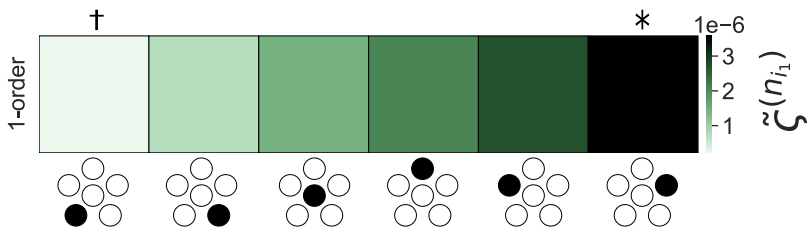

Figure 4.10: **1-order irreversibility**. We find that the hippocampal regions are more irreversible than medial regions, which in turn are more irreversible than sensory regions. There is a skew towards the right side of the brain being more irreversible than the left.

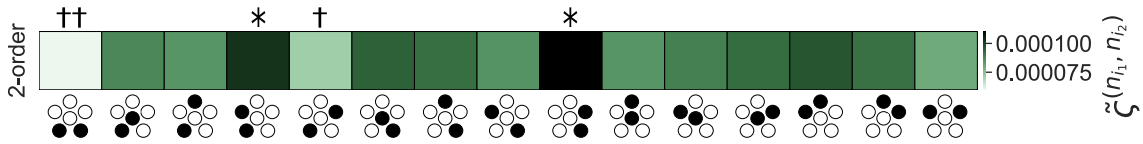

Figure 4.11: **2-order irreversibility**. We find that the corresponding hippocampal-sensory pairs on each side of the brain are the most irreversible. The least irreversible interaction is between the two sensory regions across hemispheres.

2-order irreversibility Fig. 4.11 shows the results at $k = 2$ i.e. pairs of regions. The most irreversible pairs are the hippocampal-sensory pairs on either side of the brain i.e. (ACL, HITL) and (ACR, HITR). Again, we see hemispheric symmetry. The results suggest a causal and hierarchical interaction at the pairwise level between the natural pair of sensory and processing regions in each hemisphere. Moreover, we see that the least irreversible pair is (ACL, ACR), the two sensory regions. We expect these regions to have very similar dynamics and be highly correlated, yet have mostly redundant information. This is supported by a lack of irreversibility in their interaction.

3-order irreversibility Fig. 4.12 shows the results at $k = 3$ i.e. triplets of regions. We see that the most irreversible triplets are those that include a hemispheric pair,

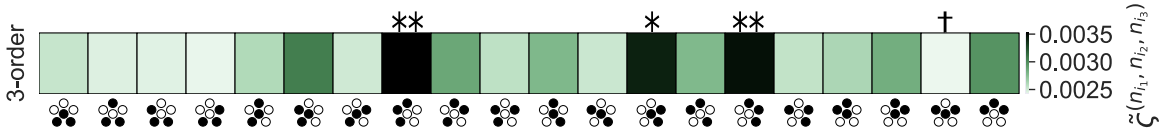

Figure 4.12: **3-order irreversibility**. We find that the corresponding hippocampal-sensory pairs alongside the VMPFC, are the most irreversible. The least irreversible interaction is between the two hippocampal regions with the MC i.e. (HITL, HITR, MC).

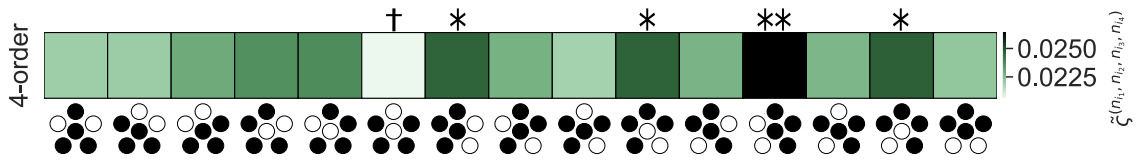

Figure 4.13: **4-order irreversibility**. We see that the most irreversible quadruplets are those that contain a hemispheric pair and both medial regions, which is in line with the results of at $k = 3$. Moreover, we find that the least irreversible tuple is the combination of the most irreversible pairs, indicating that we are capturing genuine HOIs that cannot be explained by pairwise relationships.

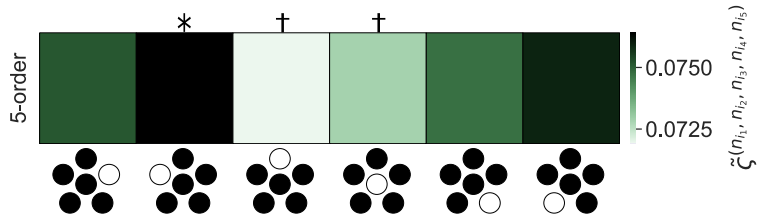

Figure 4.14: **5-order irreversibility**. At 5-order, the tuples correspond to leaving out a single region. We find that removing either of the medial regions results in significantly low irreversibility.

alongside the VMPFC i.e. (ACL, HITL, VMPFC) and (ACR, HITR, VMPFC), again showing hemispheric symmetry. On the other hand, when the VMPFC is not included, such as in the (HITR, HITL, MC) triplet, we find the least irreversible interaction. This suggests a crucial role for the VMPFC in facilitating communication across the cortex, that will be further supported at higher-orders. This is in line with the well-known, dominant role, of the prefrontal cortex in cognition dynamics during tasks [166].

4-order irreversibility Fig. 4.13 shows the results at $k = 4$ i.e. quadruplets of regions. We find that the most significantly irreversible tuples are those composed of a hemispheric pair alongside the medial pair i.e. (ACL, HITL, MC, VMPFC) and (ACR, HITR, MC, VMPFC). This again highlights the important role of medial regions in facilitating synergistic interactions. This is made even more clear when we notice that the least irreversible interaction is (ACL, ACR, HITL, HITR) which is a combination of the most irreversible pairs, reinforcing that we are capturing a higher-order effect rather than simply combining pairwise interactions.

5-order irreversibility Fig. 4.14 shows the results at $k = 5$ i.e. quintuplets of regions. Here, tuples correspond to 'leaving out a region'. We see that the interactions

with significantly low irreversibility occur when a medial region is left out, reiterating that medial regions are crucial to orchestrating HOIs across the hemispheres.

Significance testing and further analysis When measuring irreversibility in empirical time-series, there is an upward bias stemming from the finiteness of the data. More specifically, we must assess the significance of observed irreversibility to confirm that it emerges from the presence of genuine nonequilibrium interactions, and not from mere chance. First, we conduct *surrogate data testing*, where we generate a 'null time-series' by shuffling the data-points in time [3]. This breaks the temporal ordering and returns the process to detailed balance. As mentioned previously, the level of irreversibility measured in this null-series is the noise-floor, which is the irreversibility that stems from the finiteness of the data. In App. B.2, we show that all tuples have significantly higher irreversibility in the data than in the null time-series. In the published version of this work, we ran a number of additional validatory experiments, which we do not report in this thesis, such as: we considered a correction for finite data using a subsampling approach [167]; we presented an approach to measure the irreversibility of the specific k-order HOI in a k-tuple by subtracting all its sub-interactions; and we showed that limiting the support of the degree distribution makes the method more computationally efficient whilst minimally affecting the results.

4.5 Conclusions

Identifying causal and synergistic interactions in neural recordings is of fundamental importance to understand the complexity of neural dynamics. Whilst many approaches have been developed to analyse pairwise interactions [118], extending this to HOIs is a challenging and open problem. Inspired by the possible applications of NESM to neural dynamics, we suggest time-irreversibility as a correlate of a significant interaction that automatically generalises beyond pairwise relationships. Using the visibility algorithm, we were able to develop a principled approach to quantifying such irreversibility.

Applying it to MEG recordings during a long-term auditory recognition task, we found that it produces both novel and intuitive results. For example, we see hemispheric symmetry across levels, and we identify the crucial role of medial regions in facilitating dynamics across hemispheres. Our findings relate naturally to 'predictive coding' theories of cognitive processing, including that of musical sequences [168]. In brief, predictive coding is the idea that the brain operates as a Bayesian system, with

a prior world model that is updated in light of new sensory information. It is further hypothesised that this updating is facilitated through a hierarchical model of neural responses [136, 169], which, as shown in Ch. 3, are precisely the interaction structures that lead to irreversible dynamics. Furthermore, our results align with existing evidence for the hierarchical structure of the auditory system [151, 170, 171], where basic sensory processing interacts with high-level cognitive processes, facilitated by synergistic interactions that lead to irreversible dynamics. Our framework joins others that identify *functional* HOIs [132, 147] from observed dynamics, so-called *higher-order behaviours* [172], as opposed to the structural HOIs that dominate topological methods for complex systems [141]. It offers a new perspective on network neuroscience, with the potential to bring new insights into the nature of interactions in brain network dynamics.

Limitations and future directions Both the question we attempt to answer, and the specific methodological choices made in our framework introduce several limitations and opportunities for future work. First, with regards to the question itself, we note that our analysis benefits greatly from the use of a coarse parcellation into only six regions, where each has a well-known cognitive function. If we were to posit the same question using the fMRI dataset with 80 ROIs considered in Ch. 3, not only would it be computationally infeasible, but we would be unable to draw any meaningful conclusions e.g. how do we interpret a particular 27-order interaction that happened to be significantly irreversible?. As a result this question is useful in particular instances and with particular data.

Next, focusing on the framework itself, we note that we cannot compare across different values of k. This is a natural result of the divergence values increasing significantly between levels as we compute the divergence over an exponentially growing state-space. As a consequence, we implicitly define a significant k-order HOI to be one which is more significant than other potential k-order interactions. This is in contrast with other methods which define a k-order interaction to be significant if it is somehow greater than the sum of its lower-order parts [132]. It is this latter interpretation that is more in line with the mathematical theory of simplicial complexes. For instance, Lynn et al. present an approach that is able to quantify the irreversibility of entire levels k, which can then be compared across k [159]. This is a powerful approach to quantify the degree to which HOIs are present in a system, and how much information is lost by neglecting them. However, their analysis only

applies to discrete variables with multipartite dynamics and thus cannot be applied to neuroimaging in a principled manner.

Finally, we note that our measure is not directed. Unlike the regression approach in Ch. 3, or directed measures from information theory [133,152], our approach identifies a hierarchical interaction, but does not determine its direction.

Chapter 5

Discrete-state approximations of diffusion processes

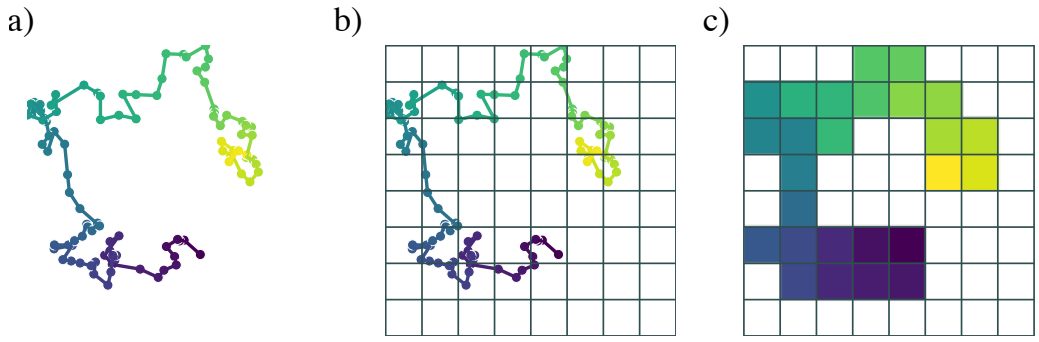

Figure 5.1: **Coarse-graining a stochastic path with discrete states.** *a*) We begin with a path, observed in discrete-time, in phase-space. *b*) We discretise phase-space into discrete voxels. *c*) By following the path through the voxels, we can treat this as an observation of a discrete-state stochastic process.

When monitoring the dynamics of stochastic systems, we typically only have access to observed paths with unknown drift and diffusion functions. In order to investigate properties of the NESS from these observed paths, we must infer a stochastic model which can then be analysed. Whilst there have been some attempts to directly infer the drift and diffusion coefficients from paths [119, 173], these methods can be particularly involved or specialised, thus a number of studies use a *discretisation* of phase-space into discrete *voxels*, where the evolution of the path can be approximated as a sequence of discrete states [3, 174–176] [Fig. 5.1]. Treating this sequence as a path from a discrete-state stochastic model, typically approximated as Markovian, we can then infer and analyse this model in lieu of the SDE. This capitalises on the fact that the maximum-likelihood estimation (MLE) of a D/CTMC from paths, typically

reduces to the simple procedure of counting observed jumps [177, 178].

Nevertheless, this approximation represents a *coarse-graining*, where parts of the underlying continuous information are lost, including those relevant to the NESS. For example, coarse-graining in space or time implies that the EPR can only be lower-bounded from observations of the coarse-grained process [45]. Moreover, it can introduce memory effects, where Markovian dynamics become non-Markovian under coarse-graining. A Markovian approximation of the EPR, for example using Eq. (2.32), is also a lower-bound on the true EPR for a non-Markovian process [179]. As a result, the consistency and convergence of discrete-state approximations to their underlying diffusions is particularly relevant, despite being often neglected in applied studies.

In this chapter, we begin by investigating the validity of a discrete-state Markovian approximation of an SDE, with a focus on the NESS. First, we use finite volume methods to derive an effective Markov process from a planar SDE, where we show that the EPR converges as the spatial discretisation goes to the continuum limit. Next, we discuss the inference of a CTMC from observed paths of an SDE, and show that the measured EPR is a weak approximation of the true value. However, we show that it can be used to test the hypothesis that a path is from a process in an NESS, which we illustrate with paths from schooling fish. In the second half of this chapter, we return to Hodge theory on graphs, and present a heuristic approach for decomposing stochastic processes into irreversible and reversible components directly from observed paths, which we illustrate on example processes and time-series from red-blood cell flickering, as well as healthy and arrhythmic heartbeats. Our results support previous findings that pathological physiological dynamics have lower time-irreversibility [180–182].

5.1 Deriving a discrete-state approximation

Non-Markovian approaches and related work It is well known that coarse-graining a D/CTMC, also known as *state-aggregating*, induces memory effects unless the chain is *lumpable* [158].[1] A similar effect occurs for diffusive processes. Consider the example in Fig. 5.2, where the red star represents a time-point where the path has entered a new state. At this point, the path is much closer to one face of the cell than the others, making it more likely to return than to transition to another adjacent cell, thus inducing memory effects into the process.

[1] For details on lumpability of Markov chains, we point the interested reader to Ref. [158].

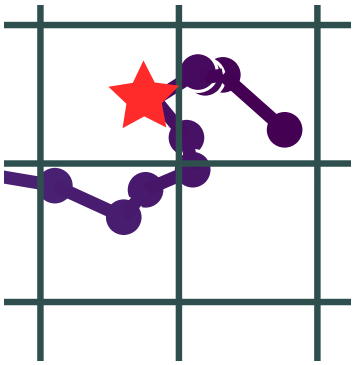

Figure 5.2: **Coarse-graining diffusive paths induces memory.** When approximating diffusion with discrete-states, we induce memory in the process. For example, the red-star represents a time-point where the path has entered a new discrete-state, but its proximity to the boundary makes it more likely to return the way it came, rather than to transition to a different cell.

We will focus on a Markovian approximation, but it is worth noting that some previous approaches have used first-passage time analysis and asymptotic expansions in an attempt to obtain the WTD of a discrete-state approximation [183–186]. Such analysis is closely related to methods for estimating the EPR using WTDs [187, 188]. However, these methods are limited in the complexity of their applications, typically focusing on one-dimensional systems, and thus are more applicable to small physical experiments, e.g. optical tweezers or beads under friction, rather than biological dynamics. In discrete-time, it is simple to infer 'higher-order'[2] Markov models from sequence data, and thus reveal the effects of memory in real stochastic dynamics [189, 190].

Other related works include the inverted problem of coarse-graining a microscopic Markov chain with an SDE in the so-called 'diffusion limit', where the EPR is lower-bounded by the approximation [191].

5.1.1 Approximation with finite volume methods

To develop a Markovian approximation, we will focus on two-dimensional diffusions with two independent sources of noise,

$$\begin{aligned} dx(t) &= f^x(x,y)\,dt + \sigma^x(x,y)\,dw^x(t), \\ dy(t) &= f^y(x,y)\,dt + \sigma^y(x,y)\,dw^y(t). \end{aligned} \tag{5.1}$$

[2] A k-order DTMC is one where $P(x_t|x_{t-1},\ldots,x_{t-k}) = P(x_t|x_{t-1},\ldots,x_0)$ i.e. there is a memory of k-steps. This is in contrast with the definition of higher-order used in Ch. 4.

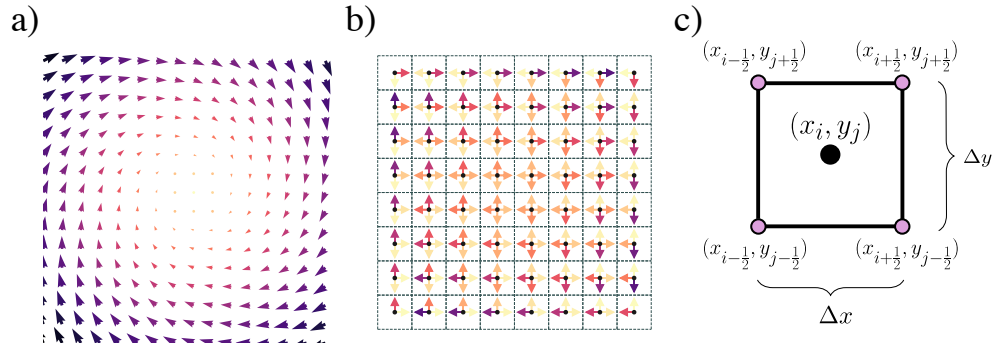

Figure 5.3: **Approximating a diffusion as a Markov process**. *a*) We aim to derive an approximation of an SDE using a finite-volume approximation. *b*) − *c*) We use a rectangular grid to coarse-grain $\mathbb{R}^2$ into a set of *volumes*, where we approximate the flux across the boundary, resulting in a CTMC.

This leads to a diagonal diffusion matrix with entries,

$$D(x,y) = \begin{pmatrix} D^x(x,y) & 0 \\ 0 & D^y(x,y) \end{pmatrix} = \frac{1}{2}\begin{pmatrix} (\sigma^x(x,y))^2 & 0 \\ 0 & (\sigma^y(x,y))^2 \end{pmatrix}. \tag{5.2}$$

and the FP equation,

$$\partial_t p(x,y,t) = -\frac{\partial J^x(x,y,t)}{\partial x} - \frac{\partial J^y(x,y,t)}{\partial y}, \tag{5.3}$$

where the flux field, $J = (J^x, J^y)$, is given by,

$$\begin{pmatrix} J^x(x,y,t) \\ J^y(x,y,t) \end{pmatrix} = \begin{pmatrix} f^x(x,y) - \partial_x D^x(x,y) \\ f^y(x,y) - \partial_y D^y(x,y) \end{pmatrix} p(x,y,t) - \begin{pmatrix} D^x(x,y)\partial_x p(x,y,t) \\ D^y(x,y)\partial_y p(x,y,t) \end{pmatrix}. \tag{5.4}$$

First, we discretise $\mathbb{R}^2$ into a grid of rectangular cells of dimension $\Delta x \times \Delta y$. We denote the midpoint of each cell B_{ij} to be (x_i, y_j), thus its boundary is the rectangle $\partial B_{ij} = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \times [y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}]$, where $x_{i\pm\frac{1}{2}} = x_i \pm \Delta x/2$, and the same for y, as shown in Fig. 5.3. The discretisation of a path proceeds by assigning a discrete-state to each voxel in the discretised space. Here, the process is in state $k = (i,j)$ at time t if the underlying path is in the corresponding cell i.e. $(x(t), y(t)) \in B_{ij}$. The corresponding MEq describes the evolution of $p_{i,j}(t)$, the probability that the process is in state (i,j).

A *finite volume approximation (FVA)* of Eq. (5.3) yields

$$\frac{dp_{i,j}}{dt} = -\frac{J^x_{i+\frac{1}{2},j} - J^x_{i-\frac{1}{2},j}}{\Delta x} - \frac{J^y_{i,j+\frac{1}{2}} - J^y_{i,j-\frac{1}{2}}}{\Delta y}, \tag{5.5}$$

where the subscript k,l implies the function is evaluated at (x_k, y_l). The expressions on the right of Eq. (5.5) represent the discrete flux through the bounding edges

of the rectangle. In practice, we enforce zero-flux at the boundary, which is chosen sufficiently far from the centre of mass of the stationary density, to simulate an infinite domain.[3]

It remains to adopt an approximation for the flux which will then lead to a MEq of the form,

$$\frac{dp_{i,j}}{dt} = \sum_{(k,l)} L_{(i,j),(k,l)} p_{k,l}. \tag{5.6}$$

We say that a numerical scheme is *valid* if the resulting MEq has $L_{(i,j),(k,l)} > 0$ for all off-diagonal entries ($(k,l) \neq (i,j)$), and probability is conserved i.e. the columns of L sum to 0.

Remark 5.1.1. In Ref. [64], Strang considered the naive centred-difference approximation in 1D, which only yields a valid MEq under step-size conditions. The problem of deriving a valid approximation is closely related to that of developing a structure-preserving scheme for the FP equation [192]. Other methods have been considered for similar discretisation problems [193], for solving nonlinear variants of the FP equation [194], and for simulating reaction-diffusion systems [195–197].

5.1.1.1 The Scharfetter-Gummel discretisation

We will focus on the *Scharfetter-Gummel discretisation (SG)*, a structure-preserving scheme which gives the following approximation for the flux across the boundary,

$$J^x_{i+\frac{1}{2},j} = \frac{f^x_{i+\frac{1}{2},j}\left(p_{i,j} - e^{-f^x_{i+\frac{1}{2},j}\Delta x/D^x_{i+\frac{1}{2},j}} p_{i+1,j}\right)}{1 - e^{-f^x_{i+\frac{1}{2},j}\Delta x/D^x_{i+\frac{1}{2},j}}}, \tag{5.7}$$

with analogous expressions for the remaining flux terms [198]. We provide a derivation in App. C.1. We opt for the SG scheme as it is known to preserve key properties of the underlying FP equation [199,200], but we note that it is just one in a larger family of methods [192].

The approximation leads to a MEq of the form

$$\frac{dp_{i,j}}{dt} = L_{(i,j)(i,j)} + \sum_{k=i\pm1} L_{(i,j)(k,j)} + \sum_{k=j\pm1} L_{(i,j)(i,k)}, \tag{5.8}$$

[3] 'Zero-flux' boundaries for the PDE can be interpreted as 'reflecting' boundaries for the CTMC or SDE. When the path hits a boundary state, it cannot go through, nor it is absorbed. One exception to this is when we consider the Kuramoto model satisfies periodic boundaries that we enforce in the discretisation.

with off-diagonal transition rates given by

$$L_{(i,j),(i+1,j)} = \frac{f^x_{i+\frac{1}{2},j} e^{-f^x_{i+\frac{1}{2},j}\Delta x/D^x_{i+\frac{1}{2},j}}}{\Delta x\left(1-e^{-f^x_{i+\frac{1}{2},j}\Delta x/D^x_{i+\frac{1}{2},j}}\right)}, \quad L_{(i,j),(i-1,j)} = \frac{f^x_{i-\frac{1}{2},j}}{\Delta x\left(1-e^{-f^x_{i-\frac{1}{2},j}\Delta x/D^x_{i-\frac{1}{2},j}}\right)},$$
$$L_{(i,j),(i,j+1)} = \frac{f^y_{i,j+\frac{1}{2}} e^{-f^y_{i,j+\frac{1}{2}}\Delta y/D^y_{i,j+\frac{1}{2}}}}{\Delta y\left(1-e^{-f^y_{i,j+\frac{1}{2}}\Delta y/D^y_{i,j+\frac{1}{2}}}\right)}, \quad L_{(i,j),(i,j-1)} = \frac{f^y_{i,j-\frac{1}{2}}}{\Delta y\left(1-e^{-f^y_{i,j-\frac{1}{2}}\Delta y/D^y_{i,j-\frac{1}{2}}}\right)}, \tag{5.9}$$

which are necessarily positive.[4] Moreover, the diagonal rates are given by,

$$\begin{aligned} L_{(i,j),(i,j)} = &- \frac{f^x_{i+\frac{1}{2},j}}{\Delta x\left(1-e^{-f^x_{i+\frac{1}{2},j}\Delta x/D^x_{i+\frac{1}{2},j}}\right)} - \frac{f^x_{i-\frac{1}{2},j} e^{-f^x_{i-\frac{1}{2},j}\Delta x/D^x_{i-\frac{1}{2},j}}}{\Delta x\left(1-e^{-f^x_{i-\frac{1}{2},j}\Delta x/D^x_{i-\frac{1}{2},j}}\right)} \\ &- \frac{f^y_{i,j+\frac{1}{2}}}{\Delta y\left(1-e^{-f^y_{i,j+\frac{1}{2}}\Delta y/D^y_{i,j+\frac{1}{2}}}\right)} - \frac{f^y_{i,j-\frac{1}{2}} e^{-f^y_{i,j-\frac{1}{2}}\Delta y/D^y_{i,j-\frac{1}{2}}}}{\Delta y\left(1-e^{-f^y_{i,j-\frac{1}{2}}\Delta y/D^y_{i,j-\frac{1}{2}}}\right)}. \end{aligned} \tag{5.10}$$

We can check that the sum,

$$L_{(i+1,j),(i,j)} + L_{(i-1,j),(i,j)} + L_{(i,j+1),(i,j)} + L_{(i,j-1),(i,j)} + L_{(i,j),(i,j)} = 0, \tag{5.11}$$

by replacing i by $i \pm 1$ and j by $j \pm 1$ into the derived rates. As a result, this scheme yields a valid MEq as it has non-negative rates between states and conserves probability.

Recovering the drift and diffusion coefficients Given L, we can recover the drift and diffusion at discrete points in the grid. To show this, we consider the rates,

$$L_{(i,j),(i+1,j)} = \frac{f^x_{i+\frac{1}{2},j} e^{-\xi}}{\Delta x(1-e^{-\xi})}, \qquad L_{(i+1,j),(i,j)} = \frac{f^x_{i+\frac{1}{2},j}}{\Delta x(1-e^{-\xi})}, \tag{5.12}$$

where $\xi = f^x_{i+\frac{1}{2},j}\Delta x/D^x_{i+\frac{1}{2},j}$. These are the transition rates from $(i+1,j)$ to (i,j) and its reverse, respectively. Dropping the subscript for convenience, we take the difference in the rates to obtain,

$$L_{(i+1,j),(i,j)} - L_{(i,j),(i+1,j)} = \frac{f^x - f^x e^{-\xi}}{\Delta x(1-e^{-\xi})} = \frac{f^x}{\Delta x}, \tag{5.13}$$

[4] In cases, where the denominator vanishes, which occurs when the drift vanishes, we can re-derive the SG discretisation to obtain $L_{(i,j),(i+1,j)} = D^x_{i+\frac{1}{2},j}/\Delta x^2$, and analogous terms for the other transitions [198].

thus we recover the drift with,

$$f^x_{i+\frac{1}{2},j} = \Delta x \left(L_{(i+1,j)(i,j)} - L_{(i,j)(i+1,j)} \right). \tag{5.14}$$

To recover the diffusion, we take the log-ratio of the rates to obtain,

$$\xi = \log \left(\frac{L_{(i+1,j)(i,j)}}{L_{(i,j)(i+1,j)}} \right) = \frac{f^x \Delta x}{D^x}. \tag{5.15}$$

We then use our previous expression for the drift to write the diffusion as,

$$D^x_{i+\frac{1}{2},j} = \frac{\Delta x^2 \left(L_{(i+1,j)(i,j)} - L_{(i,j)(i+1,j)} \right)}{\log \left(L_{(i+1,j)(i,j)} / L_{(i,j)(i+1,j)} \right)}. \tag{5.16}$$

The remaining drift and diffusion coefficients can be recovered in a similar way.

5.1.1.2 Convergence of the EPR

It is well-known that the SG scheme converges to the steady-state of the underlying FP equation and that the discrete solution converges to an ESS whenever the underlying solution does [192, 199].

We will show that the EPR of the discrete approximation converges to that of the underlying diffusion process as $\Delta x, \Delta y \to 0$. We begin by assuming that the discrete process is stationary with distribution $\pi_{i,j} = \Delta x \Delta y \, \pi(x_i, y_j) + O(\Delta x^3 \Delta y + \Delta x \Delta y^3)$, which is a midpoint approximation of the true stationary density $\pi(x, y)$. As each cell only receives transitions from the four adjacent cells, the EPR of the CTMC, given in Eq. (2.32), reduces to,

$$\Phi = \frac{1}{2} \sum_{i,j} \Phi_{(i,j),(i+1,j)} + \Phi_{(i,j),(i-1,j)} + \Phi_{(i,j),(i,j+1)} + \Phi_{(i,j),(i,j-1)}, \tag{5.17}$$

where each contribution is defined by,

$$\Phi_{(i,j)(i+1,j)} = \left(L_{(i,j),(i+1,j)} \pi_{i+1,j} - L_{(i+1,j),(i,j)} \pi_{i,j} \right) \log \frac{L_{(i,j),(i+1,j)} \pi_{i+1,j}}{L_{(i+1,j),(i,j)} \pi_{i,j}}, \tag{5.18}$$

with analogous expressions for the remaining terms. Substituting in the expressions derived from the SG discretisation, we obtain

$$\begin{aligned} \Phi_{(i,j)(i+1,j)} = & \frac{f^x_{i+\frac{1}{2},j}}{\Delta x \left(1 - e^{-f^x_{i+\frac{1}{2},j} \Delta x / D^x_{i+\frac{1}{2},j}} \right)} \left(e^{-f^x_{i+\frac{1}{2},j} \Delta x / D^x_{i+\frac{1}{2},j}} \pi_{i+1,j} - \pi_{i,j} \right) \times \\ & \left(-\frac{f^x_{i+\frac{1}{2},j} \Delta x}{D^x_{i+\frac{1}{2},j}} + \log \frac{\pi_{i+1,j}}{\pi_{i,j}} \right). \end{aligned} \tag{5.19}$$

Next, we perform an expansion to derive the leading order entropy production along an edge,

$$\Phi_{(i,j)(i+1,j)} = \Delta x \Delta y \frac{\left(-f^x_{i+\frac{1}{2},j}\pi(x_i,y_j) + D^x_{i+\frac{1}{2},j}[\partial_x \pi(x,y)]_{i,j}\right)^2}{D^x_{i+\frac{1}{2},j}\pi(x_i,y_j)} + O(\Delta x^2 \Delta y + \Delta y^2 \Delta x), \tag{5.20}$$

where we are using that the stationary distribution is convergent. Pairing the edges in the x-direction, we have that

$$\frac{1}{2}\left(\Phi_{(i,j)(i+1,j)} + \Phi_{(i,j)(i-1,j)}\right) = \Delta x \Delta y \frac{(J^x_{i,j})^2}{D^x_{i,j}\pi(x_i,y_j)} + O(\Delta x^2 \Delta y + \Delta y^2 \Delta x), \tag{5.21}$$

where J is the probability flux associated with the stationary density π. Combining this with the edges in the y-direction, we have that the EPR is,

$$\Phi = \Delta x \Delta y \sum_{(i,j)} \frac{(J^x_{i,j})^2}{D^x_{i,j}\pi(x_i,y_j)} + \frac{(J^y_{i,j})^2}{D^y_{i,j}\pi(x_i,y_j)} + O(\Delta x + \Delta y). \tag{5.22}$$

Finally, we take the limit as $\Delta x, \Delta y$ go to zero to obtain,

$$\lim_{\Delta x, \Delta y \to 0} \Phi = \int\int \frac{(J^x(x,y))^2}{D^x(x,y)\pi(x,y)} + \frac{(J^y(x,y))^2}{D^y(x,y)\pi(x,y)}\, dx\, dy, \tag{5.23}$$

which is precisely Eq. (2.18) in the case of a 2D process with diagonal diffusion matrix i.e. the EPR of the CTMC converges to that of the underlying diffusion process.

5.1.1.3 Diagonalisable and non-diagonal diffusion

The SG discretisation does not generalise to non-diagonal diffusion. However, there is a more general reason why this case is different. In the previous section, we used a *five-point stencil*, which implies that $dp_{i,j}/dt$ depends only on the four adjacent neighbours $\{p_{i+1,j}, p_{i-1,j}, p_{i,j+1}, p_{i,j-1}\}$. In App. C.2, we show that non-diagonalisable diffusion requires a *nine-point* stencil, including the additional terms $\{p_{i+1,j+1}, p_{i-1,j+1}, p_{i+1,j-1}, p_{i-1,j-1}\}$, which we illustrate with the centred-difference scheme. We show that this discretisation does not yield a valid master equation.

Nevertheless, we can treat the special case of diffusion which is *simultaneously diagonalisable*, which means that the diffusion matrix can be written as,

$$D(x,y) = V\Lambda(x,y)V^{-1}, \tag{5.24}$$

where $V \in \mathbb{R}^{N\times N}$ is a matrix whose columns are the common eigenvectors of $D(x,y)$ and $\Lambda(x,y) = \mathrm{diag}(\lambda_1(x,y), \ldots, \lambda_N(x,y))$ is the diagonal matrix of eigenvalues i.e. the

eigenvalues may change with space, but the eigenvectors remain constant. Taking the diagonalised matrix, $\Lambda(x, y) = V^{-1}D(x, y)V$, the original approach can be applied in the transformed coordinates, which we call (ν, η) in $\mathbb{R}^2$. This is equivalent to constructing a grid of *rotated rectangles* where the nodes are placed at,

$$(x_i, y_j) = v_1 i\Delta\nu + v_2 j\Delta\eta, \tag{5.25}$$

where v_1, v_2 are the eigenvectors of $D(x, y)$.

Example 5.1.1 (2-dimensional OU process with diagonalisable diffusion)**.** Consider a 2D OU process with,

$$B = \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}, \qquad D = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}, \tag{5.26}$$

i.e. where the diffusion is not diagonal, but diagonalisable and constant. We trans-

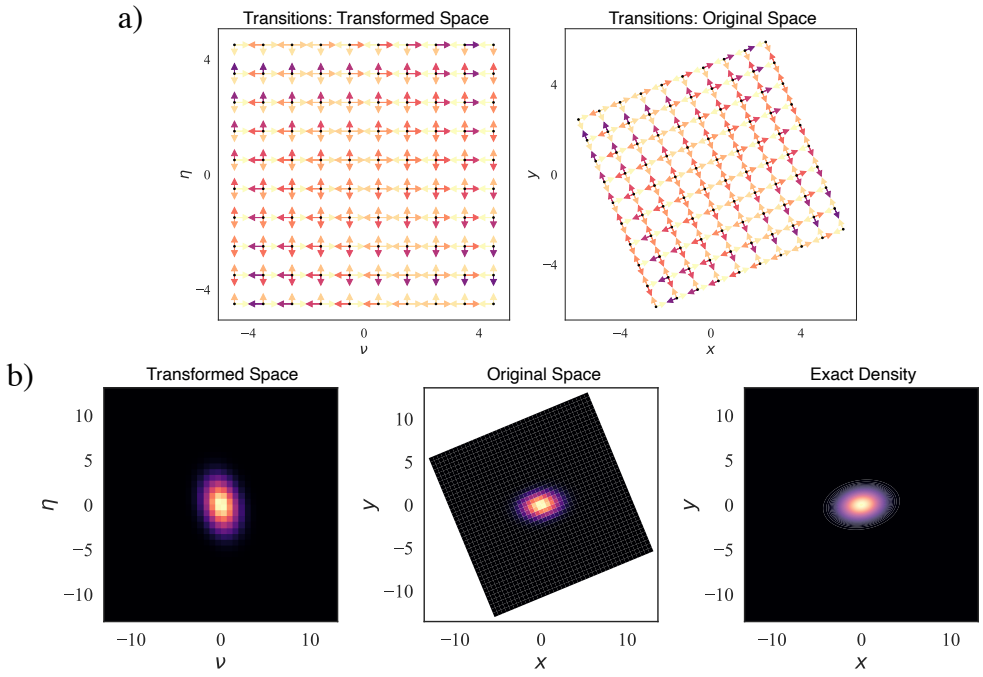

Figure 5.4: **Approximating processes with diagonalisable diffusion.** *a*) We can apply our original approach, using a square grid, in the transformed space where the diffusion is diagonal. We can then return to the original coordinates where the grid is then rotated. *b*) We illustrate this by calculating the stationary distribution of the OU process in Ex. 5.1.1, using the approximation, which we compared to the exact density.

form into the coordinate system where the diffusion is diagonal, and apply the SG discretisation with a square grid. The transition rates can then be considered as

acting on a rotated grid in the original coordinates. Fig. 5.4 illustrates the transition rates [Panel a)], and stationary distribution [Panel b)] compared to the exact density in both the original and transformed coordinates.

Remark 5.1.2. (Simultaneous diagonalisability) The condition, Eq. (5.24), is known as simultaneous diagonalisability. It implies that $D(x_1, y_1)$ and $D(x_2, y_2)$ have the same eigenvectors for all $x_1, x_2, y_1, y_2 \in \mathbb{R}$. This is equivalent to the matrices commuting [201],

$$D(x_1, y_1)D(x_2, y_2) = D(x_2, y_2)D(x_1, y_1). \tag{5.27}$$

Clearly, this is automatic when considering additive noise i.e. D is constant; or scalar multiplicative noise $D(x, y) = d(x, y)D_0$ where $d(x, y) \in \mathbb{R}$. For a process whose noise directions depend on the state, the condition is not met. A consistent approximation would require an adaptive, irregular grid. However, this severely complicates both theoretical and numerical analysis.

5.1.2 Examples and numerical experiments

We now illustrate the scheme on a number of examples, the first two have a closed-form NESS, while the third does not. In the following, we take $\Delta x = \Delta y$.

Remark 5.1.3 (Solving for the stationary distribution)**.** Solving for the steady-state of the CTMC involves solving the equation $L\pi = 0$. For a small number of discrete-states this can be calculated in an efficient and exact way, but for fine grids this calculation becomes prohibitively slow. Instead, we solve

$$\min_{\pi:\sum_j \pi_j = 1} ||L\pi||^2, \tag{5.28}$$

using LSQR [88], which capitalises on the sparsity of L. The entries of π should be guaranteed to have the same sign but numerical instabilities may occur. For extremely fine grids, the error incurred through this approximation can be comparable to the discretisation error, thus convergence analysis becomes challenging.

Example 5.1.2 (2-dimensional OU process)**.** We consider the 2D OU process,

$$d\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = -\begin{pmatrix} 2 & -\theta \\ \theta & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} dt + \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \end{pmatrix} dW(t), \tag{5.29}$$

which has covariance $S = \frac{\sigma^2}{4} I$ and $\Phi = \theta^2$.

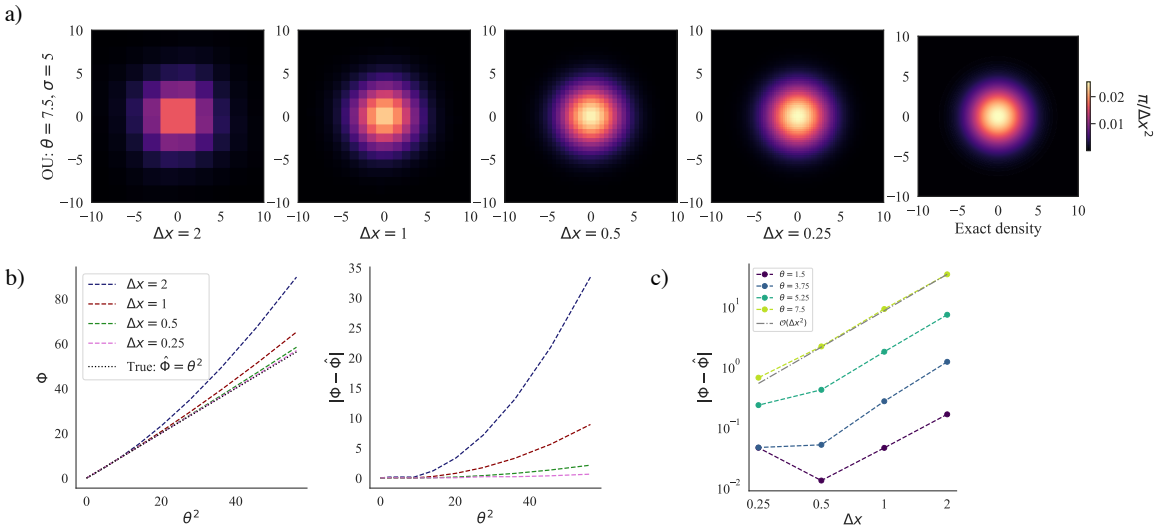

Figure 5.5: **Ex. 5.1.2. Approximating the NESS in the OU.** *a*) For a fixed value of $\theta = 7.5$, $\sigma = 5$, we calculate the stationary distribution of the OU process for decreasing step-sizes. The approximation preserves the symmetry of the stationary density and converges to the exact density. *b*) The EPR of the discrete-state approximation also increases approximately quadratically with θ, as in the exact solution. Moreover, we can see that the EPR converges as the step-size goes to 0. *c*) We plot the convergence to the true EPR as a function of the step-size Δx. Initially, we see convergence of approximately $O(\Delta x^2)$, but as Δx gets small, other errors are introduced, as discussed in Remark 5.1.3.

Fig. 5.5 shows the stationary distribution of the discrete approximation to this process for $\theta = 7.5$ and $\sigma = 5$, for a range of decreasing step-sizes. The symmetry of the distribution is preserved and converges in the limit [Panel *a*)].[5] We also see that the EPR converges to the true value and that the scaling of the EPR with θ follows the quadratic relationship (approximately) for the coarse-grained approximation [Panel *b*)].

Example 5.1.3 (The stochastic Hopf oscillator)**.** The (stochastic) *Hopf oscillator (HO)* is the Stuart-Landau oscillator with added noise, given by

$$\begin{aligned} dx(t) &= [(a - x^2 - y^2)x - \omega y]\, dt + \sigma dw_x(t), \\ dy(t) &= [(a - x^2 - y^2)y + \omega x]\, dt + \sigma dw_y(t), \end{aligned} \tag{5.30}$$

where ω is the natural frequency and σ is the noise intensity [23]. In the absence of noise, for $a < 0$ the system decays to the origin which is an attracting fixed point. At

[5] To compare between a discrete distribution that must *sum* to 1 and a continuous density that must *integrate* to 1, we normalise the discrete distribution π by the area of each volume Δx^2 i.e. the density in cell i can be approximated discretely as $\pi_i/\Delta x^2$.

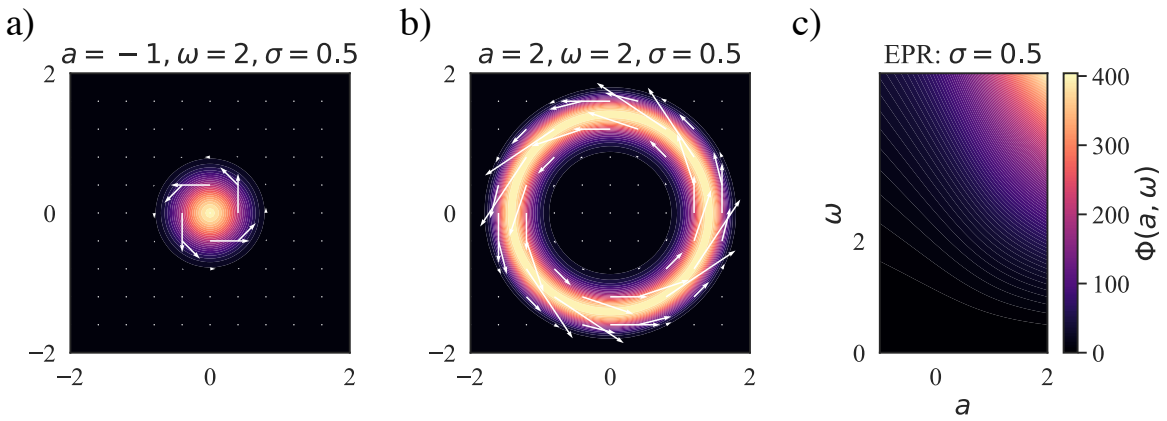

Figure 5.6: **NESS for the stochastic Hopf oscillator.** The stationary density and flux show that the Hopf oscillator converges to an NESS. *a*) For $a < 0$, it gives a distribution peaked at the origin. *b*) For $a > 0$, it is a 'Mexican-hat' distribution. Both show rotational probability flux due to the oscillatory dynamics. *c*) The EPR varies as a function of ω and a. Both ω and a drive irreversible dynamics, whilst σ drives reversible diffusion.

$a = 0$, the system goes through a Hopf bifurcation, resulting in limit-cycle oscillations for $a > 0$. This system has become a useful model of neural dynamics [23], and hair-cell bundles [202].

To derive the NESS, we consider the FP equation,

$$\partial_t p = -(\nabla \cdot f)p - (f \cdot \nabla p) + \frac{\sigma^2}{2}\nabla^2 p, \tag{5.31}$$

and replace $\nabla, \nabla\cdot$, and ∇^2 with their expressions in polar coordinates i.e. $x = r\cos\theta$ and $y = r\sin\theta$. We obtain

$$\partial_t p(r,\theta,t) = \left(4r^2 - 2a\right)p - \left(\left(a - r^2\right) r\frac{\partial p}{\partial r} + \frac{\omega}{r}\frac{\partial p}{\partial \theta}\right) + \frac{\sigma^2}{2}\left(\frac{\partial^2 p}{\partial r^2} + \frac{1}{r}\frac{\partial p}{\partial r} + \frac{1}{r^2}\frac{\partial^2 p}{\partial \theta^2}\right). \tag{5.32}$$

To solve for the steady-state, we assume a radially-symmetric solution of the form,

$$\pi(r,\theta) = \frac{1}{2\boldsymbol{\pi}}\pi(r), \tag{5.33}$$

where $\pi(r)$ satisfies

$$0 = \left(4r^2 - 2a\right)\pi + \left(\frac{\sigma^2}{2r} + r^3 - ar\right)\frac{d\pi}{dr} + \frac{\sigma^2}{2}\frac{d^2\pi}{dr^2}. \tag{5.34}$$

This equation is solved by a Boltzmann density,[6]

$$\pi(r,\theta) = \frac{1}{Z}\exp\left(\frac{r^2}{\sigma^2}\left(a - \frac{r^2}{2}\right)\right), \tag{5.35}$$

where Z is given by,

$$Z = \frac{1}{\sqrt{2}}\exp\left(\frac{a^2}{2\sigma^2}\right)\boldsymbol{\pi}^{3/2}\sigma\left(1 + \operatorname{erf}\left(\frac{a}{\sqrt{2}\sigma}\right)\right). \tag{5.36}$$

Transforming the stationary density to Cartesian coordinates, we can then perform the HHD to obtain,

$$f_{\text{rev}}(x,y) = \begin{pmatrix} x(a - x^2 - y^2) \\ y(a - x^2 - y^2) \end{pmatrix}, \qquad f_{\text{irr}}(x,y) = \begin{pmatrix} -\omega y \\ \omega x \end{pmatrix}. \tag{5.37}$$

The irreversibility of the process is driven by the frequency of oscillation and we have detailed balance only when $\omega = 0$. With Eq. (2.22), we calculate the EPR,

$$\Phi = \frac{\boldsymbol{\pi}\omega^2}{\sigma Z}\left(2\sigma + \sqrt{2\boldsymbol{\pi}}ae^{\frac{a^2}{4\sigma^2}}\left(1 + \operatorname{erf}\left(\frac{a}{\sqrt{2}\sigma}\right)\right)\right), \tag{5.38}$$

which scales quadratically in ω. Fig. 5.6 shows examples of the HO in an NESS for $a = -1$ and $a = 2$, the two regimes, and shows the EPR as a function of ω and a for a fixed value of σ.

Fig. 5.7 shows that, for fixed values of $a = 2, -1$ and $\sigma = 0.5$, the EPR of the coarse-grained process closely approximates that of the underlying diffusion. For $a = -1$, we show that it converges as the step-size decreases [Panel a)]. For $a = -2$, the convergence is not clear, even though the approximation is close to the true EPR [Panel b)]. This could be because we are outside the asymptotic regime (we use larger step sizes as the stationary distribution is much 'larger'), or because of error in solving for the stationary distribution numerically (see Remark. 5.1.3). Finally, for a fixed step-size of $\Delta x = 0.125$, we explore the EPR of the coarse-grained process as a function of a and ω [Panel c)], showing the same qualitative behaviour as the exact solution in Fig. 5.6.

Example 5.1.4 (Coupled, noisy Kuramoto oscillators)**.** In Ex. 2.2.3, we introduced a single Kuramoto oscillator with noise, and showed that it was in an NESS. Typically, Kuramoto oscillators are placed on a network with weight matrix (A_{ij}), where A_{ij} is the connection strength from oscillator j to i. We focus on a pair of Kuramoto

[6] We can see that $f^r = -\frac{dU}{dr}$ for a potential $U = \frac{1}{2}r^2(a - \frac{1}{2}r^2)$. All that remains is to check that $\pi(r) = \exp\left(-2U/\sigma^2\right)$ solves Eq. (5.34).

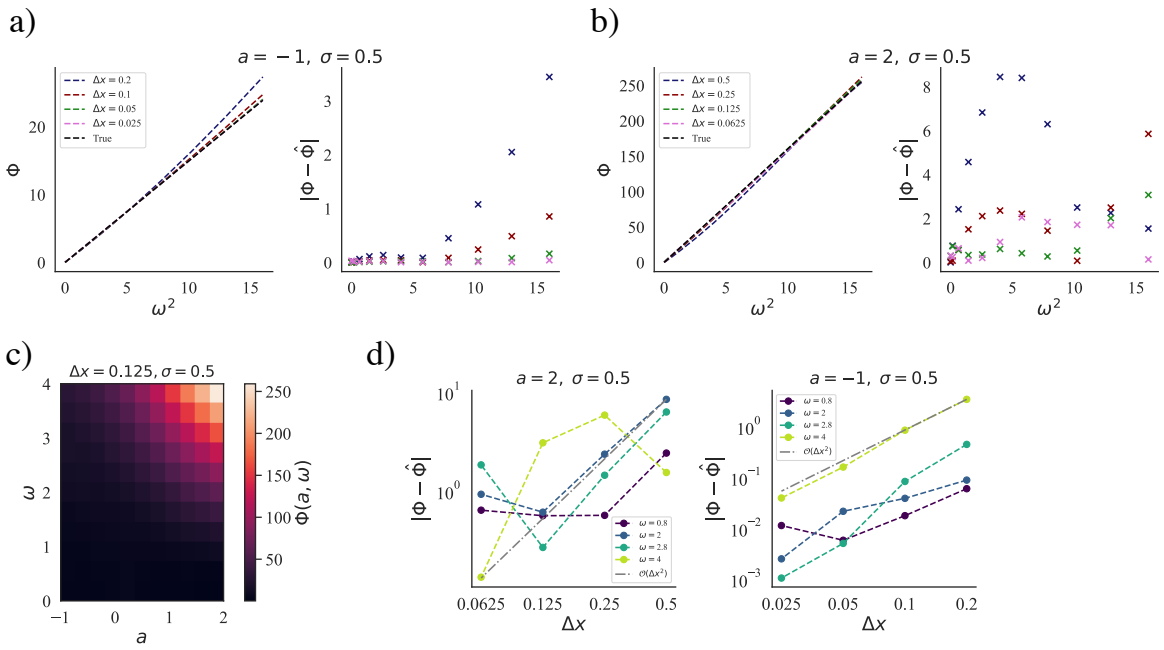

Figure 5.7: **Ex. 5.1.3. EPR in the coarse-grained Hopf oscillator.** For fixed values of a and σ, we can increase ω to increase the EPR of the HO. *a*) For $a = -1$, we see that the EPR converges as Δx goes to 0. *b*) For $a = 2$, we use larger step-sizes, and appear to be outside the asymptotic regime. We note that for some parameters, calculating the steady-state distribution numerically can induce another source of error not accounted for (see Remark 5.1.3). *c*) For a fixed step-size Δx, we show that the EPR follows the same qualitative behaviour as the true EPR of the underlying process, shown in Fig. 5.6. *d*) We plot the convergence to the true EPR as a function of the step-size Δx. As in Fig. 5.5, as Δx gets small, other errors are introduced. The convergence appears to be more robust for the fixed point attractor ($a = -1$) than for the limit-cycle oscillator ($a = 2$).

oscillators with additive noise, where their interaction is *frustrated*, a feature that can lead to nonequilibrium behaviour [203]. The process is given by,

$$\begin{aligned} d\theta_1(t) &= [\omega_1 + A_{12}\sin(\theta_2 - \theta_1 + \alpha)]\,dt + \sigma\,dw_1(t), \\ d\theta_2(t) &= [\omega_2 + A_{21}\sin(\theta_1 - \theta_2 + \alpha)]\,dt + \sigma\,dw_2(t), \end{aligned} \tag{5.39}$$

where $\theta_i \in [0, 2\boldsymbol{\pi})$ are the phases, $\omega_i > 0$ are the frequencies, and α is the frustration parameter.[7] As the stationary density is not obtainable analytically, we will employ our discrete approximation to analyse the NESS. Unlike the other examples, this process satisfies periodic boundary conditions, which we enforce in the discrete approximation. This implies that we have a flux at the boundaries. Fig. 5.8 shows the

[7]This model with a fixed phase-lag is known as the Kuramoto-Sakaguchi model and has been studied extensively [204].

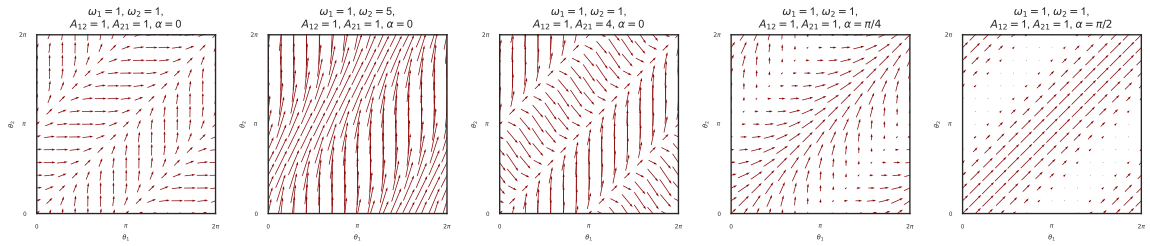

Figure 5.8: **Drift field of coupled Kuramoto oscillators.** We consider the drift field of a pair of Kuramoto oscillators. When the model is symmetric, the synchronised state is the only stable equilibrium, but this can be disrupted with asymmetry in the frequencies, interaction strengths or through frustration.

drift field for a range of different parameters. In the symmetric case, we see that the synchronised state is a stable equilibrium, which can be disrupted with asymmetry in the natural frequencies, interaction strengths or through frustration.

Remark 5.1.4. In the case $A_{12} = A_{21}$ and $\alpha = \pi/2$, the NESS is explicitly solvable with uniform stationary distribution, $\pi = 1/(4\boldsymbol{\pi}^2)$, and $\Phi = 2(\omega_1^2+\omega_2^2+A_{12}^2)/\sigma^2$. This shows that frustration can lead to an NESS, even in the case of symmetric coupling and no internal oscillations i.e. if we set $\omega_1 = \omega_2 = 0$.

Fig. 5.9 shows the stationary distribution obtained from the discrete approximation for a variety of different parameter values, which can shift and transform the distribution substantially [Panel *a*)]. Panels *b*) and *c*) show the discrete and continuous flux in the steady-state. Finally, we can approximate the EPR as a function of the parameters. We find that increased asymmetry in the network increases the EPR, in line with the results of Ch. 3, and that the EPR increases with the frequency, as in the 1D case (see Ex. 2.2.3) [Panel *d*)]. The final panel shows the EPR as a function of the frustration and noise intensity. Interestingly, the frustration that maximises the EPR appears to depend on the noise intensity.

5.2 Statistical inference for discrete-state models

Whilst we have focused on the deriving a CTMC from the drift and diffusion of an SDE, this approach is most useful when analysing paths from an unknown SDE. Consequently, we now turn to the problem of inferring a CTMC from a sequence of discrete states. This problem naturally bifurcates into two cases [178].

In the first case, we assume that we have information about K 'jumps', and our data takes the form, $\{(x_k, t_k) : k \in \{1, \ldots, K\}\}$, where x_k is the state the process

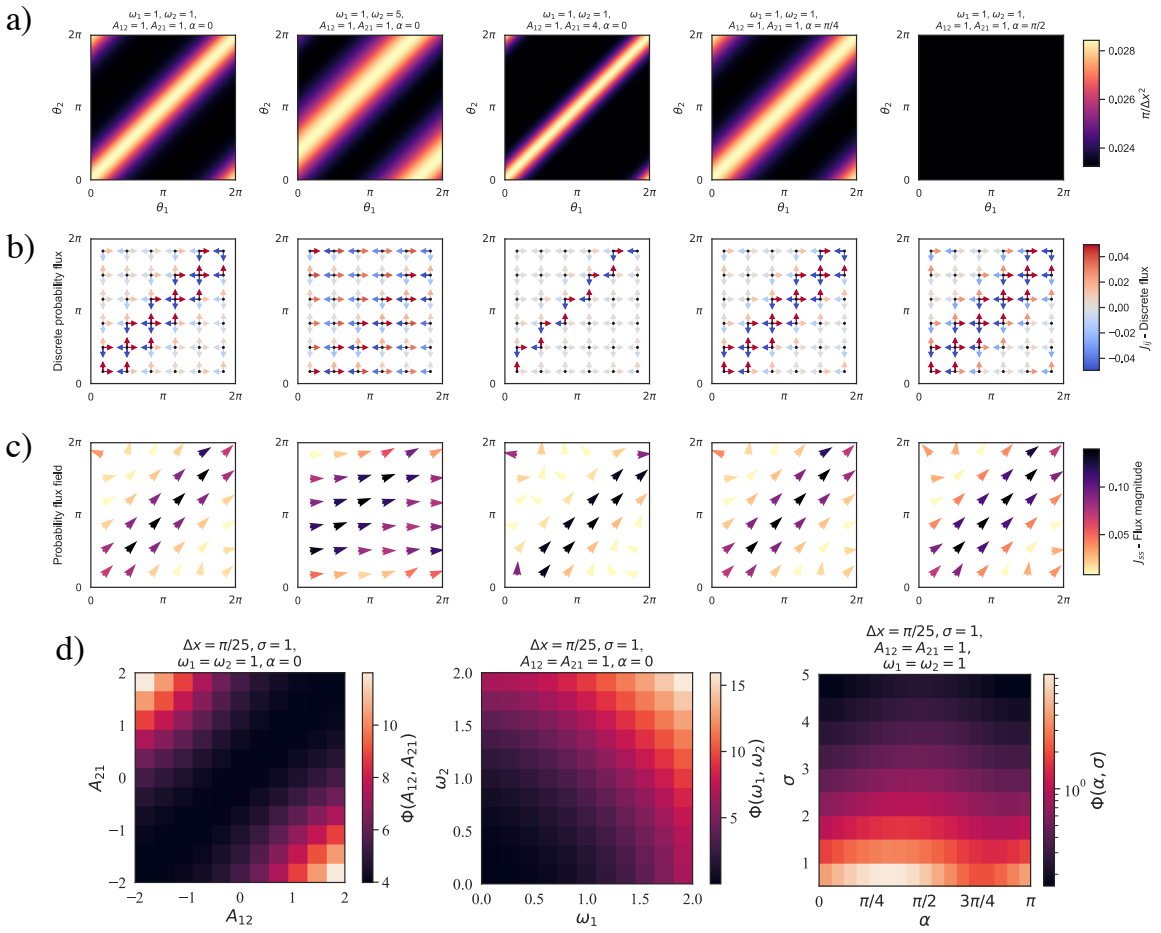

Figure 5.9: **Ex. 5.1.4. NESS for a pair of Kuramoto oscillators.** *a*) We compute the stationary distribution for the Kuramoto system using the discrete approximation for a number of different parameter settings. *b*) – *c*) We compute the probability flux in the discrete model, which aligns with the continuous flux, computed with the SG discretisation. *d*) We can compute the EPR as a function of the parameters. We find that increased asymmetry and oscillation increases the EPR. Interestingly, the effect of frustration is coupled to the noise intensity.

jumped into at 'jump time' t_k. Crucially, we know the *exact* jump times, rather than observing the process at discrete time-steps. In this case, the MLE of the transition rate from state j to state i is given by,

$$\widehat{L}_{ij} = \frac{N_{j\to i}}{T_j}, \tag{5.40}$$

where $N_{j\to i}$ is the observed number of jumps from state j to state i that occurred during the K steps, and T_j is the *holding time* i.e. the total time the process spent in state j during the path [178].

5.2.1 The embedding problem

In the second case, data come in the form $\{x_t : t = k\Delta t, k \in \{1, \ldots, K\}\}$, where we observe the state of the process at a series of discrete, equispaced snapshots i.e. a time-series. Bladt and Sørensen [178] show that Eq. (5.40) no longer maximises the likelihood function and is, instead, an approximation of the likelihood-maximising transition rate, with the accuracy depending on the time-step.

Instead, we can infer a DTMC where the MLE of the transition probability matrix is given by,

$$\widehat{P}_{ij} = \frac{N_{j\to i}}{\sum_k N_{j\to k}}, \tag{5.41}$$

where $N_{j\to k}$ is the number of observed transitions from state j to state k i.e. a straightforward counting of jumps [177, 178]. Here we find ourselves faced with the classical *embedding problem* for Markov matrices [205]. Given P, the embedding problem is to find L such that,

$$P = \exp(L\Delta t), \tag{5.42}$$

i.e. is there a CTMC whose observation in discrete-time aligns with this DTMC? Importantly, if L solves the embedding problem for $\widehat{P}$, then it is the MLE for the CTMC [178].[8]

Practical inference approach In situations with coarse temporal observations, it is recommended to use expectation-maximisation (EM) algorithms to obtain the MLE of the CTMC [178]. However, when coarse-graining a diffusion, temporal granularity is typically high, thus we opt to use Eq. (5.40) as an approximation of the MLE, which proved more robust and efficient than computing the matrix logarithm numerically.

Given our grid, we assume transitions are only possible between 'adjacent cells', thus we only count these transitions. If a transition occurs between non-adjacent cells, we discard it. We select our spatial resolution such that at least 70% of observed transitions are valid. For comparison, in App. C.3 we propose an approach for inferring $\widehat{L}$ from $\widehat{P}$ which does not discard transitions and uses non-convex optimisation over a feasible set of Laplacians with valid structure. However, the accuracy is comparable and the optimisation method is significantly less efficient. Additionally, we must also avoid singularities, thus we employ a Bayesian prior. This means we

[8] This relationship is not injective and a number of matrices L may solve Eq. (5.42). Bladt and Sørensen give a number of conditions on the matrix P that guarantee uniqueness [178].

add a 'pseudo-count' for all valid transitions i.e. initialise $\widehat{L}$ as if we have seen each possible transition once, before counting transitions in the data [190].

5.2.2 Numerical experiments

We have shown that a CTMC is a good approximation of the NESS of an underlying SDE when the rates are derived directly, but we have yet to see if this holds when performing inference from data. We investigate this using paths sampled from the examples from Sec. 5.1.2.[9]

We fix parameters $\theta = 5$ and $\sigma = \sqrt{2}$ for the OU, and $\omega = 5$, $\sigma = \sqrt{2}$, and $a = 1$ for the HO, and use $\Delta t = 10^{-3}$. We then sample 100 paths and use each one to obtain an estimate of the EPR, $\widehat{\Phi}$. This procedure is repeated for a fixed spatial step, $\Delta x = 0.25$, for a range of path lengths, and then for a fixed length with decreasing spatial step-size. Panels $a) - c)$ of Fig. 5.10 show the inferred stationary density, and EPR as a function of length and step-size for the OU, whilst Panels $d) - f)$ show the same results for the HO. We can see that the EPR estimate increases, improving towards the correct value, as we increase the path length or decrease the step-size. However, even with a small step-size and very long path, we find that we are significantly underestimating the EPR of the underlying SDE.

Remark 5.2.1. (Taking the time-step to zero)

Naively, one may think that taking Δt to zero, should also result in a converging approximation of the EPR. However, as shown in Fig. 5.11, this is not the case. We estimate the EPR from 100 sampled paths from the HO, for a series of decreasing time-steps, and see that it does not converge to the true EPR.

Whilst we would expect convergence when taking both the time and spatial resolution to the limit, taking the time resolution to the limit whilst keeping the spatial resolution fixed is a more complex scenario. It is true that as the time-step decreases initially, the transition rates are better estimated leading to an improved estimate of the EPR. However, beyond a certain point, this process begins to exhibit increasingly strong non-Markovian effects. As a result our Markovian estimator becomes increasingly inaccurate.

[9] We sample from the OU process using the exact simulation given in Eq. (3.29). For the HO, we do not use EM as it performs poorly on oscillatory models, and instead use a geometric splitting method detailed in App. C.4.1.

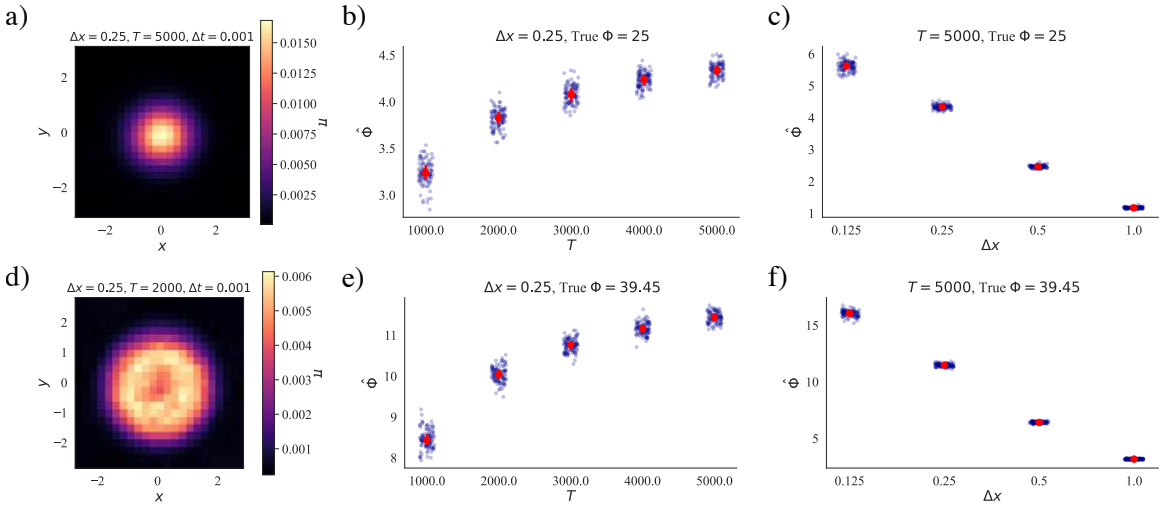

Figure 5.10: **Inferring NESS with discrete approximations.** *a)* We infer a CTMC and calculate the stationary distribution for the OU with $\theta = 5$ and $\sigma = \sqrt{2}$. It appears to be a good approximation of the true density. *b)* $-$ *c)* Using 100 sampled paths, we estimate the EPR for paths of different lengths and using different grid-sizes, where the red point represents the sample mean. We find that using longer paths and smaller grids leads to higher accuracy, but that the inferred EPR is substantially lower than the true value. *d)* $-$ *f)* We perform the same simulations for the HO with $\omega = 5$, $\sigma = \sqrt{2}$ and $a = 1$.

It is also worth noting that drift scales as $O(\Delta t)$, whilst diffusion scales as $O(\sqrt{\Delta t})$,[10] thus, as the time-step decreases, reversible diffusion dominates irreversible drift leading to increased number of reversible transitions i.e. the per-step drift-to-diffusion ratio goes to zero.

5.2.3 Determining if paths are from an NESS

5.2.3.1 The shuffling procedure

When inferring the EPR from sampled paths, by only considering finite data, we may obtain an unreliable estimate of the true EPR. As mentioned in Sec. 4.4 and in App. B.2, we naturally overestimate the irreversibility and flux, and could incorrectly determine that a reversible stationary process is out of equilibrium. For example, if we witness an odd number of transitions between two states, we will naturally observe an asymmetry, even if the true flux between the states is zero. As a result it is necessary to perform 'surrogate testing' to determine if the measured irreversibility and EPR is

[10] This scaling is most evident when looking at the Euler-Maruyama scheme in Eq. (2.10). The drift contribution to each increment is $f(x_t)\Delta t = O(\Delta t)$, whilst the diffusion is $\sigma(x_t)\sqrt{\Delta t}\xi_t = O(\sqrt{\Delta t})$ with $\xi_t \sim \mathcal{N}(0, 1)$.

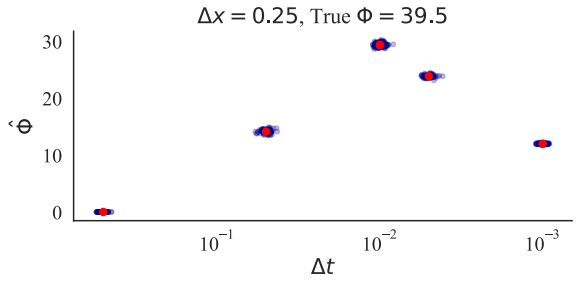

Figure 5.11: **EPR for decreasing time-steps.** We infer a CTMC from sampled paths from the HO and measure the EPR, for a series of decreasing time-steps and a fixed spatial resolution. We find that the EPR does not converge as the drift to diffusion ratio goes to zero.

a genuine statistical feature, or an artefact arising from the 'noise-floor', the level of irreversibility that is induced by finite data. A typical approach to surrogate testing for time-irreversibility and entropy production is to shuffle paths in time, breaking the temporal order, and then re-infer the EPR to obtain an estimate of the noise-floor [3,174], as we conducted in App. B.2. Whilst this approach works for some *direct* measures of time-irreversibility, it is unsuitable for this Markov inference approach. This is because during the inference, we only consider transitions between adjacent cells – choosing the resolution of our grid such that this accounts for a sufficient number of time-points. When the path is randomly shuffled, transitions between neighbouring cells become extremely rare, thus there are not enough valid transitions to infer the surrogate model.

Instead, to test if the dynamics of the stochastic data are in fact from an NESS, we analyse the path as before, with one crucial difference. For each valid transition in the path, we flip the direction of the original transition with probability 1/2, thus erasing the net flux in the system that results in entropy production and irreversibility. Then the EPR of the genuine path can be compared to a surrogate distribution obtained from different realisations of the surrogate model. The null hypothesis is that the original path is from an ESS, thus we can test if the path is, in fact, from an NESS using a one-sided $t-$test.

We illustrate this approach on paths from the OU in Ex. 5.1.2, with $\theta = 5$, an irreversible process with $\Phi = 25$, and $\theta = 0$, a reversible process. Fig. 5.12 compares between the EPR of 100 paths sampled from each OU compared with the 100 surrogate models. At $\theta = 5$, we find that the EPR of the original paths is significantly higher (one-sided independent $t-$test; ****, $p < 0.0001$), than the surrogate paths, thus supporting the conclusion that the original paths are sampled

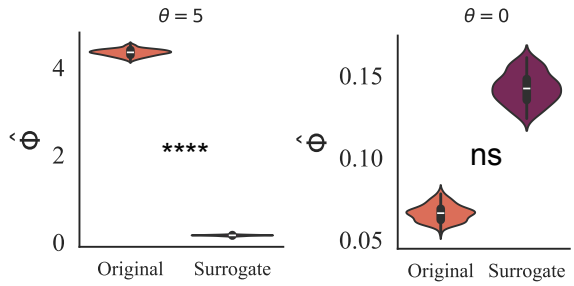

Figure 5.12: **Surrogate testing for NESS.** Using the shuffling procedure, we construct surrogate Markov models from observed sequence data, and use this to perform testing to identify if real-world paths are from an ESS or an NESS. We illustrate this with paths from the OU process at $\theta = 5$ and $\theta = 0$. Using a one-sided t-test, we find that at $\theta = 5$ the process has significant (****, $p < 0.0001$) entropy production, thus is in an NESS, whilst at $\theta = 0$, the process does not have significant entropy production.

from an NESS. At $\theta = 0$, we find that the original paths are not significantly higher than the surrogate models (one-sided independent $t-$test; ns, $p > 0.05$), thus we conclude that the original process is reversible.[11]

5.2.3.2 Applications to polarisation in schooling fish

Next, we apply this approach to the real-world case of polarisation in the dynamics of schooling fish and show that such a process is *not* in an NESS. Schooling, or flocking, is a well-known phenomenon in collective dynamics and active matter. Collective dynamics are inherently *active* at a local level [206], as energy is dissipated in order for agents to adjust their states dynamically, yet their collective behaviour is often described by ME distributions in an ESS respecting time-reversal symmetry [207–209]. As a result, it is interesting to investigate the evidence for probability flux in empirical observations of these systems.

We follow the analysis and use the dataset presented in Jhawar et al. [210], who analyse populations of $N = 15, 30$ and 60 fish moving in a circular tank, modelled as the unit disk in 2D i.e. $D_1(0) \subset \mathbb{R}^2$. In particular, the positions, $\mathbf{x}_i(t) \in \mathbb{R}^2$, and velocities, $\mathbf{v}_i(t) \in \mathbb{R}^2$, of fishes $i = 1, \ldots, N$ can be extracted from video recordings. The *group polarisation* is an order parameter describing their collective alignment

[11] We note that in this case, the surrogate paths appear to have a significantly higher EPR than the original time-series, but the result is not significant as the $t-$test is one-sided. In theory, this shuffling approach should bring a NESS closer to detailed balance. On the other hand, if a path is genuinely reversible, flipping a number of transitions is likely to *increase* the EPR artificially, thus this approach only allows for one-sided testing.

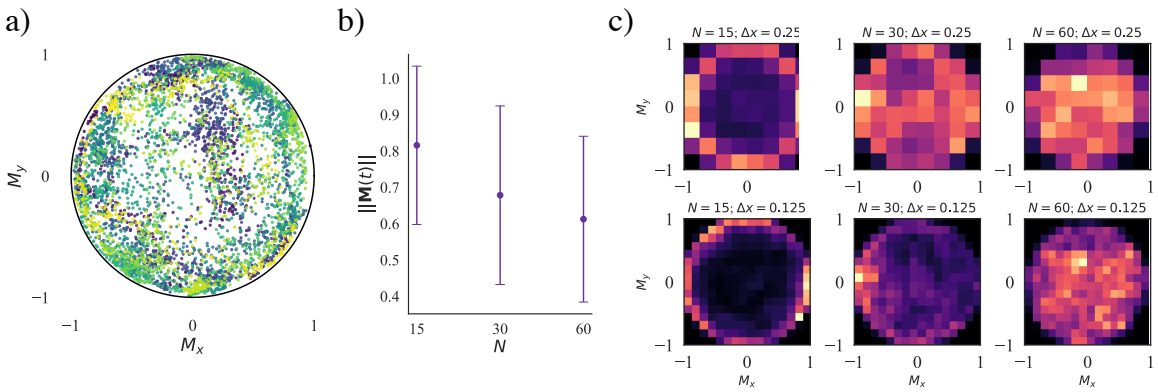

Figure 5.13: **Stationary dynamics in schooling fish.** *a*) We consider the group polarisation vector $\mathbf{M} = (M_x, M_y)$ for populations of schooling fish of size $N = 15, 30$ and 60. *b*) We find that the degree of collective alignment increases as group-size decreases, which can be seen by computing the mean and standard deviation of $||\mathbf{M}(t)||$ over time. *c*) We infer infer a CTMC from the paths and calculate its stationary density, which again illustrates the higher level of collective alignment for smaller populations of fish.

and is given by,

$$\mathbf{M}(t) = \frac{1}{N} \sum_{i=1}^{N} \widehat{\mathbf{v}}_i(t), \tag{5.43}$$

where $\widehat{\mathbf{v}}_i = \mathbf{v}_i / ||\mathbf{v}_i||$ is the normalised velocity of fish i and $\mathbf{M} \in D_1(0)$. A value of $||\mathbf{M}||$ close to 1 represents a coherent collective direction, whilst $||\mathbf{M}|| \approx 0$ implies no collective direction and isotropic individual movement.[12]

Jhawar et al. infer a 2D SDE for the dynamics of $\mathbf{M}(t)$, and then study its stationary behaviour, not including its nonequilibrium properties [210]. We will instead infer a CTMC using a square grid approximation of $D_1(0)$ [Fig. 5.13]. Fig. 5.13 shows an example path of the group polarisation vector [Panel *a*)]. Using a discrete approximation, we infer a CTMC and compute the stationary distributions for different grids and population sizes [Panel *c*)]. We see that collective alignment increases as population size decreases, as found by Jhawar et al. [210], illustrated by the mean value $||\mathbf{M}(t)||$ [Panel *b*)], and the higher stationary densities at the boundary of the unit disk [Panel *c*)].

Is this an NESS? We now investigate if the stationary dynamics of the schooling fish are best described as an NESS or an ESS. For each of the population sizes, we

[12]Note that the sum of normalised velocities is equivalent to the sum of pairwise squared chordal distances between headings [211].

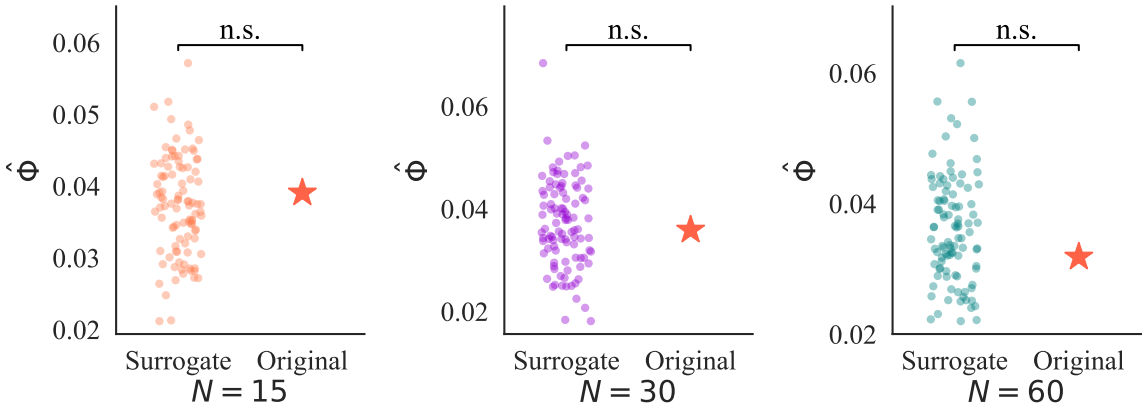

Figure 5.14: **EPR for populations of schooling fish.** By inferring a CTMC we calculate the EPR of the NESS from paths of the polarisation of schooling fish, which is represented by the single orange star in each panel. Using the shuffling procedure, we obtain a 'null' distribution of the EPR in 100 surrogate models. We use a one-sided t-test to compare the single empirical path with the null distribution, and find that the result is not significant across all population sizes ($p > 0.05$).

infer a CTMC using the original path, and then compute the EPR. For comparison, we apply the shuffling procedure detailed in Sec. 5.2.3.1 to compute the EPR of 100 surrogate models.

Fig. 5.14 shows the EPR computed using the single original path (orange star) compared with the surrogate values, for each of the system sizes. Performing a one-sided $t-$test, we determine that the EPR of the original path is not significantly different from the null distribution, which indicates that there is not enough evidence to conclude this process is in an NESS, and supports the validity of equilibrium, ME models for schooling collectives.

Remark 5.2.2 (Observations and odd variables)**.** As mentioned in Remark 2.2.3, we focus on SDEs with even variables in this thesis, and by extension in our derivation in Sec. 5.1. Whilst Jhawar et al. [210] show that the dynamics of $\mathbf{M}(t)$ do follow a 2D SDE, it is important to note that $\mathbf{M}$ is an odd variable, thus its flux should take the more complex form mentioned in Remark 2.2.3 [67]. Moreover, choosing to focus on the dynamics of the velocities implies a partial observation and a spatial coarse-graining, which we have stated can obscure currents in the NESS. In light of these limitations, it is more difficult to conclude if the system is, or is not, truly in a NESS.

5.3 Decomposing processes as flows on graphs

5.3.1 The many faces of the Helmholtz-Hodge decomposition

We have arrived at the second half of this chapter, where we will continue to consider discrete approximations of SDEs, but will instead focus on the HHD of an SDE presented in Sec. 2.2.2.2.

To briefly recap, an SDE that converges to a stationary distribution admits the following decomposition of the drift,

$$f = f_{\text{rev}} + f_{\text{irr}}, \qquad f_{\text{rev}} = D\nabla \log \pi + \nabla \cdot D, \qquad f_{\text{irr}} = J_{\text{ss}}/\pi, \tag{5.44}$$

i.e. into a reversible and irreversible component. In the case of constant, isotropic diffusion, $D(x) = \frac{\sigma^2}{2} I$, this simplifies to,

$$f = \nabla \left(\frac{\sigma^2}{2} \log \pi \right) + \frac{J_{\text{ss}}}{\pi}, \tag{5.45}$$

i.e. the drift splits into a gradient flow and a component that is DFSD. In this case, the decomposition is highly reminiscent of the *Helmholtz decomposition* of a vector field $f = \nabla\varphi + R$, commonly seen in electrodynamics and other fields [62]. For isotropic diffusion, Corr. 2.2.1 implies that this process obeys detailed balance if and only if f is conservative i.e. $f = \nabla V$ for some potential V.

Being able to perform the HHD of an irreversible process can give important insight into the structure of its governing forces, and has been used to analyse biochemical networks [212], neural networks [213], genetic circuits [214], and more [215]. However, such analyses are typically constrained to theoretical models where we have direct access to the drift and diffusion, and can compute the stationary distribution. On the other hand, computing the HHD directly from stochastic paths would require the inference of the drift and diffusion at all points in phase-space [119].

Whilst computing the HHD of an SDE is challenging, we recall that the analogous decomposition for discrete vector fields is particularly easy to compute, as outlined in Sec. 2.3. Recall that, given a graph $G = (V, E)$, the GHHD decomposes an edge-flow $X : E \to \mathbb{R}$,

$$X = \operatorname{grad} f + R, \tag{5.46}$$

where $f : V \to \mathbb{R}$, $\operatorname{grad}(\cdot)$ is the discrete gradient operator, and R is a divergence-free edge-flow. This decomposition is precisely the discrete analogue of Eq. (5.45),

although it remains to define X given a particular SDE, and then introduce a scheme to construct X from observed paths.

The key insight relating the discrete and continuous decompositions is given by the following result, as proved by Strang [64].[13][14]

Lemma 5.3.1 (Strang (2020))**.** *A CTMC obeying microscopic reversibility, with Laplacian L and associated graph $G = (V, E)$, obeys detailed balance if and only if the corresponding edge-flow, $X \in L^2(E)$,*

$$X(i,j) = \frac{1}{2}\left(\log L_{ij} - \log L_{ji}\right), \tag{5.47}$$

is conservative i.e. $X = \operatorname{grad} f$ for some $f \in L^2(V)$. The stationary distribution is then a Boltzmann distribution with $\pi_i \propto \exp(-2f_i)$.

This result indicates that, by approximating the SDE with a discrete-state approximation, we can define an edge-flow with precisely the correct properties. In other words, given paths from an SDE, we construct an approximating CTMC, compute X with Eq. (5.47), and perform the GHHD in Eq. (5.46). The reversible dynamics will be captured by the grad f component and the irreversible dynamics will be captured by R.

This choice of flow in Eq. (5.47) may appear somewhat artificial, however it is motivated by both thermodynamic and mathematical arguments. For example, it can be identified with the excess work required to move from state j to state i [216]. We define it as the (discrete) *thermodynamic force*, in analogy with the vector field of the same name in the case of the SDE (Eq. (2.23)). As a result, Lemma 5.3.1 can be seen as the exact discrete analogue to Corr. 2.2.1.

The challenge of convergence The decomposition of a CTMC into two separate processes, one which obeys detailed balance and one which does not, is not unique. The GHHD itself is defined using a choice of inner-product (see Sec. 2.3), and whilst the L^2 inner-product is a common and reasonable choice, it is not thermodynamically motivated. In fact, this choice results in a reversible chain and irreversible remainder that 'smooths the irreversibility' over all the edges in the graph, thus minimising the L^2 norm of $X - R$. However, this is not necessarily consistent with what we would expect in a physical system.

[13] This result appears as Theorem 24 in Ref. [64].

[14] For the remainder of this chapter, we will consider the Hodge decomposition with respect to the standard L^2 inner-product (see Sec. 2.3).

The convergence of the approximating CTMC to the SDE in the diffusion limit is not sufficient to guarantee that the GHHD and HHD will also converge component-wise. This problem of convergence is understudied, with only a few results contributed by Strang [64]. For example, Strang shows that CTMCs that converge to an SDE in the diffusion limit will not converge with respect to the HHD unless the family of transition graphs associated to the CTMC converge to $\mathbb{R}^N$. However, this requirement is satisfied for the rectangular lattice, provided the SDE has diagonal diffusion, where Strang shows convergence between the components of the HHD.[15]

In this section, unlike in the previous half of this chapter, we will not focus on convergence in the limit, but instead will develop a heuristic, approximate, and data-driven method for decomposing SDEs from path data using a discrete-state approximation and the GHHD. The goal is not to recover the *exact* components of the HHD, but instead to give some insight into the structure, location, and strength of irreversible currents from stochastic paths; as well as to draw an interesting relationship between two seemingly distant areas of mathematics.

5.3.2 Data-driven grids and the pipeline

Unlike Sec. 5.1.1, we will not use a square lattice, but instead will construct a grid from the points of the path using a data-driven procedure, illustrated in Fig. 5.15.

We begin with an ensemble of paths in $\mathbb{R}^N$, sampled from the same stationary SDE. In the first case, we assume that the noise is isotropic, and so admits the HHD in Eq. (5.45), but we will later consider the case of non-isotropic diffusion.

Step 1. We construct a data-driven *tessellation* of phase-space using the points in the dataset. Ignoring time, we consider the union of all the paths to be a point-cloud in $\mathbb{R}^N$ [Panel *a*), Fig. 5.15]. Following Gosztolai et al. [217], we perform 'farthest point sampling' [218], which subsamples points from the cloud and yields a uniform distribution over the manifold explored by the paths [purple points, Panel *b*)].[a]
Step 2. Next, we triangulate these points with the *Delauney triangulation*, which yields an irregular triangular grid containing the point-cloud [220]. We calculate the midpoint of each triangle [orange crosses, Panel *c*)].
Step 3. Next, we can infer the rates of the CTMC from the observed paths with Eq. (5.40), as in Sec. 5.2, using each triangular bin as a discrete state, and only allowing transitions between adjacent triangles [Panel *d*)].[b]
Step 4. At this point, we have a triangulation with a CTMC that acts between triangles. To obtain the transition graph of the CTMC, where nodes are states

[15] See Sec. 4.8.1 in Ref. [64].

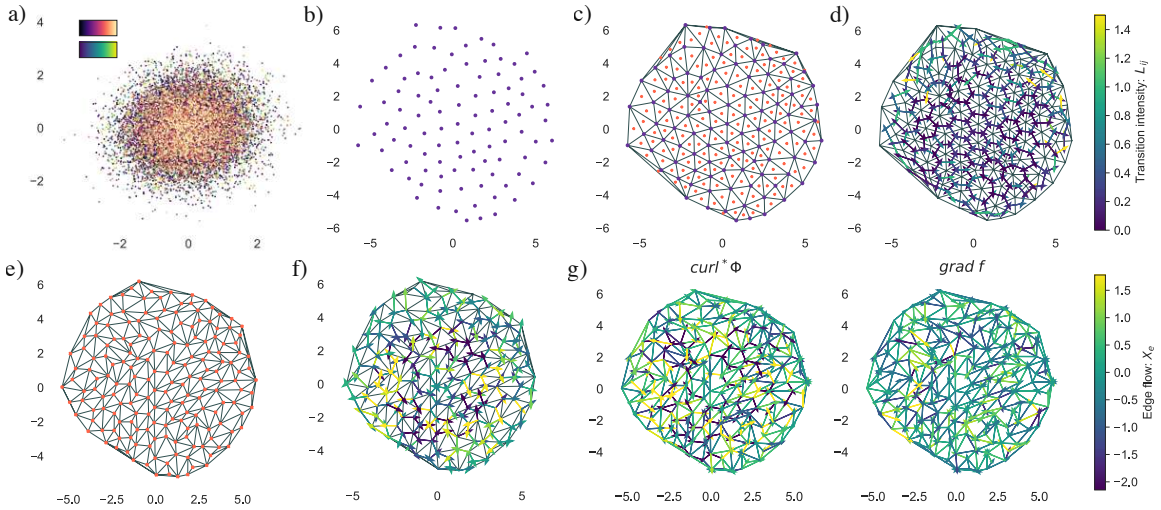

Figure 5.15: **Pipeline for decomposing flows.** *a*) We begin by considering an ensemble of paths as a point-cloud. *b*) We subsample using 'further point subsampling' to obtain a uniform density sample over the cloud. *c*) Using the Delauney triangulation, we create a tessellation. *d*) Treating each triangle as a discrete state, we infer a CTMC from the paths using MLE. *e*) We take the dual tessellation, such that nodes are states and edges represent transitions. *f*) We define an edge-flow with Eq. (5.47). *g*) We perform the GHHD with LSQR to obtain two components.

and edges represent transitions, we must take the *dual complex* of our triangular grid [orange crosses, Panel *e*)] [220].

Step 5. We can then define the edge-flow, X, with Eq. (5.47) [Panel *f*)], and decompose it using the GHHD, to yield the two components [Panel *g*)].[c] In practice, we compute the decomposition with LSQR, as described in Sec. 2.3.2.

[a] It is common to find that data, in the form of point-clouds or paths, live in a low-dimensional manifold embedded in a higher dimensional ambient space [219]. This is known as the 'manifold hypothesis'. These sampling methods are designed to identify this latent structure in an unsupervised fashion. We will return to the manifold hypothesis and its implications for neural network dynamics in Ch. 7.

[b] As in Sec. 5.2.2, we choose the number of subsampled points, which determines our spatial resolution, such that at least 70% of transitions are valid, and thus used in the inference.

[c] The nullity of the 1-Laplacian can be though of as the number of 'holes' in the complex [84]. We construct a triangulation with no holes thus each edge-flow has no harmonic component i.e. $R = \text{curl}^* \phi$ for $\phi \in L^2(T)$.

5.3.3 Numerical experiments

We investigate the performance of our approach with numerical experiments using sampled paths from both solvable and nonlinear diffusions.

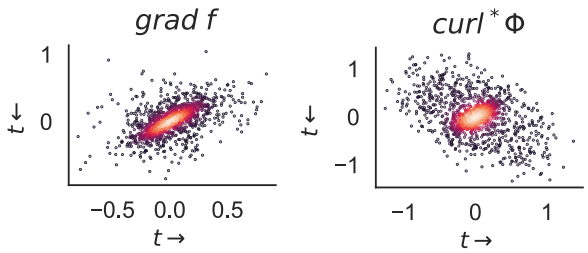

Figure 5.16: **Correlations of edge-flows under time-reversal.** We infer a CTMC and edge-flow from a long path from the OU process ($\theta = 5$, $\sigma = 1$), and then do the same for its time-reversal. We then look at the correlation of the flow along each edge in both of these cases. We find a positive correlation for the gradient component, and a negative correlation for the curl component, supporting that they are approximately even and odd under time-reversal respectively.

Components under time-reversal The irreversible and reversible components of the HHD are odd and even under time-reversal respectively [55]. As a first assessment of the method, we sample a single, long trajectory from the OU process in Ex. 5.1.1 for $\theta = 5$ and $\sigma = 1$. We then infer the CTMC, and decompose the edge-flow for both the original path and its time-reversal. We then look at the correlation between the gradient component in the original and reversed path, along each edge, and the same for the curl component.[16]

We see that the gradient component is positively correlated ($r = 0.495$, ****) and the curl component is negatively correlated ($r = -0.397$, ****) supporting the interpretation that the components are approximately even and odd under time-reversal, reflecting the components of the underlying SDE [Fig. 5.16].

Quantifying irreversibility A natural first application is to measure the degree of irreversibility with the proportion of the edge-flow that is in the divergence-free component ('curl part'),

$$P_c(R, X) = \frac{||R||_2}{||R||_2 + ||X - R||_2}, \tag{5.48}$$

where $||\cdot||_2$ is the L_2 norm of the edge-flow as a vector.

In cases where we can perform the exact HHD of the underlying SDE, we can compute a comparable quantity,

$$P_c(f_{\text{irr}}, f) = \langle \frac{||f_{\text{irr}}(x)||_2}{||f_{\text{irr}}(x)||_2 + ||f(x) - f_{\text{irr}}(x)||_2} \rangle, \tag{5.49}$$

[16] We note that the representation of paths as a point-cloud is invariant under time-reversal, thus the grid that we use is the same in both cases. The only thing that changes is the pattern of transitions in the discretised path.

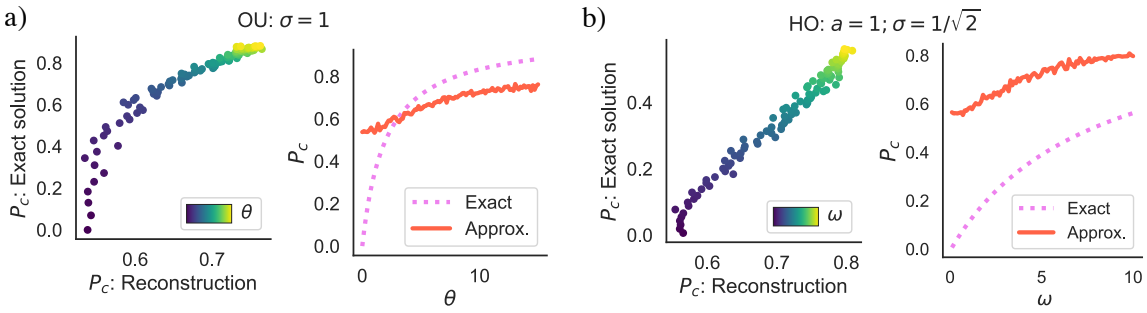

Figure 5.17: P_c **for the discrete approximation of solvable processes.** We apply our pipeline to ensembles of paths from the *a*) OU and *b*) HO processes for increasing values of the irreversibility parameters, θ and ω. We see that P_c in the discrete approximation increases with these parameters, as it does in the analytical solution. This shows that the pipeline is correctly capturing the increasing irreversibility. There is numerical error between the underlying decomposition and the reconstructed components.

where $\langle \cdot \rangle$ represents the average over a discrete grid of points. We can then directly compare the exact solution to the discrete approximation.

Solvable examples We begin by considering solvable examples i.e. situations where the NESS and HHD admit an analytical solution. In particular, we will consider Ex. 5.1.2, the OU process with $\sigma = 1$, and Ex. 5.1.3, the HO with $a = 1$ and $\sigma = 1/\sqrt{2}$ i.e. a noisy planar limit-cycle oscillator.

In both cases, the irreversible component is a linear component given by a skew-symmetric part whose strength is given by $\theta \in [0, 15]$ (OU) and $\omega \in [0, 10]$ (HO) respectively. We sweep this parameter through the ranges, sampling an ensemble of 500 paths, each with length 10,000. With this ensemble, we apply our method and compute P_c for both the discrete approximation, and the analytical HHD.

Fig. 5.17 shows P_c in the discrete approximation and the exact HHD, computed on a fine grid, for the OU [Panel *a*)] and the HO [Panel *b*)]. We see that, as in the exact HHD, P_c in the discrete approximation increases as a function of the irreversibility parameter for both processes. This implies that the decomposition approach is able to capture the level of irreversibility in the process. Due to both finite data, the discrete grid, and possible discrepancies in the convergence of the decompositions, there is a significant degree of imprecision in P_c between the discrete and analytical decompositions. In these and subsequent numerical experiments, and in the data analysis, we use 250 subsampled points to construct the tessellation.

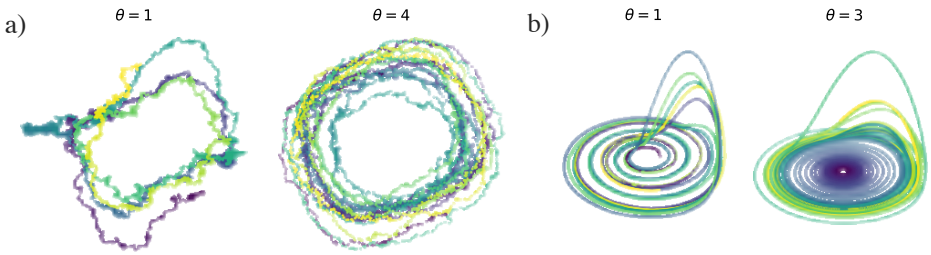

Figure 5.18: **Example paths from processes with rotation** θ. Sampled paths from *a*) the VDP for $\theta = 1$ and $\theta = 4$, and *b*) the Rössler attractor for $\theta = 1$ and $\theta = 3$.

Unsolvable processes Next, we turn to a pair of processes where there is no analytical form for the NESS or the HHD.

Example 5.3.1 (The stochastic, modified van der Pol oscillator)**.** The *van der Pol oscillator (VDP)* is a nonlinear, limit-cycle oscillator with damping [221]. We consider a modified version that includes both isotropic noise, and an additional rotational parameter,

$$\begin{aligned} dx(t) &= [\theta y]\, dt + \sigma dw_x(t), \\ dy(t) &= [-\theta x + \mu y(1 - x^2)]\, dt + \sigma dw_y(t). \end{aligned} \tag{5.50}$$

Whilst there is no analytical form for the NESS, we assume that θ determines the irreversibility of the process, as it drives irreversible rotation in phase-space, in a manner analogous to the OU and HO processes. In the typical VDP, $\theta = 1$. The parameter μ controls the damping, and we set it to 1. Panel *a*) of Fig. 5.18 shows example paths from this process for $\theta = 1$ and $\theta = 4$. We sample from this process using a splitting method (see App. C.4.1).

Example 5.3.2 (The stochastic, modified Rössler attractor)**.** The Rössler attractor is a three-dimensional system that has a chaotic attractor. We consider a modified version with isotropic noise, and an additional rotational parameter,

$$\begin{aligned} dx(t) &= [-\theta y - z]\, dt + \sigma dw_x(t), \\ dy(t) &= [\theta x + a y]\, dt + dw_y(t), \\ dz(t) &= [b + z(x - c)]\, dt + dw_z(t). \end{aligned} \tag{5.51}$$

In the absence of noise, for $\theta = 1$, $a = 0.2$, $b = 0.2$, and $c = 5.7$, the system has a characteristic chaotic attractor. The parameter θ drives rotation in the $xy-$plane, and thus serves as a proxy for the irreversibility of the process. Panel *b*) of Fig. 5.18

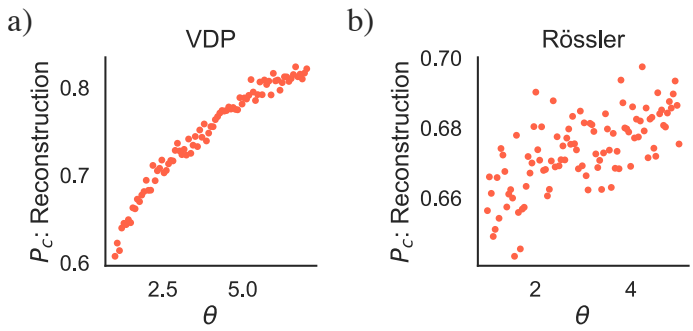

Figure 5.19: P_c **for the discrete approximation of unsolvable processes.** We apply our pipeline to ensembles of paths from the *a*) VDP and *b*) Rössler processes for increasing values of the irreversibility proxy, θ. We see that P_c in the discrete approximation increases with θ, again implying that the pipeline is capturing the increasing irreversibility.

shows example paths sampled for $\theta = 1$ and $\theta = 3$. We sample from this process using the tamed Euler-Maruyama method [222].

For these processes, we again sweep the parameter $\theta \in [1, 7]$ (VDP) and $\theta \in [1, 5]$ (Rössler), sampling 500 paths of length 10,000. For the VDP we take $\sigma = 1/\sqrt{2}$, whilst for the Rössler system we take $\sigma = 1/\sqrt{20}$.[17]

Fig. 5.19 show P_c as a function of θ for both processes. We find a strong position correlation in both cases (VDP: $r = 0.969$, ****; Rössler: $r = 0.677$, ****), which implies that the irreversibility measured by our method increases with θ as expected. The results for the Rössler system show that the approach can be scaled to three-dimensions, except that the larger cells in the tessellation will lead to larger errors, as seen in the noisier relationship between θ and P_c.

5.3.4 Applications to physiological dynamics

Now that we have shown that our approach is able to measure the level of irreversibility of an NESS from an ensemble of paths, we can apply it to real-world data. In particular, we will focus on flickering experiments from red blood cells (RBCs), differentiating between healthy and 'Adenosine triphosphate (ATP)'-depleted[18] cells (which we call 'passive'), and healthy and arrhythmic heartbeats captured by an electrocardiogram (ECG). In both cases, previous studies have shown higher levels of

[17] We take a low noise-intensity for the Rössler system as the chaotic attractor is fragile under noise. Increasing the noise too far will prevent the attractor from emerging, and thus we cannot argue that the process is stationary.

[18] ATP is a molecule used to fuel energetic expenditure in cells.

irreversibility or EPR in the healthy dynamics compared to the pathological case i.e. lower irreversibility in passive RBCs [180, 223] and arrhythmic ECGs [181, 182].

Flickering experiments in red-blood cells The dissipation of energy in RBCs results in active (irreversible) fluctuations of their inner and outer membranes – so-called 'flickering' [180, 223, 224]. In order to differentiate active fluctuations from those which are purely thermal (and reversible), healthy RBCs can be compared with those whose ATP is depleted (passive) (for details on the experimental set-up see Ref. [224]). Using optical microscopy, the fluctuations of the outer membrane can be recorded and their dynamics can be analysed using the formalism of SDEs.

This experimental procedure results in a univariate time-series describing the displacement of a single (fixed) point on the outer cell-membrane. Despite the fact that the data is one-dimensional, we assume that its dynamics are coupled to an inner membrane whose displacement is not observed.[19] We consider the dynamics of five healthy RBCs, compared to a single passive RBC. For each cell we have 512 paths, each with 10,000 time-steps. The dataset is available at Ref. [225].

Healthy and arrhythmic heartbeats The electrical signals governing the dynamics of the human heart can be captured with an ECG. The nonequilibrium nature of cardiac dynamics has been closely linked to healthy biological function, with lower irreversibility being associated with ageing and disease [181, 182]. We consider data from the 'PhysioNet' database and compare healthy heartbeats to those from patients with a *myocardial infarction* (a type of arrhythmia) [226].

Each ECG is a univariate time-series, however the limit-cycle nature of cardiac dynamics imply a second unobserved phase-space variable. The dataset contains 4046 examples of healthy heartbeats, and 10,506 abnormal examples, from which we randomly sample a subset of 4046. The signals are taken with a frequency of 125 Hz resulting in 188 time-steps each.[20]

5.3.4.1 Pipeline modifications for real-world data

Before we can apply our pipeline to the real-world paths, we have to introduce a number of additional considerations.

[19] This dataset has previously been used to validate techniques for measuring the EPR [180].
[20] The dataset is available at Ref. [227] with further details on preprocessing in Ref. [228]

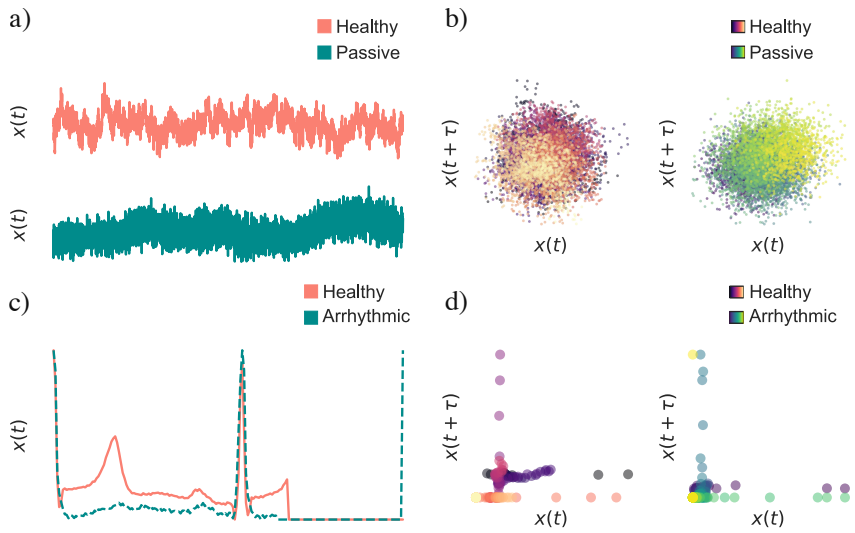

Figure 5.20: **Time-delay embedding for phase-space reconstruction.** Example paths of healthy and pathological dynamics, and phase-space reconstruction with a time-delay embedding for $a - b$) RBCs, and $c - d$) human heartbeats.

Time-delay embedding for phase-space reconstruction In both datasets, paths take the form of a univariate time-series, which we consider to be the observation of a single coordinate from a planar SDE in an NESS. However, our approach does not apply in a single dimension, and we note that univariate diffusions are always reversible in the absence of additional boundary conditions (see Corr. 2.2.2).

Instead, following Takens' embedding theorem [229], we employ the common procedure of reconstructing unobserved variables using a *time-delay embedding* [230]. More explicitly, we consider the two-dimensional system,

$$y(t) = (x(t), x(t + \tau)), \tag{5.52}$$

where $\tau > 0$ is the time-lag parameter. Following convention, we choose τ as the first zero-crossing of the autocorrelation function. Fig. 5.20 shows example paths from both datasets and a phase-space reconstruction with a time-delay embedding.[21]

Correcting for non-isotropic diffusion Our method assumes isotropic diffusion so that the discrete and continuous HHD take the same form. However, in real-world data this assumption is not always valid. Nevertheless we note that a process $x(t)$ with spatially-constant, yet non-isotropic, diffusion, given by matrix D, can be

[21]Not reported in this thesis, we conducted ablation experiments, varying the number of subsampled points, the number of paths in the ensemble, the length of paths, and the time-delay parameter using both sampled paths and real-world data. The results did not change qualitatively or in significance.

transformed into a process $y(t) = D^{-1/2}x(t)$, where $y(t)$ has isotropic diffusion. In order to infer the diffusion from the path data, we use a MLE of the diffusion tensor.

More specifically, given a tessellation of phase-space into triangular bins, which we obtain with our pipeline (Fig. 5.15), the diffusion tensor can be estimated with,

$$\widehat{D}(\Delta_{\bar{x}}) = \frac{1}{N(\Delta_{\bar{x}})} \sum_{i,r\,:\,x_{t_i,r}\in\Delta_{\bar{x}}} \frac{[x_{t_{i+1},r} - x_{t_i,r}] \otimes [x_{t_{i+1},r} - x_{t_i,r}]}{2\Delta t}, \tag{5.53}$$

where $N(\Delta_{\bar{x}})$ represents the number of points in each bin $\Delta_{\bar{x}}$ identified by its mid-point $\bar{x}$, $\{x_{t_1,r}, \ldots, x_{t_T,r}\}_{r=1}^{Q}$ is the ensemble of Q paths, $\otimes$ is the outer product, and Δt is the time-step between observations [119]. Whilst this is a spatially-dependent estimator, we take the mean value over the triangles, $D = \langle D(\Delta)\rangle$, to be the diffusion matrix.

Remark 5.3.1 (Spatial dependence in diffusion)**.** In this chapter, we first assumed diagonal diffusion (when using square grids) and now constant diffusion (using data-driven grids), as generalised diffusion poses a major mathematical challenge. In some physical systems, such as the dynamics of soft-matter in an inhomogeneous medium, spatial-dependence in the diffusion tensor is a central feature [231]. In other real-world systems, diffusion can represent measurement error, where it commonly considered to be state-independent and temporally uncorrelated. In App. C.5, we briefly investigate the validity of this assumption by assessing if there is spatially-dependent diffusion in the real-world data considered in this section.

Bootstrapping paths for significance testing Our method returns a single value P_c from an ensemble of paths. As a result, despite having two datasets that include a large number of paths, this does not allow for significance testing between distributions. To combat this, we perform *ensemble bootstrap sampling*, an approach which is common in such situations [3, 174]. We construct bootstrapped ensembles by sampling paths from the original ensemble with replacement. Each bootstrapped ensemble has the same number of paths as the original ensemble. We repeat this 100 times, which gives a distribution over the value of P_c.

Results Fig. 5.21 shows the distributions of P_c using bootstrapped ensembles for the healthy and passive RBCs [Panel *a*)], and the healthy and arrhythmic heartbeats. We find that P_c is significantly higher for the healthy cells compared to the passive cell ($p < 0.0001$, independent t-test). We also find that P_c is significantly higher in healthy heartbeats compared to arrhythmic ones ($p < 0.0001$, independent t-test).

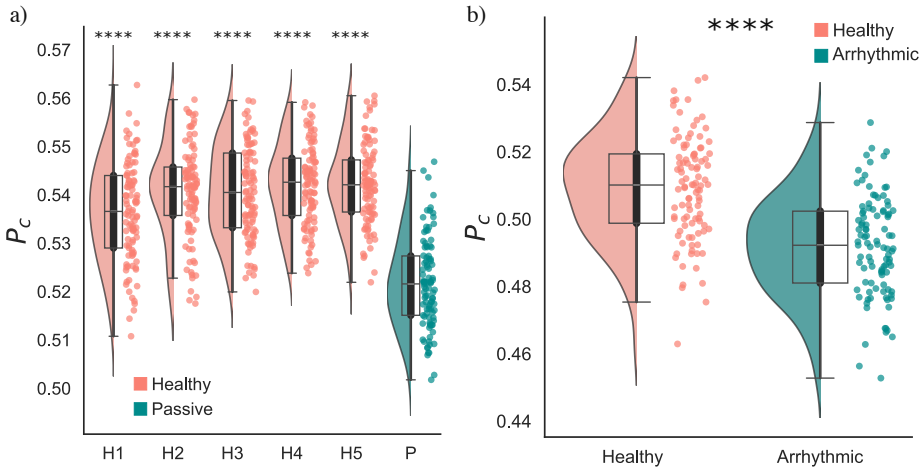

Figure 5.21: P_c **for RBCs and heartbeat data.** We measure P_c for a) the healthy and passive RBCs using bootstrapped ensembles, and for b) the healthy and arrhythmic heartbeats. We find that the healthy RBCs have higher irreversibility than the passive cells (****, independent $t-$test between each healthy cell and the passive cell), and the healthy heartbeats have higher irreversibility than the arrhythmic ones (****, independent $t-$test).

These results support previous findings that irreversible dynamics are a fundamental feature of healthy physiological dynamics, that are impaired in pathological conditions [180–182].

5.4 Conclusions

In this chapter, we have focused on the approximation of SDEs in the NESS with discrete-state stochastic processes. In particular, our results can be divided into two parts. In the first, we derive an effective CTMC from the drift and diffusion fields of an SDE using the SG FVA on a rectangular grid. We focus on the validity of this approximation, showing convergence of the EPR in the limit and confirming this with numerical experiments. Additionally, we consider the statistical inference of a CTMC from the coarse-graining of an SDE with discrete states, linking this to the classical embedding problem for Markov chains. Our numerical results show that this approach yields a poor estimate of the true EPR of the underlying SDE, but that it can be used to perform hypothesis testing for SDE path data. We illustrate this with paths from schooling fish, which we confirm are not from an NESS.

The second half of the chapter focuses on the relationship between the discrete and continuous forms of the HHD. In particular, we develop a pipeline for approximating an SDE with a CTMC, and then apply the discrete HHD in an effort to recover the

components of the continuous HHD. Whilst we cannot assure convergence, we use numerical experiments to show that the relative size of the irreversible component in the discrete approximation aligns with the corresponding component in the underlying drift. Finally, we apply this approach to path data from flickering RBCs and healthy and arrhythmic heartbeats, where we find higher irreversibility in healthy dynamics when compared to pathological cases. This supports previous findings for both RBCs [180] and cardiac dynamics [181, 182], and demonstrates the usefulness of the approach.

These results investigating the consistency of using discrete approximations to approximate continuous data are highly relevant, due to the plethora of studies that use this approach on biophysical data [3, 174–176]. Although we derive a convergent discrete-state approximation, our statistical analyses shows that the estimate of the EPR is a coarse lower-bound, even when using long and finely sampled paths. Furthermore, we discuss various approaches for the inference of these processes, methods to test for significance, and other practical considerations.

Limitations and future directions As we showed, using discrete-states to approximate SDEs can be coarse and imprecise. Alternative approaches that have shown more promise include sparse regression onto bases of functions [119], where the exact terms in the inferred drift and diffusion can be recovered. Whilst more involved, this family of approaches appears to yield improved estimators of the EPR than discretised approaches. Moreover, they have been extended to underdamped SDEs (e.g. the hypoelliptic case) [232], which is often used for modelling collective movement, whereas our analyses focuses on overdamped systems.

Another key assumption of our analysis is Markovianity. As mentioned in Sec. 5.1, coarse-graining an SDE with discrete states leads to a non-Markovian process, which we are approximating as Markovian. This leads to errors in the EPR estimate [45] and can lead to artefactual scaling relationships [179]. Nevertheless, the inference of higher-order or semi-Markov chains is either significantly more challenging or requires a lot more data.[22]

We note that the convergence between the discrete and continuous forms of the HHD is still not well understood. The few results that have been presented are found in Ref. [64], but these are restricted to square grids and diagonal diffusion, as are many of our results and methods in this chapter. Finally, we focused on P_c, the relative

[22] A simple approach for inferring a higher-order Markov model is simply to infer a Markovian model over the set of possible 2-step transitions [189]. However, this dramatically increases the state-space, necessitating much longer paths for estimation.

size of the irreversible component, as a simple first measurement. A number of more sophisticated approaches have been developed for the visualisation and interpretation of each component in the discrete HHD, including deriving node/state rankings [90], visualising the rotational component as circular trade-offs [233], and using the components to perform embeddings [234], which could also find future applications in the analysis of stochastic processes.

This approach is notably more mathematically complex than our approach for identifying irreversibility in Ch. 4. Nevertheless, in this set up we gain the additional insight of a stochastic model of the process, albeit discrete, which can be analysed and sampled from. In the case of the visibility graph, one obtains only a measure of irreversibility with no insight into the dynamics of the process. Both methods become computationally restrictive for high-dimensional data which, as discussed, is a natural restriction for focussing on irreversibility in a 'model-free' framework.

Chapter 6

Vortex-like dynamics in brain network oscillations

Networks of coupled oscillators have become useful mathematical models for investigating collective phenomena in complex systems, including synchronisation, the formation of waves, and the transition to turbulence [66, 235]. The study of neural dynamics is a prominent example, with coupled oscillators being used to model and analyse large-scale brain network dynamics [23, 24, 236], and where the emergent spatiotemporal patterns have been labelled as 'metastable' [237], 'turbulent' [238], or 'chimera-states' [239]. Moreover, prominent theories of neural computation focus on the coherence and incoherence of brain network oscillations across scales [240–242].

One collective phenomenon that has long been studied in statistical mechanics is the formation of 'spiral-like waves' around 'topological defects', which have found applications in condensed matter theory, material science, and quantum mechanics [243]. Most famously, these patterns – which we will refer to as *vortices* – emerge in the two-dimensional XY model [243, 244]. Recently, there have been attempts to identify the presence of vortices in brain dynamics, specifically cortical oscillations, and provide evidence for their existence as well as their possible functional mechanisms [245, 246]. These approaches use the NFT framework [25] which models brain activity as a field over a continuous cortical geometry, in contrast with network neuroscience, where brain activity is modelled as signals evolving over a structural network of discrete regions (see Sec. 1.2.2) [19, 23].

The classical XY model considers *spins*, also known as rotors or clocks organised into a two-dimensional lattice. As such, vortices can be identified by comparing spins around each elementary *plaquette* i.e. each little square in the lattice [247]. The two-dimensional lattice can be considered to be an approximation of a plane. Thus, this approach could be easily translated to the study of neural data, provided we are

working in the NFT framework and we are modelling the cortex as a rectangular grid approximating a plane, and neural dynamics as coupled oscillators [246]. However, the regular lattice is not a good representation of the cortex. If we instead model large-scale brain dynamics as phase-oscillators on the connectome, which is an irregular (complex) network, then there is no natural analogue of the square plaquette, and no obvious way to define or identify a vortex.

In this chapter, we tackle this problem by presenting an approach that is able to identify vortex-like patterns in complex networks. In particular, using discrete Hodge theory, introduced in Sec. 2.3, we provide an efficient method that identifies the location and strength of vortices, and that is general to the structure of the underlying interactions. From this, we are able to define a measure of *vorticity* for a phase-configuration, which we call the 'Hodge vorticity', as well as a measure of *complexity* for the vortex-like dynamics, using an expansion into structurally-derived modes.

We validate our approach with numerical experiments using the XY model on a lattice, showing that Hodge vorticity mirrors the canonical measure. We also study the XY model on a range of structural brain networks under increasing temperature, showing evidence for a transition point that is similar to the behaviour on the lattice. Next, we apply our method to fMRI recorded in human participants during either rest or whilst performing one of seven cognitive tasks, obtained as part of the HCP [125, 126]. Our results show higher Hodge vorticity in task-activated dynamics compared to the resting state. Finally, using the eigenmodes of the 1-Laplacian, we represent the brain network activity as an 'edge-signal', expand it in a basis of structurally-derived patterns, and introduce a measure of complexity for this representation. This expansion directly extends previous approaches which derive these modes from the graph Laplacian [85] or the Laplace-Beltrami operator [27]. We find that the complexity of this representation is higher in tasks compared to the resting state.

6.1 Vortices and the XY model

To begin with, we introduce the XY model, which is the prototypical system displaying vortices. We consider a system with N spins, where each spin has a state $\mathbf{s}_i = (\cos(\theta_i), \sin(\theta_i))$ with $\theta_i \in [0, 2\pi)$ for $i = 1, \ldots, N$. The angle θ_i is known as the *phase*. Together, these states form a *configuration*, $\{\mathbf{s}\} = (\mathbf{s}_1, \ldots, \mathbf{s}_N)$. In its most common form, the XY model is a static equilibrium model, and thus defined by a

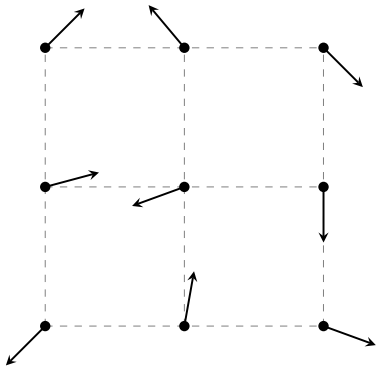

Figure 6.1: XY **on the lattice.** We will first consider configurations of the XY model on the regular square lattice, as shown here for $L = 3$.

Hamiltonian which gives the energy of each configuration.[1] In the case of the XY model with no external fields, it is given by,

$$H(\{\mathbf{s}\}) = -\sum_{i,j} J_{ij} \cos(\theta_i - \theta_j)\,. \tag{6.1}$$

Here J_{ij} represents the connectivity between oscillators i and j, which we assume to be symmetric $J_{ij} = J_{ji}$.

As discussed in Sec. 2.2.1, such Hamiltonian systems correspond to an ESS with a Boltzmann distribution,

$$P(\{\mathbf{s}\}) = \frac{e^{-\beta H(\{\mathbf{s}\})}}{Z}, \tag{6.2}$$

where $\beta \geq 0$ is the inverse of the thermodynamic temperature, and Z is a normalising factor [52]. Unlike much of the rest of this thesis, in this chapter we focus on dynamics in the ESS, which are reversible.

This formulation allows J_{ij} to represent a general coupling structure, such as the adjacency or weight matrix of an undirected network. We will begin with the case where J_{ij} encodes nearest-neighbour interactions in a two-dimensional square lattice with $N = L \times L$ sites without periodic boundaries, illustrated for $L = 3$ in Fig 6.1.

[1] By adding a kinetic energy term, one can define an alternative XY model as a Hamiltonian dynamical system [248]. This model is closely related to the Kuramoto model of Ex. (2.2.3), and is known as the 'inertial' or 'second-order' Kuramoto model [249].

6.1.1 Sampling from the XY model

In order to study the behaviour of the XY model numerically, we draw samples from its stationary distribution using MCM. In particular, we use the *Metropolis-Hastings (MH)* algorithm, which goes as follows [52].

We begin with a fixed inverse temperature, β. We then:

1. Pick a random site i, which has phase θ_i.
2. Propose a new phase, θ_i',
$$\theta_i' = \theta_i + \delta\,,$$
where $\delta \sim U(-\Delta, \Delta)$.
3. We compute the energy difference,
$$\Delta E = H(\{\mathbf{s}'\}) - H(\{\mathbf{s}\})\,,$$
where $\{\mathbf{s}'\}$ is the configuration that uses the proposed phase θ_i' for site i.
4. We accept the proposed phase change for site i with probability,
$$p = \min\{1, e^{-\beta\Delta E}\}\,.$$

One *sweep* of the MH algorithm is made up of n sequential proposals and generates a single sampled configuration i.e. a 'sample'. For the remainder of this chapter, we take $\Delta = \pi/2$.[2] The MH algorithm generates correlated samples, thus we have a 'thermalisation period' at the beginning of each 'run' i.e. a fixed number of samples before we begin to take measurements, followed by the 'measurement period', where we record global quantities after each sweep. Each run is composed of n_T thermalisation samples and n_M measurement samples.

Fig. 6.2 shows a sampled configuration for a lattice with $L = 25$ at increasing temperatures $\beta = 10, 2$, and 0.5. At low temperature, the spins are largely aligned in a coherent direction. As the temperature increases, they become increasingly disordered.

Thermodynamic measurements As is often the case with statistical mechanics models, we are interested in the averaged *global* properties of a configuration, $\{\mathbf{s}\}$, which we now introduce [52].

[2]The step-size Δ affects the efficiency of the sampler. For large $\Delta \approx \pi$, many proposals will be rejected, thus it is inefficient. For small $\Delta \approx 0$, many proposals will be accepted but the phase only changes slightly. We choose $\Delta = \pi/2$ as a sensible intermediate choice.

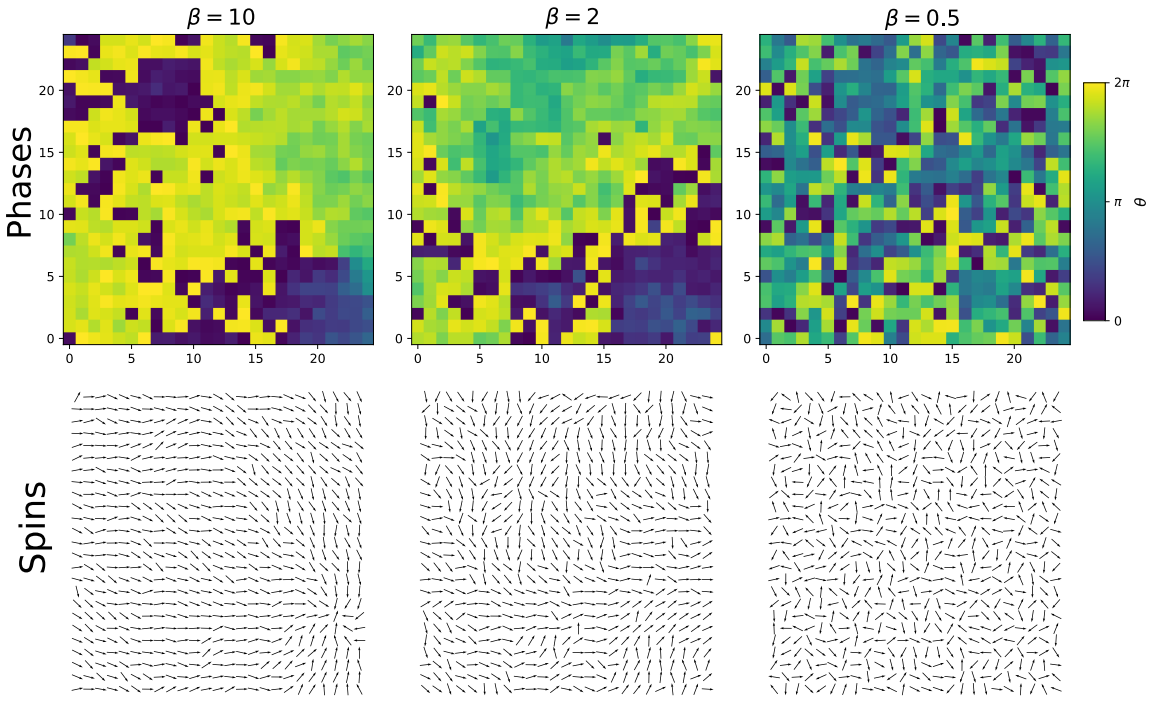

Figure 6.2: XY **configurations at different temperatures.** We sample a configuration from the XY model with $L = 25$ for increasing temperatures $\beta \in \{10, 2, 0.5\}$. At low temperature, the model is close to synchronisation. At high temperature, it is disordered.

The *energy density*, E, is given by,

$$E(\{\mathbf{s}\}) = \frac{H(\{\mathbf{s}\})}{N}\,, \tag{6.3}$$

which we average over the measurement configurations sampled during the measurement period at a given β, i.e. $\langle E \rangle = \frac{1}{n_M} \sum_{l=1}^{n_M} E_l$ where E_l is the energy of the l-th measurement sample.

The *specific heat capacity*, C_v, is given by,

$$C_v = \beta^2 N \left(\langle E^2 \rangle - \langle E \rangle^2 \right)\,. \tag{6.4}$$

The *magnetisation*, m, and the magnetisation vector, $\mathbf{m}$, quantify the degree to which the spins are aligned in a single configuration. The vector is given by,

$$\mathbf{m}(\{\mathbf{s}\}) = \frac{1}{N} \sum_{i=1}^{N} \left(\cos\theta_i, \sin\theta_i \right), \tag{6.5}$$

and the magnetisation is given by its magnitude,

$$m = ||\mathbf{m}|| = \sqrt{\left(\frac{1}{N} \sum_{i=1}^{N} \cos\theta_i \right)^2 + \left(\frac{1}{N} \sum_{i=1}^{N} \sin\theta_i \right)^2}\,, \tag{6.6}$$

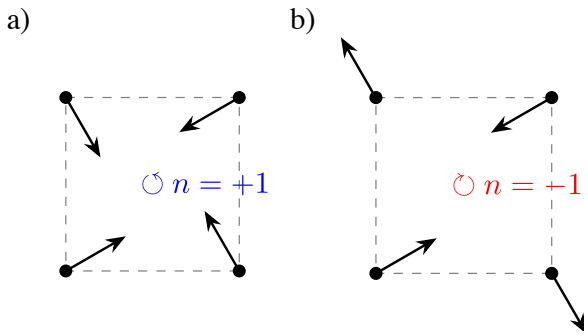

Figure 6.3: **Vortices and anti-vortices.** *a*) An example configuration forming a *vortex*. As you traverse the plaquette counter-clockwise, the spin goes through a rotation counter-clockwise. *b*) An example configuration forming an *anti-vortex*. As you traverse the plaquette counter-clockwise, the spin goes through a rotation clockwise.

which is averaged to obtain $\langle m \rangle$.

Finally, we have the *susceptibility*, χ, given by,

$$\chi = \beta N \left(\langle m^2 \rangle - \langle m \rangle^2 \right). \tag{6.7}$$

6.1.2 Vortices on a lattice: formation and identification

To define a vortex, we begin by considering the continuum limit of the XY model. Spin configurations are defined by an angular field $\theta : \Omega \subseteq \mathbb{R}^2 \to [0, 2\pi)$. A point $x \in \Omega$ is called a *vortex of charge* n if, for any sufficiently small closed curve $\Gamma \subset \Omega \backslash \{x\}$ which encircles x once,[3]

$$\oint_\Gamma \nabla\theta \, d\mathbf{l} = 2\pi n \,, \tag{6.8}$$

where $n \in \mathbb{Z}$ is called the *winding number* [243].[4] In the XY model, stable vortices of charge $n = \pm 1$ emerge, which we call a *vortex* (+1) and an *anti-vortex* (-1), respectively.

In the lattice, a(n) (anti-)vortex can form at the centre of any single elementary plaquette, if the spins on the boundaries of the plaquette (i.e. the four corners of the plaquette in a square lattice) have a discrete winding number of ± 1, which we define

[3] More specifically, the curve, Γ, has *winding number* one around x.

[4] This integral is *quantised* i.e. can only take integer values of 2π, because it is a map from the unit circle, S^1, (i.e. $\Gamma \cong S^1$) to itself (i.e. the spins take values in S^1). As we follow a parametrisation around Γ, the spin is also rotating around S^1. At the end point, which coincides with the start point, the spin must agree with its starting phase, thus it must make an integer number of revolutions, in either direction, along the curve.

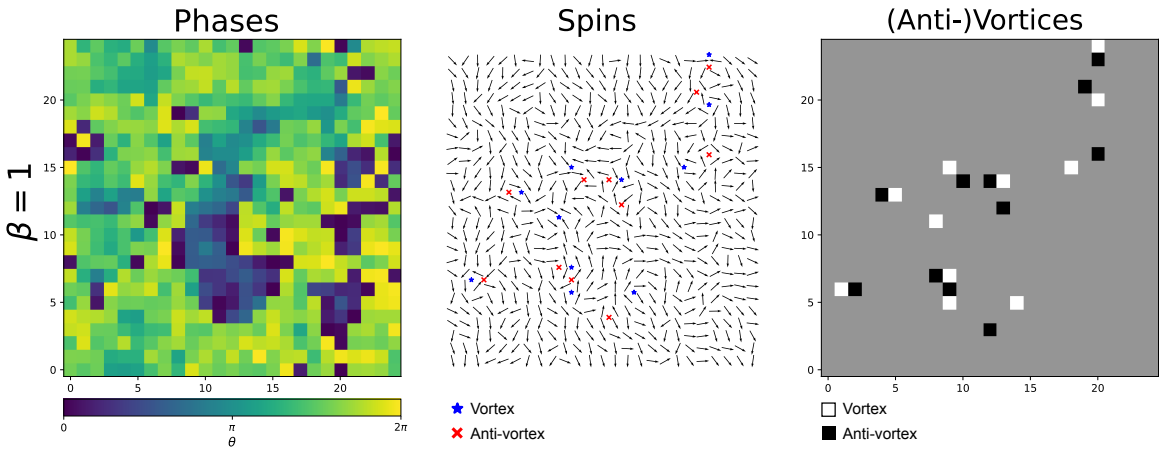

Figure 6.4: **Vortices on the lattice**. Using the method detailed in Sec. 6.1.2, we can identify vortices and anti-vortices in a configuration of the XY model on a lattice.

as follows. We first define the function wrap : $\mathbb{R} \to (-\pi, \pi]$,

$$\text{wrap}(\alpha) = \alpha - 2\pi \cdot \text{round}\left(\frac{\alpha}{2\pi}\right), \tag{6.9}$$

where round$(\cdot)$ denotes rounding to the nearest integer.[5] With this function, we can compute the difference between two phases, along either dimension,

$$\Delta_x \theta_{i,j} = \text{wrap}\left(\theta_{i+1,j} - \theta_{i,j}\right), \qquad \Delta_y \theta_{i,j} = \text{wrap}\left(\theta_{i,j+1} - \theta_{i,j}\right), \tag{6.10}$$

which, by the definition of wrap$(\cdot)$, takes the minimal angular difference. We then sum the angle differences counter-clockwise about each plaquette,[6]

$$W_{i,j} = (\Delta_x \theta_{i,j} - \Delta_x \theta_{i,j+1}) + (\Delta_y \theta_{i+1,j} - \Delta_y \theta_{i,j}). \tag{6.11}$$

Finally, we can extract the winding number,[7]

$$n_{i,j} = \text{round}\left(\frac{W_{i,j}}{2\pi}\right). \tag{6.12}$$

As before, if $n_{i,j} = \pm 1$ we identify it as a vortex (+1) or anti-vortex (-1) [247]. Fig. 6.3 shows an example of a vortex [Panel a)] and an anti-vortex [Panel b)] respectively on a square plaquette. Stable vortices and anti-vortices emerge in the XY model on the lattice, as shown in Fig. 6.4 for a configuration sampled at $\beta = 1$.

[5] For any integer n, we define round$(n + 1/2) := n$.

[6] It is a standard convention that we measure the phase differences when traversing the plaquette counter-clockwise. If the spin also rotates counter-clockwise, then it will have a positive winding number. If it rotates clockwise, then it will have a negative winding number.

[7] $W_{i,j} \in 2\pi\mathbb{Z}$ by definition. In practice, we round to account for numerical errors.

We can measure the *vorticity* of a lattice configuration with,

$$V = \sum_{i,j} \frac{|n_{i,j}|}{N}, \tag{6.13}$$

where $V \in [0, 1]$ is the proportion of sites where a vortex or anti-vortex has formed.[8]

6.2 Normalised Hodge Laplacians

Our framework for measuring vorticity on complex networks will employ the discrete Hodge theory introduced in Sec. 2.3. Before introducing how this theory can be applied to detect vortices, we must first introduce *normalised* Hodge Laplacians and their eigendecompositions.

First, as mentioned in Ex. 2.3.2, we note that when using the 'combinatorial' form of the 0-Laplacian, its eigendecomposition often suffers from localisation, meaning eigenvectors are localised in only a few high-degree nodes [86], and normalised variants must be introduced, which take into account node degrees. It is reasonable to expect that a similar phenomenon occurs in the 1-Laplacian for high-degree nodes, edges, or triangles, which we discuss in further detail in Sec. 6.2.2.1. In order to circumvent this problem, we introduce a normalised form of 1-Laplacian, taking inspiration from Schaub et al. [234].

We begin with an oriented graph $G = (V, E)$ with N nodes, M edges, and an induced set of Θ triangles, which we denote by T. Similarly to the case of the 0-Laplacian in Ex. 2.3.2, we weight the inner products for functions on nodes, edges, and triangles, by the degree of the simplices,

$$\langle x, y \rangle = x^\top D_i y, \tag{6.14}$$

where x and y are $(i-1)$ cochains, and D_i is the diagonal degree matrix for $(i-1)$ simplices, for $i = 0, 1, 2$. Following Schaub et al. [234], we define these degree matrices as follows. First, $D_2 \in \mathbb{R}^{M \times M}$ is the diagonal matrix,

$$(D_2)_{(i,j),(i,j)} = \max\left(\deg(i, j), 1\right), \tag{6.15}$$

where $\deg(i, j)$ is the number of triangles containing the edge (i, j). The matrix $D_1 = 2 \cdot \mathrm{diag}(|B_1| D_2 \mathbf{1})$ where $\mathbf{1}$ is a vector of M ones i.e. this is a diagonal matrix of

[8] At the boundary, we use a periodic boundary condition (PBC). It is worth noting that we do not use a PBC when coupling spins on the lattice and generating samples. This is because a PBC can result in *global winding* around the torus – this winding is measured by the Hodge vorticity, that we have yet to introduce, and not by the canonical measure of vorticity, and thus can create disparities.

weighted node degrees, with the weight of an edge being equal to the maximum of 1 and the number of triangles that include the edge. Finally, we have $D_3 = \frac{1}{3}I$.

Remark 6.2.1. There are different notions of degree for simplicial complexes. We choose to follow Schaub et al., as they specifically justify their choice by considering a RW on a simplicial complex and then derived the associated, normalised 1-Laplacian [234]. This analysis draws a clear analogy with the derivation for the 0-Laplacian, which can be obtained by considering a RW on a graph with equal waiting-times (see Ex. 2.2.4).

In order to write the normalised 1-Laplacian in matrix form, we will need to compute the adjoint of $(B_i^\top)^*$,

$$(B_i^\top)^* = D_i^{-1} B_i D_{i+1}, \tag{6.16}$$

for $i = 1, 2$. These can then be substituted into the equation for the 1-Laplacian,

$$\widehat{L}_1 = B_1^\top (B_1^\top)^* + (B_2^\top)^* B_2^\top, \tag{6.17}$$

$$= B_1^\top D_1^{-1} B_1 D_2 + D_2^{-1} B_2 D_3 B_2^\top. \tag{6.18}$$

We notice that grad $\cong B_1^\top$ and curl $\cong B_2^\top$ remain unchanged from the standard inner product (see Sec. 2.3.1), as they do not depend on the inner product, whilst we now have that $-\text{div} \cong D_1^{-1} B_1 D_2$ and $\text{curl}^* \cong D_2^{-1} B_2 D_3$. Finally, we note that this inner product induces a set of norms on cochains, which we define as $||\cdot||_V$, $||\cdot||_E$, and $||\cdot||_T$.

6.2.1 The normalised Hodge decomposition

In order to compute the GHHD associated with this normalised Laplacian, we represent an edge-flow by a vector in $\mathbb{R}^M$, and the operators by real matrices. However, if we represent functions on simplices using standard bases, then the Euclidean spaces do not have the standard L^2 inner product structure. This is because the bases are not *orthonormal*, despite being orthogonal. Instead, we express functions in the orthonormal bases,

$$\widetilde{f}_v = \frac{f_v}{\sqrt{(D_1)_{v,v}}}, \qquad \widetilde{X}_e = \frac{X_e}{\sqrt{(D_2)_{e,e}}}, \qquad \widetilde{\phi}_t = \frac{\phi_t}{\sqrt{(D_3)_{t,t}}},$$

where $\{f_v\}$, $\{X_e\}$, and $\{\phi_t\}$ are the standard bases i.e. each simplex is represented by a vector of zeros with a one in the corresponding index. In this orthonormal basis,

one can check that grad $\cong D_2^{1/2} B_1^\top D_1^{-1/2}$ and curl $\cong D_3^{1/2} B_2^\top D_2^{-1/2}$ respectively, and $H(E)$ identifies with $\mathbb{R}^M$ as a Hilbert space. Therefore the GHHD given in Eq. (2.61) can be applied to $\mathbb{R}^M$,

$$\mathbb{R}^M = \mathrm{im}(D_2^{1/2} B_1^\top D_1^{-1/2}) \oplus \mathrm{im}(D_2^{-1/2} B_2 D_3^{1/2}) \oplus \ker(\widetilde{L}_1)\,, \tag{6.19}$$

$$= \mathrm{im}(D_2^{1/2} B_1^\top) \oplus \mathrm{im}(D_2^{-1/2} B_2) \oplus \ker(\widetilde{L}_1)\,, \tag{6.20}$$

where $\widetilde{L}_1$ is a symmetrised version of the 1-Laplacian in matrix form,

$$\widetilde{L}_1 = D_2^{1/2} \widehat{L}_1 D_2^{-1/2}\,, \tag{6.21}$$

$$= D_2^{1/2} B_1^\top D_1^{-1} B_1 D_2^{1/2} + D_2^{-1/2} B_2 D_3 B_2^\top D_2^{-1/2}\,, \tag{6.22}$$

which corresponds to transforming $\widehat{L}_1$ from the standard to the orthonormal basis. We perform the normalised Hodge decomposition as follows.

Let $X \in \mathbb{R}^M$ be an edge-flow written in the standard basis.

1. Transform it into the orthonormal basis, $Y = D_2^{1/2} X$.
2. Solve
$$f = \mathrm{argmin}_{g \in \mathbb{R}^M} ||D_2^{1/2} B_1^\top g - Y||_2^2, \tag{6.23}$$
with LSQR, where $||\cdot||_2$ is the standard L^2 norm on $\mathbb{R}^M$.
3. The divergence free component in the standard basis is given by,[a]
$$R = D_2^{-1/2}(Y - D_2^{1/2} B_1^\top f) = X - B_1^\top f \tag{6.24}$$

[a] We will not need to split the divergence-free component. For completeness, it can be decomposed by solving $\phi = \mathrm{argmin}_{\psi \in \mathbb{R}^\Theta} ||D_2^{-1/2} B_2 \psi - Y||^2$, where the harmonic part is the remainder $X_H = X - B_1^\top f - D_2^{-1} B_2 \phi$.

Remark 6.2.2. The matrix $\widetilde{L}_1$ is to $\widehat{L}_1$ what $\widetilde{L}_0$ is to $\widehat{L}_0$ in Ex. 2.3.2 i.e. a symmetrised version of a Laplacian derived from a weighted inner product. Whilst our choice of $\widetilde{L}_1$ coincides with that of Schaub et al. [234], they derive a different $\widehat{L}_1$ by considering a RW on a simplicial complex, rather than from choosing an inner product structure. As we will show in Sec. 6.3, our choice of 1-Laplacian and edge-flow satisfy important properties that are relevant to the detection of vortices. Finally Schaub et al. assume that empirical edge-signals are written in the orthonormal basis [234], whilst we assume they are written in the standard basis, and we then shift coordinates before computing the decomposition.

6.2.2 Eigendecomposition of the normalised 1-Laplacian

In Sec. 6.5, we will consider an expansion of edge-signals into the eigenbasis of $\widetilde{L}_1$, thus it is worth considering the implications of the GHHD on the associated eigendecomposition.

First, we note that the eigenmodes of $\widetilde{L}_1$ can be partitioned into three subsets, the *gradient eigenmodes*, the *curl eigenmodes*, and the *harmonic eigenmodes*. These span $\mathrm{im}(D_2^{1/2}B_1^\top)$, $\mathrm{im}(D_2^{-1/2}B_2)$, and $\ker(\widetilde{L}_1)$, respectively.

The number of gradient modes is $N-\beta_0$, where β_0 is the zeroth Betti number i.e. the number of connected components. This stems from the fact that the gradient eigenmodes can be derived from the eigenmodes of a 0-Laplacian (with the corresponding inner product structure),

$$G_1 = D_1^{-1/2}B_1D_2B_1^\top D_1^{-1/2}. \tag{6.25}$$

More specifically, if v satisfies,

$$G_1 v = \lambda v, \tag{6.26}$$

then $\nu = D_2^{1/2}B_1^\top D_1^{-1/2}v$ is a gradient eigenmode of $\widetilde{L}_1$ with eigenvalue λ (Theorem 3.7 in Ref. [234]). This implies that the gradient eigenmodes are edge-projections of node-level modes. These modes are similar to, but not identical due to the choice of inner product, the typical modes used in previous studies of neural dynamics [85,250].

The number of harmonic modes is equal to the nullity of $\widetilde{L}_1$, which is β_1, the first Betti number i.e. the number of one-dimensional holes in the complex [84]. Each of these modes has eigenvalue $\lambda = 0$. Finally, the number of curl modes is $M-N+\beta_0-\beta_1$, coming to a total of M eigenmodes.

We note that $\widetilde{L}_1$ can be written as $\widetilde{L}_1 = \widetilde{L}_1^l + \widetilde{L}_1^u$, where

$$\widetilde{L}_1^l = D_2^{1/2}B_1^\top D_1^{-1}B_1D_2^{1/2}, \qquad \widetilde{L}_1^u = D_2^{-1/2}B_2D_3B_2^\top D_2^{-1/2}, \tag{6.27}$$

where we refer to $\widetilde{L}_1^l$ and $\widetilde{L}_1^u$ as the *lower* and *upper* Laplacians, respectively. Curl modes are (non-trivial) eigenmodes of $\widetilde{L}_1^u$, whilst they are in the null-space of $\widetilde{L}_1^l$. Gradient modes are (non-trivial) eigenmodes of $\widetilde{L}_1^l$, whilst they are in in the null-space of $\widetilde{L}_1^u$. Harmonic modes are in the null-space of both.

Computing the eigendecomposition In practice, computing the eigendecomposition of an edge-based Laplacian can be challenging for large, dense networks. In order to make this procedure more tractable, we utilise the structure of the GHHD.

First, gradient eigenmodes can be computed via the eigendecomposition of G_1. Assuming N is not prohibitively large (whilst M might be), we perform the eigendecomposition with standard numerical methods. This corresponds to solving for modes defined on the nodes, and then 'lifting' them onto the edges.

Next, we compute harmonic modes by solving $\widetilde{L}_1 v = 0$. Due to numerical errors, eigenmodes with $\lambda < \epsilon$ for some tolerance $\epsilon \approx 10^{-9}$, are considered to be harmonic. A number of numerical methods exist for calculating eigenmodes in the case of sparse, positive-semi-definite matrices, such as locally optimal block preconditioned conjugate gradient method (LOBPCG) [251].

Finally, there are two approaches to computing the curl modes, depending on the density of triangles in the complex. Assuming that $\Theta < M$ i.e. there are fewer triangles than edges, then we can compute the eigendecomposition of the $\Theta \times \Theta$ matrix,

$$G_2 = D_3^{1/2} B_2^\top D_2^{-1} B_2 D_3^{1/2}. \tag{6.28}$$

For every vector v such that $G_2 v = \lambda v$ i.e. eigenmodes of G_2, we have that $\nu = D_2^{-1/2} B_2 D_3^{1/2} v$ is a curl eigenmode of $\widetilde{L}_1$ with eigenvalue λ (Theorem 3.7 in Ref. [234]). Analogous to the case of the gradient modes, this implies that the curl modes can be derived by projecting a mode defined on the triangles, onto the edges.

However, if $\Theta > M$ this does not offer any computational advantage. In this case, we recall that curl modes are non-trivial eigenmodes of $\widetilde{L}_1^u$. As the gradient and harmonic modes of $\widetilde{L}_1$ are eigenmodes with $\lambda = 0$ for $\widetilde{L}_1^u$, we only solve for the non-trivial eigenmodes. LOBPCG allows us to add the constraint that the calculated modes are orthogonal to the set of precalculated gradient and harmonic modes, and thus allows us to compute only the curl modes.[9]

6.2.2.1 Localisation in $\widetilde{L}_1$

A number of studies have used the combinatorial 1-Laplacian, L_1, for signal processing [92], clustering [252], and community detection [253], without considering possible normalisations. On one hand, when performing geometric or topological data analysis, simplicial complexes are often constructed from point-cloud data and thus their

[9] We can confirm that $\widetilde{L}_1^u$ is also positive semi-definite as $\widetilde{L}_1^u = \Lambda\Lambda^\top$ for $\Lambda = D_2^{-1/2} B_2 D_3^{1/2}$, thus we can use LOBPCG.

structure is relatively uniform, which mitigates concerns about localisation. On the other hand, when considering heterogenous complex networks, such as the connectome, L_1 may suffer from eigenmodes that are concentrated around high-degree nodes, edges, and triangles, in analogy with the well-known case of the 0-Laplacian [86].

To illustrate this effect, we compare the eigenmodes of L_1 and $\widetilde{L}_1$. Even though these modes live on the edges, we project them onto the nodes so that they can be visualised more easily. This is motivated by the case of the gradient eigenmodes which are derived from node-level modes, where localisation is known to occur. We begin with gradient modes and consider the combinatorial 0-Laplacian, $L_0 = D - A$, and note that the gradient modes of the combinatorial 1-Laplacian, L_1, are given by,

$$L_1(B_1^\top v) = \lambda(B_1^\top v), \tag{6.29}$$

where $L_0 v = \lambda v$ i.e. we obtain a gradient eigenmode of L_1 by taking the gradient of an eigenmode of L_0.[10]

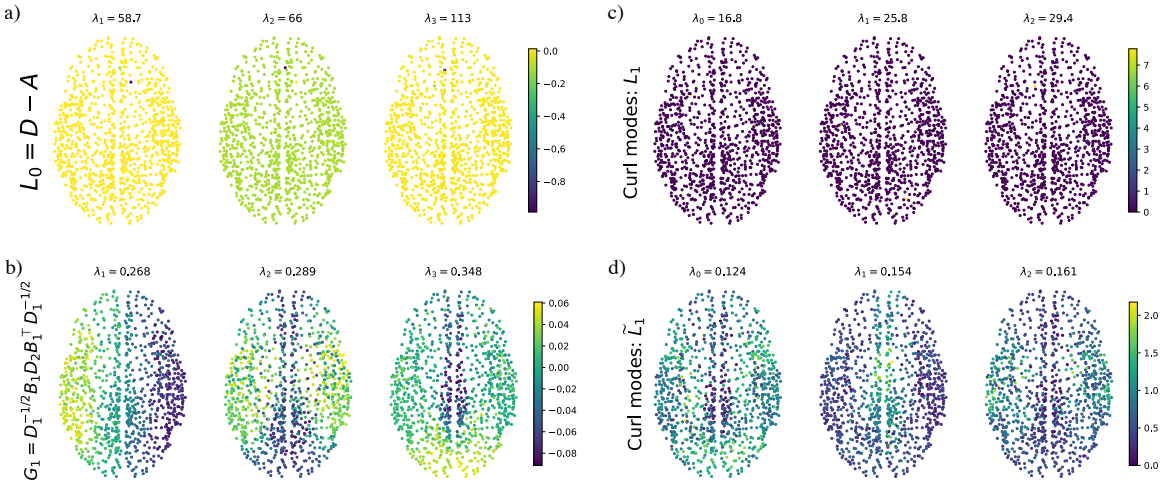

Figure 6.5: **Eigenmodes and localisation in the connectome.** We consider the first three non-trivial eigenmodes of *a*) the combinatorial Laplacian, L_0, and of *b*) G_1, for the Schaefer parcellation with 1000 regions (S1000) structural connectome, and plot these modes using the (x, y) position of each region in MNI space [254]. We see that the combinatorial Laplacian suffers from localisation, whilst the normalised Laplacian reflects the underlying structure of the connectome. For the divergence free component, we consider the first three curl modes of *c*) L_1 and *d*) $\widetilde{L}_1$, and project these onto the nodes, by summing the absolute value of the flow on all incident edges. Again, we observe localisation in the combinatorial Laplacian, and not in the normalised one.

[10] This can be shown by noting that $L_0 = B_1 B_1^\top$ and thus for any eigenmode v of L_0 with eigenvalue λ, we have $B_1^\top B_1 (B_1^\top v) = B_1^\top B_1 B_1^\top v = \lambda B_1^\top v$ i.e. $B_1^\top v$ is a gradient eigenmode of L_1.

Panel a) of Fig. 6.5 shows the first three (non-trivial) eigenmodes of L_0 for the S1000 structural connectome, plotted by the (x, y) position of the region in the brain.[11] We can see that the eigenmodes do not reflect the structure of the connectome, and instead are highly localised on a few nodes each. The projection of these modes onto the edges will also be highly localised, with non-trivial flows only on the edges incident to high-degree nodes.

On the other hand, gradient eigenmodes of $\widetilde{L}_1$ correspond to edge-projections of node-level modes of G_1 (see Eq. (6.26)). Panel b) shows the first three eigenmodes of G_1, which reflect known connectome structure e.g. splitting the connectome into the left and right hemisphere, and are very similar to the canonical modes derived from $\widetilde{L}_0$ and used in previous studies [85, 250, 255].

Next, we focus on the curl eigenmodes which span the kernel of the divergence.[12] We project these edge-modes onto the nodes by summing the absolute value of the flow on each edge incident to a particular node. Panels c) and d) show this projection for the first three curl modes of L_1 and $\widetilde{L}_1$ respectively. We see that there is localisation in the curl modes of L_1, but not in the normalised form $\widetilde{L}_1$, where the modes reflect the connectome structure.

With the exception of Schaub et al. [234], the practical use of normalised Hodge Laplacians has been notably neglected. As mentioned, this is due to such considerations being less important when performing geometric data analysis, but are necessary when studying signals on heterogeneous complex networks.

6.3 Detecting vortices with the Hodge decomposition

Equipped with the normalised GHHD associated with $\widetilde{L}_1$, we can now introduce our approach for measuring Hodge vorticity. Our method is formulated for a general network, but we will first illustrate it on the lattice, where we can compare it with the canonical approach defined in Sec. 6.1.2.

We begin with a graph $G = (V, E)$ with N nodes and M edges, with a phase-configuration, $\{\mathbf{s}\}$, on the nodes. The phases can be thought of as a potential on the nodes, $\theta : V \to [0, 2\pi) \subset \mathbb{R}$. From this potential, we define an edge-flow,

$$X(i, j) = \text{wrap}(\theta_j - \theta_i) = \text{wrap}(\text{grad}\,\theta)\,, \tag{6.30}$$

[11] We are using 'Montreal Neuroimaging Institute' (MNI) coordinates [254].

[12] Due to the high-density of triangles, we find that there is no harmonic component for the S1000 connectome.

where the wrap function is defined in Eq. (6.9), which naturally satisfies the alternating property, provided $\theta_i - \theta_j \neq \pm\pi$, which is generically true. The flow on an edge is then the smallest phase difference between the connected nodes.[13] Thus, we can decompose this edge-flow into,

$$X = \operatorname{grad} f + R, \tag{6.31}$$

where $f \in H(V)$ and $R \in \ker(\operatorname{div})$. This choice of edge-flow is motivated by an important property, which makes it clear that the divergence-free component, R, identifies the vortices. First, we introduce a few important definitions.

Let $G = (V, E)$ be a graph equipped with an arbitrary edge orientation. A *path* is a list of nodes $i_0, \ldots, i_n$ such that $(i_k, i_{k+1}) \in E$ i.e. there is an edge between each i_k and i_{k+1} for $k = 0, \ldots, n-1$, regardless of its orientation. A *cyclic path* is a path where $i_0 = i_n$ i.e. the first and final node coincide. An *oriented cycle* is a formal integer linear combination of edges,

$$\sigma = \sum_{k=1}^{n} m_k e_k \,, \tag{6.32}$$

satisfying the flow conservation condition,

$$\sum_{k:t(e_k)=v} m_k - \sum_{k:s(e_k)=v} m_k = 0, \tag{6.33}$$

for all $v \in V$, where $s(e_k)$ and $t(e_k)$ represent the source and target nodes of edge e_k respectively. We denote the set of all oriented cycles on G by Z_1. Each cyclic path induces an oriented cycle by counting the oriented multiplicity m_k with which each edge appears as a pair of adjacent nodes along the path.

We then have the following two results.

Lemma 6.3.1. *For any oriented cycle, $\sigma = \sum_{k=1}^{n} m_k e_k \in Z_1$, and any $f \in H(V)$, we have that*

$$grad(f)(\sigma) := \sum_{k=1}^{n} m_k grad(f)(e_k) = \sum_{k=1}^{n} m_k \big[f(t(e_k)) - f(s(e_k))\big] = 0 \,, \tag{6.34}$$

i.e. the net-change of any gradient flow along an oriented cycle must be zero. Conversely, if $Y \in H(E)$ satisfies that $Y(\sigma) = 0$ for all oriented cycles $\sigma \in Z_1$, then $Y = grad\, f$ for some $f \in H(V)$.

[13] The flow is defined as a difference along edges. However, this flow is *not* a gradient flow due to the wrap function, and thus has a divergence-free component which captures circulation around cycles in the graph.

This result is foundational and has been presented in a number of equivalent forms [64, 91, 256]. It is the discrete analogue of the statement that a vector field is conservative if and only if it is path independent.

Lemma 6.3.2. *For any oriented cycle, $\sigma = \sum_{k=1}^{n} m_k e_k \in Z_1$, and $X \in H(E)$ defined as in Eq. (6.30), we have that*

$$X(\sigma) := \sum_{k=1}^{n} m_k X(e_k) \in 2\pi\mathbb{Z}\,, \tag{6.35}$$

i.e. the net-change of X along any cyclic path must be an integer multiple of 2π.

Proof. Using the definition of the wrap function in Eq. (6.9), the phase difference along any edge can be written as,

$$X(e_k) = \theta_{t(e_k)} - \theta_{s(e_k)} + 2\pi a_k, \tag{6.36}$$

where $a_k \in \mathbb{Z}$. Evaluating X over the oriented cycle σ, we can write the net-change of X along a cyclic path as,

$$X(\sigma) = \sum_{k=1}^{n} m_k \left[\theta_{i_k} - \theta_{i_{k-1}}\right] + 2\pi \sum_{k=1}^{n} m_k a_k. \tag{6.37}$$

The first sum gives a discrete gradient around an oriented cycle, which vanishes by Lemma 6.3.1. The second term is an integer multiple of 2π, thus $X(\sigma) \in 2\pi\mathbb{Z}$. □

Combining these results, the following corollary is immediate.

Corollary 6.3.1. *Let $X \in H(E)$ be defined as in Eq. (6.30) with the GHHD, $X = \mathrm{grad}\, f + R$, where $R \in \ker(div)$. Then for any oriented cycle, $R(\sigma) = X(\sigma) \in 2\pi\mathbb{Z}$.*

The winding of the divergence-free component of the phase difference edge-flow is *quantised*, much like the integral in Eq. (6.8). Given our choice of edge-flow, we expect that the gradient component will contain the small phase differences consistent with the almost synchronised spins, whilst the divergence-free component will contains the vortices.

For visualisation purposes, we can also project the components onto the nodes by taking the sum of the incident edge-flows in absolute value,

$$r(i) = \sum_{j:(i,j)\in E} |R(i,j)|, \tag{6.38}$$

$$g(i) = \sum_{j:(i,j)\in E} |(\mathrm{grad}\, f)(i,j)|. \tag{6.39}$$

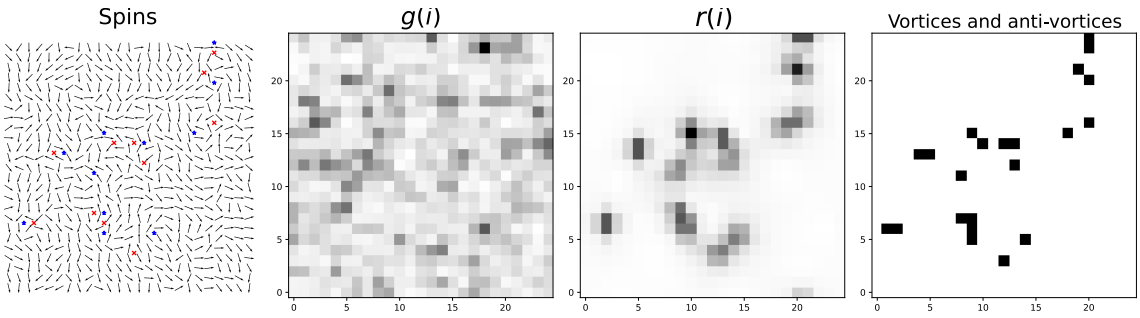

Figure 6.6: **Identifying vortices with the Hodge decomposition**. We illustrate our method on a spin configuration on the lattice. We define the edge-flow and then decompose it into gradient and divergence-free components, projected here onto the edges. The divergence-free component identifies the locations of the vortices and the anti-vortices. As in Fig. 6.4, a vortex is represented by a blue '$\star$' whilst an anti-vortex is represented by a red '$\times$'.

To illustrate that our approach correctly identifies the location of the vortices, Fig. 6.6 shows a spin configuration, the projections of the gradient and divergence-free (rotational) components, and the locations of the vortices (computed with the canonical approach). For the purposes of measuring the overall level of vorticity, it is not necessary to differentiate between vortices and anti-vortices (see Eq. (6.13)), thus we shall henceforth refer to them simply as 'vortices'.[14] Fig. 6.6 shows that the divergence-free component, and its projection $r(i)$, accurately reflect the location of the vortices. Crucially, it does so without a choice of cycle basis (or a notion of an elementary plaquette).[15]

Vorticity and vortices over time In order to apply this method to neural recordings, it is useful to define the overall vorticity of a configuration, which we do by taking the size of the rotational component relative to the whole flow, with respect to the edge-flow norm,

$$V_H = \frac{||R||_E}{||X||_E}. \tag{6.40}$$

We refer to V_H as the *Hodge vorticity*.

Additionally, when computing the Hodge vorticity of neural recordings, we do not have a static phase-configuration, but instead a multivariate time-series $\{\theta_i(t_k) :$

[14] On the lattice, this difference can be determined from the sign of the rotational component. Nevertheless, it is unclear differentiating between vortices and anti-vortices can be generalised beyond the lattice.

[15] We notice a 'blurring' effect. Some of this stems from the projection onto the edges, but some is inherent to the GHHD.

$i = 1, \ldots, N; k = 0, \ldots, \tau\}$ (details on phase-extraction will be given in Sec. 6.5). In this case, we consider each time-step to be an independent phase-configuration, and simply average the Hodge vorticity over the time-steps,

$$\bar{V}_H = \langle V_H(t) \rangle_t = \left\langle \frac{||R(t)||_E}{||X(t)||_E} \right\rangle_t, \tag{6.41}$$

where $\langle \cdot \rangle_t$ is the mean over the time-steps.

6.3.1 Expansion into eigenmodes of $\widetilde{L}_1$

In addition to computing the overall level of vorticity, we expand the edge-flow time-series into a basis of structurally-derived edge-level modes, namely the eigenmodes of $\widetilde{L}_1$. As mentioned, this is analogous to previous approaches that use eigenmodes of the node-level Laplacian (or Laplace-Beltrami operator) to analyse neural dynamics and perform graph signal processing [27, 85], but at the edge-level.

Given a phase time-series, $\{\theta_i(t_k)\}_{k=1}^{\tau}$, we calculate the edge-flow, $\{X(t_k)\}_{k=1}^{\tau}$, as in Eq. (6.30). We then transform the edge-signal into the orthonormal basis with $Y(t) = D_2^{1/2} X(t)$, which is a multivariate time-series in $\mathbb{R}^M$. Let U be the $M \times Q$ matrix where the columns are a collection of Q normalised eigenmodes of $\widetilde{L}_1$. We can project the edge-signal onto these eigenmodes,

$$\Upsilon_U(t) = U^\top Y(t), \tag{6.42}$$

where $\Upsilon_U(t) \in \mathbb{R}^Q$, and where the k-th entry, $\Upsilon_U^k(t)$ represents the *amplitude* of the k-th eigenmode at time t. As highlighted in Sec. 6.2.2, the projection onto the divergence-free eigenmodes captures the vortex-like patterns in the signal, whilst the projection onto the gradient modes captures the synchronised activity.

An expansion into eigenmodes of the Laplacian is analogous to performing a discrete Fourier transform on the signal [92, 257]. In this context, lower eigenvalues correspond to lower frequencies, which are associated with the large-scale, dominant patterns of co-activation that emerge from the structure of the graph. As a result, we can construct U by ordering modes by increasing eigenvalue, where truncation corresponds to filtering out high-frequency (potentially noisy) modes in the signal.

In order to measure the overall strength of a given mode in the expansion, we define the *energy* of mode k to be,

$$E_U(k) = \langle |\Upsilon_U^k(t)|^2 \rangle_t, \tag{6.43}$$

from which we can define a probability distribution,

$$P_U(k) = \frac{E_U(k)}{\sum_{l=1}^{Q} E_U(l)}. \tag{6.44}$$

Finally, we compute the Shannon entropy of this distribution,

$$\mathcal{E}_U = -\sum_{k=1}^{Q} P_U(k) \log\left(P_U(k)\right), \tag{6.45}$$

which measures the *eigenmode complexity* of the distribution. This measures how 'spread out' the power of the signal is over the various modes in U, with a low entropy implying that much of the signal can be captured with only a few modes, whilst a high entropy implies a more complex signal [258].

6.4 Numerical experiments

6.4.1 XY simulations on the lattice

The XY model on the lattice goes through the famous *Berezinskii-Kosterlitz-Thouless phase transition (BKT-PT)* at the critical temperature $\beta \approx 1.1199$ [259,260]. Simulations of this model on the lattice typically sample configurations, using MCM, along a range of increasing temperatures that goes through this critical point, recording the basic thermodynamic quantities introduced in Sec. 6.1.1 [244, 259]. The behaviour of these quantities, including the canonical vorticity, has a characteristic shape as a function of the temperature around the critical point.

We perform this same experiment, whilst also measuring the Hodge vorticity of the configurations at each temperature, which we compare to the canonical measure. We consider the XY model on a 100×100 lattice without periodic boundaries at a range of temperatures $\beta^{-1} \in (0, 2]$. At each temperature, we use the MH algorithm to perform sweeps with $n_T = 100,000$ and $n_M = 500,000$, where we take measurements of the energy density, magnetisation, susceptibility, heat capacity, vorticity, and Hodge vorticity.

Fig. 6.7 shows (the distribution of)[16] each of these measurements as a function of β^{-1}. We can see the characteristic increase in magnetisation, drop in energy, peak in susceptibility and heat capacity, that is known to occur near the critical

[16]The energy, magnetisation, and vorticity measurements are computed for each configuration, thus we have a distribution over a number of measuring sweeps at each temperature. The susceptibility and heat capacity are defined at the level of the distribution, thus we have a single value at each temperature.

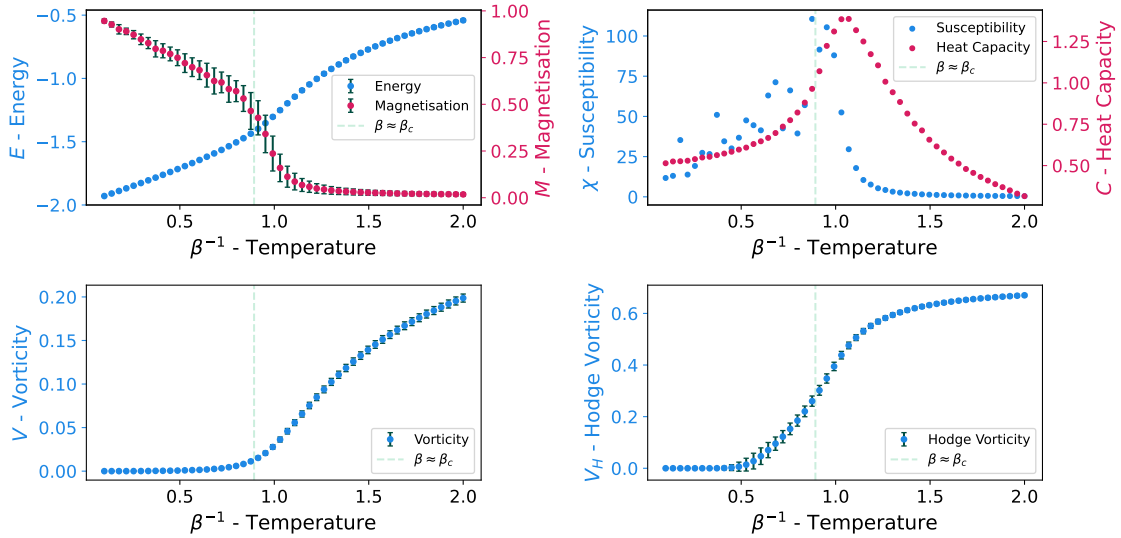

Figure 6.7: **Monte-Carlo simulations of the XY model on the lattice under increasing β^{-1}.** We draw samples from the XY model on a 100×100 lattice for increasing temperature and observe a well-known critical phase-transition. As we move through the critical temperature, the magnetisation decreases, the energy increases, and the susceptibility and heat capacity peak. The traditional and Hodge vorticity qualitatively agree, increasing rapidly as we move through the critical point.

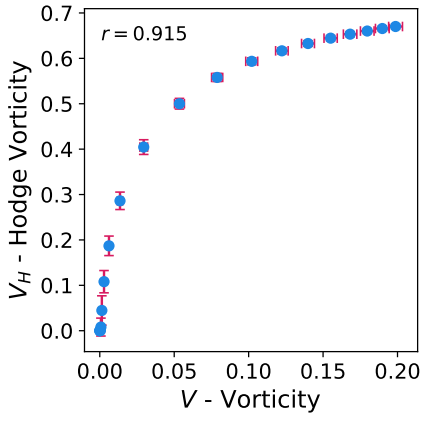

Figure 6.8: **Correlation between vorticity and Hodge vorticity.** Using the same data as in Fig. 6.7, we plot the vorticity, V, against the Hodge vorticity, V_H, across the temperature range. We find a highly correlated, monotonic, yet nonlinear, relationship.

temperature [244, 259]. Most importantly, we see qualitative agreement between the traditional vorticity and the Hodge vorticity, which are both near zero before the critical temperature, and then increase rapidly close to the critical temperature.[17] This agreement supports the idea that Hodge vorticity is capturing the same phenomenon as the traditional counting measurement, despite slight differences in the way each measurement is defined. Moreover, Fig. 6.8 plots the vorticity against the Hodge vorticity across the temperature range. This shows high correlation and a monotonic relationship. However, the relationship is nonlinear suggesting that the two measures are not perfectly equivalent, perhaps due to the normalisations used when defining the Hodge vorticity.

6.4.2 XY simulations on the connectome

Next, we will consider simulations of the XY model on the human connectome parcellated into different atlases, and investigate the effect of increasing temperature on the thermodynamic measurements and the Hodge vorticity. In this case, the canonical approach for measuring vorticity is no longer defined. Whilst the specific properties of the BKT-PT transition do not extend to complex networks, specifically 'vortex unbinding' [260], previous studies have noted a similar temperature transition point on complex networks [261], which we refrain from labelling as 'critical'.

Structural connectomes and brain atlases We consider three parcellations of the human connectome. The first is the DK80, introduced in Sec. 3.4.1. The second and third are the Schaefer parcellation with 100 regions (S100), and with 1000 regions (S1000) [262]. Whilst the structural networks extracted from DTI are typically weighted, we binarise the connectomes to form unweighted, undirected graphs.

We consider the XY model on each of the three connectomes. Depending on the number of regions, we select a different temperature range, and take 50 equispaced values. For the DK80 and S100, we have the range $\beta^{-1} \in [10, 100]$; for the S1000, we have the range $\beta^{-1} \in [10, 500]$. We use the MH algorithm with $n_T = 10,000$ and $n_M = 75,000$ measurement sweeps.

Fig. 6.9 shows the thermodynamic measurements as a function of β^{-1} for the DK80 [Panel a)], the S100 [Panel b)], and the S1000 connectomes [Panel c)]. In

[17] There is a slightly disagreement between the measures in that the Hodge vorticity begins to increase before the traditional measure. This, most probably, occurs due to the fact the Hodge vorticity is a smooth measurement whereas the traditional measurement is a discrete counting function. Before the emergence of the first vortex, the traditional vorticity is zero whereas the Hodge vorticity may detect the beginnings of rotational patterns.

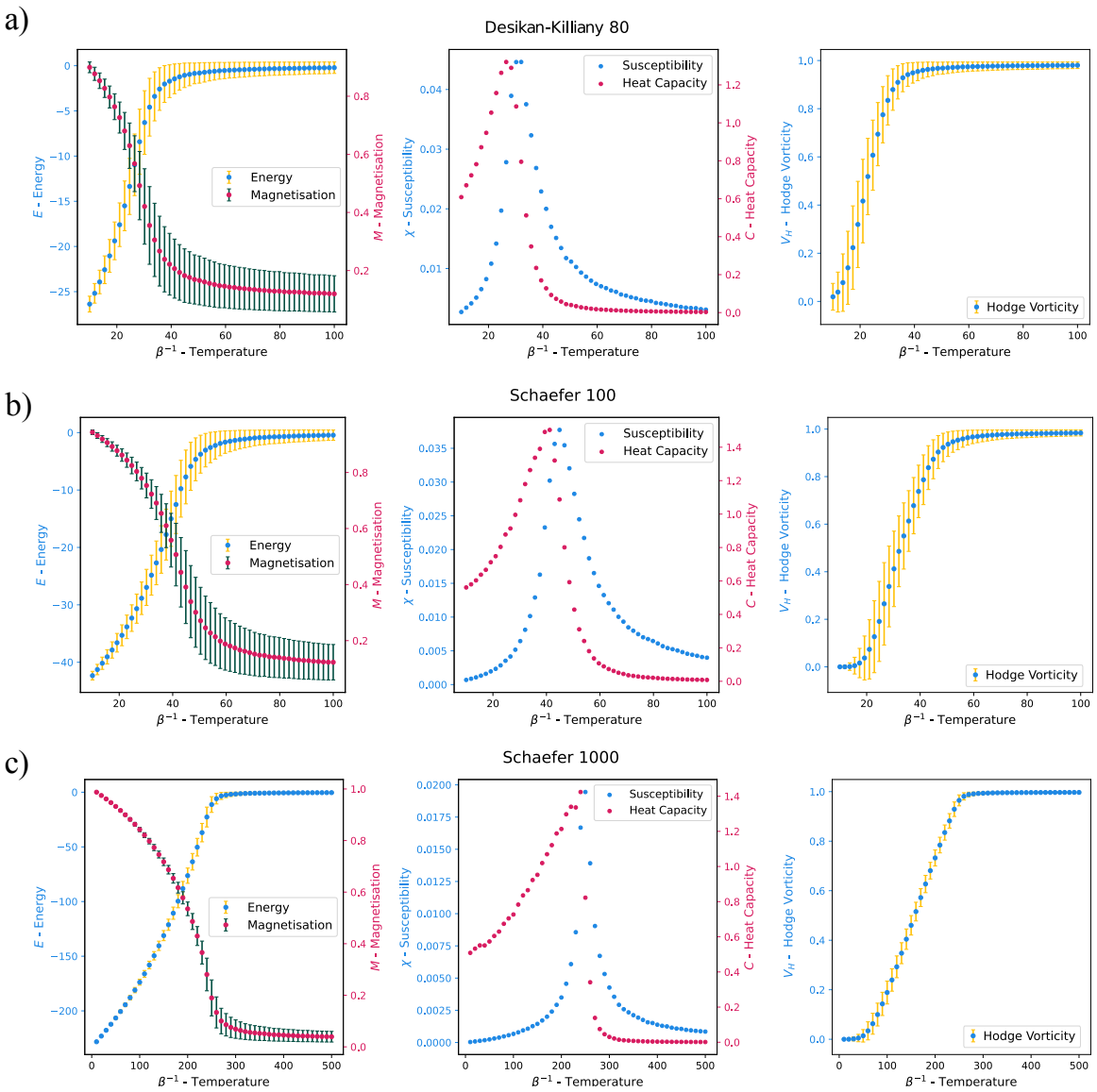

Figure 6.9: **Monte-Carlo simulations of the XY model on various connectomes under increasing β^{-1}.** We draw samples from the XY model on the *a*) DK80, *b*) S100, and *c*) S1000 connectomes for increasing temperature. We see a similar transition occur to that observed on the lattice in Fig. 6.7. Here, the canonical measure of vorticity cannot be applied, but the Hodge vorticity can. We see it increase around the transition point, as both vorticity measures did on the lattice.

each case, we see that a transition occurs around a particular temperature, which appears to shift depending on the number of nodes in the network. This transition is very similar to that seen on the lattice and features the same characteristic drop in magnetisation, increase in energy, and peak in susceptibility and heat capacity. Most interestingly, we also see that, during the transition, the Hodge vorticity increases in a fashion that mirrors the behaviour of both the Hodge and canonical vorticity on

the lattice.

This transition has been observed on complex networks with different topologies [261,263,264], but without a notion of vorticity. Our results support the idea that our method generalises vorticity to the case of complex networks, and provides evidence that the topology of the connectome can support the formation of vortex-like patterns.

6.5 Applications to neuroimaging data

We now move away from simulations of the XY model to focus on empirical recordings of neural dynamics in the form of fMRI. We apply our approach to this data to measure the vorticity and complexity of neural dynamics in both resting and task-based activity.

Data and preprocessing We consider fMRI BOLD recordings taken from 100 unrelated adult participants from the HCP database [125]. In particular, we consider activity in resting-state and in seven cognitive tasks, which are designated as: 'Emotion', 'Gambling', 'Language', 'Motor', 'Relational', 'Social', and 'Working Memory' (WM) [126]. We parcellate the recordings into the S1000 brain atlas, and follow the standard HCP recommendations on signal preprocessing (for further details see Ref. [238,265,266]).[18] This yields a multivariate time-series with $N = 998$ and $\tau = 176$ time-steps [Panel a), Fig. 6.10].[19]

Phase extraction As our focus is on brain network oscillations, we must extract the instantaneous phase of the signal. We filter the signal in the range [0.008, 0.08] Hz with a Butterworth bandpass filter, and use the *Hilbert transform* to convert the real-valued signal into a complex-valued signal, and then to extract the instantaneous phase at each time-point for each region i.e. we extract $\{\theta_i(t_k) : i = 1, \ldots, N; k = 1, \ldots, \tau\}$ [269] [Panel b), Fig. 6.10].[20]

[18] When parcellating neuroimaging data into the S1000 atlas with standard neuroimaging packages, it is known that there are two regions which yield errors and thus their time-series cannot be extracted [267, 268]. As a result, we consider the activity in 998 regions.

[19] As the duration between time-steps is 1 second, this implies we have 2 minutes and 56 seconds of recording per participant.

[20] Filtering in this frequency range and then applying the Hilbert transform is the standard approach in studies of turbulence and vorticity [238, 245], and modelling fMRI with Kuramoto oscillators [236], as this is the range where most resting-state activity is localised [23].

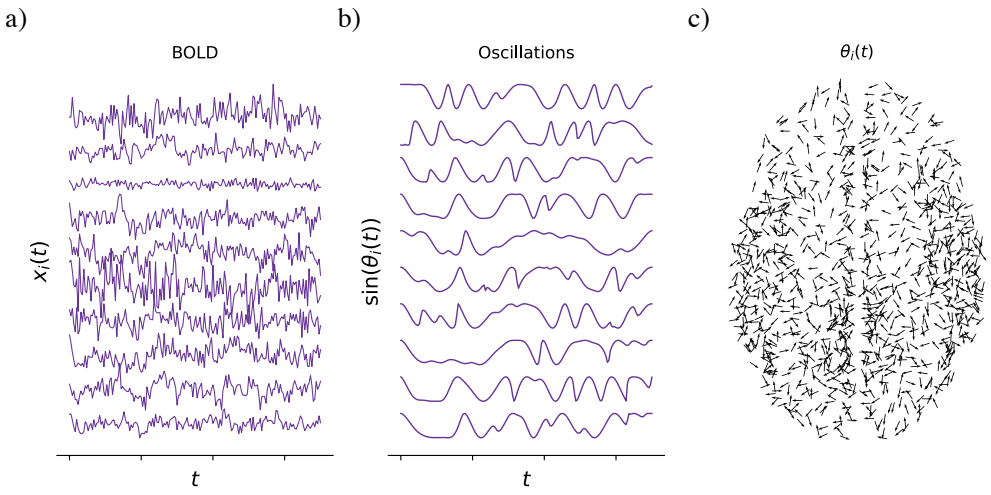

Figure 6.10: **Phase extraction of neural recordings.** *a*) We begin with BOLD activity, plotted here for 10 random regions. *b*) Using the Hilbert transform, we extract the instantaneous phase of the signal in each region. *c*) Each time-step can be thought of as a phase-configuration on a complex network, much like a sample from the XY model.

Remark 6.5.1 (The Hilbert transform)**.** Given a real-valued signal $x(t)$, the Hilbert transform creates a version of the signal that is phase-shifted by $\pi/2$ at every frequency,

$$\widehat{x}(t) = \mathcal{H}(x(t)) = -\frac{1}{\pi} \lim_{\epsilon \to 0} \int_{\epsilon}^{\infty} \frac{x(t+\tau) - x(t-\tau)}{\tau} \, d\tau, \tag{6.46}$$

which is known as the *analytical signal*, and from which we define the complex signal $z(t) = x(t) + \mathbf{i}\widehat{x}(t)$. From this, we extract the instantaneous amplitude, $A(t)$, and the instantaneous phase $\theta(t)$. Performing this for each region in turn, we obtain a time-series of phase-configurations on a complex network, where the underlying structure is given by the unweighted, undirected connectivity of the S1000 atlas [Panel *c*), Fig. 6.10]. In practice, we use the `scipy.signal.hilbert` method.

6.5.1 Hodge vorticity during cognitive tasks

First, we measure the overall Hodge vorticity for each participant in each cognitive task using Eq. (6.41). Fig. 6.11 shows the distribution of the average Hodge vorticity, $\langle V_H \rangle_t$, for each of the 100 participants in each of the eight tasks. We find that the overall level of Hodge vorticity is higher in each of the cognitive tasks compared to the baseline of the resting state (with a range of significance). This result complements previous studies which found higher levels of turbulent-like activity during tasks [238]

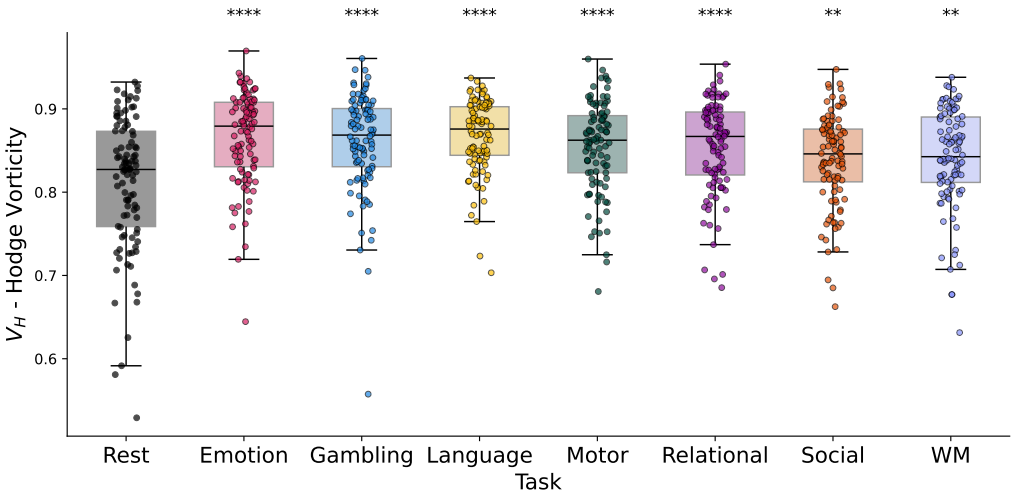

Figure 6.11: **Hodge vorticity of neural dynamics during cognitive tasks**. We measure the average Hodge vorticity, V_H, from fMRI recordings of 100 participants in resting state and seven cognitive tasks. We find higher vorticity in each task when compared to resting state, using paired t-tests. Significance is denoted as follows: $p > 0.05$ (n.s.), $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***), $p < 0.0001$ (****).

alongside fluctuations in the vorticity [245, 246]. This provides evidence that vortex-like dynamics emerge during, and are correlated with, cognitive processing.

Next, we project the divergence-free component of the neural recordings, $R(t)$, which is a signal on the edges of the connectome, onto the nodes using Eq. (6.38). This gives us a 'Hodge vorticity map' $\{r_i\}$ i.e. a potential r on the nodes that describes the overall level of a Hodge vorticity for each region in the S1000 atlas. Fig. 6.12 shows this regional Hodge vorticity map projected onto an illustration of the flattened cortex of the left and right hemisphere, with the colour denoting $r(i)$, averaged over the time-series and all participants.[21] We observe a consistent difference between resting state and the cognitive tasks. Across the tasks, the maps look similar in a number of places, suggesting that specific regions are systematically implicated in vortex-like patterns during cognitive processing. This similarity could emerge as a result of the structure of the underlying network, which we assume to be the same for each task. Nevertheless, some unique patterns emerge for each task, consistent with the idea of specialised function among brain regions.

[21] In order to maximise contrast in a given map, we use a different colour-scale for each task.

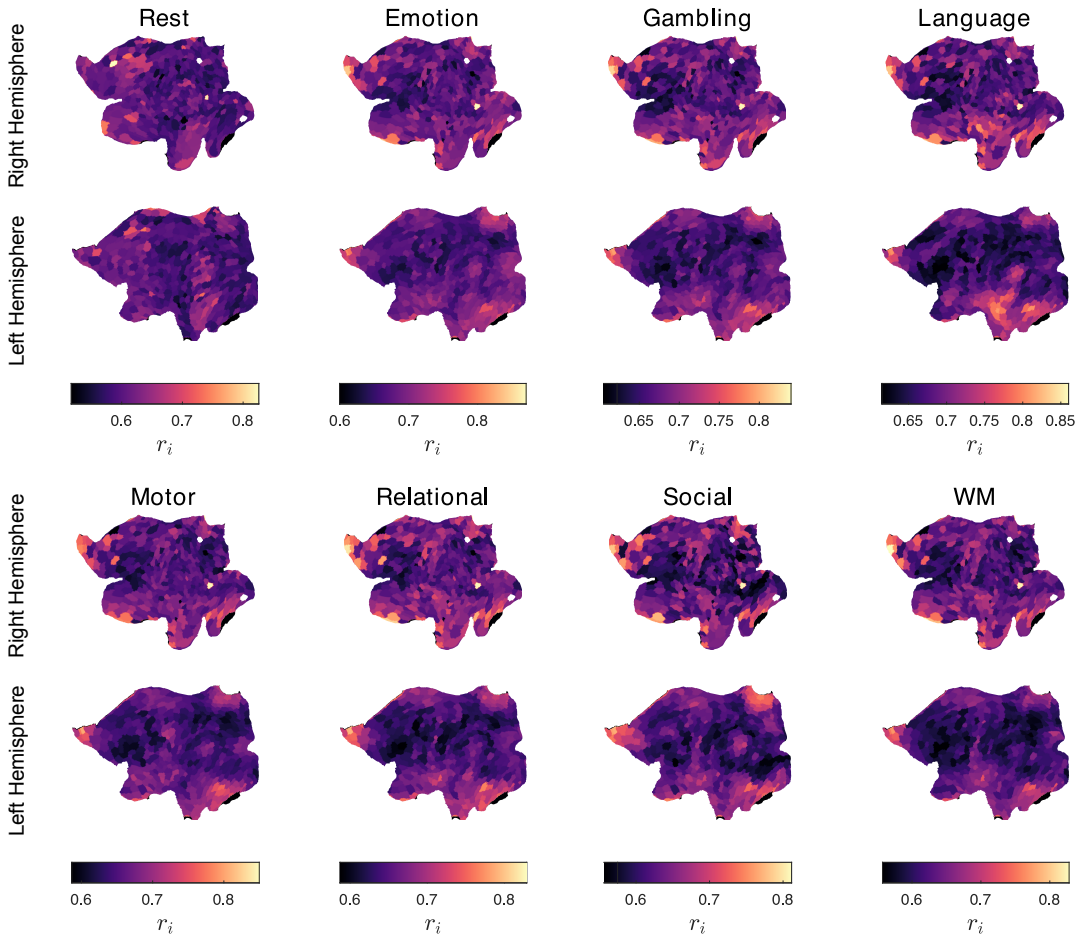

Figure 6.12: **Brain maps of average regional vorticity, $r(i)$, during cognitive tasks.** We compute the average overall vorticity incident to each individual region and produce a 'vorticity map' for each task. We project this onto a flattened illustration of the left and right cortex.

6.5.2 Eigenmode expansion of edge-based neural signals

In order to expand the neural signals in the eigenmode basis of $\widetilde{L}_1$, we first need to compute the eigenmodes of the S1000 connectome. The connectome has $N = 998$ regions and $M = 232,384$ edges, which induce $\Theta = 24,194,608$ triangles. As mentioned before, we find that there are no harmonic modes, which implies that the clique complex of triangles has no holes and $\beta_1 = 1$. As the graph is connected, we have $N-1$ non-trivial gradient modes which can be computed efficiently at the node-level and then projected as described in Sec. 6.2.2. This leaves the $M-(N-1)$ curl modes, which is a computationally impractical number of modes to calculate. Using the LOBPCG method, we compute the first 1000 curl modes i.e. the 1000 modes with the smallest eigenvalues.

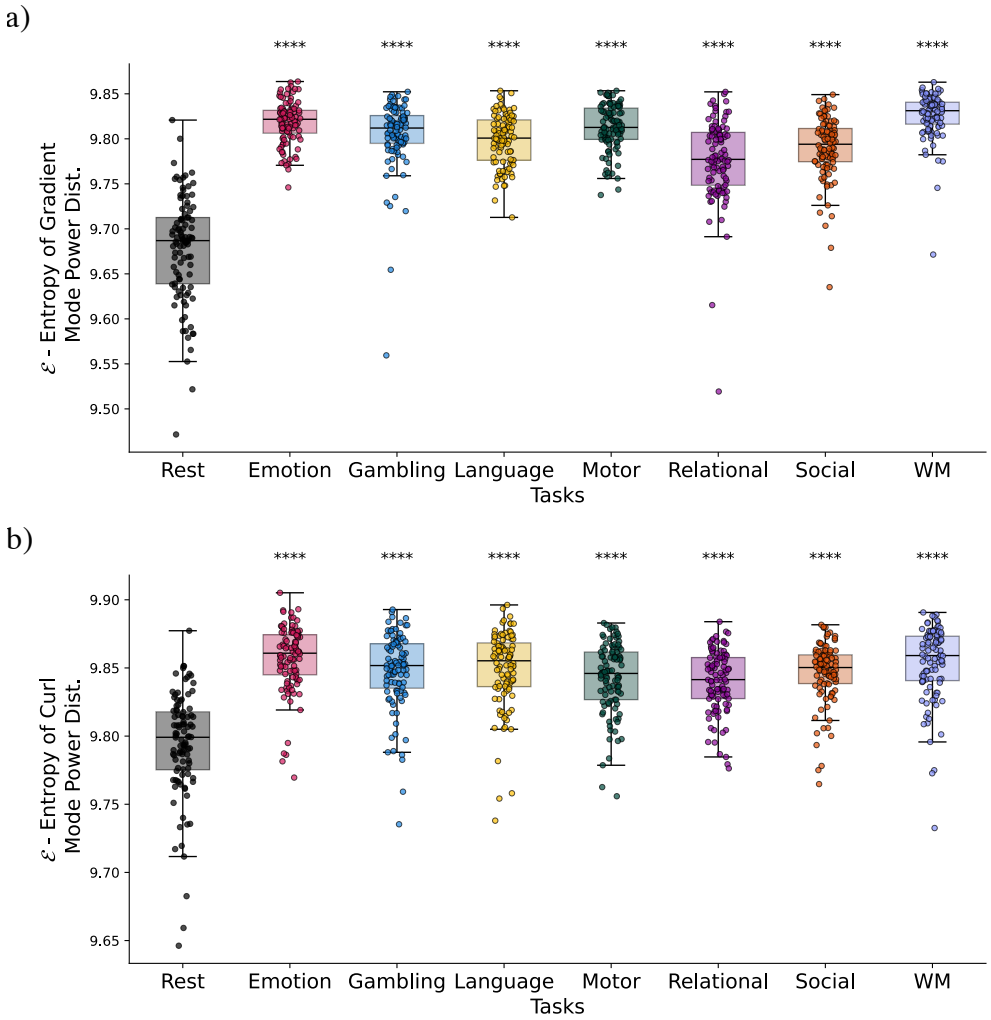

Figure 6.13: **Entropy of eigenmode power distributions in cognitive tasks.** For each participant in each task and resting-state, we project the neural signals onto the basis of gradient modes and a partial basis of the first 1000 curl modes. We then calculate the power distribution and its entropy, which we plot for the *a*) gradient modes *b*) curl modes. We find higher entropy for the expansion into both gradient and curl modes for each task compared to resting state.

As described in Sec. 6.3.1, for each participant in each task, we compute the edge-flow and project it onto either the curl basis or the gradient basis. We then compute the power distribution and its entropy, which is a measure of the complexity of the representation in the fundamental modes of the system. Fig. 6.13 shows the entropy for each participant and each task for the gradient basis [Panel *a*)] and the curl basis [Panel *b*)]. We see that there is a higher entropy in every task compared to resting state (at the highest level of significance, $p < 0.0001$). This result suggests that both the synchronised gradient flows captured by the gradient modes, and the vortex-

like dynamics captured by the curl modes, are more complex in task-based neural activity than at rest. We expect that this result emerges as a consequence of cognitive processing. Our results are consistent with a variety of other studies indicating that task-activated dynamics are significantly more complex than resting-state such as those that focus on statistical-physics inspired measures such as turbulence [238] and irreversibility (see Ch. 3), temporal complexity [270], and representational complexity in Laplace eigenbases [255].

6.6 Conclusions

Large-scale brain network oscillations display a number of emergent features, such as transient synchronisation, turbulent-like dynamics, and the emergence of vortex-like dynamics. Whilst the latter has been of recent interest [245,246], it has been restricted to a continuous, geometric view of the brain, which is inconsistent with the paradigm of network neuroscience [19, 23]. Using the framework of discrete Hodge theory, we have introduced a novel approach for measuring vorticity in oscillator networks that can be applied to arbitrary network topologies, without a choice of cycle basis, and thus can be applied to neuroimaging data.

To justify that our approach truly captures the emergence of vortices, we validate our method with simulations of the XY model on the lattice, and show that our results align with the canonical method, including during the BKT-PT i.e. increasing temperature through the critical point. Our approach can be also be applied to complex networks, which we confirmed with numerical experiments of the XY model on a range of connectomes. Next, we turn to the analysis of task-activated and resting-state fMRI data. With our method, we show that the Hodge vorticity of task-activated dynamics is higher than in the resting state, which is in-keeping with previous results surrounding the vorticity of different tasks [246], the 'turbulence' of task-activated dynamics [238], and other measurements from statistical physics. Finally, using an expansion into an appropriate eigenbasis, we devise a measure of complexity for the vortex-like dynamics, and show evidence for higher complexity in task-activated dynamics compared to resting state.

Our work offers a unique framework for probing an interesting phenomenon in oscillator networks and neural dynamics using topological signal processing. It introduces links between two seemingly disparate areas of mathematics, namely discrete Hodge theory and the statistical mechanics of phase-oscillators. In contrast with the Kuramoto and Ising models, the XY model has been relatively understudied in the

realm of complex networks [261], and our method provides a way for one its principal features, the emergence of vortices, to be extended to complex network topologies. This holds promise for future applications in a number of domains where the dynamics of oscillator networks are of interest, including neuroscience, power grids, and the study of collective behaviour.

Limitations and future directions Whilst we present a novel approach for detecting vortices and measuring vorticity in neural data, this is a first study, and a number of important questions remain. We focused on the aggregate level of Hodge vorticity in cognitive states, showing significant differences between tasks and resting-state. This falls short of revealing the cognitive mechanisms that lead to the emergence of vortices, or their implications for theories of neural computation based on oscillatory coherence [241, 242]. Moreover, unlike studies that use the NFT approach [246], we are not able to 'pinpoint' or track an individual vortex as it moves in space and time, or to differentiate between a vortex and an anti-vortex. Previous evidence suggests that the difference between vortices and anti-vortices has some relationship with cognitive processing [246]. This distinction is absent from our framework and it is unclear if such a differentiation is well-posed for oscillators on complex networks.

Furthermore, we focused on static XY dynamics in an ESS, even though neural recordings are more consistent with dynamic models, such as the Kuramoto or dynamic XY models [249], and are more accurately approximated by a process in an NESS, as is the focus in the rest of this thesis. Finally, the relationship between vorticity and other dynamic features such as synchronisation, chimera states, and turbulence, remains to be investigated in both theoretical models of oscillator networks and empirical neural recordings.

Chapter 7

Embedding latent dynamics in neural networks

7.1 Introduction

Throughout this thesis, particularly in Ch. 3, 4, and 6, we have considered how the spatiotemporal dynamics of large-scale brain network activity can be analysed from the perspective of statistical mechanics and complex systems. However, we have yet to consider the mechanisms of *computation* in neural systems, which occurs at the level of individual neurons. Understanding how networked populations of neurons encode information and perform computations is one of the most significant problems in neuroscience [271].

Central to the connectionist perspective, which we outlined in Sec. 1.2.1, is the idea that cognition is performed by distributed networks of interacting neurons. However, unlike the 'feed-forward' architectures that dominate modern machine learning (ML) [Panel *a*), Fig. 7.1], the neuronal networks of the brain are *recurrent* [Panel *b*), Fig. 7.1], leading to a rich repertoire of time-evolving dynamics.

As a result, classes of *recurrent neural networks (RNNs)* have become the prototypical models for dynamic computation in biological neural circuits [272]. These networks, and variants thereof, have been used extensively to model neural data [273], develop theories of neural computation [274], as well as perform sequence learning tasks, such as time-series prediction.

In order to describe the aims and contribution of this chapter, we must first introduce Hopfield's seminal model of associative memory, and the emerging theory of neural manifolds.

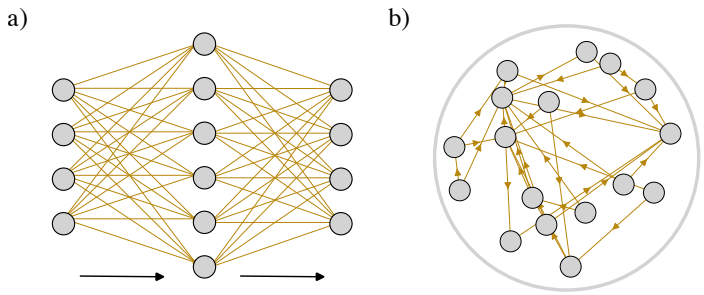

Figure 7.1: **Feed-forward and recurrent neural networks.** *a*) The feed-forward networks of modern deep learning are organised into layers, where connections can only go forward between layers. This leads to them being static function approximators. The dimension of the first (input) layer is the dimension of the input space, and the dimension of the final (output) layer is the dimension of the output space. *b*) Recurrent neural networks take the form of a general weighted network. Their recurrence leads to states that evolve in time. These networks are more consistent with the architecture of the brain.

7.1.1 Hopfield's model of associative memory

In 1982, Hopfield proposed a simple model of an RNN with discrete-state dynamics that could be trained to retrieve a particular pattern [34]. In 1984, he extended this model to continuous-state neurons [275], the focus of this chapter.

In the continuous version of the model, the system is composed of n neurons, each with internal state u_i, representing the 'firing rate' of the neuron, and output state $v_i = h(u_i)$, where h is a monotonic nonlinear activation function. For the remainder of this chapter, we will take $h(x) = \tanh(x)$ – a common choice. The dynamics of the internal state are given by the SDE,

$$d\mathbf{u}(t) = \mathbf{F}(\mathbf{u})\, dt + B\, d\mathbf{w}(t), \tag{7.1}$$

where $\mathbf{u} = (u_1, .., u_n) \in \mathbb{R}^n$, $B \in \mathbb{R}^{n\times d}$ and $\mathbf{w}(t)$ is a Wiener process with d independent sources of noise. We also define $\Upsilon = BB^\top/2$ to be the positive semidefinite diffusion matrix of rank d.[1] We define $\mathbf{F} = (F_1, \ldots, F_n)$ to be the drift function, with

$$F_i = -u_i + \sum_{j=1}^{n} W_{ij} v_j + I_i, \qquad\qquad i = 1, \ldots, n. \tag{7.2}$$

The term $-u_i$ is the 'leak' term. The input current to neuron i is given by I_i. Finally, $\sum_j W_{ij} v_j$ represents the contribution to neuron i from neighbouring neurons where

[1] In this chapter, notation will vary somewhat from previous chapters. In particular, D will be introduced as a diagonal preconditioning matrix for the gradient-flow, thus we use Υ to be the diffusion matrix.

the connectivity from neuron j to neuron i is given by W_{ij}. Unlike the settings in previous chapters, the weights between neurons, $W_{ij} \in \mathbb{R}$, can take negative values as this is consistent with *inhibitory connectivity* at the neuronal level [21].

Hopfield enforces that connectivity between neurons is symmetric i.e. $W = W^\top$. This restriction prompts us to introduce the *energy potential*,

$$E(\mathbf{v}) = -\frac{1}{2}\sum_{j,i}^{n} W_{ij} v_i v_j - \sum_i I_i v_i + \sum_i^n \int_0^{v_i} h^{-1}(\omega)\, d\omega, \tag{7.3}$$

and $\widetilde{E}(\mathbf{u}) = E(\mathbf{h}(\mathbf{u}))$. When $W = W^\top$, the drift can be written in the form,

$$\mathbf{F}(\mathbf{u}) = -\nabla_{\mathbf{v}} E(\mathbf{v}) = -\mathbf{D}(\mathbf{u})\nabla_{\mathbf{u}}\widetilde{E}(\mathbf{u}), \tag{7.4}$$

where $\mathbf{D}(\mathbf{u}) = \operatorname{diag}\left(\frac{1}{h'(u_1)}, \ldots, \frac{1}{h'(u_n)}\right)$ is a positive-definite matrix-valued function. As a result, the RNN is constrained to perform 'generalised' gradient descent on the energy landscape, eventually arriving at a local minimum of E.

Remark 7.1.1. The prefactor $\mathbf{D}(\mathbf{u})$ causes the gradient to be scaled unevenly in different directions, thus the system does not follow the Euclidean direction of steepest descent on the landscape. Instead, the drift is a gradient flow with respect to the Riemannian metric defined by $\mathbf{D}(\mathbf{u})^{-1}$. Practically, we would see that the system descends the landscape monotonically, but more quickly in directions where $h'(u_i)$ is small.

Hopfield's model allows for the manipulation of the energy landscape via training, such that a desired 'memory' can be encoded as a local energy minimum. The result is an autonomous neural circuit that will 'retrieve the memory given initial information' – this models the function of *associative memory* in the brain i.e. when one is able to recover a memory based on partial 'associated' information. In this case, the network converges to the corresponding energy minimum from an appropriate initial state. Hopfield's results laid the groundwork for the study of *attractor neural networks*, and their possible implications for neuroscience [276, 277].

Example 7.1.1 (A continuous Hopfield network)**.** In order to visualise memory retrieval in a typical, continuous Hopfield network, we organise the $n = m^2$ neurons into a square grid, where neural states correspond to simple 'pictures'. We consider an example with $m = 32$. Panel *a*) of Fig. 7.2 shows the 'Cat-Shark' memory, which corresponds to a vector $(v_1^\zeta, \ldots, v_n^\zeta) \in \{-1, 1\}^n \subset \mathbb{R}^n$. We then set the weights of the neural network to be,

$$W_{ij} = v_i^\zeta v_j^\zeta, \tag{7.5}$$

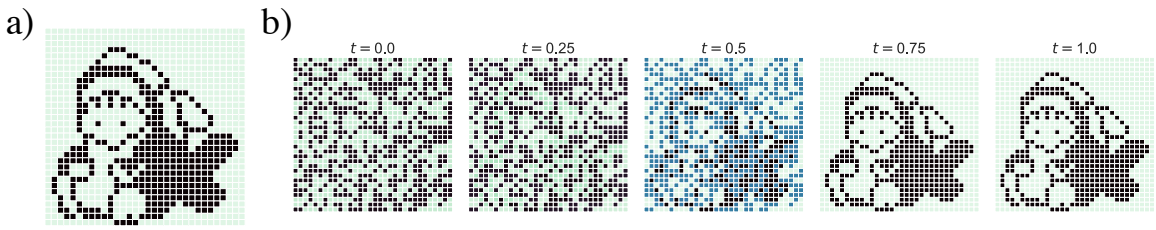

Figure 7.2: **Memory retrieval in a Hopfield network.** *a*) We organise the neurons into a grid such that neuronal states correspond to pixelated images. We train the Hopfield network to encode a binarised, pixel-image of a cat dressed as a shark. *b*) This shows a (noiseless) trajectory starting from a random, but nearby, initial condition. The network is able to recover the memory.

The 'Cat-Shark' pixel-image is used with permission from author, Pauline Thibaut.

which is Hopfield's symmetric learning rule.[2]

We can integrate Eq. (7.1) from a sufficiently close initial condition to simulate the activity of the RNN. Panel *b*) shows a trajectory of the system (in the absence of the noise), and shows that the network is able to correctly recover the memory.

Despite Hopfield's application to associative memory, limiting to symmetric connectivity severely constrains the repertoire of dynamics within the system, and precludes the existence of limit-cycles and other dynamic phenomena. Moreover, asymmetry is ubiquitous in biological neural circuits, thus the symmetric assumption limits the neuroscientific plausibility of such a model. Whilst numerous studies have analysed asymmetric Hopfield networks, these consider the discrete, spin-glass form of the model and analyse the attracting states emerging from random connectivity [279–281]. In the continuous model, it is known that asymmetric connectivity disrupts the energy landscape eroding the ability of the network to preserve attracting states [213]. Moreover, the dichotomy of symmetric and asymmetric connectivity is closely related to the necessary departure from equilibrium to nonequilibrium statistical mechanics, which has been the focus of this thesis, and has gained traction in recent research into neuroscience, neural computation, and associative memory [213, 282–286].

Remark 7.1.2. Hopfield's discrete, stochastic model [34] is equivalent to a symmetric Ising model, and therefore is in thermodynamic equilibrium [285]. In the case of the continuous model, we note that whilst gradient-flow dynamics are reversible

[2]This is a *Hebbian* learning rule, meaning that it follows Hebb's prominent theory of neuronal interactions [278]. It is most commonly summarised as 'neurons that fire together wire together'. In this context, this means that neurons that have the same sign in the memory should have a positive interaction and those that have different signs should have a negative interaction.

(Ex. 2.2.2), this model is only a gradient flow in the appropriate Riemannian metric, thus does not preserve detailed balance in Euclidean coordinates. Nevertheless, it is constrained to descend an energy function and cannot admit rich, time-dependent, nonequilibrium behaviour without asymmetry.

7.1.2 Neural manifolds

In an entirely separate development, dimensionality reduction techniques have shown the emergence of low-dimensional, latent structure in high-dimensional neural activity associated with tasks like motor control, navigation and working memory [287, 288]. These so-called *neural manifolds* indicate that high-dimensional brain network activity is, in fact, constrained to low-dimensional subspaces where the dynamics are often more mechanistically interpretable. Fig. 7.3 shows an example of the emergence of a toroidal neural manifold in the dynamics of grid cells in the entorhinal cortex of a mouse during a navigation task when projected with dimensionality reduction methods [289].

These low-dimensional dynamics appear to be responsible for computation and also emerge in RNNs [290], typically associated with *low-rank* structure in the connectivity [291, 292].[3] Neural data analysis has provided significant evidence that computational dynamics are encoded at the level of latent manifolds in empirical brain network activity [287, 288, 295]. As a result, understanding how such cognitive processes can be encoded in RNNs via neural manifolds is crucial to understanding the mechanisms and limits of neural computation.

Aims and contributions In this chapter, we will combine asymmetric attractor neural networks with the theory of neural manifolds by proposing a framework for training RNNs with the objective of representing any target SDE in a latent affine subspace. As a result, we show that sufficiently large RNNs can encode arbitrary dynamics in this latent space, provided asymmetric connectivity is permitted. We illustrate this by embedding nonlinear and nonequilibrium systems, such as chaotic attractors, within the RNN.

Next, to illustrate how the theory can be linked to associative memory, we introduce two classes of SDEs, both of which have a number of attracting states. The first is the gradient of a tiltable energy potential, which allows the network to choose

[3] Neural manifolds are just a single example forming part of the more general hypothesis that real-world network dynamics are effectively low-dimensional [293, 294], which in turn can be seen as a special case of the well-known *manifold hypothesis* for high-dimensional data [219].

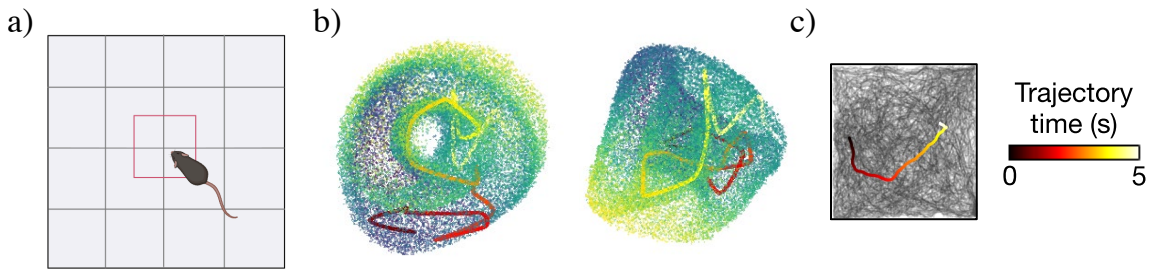

Figure 7.3: **Neural manifold structure in grid cell activity.** An example of toroidal manifold structure in neural population dynamics from the medial entorhinal cortex. *a*) A single mouse during free exploration of a square environment *b*) Two views of a three-dimensional UMAP† embedding of the activity of 149 grid cells recorded in the mouse. Each point is a state, and is coloured by its projection on the first principal component identified in the population activity *c*) A top-down view of the mouse's environment.*

*Adapted from Ref. [289] under CC BY 4.0. †Uniform Manifold Approximation and Projection (UMAP) is a form of nonlinear dimensionality reduction.

between energy minima using input-driven switching. The second is a nonequilibrium stationary diffusion that cycles autonomously between attractors in a specified order using irreversible currents. We present these as illustrative models of both input-driven associative and sequential (episodic) associative memory [296]. We apply our framework to show that such dynamics can be encoded in a latent subspace of an asymmetric RNN.

In order to investigate how the dynamics are encoded in the network, we introduce two decompositions of the RNN, the first based on the asymmetry of the connectivity, and the second on the time-irreversibility of the RNN (i.e. the HHD). This theoretical model suggests that populations of neurons with asymmetric connectivity are capable of encoding a wide array of dynamical systems, including attractor-switching and -cycling dynamics, thus extending Hopfield's model and integrating it with the theory of neural manifolds.

7.2 The 'drift-diffusion matching' framework

We now introduce our framework for training RNNs to encode target dynamics over a latent neural manifold. Consider once more an RNN of the form,

$$\begin{aligned} d\mathbf{u}(t) &= \mathbf{F}(\mathbf{u})\,dt + B\,d\mathbf{w}(t), \\ F_i &= -u_i + \sum_j W_{ij} v_j + I_i, \end{aligned} \tag{7.6}$$

for $i = 1, \ldots, n$, and consider the latent *affine subspace*,

$$\mathcal{A} = \{\mathbf{u} \in \mathbb{R}^n : \mathbf{u} = \Gamma \mathbf{y} + \mathbf{b}, \mathbf{y} \in \mathbb{R}^k\} \tag{7.7}$$

for some $\Gamma \in \mathbb{R}^{n \times k}$ and $\mathbf{b} \in \mathbb{R}^n$ for some latent dimension $k \ll n$. We first ask whether $\{W, I, B, \Gamma, \mathbf{b}\}$ can be chosen such that the dynamics of our latent parametrisation, $\mathbf{y}(t)$, follow a target SDE of the form,

$$d\mathbf{y}(t) = \mathbf{f}(\mathbf{y})\, dt + \sigma\, d\mathbf{w}(t), \tag{7.8}$$

where $\mathbf{f}$ is the drift of the target process, and $\sigma\, d\mathbf{w}(t)$ is isotropic noise in $\mathbb{R}^k$.[4]

RNNs are often trained through 'backpropagation through time' [297], a training procedure that is known to be computationally expensive, especially for stochastic data, and numerically unstable, often suffering from vanishing and exploding gradients [298]. To avoid this, we introduce the *drift-diffusion matching (DDM)* framework, an approach that directly aligns the drift and diffusive terms in the affine subspace to the target SDE.

First, we assume that Γ has full rank k, thus it has an MP inverse such that $\Gamma^\dagger \Gamma = \mathbf{I}_k$, where $\mathbf{I}_k$ is the identity in $\mathbb{R}^k$. Using Itô's rule [56], we have that the drift in the affine subspace is given by,

$$\widehat{\mathbf{f}}(\mathbf{y}) = \Gamma^\dagger \left(\mathbf{F}(\Gamma \mathbf{y} + \mathbf{b})\right), \tag{7.9}$$

whilst the diffusion matrix is,

$$\Sigma = \Gamma^\dagger B B^\top (\Gamma^\dagger)^\top. \tag{7.10}$$

However, simply matching $\widehat{\mathbf{f}} = \mathbf{f}$ and $\Sigma = \sigma^2 \mathbf{I}_k$, will not guarantee that projections of paths from the RNN will behave as paths from the target SDE. This is due to the presence of the drift and diffusion terms in directions orthogonal to the affine subspace. In order to guarantee that DDM results in an RNN whose projected dynamics recover paths from the target SDE, we must ensure there is a one-to-one relationship between the latent point $\mathbf{y}(t)$ and the high-dimensional $\mathbf{u}(t)$.

Each point in the trajectory can be uniquely decomposed as,

$$\mathbf{u}(t) = \mathbf{u}_{||}(t) + \mathbf{u}_{\perp}(t), \tag{7.11}$$

where $\mathbf{u}_{||} = \Gamma \mathbf{y} + \mathbf{b} \in \mathcal{A}$, whilst $\mathbf{u}_\perp \perp \mathrm{Im}(\Gamma)$ i.e. $\mathbf{u}_\perp$ is orthogonal to the image of Γ. As a result, $\Gamma^\dagger(\mathbf{u} - \mathbf{b}) = \mathbf{y}$ for all different $\mathbf{u}_\perp \perp \mathrm{Im}(\Gamma)$, thus it is not injective.

[4] For simplicity, we will assume that our target process has isotropic diffusion in the chapter. Nevertheless, extending DDM to (spatially-constant) non-isotropic diffusion is simple.

To prevent dynamics in directions that are not 'visible' under the projection, we enforce a *low-rank parametrisation* of the RNN,

$$W = \Gamma W_s, \qquad I = \Gamma I_s + \mathbf{b}, \qquad B = \Gamma B_s, \tag{7.12}$$

where $W_s \in \mathbb{R}^{k\times n}$, $I_s \in \mathbb{R}^k$, $B_s \in \mathbb{R}^{k\times d}$, where we recall that d is the number of independent noise sources. This parameterisation ensures that, for any $\mathbf{u} \in \mathcal{A}$, we have that,

$$\mathbf{F}(\mathbf{u}) = -\mathbf{u} + \Gamma W_s \mathbf{h}(\mathbf{u}) + \Gamma I_s + \mathbf{b} \in \mathrm{Im}(\Gamma), \tag{7.13}$$

where $\mathbf{h}(\mathbf{u}) = (h(u_1), \ldots, h(u_n))$. Moreover, we have that $\Gamma B_s \, d\mathbf{w}(t) \in \mathrm{Im}(\Gamma)$. Together, these imply that, given $\mathbf{u}(0) \in \mathcal{A}$, the drift and diffusion are tangent to the subspace, and thus $\mathbf{u}(t) \in \mathcal{A}$ for all $t > 0$ i.e the affine subspace is *invariant* under the dynamics. When the network state begins outside the subspace, the leak term, $-\mathbf{u}$, will cause the process to drift to the origin along the directions orthogonal to $\mathrm{Im}(\Gamma)$.

Under this parametrisation, the dynamics in the affine subspace simplify significantly,

$$\widehat{\mathbf{f}}(\mathbf{y}) + \mathbf{y} = W_s \mathbf{h}(\Gamma \mathbf{y} + \mathbf{b}) + I_s, \tag{7.14}$$

where we have moved the leak term to the left-hand side (LHS) for convenience. Here, we notice the RHS is equivalent to a two-layer *perceptron* with a k-neuron input layer, an n-neuron hidden layer, and a k-neuron output layer. The input-to-hidden weights and biases are given by Γ and $\mathbf{b}$ respectively, whilst the hidden-to-output weights and biases are given by W_s and I_s, and where $\mathbf{h}(\cdot)$ plays the role of the element-wise activation function. The diffusion matrix in the affine subspace also simplifies to $B_s B_s^\top/2$.

Remark 7.2.1 (Multilayer perceptrons)**.** A *multilayer perceptron* is simply a feed-forward deep neural network with a nonlinear activation function [18]. More explicitly, given an input $\mathbf{x} \in \mathbb{R}^m$, an L-layer perceptron is given by,

$$\begin{aligned} \mathbf{\Lambda}^{(0)} &= \mathbf{x}, \\ \mathbf{\Lambda}^{(l)} &= \mathbf{h}(\mathbf{W}^{(l)}\mathbf{\Lambda}^{(l-1)} + \mathbf{b}^{(l)}), \qquad \text{for } l = 1, \ldots, L \\ \mathbf{z} &= \mathbf{\Lambda}^{(L)} \end{aligned} \tag{7.15}$$

where $\mathbf{\Lambda}^{(l)}$ is the output of the $l-$th layer, $(\mathbf{W}^{(l)}, \mathbf{b}^{(l)})$ is the weight matrix and bias vector between the layer $l-1$ and layer l, respectively. The final output is $\mathbf{z} \in \mathbb{R}^q$ where q is the number of neurons in the output layer.

Given some target function $g : \mathbb{R}^m \to \mathbb{R}^q$, training describes the procedure of finding the set of weights and biases $\{(\mathbf{W}^{(l)}, \mathbf{b}^{(l)})\}_{l=1}^{L}$ that minimises the difference, according to a particular loss function, between the parameterisation defined by the perceptron and the target. It is well-known that this parametrisation, including for $L = 2$ and $h(x) = \tanh(x)$, is a *universal approximator* i.e. it can approximate any continuous function to arbitrary accuracy [299]. Moreover, it can be efficiently optimised via backpropagation of error (BOE) [18].[5]

Given that the projected drift is parametrised by a perceptron, we can set up the DDM loss-function to be,

$$\mathcal{L}_{\text{DDM}} = ||\mathbf{f}(\mathbf{y}) + \mathbf{y} - W_s\mathbf{h}(\Gamma\mathbf{y} + \mathbf{b}) - I_s|| + \lambda_{\text{diff}}||\sigma^2\mathbf{I}_k - B_sB_s^\top||, \tag{7.16}$$

where λ_{diff} is a hyper-parameter controlling the relative importance of the drift and diffusion losses, and where we optimise over the parameters $\{\Gamma, \mathbf{b}, W_s, I_s, B_s\}$, for a total of $2nk + n + k + kd$ free parameters.

Fig. 7.4 illustrates the DDM framework. We have shown that training a low-rank RNN is equivalent to training a two-layer perceptron, which can be optimised with BOE to approximate the target drift and diffusion [Panels *a*) and *b*)]. Using these parameters, we can reconstruct the RNN and sample paths of neural activity. When these paths are projected into the learnt latent space, they behave like paths from the target SDE [Panel *c*)], thus the RNN has been trained to encode a particular target dynamics in a neural manifold.[6]

7.2.1 A symmetric-asymmetric decomposition

Symmetric networks define latent (generalised) gradient flows Previously, we noted that the drift of an RNN with symmetric connectivity can be written in the form,

$$\mathbf{F}(\mathbf{u}) = -\mathbf{D}(\mathbf{u})\nabla_{\mathbf{u}}\widetilde{E}(\mathbf{u}), \tag{7.17}$$

i.e. as a generalised gradient flow that monotonically descends the energy landscape, defined in Eq. (7.3), to a local minimum. Now we focus on the case of a low-rank, but

[5] In the past few decades, these architectures and their training have become extremely well-known in ML. We will not introduce them in further detail, nor will we review the proofs, but an interested reader may consult Ref. [18].

[6] In this case, the neural manifold is simply an affine subspace, thus we could also call it a *neural subspace*.

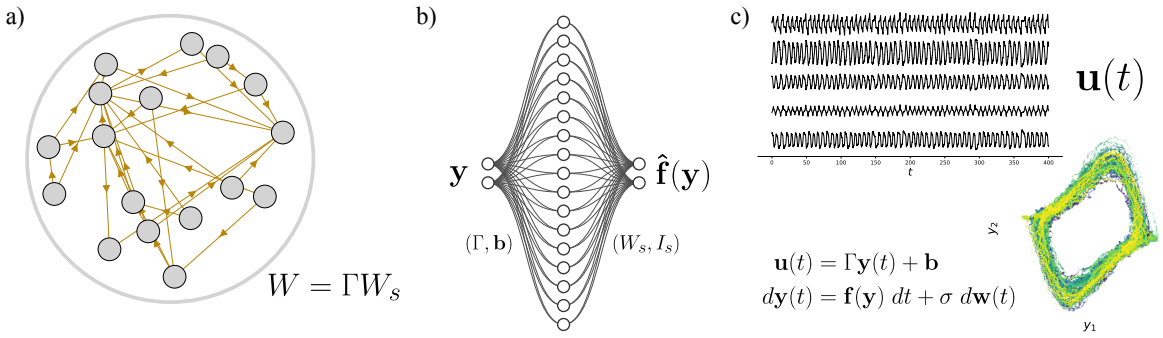

Figure 7.4: **Drift-diffusion matching framework**. *a*) We assume that our RNN has a low-rank structure $W = \Gamma W_s$. *b*) Its projected drift is exactly parametrised by a two-layer perceptron, which we can train to approximate our target drift using BOE *c*) Taking the parameters of the perceptron, we can reconstruct our RNN and sample paths. When projected into the latent space, we recover paths from the target SDE, illustrated here with the VDP oscillator of Ex. 5.3.1 i.e. $k = 2$, and five randomly selected neurons from the network with $n = 64$.

symmetric, network projected onto the affine subspace, $\mathcal{A}$, given the parametrisation in Eq. (7.12).

We define an additional energy function on $\mathcal{A}$, $S(\mathbf{y}) = \widetilde{E}(\Gamma\mathbf{y} + \mathbf{b})$ whose gradient is therefore $\nabla_{\mathbf{y}} S(\mathbf{y}) = \Gamma^\top \nabla_{\mathbf{u}} \widetilde{E}(\Gamma\mathbf{y} + \mathbf{b})$. The dynamics in the affine subspace are given by,

$$\widehat{\mathbf{f}}(\mathbf{y}) = -\Gamma^\dagger \mathbf{D}(\Gamma\mathbf{y} + \mathbf{b})(\Gamma^\dagger)^\top \nabla_{\mathbf{y}} S(\mathbf{y}) = -\mathbf{K}(\mathbf{y}) \nabla_{\mathbf{y}} S(\mathbf{y}), \tag{7.18}$$

where $\mathbf{K}(\mathbf{y})$ is positive definite i.e. the dynamics in $\mathcal{A}$ also admit an energy function.[7] As a result, a symmetric RNN induces an energy potential over the subspace and precludes limit-cycles or chaotic attractors. Therefore, any dynamics which go beyond monotonic decreases in energy must be encoded via the asymmetry of the connectivity matrix.

A symmetric-asymmetric decomposition of an RNN The previous observation motivates a decomposition of a trained RNN into a symmetric and asymmetric component. We consider $W = W_{\text{sym}} + W_{\text{asym}}$, and define,

$$\widehat{\mathbf{f}} = \widehat{\mathbf{f}}_{\text{sym}} + \widehat{\mathbf{f}}_{\text{asym}}, \quad \widehat{\mathbf{f}}_{\text{sym}}(\mathbf{u}) = -\mathbf{u} + W_{\text{sym}}\mathbf{h}(\mathbf{u}) + I, \quad \widehat{\mathbf{f}}_{\text{asym}}(\mathbf{u}) = W_{\text{asym}}\mathbf{h}(\mathbf{u}), \tag{7.19}$$

where the symmetric component, $\widehat{\mathbf{f}}_{\text{sym}}$, descends an energy landscape, whilst the asymmetric component, $\widehat{\mathbf{f}}_{\text{asym}}$, encodes rotational dynamics. Whilst it is reminiscent of the

[7] As the system is not precisely the gradient of an energy potential, it is more accurate to call this a *Lyapunov* function.

HHD of a vector field or SDE, it is not identical to either, as the symmetric component is neither a conservative field, nor is it equal to the time-reversible component of the stochastic process.

At first glance, it is tempting to use the unique decomposition of the network into a symmetric and antisymmetric component,

$$W = \underbrace{\frac{1}{2}\left(W + W^\top\right)}_{C} + \frac{1}{2}\left(W - W^\top\right), \tag{7.20}$$

as W_{sym} and W_{asym}. However, with this decomposition, the drift components, $\widehat{\mathbf{f}}_{\text{sym}}$ and $\widehat{\mathbf{f}}_{\text{asym}}$, are no longer confined to the affine subspace. As a result, the projected dynamics of the symmetric network will still admit an energy function on $\mathbb{R}^k$, but paths are no longer confined to the subspace. This means that the projected dynamics of the symmetric component of the network do not look 'attractor-like' over the subspace and, in general, the projected dynamics do not converge to the minimum of the projected energy function $S(\mathbf{y})$. In order to guarantee that the paths are confined to the subspace, we require that $W_{\text{sym}} = \Gamma A$ for some A. Even if $W = \Gamma W_s$, this does not hold in general, thus this decomposition is not as intuitive as it first appears.

To tackle this problem, we split the network as,

$$W = \Gamma(\Omega + \Pi), \tag{7.21}$$

where $\Gamma\Omega = (\Gamma\Omega)^\top$ i.e. it is symmetric. Let us note that $\Gamma\Pi$ is asymmetric, unless $W = W^\top$, but not antisymmetric, except in the special case when this decomposition coincides with Eq. (7.20). Such a decomposition is clearly not unique, thus we define the 'best' decomposition to be,

$$\Omega = \text{argmin}_{A \in \mathbb{R}^{k \times n}} ||\Gamma A - C||_F^2, \tag{7.22}$$

i.e. our symmetric part is the most similar to the canonical decomposition, with respect to the Frobenius norm. This problem has the closed form solution (see App. D.1),

$$\Omega = \Gamma^\dagger C \Gamma \Gamma^\dagger. \tag{7.23}$$

This improved decomposition, which we denote the *symmetric-asymmetric decomposition (SAD)*, gives two drift components, both constrained to the affine subspace, and where $\widehat{\mathbf{f}}_{\text{sym}}$ descends the energy function $S(\mathbf{y})$ associated with the network $W_{\text{sym}} = \Gamma\Omega$.

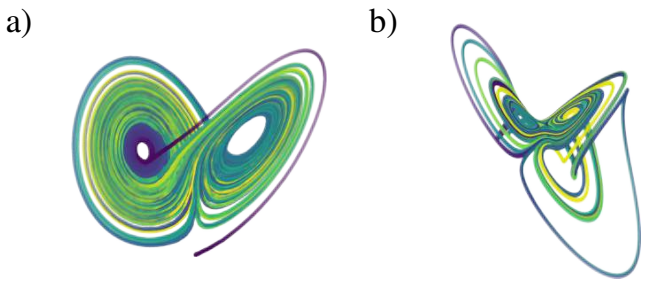

Figure 7.5: **Trajectories from chaotic systems.** *a*) A trajectory from the Lorenz attractor in Ex. 7.2.1. *b*) A trajectory from the Dadras attractor in Ex. 7.2.2.

7.2.2 Examples and numerical experiments

To illustrate the DDM framework and the network decomposition, we first consider three well-known nonlinear systems with additive noise. First, we have the VDP oscillator, introduced in Ex. 5.3.1. Next, we introduce two three-dimensional chaotic systems. Further details on parameters and numerical implementation is given in App. D.5.

Example 7.2.1. (The Lorenz attractor) The stochastic *Lorenz attractor (LA)* is a three-dimensional chaotic system given by,

$$d\begin{pmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \end{pmatrix} = \begin{pmatrix} \varsigma(y_2 - y_1) \\ y_1(\rho - y_3) - y_2 \\ y_1 y_2 - \beta y_3 \end{pmatrix} dt + \sigma d\begin{pmatrix} w_1(t) \\ w_2(t) \\ w_3(t) \end{pmatrix}. \tag{7.24}$$

The system has a characteristic strange attractor with two scrolls when $\varsigma = 10$, $\rho = 28$, and $\beta = 8/3$ [221]. Panel *a*) of Fig. 7.5 shows a trajectory from the attractor for $\sigma = 0$.

Example 7.2.2. (The Dadras attractor) The stochastic *Dadras attractor (DA)* is a three-dimensional chaotic system given by,

$$d\begin{pmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \end{pmatrix} = \begin{pmatrix} y_2 - py_1 + oy_2y_3 \\ ry_2 - y_1y_3 + y_3 \\ cy_1y_2 - ey_3 \end{pmatrix} dt + \sigma d\begin{pmatrix} w_1(t) \\ w_2(t) \\ w_3(t) \end{pmatrix}. \tag{7.25}$$

The system has a strange attractor with three scrolls when $p = 3$, $o = 2.7$, $r = 1.7$, $c = 2$, and $e = 9$ [300]. Panel *b*) of Fig. 7.5 shows a trajectory from the attractor for $\sigma = 0$.

Numerical experiments To illustrate the DDM framework and the network decomposition, we train an RNN with 64, 512, and 1024 neurons for the VDP, LA, and

DA systems respectively, using the DDM framework. Using the fitted parameters, we sample paths from the RNN as an autonomous SDE. Projecting the dynamics into the learnt affine subspace, we recover paths from each of the target SDEs. Next, we decompose the RNN connectivity with the SAD as described in Sec. 7.2.1, where we can investigate which features of the dynamics are encoded in each part of the network respectively.

Fig. 7.6 shows example traces from five randomly selected neurons in each of the RNNs, alongside the phase-plane dynamics in the learnt latent space. We see that the system is able to encode all three nonlinear systems and that the paths of individual neurons reflect the qualitative structure of the embedded attractor.

Fig. 7.7 shows the results of the decomposition for the VDP, LA, and DA examples. This decomposition gives us two vector fields, one of which has an associated energy function. Panels *a*-*c*) show both components for each system, whilst Panels *d*-*e*) show characteristic isopotentials of the energy function for the LA and DA systems. It appears that for the VDP and DA systems, the symmetric component of the network encodes attracting dynamics that maintain the process near the attractor, whilst the asymmetric component drives nonlinear rotation in phase-space. On the other hand, for the LA, the energy gradient appears to drive the process away from the origin, with the asymmetric, rotational component preventing it from diverging.

7.2.3 Reconstructing the subspace from paths

Another pertinent question is whether or not we can reconstruct the appropriate affine subspace given only paths from the RNN. When the dynamics that are entirely constrained to an affine subspace, it is possible to exactly identify the subspace, up to numerical accuracy. The solution comes from a seminal result by Pearson, and is known as *principal component analysis (PCA)* [301]. Additionally, if the path data has measurement error or noise in the form of isotropic Gaussian variation in the directions orthogonal to the subspace – this is different to the noise in the stochastic dynamics – then it is not entirely constrained and an exact recovery is not possible, but PCA becomes the MLE of the subspace. Subspace identification through PCA is extremely common in the analysis of neural data and has been used in countless studies to identify neural manifolds in brain network activity [295, 302, 303], as well as to identify latent subspaces in the activity of trained RNNs [290, 304].

This method is standard and the proofs have been reviewed at length (e.g. Ref. [305]), thus we will only outline the approach in brief, and illustrate how it applies to our setting. Given trajectory data of the form $\{\mathbf{u}(t_j) : j = 1, \ldots, T\}$, we assume the data

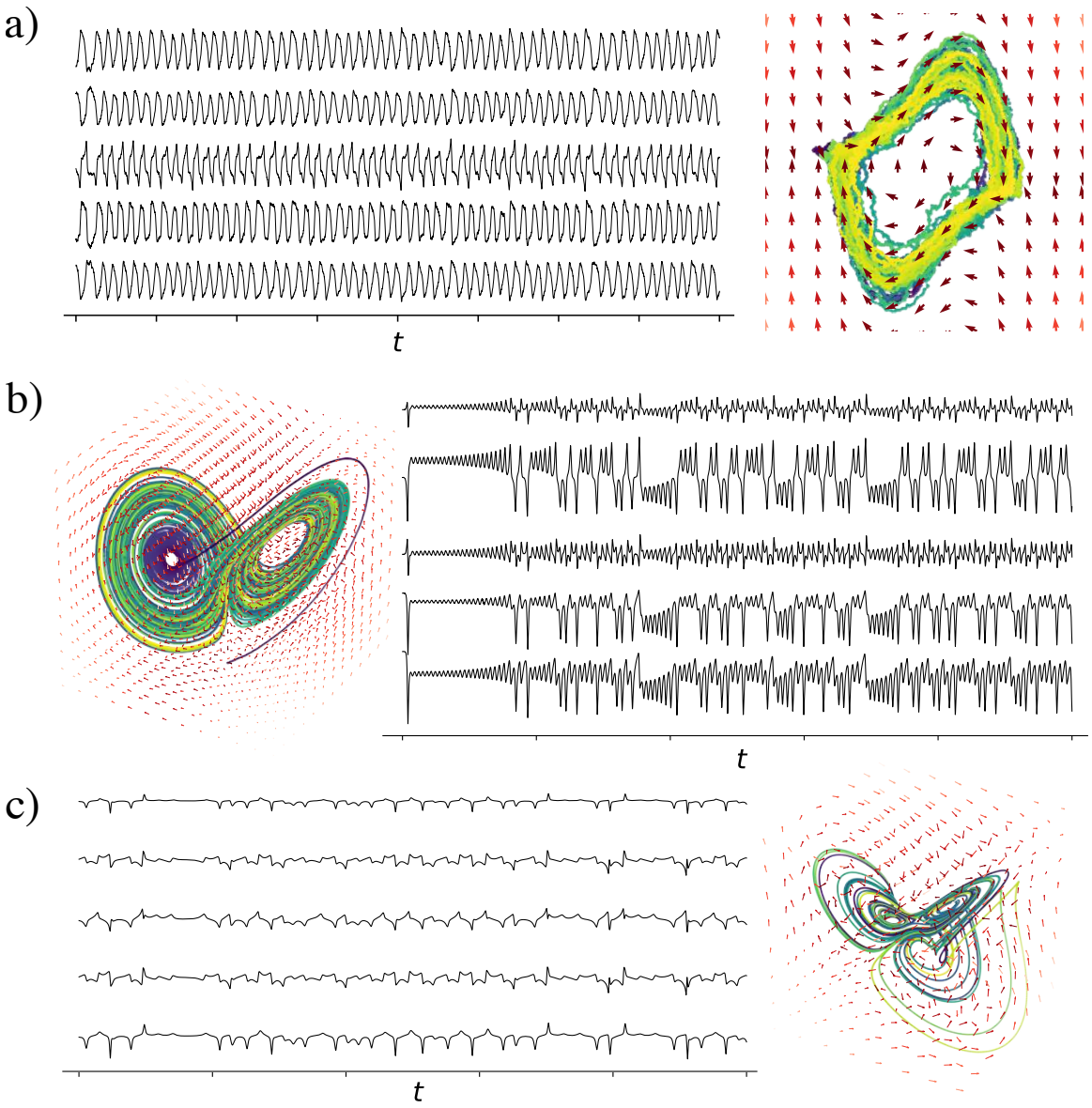

Figure 7.6: **Embedding attractors in neural networks.** *a*) Traces from 5 randomly selected neurons in a 64-neuron RNN trained to embed a stochastic VDP oscillator in a latent subspace. In the subspace, we see the characteristic VDP limit-cycle. $b-c$) Traces from a 512-neuron RNN trained to embed the LA, and a 1024-neuron RNN trained to embed the DA.

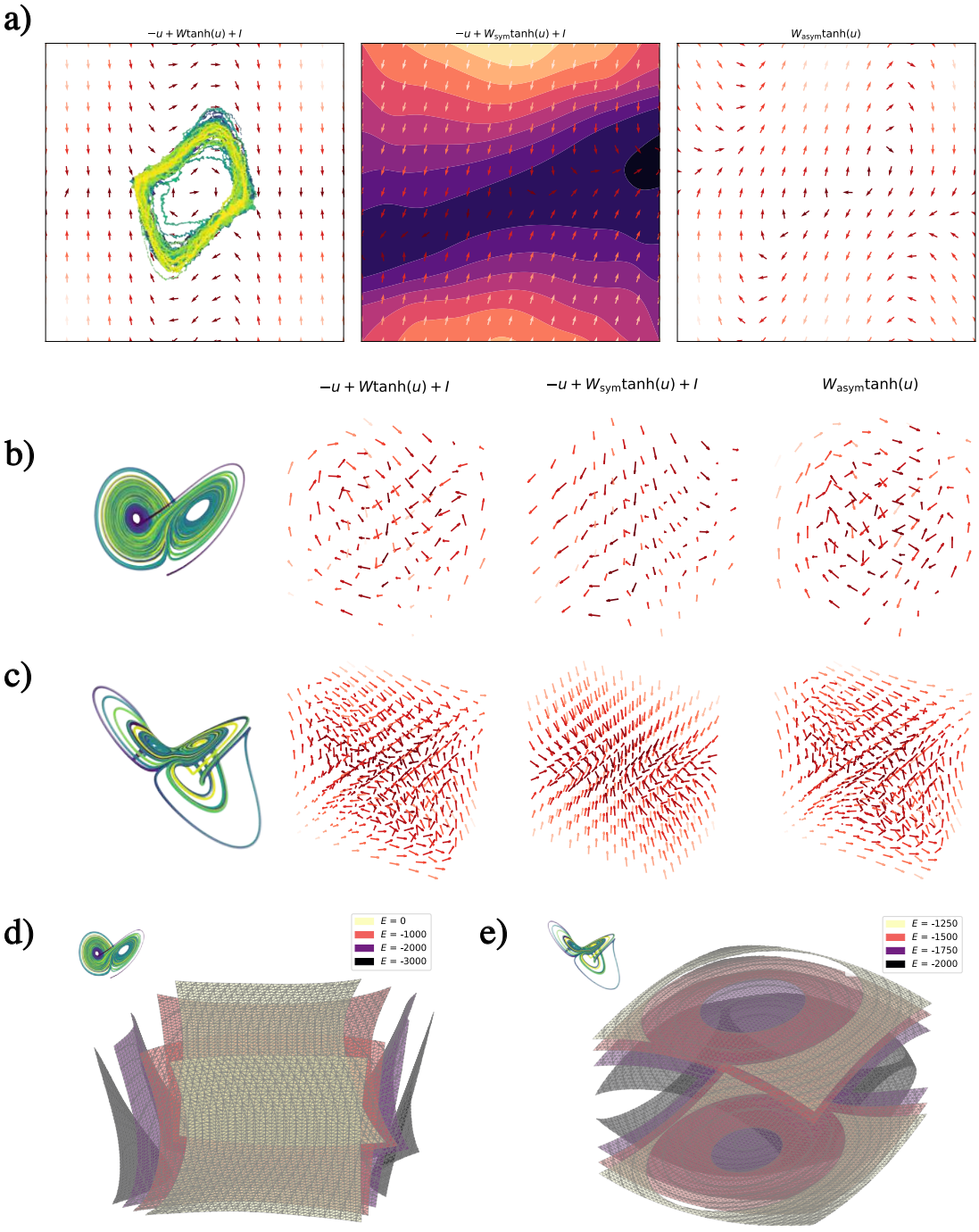

Figure 7.7: **Symmetric-asymmetric decomposition of RNNs.** *a*) Decomposition of RNN trained to encode the VDP. The symmetric dynamics descend the contoured energy function. This appears to keep the process near the limit cycle, whilst the asymmetric component provides rotational dynamics. $b-c$) Decompositions of RNNs trained to encode the LA and DA. For the DA, the symmetric dynamics keep the process near the attractor, whilst the asymmetric component provides rotational dynamics. For the LA, it appears the symmetric component pushes the dynamics away from the origin, and the asymmetric component provides a restoring force. $d-e$) This is confirmed by plotting isopotentials of the energy functions for the LA and DA respectively.

lives in an affine subspace of dimension k, where k typically begins as an unknown, and can be written in the form,

$$\mathbf{u} = \Psi \boldsymbol{\upsilon} + \boldsymbol{\mu}, \tag{7.26}$$

where $\Psi \in \mathbb{R}^{n\times k}$ has orthonormal columns, $\boldsymbol{\mu} \in \mathbb{R}^n$, and $\boldsymbol{\upsilon}(t) \in \mathbb{R}^k$ is a parametrisation. Whilst it is tempting to associate Ψ with Γ and $\boldsymbol{\mu}$ with $\mathbf{b}$, we note that this is not necessarily the case. Whilst the subspace can be identified, the exact parametrisation cannot be as, for any invertible matrix A and constant vector $\mathbf{c}$, we have indistinguishable parametrisations,

$$\mathbf{u} = \Gamma\mathbf{y} + \mathbf{b} = \underbrace{(\Gamma A^{-1})}_{\widetilde{\Gamma}}\underbrace{(A(\mathbf{y}-\mathbf{c}))}_{\widetilde{\mathbf{y}}} + \underbrace{\mathbf{b}+\Gamma\mathbf{c}}_{\widetilde{\mathbf{b}}} = \widetilde{\Gamma}\widetilde{\mathbf{y}} + \widetilde{\mathbf{b}}. \tag{7.27}$$

Thus, the parametrisation we obtain is equivalent to the original parametrisation, up to an affine transformation.

We stack our time-series into the matrix $\mathbf{U} \in \mathbb{R}^{n\times T}$ and center it, $\mathbf{X} = \mathbf{U} - \bar{\mathbf{u}}$, where,

$$\bar{\mathbf{u}} = \frac{1}{T}\left(\sum_j u_1(t_j), \ldots, \sum_j u_n(t_j)\right). \tag{7.28}$$

This is equivalent to choosing $\boldsymbol{\mu} = \mathbf{0}$. Next, we take the singular value decomposition (SVD) of $\mathbf{X}$. For long paths, it is more efficient to compute the covariance matrix, $\Theta = \mathbf{X}\mathbf{X}^\top$ and compute its eigendecomposition. Assuming that the dynamics are exactly constrained to the subspace, then this matrix will have k positive eigenvalues, whose eigenvectors form the columns of Ψ, which we call the *principal components (PCs)*, and where we order the columns by decreasing eigenvalue magnitude, $\{\lambda_1, \ldots, \lambda_k\}$ and $\lambda_j = 0$ for $j \in \{k+1, \ldots, n\}$.[8] The zero eigenvalues correspond to the directions orthogonal to the subspace, in which the dynamics do not evolve.[9] We then project the time-series to obtain the parameterisation,

$$\boldsymbol{\upsilon} = \Psi^\top \mathbf{X}. \tag{7.29}$$

Panels *a*) and *b*) of Fig. 7.8 show the traces and phase-space dynamics of the 3 non-trivial PCs for a trajectory from the 512-neuron RNN trained to embed the LA.

[8] Covariance matrices are symmetric positive semi-definite, thus have non-negative eigenvalues.

[9] For noisy data, these eigenvalues are non-zero but very small. In this case, the dimension of the subspace, k, can be approximately identified by considering the *proportion of the variance explained* by each PC, $\vartheta_l = \lambda_l / \sum_{j=1}^n \lambda_j$, where we 'cut-off' after k components, if $\sum_{j=1}^k \vartheta_j > c$, where c is a threshold e.g. $c = 0.7$ [305].

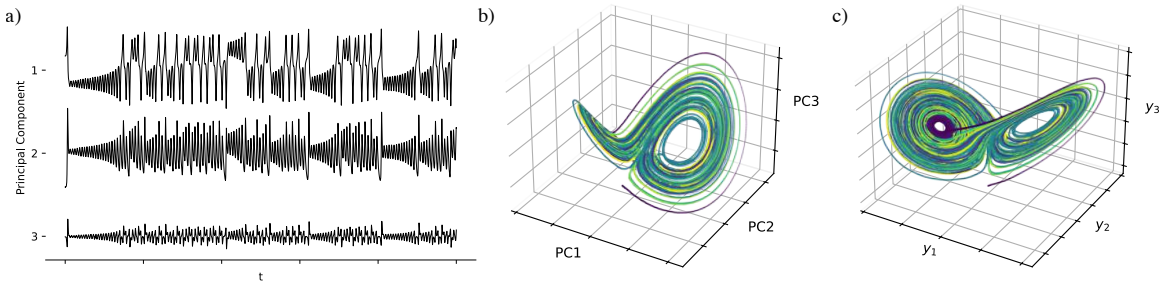

Figure 7.8: **Reconstructing the latent subspace from paths**. *a*) Paths from a low-rank RNN trained to embed a $k-$dimensional system only have k non-zero principal components. *b*) With PCA, we can identify the affine subspace in which the dynamics are constrained. *c*) The PCA representation is equivalent to original coordinates up to an affine linear transformation. Given $(\Gamma, \mathbf{b})$ can solve for this transformation to recover the original coordinates.

Given Γ and $\mathbf{b}$, we can find the affine transformation that turns $\boldsymbol{\upsilon}(t)$ to $\mathbf{y}(t)$. We can solve $\Gamma\Xi = \Psi$ and $\Gamma\boldsymbol{\chi} = \bar{\mathbf{u}} + \mathbf{b}$ as a least-squares problem, and apply the transformation $\mathbf{y}(t) = \Xi\boldsymbol{\upsilon}(t) + \boldsymbol{\chi}$, to obtain the original parametrisation, as shown in Panel *c*) of Fig. 7.8.

7.3 Diffusions and networks that switch and cycle between attractors

Attractor-like dynamics are frequently cited as both emergent features in neural data, and as a possible mechanism for computation via dynamics [276, 277]. These attractors are most easily understood in terms of energy landscapes, where the dynamics perform some kind of gradient descent, arriving at a local minima in the form of a fixed point or other low-dimensional manifold e.g. a line, a ring, or a torus [Fig. 7.3]. On the other hand, in neural systems, asymmetry and non-reciprocity are ubiquitous, typically disrupting the dynamics associated with symmetric models and their energy landscapes [213, 306]. Designing learning systems that utilise such dynamics has, so far, been elusive.

Remark 7.3.1. A range of 'learning' mechanisms, in both neuroscience and ML, rely on attractor-like dynamics over energy landscapes, including stochastic gradient descent [18], the 'free-energy' principle [307], and energy-based models [308].

Remark 7.3.2. Our construction and its applications to chaotic systems draw interesting, but unexplored, parallels to the work of Brunton et al. [309], in which chaotic dynamics are represented as a linear system driven by a nonlinear forcing term.

We now introduce two classes of SDEs that evolve over an underlying energy potential and thus have attractor-like dynamics, which we encode with the DDM framework. First, we introduce a gradient-flow system with additional inputs that allow the SDE to *switch* between the minima of the energy potential. At the level of the RNN, the inputs become 'network inputs' that represent sensory information or the output of an upstream neural circuit. These inputs prompt the RNN to retrieve a particular memory in light of additional 'information'. In Hopfield's model, the initial network state serves as the input and the system cannot transition between memories, thus our approach represents an extension.

Next, we introduce a class of stationary nonequilibrium diffusions that use irreversible, rotational currents to cycle between energy minima in a preferred direction. In this case, our model attempts to describe the ability of a memory system to autonomously transition between a series of related memories in a preferred direction. We show that both these dynamics can be encoded within an RNN, as illustrative examples of how a neural circuit could encode more complex forms of associative memory.

7.3.1 Input-driven, attractor-switching dynamics

We now consider an extended RNN of the form,

$$\begin{aligned} d\mathbf{u}(t) &= \left(\mathbf{F}(\mathbf{u}) + G\mathbf{s}(t)\right)\, dt + B\, d\mathbf{w}(t), \\ F_i &= -u_i + \sum_j W_{ij} v_j + I_i + (\Gamma \mathbf{d})_i\,, \end{aligned} \tag{7.30}$$

where we have an additional *input* term $G\mathbf{s}(t) + \Gamma\mathbf{d}$, where $\mathbf{s}(t) \in \mathbb{R}^l$ is the low-dimensional input signal, and $G \in \mathbb{R}^{n\times l}$ represents a connectivity matrix that projects the input into the RNN. In Eq. (7.30), we assume that the additional bias term is low-rank i.e. $\Gamma\mathbf{d}$. Moreover, we assume that our input connectivity is also low-rank and of the form $G = \Gamma G_s$. As a result, the dynamics remain constrained to the affine subspace.

Next, we construct the target process. We assume that the RNN, in the absence of inputs, encodes a gradient-flow on an energy potential, $V : \mathbb{R}^k \to \mathbb{R}$, with a number of minima of approximately equal energy, and isotropic noise,

$$d\mathbf{y}(t) = -\nabla V(\mathbf{y})\, dt + \sigma\, d\mathbf{w}(t), \tag{7.31}$$



Figure 7.9: **'Tilting' the energy landscape.** Our approach uses network inputs to 'tilt' the energy landscape, encouraging the process to push a path into a particular energy minimum, by making it a global minimum. This is a soft version of the child's 'labyrinth' game, where the board can be tilted such that the ball rolls into a specific hole – though in the actual game, one tries rather hard to avoid them.

Photo by user Alexas_Fotos, reproduced under a Pixabay License.

which is in an ESS, given by the Boltzmann distribution (see Ex. 2.2.1). Including the network inputs, the dynamics in the affine subspace are given by,

$$d\mathbf{y}(t) = (-\nabla V(\mathbf{y}) + G_s\mathbf{s}(t) + \mathbf{d})\ dt + \sigma\, d\mathbf{w}(t). \tag{7.32}$$

For simplicity, we focus on a constant input over time,

$$\mathbf{c} = G_s\mathbf{s} + \mathbf{d}. \tag{7.33}$$

Here, we see that the drift can be written as the gradient of a modified potential function,

$$\mathbf{f}(\mathbf{y}) = -\nabla[V(\mathbf{y}) - \mathbf{c}\cdot\mathbf{y}]. \tag{7.34}$$

This modified potential illustrates that adding a drift term that is constant in $\mathbf{y}$, corresponds to 'tilting' the potential to prioritise a particular minimum, making it a global minimum.

Remark 7.3.3. This approach can be understood by analogy with a child's 'labyrinth' game [Fig. 7.9], where the objective is to tilt the board, which serves as the energy potential, such that the ball rolls into a particular hole i.e. the chosen energy minimum.

Fitting the inputs Next, we must pick a collection of input values $\{\mathbf{c}_j\}_{j=1}^m$, where each corresponds to a minimum j, and causes the potential to tilt in this direction. Following this, we must choose a way to represent each minimum via our network inputs $\mathbf{s}$, which we call the 'coding scheme'. Finally, we can fit the input connectivity and bias G_s and $\mathbf{d}$ with *ridge regression* [310], such that they project each coded input, $\mathbf{s}_j$, to the corresponding $\mathbf{c}_j$.

We assume that the potential has a set of m deep, well-separated minima, $\{\boldsymbol{\mu}_j\}_{j=1}^m$. For each minimum i^* there is a convex subset $\mathcal{C}_{i^*} \subset \mathbb{R}^k$, where $\mathbf{c} \in \mathcal{C}_{i^*}$ makes $\boldsymbol{\mu}_{i^*}$ the global minimum. More explicitly, defining $V_j^0 = V(\boldsymbol{\mu}_j)$, if

$$\mathbf{c} \cdot (\boldsymbol{\mu}_i^* - \boldsymbol{\mu}_j) > V_{i^*}^0 - V_j^0, \tag{7.35}$$

for all $j \neq i^*$, then $\boldsymbol{\mu}_i^*$ is the global minimum. In theory, we can select any value $\mathbf{c} \in \mathcal{C}_{i^*}$, but in practice we would like to choose one, so we solve the following quadratic convex problem,

$$\begin{aligned} \mathbf{c}_i^* &= \text{argmin}_{\mathbf{c}\in\mathbb{R}^k} \frac{1}{2}||\mathbf{c}||_2^2, \\ \text{s. t. } \ \mathbf{c} \cdot (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) &\geq V_i^0 - V_j^0 + \delta, \ \text{for } j \neq i, \end{aligned} \tag{7.36}$$

where $\delta > 0$ is a small parameter used to weaken the strict inequality.[10]

The remainder of the procedure depends on the coding scheme, thus we will illustrate this with a pair of examples.

Example 7.3.1 (Gaussian wells)**.** Before we introduce coding schemes and the fitting procedure for the input-driven RNN, we will introduce a simple example of a potential, which will allow us to construct flexible examples. In the remainder of this chapter, we choose the energy potential to be a sum of *Gaussian wells*,

$$V(\mathbf{y}) = -\sum_{j=1}^{m} a_j \exp\left(-\frac{||\mathbf{y} - \boldsymbol{\mu}_j||^2}{2\nu_j^2}\right), \tag{7.37}$$

where the m minima can be placed at arbitrary positions $\boldsymbol{\mu}_j \in \mathbb{R}^k$, and we can specify the depth, a_j, and breadth, ν_j, of each well. In the following two examples, we will we place four identical minima at points in the plane [Fig. 7.10 & 7.11].

Example 7.3.2 ('One-hot' inputs and a four-well potential)**.** In this example, we consider a four-well potential in $\mathbb{R}^2$. For each of the four minima, we compute $\mathbf{c}_i^*$

[10] This can be solved efficiently using standard convex optimisation algorithms. In practice, we use `cvxpy`. We take $\delta = 0.125$.

as in Eq. (7.36). We choose a one-hot encoding, which is the simplest scheme and is extremely common in ML [18]. This takes $l = k = 4$ and defines the inputs corresponding to each well to be,

$$\mathbf{s}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \qquad \mathbf{s}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \qquad \mathbf{s}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \qquad \mathbf{s}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \tag{7.38}$$

i.e. $\mathbf{s}_i = \mathbf{e}_i$ is the input representing the minimum i, where $\mathbf{e}_i$ is the unit vector in the direction i.

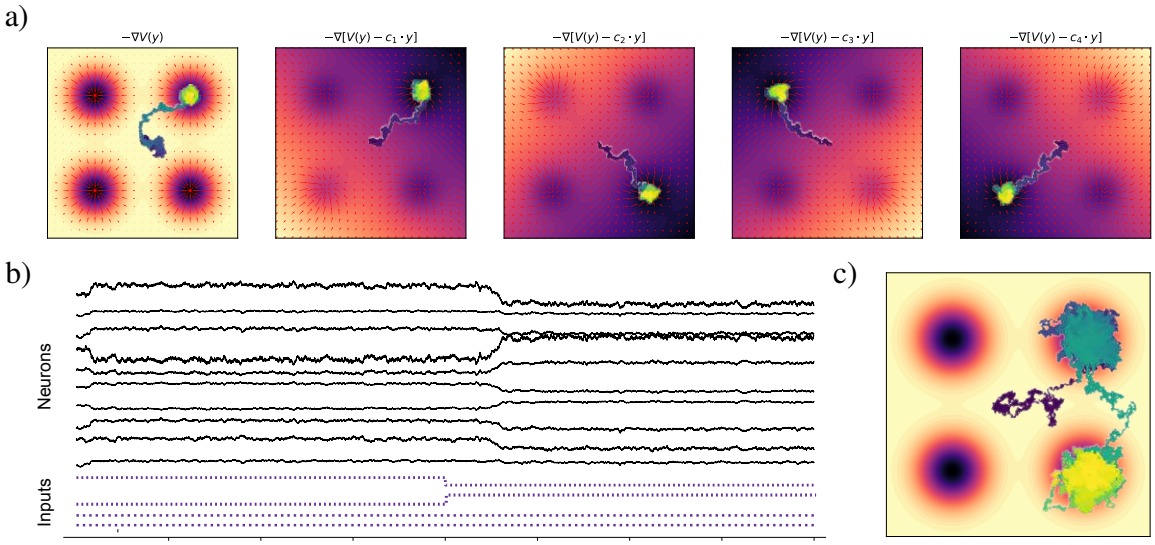

Figure 7.10: **Input-driven attractor-switching with one-shot coding.** *a*) We consider a four-well energy potential in the plane. In the absence of inputs, the paths converge to a minimum at random. Using the inputs, we tilt the potential towards each of the minima in turn. *b*) We train a 256-neuron RNN to embed the gradient-flow dynamics and fit the inputs. We plot traces from 10 random neurons under a piecewise constant input that switches from one minimum to another, and see that the RNN dynamics react to this change. *c*) Projecting the RNN path onto the latent space, we see the attractor-switching that occurs due to the change in input.

Next, we want to solve for G_s and $\mathbf{d}$, such that they map each input $\mathbf{s}_i$ to $\mathbf{c}_i^*$. We use ridge regression, which is given by the optimisation problem,

$$\min_{G_s, \mathbf{d}} \sum_{i=1}^{m} ||G_s \mathbf{s}_i + \mathbf{d} - \mathbf{c}_i^*||_2^2 + \lambda ||G_s||_F^2, \tag{7.39}$$

where λ is called the *ridge penalty*[11] [310]. This problem has an exact solution, which we give in App. D.2.

[11] We take $\lambda = 10^{-6}$.

We train a 256-neuron RNN using DDM to approximate the gradient-flow of the potential i.e. we embed the SDE from Eq. (7.31) into the RNN. A path from the RNN, in the absence of further inputs, will converge to a minimum at random, as shown in the first case of Panel *a*) in Fig. 7.10. When the network is perturbed by the input,

$$G\mathbf{s}_i + \Gamma\mathbf{d} = \Gamma G_s \mathbf{s}_i + \Gamma\mathbf{d}, \tag{7.40}$$

its dynamics over the affine subspace converge to the corresponding minimum i, as shown in Panel *a*) of Fig. 7.10, for each of the minima in turn. Panel *a*) also shows a contour of the modified potential, which illustrates how the network input creates the desired tilting effect.

We can also consider a piecewise-constant input, where the network first converges to one minimum and then to another. Panel *b*) shows a sample path from 10 randomly selected neurons, as well as the four-dimensional input channel. Halfway through the path, the input switches from $\mathbf{s} = \mathbf{s}_1$ to $\mathbf{s} = \mathbf{s}_2$, causing the network dynamics to shift between attractors. Panel *c*) shows the dynamics in the affine subspace, where the path first converges to minimum 1 before switching to minimum 2.

Example 7.3.3 (Random Gaussian inputs with additive noise)**.** We consider the same four-well potential as in the previous example. Instead of a one-hot encoding, we represent each minimum with a (quenched) random Gaussian vector. We also add noise to the input signal. More explicitly, we take

$$\widetilde{\mathbf{s}}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \qquad\qquad \mathbf{s}_i = \frac{\widetilde{\mathbf{s}}_i}{||\widetilde{\mathbf{s}}_i||}. \tag{7.41}$$

We then train the RNN and solve for G_s and $\mathbf{d}$ as before. The RNN receives inputs subject to Gaussian noise i.e. $\mathbf{s}(t) = \mathbf{s}_i + \varsigma_t$, where $\varsigma_t \sim \mathcal{N}(\mathbf{0}, \varsigma^2\mathbf{I})$.

Fig. 7.11 shows the dynamics of ten randomly selected neurons from the RNN, as well as the input channels using a Gaussian coding scheme, in the presence of additive noise [Panel *a*)]. If the noise intensity is small relative to the difference between the Gaussian input codes, then the network behaves as desired, and switching the input results in a switching between the minima in the latent space [Panel *b*)].

The set-up considered in this section represents a simple, but effective, computational model of associative memory. Instead of considering the initial state as the system's input, as in Hopfield's model, we instead integrate a specific input signal that is projected into the network. This can be interpreted as sensory information or

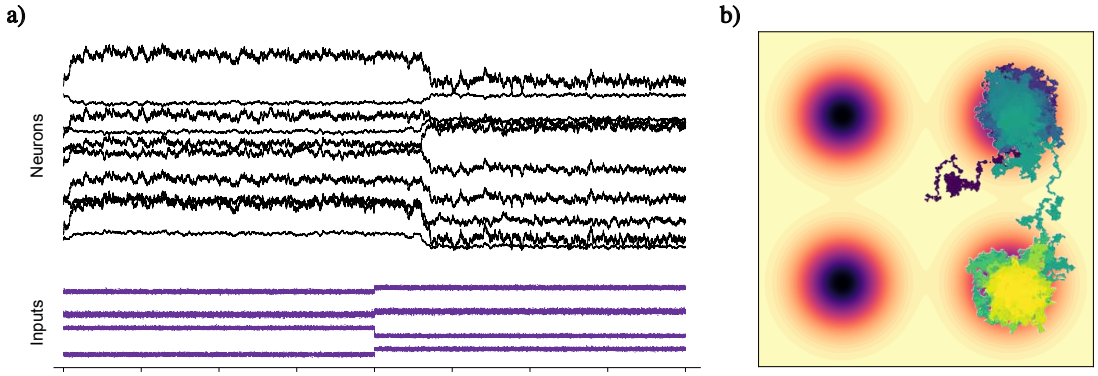

Figure 7.11: **Attractor-switching with random Gaussian codes.** *a*) We fit the input-coupling using random Gaussian input codes, and then perturb the network with these inputs subject to additive noise. The network is able to differentiate between the inputs and switch between the attractors. *b*) Projecting the RNN path onto the latent space, we see the expected attractor-switching behaviour.

the output of a previous neural computation from an upstream circuit, which drives the system to a particular memory. In response to this input of information, the RNN converges to the *appropriate* memory i.e. the global energy minimum.

7.3.2 Attractor-cycling, irreversible dynamics

In this section, we design a class of planar SDEs, which not only converge to energy minima, but also cycle between these minima in a preferred direction using irreversible currents. To begin with, we take the gradient-flow of an energy potential and add a rotational component, $\mathbf{R}(\mathbf{y})$, to the drift. Thus, we consider a process of the form,

$$d\mathbf{y}(t) = (-\nabla V(\mathbf{y}) + \mathbf{R}(\mathbf{y}))\ dt + \sigma\, d\mathbf{w}(t). \tag{7.42}$$

However, if we add an arbitrary, or poorly designed, rotational component, then this will disrupt the gradient-flow and erase the attracting dynamics. To prevent this, we enforce that the Eq. (7.42) must still converge to the Boltzmann distribution associated with V i.e. the steady-state when $\mathbf{R} \equiv \mathbf{0}$. With this motivation, we derive the following result.

Lemma 7.3.1. *The SDE in Eq. (7.42) admits a stationary Boltzmann density,*

$$\pi \propto \exp\left(-\frac{2V}{\sigma^2}\right), \tag{7.43}$$

if and only if $\nabla \cdot (\mathbf{R}\pi) = 0$ *i.e.* $\mathbf{R}$ *is DFSD.*[12]

[12] A function g with $\nabla \cdot g = 0$ is *divergence-free*. A function g is DFSD (with density π) if $\nabla \cdot (g\pi) = 0$. DFSD is defined in Sec. 2.2.2.2, in the context of the HHD.

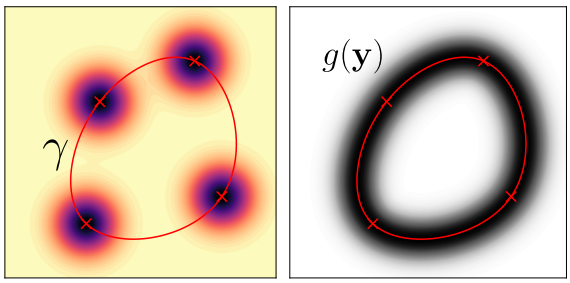

Figure 7.12: **Constructing rotational planar diffusions.** To construct an SDE with rotation between energy minima, we begin with the gradient of the energy potential. We then draw a closed curve, γ, through the minima in the direction of rotation. From this, we can define a rotational drift which decays with distance from the curve due to the weighting kernel $g(\mathbf{y})$.

Proof. The Boltzmann distribution, π, is the stationary density if it solves the stationary FP equation,

$$\nabla \cdot [f\pi - \frac{\sigma^2}{2}\nabla\pi] = 0. \tag{7.44}$$

Substituting in the drift and the Boltzmann density on the LHS, we obtain,

$$\nabla \cdot [f\pi + \frac{\sigma^2}{2}\frac{2}{\sigma^2}(\nabla V)\pi] = \nabla \cdot [(-\nabla V)\pi + \mathbf{R}\pi + (\nabla V)\pi] = \nabla \cdot (\mathbf{R}\pi),$$

i.e. π solves the solves the stationary FP if and only if $\nabla \cdot (\mathbf{R}\pi) = 0$. □

This condition is closely related to the HHD. In this case, we have simply that $\mathbf{f}_{\text{rev}} = -\nabla V$ whilst $\mathbf{f}_{\text{irr}} = \mathbf{R}$. By adding a rotational component that does not disrupt the Boltzmann density, we construct stationary diffusions that still descend the energy landscape whilst being driven rotationally by stationary probability flux, $J_{\text{ss}} = \mathbf{R}\pi$. As a result, this SDE is in an NESS. We will now focus on constructing planar examples of these rotational diffusion processes, which we can then embed into an RNN with the DDM framework.

7.3.2.1 Constructing rotational planar diffusions

We begin with a Gaussian potential, as in Ex. 7.3.1, on $\mathbb{R}^2$. Our goal is to design a rotational component that is DFSD. Notice that we can construct a differentiable, closed planar curve, γ, which connects points in the order that we would like the field to rotate [Fig. 7.12].

We want the rotational field to provide a force in the direction of this closed curve, acting along all concentric curves with the same shape. For example, if the curve is a circle, this is the set of circles with the same centre but different radii. To do this, we

consider $\gamma = \partial O$ to be the boundary of a bounded planar region $O \subset P$, and define the *signed distance* to be,

$$\phi(\mathbf{y}) = \begin{cases} -\text{dist}(\mathbf{y}, \gamma) & \text{if } \mathbf{y} \in O \\ \text{dist}(\mathbf{y}, \gamma) & \text{if } \mathbf{y} \in P \backslash O \end{cases} \tag{7.45}$$

where

$$\text{dist}(\mathbf{y}, \gamma) = \inf_{\mathbf{p} \in \partial O} ||\mathbf{y} - \mathbf{p}||. \tag{7.46}$$

This defines a potential where each level-set is a closed curve with the same shape as γ, up to an inflation or deflation. We would like our rotational field to be tangent to these level sets. We obtain such a field by rotating $\nabla\phi$ by $\pi/2$ via the antisymmetric matrix $\mathbf{M}$, given by,

$$\mathbf{M} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{7.47}$$

To prevent the rotational field from dominating the gradient dynamics far from the minima, we also introduce a weighting kernel,

$$g(\mathbf{y}) = \exp\left(-\frac{\phi^2(\mathbf{y}, \gamma)}{2\xi^2}\right), \tag{7.48}$$

which decays exponentially with the distance from the closed curve γ at a rate inversely proportional to ξ^2 [Fig. 7.12]. Putting together the components, we have the rotational vector field,

$$\mathbf{R}(\mathbf{y}) = \frac{\alpha}{\pi(\mathbf{y})} g(\mathbf{y}) \mathbf{M} \nabla\phi(\mathbf{y}), \tag{7.49}$$

where α is the scalar strength of rotation, and where $\mathbf{R}$ is DFSD (see App. D.3 for a proof). This gives a combined field,

$$\mathbf{f} = -\nabla V + \mathbf{R}, \tag{7.50}$$

that has both attractive and cycling dynamics.

7.3.2.2 Examples and numerical experiments

We now turn to concrete examples, with three, four, and five minima, in the plane, and then embed these SDEs in RNNs. For the potential, we choose Gaussian wells (Ex. 7.3.1). In the case of the three minima, we choose the curve γ to be the intersecting circle, whilst for four and five minima we compute it numerically with cubic

splines.[13] Fig. 7.13 illustrates the example processes for a) three, b) four, and c) five minima, including an example path. The paths transiently explore the stationary distribution, spending significant time at each attractor, and transitioning between them in the direction of the stationary flux. Further details on parameters and numerical implementation is given in App. D.5.

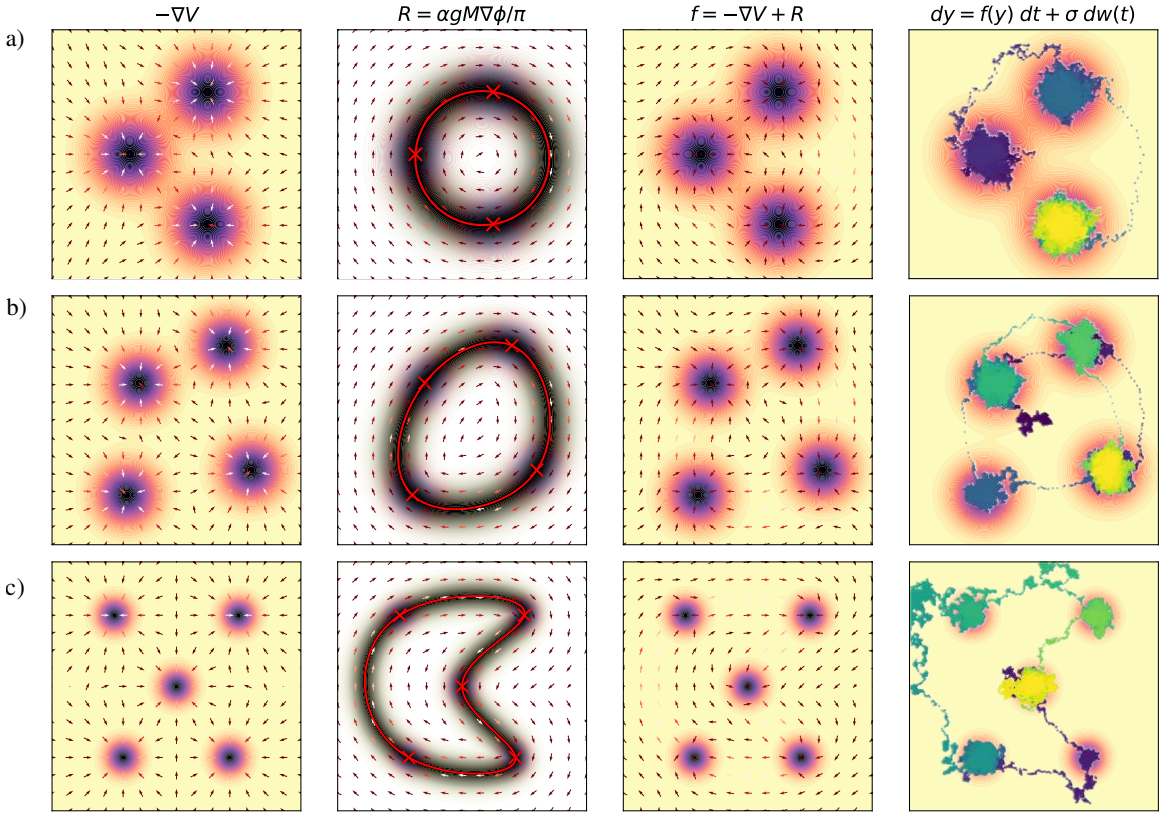

Figure 7.13: **Attractor-cycling planar diffusions.** We construct example diffusion processes that cycle a) three, b) four, and c) five attractors with irreversible currents. These are composed of a gradient flow (first column) plus a rotational component (second column), which includes a decaying kernel g (contoured). Combined, this defines a vector field that both converges to, and cycles between attractors (third and fourth column).

Next, we use DDM to train three different RNNs, each with 128 neurons, to approximate each example in a latent affine subspace. Once the model is trained, we sample paths from the RNN and project them into the learnt subspace. Fig. 7.14 shows example traces from ten randomly selected neurons in each network, as well as a

[13] We compute interpolating closed curves with cubic splines with periodic boundary conditions. More specifically, we use `CubicSpline` from `scipy.interpolate`. We then parametrise the curve, $\gamma(t)$, with 500 time-points and construct a `matplotlib` path. We can then measure the distance between any point, p, and each discrete point in the path, taking the minimum to be an approximation of the distance between p and γ.

projection of each path and drift field onto the affine subspace, where we can see that the RNN is able to encode the autonomous attractor-cycling dynamics. Moreover, as in the case of the nonlinear systems [Fig. 7.6], the activity of the individual neurons reflects the character of the encoded process i.e. here we see neuronal activity jumping between piecewise-approximately-constant segments.

7.4 Neural networks in the NESS

Now that we have embedded an SDE which is in an NESS into an RNN, it is worth considering the steady-state properties of the network itself. Computing properties of the NESS for a high-dimensional network is usually challenging, except in the case of the simple, solvable processes considered in Ch. 3. However, in this case where we have a high-dimensional RNN with an embedded solvable, stationary SDE, the situation simplifies significantly.

We consider, as before, an RNN with state $\mathbf{u}(t) \in \mathcal{A}$, where the latent dynamics follow the SDE,

$$d\mathbf{y}(t) = \mathbf{f}(\mathbf{y})\, dt + \sigma\, d\mathbf{w}(t), \tag{7.51}$$

which has a stationary density $\pi_{\mathbf{y}}(\mathbf{y})$ over $\mathbb{R}^k$. From this, we can derive the stationary density, $\pi_{\mathbf{u}}(\mathbf{u})$ over $\mathbb{R}^n$, which is given by the *image measure* [311],[14]

$$\pi_{\mathbf{u}}(\mathbf{u}) = \frac{\pi_{\mathbf{y}}(\Gamma^\dagger(\mathbf{u}-\mathbf{b}))}{\sqrt{\det(\Gamma^\top\Gamma)}}. \tag{7.52}$$

Assuming that the HHD of the latent process is given by $\mathbf{f} = \mathbf{f}_{\text{rev}} + \mathbf{f}_{\text{irr}}$, we can decompose the network drift, $\mathbf{F}$, as,

$$\begin{aligned} \mathbf{F} &= \mathbf{F}_{\text{rev}} + \mathbf{F}_{\text{irr}}, \\ \mathbf{F}_{\text{irr}}(\mathbf{u}) &= \Gamma \mathbf{f}_{\text{irr}}(\Gamma^\dagger(\mathbf{u}-\mathbf{b})), \\ \mathbf{F}_{\text{rev}}(\mathbf{u}) &= \Gamma \mathbf{f}_{\text{rev}}(\Gamma^\dagger(\mathbf{u}-\mathbf{b})), \end{aligned} \tag{7.53}$$

which additionally implies that the flux and EPR also converge (see App. D.4).

We illustrate this by returning to the example process with three minima given in Sec. 7.3.2, and train a 128-neuron RNN to embed the process. Given that we have the HHD of the latent process, $\mathbf{f}_{\text{rev}} = -\nabla V$ and $\mathbf{f}_{\text{irr}} = \mathbf{R}$, we can define a decomposition of the high-dimensional dynamics, using Eq. (7.53), in the ambient space, and sample

[14] From this point on, we are implicitly constraining $\mathbf{u} \in \mathcal{A}$ to avoid Dirac deltas in the expression.

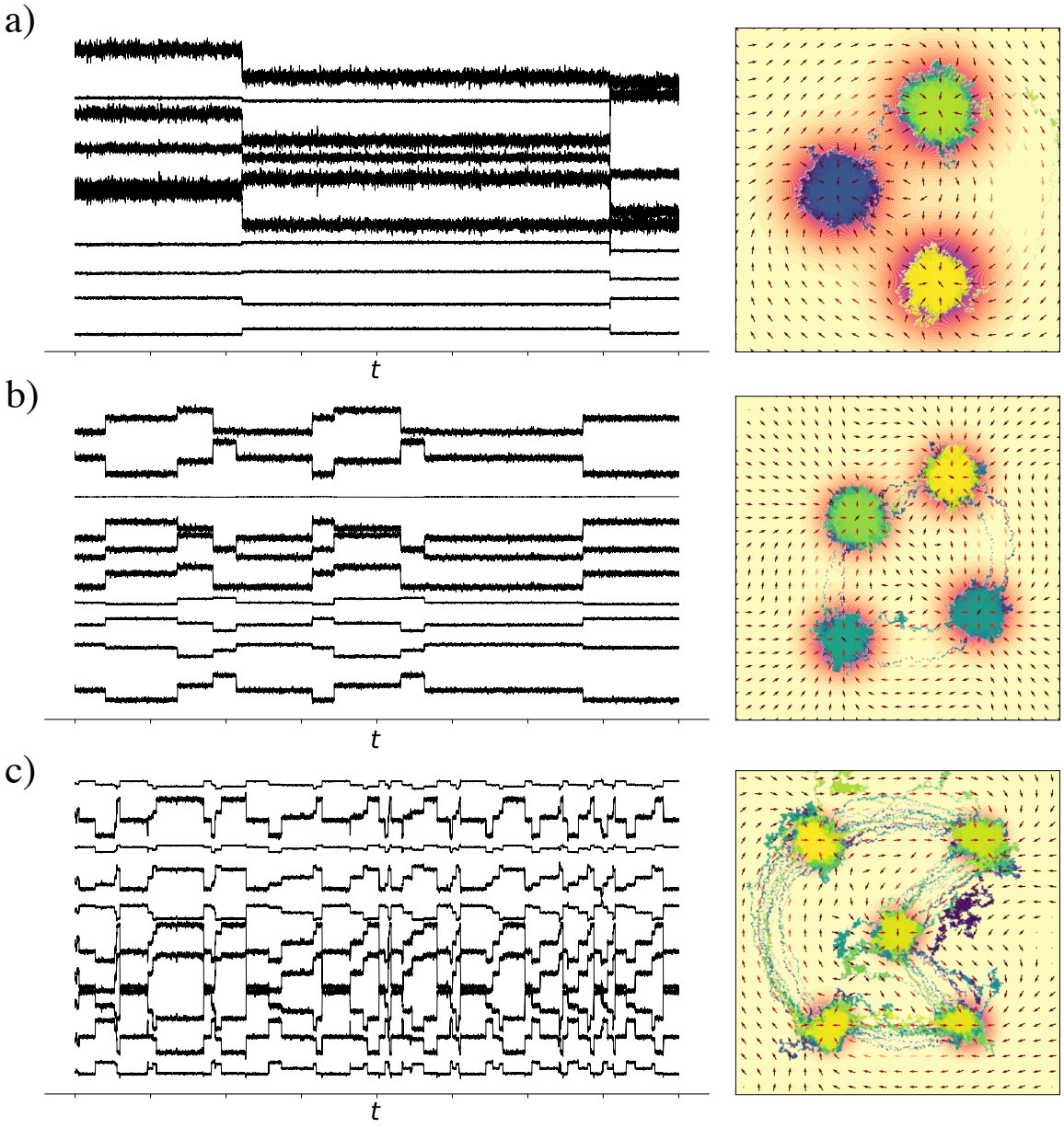

Figure 7.14: **Attractor-cycling in RNNs.** We train 128-neuron RNNs to approximate the planar diffusion processes from Fig. 7.13 for *a*) three, *b*) four, and *c*) five minima, respectively. We sample paths from the RNN, and project both the paths and the learnt drift onto the learnt affine subspace. For each RNN, we plot traces from ten random neurons. The attractor cycling dynamics are visible in the latent space, as well as in the traces.

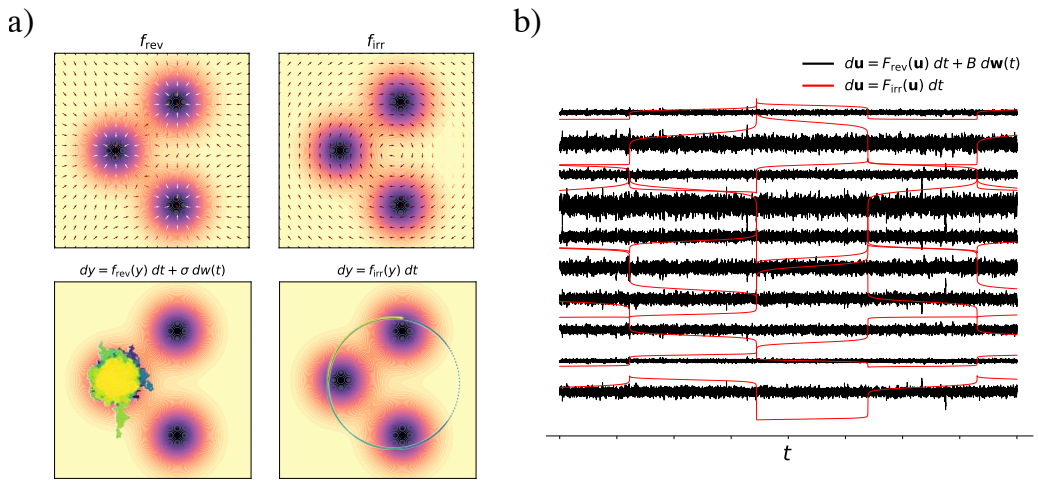

Figure 7.15: **The HHD of a latent, embedded process.** *a*) When we embed a process with a known NESS, we can perform the HHD. In this example with three minima, we see that the reversible dynamics balance the diffusion and maintain the process on the stationary distribution, whilst the rotational dynamics create rotation around the curve γ. *b*) We train an RNN with 128 neurons to approximate the process in a latent affine subspace. Using the HHD and Eq. (7.53), we can perform the HHD of the high-dimensional dynamics. We see that the reversible dynamics in the ambient space keep the process near the minima, balancing the noise, whilst the irreversible dynamics show limit-cycle behaviour in the ambient space.

paths from each component. Panel *a*) of Fig. 7.15 shows the HHD of the example with three minima, including a path sampled from each component. Panel *b*) shows a path sampled from each component of the HHD applied to the network in the high-dimensional ambient space. As with the dynamics at the latent level, we see that the reversible dynamics keep the process near the minima, balancing the noise, whilst the irreversible dynamics show limit-cycle behaviour.

7.4.1 A Helmholtz-Hodge decomposition for RNNs

Whilst Eq. (7.53) does give a decomposition of the high-dimensional network dynamics into two components, the decomposition is not directly related to the learnt network connectivity or input currents, only to the learnt affine subspace. As a result, each component is a high-dimensional vector field that we cannot visualise, and thus is challenging to interpret.

In order to turn this decomposition of a vector field into a decomposition of the

RNN itself, we assume that the components can be written in the form,

$$\begin{aligned}\mathbf{F}_{\text{irr}}(\mathbf{u}) &= -\varrho\mathbf{u} + W_{\text{irr}}h(\mathbf{u}) + I_{\text{irr}},\\ \mathbf{F}_{\text{rev}}(\mathbf{u}) &= -(1-\varrho)\mathbf{u} + W_{\text{rev}}h(\mathbf{u}) + I_{\text{rev}},\end{aligned} \tag{7.54}$$

i.e. each component is associated with a part of the network connectivity, bias, and leak term. Here we are assuming that $W = W_{\text{rev}} + W_{\text{irr}}$, $I = I_{\text{rev}} + I_{\text{irr}}$, and $\varrho \in [0, 1]$. We also assume that these parts respect the parameterisation,

$$\begin{aligned} W_{\text{rev}} &= \Gamma W_{s,\text{rev}}, & W_{\text{irr}} &= \Gamma W_{s,\text{irr}},\\ I_{\text{rev}} &= \Gamma I_{s,\text{rev}} + (1-\varrho)\mathbf{b}, & I_{\text{irr}} &= \Gamma I_{s,\text{irr}} + \varrho\mathbf{b},\end{aligned} \tag{7.55}$$

which ensures that each component is tangent to the affine subspace, and thus does not cause the process to drift away from it.

Eq. (7.54) implies a decomposition of the RNN itself into reversible and irreversible components, yet we have yet to describe how to solve for these components. First, we write the projected dynamics in the affine subspace,

$$\widehat{\mathbf{f}}_{\text{irr}}(\mathbf{y}) = -\varrho\mathbf{y} + W_{s,\text{irr}}h(\Gamma\mathbf{y} + \mathbf{b}) + I_{s,\text{irr}}, \tag{7.56}$$

$$\widehat{\mathbf{f}}_{\text{rev}}(\mathbf{y}) = -(1-\varrho)\mathbf{y} + W_{s,\text{rev}}h(\Gamma\mathbf{y} + \mathbf{b}) + I_{s,\text{rev}}, \tag{7.57}$$

where we notice that these can be seen as a pair of two-layer perceptrons, each approximating a component of the drift. However, the weights and biases of the first layer, $(\Gamma, \mathbf{b})$, are both shared and fixed. The *sum* of the weights and biases in the second layer are also fixed i.e. $W_s = W_{s,\text{rev}} + W_{s,\text{irr}}$ and $I_s = I_{s,\text{rev}} + I_{s,\text{irr}}$. Finally, we have a trainable parameter $\varrho \in [0, 1]$.

In order to obtain the components of the decomposition of the RNN, we must, again, train a two-layer perceptron, this time to approximate the irreversible component under these constraints: the weights and biases of the first layer are fixed; the 'twin' perceptron defined by the same parameters in Eq. (7.57) approximates the reversible component. More explicitly, we introduce the HHD-DDM loss-function,

$$\begin{aligned}\mathcal{L}_{\text{HHD-DDM}} = &||\mathbf{f}_{\text{irr}}(\mathbf{y}) + \varrho\mathbf{y} - W_{s,\text{irr}}\mathbf{h}(\Gamma\mathbf{y} + \mathbf{b}) - I_{s,\text{irr}}||\\ &+ ||\mathbf{f}_{\text{rev}}(\mathbf{y}) + (1-\varrho)\mathbf{y} - (W_s - W_{s,\text{irr}})\mathbf{h}(\Gamma\mathbf{y} + \mathbf{b}) - (I_s - I_{s,\text{irr}})||,\end{aligned} \tag{7.58}$$

where we optimise over $W_{s,\text{irr}}$, $I_{s,\text{irr}}$, and ρ; whilst $\{\Gamma, \mathbf{b}, W_s, I_s, B_s\}$ are fixed from training the original RNN.

Under these additional constraints, the two-layer perceptron is no longer guaranteed to be a universal approximator. However, in practice, we find that this parameterisation is able to approximate the components with very high accuracy. This is

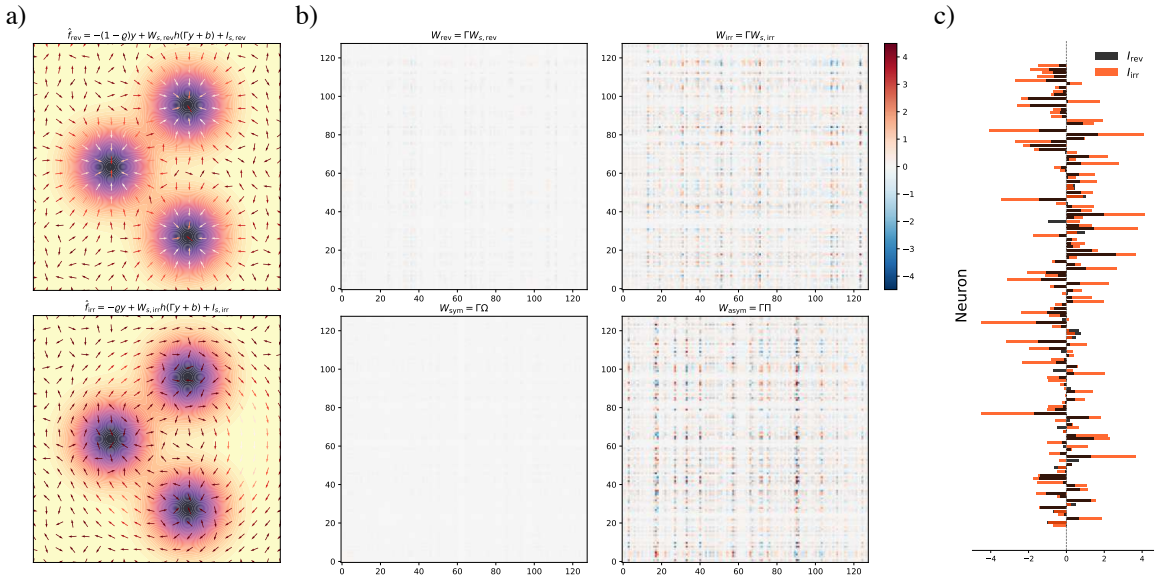

Figure 7.16: **The HHD of an RNN.** *a*) Using the HHD-DDM loss, we train twin perceptrons to approximate the components of the HHD of the latent process. *b*) We decompose W using both the HHD, $W = W_{\text{rev}} + W_{\text{irr}}$, and the SAD from Sec. 7.2.1, $W = W_{\text{sym}} + W_{\text{asym}}$. *c*) We also decompose the input current vector, $I = I_{\text{irr}} + I_{\text{rev}}$.

most likely due to the fact that $(\Gamma, \mathbf{b})$ are optimised for the full drift field and therefore remain well-aligned with this downstream 'fine tuning'. This procedure yields a HHD-like decomposition of the network structure that we can compare to the SAD from Sec. 7.2.1.

Using the same three-minima example as previously, we perform the HHD of the 128-neuron RNN. Panel *a*) of Fig. 7.16 shows the projected drift of each component of the RNN i.e. the vector fields learnt by the twin perceptron, which closely approximates the target components (see Fig. 7.15). Panel *b*) shows the HHD components (top-row) and SAD components (bottom row) of the connectivity. In both decompositions, we find that the irreversible/asymmetric connectivity has significantly larger weights, in absolute value, hinting that irreversible cycling dynamics require far higher network complexity. We also see that the decompositions do not necessarily align. Panel *c*) shows the decomposition of the RNN input current. As with the connectivity, the irreversible current is larger than the reversible current for most neurons (in absolute value), and their signs typically agree.

Does more irreversibility in the latent process imply more asymmetry? In Ch. 3, we showed that more asymmetric networks result in more irreversible dynamics, and that more irreversible neural dynamics could be explained by more asymmetric

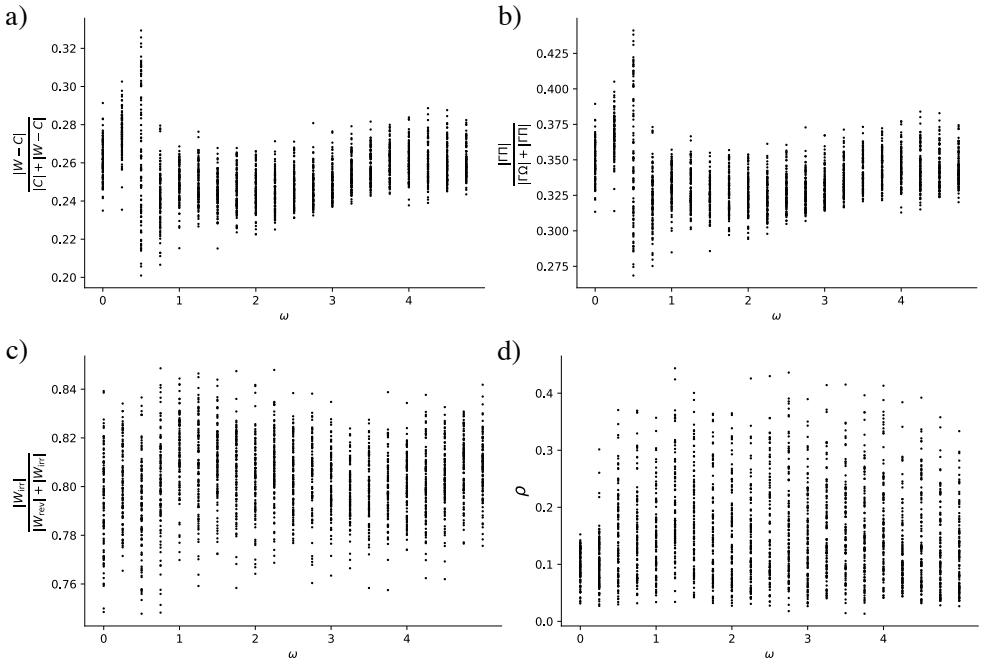

Figure 7.17: **Network asymmetry under increasing irreversibility.** For a range of values $\omega \in [0, 5]$, we train an ensemble of 100 RNNs. We then compute the relative energy for *a*) $(W - C, C)$, *b*) $(\Gamma\Pi, \Gamma\Omega)$, and *c*) $(W_{\text{irr}}, W_{\text{rev}})$. In *d*) we also plot ϱ from the network HHD in Eq. (7.54).

effective networks. The HHD of the RNN allows us to examine this hypothesis in a different context. We ask the question: does encoding a more irreversible latent process result in a more asymmetric RNN?

To investigate this, we return to the stochastic HO from Ex. 5.1.3. The EPR, Φ, and therefore the irreversibility, of this nonlinear, planar process increases quadratically with the frequency of rotation ω i.e. $\Phi \propto \omega^2$ (see Eq. (5.38)). We can split the RNN into symmetric and asymmetric parts using the three decompositions introduced in this chapter; namely the canonical symmetric-antisymmetric decomposition (Eq. (7.20)), the SAD (Eq. (7.21)), and the HHD (Eq. (7.54)).

Whilst we introduced a number of measures of asymmetry in Ch. 3, such measures are defined for positive, weighted networks, and thus are unsuitable for RNNs, which admit negative weights. Instead, we use the *relative energy*, which is defined to be,

$$\mathcal{E}(W_1, W_2) = \frac{||W_1||_F}{||W_1||_F + ||W_2||_F}, \tag{7.59}$$

for a pair (W_1, W_2) with $W = W_1 + W_2$.

For a range of different values $\omega \in [0, 5]$, we train an ensemble of 100 RNNs to encode the HO. We decompose the RNN connectivity with each of the three decom-

positions, and compute the relative energy of the asymmetric/irreversible component. Fig. 7.17 shows the relative energy of $(W - C, C)$ [Panel a)], $(\Gamma\Pi, \Gamma\Omega)$ [Panel b)], and $(W_{\text{irr}}, W_{\text{rev}})$ [Panel c)]. Panel d) shows the fitted value of ϱ from the network HHD in Eq. (7.54). We find that there is no clear relationship between the relative energy in the asymmetric/irreversible component and the EPR. This could suggest that irreversible currents are also parametrised in the bias terms, and that the projection matrix Γ induces asymmetries that are not directly related to the rotational forces in the target vector field. This result contrasts with the results of Ch. 3, which imply that the EPR is driven directly by network asymmetry – although this was not in the case of a low-rank system with low-dimensional dynamics and instead focused on positively-weighted networks.

Whilst it is clear that asymmetry is crucial for complex neural computation, in the case of RNNs we cannot draw the conclusion that the degree of asymmetry – and the effective hierarchy it induces – drives the degree of irreversibility, as we did in Ch. 3.

7.5 Conclusions

In this chapter, we have focused on the dynamics of asymmetric RNNs and their intersection with the theory of neural manifolds. In particular, we presented the DDM framework, which allows us to train RNNs to embed arbitrary latent SDEs in an affine subspace. We illustrated this by embedding a range of nonlinear processes. Next, we focused on applying our framework to extend Hopfield's model of associative memory. Specifically, we introduced an approach to train an RNN to encode a model of input-driven associative memory, where the network recovered persistent states based on external input signals. We further designed processes that autonomously cycled between attractors, and encoded them within RNNs. Finally, we presented two decompositions of the RNN connectivity – one based on asymmetry and one based on time-irreversibility – to investigate how the nonlinear dynamics were represented by the system.

These results extend our central theme of nonequilibrium brain network dynamics to the realm of neural computation. We leverage the framework of NESM to design asymmetric and irreversible RNNs, capable of encoding more complex dynamics than their symmetric counterparts. Our focus on low-rank networks and their dynamics in low-dimensional latent subspaces offers a direct link between RNNs and the theory of neural manifolds, building on a number of previous results and approaches [287, 290–

292, 312]. Our work underscores the ubiquity and necessity of asymmetric structure and nonequilibrium dynamics in models of neural computation that are capable of complex behaviour, and we offer a practical framework for further investigation.

Limitations and future directions Whilst interesting, our approach brings with it a number of limitations. First, unlike Hopfield's model which employs a Hebbian learning rule, the DDM framework trains RNNs via BOE. This makes training quick and easy, but is not biologically-motivated nor plausible. In this model, neural manifolds do not *emerge* spontaneously due to self-organisation, perhaps as the result of a learning rule. Instead, they are created by enforced structure in the connectivity. Most approaches enforce latent manifold structure through low-rank connectivity [291, 312], or by assuming certain symmetry properties [313, 314], but there have been some attempts to develop biologically-inspired learning rules for RNNs and other neural network models [274, 306, 315, 316]. Nevertheless, the mechanisms by which biological neural circuits *learn* to encode complex dynamics over neural manifolds remain unclear and require the development of learning rules which produce asymmetric connectivity and emergent latent structure.

Next, our framework encodes dynamics in an affine linear subspace, suggesting that the relationship between neuronal states and neural manifolds is linear. Whilst linear methods have proved useful when probing neural data [303], state-of-the-art approaches reveal neural manifolds with nonlinear methods [217, 287, 317]. A more general relationship between neural firing and the effective computational manifold allows for a much richer latent geometry in the system, which would be more consistent with empirical data. However, this produces major mathematical challenges to our framework as nonlinear maps introduce troublesome Jacobian and Hessian terms in the projected dynamics. In order to build a closer relationship between our framework and modern approaches to neural manifolds, future work should focus on the inverse problem of reconstructing low-dimensional vector fields or operators from observed trajectories over neural manifolds [11, 318], in line with some recent approaches that have attempted to simultaneously learn both the manifold and its dynamics [217, 319, 320].

Whilst we considered one case of an input-driven circuit, we mostly focus on autonomous RNN dynamics, which encoded stationary processes. Integrating an RNN within a more complete and time-dependent neural circuit, which includes both sensory input and an upstream output, is crucial for relating this model to further neuroscientific applications, such as motor control.

Finally, we focus on a 'firing-rate model', which assumes that information in neural circuits is encoded in the instantaneous firing rates of individual neurons [21]. This is in contrast with alternative models that focus on the binary dynamics of neurons, typically using spin-glass models [33], or spiking neural networks, which have biophysically-consistent dynamics for each individual neuron [13] and typically assume that exact spike-times encode information.

Chapter 8

Discussion

In this thesis, we have approached neural systems from the perspective of complex systems and statistical physics. We placed a particular focus on techniques from *nonequilibrium* statistical physics, modelling the brain and neural recordings as a stochastic system in a nonequilibrium steady-state (NESS), building on recent interest in this direction [3, 36–40]. Our main contributions can be summarised as follows:

- In Ch. 3, we focused on the network structural origins of NESSs. Introducing measures of directedness and using prototypical networked processes, we showed that hierarchically asymmetric structure drives networks out of equilibrium. Following this, we introduced a model-based approach to infer an asymmetric effective network and the entropy production rate (EPR) from functional neuroimaging data, which we illustrated on functional magnetic resonance imaging (fMRI) from humans during tasks and at rest. We found that the elevated EPR seen during tasks can be explained by a reorganisation of effective brain interactions into a more hierarchically asymmetric structure.

- In Ch. 4, we turned to the study of higher-order interactions (HOIs), a topic of burgeoning interest in complex systems [140]. In an effort to identify HOIs in neural recordings, we introduced the directed multiplex visibility graph irreversibility (DiMViGI) framework for measuring the irreversibility of all subinteractions in a complex system, from which we were able to identify those that are significant. We applied this approach to magnetoencephalography (MEG) data from participants engaging in an auditory-recognition task. Our framework revealed plausible interaction dynamics across levels, such as hemispheric symmetry, synergistic dynamics between sensory and processing regions, and a facilitating role for medial regions in auditory cognition.

- In Ch. 5, we shifted the focus from neural systems to the more general problem of approximating a nonequilibrium diffusion with discrete states. This was motivated by a number of empirical studies in biophysics that employed this approach [3, 174–176]. We began by deriving a continuous-time Markov chain (CTMC) from an stochastic differential equation (SDE) using a finite volume approximation (FVA), where we showed that the scheme yielded a valid master equation (MEq), and that the EPR converged in the limit. Next, we discussed practical approaches for inferring discrete-state models from path data, showing that this yielded a poor approximation of the EPR in practice. Nevertheless, we introduced an approach to test if a process is in an NESS, which we applied to polarisation data from schooling fish, concluding that their dynamics were not irreversible. In the second half of this chapter, we attempted to reconcile the discrete Helmholtz-Hodge decomposition (HHD) for signals on graphs and simplicial complexes [92], and the HHD of an SDE into reversible and irreversible components [55]. Centring on a result from Strang [64], we introduced a framework for decomposing an SDE directly from path data, where we could measure the relative size of the irreversible component. We validated this approach with numerical experiments before turning to data from flickering red blood cells (RBCs) and healthy and arrhythmic heartbeats. Our results showed higher irreversibility for healthy dynamics, consistent with previous results for physiological systems [180–182].

- In Ch. 6, we focused on oscillations in large-scale neural dynamics. Recent work has investigated the evidence for vortex-like dynamics in the human brain by using a continuous approximation for the cortex and extending basic methods from statistical mechanics [245, 246]. We proposed a mathematical framework for identifying vortices on *complex* networks, without a choice of cycle basis, using the normalised Hodge decomposition. We showed that this can be used to identify vortices and measure 'Hodge vorticity' on the lattice, using the XY model as an example. Using Monte-Carlo simulations, we showed evidence for a transition point in Hodge vorticity under increasing temperature on the lattice i.e. during the Berezinskii-Kosterlitz-Thouless phase transition (BKT-PT), and on various connectomes of different scales. Furthermore, using an expansion into eigenmodes of the 1-Laplacian, we introduced a measure of complexity for neural 'edge-signals', building on classical methods defined at the level of the nodes [85]. Finally, we applied our framework to human fMRI recordings,

showing higher vorticity and higher complexity in tasks compared to resting state.

- In Ch. 7, we studied recurrent neural networks (RNNs) as models of neural computation. In an effort to extend symmetric RNNs, such as Hopfield's model of associative memory [275], to nonequilibrium dynamics and the theory of neural manifolds [287], we introduced the drift-diffusion matching (DDM) framework. This allowed us to train RNNs to embed an arbitrary SDE in a latent affine subspace using basic machine learning (ML) techniques, which we illustrated with typical nonlinear systems. Next, we focused on developing models of *sequential* associative memory by designing target processes for the DDM framework. Specifically, we designed a class of processes that was able to switch between 'memories' due to an input signal given to the RNN and another that autonomously cycled between 'memories' due to irreversible currents. Finally, we introduced two decompositions of the RNNs based on both the asymmetry and the time-irreversibility of the network, in order to reveal how the dynamics were encoded in the connectivity matrices.

This thesis builds on a plethora of work that models information processing in the brain as a complex and dynamic network [19, 23]. Nonequilibrium statistical physics has often been posited as the principal mathematical framework for describing self-organisation in complex systems [4, 6], including in biology [41]. However, in the study of neural systems, it is mostly *equilibrium* approaches that have dominated the literature [33, 285]. This contradiction is explained by a number of basic facts: the Ising model is an elegant approximation of neural spike-trains; gradient-flows are an intuitive model of learning dynamics; and determining the direction of information flow in the human brain is challenging in both structural and dynamic data. Nevertheless, the maturing area of nonequilibrium brain dynamics provides increased evidence for the limitations of these equilibrium assumptions. Both within and beyond neuroscience, irreversible and stochastic thermodynamics are finding diverse and increasingly relevant applications, such as *diffusion models* in ML [321], which are the state-of-the-art in image generation, the study of the thermodynamic origins of life [322], and the thermodynamics of computation in both abstract [323] and biological systems [324].

Nevertheless, focusing on a nonequilibrium perspective introduces multiple mathematical and computational challenges that must be weighed against their advantages. Here we discuss a couple of important considerations:

- Whilst human structural brain data remain undirected, fitting asymmetric whole-brain models will either be restricted to oversimplified linear models, such as in Ch. 3, or rely on heuristic and inaccurate parameter inference approaches. Studies of structure-function coupling in the brain are at their best when they draw relationships between separate sources of structural imaging data and functional recordings. This can be marred by a focus on *effective* connectivity, where the structure is also derived from functional recordings – a potential example of 'double-dipping' [325].

- Detecting the presence of phase-space currents, which are central to the NESS, with a model-free approach often relies on a *reconstruction* of the stochastic process by either inferring the SDE itself, or by a discrete-state approximation. This is computationally intensive and requires the process to repeatedly visit all areas of the phase-space. As a result, such approaches are limited to low-dimensional dynamics, and are often unsuitable for neural data without aggressive dimensionality reduction [3]. It is for this reason that Ch. 5 fails to find applications to neural systems. More generally, the majority of methods designed to capture nonequilibrium fluctuations in biological time-series cannot be applied to neuroscientific data [43].

- The EPR has proved to be an interesting quantity that varies across conditions in neural recordings [3, 36]. Yet, at its core, it is a measure of heat dissipation that is most relevant when considering biophysical systems at the microscopic scale. Network neuroscience has no shortage of 'measures' derived from statistical mechanics, nonlinear dynamics, or information theory that vary between conditions, yet give only debatable neuroscientific insight [326, 327]. For a quantity as challenging to compute as the EPR, it is pertinent to consider whether or not it is worth the trouble. In this light, some studies in systems neuroscience implicitly take a nonequilibrium perspective, such as inferring asymmetric hidden Markov models [328, 329] or directed functional networks [330], with little to no mention of irreversibility, entropy, or physics.

As of yet, we are far from a complete mathematical theory of how the brain performs neural computation, cognition, and consciousness. In this thesis, we conducted a number of studies with a focus on nonequilibrium stochastic brain network dynamics at a range of different scales and from a number of perspectives. We considered both theoretical models of neural systems and practical frameworks for the analysis of

neural data, whilst trying to integrate this work within the wider scope of complexity in biological systems. There is little doubt that the dynamics of the brain are irreversible and out of equilibrium. It is yet to be decided how to best use this property and how fundamental it will be to furthering our understanding of the brain.

Appendix A

Appendices for Ch. 3

A.1 Proof of Lemma 3.1.1

We begin with the irreciprocity. We denote the reciprocity of A by $r\,(= r(1))$. The irreciprocity of the interpolated network is given by,

$$1 - r(\epsilon) = 1 - \frac{\sum_{i,j} \min(\hat{A}_{ij}(\epsilon), \hat{A}_{ji}(\epsilon))}{\sum_{i,j} \hat{A}_{ij}(\epsilon)}, \tag{A.1}$$

$$= \frac{\sum_{i,j} \hat{A}_{ij}(\epsilon) - \min(\hat{A}_{ij}(\epsilon), \hat{A}_{ji}(\epsilon))}{\sum_{i,j} \hat{A}_{ij}(\epsilon)}. \tag{A.2}$$

The interpolation preserves the row sums (even excluding the diagonal), so the denominator, which we denote k, is the same as in the irreciprocity of the original network A.

$$1 - r(\epsilon) = \frac{\sum_{i,j} \hat{A}_{ij}(\epsilon) - \min(\hat{A}_{ij}(\epsilon), \hat{A}_{ji}(\epsilon))}{k}, \tag{A.3}$$

$$k(1 - r(\epsilon)) = \sum_{i,j} \hat{A}_{ij}(\epsilon) - \min(\hat{A}_{ij}(\epsilon), \hat{A}_{ji}(\epsilon)). \tag{A.4}$$

Then, using the defintion of $\hat{A}_{ij}(\epsilon)$ and cancelling the symmetric terms, we obtain,

$$k(1 - r(\epsilon)) = \sum_{i,j} \epsilon A_{ij} - \epsilon \min(A_{ij}, A_{ji}). \tag{A.5}$$

Finally, we get,

$$1 - r(\epsilon) = \frac{\epsilon \sum_{i,j} A_{ij} - \min(A_{ij}, A_{ji})}{k} = \epsilon(1 - r). \square \tag{A.6}$$

Next, we consider the trophic directedness. As defined in Eq. (3.5), the symmetric pseudo-Laplacian of the network is given by,

$$\Lambda = \mathrm{diag}\left(\sum_i A_{ij} + A_{ji}\right) - A - A^\top, \tag{A.7}$$

which remains invariant under the interpolation,

$$\hat{\Lambda}(\epsilon) = \text{diag}\left(\sum_i \hat{A}(\epsilon)_{ij} + \hat{A}(\epsilon)_{ji}\right) - \hat{A}(\epsilon) - \hat{A}(\epsilon)^\top, \tag{A.8}$$

$$= \Lambda. \tag{A.9}$$

We notice that

$$\hat{v}_j = \sum_i \hat{A}_{ji}(\epsilon) - \hat{A}_{ij}(\epsilon) = \epsilon \sum_i A_{ji} - A_{ij} = \epsilon v_j. \tag{A.10}$$

Therefore the trophic levels, $\hat{h}(\epsilon)$, for the interpolated network are given by the solution to the equation,

$$\Lambda \hat{h}(\epsilon) = \epsilon v, \tag{A.11}$$

which can be written as $\hat{h}(\epsilon) = \epsilon h$ where h are the trophic levels of the original network A.

Finally, we use the alternative, but equivalent, formulation for $1 - F_0$, shown by Mackay et al. [89], given by,

$$1 - F_0 = \frac{\sum_{i,j} A_{ij}(h_i - h_j)}{\sum_{i,j} A_{ij}}. \tag{A.12}$$

The denominator, which we again denote k, is preserved under the interpolation. Expanding the interpolation, we obtain,

$$k(1 - F_0(\epsilon)) = \epsilon(1-\epsilon)\frac{1}{2}\sum_{i,j}(A_{ij} + A_{ji})(h_i - h_j) + \epsilon^2 \sum_{i,j} A_{ij}(h_i - h_j). \tag{A.13}$$

Notice that the sum $\sum_{i,j}(A_{ij} + A_{ji})(h_i - h_j)$ vanishes as the term for (i, j) cancels with the term for (j, i). Therefore, we get that the trophic directedness is given by,

$$1 - F_0(\epsilon) = \epsilon^2(1 - F_0), \qquad \sqrt{1 - F_0(\epsilon)} = \epsilon\sqrt{1 - F_0}. \square \tag{A.14}$$

A.2 Algorithm for sampling hierarchically directed networks

In Sec. 3.1.1.1 and 3.3.1, we consider an interpolation of a hierarchical, directed, weighted network. This network is generated using the algorithm in Ref. [96], which is an extension of the *preferential attachment* algorithm from Ref. [331].

We consider a growing network where, at each step, a new node is added to the network. The new node j connects to an existing node i with probability proportional to d_i, the degree of node i. More specifically, if the node has $d_i = 0$, it is connected with a constant probability $0 < p_0 \ll 1$. Otherwise, it is connected with probability $\min\left(1, \mu + \frac{d_i}{\sum_k d_k}\right)$. We then sample the weight of the edge from $U(0,1)$, and introduce a weak reciprocal link with weight $W_{ji} = W_{ij}/\gamma$ for $\gamma \gg 1$ such that $0 \leq W_{ji} \leq W_{ij}$. This algorithm produces a strongly connected graph where each edge is reciprocated.

A.3 Proof of Lemma 3.3.2

First, we note that if A is normal, then $B = \Theta(\gamma A - I)$ is also normal,

$$BB^\top = \Theta^2(I - \gamma A - \gamma A^\top + \gamma^2 AA^\top), \tag{A.15}$$

$$= \Theta^2(I - \gamma A - \gamma A^\top + \gamma^2 A^\top A) = B^\top B. \tag{A.16}$$

Next, we note that the interpolation $\hat{A}(\epsilon)$ gives rise to an interpolation on B, $\hat{B}(\epsilon)$, which we can write in the form,

$$\hat{B}(\epsilon) = \bar{B} + \epsilon K, \tag{A.17}$$

where $\bar{B} = \frac{1}{2}\left(B + B^\top\right)$ and $K = \frac{1}{2}\left(B - B^\top\right) = \frac{1}{2}\left(A - A^\top\right)$. Our proof requires an intermediate proposition.

Proposition A.3.1. *If B is normal, then the covariance of the Ornstein-Uhlenbeck process (OU) process with $\hat{B}(\epsilon)$ and $D = \frac{\sigma^2}{2} I$, is $S = \frac{\sigma^2}{2}\bar{B}^{-1}$, which is independent of ϵ.*

Assuming that this is true, we compute the EPR. We have that $\hat{B}(\epsilon)D^{-1}\hat{Q}(\epsilon) = \epsilon K + \epsilon^2 K^2 \bar{B}^{-1}$, thus,

$$\hat{\Phi}(\epsilon) = \mathrm{Tr}(\epsilon K + \epsilon^2 K^2 \bar{B}^{-1}) = \epsilon^2 \,\mathrm{Tr}(K^2 \bar{B}^{-1}) = \epsilon^2 \Phi(1), \tag{A.18}$$

i.e. we have that the EPR along the interpolation is quadratic in ϵ. It remains to prove Prop. A.3.1. Let $\hat{S}(\epsilon)$ be the covariance associated with $\hat{B}(\epsilon)$, which satisfies,

$$\hat{B}(\epsilon)\hat{S}(\epsilon) + \hat{S}(\epsilon)\hat{B}(\epsilon)^\top = \sigma^2 I, \tag{A.19}$$

where it is clear that $\hat{S}(0) = \frac{\sigma^2}{2}\bar{B}^{-1}$. We assume that $\hat{S}(\epsilon) = \frac{\sigma^2}{2}\bar{B}^{-1}$,

$$(\bar{B} + \epsilon K)\frac{\sigma^2}{2}\bar{B}^{-1} + \frac{\sigma^2}{2}\bar{B}^{-1}(\bar{B} - \epsilon K) = \sigma^2 I. \tag{A.20}$$

This holds if and only if,

$$K\bar{B}^{-1} = \bar{B}^{-1}K, \tag{A.21}$$

which we will show holds if B is normal. First we show that $\bar{B}K = K\bar{B}$ i.e. the symmetric and antisymmetric components of a normal matrix must commute. We have

$$BB^{\top} = (\bar{B} + K)(\bar{B} - K) = \bar{B}^2 + K\bar{B} - \bar{B}K - K^2, \tag{A.22}$$

$$B^{\top}B = (\bar{B} - K)(\bar{B} + K) = \bar{B}^2 - K\bar{B} + \bar{B}K - K^2, \tag{A.23}$$

thus $K\bar{B} - \bar{B}K = -K\bar{B} + \bar{B}K$ i.e. $K\bar{B} = \bar{B}K$. Finally, we show that K must commute with $\bar{B}^{-1}$ as well,

$$\bar{B}^{-1}(\bar{B}K) = \bar{B}^{-1}(K\bar{B}), \qquad K = \bar{B}^{-1}K\bar{B}, \qquad K\bar{B}^{-1} = \bar{B}^{-1}K, \tag{A.24}$$

and so we have the result. □

Appendix B

Appendices for Ch. 4

B.1 Participant-level results with DiMViGI

In Sec. 4.4, we present the results of the DiMViGI framework applied at the 'cohort-level' i.e. all participants and trials were used to compute a single in- and out-degree distribution. Here, we present the 'participant-level' analysis, where we construct the degree distributions for each participant individually using their associated trials, yielding a distribution of the irreversibility for each tuple. Fig. B.1 shows the distribution of the irreversibility results for each tuple over the 51 participants. We find that the results at the participant level broadly agree with those at the cohort-level. Fig. B.2 reports the results of paired $t-$tests for all possible pairwise comparisons between tuples at an order k for $k = 1, ..., 5$, with a Bonferroni correction for multiple comparisons. We find a range of significant differences.

B.2 Surrogate testing with shuffled time-series

Finite-length time-series have a level of irreversibility that emerges from their finiteness as opposed to the temporal organisation of the data, which we call the 'noise-floor' [3,174]. To validate that the irreversibility is significantly above this noise-floor we must compare them to surrogate data. We randomly shuffled the time-series in time and applied DiMViGI. This means that the density over states is fixed but the ordering is randomised. This restores detailed balance [3]. Fig. B.3 shows a comparison of the measurements in the original MEG time-series and in the surrogate. For each tuplet the difference between the data and surrogate is significant ($p < 0.0001$, Bonferroni corrected). This shows that the irreversibility measured is a significant statistical feature of the data.

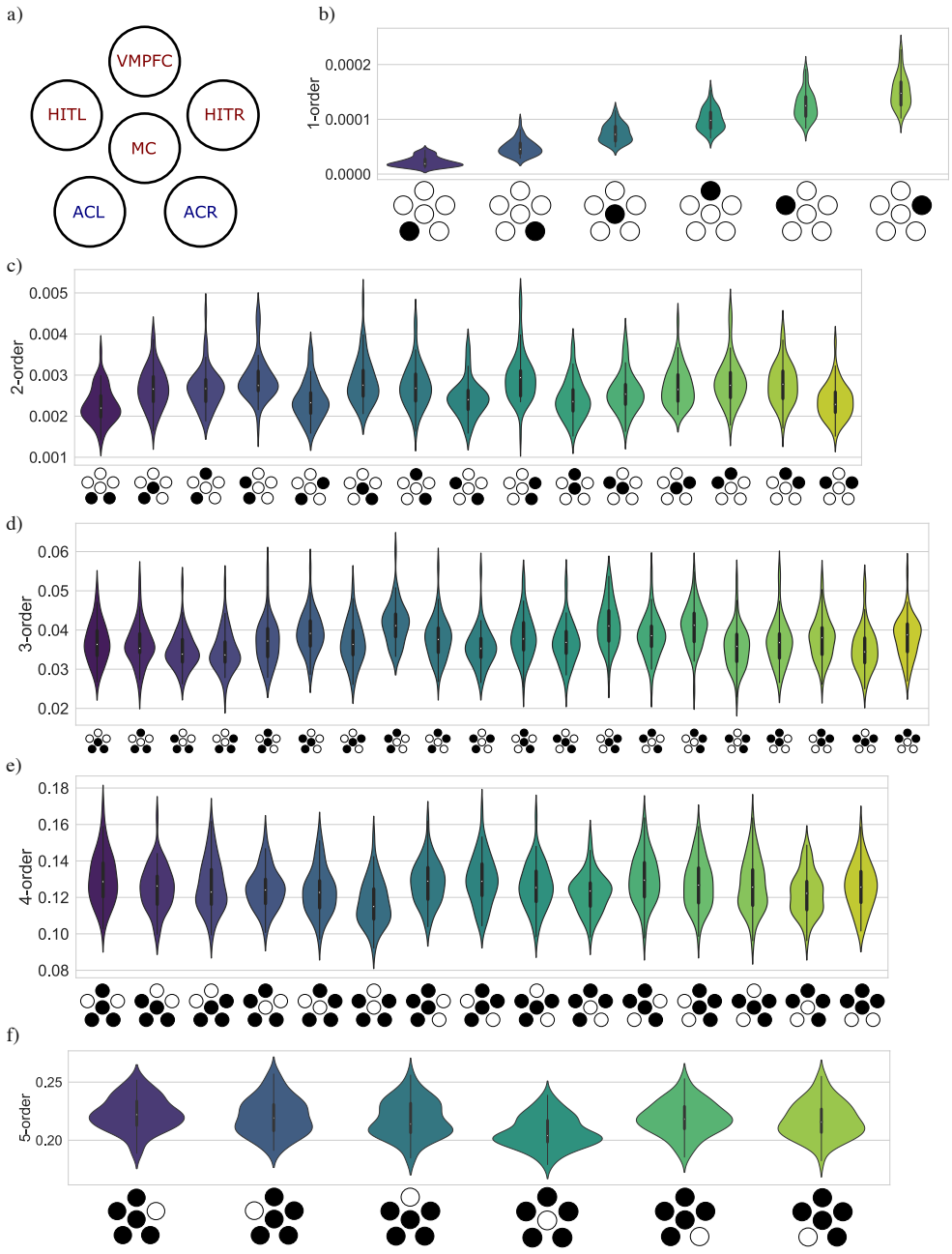

Figure B.1: **Participant distribution of irreversibility for each tuple.** *a*) Schematic representation of the 6 ROIs in the MEG recordings. The icon in the following panels indicates which regions are included in each tuple. *b*) 1-order irreversibility distribution for each region of interest (ROI) in isolation. The results follow the same hierarchy as the cohort-level analysis in the main manuscript. $c) - f)$ k-order irreversibility distribution for each k-tuple of ROIs.

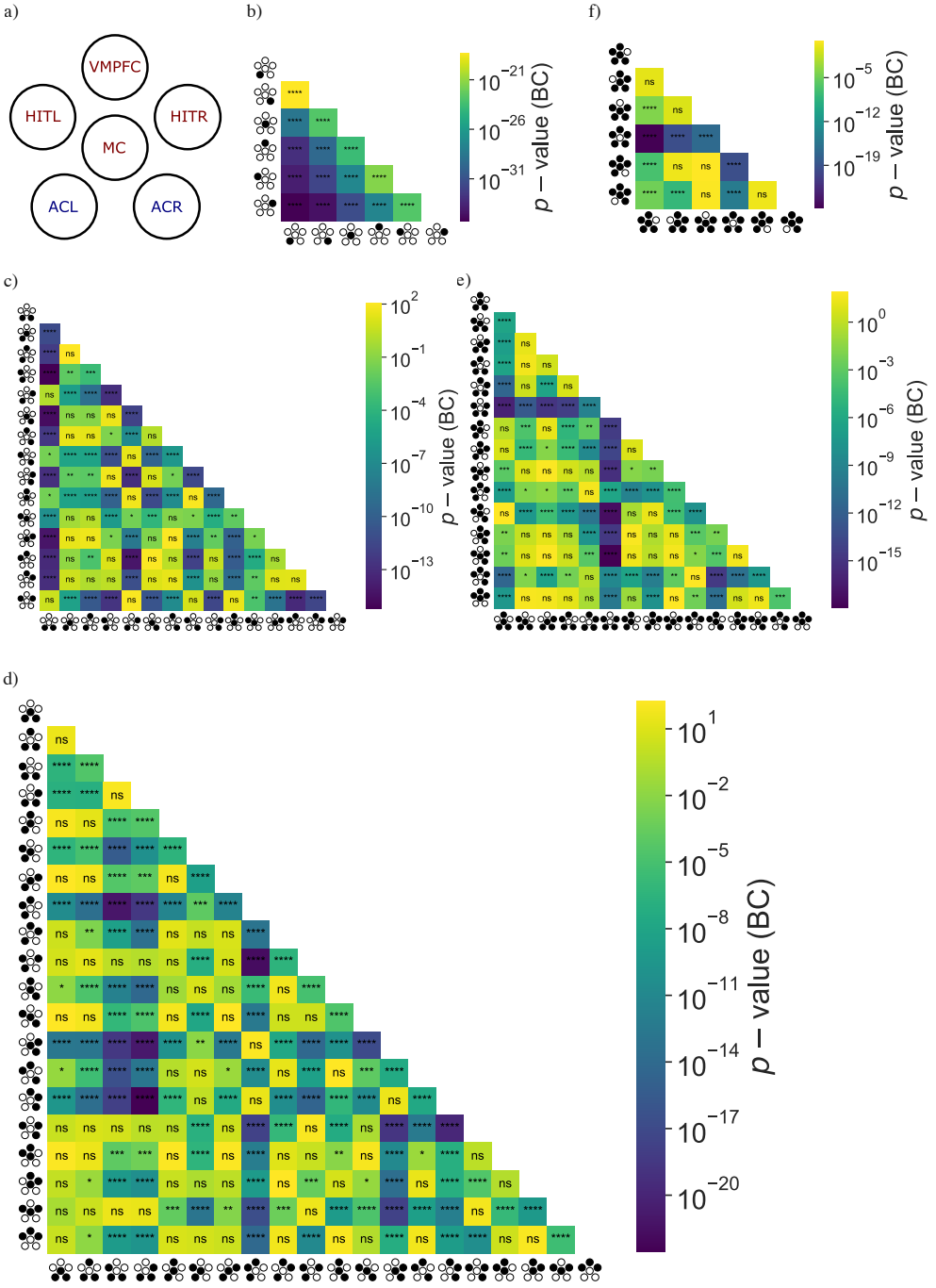

Figure B.2: **Results of paired $t-$tests between all pairs of $k-$tuples at each order k.** The Bonferroni-corrected significance of each comparison is denoted as follows: (ns) if $p > 0.05$; (*) if $p < 0.05$; (**) if $p < 0.01$; (***) if $p < 0.001$ and (****) if $p < 0.0001$. Panel a) shows a schematic representation of the 6 ROIs in the MEG recordings. The remaining panels $b) - f$) show the results for $k = 1, ..., 5$. There is a mixture of significant and not significant differences depending on the number of ROIs in common between the compared tuples.

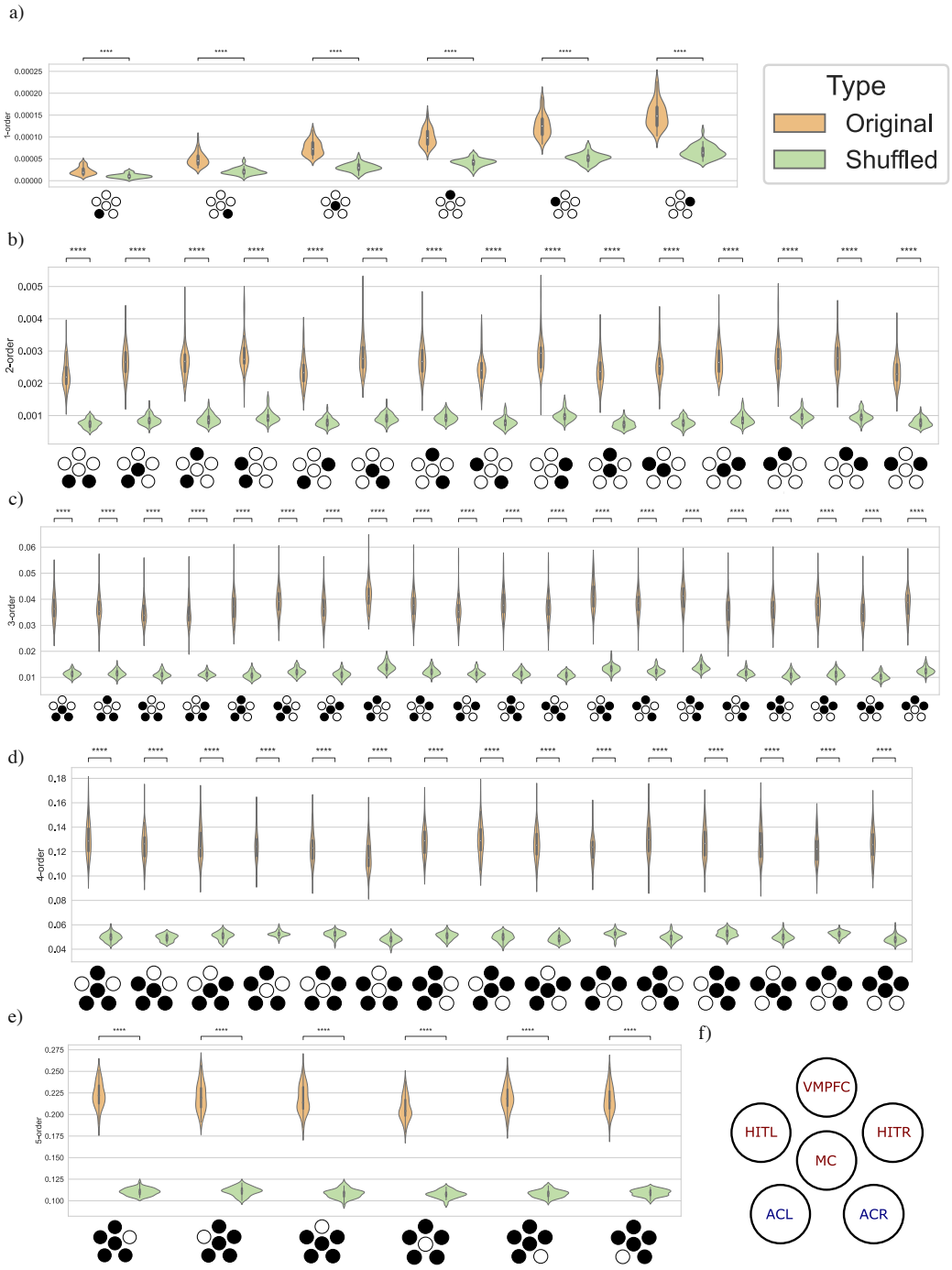

Figure B.3: **Comparison of original neural recording against surrogate data obtained via shuffling the time-series in time.** The difference between the original and shuffled data for each tuple at each level is significant. We label (****) if $p < 0.0001$. $a) - e)$ For $k = 1, ..., 5$ we find that the measured irreversibility is significantly above the noise-floor. $f)$ Schematic representation of the 6 ROIs in the MEG recordings.

Appendix C

Appendices for Ch. 5

C.1 Deriving the Scharfetter-Gummel discretisation

To derive the Scharfetter-Gummel discretisation (SG) discretisation in Eq. (5.7), we follow the analysis of Ref. [198], extending to the 2D case with non-constant drift. We begin with the FVA of the Fokker-Planck (FP) equation

$$\frac{dp_{i,j}}{dt} = -\frac{J^x_{i+\frac{1}{2},j} - J^x_{i-\frac{1}{2},j}}{\Delta x} - \frac{J^y_{i,j+\frac{1}{2}} - J^y_{i,j-\frac{1}{2}}}{\Delta y}. \tag{C.1}$$

Next, we assume that the flux $J^x_{i+\frac{1}{2}}$, drift $f^x_{i+\frac{1}{2}}$ and diffusion $D^x_{i+\frac{1}{2}}$ are constant over the interval $[x_i, x_{i+1}]$. As a result, $p^j(x) := p(x, y_j)$ satisfies the first-order ordinary differential equation (ODE) with constant coefficients

$$J^x_{i+\frac{1}{2},j} = f^x_{i+\frac{1}{2},j} p^j(x) - D^x_{i+\frac{1}{2},j} \frac{dp^j}{dx}(x), \tag{C.2}$$

with boundary values $p^j(x_i) = p_{i,j}$ and $p^j(x_{i+1}) = p_{i+1,j}$. For readability, we now drop the sub- and superscripts. We can solve this boundary value problem using an integrating factor,

$$Je^{-f(x-x_i)/D} = \left(fp(x) - D\frac{dp}{dx} \right) e^{-f(x-x_i)/D}, \tag{C.3}$$

$$= -D\frac{d}{dx}\left(e^{-f(x-x_i)/D} p(x) \right). \tag{C.4}$$

We then integrate this over the range $[x_i, x_{i+1}]$ where $\Delta x = x_{i+1} - x_i$. This yields

$$\int_{x_i}^{x_i+\Delta x} Je^{-f(x-x_i)/D} = -D\left(p(x) e^{-f(x-x_i)/D} \right)\Big|_{x_i}^{x_{i+1}}, \tag{C.5}$$

$$J\frac{D}{f}\left(1 - e^{-f\Delta x/D} \right) = D\left(p_{i,j} - p_{i+1,j} e^{-f\Delta x/D} \right). \tag{C.6}$$

Reintroducing the subscripts, we obtain

$$J^x_{i+\frac{1}{2},j} = \frac{f^x_{i+\frac{1}{2},j}\left(p_{i,j} - e^{-f^x_{i+\frac{1}{2},j}\Delta x/D^x_{i+\frac{1}{2},j}}p_{i+1,j}\right)}{1 - e^{-f^x_{i+\frac{1}{2},j}\Delta x/D^x_{i+\frac{1}{2},j}}}, \tag{C.7}$$

as in Eq. (5.7). A similar derivation can be made for the flux terms $J^x_{i-\frac{1}{2},j}$, $J^y_{i,j+\frac{1}{2}}$, $J^y_{i,j-\frac{1}{2}}$. These can then be substituted into Eq. (5.3) where the coefficients of $p_{i+1,j}$, $p_{i-1,j}$, $p_{i,j+1}$, $p_{i,j-1}$ and $p_{i,j}$ can be derived.

Whilst not reported here, we found that other structure-preserving schemes such as the *Chang-Cooper* discretisation [194] yielded valid master equations for these diffusions, whilst the centred-difference scheme only did so under step-size conditions.

C.2 Non-diagonalisable diffusion

We will show that using a non-diagonal diffusion matrix requires a nine-point stencil, including the additional terms $\{p_{i+1,j+1}, p_{i-1,j+1}, p_{i+1,j-1}, p_{i-1,j-1}\}$. However, we show that this discretisation does not yield a valid master equation.

For simplicity, we will consider a SDE with no drift and non-diagonal, spatially-dependent diffusion, where $D(x,y)$ is given by,

$$D(x,y) = \begin{pmatrix} D^x(x,y) & D^{xy}(x,y) \\ D^{xy}(x,y) & D^y(x,y) \end{pmatrix}, \tag{C.8}$$

where D is positive definite. The FP flux is then given by $J = -\nabla\cdot(Dp)$. This process cannot be discretised with a generalisation of the SG discretisation due to these off-diagonal entries producing cross derivatives. We first note that,

$$\nabla\cdot(Dp) = \begin{pmatrix} \partial_x\left(D^x p\right) + \partial_y\left(D^{xy}p\right) \\ \partial_x\left(D^{xy}p\right) + \partial_y\left(D^y p\right) \end{pmatrix}. \tag{C.9}$$

To obtain an expression for the discrete flux, we will use traditional centred differences [332], for example we have,

$$\partial_x\left(D^x p\right)|_{i+\frac{1}{2},j} \approx \frac{D^x_{i+1,j}p_{i+1,j} - D^x_{i,j}p_{i,j}}{\Delta x}, \tag{C.10}$$

$$\partial_y\left(D^{xy}p\right)|_{i+\frac{1}{2},j} \approx \frac{D^x_{i+\frac{1}{2},j+1}p_{i+\frac{1}{2},j+1} - D^x_{i+\frac{1}{2},j-1}p_{i+\frac{1}{2},j-1}}{2\Delta x}, \tag{C.11}$$

$$\approx \frac{D^x_{i+\frac{1}{2},j+1}\left(p_{i+1,j+1}+p_{i,j+1}\right) - D^x_{i+\frac{1}{2},j-1}\left(p_{i+1,j-1}+p_{i,j-1}\right)}{4\Delta x}. \tag{C.12}$$

Repeating this calculation at each boundary, we can obtain the transition rates between states. There are transitions between the four adjacent cells with rates,

$$L_{(i,j)(i+1,j)} = \frac{D^{xx}_{i+1,j}}{\Delta x^2} + \frac{D^{xy}_{i+1,j+\frac{1}{2}} - D^{xy}_{i+1,j-\frac{1}{2}}}{4\Delta x \Delta y}, \tag{C.13}$$

and analogous terms for $(i-1,j)$ and $(i,j\pm1)$. Additionally, there are terms between cells meeting at the corner i.e.,

$$L_{(i,j)(i+1,j+1)} = \frac{D^{xy}_{i+\frac{1}{2},j+1} + D^{xy}_{i+1,j+\frac{1}{2}}}{4\Delta x \Delta y}, \qquad L_{(i,j)(i-1,j+1)} = -\frac{D^{xy}_{i-\frac{1}{2},j+1} + D^{xy}_{i-1,j+\frac{1}{2}}}{4\Delta x \Delta y}, \tag{C.14}$$

$$L_{(i,j)(i+1,j-1)} = -\frac{D^{xy}_{i+\frac{1}{2},j-1} + D^{xy}_{i+1,j-\frac{1}{2}}}{4\Delta x \Delta y}, \qquad L_{(i,j)(i-1,j-1)} = \frac{D^{xy}_{i-\frac{1}{2},j-1} + D^{xy}_{i-1,j-\frac{1}{2}}}{4\Delta x \Delta y}.$$

Examining these rates we see that the positivity constraint is violated e.g. if we take D^{xy} to be constant, then if $L_{(i,j)(i+1,j+1)} > 0$ then $L_{(i,j)(i-1,j+1)} < 0$. As a result, we cannot define a valid CTMC with this rectangular grid unless $D^{xy} = 0$.

C.3 Optimisation-based solution to the embedding problem

The embedding problem, detailed in Sec. 5.2.1 aims to find a Laplacian, L, whose observation in discrete-time has transition probability matrix (TPM),

$$P = \exp\left(L\Delta t\right). \tag{C.15}$$

Whilst L is assumed to only admit certain transitions, P can have non-zero entries in any position. We can formulate a solution to the embedding problem as a non-convex optimisation problem,

$$L^* = \arg\min_{L\in\mathcal{L}} ||P - \exp\left(L\Delta t\right)||_F, \tag{C.16}$$

where $\mathcal{L}$ is the set of valid Laplacian matrices which conserve probability, and only have non-zero transition rates for valid transitions. Due to presence of the matrix-exponential, this problem is non-convex. We can compare this approach to the MLE in Eq. (5.40), infer the Laplacian, and solve for the stationary distribution. In practice, we use the standard `L-BFGS-B` method from `scipy.optimize`, but we find that this approach is far more inefficient than the direct estimation using Eq. (5.40).

Fig. C.1 shows the result of a numerical experiment comparing the direct MLE with the optimisation-based approach for a trajectory from *a*) the OU, and *b*) the HO, compared to the exact stationary density, using the same parameters as in Sec. 5.1.2. We find that the methods yield comparably accurate stationary distributions.

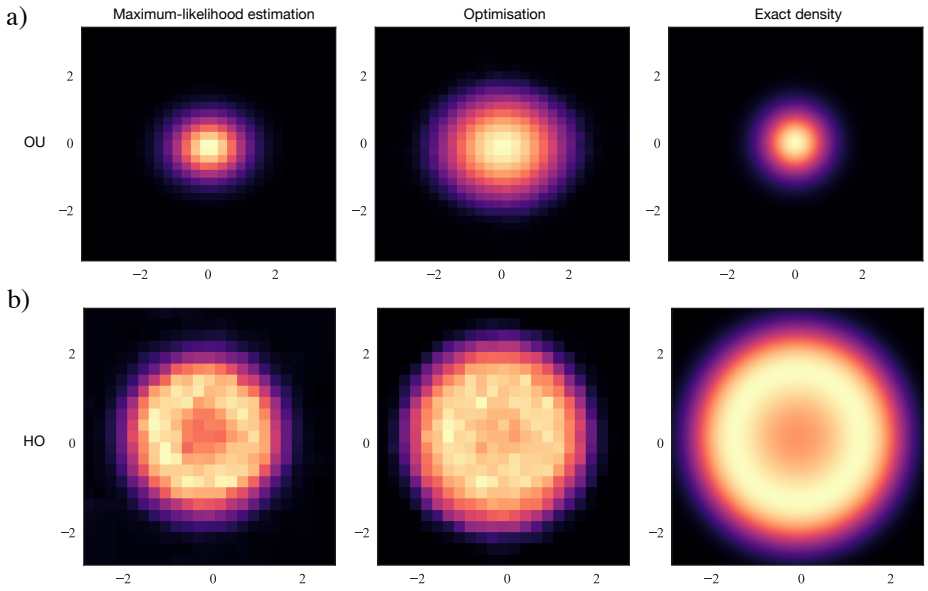

Figure C.1: **Maximum-likelihood vs. optimisation for the embedding problem.** We compare the maximum-likelihood estimation (MLE) estimator of Eq. (5.40) with the non-convex optimisation approach in Eq. (C.16). We reconstruct the Laplacian matrix, and then solve for the stationary distribution, which we compare with the exact density for the *a*) OU, and *b*) Hopf oscillator (HO), given in Ex. 5.1.2 and 5.1.3.

C.4 Sampling for oscillatory and chaotic processes

Whilst the Euler-Maruyama scheme is commonly used to sample paths from SDEs, it is not *mean-square convergent* is certain cases [333]. To sample from such SDEs, we employ alternative methods.

C.4.1 Splitting method for stochastic oscillators

Given a nonlinear SDE,

$$\frac{dx}{dt} = f(x) + \sigma\, dw(t), \tag{C.17}$$

we 'split' the drift term into a nonlinear, deterministic part and a linear stochastic part [334],

$$x^{[1]}(t) = Ax(t) + \Sigma\, dw(t), \qquad x^{[2]}(t) = G(x), \tag{C.18}$$

where $F(x) = Ax + G(x)$. The linear part, $x^{[1]}$, defines an OU process which can be simulated exactly (see 3.2.1). If the splitting is chosen such that the ODE for $x^{[2]}$ can

be solved explicitly, then the solution of the full process can be approximated with the Strang approximation,

$$x(t+h) \approx \varphi_1^{[h/2]} \circ \varphi_2^{[h]} \circ \varphi_1^{[h/2]}(x(t)), \tag{C.19}$$

where $\varphi_i^{[h]}$ is a simulation from the exact h-time transition kernel of $x^{[i]}$ [334].

The Hopf oscillator For the HO, we use the decomposition,

$$dx^{[1]}(t) = \begin{pmatrix} a & -\omega \\ \omega & a \end{pmatrix} x^{[1]}(t)\, dt + \Sigma\, dW(t), \quad dx^{[2]}(t) = \begin{pmatrix} -(x^2+y^2)x \\ -(x^2+y^2)y \end{pmatrix} dt, \tag{C.20}$$

where $x^{[2]} = (x, y)$. The nonlinear part can be solved easily in polar coordinates.

The van der Pol oscillator For the van der Pol oscillator (VDP), we use the decomposition,

$$x^{[1]}(t) = \begin{pmatrix} 0 & \theta \\ -\theta & \mu \end{pmatrix} x^{[1]}(t) + \Sigma\, dw(t), \qquad x^{[2]}(t) = \begin{pmatrix} 0 \\ -\mu x^2 y \end{pmatrix}, \tag{C.21}$$

where the nonlinear part can be solved trivially due to the constancy of the first component x.[1]

C.5 Inferring the spatial dependence of the diffusion in real-world data

In Sec. 5.3, we assume that diffusion is not spatially-dependent and thus can be represented by a single matrix D. Here, we briefly assess whether this is the case in the real-world data by leveraging the spatially-dependent estimator $\hat{D}$ of Eq. 5.53. Panel *a*) of Fig. C.2 shows the coefficients of the diffusion estimator in each of the triangular bins, estimated using the RBC 'Healthy 1'. The diffusion appears to be spatially-independent, with more variation at the borders where the paths have fewer points leading to higher estimation error. Moreover, the diffusion is close to isotropic. Panel *b*) shows the diffusion estimator calculated on the healthy heartbeats. This appears to show more spatial dependence whilst remaining mostly isotropic. Nevertheless, by examining the stationary distribution with MLE, we can see that the paths spend relatively little time in the parts of phase-space that diffusion variation which indicates that this variation could arise from the numerical error associated with having only a few samples in this area of phase-space rather than a genuine spatial dependence [Panel *c*)].

[1] Here $x(t)$ is not the full state vector, but the first component i.e. (x, y).

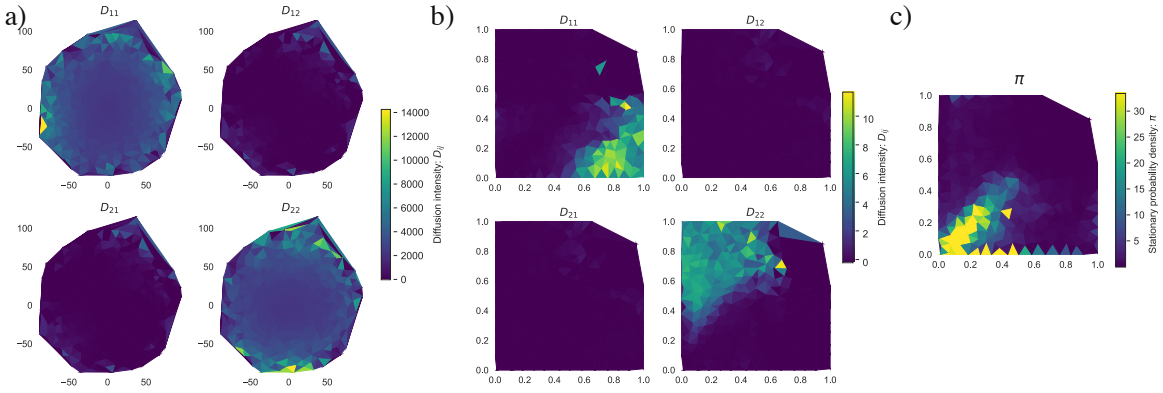

Figure C.2: **Spatially-dependent diffusion in real-world paths.** *a*) An example of the spatially-dependent diffusion estimator in the RBC 'Healthy 1'. *b*) An example of the spatially-dependent diffusion estimator in the healthy heartbeats. *c*) A MLE of the stationary probability distribution π in the healthy heartbeats.

Appendix D

Appendices for Ch. 7

D.1 Optimal decomposition of a network under constraints

Let $W \in \mathbb{R}^{n\times n}$ be a matrix that is low-rank and has the form $W = \Gamma W_s$, where $\Gamma \in \mathbb{R}^{n\times k}$ and $W_s \in \mathbb{R}^{k\times n}$. Γ has full column rank k. We want to find the 'best' decomposition,

$$W = \Gamma(\Omega + \Pi), \tag{D.1}$$

where $\Gamma\Omega$ is symmetric. We will define this optimal decomposition to have,

$$\tilde{\Omega} = \text{argmin}_{\Omega}||\Gamma\Omega - C||_F^2, \tag{D.2}$$

where $C = \frac{1}{2}\left(W + W^\top\right)$ i.e. the matrix that is most similar to the symmetrised version of W, but which respects the parametrisation.

Let P be the projector matrix onto $\mathcal{S} = \{X \in \mathbb{R}^{n\times n} : X \in \text{range}(\Gamma)\}$, i.e. $P = \Gamma\Gamma^\dagger$. Due to the symmetry of any feasible X, we have that both the columns and rows of X live in the range of Γ. This implies that,

$$PX = X, \qquad XP = X, \tag{D.3}$$

which implies that $X = PXP$ for any feasible X. As we are trying to minimise the Frobenius norm, the optimal solution $X = \Gamma\Omega$, is the projection of C onto the linear subspace,

$$\mathcal{Q} = \{X \in \mathbb{R}^{n\times n} : X = X^\top, X \in \text{range}(\Gamma)\}. \tag{D.4}$$

As C is already symmetric, the solution is simply the projection of C onto $\mathcal{S}$,

$$X = PCP = \Gamma\Gamma^\dagger C\Gamma\Gamma^\dagger, \tag{D.5}$$

which gives $\tilde{\Omega} = \Gamma^\dagger C\Gamma\Gamma^\dagger$.

D.2 Mapping coded inputs to network inputs with ridge-regression

As part of Ex. 7.3.2, we need to solve the ridge regression optimisation problem in Eq. (7.39). This has the following exact solution.

We combine our $\mathbf{c}^*$ values into a matrix $\mathbf{C} \in \mathbb{R}^{m \times k}$, and combine the codes $\mathbf{s}_i$ into a matrix $\mathbf{S} \in \mathbb{R}^{m \times m}$, and augment it with an additional column of ones to form $\mathbf{S}^* = [\mathbf{S}, \mathbf{1}] \in \mathbb{R}^{m \times (m+1)}$. We then aim to fit the model $\mathbf{C} \approx \mathbf{S}\mathbf{G}_\mathbf{s} + \mathbf{1}\mathbf{d}^\top$, by minimising,

$$\Theta^* = \text{argmin}_\Theta ||\mathbf{S}^*\Theta - \mathbf{C}||^2 + \lambda ||\Theta||^2, \tag{D.6}$$

where $\Theta = [G_s, \mathbf{d}]$. This is equivalent to Eq. (7.39). This has exact solution [310],

$$\Theta^* = ((\mathbf{S}^*)^\top \mathbf{S}^* + \lambda \mathbf{I})^{-1} (\mathbf{S}^*)^\top \mathbf{C}. \tag{D.7}$$

D.3 Designed rotational drift is DFSD

In Sec. 7.3.2.1, we design a rotational field of the form,

$$\mathbf{R}(\mathbf{y}) = \frac{\alpha}{\pi(\mathbf{y})} g(\mathbf{y}) \mathbf{M} \nabla \phi(\mathbf{y}), \tag{D.8}$$

which we now show to be divergence free with respect to the stationary density (DFSD). It is sufficient to show that,

$$\boldsymbol{\zeta}(\mathbf{y}) = g(\mathbf{y}) \mathbf{M} \nabla \phi(\mathbf{y}), \tag{D.9}$$

is divergence-free. Applying the product rule we have that,

$$\nabla \cdot \boldsymbol{\zeta} = (\nabla g) \cdot (\mathbf{M} \nabla \phi) + g(\nabla \cdot (\mathbf{M} \nabla \phi)). \tag{D.10}$$

We have that,

$$\nabla g = -\frac{\phi}{\xi^2} g \nabla \phi, \tag{D.11}$$

which implies that,

$$(\nabla g) \cdot (\mathbf{M} \nabla \phi) = -\frac{\phi}{\xi^2} g \nabla \phi \cdot (\mathbf{M} \nabla \phi) = 0,$$

as $\mathbf{v} \cdot \mathbf{M}\mathbf{v} = 0$ for antisymmetric $\mathbf{M}$. Next, we have that,

$$\nabla \cdot (\mathbf{M} \nabla \phi) = \sum_{i,j} \mathbf{M}_{ij} \partial_i \partial_j \phi, \tag{D.12}$$

which is the sum-product of a symmetric (Hessian) and antisymmetric matrix, which is zero. Thus $\nabla \cdot \boldsymbol{\zeta} = 0$. $\square$

D.4 Flux and EPR in a low-rank RNN

As considered in Sec. 7.4, we consider an RNN with state $\mathbf{u}(t) \in \mathcal{A}$, where the latent dynamics follow the SDE,

$$d\mathbf{y}(t) = \mathbf{f}(\mathbf{y})\, dt + B_s\, d\mathbf{w}(t), \tag{D.13}$$

which has a stationary density $\pi_{\mathbf{y}}(\mathbf{y})$. In the network space, the stationary density is given by,

$$\pi_{\mathbf{u}}(\mathbf{u}) = \frac{\pi_{\mathbf{y}}(\Gamma^\dagger(\mathbf{u}-\mathbf{b}))}{\sqrt{\det(\Gamma^\top\Gamma)}}, \tag{D.14}$$

where we are restricting to $\mathbf{u} \in \mathcal{A}$ to avoid Dirac-delta terms. Similarly, the flux is related by,

$$\mathbf{J}_{\mathbf{u}}(\mathbf{u}) = \frac{\Gamma\mathbf{J}_{\mathbf{y}}(\Gamma^\dagger(\mathbf{u}-\mathbf{b}))}{\sqrt{\det\left(\Gamma^\top\Gamma\right)}}. \tag{D.15}$$

Finally, we can compute the EPR. The traditional formula for EPR, Eq.(2.18), assumes that the diffusion matrix is full-rank [57]. This does not hold in our case where $2\Upsilon = BB^\top = \Gamma B_s B_s^\top \Gamma^\top$, which is a symmetric positive semidefinite matrix with rank $k < n$. However, as mentioned in Remark. 2.2.2, we can extend the EPR formula to this case, where it is given by,

$$\Phi = \int_{\mathbb{R}^n} \mathbf{F}_{\text{irr}}^\top \Upsilon^\dagger \mathbf{F}_{\text{irr}} \pi_{\mathbf{u}}\, d\mathbf{u}, \tag{D.16}$$

i.e. where Υ^{-1} has been replaced by the Moore-Penrose (MP) inverse. This is valid if and only if $\mathbf{F}_{\text{irr}}(\mathbf{u}) \in \text{Range}(B(\mathbf{u}))$ where $B = \Gamma B_s$ [55]. We consider spatially-constant diffusion, and as we have already shown that $\mathbf{F}_{\text{irr}}$ lies in the image of Γ (Eq. (7.53)), it is sufficient for B_s to be of full row-rank k, guaranteeing that the noise is *elliptic*. A necessary requirement for this is that $d \geq k$ i.e. we have more independent noise-sources in the network than dimensions in the latent process. When B_s has full row-rank, we have that the EPR of the latent process and the full network are identical,

$$\Phi_{\mathbf{u}} = \Phi_{\mathbf{y}}. \tag{D.17}$$

D.5 Parameters and details for numerical experiments

Here we provide details on parameters for the numerical simulations. Throughout, we train the perceptrons with `PyTorch` using the Adam optimiser with learning rate $= 0.001$. We fix $\lambda_{\text{diff}} = 20$.

The parameters Γ and W_s used standard `PyTorch` initialisations i.e. they are sampled using a Kaiming uniform initialisation with $\Gamma_{ij} \sim U(-\sqrt{m}, \sqrt{m})$ where m is the number of input dimensions – thus $m = k$ for Γ and $m = n$ for W_s. The parameters $\mathbf{b}$ and I_s are initialised as zero vectors. The entries of B_s are initialised using a standard normal distribution i.e. $B_{s,ij} \sim \mathcal{N}(0, 1)$.

D.5.1 Nonlinear systems and RNN representations

In Sec. 7.2.2, we use the following parameters.

- **Van der Pol.** We take $\sigma = 0.25$. We train an RNN with $N = 64$, and 25,000 uniform samples from $[-4, 4]^2$ for 30,000 epochs.
- **Lorenz.** We take $\sigma = 0.25$. We train an RNN with $N = 512$, and 125,000 uniform samples from $[-25, 25] \times [-30, 30] \times [0, 50]$ with $\lambda_{\text{diff}} = 20$ for 100,000 epochs.
- **Dadras.** We take $\sigma = 0.01$. We train an RNN with $N = 1024$, and 150,000 uniform samples from $[-20, 20] \times [-12, 10] \times [-15, 15]$ for 100,000 epochs.

D.5.2 Switching and cycling attractors

In Sec. 7.3, we use the following parameters.

- **Input-driven switching.** We place four minima at $(\pm 3, \pm 3) \in \mathbb{R}^2$, with Gaussian well parameters given by $a_j = 0.125$ and $\nu_j = 1$, and noise intensity $\sigma = 0.1$. We train the RNN with $N = 256$ using 25,000 samples from $[-4, 4]^2$ for 30,000 epochs.

Next, we have the three examples of autonomously cycling diffusions.

- **Three minima.** We place three minima at $\boldsymbol{\mu} = \{(1, 2), (-1.5, 0), (1, -2.25)\}$, with $a_j = 0.25$ and $\nu_j = 0.75$. The process has $\xi = 0.2$, $\alpha = 1.5$ and $\sigma = 0.2$. We train an RNN with $N = 128$ and 25,000 samples from $[-4, 4]^2$ for 30,000 epochs.

- **Four minima.** We place four minima at $\boldsymbol{\mu} = \{(2,3),(3,-2),(-2,-3),(-1.5,1.5)\}$, with $a_j = 0.25$ and $\nu_j = 0.75$. The process has $\xi = 0.6$, $\alpha = 2.5$ and $\sigma = 0.225$. We train an RNN with $N = 128$ and 25,000 samples from $[-4,4]^2$ for 30,000 epochs.

- **Five minima.** We place five minima at $\boldsymbol{\mu} = \{(3.5,4),(0,0),(3,-4),(-3,4),(-3.5,4)\}$, with $a_j = 0.25$ and $\nu_j = 0.6$. The process has $\xi = 0.6$, $\alpha = 2.5$ and $\sigma = 0.225$. We train an RNN with $N = 128$ and 40,000 samples from $[-8,8]^2$ for 40,000 epochs.

In Sec. 7.4, we retrain the second layers using 25,000 additional samples and 30,000 epochs to obtain the HHD.

Bibliography

[1] P. Anderson, "More is different: Broken symmetry and the nature of the hierarchical structure of science," *Science*, vol. 177, no. 4047, pp. 393–396, 1972.

[2] K. Friston, "Functional and effective connectivity: A review," *Brain Connectivity*, vol. 1, no. 1, 2011.

[3] C. W. Lynn, E. J. Cornblath, L. Papadopoulos, and D. S. Bassett, "Broken detailed balance and entropy production in the human brain," *Proceedings of the National Academy of Sciences*, vol. 118, no. 47, 2021.

[4] H. Haken, *Synergetics: An Introduction Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry and Biology*. Springer, 1978.

[5] N. Wiener, *Cybernetics: Or Control and Communication in the Animal and the Machine*. Hermann & Cie, 1948.

[6] I. Prigogine, *Introduction to Thermodynamics of Irreversible Processes*. California, United States: John Wiley & Sons, 1968.

[7] M. E. J. Newman, *Networks*. Oxford University Press, 2018.

[8] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.

[9] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[10] J. A. Thomas and T. M. Cover, *Elements of Information Theory*. Wiley, 2006.

[11] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press, 2022.

[12] D. C. Krakauer, *The Complex World: An Introduction to the Foundations of Complexity Science*. SFI Press, 2024.

[13] E. M. Izhikevich, *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. The MIT Press, 2006.

[14] S. Coombes and K. C. A. Wedgwood, *Neurodynamics: An Applied Mathematics Perspective*. Springer, 2023.

[15] C. Golgi, "Sulla struttura della sostanza grigia del cervello," *Gazzetta Medica Italiana. Lombardia*, vol. 33, pp. 244–246, 1873.

[16] S. R. y Cajal, "Estructura de los centros nerviosos de las aves," *Revista Trimestral de Histología Normal y Patológica*, vol. 1, pp. 1–27, 1888.

[17] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *Journal of Physiology*, vol. 117, no. 4, 1952.

[18] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[19] D. S. Bassett and O. Sporns, "Network neuroscience," *Nature Neuroscience*, vol. 20, p. 353–364, 2017.

[20] A. Fornito, A. Zalesky, and E. T. Bullmore, *Fundamentals of Brain Network Analysis*. Academic Press, 2016.

[21] P. Dayan and L. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press, 2005.

[22] B. Ermentrout and D. H. Terman, *Mathematical Foundations of Neuroscience*. Springer, 2010.

[23] G. Deco and M. Kringelbach, *Whole-brain modelling: Cartography of the dynamics of mind*. Oxford University Press, 2026.

[24] P. Ashwin, S. Coombes, and R. Nicks, "Mathematical frameworks for oscillatory network dynamics in neuroscience," *The Journal of Mathematical Neuroscience*, vol. 6, pp. 1–92, 2016.

[25] B. J. Cook, A. D. H. Peterson, W. Woldman, and J. R. Terry, "Neural field models: A mathematical overview and unifying framework," *Mathematical Neuroscience and Applications*, vol. 2, no. 2, pp. 1–67, 2022.

[26] S. Coombes, "Waves, bumps, and patterns in neural field theories," *Biological Cybernetics*, vol. 93, pp. 91–108, 2005.

[27] J. C. Pang, K. M. Aquino, M. Oldehinkel, P. A. Robinson, B. D. Fulcher, M. Breakspear, and A. Fornito, "Geometric constraints on human brain function," *Nature*, vol. 618, p. 566–574, 2023.

[28] D. Papo and J. Buldú, "Does the brain behave like a (complex) network? I. Dynamics," *Physics of Life Reviews*, vol. 48, pp. 47–98, 2024.

[29] R. C. Balescu, *Equilibrium and Nonequilibrium Statistical Mechanics*. Wiley-Blackwell, 1975.

[30] E. Jaynes, "Information theory and statistical mechanics," *Physical Review*, vol. 106, no. 4, pp. 620–630, 1957.

[31] E. Jaynes, "Information theory and statistical mechanics II," *Physical Review*, vol. 108, no. 2, pp. 171–190, 1957.

[32] J. W. Gibbs, *Elementary Principles in Statistical Mechanics*. Charles Scribner's Sons, 1902.

[33] L. Meshulam and W. Bialek, "Statistical mechanics for networks of real neurons," *Reviews of Modern Physics*, vol. 97, no. 045002, 2025.

[34] J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the National Academy of Sciences*, vol. 79, pp. 2554–2558, 1982.

[35] D. H. Ackley, G. Hinton, and T. J. Sejnowski, "A learning algorithm for Boltzmann machines," *Cognitive Science*, vol. 9, pp. 147–169, 1985.

[36] M. Gilson, E. Tagliazucchi, and R. Cofré, "Entropy production of multivariate Ornstein-Uhlenbeck processes correlates with consciousness levels in the human brain," *Physical Review E*, vol. 107, no. 024121, 2023.

[37] G. Deco, Y. Sanz-Perl, H. Bocaccio, E. Tagliazucchi, and M. L. Kringelbach, "The INSIDEOUT framework provides precise signatures of the balance of intrinsic and extrinsic dynamics in brain states," *Communications Biology*, vol. 5, no. 572, 2022.

[38] K. Shinozuka, P. Tewarie, A. Luppi, C. Lynn, L. Roseman, S. D. Muthukumaraswamy, D. J. Nutt, R. Carhart-Harris, G. Deco, and M. L. Kringelbach, "LSD flattens the hierarchy of directed information flow in fast whole-brain dynamics," *Imaging Neuroscience*, vol. 3, 2025.

[39] J. Cruzat, R. Herzog, P. Prado, Y. Sanz-Perl, R. Gonzalez-Gomez, S. Moguilner, M. Kringelbach, G. Deco, E. Tagliazucchi, and A. Ibañez, "Temporal irreversibility of large-scale brain dynamics in Alzheimer's disease," *Journal of Neuroscience*, vol. 43, no. 9, p. 1643–1656, 2023.

[40] E. Guzmán, Y. Sanz Perl, J. Vohryzek, A. Escrichs, D. Manasova, B. Türker, E. Tagliazucchi, M. Kringelbach, J. Sitt, and G. Deco, "The lack of temporal brain dynamics asymmetry as a signature of impaired consciousness states," *Interface Focus*, vol. 13, no. 3, 2023.

[41] E. Schrödinger, *What is Life? The Physical Aspect of the Living Cell and Mind.* Cambridge, United Kingdom: Cambridge University Press, 1944.

[42] X. Fang, K. Kruse, T. Lu, and J. Wang, "Nonequilibrium physics in biology," *Reviews of Modern Physics*, vol. 91, no. 045004, 2019.

[43] F. S. Gnesotto, F. Mura, J. Gladrow, and C. P. Broedersz, "Broken detailed balance and non-equilibrium dynamics in living systems: a review," *Reports on Progress in Physics*, vol. 81, no. 066601, 2018.

[44] U. Seifert, *Stochastic Thermodynamics.* Cambridge University Press, 2025.

[45] M. Esposito, "Stochastic thermodynamics under coarse graining," *Physical Review E*, vol. 85, no. 041125, 2012.

[46] D. H. Wolpert, "The stochastic thermodynamics of computation," *Journal of Physics A: Mathematical and Theoretical*, vol. 52, no. 193001, 2019.

[47] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa, "Thermodynamics of information," *Nature Physics*, vol. 11, pp. 131–139, 2015.

[48] S. H. Strogatz, "Exploring complex networks," *Nature*, vol. 410, p. 268–276, 2001.

[49] C. Bick, E. Gross, H. A. Harrington, and M. T. Schaub, "What are higher-order networks?," *SIAM Review*, vol. 65, no. 3, pp. 686–731, 2023.

[50] R. Brown, "A brief account of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies," *Philosophical Magazine*, vol. 4, no. 161, 1828.

[51] D. Chandler, *Introduction to Modern Statistical Mechanics*. Oxford University Press, 1987.

[52] M. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics*. Oxford University Press, 1999.

[53] A. Borst and F. E. Theunissen, "Information theory and neural coding," *Nature Neuroscience*, vol. 2, pp. 947–957, 1999.

[54] G. Tononi, O. Sporns, and G. M. Edelman, "A measure for brain complexity: relating functional segregation and integration in the nervous system," *Proceedings of the National Academy of Sciences*, vol. 91, no. 11, pp. 5033–5037, 1994.

[55] L. D. Costa and G. A. Pavliotis, "The entropy production of stationary diffusions," *Journal of Physics A: Mathematical and Theoretical*, vol. 56, no. 36, p. 365001, 2023.

[56] G. Pavliotis, *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*. Springer, 2014.

[57] D.-Q. Jiang, M. Qian, and M.-P. Qian, *Mathematical Theory of Nonequilibrium Steady States: On the Frontier of Probability and Dynamical Systems*. Springer, 2004.

[58] E. Roldán, *Irreversibility and Dissipation in Microscopic Systems*. Springer, 2014.

[59] U. Seifert, "Stochastic thermodynamics, fluctuation theorems and molecular machines," *Reports on Progress in Physics*, vol. 475, no. 12, 2012.

[60] E. H. Feng and G. E. Crooks, "Length of time's arrow," *Physical Review Letters*, vol. 101, no. 090602, 2008.

[61] P. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*. Springer, 1992.

[62] E. Glötzl and O. Richters, "Helmholtz decomposition and potential functions for n-dimensional analytic vector fields," *Journal of Mathematical Analysis and Applications*, vol. 525, no. 2, p. 127138, 2023.

[63] H. Qian, M. Qian, and X. Tang, "Thermodynamics of the general diffusion process: Time-reversibility and entropy production," *Journal of Statistical Physics*, vol. 107, no. 5, 2002.

[64] A. Strang, *Applications of the Helmholtz-Hodge decomposition to networks and random processes*. PhD thesis, Case Western Reserve University, August 2020.

[65] R. Penrose, "A generalized inverse for matrices," *Proceedings of the Cambridge Philosophical Society*, vol. 51, p. 406–413, 1955.

[66] Y. Kuramoto, *Chemical Oscillations, Waves, and Turbulence*. Springer-Verlag, 1984.

[67] C. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. Springer, 2003.

[68] R. E. Spinney and I. J. Ford, "Entropy production in full phase space for continuous stochastic dynamics," *Physical Review E*, vol. 85, no. 051113, 2012.

[69] H. B. G. Casimir, "On Onsager's principle of microscopic reversibility," *Reviews of Modern Physics*, vol. 17, no. 343, 1945.

[70] J. S. W. Lamb and J. A. G. Roberts, "Time-reversal symmetry in dynamical systems: a survey," *Physica D: Nonlinear Phenomena*, vol. 112, no. 1–2, pp. 1–39, 1998.

[71] D. Ruelle, "Entropy production in nonequilibrium statistical mechanics," *Communications in Mathematical Physics*, vol. 189, pp. 365–371, 1997.

[72] J. K. Hale, *Asymptotic Behavior of Dissipative Systems*, vol. 25 of *Mathematical Surveys and Monographs*. American Mathematical Society, 1988.

[73] C. Van den Broeck and M. Esposito, "Three faces of the second law. II. Fokker–Planck formulation," *Physical Review E*, vol. 82, no. 1, p. 011144, 2010.

[74] M. Esposito and C. Van den Broeck, "Three faces of the second law. I. Master equation formulation," *Physical Review E*, vol. 82, no. 1, p. 011143, 2010.

[75] R. Balescu, *Statistical Dynamics: Matter Out of Equilibrium*. Imperial College Press, 1997.

[76] J. R. Norris, *Markov Chains*. Cambridge University Press, 1998.

[77] J. Schnakenberg, "Network theory of microscopic and macroscopic behavior of master equation systems," *Reviews of Modern Physics*, vol. 48, no. 571, 1976.

[78] D. M. Busiello, D. Gupta, and A. Maritan, "Entropy production in systems with unidirectional transitions," *Physical Review Research*, vol. 2, no. 023011, 2020.

[79] N. Masuda, M. Porter, and R. Lambiotte, "Random walks and diffusion on networks," *Physics Reports*, vol. 716-717, pp. 1–58, 2017.

[80] R. Lambiotte, J.-C. Delvenne, and M. Barahona, "Random walks, Markov processes and the multiscale modular organization of complex networks," *IEEE Transactions on Network Science and Engineering*, vol. 1, no. 2, pp. 76–90, 2014.

[81] J.-C. Delvenne, S. N. Yaliraki, and M. Barahona, "Stability of graph communities across time scales," *Proceedings of the National Academy of Sciences*, vol. 107, no. 29, pp. 12755–12760, 2010.

[82] J.-C. Delvenne, R. Lambiotte, and L. E. C. Rocha, "Diffusion on networked systems is a question of time or structure," *Nature Communications*, vol. 6, no. 7366, 2015.

[83] R. Lambiotte, "Continuous-time random walks and temporal networks," in *Temporal Network Theory*, pp. 225–239, Springer, 2023.

[84] L.-H. Lim, "Hodge Laplacians on graphs," *SIAM Review*, vol. 62, no. 3, pp. 685–715, 2020.

[85] S. Atasoy, I. Donnelly, and J. Pearson, "Human brain networks function in connectome-specific harmonic waves," *Nature Communications*, vol. 7, no. 10340, 2016.

[86] T. Martin, X. Zhang, and M. E. J. Newman, "Localization and centrality in networks," *Physical Review E*, vol. 90, no. 052808, 2014.

[87] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, 2007.

[88] C. C. Paige and M. A. Saunders, "LSQR: Sparse linear equations and least squares problems," *ACM Transactions on Mathematical Software*, vol. 8, no. 2, pp. 195—209, 1982.

[89] R. MacKay, S. Johnson, and B. Sansom, "How directed is a directed network?," *Royal Society Open Science*, vol. 7, no. 9, 2020.

[90] A. Strang, K. C. Abbott, and P. J. Thomas, "The network HHD: Quantifying cyclic competition in trait-performance models of tournaments," *SIAM Review*, vol. 64, no. 2, 2022.

[91] X. Jiang, L.-H. Lim, Y. Yao, and Y. Ye, "Statistical ranking and combinatorial Hodge theory," *Mathematical Programming*, vol. 127, p. 203–244, 2011.

[92] M. Schaub, J. Seby, F. Frantzen, T. Roddenberry, Y. Zhu, and S. Segarra, "Signal processing on simplicial complexes," in *Higher Order Systems*, Springer, 2022.

[93] M. Fruchart, R. Hanai, P. B. Littlewood, and V. Vitelli, "Non-reciprocal phase transitions," *Nature*, vol. 592, pp. 363–369, 2021.

[94] A. Ivlev, J. Bartnick, M. Heinen, C. Du, V. Nosenko, and H. Löwen, "Statistical mechanics where Newton's third law is broken," *Physical Review X*, vol. 5, no. 011035, 2015.

[95] P. Kale, A. Zalesky, and L. Gollo, "Estimating the impact of structural directionality: How reliable are undirected connectomes?," *Network Neuroscience*, vol. 2, no. 2, pp. 259–284, 2018.

[96] M. Asllani, R. Lambiotte, and T. Carletti, "Structure and dynamical behavior of non-normal networks," *Science Advances*, vol. 4, no. 12, 2018.

[97] M. Asllani, J. D. Challenger, F. S. Pavone, L. Sacconi, and D. Fanelli, "The theory of pattern formation on directed networks," *Nature Communications*, vol. 5, no. 4517, 2014.

[98] R. Muolo, T. Carletti, J. P. Gleeson, and M. Asllani, "Synchronization dynamics in non-normal networks: The trade-off for optimality," *Entropy*, vol. 23, no. 36, 2020.

[99] S. Johnson, "Digraphs are different: why directionality matters in complex systems," *Journal of Physics: Complexity*, vol. 1, no. 015003, 2020.

[100] S. Johnson, V. Domínguez-García, L. Donetti, and M. A. Muñoz, "Trophic coherence determines food-web stability," *Proceedings of the National Academy of Sciences*, vol. 111, no. 50, pp. 17923–17928, 2014.

[101] C. D. Bacco, D. B. Larremore, and C. Moore, "A physical model for efficient ranking in networks," *Science Advances*, vol. 4, no. 7, 2018.

[102] D. Garlaschelli and M. I. Loffredo, "Patterns of link reciprocity in directed networks," *Physical Review Letters*, vol. 93, no. 268701, 2004.

[103] T. Squartini, F. Picciolo, F. Ruzzenenti, and D. Garlaschelli, "Reciprocity of weighted networks," *Scientific Reports*, vol. 3, no. 2729, 2013.

[104] G. Hennequin, T. P. Vogels, and W. Gerstner, "Non-normal amplification in random balanced neuronal networks," *Physical Review E*, vol. 86, no. 011909, 2012.

[105] G. Baggio, V. Rutten, G. Hennequin, and S. Zampieri, "Efficient communication over complex dynamical networks: The role of matrix non-normality," *Science Advances*, vol. 6, no. 22, 2020.

[106] R. M. Gray, *Toeplitz and Circulant Matrices: A Review*. Now Publishers, 2006.

[107] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Transactions on Information Theory*, vol. 37, no. 1, 1991.

[108] C. Godrèche and J. Luck, "Characterising the nonequilibrium stationary states of Ornstein-Uhlenbeck processes," *Journal of Physics A: Mathematical and Theoretical*, vol. 52, no. 3, 2018.

[109] F. Rotella and P. Borne, "Explicit solution of Sylvester and Lyapunov equations," *Mathematics and Computers in Simulation*, vol. 31, no. 3, pp. 271–281, 1989.

[110] V. Simoncini, "Computational methods for linear matrix equations," *SIAM Review*, vol. 58, no. 3, 2016.

[111] A. Schwarze and M. Porter, "Motifs for processes on networks," *SIAM Journal on Applied Dynamical Systems*, vol. 20, no. 4, 2021.

[112] V. R. Nareddy and J. Machta, "Kinetic Ising models with self-interaction: Sequential and parallel updating," *Physical Review E*, vol. 101, no. 012122, 2020.

[113] M. Aguilera, S. A. Moosavi, and H. Shimazaki, "A unifying framework for mean-field theories of asymmetric kinetic Ising systems," *Nature Communications*, vol. 58, 2021.

[114] J. A. Pachter, Y.-J. Yang, and K. A. Dill, "Entropy, irreversibility and inference at the foundations of statistical physics," *Nature Review Physics*, vol. 6, p. 382–393, 2024.

[115] R. J. Glauber, "Time-dependent statistics of the Ising model," *Journal of Mathematical Physics*, vol. 4, pp. 94–307, 1963.

[116] M. Aguilera, M. Igarashi, and H. Shimazaki, "Nonequilibrium thermodynamics of the asymmetric Sherrington-Kirkpatrick model," *Nature Communications*, vol. 14, no. 3685, 2023.

[117] M. Timme and J. Casadiego, "Revealing networks from dynamics: an introduction," *Journal of Physics A: Mathematical and Theoretical*, vol. 47, no. 343001, 2014.

[118] O. M. Cliff, A. G. Bryant, J. T. Lizier, N. Tsuchiya, and B. D. Fulcher, "Unifying pairwise interactions in complex dynamics," *Nature Computational Science*, vol. 3, pp. 883–893, 2023.

[119] A. Frishman and P. Ronceray, "Learning force fields from stochastic trajectories," *Physical Review X*, vol. 10, no. 021009, 2020.

[120] J. Li, J. M. Horowitz, T. R. Gingrich, and N. Fakhri, "Quantifying dissipation using fluctuating currents," *Nature Communications*, vol. 10, no. 1666, 2019.

[121] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications*. Springer, 2017.

[122] A. C. Rencher and W. F. Christensen, *Methods of Multivariate Analysis*. John Wiley & Sons, 2012.

[123] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[124] D. Benozzo, G. Baggio, G. Baron, A. Chiuso, S. Zampieri, and A. Bertoldo, "Analyzing asymmetry in brain hierarchies with a linear state-space model of resting-state fMRI data," *Network Neuroscience*, pp. 1–24, 2023.

[125] D. Van Essen, S. Smith, D. Barch, T. Behrens, E. Yacoub, K. Ugurbil, and for the WU-Minn HCP Consortium, "The WU-Minn Human Connectome Project: An overview," *NeuroImage*, vol. 80, pp. 62–79, 2013.

[126] D. M. Barch, G. C. Burgess, M. P. Harms, S. E. Petersen, B. L. Schlaggar, M. Corbetta, M. F. Glasser, S. Curtiss, S. Dixit, C. Feldt, D. Nolan, E. Bryant, T. Hartley, O. Footer, J. M. Bjork, R. Poldrack, S. Smith, H. Johansen-Berg, A. Z. Snyder, and D. C. V. Essen, "Function in the human connectome: Task-fMRI and individual differences in behavior," *NeuroImage*, vol. 80, pp. 169–189, 2013.

[127] R. Desikan, F. Ségonne, B. Fischl, B. Quinn, B. Dickerson, D. Blacker, R. Buckner, A. Dale, R. Maguire, B. Hyman, M. Albert, and R. Killiany, "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," *Neuroimage*, vol. 31, no. 3, pp. 968–980, 2006.

[128] G. Deco, D. Vidaurre, and M. L. Kringelbach, "Revisiting the global workspace orchestrating the hierarchical organization of the human brain," *Nature Human Behaviour*, vol. 5, pp. 497–511, 2021.

[129] E. Nozari, M. A. Bertolero, J. Stiso, L. Caciagli, E. J. Cornblath, X. He, A. S. Mahadevan, G. J. Pappas, and D. S. Bassett, "Macroscopic resting-state brain dynamics are best described by linear models," *Nature Biomedical Engineering*, 2023.

[130] B. T. T. Yeo, F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zöllei, J. R. Polimeni, B. Fischl, H. Liu, and R. L. Buckner, "The organization of the human cerebral cortex estimated by intrinsic functional connectivity," *Journal of Neurophysiology*, vol. 106, no. 3, pp. 1125–1165, 2011.

[131] L. Chanes and L. F. Barrett, "Redefining the role of limbic areas in cortical processing," *Trends in Cognitive Sciences*, vol. 20, no. 2, 2016.

[132] A. Santoro, F. Battiston, G. Petri, and E. Amico, "Higher-order organization of multivariate time series," *Nature Physics*, vol. 19, p. 221–229, 2023.

[133] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, pp. 424–438, 1969.

[134] O. Shirokikh, G. Pastukhov, A. Semenov, S. Butenko, A. Veremyev, E. L. Pasiliao, and V. Boginski, "Networks of causal relationships in the U.S. stock market," *Dependence Modelling*, vol. 10, no. 1, 2022.

[135] A. Scherbina and B. Schlusche, "Follow the leader: Using the stock market to uncover information flows between firms," *Review of Finance*, vol. 24, no. 1, p. 189–225, 2020.

[136] K. Friston, "Hierarchical models in the brain," *PLOS Computational Biology*, vol. 4, no. 11, 2008.

[137] S. Dehaene, M. Kerszberg, and J.-P. Changeux, "A neuronal model of a global workspace in effortful cognitive tasks," *Proceedings of the National Academy of Sciences*, vol. 95, no. 24, pp. 14529–14534, 1998.

[138] S. Johnson and N. S. Jones, "Looplessness in networks is linked to trophic coherence," *Proceedings of the National Academy of Sciences*, vol. 114, no. 22, pp. 5618–5623, 2017.

[139] J. Hertz, Y. Roudi, and J. Tyrcha, "Ising models for inferring network structure from spike data," in *Principles of Neural Coding*, Taylor & Francis, 2013.

[140] R. Lambiotte, M. Rosvall, and I. Scholtes, "From networks to optimal higher-order models of complex systems," *Nature Physics*, vol. 15, p. 313–320, 2019.

[141] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri, "Networks beyond pairwise interactions: Structure and dynamics," *Physics Reports*, vol. 874, pp. 1–92, 2020.

[142] M. R. R. Tabar, F. Nikakhtar, L. Parkavousi, A. Akhshi, U. Feudel, and K. Lehnertz, "Revealing higher-order interactions in high-dimensional complex systems: A data-driven approach," *Physical Review X*, vol. 14, no. 011050, 2024.

[143] Z. Liu, R. Peach, P. A. M. Mediano, and M. Barahona, "Interaction measures, partition lattices and kernel tests for high-order interactions," in *Advances in Neural Information Processing Systems*, vol. 36, pp. 36991–37012, 2023.

[144] Z. Liu, M. Barahona, and R. Peach, "Information-theoretic measures on lattices for higher-order interactions," in *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*, vol. 258 of *Proceedings of Machine Learning Research*, pp. 2206–2214, PMLR, 2025.

[145] P. A. Mediano, F. E. Rosas, A. I. Luppi, R. L. Carhart-Harris, D. Bor, A. K. Seth, and A. B. Barrett, "Towards an extended taxonomy of information dynamics via integrated information decomposition," *arXiv*, vol. 2109.13186, 2021.

[146] P. A. M. Mediano, F. E. Rosas, A. I. Luppi, H. J. Jensen, A. K. Seth, A. B. Barrett, R. L. Carhart-Harris, and D. Bor, "Greater than the parts: a review of the information decomposition approach to causal emergence," *Philosophical Transactions of the Royal Society A*, vol. 380, no. 2227, 2022.

[147] F. E. Rosas, P. A. Mediano, M. Gastpar, and H. J. Jensen, "Quantifying high-order interdependencies via multivariate extensions of the mutual information," *Physical Review E*, vol. 100, no. 032305, 2019.

[148] G. Petri, P. Expert, F. Turkheimer, R. Carhart-Harris, D. Nutt, P. J. Hellyer, and F. Vaccarino, "Homological scaffolds of brain functional networks," *Journal of the Royal Society Interface*, vol. 11, no. 101, 2014.

[149] T. F. Varley and O. Sporns, "Network analysis of time series: Novel approaches to network neuroscience," *Frontiers in Neuroscience*, vol. 15, no. 787068, 2022.

[150] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuño, "From time series to complex networks: The visibility graph," *Proceedings of the National Academy of Sciences*, vol. 105, no. 13, pp. 4972–4975, 2008.

[151] L. Bonetti, G. F. Rubio, F. Carlomagno, M. Dietz, D. Pantazis, P. Vuust, and M. Kringelbach, "Spatiotemporal brain hierarchies of auditory memory recognition and predictive coding," *Nature Communications*, vol. 15, no. 4313, 2024.

[152] T. Schreiber, "Measuring information transfer," *Physical Review Letters*, vol. 85, no. 461, 2000.

[153] G. Sugihara, R. May, H. Ye, C.-h. Hsieh, E. Deyle, M. Fogarty, and S. Munch, "Detecting causality in complex ecosystems," *Science*, vol. 338, no. 6106, pp. 496–500, 2012.

[154] H. Ye, E. R. Deyle, L. J. Gilarranz, and G. Sugihara, "Distinguishing time-delayed causal interactions using convergent cross mapping," *Scientific Reports*, vol. 5, p. 14750, 2015.

[155] A. Crisanti, A. Puglisi, and D. Villamaina, "Nonequilibrium and information: The role of cross correlations," *Physical Review E*, vol. 85, no. 061127, 2012.

[156] I. Z. Steinberg, "On the time reversal of noise signals," *Biophysical Journal*, vol. 50, no. 1, pp. 171–179, 1986.

[157] R. S. Liptser and A. N. Shiryaev, *Statistics of Random Processes: I. General Theory*. Springer, 2001.

[158] J. G. Kemeny and J. L. Snell, *Finite Markov Chains*. Springer-Verlag, 1960.

[159] C. W. Lynn, C. M. Holmes, W. Bialek, and D. J. Schwab, "Decomposing the local arrow of time in interacting systems," *Physical Review Letters*, vol. 129, no. 118101, 2022.

[160] T. F. Varley, "Information theory for complex systems scientists: What, why, and how," *Physics Reports*, vol. 1148, pp. 1–55, 2025.

[161] L. Lacasa, V. Nicosia, and V. Latora, "Network structure of multivariate time series," *Scientific Reports*, vol. 5, no. 15508, 2015.

[162] L. Lacasa, A. Nuñez, E. Roldán, J. Parrondo, and B. Luque, "Time series irreversibility: a visibility graph approach," *The European Physical Journal B*, vol. 85, no. 217, 2012.

[163] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[164] K.-S. Chan, L.-H. Ho, and H. Tong, "A note on time-reversibility of multivariate linear processes," *Biometrika*, vol. 93, no. 1, pp. 221–227, 2006.

[165] R. Nartallo-Kaluarachchi, L. Bonetti, G. Fernandez-Rubio, P. Vuust, G. Deco, M. L. Kringelbach, R. Lambiotte, and A. Goriely, "Multilevel irreversibility reveals higher-order organisation of non-equilibrium interactions in human brain dynamics." https://doi.org/10.5281/zenodo.13939016, 2024.

[166] G. Deco, Y. S. Perl, A. Ponce-Alvarez, E. Tagliazucchi, P. C. Whybrow, J. Fuster, and M. L. Kringelbach, "One ring to rule them all: The unifying role of prefrontal cortex in steering task-related brain dynamics," *Progress in Neurobiology*, vol. 227, no. 102468, 2023.

[167] A. Treves and S. Panzeri, "The upward bias in measures of information derived from limited data samples," *Neural Computation*, vol. 7, no. 2, pp. 399—-407, 1995.

[168] S. Koelsch, P. Vuust, and K. Friston, "Predictive processes and the peculiar case of music," *Trends in Cognitive Sciences*, vol. 23, no. 1, pp. 63–77, 2019.

[169] K. Friston, "A theory of cortical responses," *Philosophical Transactions of the Royal Society B*, vol. 360, no. 1456, pp. 815–836, 2005.

[170] T. A. Hackett, "Information flow in the auditory cortical network," *Hearing Research*, vol. 271, no. 1, pp. 133–146, 2011.

[171] K. Okada, F. Rong, J. Venezia, W. Matchin, I.-H. Hsieh, K. Saberi, J. T. Serences, and G. Hickok, "Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech," *Cerebral Cortex*, vol. 20, no. 10, pp. 2486–2495, 2010.

[172] F. E. Rosas, P. A. M. Mediano, A. I. Luppi, T. F. Varley, J. T. Lizier, S. Stramaglia, H. J. Jensen, and D. Marinazzo, "Disentangling high-order mechanisms and high-order behaviours in complex systems," *Nature Physics*, vol. 18, pp. 476–477, 2022.

[173] R. Friedrich, J. Peinke, M. Sahimi, and M. R. R. Tabar, "Approaching complexity by stochastic methods: From biological systems to turbulence," *Physics Reports*, vol. 506, no. 5, pp. 87–162, 2011.

[174] C. Battle, C. P. Broedersz, N. Fakhri, V. F. Geyer, J. Howard, C. F. Schmidt, and F. C. Mackintosh, "Broken detailed balance at mesoscopic scales in active biological systems," *Science*, vol. 352, no. 6285, pp. 604–607, 2016.

[175] J. Paijmans, D. K. Lubensky, and P. R. ten Wolde, "A thermodynamically consistent model of the post-translational Kai circadian clock," *PLOS Computational Biology*, vol. 13, no. 3, p. e1005415, 2017.

[176] J. C. Kimmel, A. Y. Chang, A. S. Brack, and W. F. Marshall, "Inferring cell state by quantitative motility analysis reveals a dynamic state system and broken detailed balance," *PLOS Computational Biology*, vol. 14, no. 1, p. e1005927, 2018.

[177] P. Billingsley, *Statistical Inference for Markov Processes*. The University of Chicago Press, 1961.

[178] M. Bladt and M. Sørensen, "Statistical inference for discretely observed Markov jump processes," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, no. 3, p. 395–410, 2005.

[179] T. Schwarz, A. B. Kolomeisky, and A. Godec, "Consistent time reversal and reliable and accurate inference in the presence of memory," *arXiv*, vol. 2410.11819, 2024.

[180] I. D. Terlizzi, M. Gironella, D. Herraez-Aguilar, T. Betz, F. Monroy, M. Baiesi, and F. Ritort, "Variance sum rule for entropy production," *Science*, vol. 383, no. 6686, pp. 971–976, 2024.

[181] C. K. Peng, S. V. Buldyrev, J. M. Hausdorff, S. Havlin, J. E. Mietus, M. Simons, H. E. Stanley, and A. L. Goldberger, "Non-equilibrium dynamics as an indispensable characteristic of a healthy biological system," *Integrative Psychological and Behavioral Science*, vol. 29, no. 3, pp. 283–293, 2019.

[182] M. Costa, A. L. Goldberger, and C.-K. Peng, "Broken asymmetry of the human heartbeat: loss of time irreversibility in aging and disease," *Physical Review Letters*, vol. 95, no. 198102, 2005.

[183] R. Gernert, C. Emary, and S. H. L. Klapp, "Waiting time distribution for continuous stochastic systems," *Physical Review E*, vol. 90, no. 062115, 2014.

[184] G. Falasco and M. Esposito, "Local detailed balance across scales: From diffusions to jump processes and beyond," *Physical Review E*, vol. 103, no. 042114, 2021.

[185] A. Ghosala and G. Bisker, "Inferring entropy production rate from partially observed Langevin dynamics under coarse-graining," *Physical Chemistry Chemical Physics*, vol. 24, no. 39, pp. 24021–24031, 2022.

[186] E. Meyberg, J. Degünther, and U. Seifert, "Entropy production from waiting-time distributions for overdamped Langevin dynamics," *Journal of Physics A: Mathematical and Theoretical*, vol. 57, no. 25LT01, 2024.

[187] I. A. Martínez, G. Bisker, J. M. Horowitz, and J. M. R. Parrondo, "Inferring broken detailed balance in the absence of observable currents," *Nature Communications*, vol. 10, no. 3542, 2019.

[188] D. J. Skinner and J. Dunkel, "Estimating entropy production from waiting time distributions," *Physical Review Letters*, vol. 127, no. 198101, 2021.

[189] M. Rosvall, A. V. Esquivel, A. Lancichinetti, J. D. West, and R. Lambiotte, "Memory in network flows and its effects on spreading dynamics and community detection," *Nature Communications*, vol. 5, no. 4630, 2014.

[190] R. Sahasrabuddhe, R. Lambiotte, and M. Rosvall, "Concise network models of memory dynamics reveal explainable patterns in path data," *Science Advances*, vol. 11, no. eadw4544, 2025.

[191] D. M. Busiello, J. Hidalgo, and A. Maritan, "Entropy production for coarse-grained dynamics," *New Journal of Physics*, vol. 21, no. 073004, 2019.

[192] M. Heida, M. Kantner, and A. Stephan, "Consistency and convergence for a family of finite volume discretizations of the Fokker–Planck operator," *ESAIM: Mathematical Modelling and Numerical Analysis*, vol. 55, p. 3017–3042, 2021.

[193] V. Holubec, K. Kroy, and S. Steffenoni, "Physically consistent numerical solver for time-dependent Fokker-Planck equations," *Physical Review E*, vol. 99, p. 032117, 2019.

[194] L. Pareschi and M. Zanella, "Structure preserving schemes for nonlinear Fokker–Planck equations and applications," *Journal of Scientific Computing*, vol. 74, p. 1575–1600, 2018.

[195] S. Engblom, L. Ferm, A. Hellander, and P. Lötstedt, "Simulation of stochastic reaction-diffusion processes on unstructured meshes," *SIAM Journal on Scientific Computing*, vol. 31, no. 3, 2009.

[196] L. Meinecke, S. Engblom, A. Hellander, and P. Lötstedt, "Analysis and design of jump coefficients in discrete stochastic diffusion models," *SIAM Journal on Scientific Computing*, vol. 38, no. 1, 2016.

[197] S. A. Isaacson and Y. Zhang, "An unstructured mesh reaction-drift-diffusion master equation with reversible reactions," *Bulletin of Mathematical Biology*, vol. 87, no. 13, 2025.

[198] W. R. Frensley, "Scharfetter-Gummel discretization scheme for drift-diffusion equations," tech. rep., The University of Texas at Dallas, 2004.

[199] M. Bessemoulin-Chatard, "A finite volume scheme for convection–diffusion equations with nonlinear diffusion derived from the Scharfetter–Gummel scheme," *Numerische Mathematik*, vol. 121, pp. 637–670, 2012.

[200] A. Schlichting and C. Seis, "The Scharfetter–Gummel scheme for aggregation–diffusion equations," *IMA Journal of Numerical Analysis*, vol. 42, pp. 2361–2402, 2022.

[201] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 2 ed., 2013.

[202] J. Sheth, S. W. F. Meenderink, P. M. Quiñones, D. Bozovic, and A. J. Levine, "Nonequilibrium limit-cycle oscillators: Fluctuations in hair bundle dynamics," *Physical Review E*, vol. 97, no. 062411, 2018.

[203] S. Gupta, A. Campa, and S. Ruffo, "Kuramoto model of synchronization: equilibrium and nonequilibrium aspects," *Journal of Statistical Mechanics: Theory and Experiment*, p. R08001, 2014.

[204] H. Sakaguchi and Y. Kuramoto, "A soluble active rotator model showing phase transitions via mutual entrainment," *Progress of Theoretical Physics*, vol. 76, no. 3, pp. 576–581, 1986.

[205] J. F. C. Kingman, "The imbedding problem for finite Markov chains," *Z. Wahrscheinlichkeitstheorie verw Gebiete*, vol. 1, pp. 14–24, 1962.

[206] J. Toner and Y. Tu, "Flocks, herds, and schools: A quantitative theory of flocking," *Physical Review E*, vol. 58, no. 4, 1998.

[207] W. Bialek, A. Cavagna, I. Giardina, T. Mora, E. Silvestri, M. Viale, and A. M. Walczak, "Statistical mechanics for natural flocks of birds," *Proceedings of the National Academy of Sciences*, vol. 109, no. 13, pp. 4786–4791, 2012.

[208] T. Mora, A. M. Walczak, L. D. Castello, F. Ginelli, S. Melillo, L. Parisi, M. Viale, A. Cavagna, and I. Giardina, "Local equilibrium in bird flocks," *Nature Physics*, vol. 12, p. 1153–1157, 2016.

[209] E. Fodor, C. Nardini, M. E. Cates, J. Tailleur, P. Visco, and F. van Wijland, "How far from equilibrium is active matter?," *Physical Review Letters*, vol. 117, no. 038103, 2016.

[210] J. Jhawar, R. G. Morris, U. R. Amith-Kumar, M. D. Raj, T. Rogers, H. Rajendran, and V. Guttal, "Noise-induced schooling of fish," *Nature Physics*, vol. 16, pp. 488–493, 2020.

[211] R. Sepulchre, D. A. Paley, and N. E. Leonard, "Stabilization of planar collective motion: All-to-all communication," *IEEE Transactions on Automatic Control*, vol. 52, no. 5, 2007.

[212] J. Wang, L. Xu, and E. Wang, "Potential landscape and flux framework of nonequilibrium networks: Robustness, dissipation,and coherence of biochemical oscillations," *Proceedings of the National Academy of Sciences*, vol. 105, no. 34, pp. 12271–12276, 2008.

[213] H. Yan, L. Zhao, L. Hu, X. Wang, E. Wang, and J. Wang, "Nonequilibrium landscape theory of neural networks," *Proceedings of the National Academy of Sciences*, vol. 100, no. 45, 2013.

[214] J. Wang, L. Xu, E. Wang, and S. Huang, "The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation," *Biophysical Journal*, vol. 99, p. 29–39, 2010.

[215] J. Wang, "Landscape and flux theory of non-equilibrium dynamical systems with application to biology," *Advances in Physics*, vol. 64, no. 1, pp. 1–137, 2015.

[216] H. Qian, "Nonequilibrium steady-state circulation and heat dissipation functional," *Physical Review E*, vol. 64, no. 022101, 2001.

[217] A. Gosztolai, R. L. Peach, A. Arnaudon, M. Barahona, and P. Vandergheynst, "MARBLE: interpretable representations of neural population dynamics using geometric deep learning," *Nature Methods*, vol. 22, p. 612–620, 2025.

[218] C. Moenning and N. A. Dodgson, "Fast marching farthest point sampling," tech. rep., Computer Laboratory, University of Cambridge, 2003.

[219] C. Fefferman, S. Mitter, and H. Narayanan, "Testing the manifold hypothesis," *Journal of the American Mathematical Society*, vol. 29, pp. 983–1049, 2016.

[220] H. Edelsbrunner and J. L. Harer, *Computational Topology: An Introduction*. American Mathematical Society, 2010.

[221] J. C. Sprott, *Elegant Chaos: Algebraically Simple Chaotic Flows*. World Scientific, 2010.

[222] M. Hutzenthaler, A. Jentzen, and P. E. Kloeden, "Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients," *Annals of Applied Probability*, vol. 22, no. 4, pp. 1611–1641, 2012.

[223] H. Turlier and T. Betz, "Unveiling the active nature of living-membrane fluctuations and mechanics," *Annual Review of Condensed Matter Physics*, vol. 10, pp. 213–232, 2019.

[224] R. Rodríguez-García, I. López-Montero, M. Mell, G. Egea, N. S. Gov, and F. Monroy, "Direct cytoskeleton forces cause membrane softening in red blood cells," *Biophysical Journal*, vol. 108, no. 12, pp. 2794–2806, 2015.

[225] I. D. Terlizzi, M. Gironella, D. Herraez-Aguilar, T. Betz, F. Monroy, M. Baiesi, and F. Ritort, "Data from: Variance sum rule for entropy production." https://doi.org/10.5281/zenodo.10472505, 2024.

[226] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, p. 215–220, 2000.

[227] S. Fazeli, "ECG heartbeat categorization dataset," *Kaggle*, 2018.

[228] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, "ECG heartbeat classification: A deep transferable representation," in *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, pp. 443–444, 2018.

[229] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Lecture Notes in Mathematics*, Springer-Verlag, 1981.

[230] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*. Cambridge University Press, 2003.

[231] A. W. C. Lau and T. C. Lubensky, "State-dependent diffusion: Thermodynamic consistency and its path integral formulation," *Physical Review E*, vol. 76, no. 011123, 2007.

[232] D. B. Brückner, P. Ronceray, and C. P. Broedersz, "Inferring the dynamics of underdamped stochastic systems," *Physical Review Letters*, vol. 125, no. 058103, 2020.

[233] A. Strang, D. Sewell, A. Kim, K. Alcedo, and D. Rosenbluth, "Principal trade-off analysis," *Information Visualization*, vol. 23, no. 3, 2024.

[234] M. T. Schaub, A. R. Benson, P. Horn, G. Lippner, and A. Jadbabaie, "Random walks on simplicial complexes and the normalized Hodge 1-Laplacian," *SIAM Review*, vol. 62, no. 2, pp. 353–391, 2020.

[235] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou, "Synchronization in complex networks," *Physics Reports*, vol. 469, no. 3, pp. 93–153, 2008.

[236] J. Cabral, E. Hugues, O. Sporns, and G. Deco, "Role of local network oscillations in resting-state functional connectivity," *Neuroimage*, vol. 57, no. 1, pp. 130–139, 2011.

[237] F. Hancock, F. E. Rosas, A. I. Luppi, M. Zhang, P. A. M. Mediano, J. Cabral, G. Deco, M. L. Kringelbach, M. Breakspear, J. A. S. Kelso, and F. E. Turkheimer, "Metastability demystified — the foundational past, the pragmatic present and the promising future," *Nature Reviews Neuroscience*, vol. 26, p. 82–100, 2025.

[238] G. Deco and M. L. Kringelbach, "Turbulent-like dynamics in the human brain," *Cell Reports*, vol. 33, no. 10, 2020.

[239] K. Bansal, J. O. Garcia, S. H. Tompson, T. Verstynen, J. M. Vettel, and S. F. Muldoon, "Cognitive chimera states in human brain networks," *Science Advances*, vol. 5, no. 4, 2019.

[240] G. Deco and M. L. Kringelbach, "Metastability and coherence: Extending the communication through coherence hypothesis using a whole-brain computational perspective," *Trends in Neurosciences*, vol. 39, no. 3, pp. 125–135, 2016.

[241] P. Fries, "A mechanism for cognitive dynamics: neuronal communication through neuronal coherence," *Trends in Cognitive Sciences*, vol. 9, no. 10, pp. 474–480, 2005.

[242] P. Fries, "Rhythms for cognition: Communication through coherence," *Neuron*, vol. 88, no. 1, pp. 220–235, 2015.

[243] P. M. Chaikin and T. C. Lubensky, *Principles of Condensed Matter Physics*. Cambridge University Press, 2012.

[244] J. Tobochnik and G. V. Chester, "Monte Carlo study of the planar spin model," *Physical Review B*, vol. 20, no. 9, 1979.

[245] G. Deco, Y. S. Perl, J. Feng, and M. L. Kringelbach, "The turbulent brain: Modelling vortex interactions for understanding human cognition," *Network Neuroscience*, pp. 1–25, 2026.

[246] Y. Xu, X. Long, J. Feng, and P. Gong, "Interacting spiral wave patterns underlie complex brain dynamics and are related to cognitive processing," *Nature Human Behaviour*, vol. 7, p. 1196–1215, 2023.

[247] M. J. P. Gingras and D. A. Huse, "Topological defects in the random-field XY model and the pinned vortex lattice to vortex glass transition in type-II superconductors," *Physical Review B*, vol. 53, no. 15193, 1996.

[248] X. Leoncini, A. D. Verga, and S. Ruffo, "Hamiltonian dynamics and the phase transition of the XY model," *Physical Review E*, vol. 57, no. 6377, 1998.

[249] F. A. Rodrigues, T. K. D. Peron, P. Ji, and J. Kurths, "The Kuramoto model in complex networks," *Physics Reports*, vol. 610, pp. 1–98, 2016.

[250] S. Atasoy, G. Deco, M. L. Kringelbach, and J. Pearson, "Harmonic brain modes: A unifying framework for linking space and time in brain dynamics," *The Neuroscientist*, vol. 24, no. 3, 2017.

[251] A. V. Knyazev, "Toward the optimal preconditioned eigensolver: Locally optimal block preconditioned conjugate gradient method," *SIAM Journal on Scientific Computing*, vol. 23, no. 2, 2001.

[252] V. P. Grande and M. T. Schaub, "Topological point cloud clustering," in *Proceedings of the 40th International Conference on Machine Learning (ICML 2023)*, 2023.

[253] S. Krishnagopal and G. Bianconi, "Spectral detection of simplicial communities via Hodge Laplacians," *Physical Review E*, vol. 104, no. 064303, 2021.

[254] R. A. Poldrack, J. A. Mumford, and T. E. Nichols, *Handbook of Functional MRI Data Analysis*. Cambridge University Press, 2024.

[255] K. Glomb, M. L. Kringelbach, G. Deco, P. Hagmann, J. Pearson, and S. Atasoy, "Functional harmonics reveal multi-dimensional basis functions underlying cortical organization," *Cell Reports*, 2021.

[256] N. Biggs, *Algebraic Graph Theory*. Cambridge University Press, 2012.

[257] A. Ortega, P. Frossard, J. Kovačević, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[258] J. Vohryzek, A. I. Luppi, S. Atasoy, G. Deco, R. L. Carhart-Harris, C. Timmermann, and M. L. Kringelbach, "N,N-dimethyltryptamine effects on connectome harmonics, subjective experience and comparative psychedelic experiences," *Neuropsychopharmacology*, vol. 50, 2025.

[259] M. Hasenbusch, "The two-dimensional XY model at the transition temperature: a high-precision Monte Carlo study," *Journal of Physics A: Mathematical and General*, vol. 38, no. 5869, 2005.

[260] J. M. Kosterlitz and D. J. Thouless, "Ordering, metastability and phase transitions in two-dimensional systems," *Journal of Physics C: Solid State Physics*, vol. 6, no. 1181, 1973.

[261] B. J. Kim, H. Hong, P. Holme, G. S. Jeon, P. Minnhagen, and M. Y. Choi, "XY model in small-world networks," *Physical Review E*, vol. 64, no. 056135, 2001.

[262] A. Schaefer, R. Kong, E. M. Gordon, T. O. Laumann, X.-N. Zuo, A. J. Holmes, S. B. Eickhoff, and B. T. T. Yeo, "Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI," *Cerebral Cortex*, vol. 28, no. 9, pp. 3095–3114, 2018.

[263] W. Kwak, J.-S. Yang, J. il Sohn, and I. mook Kim, "Critical behavior of the XY model on growing scale-free networks," *Physical Review E*, vol. 75, no. 061130, 2007.

[264] J.-S. Yang, W. Kwak, K.-I. Goh, and I.-m. Kim, "Critical behavior of the XY model on static scale-free networks," *Europhysics Letters*, vol. 84, no. 3, p. 36004, 2008.

[265] M. F. Glasser, S. N. Sotiropoulos, J. A. Wilson, T. S. Coalson, B. Fischl, J. L. Andersson, J. Xu, S. Jbabdi, M. Webster, J. R. Polimeni, D. C. V. Essen, M. Jenkinson, and for the WU-Minn HCP Consortium, "The minimal preprocessing pipelines for the Human Connectome Project," *NeuroImage*, vol. 18, pp. 105–124, 2013.

[266] S. M. Smith, C. F. Beckmann, J. Andersson, E. J. Auerbach, J. Bijsterbosch, G. Douaud, E. Duff, D. A. Feinberg, L. Griffanti, M. P. Harms, M. Kelly, T. Laumann, K. L. Miller, S. Moeller, S. Petersen, J. Power, G. Salimi-Khorshidi, A. Z. Snyder, A. T. Vu, M. W. Woolrich, J. Xu, E. Yacoub, K. Uğurbil, D. C. V. Essen, M. F. Glasser, and for the WU-Minn HCP Consortium, "Resting-state fMRI in the Human Connectome Project," *NeuroImage*, vol. 80, pp. 144–68, 2013.

[267] NetNeuroLab, "neuromaps issue #66." https://github.com/netneurolab/neuromaps/issues/66, n.d. Accessed: 2026-03-23.

[268] Thomas Yeo Lab, "Cbig issue #43." https://github.com/ThomasYeoLab/CBIG/issues/43, n.d. Accessed: 2026-03-23.

[269] M. X. Cohen, *Analyzing Neural Time Series Data: Theory and Practice*. MIT Press, 2014.

[270] A. Omidvarnia, R. Liégeois, E. Amico, M. G. Preti, A. Zalesky, and D. V. D. Ville, "On the spatial distribution of temporal complexity in resting state and task functional MRI," *Entropy*, vol. 24, no. 8, p. 1148, 2022.

[271] M. W. Mathis, A. P. Rotondo, E. F. Chang, A. S. Tolias, and A. Mathis, "Decoding the brain: From neural representations to mechanistic models," *Cell*, vol. 187, no. 21, pp. 5814–5832, 2024.

[272] S. Vyas, M. D. Golub, D. Sussillo, and K. V. Shenoy, "Computation through neural population dynamics," *Annual Review of Neuroscience*, vol. 43, 2020.

[273] D. Durstewitz, G. Koppe, and M. I. Thurm, "Reconstructing computational system dynamics from neural data with recurrent neural networks," *Nature Reviews Neuroscience*, vol. 24, p. 693–710, 2023.

[274] D. Sussillo and L. F. Abbott, "Generating coherent patterns of activity from chaotic neural networks," *Neuron*, vol. 63, no. 4, pp. 544–557, 2009.

[275] J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons," *Proceedings of the National Academy of Sciences*, vol. 81, pp. 3088–3092, 1984.

[276] D. J. Amit, *Modeling Brain Function: The World of Attractor Neural Networks*. Cambridge University Press, 1989.

[277] M. Khona and I. R. Fiete, "Attractor and integrator networks in the brain," *Nature Reviews Neuroscience*, vol. 23, p. 744–766, 2022.

[278] D. Hebb, *The Organization of Behavior: A Neuropsychological Theory*. Wiley, New York, 1949.

[279] H. Sompolinsky and I. Kanter, "Temporal association in asymmetric neural networks," *Physical Review Letters*, vol. 57, no. 2861, 1986.

[280] J. A. Hertz, G. Grinstein, and S. A. Solla, "Irreversible spin glasses and neural networks," in *Heidelberg Colloquium on Glassy Dynamics*, Springer, 2006.

[281] A. Crisanti and H. Sompolinsky, "Dynamics of spin systems with randomly asymmetric bonds: Langevin dynamics and a spherical model," *Physical Review A*, vol. 36, no. 4922, 1987.

[282] A. K. Behera, M. Rao, S. Sastry, and S. Vaikuntanathan, "Enhanced associative memory, classification, and learning with active dynamics," *Physical Review X*, vol. 13, no. 041043, 2023.

[283] T. Spisak and K. Friston, "Self-orthogonalizing attractor neural networks emerging from the free energy principle," *Neurocomputing*, vol. 682, no. 133472, 2026.

[284] M. Aguilera, D. D. Martino, I. Garashchuk, and D. Sinelshchikov, "Nonequilibrium thermodynamics of associative memory continuous-time recurrent neural networks," in *ALIFE 2025: Ciphers of Life: Proceedings of the Artificial Life Conference 2025*, 2025.

[285] A. Coolen, "Statistical mechanics of recurrent neural networks I—Statics," in *Handbook of Biological Physics*, vol. 4, pp. 553–618, Elsevier, 2001.

[286] A. Coolen, "Statistical mechanics of recurrent neural networks II — Dynamics," in *Handbook of Biological Physics*, vol. 4, pp. 619–684, Elsevier, 2001.

[287] M. G. Perich, D. Narain, and J. A. Gallego, "A neural manifold view of the brain," *Nature Neuroscience*, vol. 28, p. 1582–1597, 2025.

[288] J. A. Gallego, M. G. Perich, L. E. Miller, and S. A. Solla, "Neural manifolds for the control of movement," *Neuron*, vol. 94, no. 5, pp. 978–984, 2017.

[289] R. J. Gardner, E. Hermansen, M. Pachitariu, Y. Burak, N. A. Baas, B. A. Dunn, M.-B. Moser, and E. I. Moser, "Toroidal topology of population activity in grid cells," *Nature*, p. 123–128, 2022.

[290] D. Sussillo and O. Barak, "Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks," *Neural Computation*, vol. 25, no. 3, pp. 626–649, 2013.

[291] F. Mastrogiuseppe and S. Ostojic, "Linking connectivity, dynamics, and computations in low-rank recurrent neural networks," *Neuron*, vol. 99, no. 3, pp. 609–623, 2018.

[292] V. Schmutz, A. Haydaroğlu, S. Wang, Y. Feng, M. Carandini, and K. D. Harris, "High-dimensional neuronal activity from low-dimensional latent dynamics: a solvable model," in *39th Conference on Neural Information Processing Systems (NeurIPS 2025)*, vol. 38, 2025.

[293] V. Thibeault, A. Allard, and P. Desrosiers, "The low-rank hypothesis of complex systems," *Nature Physics*, vol. 20, pp. 294—-302, 2024.

[294] B. Prasse and P. V. Mieghem, "Predicting network dynamics without requiring the knowledge of the interaction graph," *Proceedings of the National Academy of Sciences*, vol. 119, no. 44, 2022.

[295] R. Mitchell-Heggs, S. Prado, G. P. Gava, M. A. Go, and S. R. Schultz, "Neural manifold analysis of brain circuit dynamics in health and disease," *Journal of Computational Neuroscience*, vol. 51, pp. 1–21, 2023.

[296] D. V. Buonomano, G. Buzsáki, L. Davachi, and A. C. Nobre, "Time for memories," *Journal of Neuroscience*, vol. 43, no. 45, pp. 7565–7574, 2023.

[297] P. J. Werbos, "Backpropagation through time: what it does and how to do it," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, 1990.

[298] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994.

[299] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.

[300] S. Dadras and H. R. Momeni, "A novel three-dimensional autonomous chaotic system generating two, three and four-scroll attractors," *Physics Letters A*, vol. 373, pp. 3637–3642, 2009.

[301] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 1901.

[302] M. M. Churchland, J. P. Cunningham, M. T. Kaufman, J. D. Foster, P. Nuyujukian, S. I. Ryu, and K. V. Shenoy, "Neural population dynamics during reaching," *Nature*, vol. 487, p. 51–56, 2012.

[303] J. P. Cunningham and B. M. Yu, "Dimensionality reduction for large-scale neural recordings," *Nature Neuroscience*, vol. 17, p. 1500–1509, 2014.

[304] C. J. Cueva, A. Ardalan, M. Tsodyks, and N. Qian, "Recurrent neural network models for working memory of continuous variables: activity manifolds, connectivity patterns, and dynamic codes," *arXiv*, vol. 2111.01275, 2021.

[305] R. Vidal, Y. Ma, and S. S. Sastry, *Generalized Principal Component Analysis*. Springer, 2016.

[306] H. Ninou, J. Kadmon, and N. A. Cayco-Gajic, "Curl descent: Non-gradient learning dynamics with sign-diverse plasticity," in *39th Conference on Neural Information Processing Systems (NeurIPS 2025)*, 2025.

[307] K. Friston and P. Ao, "Free energy, value, and attractors," *Computational and Mathematical Methods in Medicine*, vol. 2012, no. 937860, 2011.

[308] Y. LeCun, S. Chopra, R. Hadsell, M. Ranzato, and F. J. Huang, "Energy-based models," in *Predicting Structured Data*, MIT Press, 2007.

[309] S. L. Brunton, B. W. Brunton, J. L. Proctor, E. Kaiser, and J. N. Kutz, "Chaos as an intermittently forced linear system," *Nature Communications*, vol. 8, no. 19, 2017.

[310] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2009.

[311] V. I. Bogachev, *Measure Theory*. Springer, 2007.

[312] E. Pollock and M. Jazayeri, "Engineering recurrent neural networks from task-relevant manifolds and dynamics," *PLOS Computational Biology*, vol. 16, no. 8, p. e1008128, 2020.

[313] R. B. Yishai, R. L. Bar-Or, and H. Sompolinsky, "Theory of orientation tuning in visual cortex," *Proceedings of the National Academy of Sciences*, vol. 92, pp. 3844–3848, 1995.

[314] A. D. Bernardo, A. Valente, F. Mastrogiuseppe, and S. Ostojic, "Shaping manifolds in equivariant recurrent neural networks," *arXiv*, vol. 2511.04802v2, 2025.

[315] K. Rajan, C. D. Harvey, and D. W. Tank, "Recurrent network models of sequence generation and memory," *Neuron*, vol. 90, no. 1, pp. 128–142, 2016.

[316] J. C. Whittington and R. Bogacz, "Theories of error back-propagation in the brain," *Trends in Cognitive Sciences*, vol. 23, no. 3, pp. 235–250, 2019.

[317] C. Pandarinath, D. J. O'Shea, J. Collins, R. Jozefowicz, S. D. Stavisky, J. C. Kao, E. M. Trautmann, M. T. Kaufman, S. I. Ryu, L. R. Hochberg, J. M. Henderson, K. V. Shenoy, L. F. Abbott, and D. Sussillo, "Inferring single-trial neural population dynamics using sequential auto-encoders," *Nature Methods*, vol. 15, p. 805–815, 2018.

[318] S. L. Brunton, M. Budisic, E. Kaiser, and J. N. Kutz, "Modern Koopman theory for dynamical systems," *SIAM Review*, vol. 64, no. 2, 2022.

[319] K. Champion, B. Lusch, J. N. Kutz, and S. L. Brunton, "Data-driven discovery of coordinates and governing equations," *Proceedings of the National Academy of Sciences*, vol. 116, no. 45, pp. 22445–22451, 2019.

[320] P. Muratore and M. W. Mathis, "Extracting governing equations from latent dynamics via multi-view contrastive learning," *arXiv*, vol. 2606.13260, 2026.

[321] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *Proceedings of the 32nd International Conference on Machine Learning*, 2015.

[322] J. England, *Every Life is on Fire: How Thermodynamics Explains the Origins of Living Things*. Basic Books, 2020.

[323] D. Wolpert, J. Korbel, C. Lynn, F. Tasnim, J. Grochow, G. Kardeş, J. Aimone, V. Balasubramanian, E. de Giuli, D. Doty, N. Freitas, M. Marsili, T. E. Ouldridge, A. Richa, P. Riechers, E. Roldán, B. Rubenstein, Z. Toroczkai, and J. Paradiso, "Is stochastic thermodynamics the key to understanding the energy costs of computation?," *Proceedings of the National Academy of Sciences*, vol. 121, p. 45, 2024.

[324] K. Friston, "The free-energy principle: a unified brain theory?," *Nature Reviews Neuroscience*, vol. 11, pp. 127–138, 2010.

[325] N. Kriegeskorte, W. K. Simmons, P. S. F. Bellgowan, and C. I. Baker, "Circular analysis in systems neuroscience – the dangers of double dipping," *Nature Neuroscience*, vol. 12, p. 535–540, 2009.

[326] C. J. Stam, "Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field," *Clinical Neurophysiology*, vol. 116, p. 2266–2301, 2005.

[327] N. M. Timme and C. Lapish, "A tutorial for information theory in neuroscience," *eNeuro*, vol. 5, no. 3, 2018.

[328] D. Vidaurre, S. M. Smith, and M. W. Woolrich, "Brain network dynamics are hierarchically organized in time," *Proceedings of the National Academy of Sciences*, vol. 114, no. 48, pp. 12827–12832, 2017.

[329] M. W. J. van Es, C. Higgins, C. Gohil, A. J. Quinn, D. Vidaurre, and M. W. Woolrich, "Large-scale cortical functional networks are organized in structured cycles," *Nature Neuroscience*, vol. 28, pp. 2118–2128, 2025.

[330] M. D. Greaves, L. Novelli, S. M. L., A. Zalesky, and A. Razi, "Structurally informed models of directed brain connectivity," *Nature Reviews Neuroscience*, 2024.

[331] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[332] R. J. LeVeque, *Finite Difference Methods for Ordinary and Partial Differential Equations*. SIAM, 2007.

[333] M. Hutzenthaler, A. Jentzen, and P. E. Kloeden, "Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients," *Proceedings of the Roual Society A*, vol. 487, p. 1563–1576, 2011.

[334] E. Buckwar, A. Samson, M. Tamborrino, and I. Tubikanec, "A splitting method for SDEs with locally Lipschitz drift: Illustration on the FitzHugh-Nagumo model," *Applied Numerical Mathematics*, vol. 179, pp. 191–220, 2022.