INVESTIGATION OF SPATIOTEMPORAL CHAOS USING PERSISTENT HOMOLOGY

A Dissertation
Presented to
The Academic Faculty

By

Brett Tregoning

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Physics
Center for Nonlinear Science

Georgia Institute of Technology

May 2022

INVESTIGATION OF SPATIOTEMPORAL CHAOS USING PERSISTENT HOMOLOGY

Thesis committee:

Dr. Michael F. Schatz
School of Physics
Georgia Institute of Technology

Dr. Mark R. Paul
Department of Mechanical Engineering
Virginia Tech

Dr. Predrag Cvitanović
School of Physics
Georgia Institute of Technology

Dr. Konstantin Mischaikow
Department of Mathematics
Rutgers University

Dr. Kurt Wiesenfeld
School of Physics
Georgia Institute of Technology

Date approved: December 17, 2021

For my fiancée, Arianna

ACKNOWLEDGMENTS

This Ph.D. would not have been possible without the support and patience of numerous people. First and foremost, my advisor Mike Schatz has been a pleasure to work with and has taught me so much about being a scientist, teacher, and professional. Mike has provided me with many opportunities to collaborate and interact with the scientific community, which has been essential for my academic growth. He is dedicated to his students and always makes us a priority. Roman Grigoriev is another mentor who has been extremely helpful in the productivity in my research. Just when I think I have hit a wall, Roman routinely has another idea for how to proceed. Roman has also been an excellent teacher in numerous classes that I attended at Georgia Tech.

My fellow members of the Schatz lab have also done a lot to support me over these years. Logan Kageorge is a dear friend in and outside of the lab. He has always been willing to help me with some problem with an experiment or MATLAB and send me the passwords to various lab accounts when I forget them. Chris Crowley's practical knowledge of experimental measurements has been invaluable and he tackled any problem I brought to him as if it were his own. Bala Suri's time in the lab in mine overlapped for about a year, but he has and continues to teach me a lot about being a successful scientist. Wesley Toler has always been a helpful and friendly face in the lab. In the wider Schatz / Grigoriev group, Patrick Reinbold was a good friend had lots of funny stories once you got him talking. Joshua Pughe Sanford has been a wonderful collaborator. He brings enthusiasm and know-how to every project and has helped me tackle some sticky symmetry problems throughout this research. Dmitriy Zhigunov and Matthew Golden have been great collaborators as well.

I have been fortunate to collaborate with many hardworking undergraduate students. Alexander George-Kennedy and Gautham Gorti have brought energy to our collaborations and have provided me with great insights into various subjects. Namo Wichitrnithed is an

excellent, hardworking student who produced a great deal of quality work in a short time.

Collaborators from outside of Georgia Tech have been invaluable to my progress, as well. Mark Paul has taught me a lot about the physics of Rayleigh-Bénard convection. Konstantin Mischaikow has been an enthusiastic collaborator always ready to provide an easy-to-understand mathematical insight about persistent homology. I would like to thank both Mark and Konstantin for serving on my thesis committee. Miro Kramár has been an excellent collaborator when it comes to applying persistent homology. His mind for dynamics and mathematics has led our research in many interesting directions. Rachel Levanger's support was essential for getting my persistent homology research going and she continues to be a good mentor. Also, Saikat Mukherjee has been a consistently excellent research partner, always happy to discuss a problem and meet up at the annual DFD meetings.

I have known several faculty at Georgia Tech who have taught great courses and provided excellent physical insights, including Predrag Cvitanović, Kurt Wiesenfeld, and Minami Yoda. I would like to thank Predrag and Kurt for serving on my thesis committee.

In completing this program, my friends have helped me immensely. First and foremost are fellow physicists my roommates past and present: Hector Velasco-Perez, Noah DeTal, and Derek Hart. They always been there to listen to me talk through a problem and we've had a lot of good times together.

My family has supported me unconditionally throughout this process, and I will be forever grateful. My parents, Christina and Doug Tregoning, were always willing to love and support me throughout this process and I love them dearly. They have provided so much emotional, logistical (helping me move and getting things for the apartment!), and financial support and I am extremely grateful for that. They have always believed that I can achieve a great deal in my life and this helps me stay confident in the face of whatever challenges I face. My brother Kyle and my sister Kira have always been there for me, so I thank them as well. My wonderful grandmother Nannie has always wanted to know how

I am doing throughout this process and kept me in her thoughts, so I am grateful to her. Finally, my fiancée Arianna has been with me almost this whole process, putting up with the fact that I have plenty of stress and not so much time. She has been patient, thoughtful, loving, and most importantly she's pushed me when I needed it. I am looking forward to starting our life together when all this is over.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

CHomP Computational Homology Project

ECS exact coherent structures

GUDHI Geometry Understanding in Higher Dimensions

RBC Rayleigh-Bénard convection

RPO relative periodic orbit

SDC spiral defect chaos

TDA Topological Data Analysis

TW traveling wave

SUMMARY

Spatiotemporal chaos, especially fluid turbulence, is ubiquitous in nature but can be difficult to characterize because analytic solutions of the strongly nonlinear partial differential equations that govern the behavior are often intractable. Additionally, turbulence's sensitive dependence on initial conditions renders prediction difficult because two nearby initial conditions diverge exponentially in time. Statistical studies of turbulence that avoid use of the governing equations and exact knowledge of flow states have yielded insights, but are limited. However, recent advances in numerical techniques have renewed work toward a deterministic, dynamical understanding of turbulence based on quantities computed using the governing equations. In a dynamical understanding of turbulence, turbulent flows traverse a high-dimensional state space as they evolve in time. However, the dynamics tend toward a relatively low-dimensional inertial manifold embedded in this state space. The behavior of turbulent flows as they traverse this inertial manifold can yield insights into the aspects of the governing equations that structure the dynamics.

It is not always straightforward to detect underlying dynamical behavior of a turbulent flow. The methods of analysis currently available can be difficult to work with, computationally inefficient, or fully unable to quantify relevant phenomena. However, many such flows exhibit topological structures that provide insights into their underlying dynamics. An increasingly relevant tool for quantifying the topological structures of flows is persistent homology, a technique of topological data analysis.

In this thesis, persistent homology is used to investigate the dynamics of two different spatiotemporally chaotic fluid flows. In the Kuramoto-Sivashinsky equation, a popular "toy model" system that mimics the spatiotemporal chaos exhibited by fully turbulent fluid flows with continuous symmetries, persistent homology is used to detect and quantify shadowing of exact coherent structures (ECS). ECS are invariant solutions of the governing equation which exhibit simple time dependence. These solutions are embedded in or near the in-

ertial manifold and structure its geometry, influencing turbulent flows. Turbulence often "shadows" ECS, i.e. they closely approach the solution and, where applicable, co-evolve alongside it. In this work, I show persistent homology can be used to detect shadowing of ECS with greater operational simplicity and computational efficiency than other methods used in the Kuramoto-Sivashinsky equation. The results in this simplified setting indicate the method's utility in fully-developed turbulent flows with continuous symmetries, such as pipe flow or Taylor-Couette flow.

In Rayleigh-Bénard convection, persistent homology is used to detect and quantify plumes, which are observable pattern features in experiments and simulations. In simulations, plumes indicate spatial regions of the convective flow in which the leading Lyapunov vector magnitude, a fundamental quantity that characterizes the dynamics of the flow, is high. The leading Lyapunov vector is the eigenvector partner of the flow's largest unstable Lyapunov exponent, so it indicates spatial regions of the flow in which perturbations grow most rapidly. A long-term goal is to use plumes to connect dynamics in simulations, where the leading Lyapunov vector can be computed, to experiments, where this quantity cannot be observed. This thesis advances research in both of these topics and demonstrates that persistent homology is a powerful tool for analyzing topological structure associated with the dynamics of spatiotemporally chaotic flow.

CHAPTER 1
INTRODUCTION

1.1 Paradigm problems in spatiotemporal complexity

Systems that feature behavior that is complex in space and time are omnipresent across a range of scientific disciplines [1]. In biology, spatiotemporally complex patterns are formed on organisms from biochemical processes [2]. In oscillatory chemical reactions, reagents can form complex spatial patterns that evolve in time as reaction fronts propagate [3]. In precise nonlinear optics experiments, "space-time chaos" has been observed [4]. Fluids that are driven far from equilibrium by numerous mechanisms including surface tension gradients [5], applied electric or magnetic fields [6], thermal gradients [7] and mechanical shears [8] exhibit a wide range of spatiotemporally complex behavior. Spatiotemporal complexity of fluid systems driven far from equilibrium includes turbulence, a ubiquitous, widely studied phenomenon that has resisted understanding for centuries [9].

It is typically not feasible to make well-resolved measurements of natural fluid systems driven far from equilibrium, so many scientists focus on relatively simple laboratory experiments and their theoretical descriptions that share key features with natural systems of interest. Such experiments can typically be highly resolved. Some examples include Taylor-Couette flow [10], Rayleigh-Bénard convection [11], and quasi-2D Kolmogorov-like flow [12]. These experiments can be described completely by theoretical descriptions that make few simplifying approximations, such as the Navier-Stokes equation and incompressibility condition in Taylor-Couette flow [13], the Boussinesq equations in Rayleigh-Bénard convection [14], and a depth-averaged description for quasi-2D Kolmogorov-like flow [15]. Additionally, highly reduced "toy" models are studied because they reproduce key behaviors of spatiotemporally complex fluid systems without serving as complete de-

scription of the flow, like the Swift-Hohenberg equation [16] the Lorenz equations [17], and the 1-D Kuramoto-Sivashinsky equation [18]. These model systems are low-dimensional, making their analysis more tractable than even theoretical descriptions of paradigm lab experiments. Frequently, such model systems are used as testing grounds for novel methods of analysis before they are undertaken in more complete theoretical descriptions. These paradigm systems, from simple models to experimentally-realizable fluid flows, provide a rich landscape of scientific problems of varying complexity, and their study continues to provide insights into the spatiotemporal complexity of natural systems.

1.1.1 Statistics of turbulence

Fluid turbulence is a subset of spatiotemporally chaotic behavior seen in a variety of paradigm fluid flows, the formal study of which dates back hundreds of years to at least Leonardo Da Vinci [9]. Not all spatiotemporally chaotic fluid flows would be traditionally classified as turbulence. One such example is spiral defect chaos in Rayleigh-Bénard convection. Though this flow is spatiotemporally chaotic, it does not display self-similarity over a range of spatial scales, a hallmark of turbulence.

Traditionally, turbulence has been characterized using statistical approaches that are not dependent on the governing equations, and thus not dependent on any one paradigm problem. The Navier-Stokes equation, the physical law that governs a fluid flow's velocity field, is a strongly nonlinear partial differential equation that is difficult to work with [19]. Analytic solutions of the Navier-Stokes equation are generally intractable. Thus, most early study of turbulence gained insights through statistical methods. Kolmogorov used a statistical approach to show that energy in a turbulent flow scales with the wavenumber to the power of $-\frac{5}{3}$ in the famous scaling law [20]. This result is valid over the "inertial range", i.e. the spatial scales in which inertial effects dominate viscous effects. Another statistically-derived result is Von Karman's "law of the wall" [21] which states that the average flow velocity in turbulence is proportional to the logarithm of the distance from a

boundary.

Though statistical approaches have provided historical insights into turbulent flow, they are limited. Statistical treatments assume self-similarity and spatial uniformity over a range of spatial scales, which is often false in practice even in fully turbulent flows. The observation of coherent structures, large-scale spatial features in turbulence that recur over time, show that there is more to turbulent flows than an isotropic energy cascade [22]. Theories of turbulence that incorporate the governing equation are necessary for a complete understanding.

1.2 Dynamical view of spatiotemporal chaos

In more recent years since the dominance of statistical theories in the study of turbulence, computing power has advanced. It is now possible to numerically simulate strongly nonlinear equations such as Navier-Stokes and other models, such as the Kuramoto-Sivashinsky equation. In numerical simulations, the precise state of the flow is known, allowing investigation of the underlying dynamical structures of the governing equations [23, 24, 25]. Two frameworks for understanding spatiotemporal chaos in paradigm problems from a dynamical point of view are exact coherent structures and the Lyapunov spectrum. In the following subsections, these concepts will be briefly introduced and contextualized in this thesis.

1.2.1 Exact coherent structures in the Kuramoto-Sivashinsky equation

A theory made possible by the complete knowledge of the spatiotemporally chaotic flow state afforded by numerical simulations is that of exact coherent structures (ECS) [25], which are low-dimensional stationary or time-periodic solutions to the governing equations that are embedded in a high-dimensional chaotic attractor. They are found by numerical continuation of simpler states into parameter regimes of increased driving [24] or by solution searches from a nearby initial condition [26]. Chaotic trajectories evolving in time throughout the attractor are thought to pass close to or even "shadow" ECS along their

stable and unstable manifolds [27, 28, 13]. The ECS and their manifolds describe a deterministic picture of chaotic fluid flows that evolve according to their position in state space, differentiating this approach from a statistical one. The challenge for ECS theory then, is finding dynamically relevant ECS and developing metrics that indicate when close passes and shadowing are occurring. Paradigm fluid systems where ECS have been found include plane Couette [24, 29], Taylor-Couette [13], and plane Poiseuille flow [30]. Additionally, ECS are often computed in model systems, such as the 1-D Kuramoto-Sivashinsky equation (KS) [26]. This model provides an interesting case study for ECS in a system with continuous symmetry called relative periodic orbits, which are present in laboratory fluid systems such as Taylor-Couette flow [13] and pipe flow [31]. In this thesis, a novel method for quantification of shadowing of ECS in the KS equation using persistent homology is described. This technique is invariant to the continuous symmetry of KS, so it is anticipated that insights gained in this setting will be applicable to laboratory fluid flows with continuous symmetries.

1.2.2 Lyapunov spectrum in spiral defect chaos

Fully resolving a fluid velocity field to detect close passes and shadowing to ECS is not always feasible in laboratory experiments. One such example is Rayleigh-Bénard convection (RBC), the buoyancy-driven flow between a hot bottom plate and cool top plate [32, 33]. The high aspect ratio, spatiotemporally chaotic regime called spiral defect chaos (SDC) is most commonly imaged from above using a shadowgraph technique that provides a proxy for the temperature variation in the flow [11]. This technique is noninvasive, unlike techniques used in other systems that involve seeding a flow with particles that show up easily on camera, but it cannot resolve the velocity of the flow field. Without a fully resolved flow field, it is difficult to detect close passes to and shadowing of ECS in experiments, necessitating an approach that emphasizes features that can be observed in the convection roll patterns captured by shadowgraph images from SDC experiments [34, 35].

In this thesis, observable features of spiral defect chaos are quantified using persistent homology and comparisons to statistics of those features in numerical simulations are made to gain insights into the dynamics of Rayleigh-Bénard convection. The observable features in question are convective plumes, which have been shown by Levanger *et al.* [36] to indicate spatial regions of simulated flows where the leading Lyapunov vector magnitude is high. The leading Lyapunov vector is a dynamical quantity, computed by simulating the governing equations, that indicates regions of the flow in which perturbations grow quickly. Thus, connection of experimental observable flow features to the Lyapunov vector could prove invaluable for characterizing dynamics in SDC.

1.3 Persistent homology analysis of spatiotemporal chaos

For the study of ECS in KS and the study of the Lyapunov spectrum in SDC, the analytical tool that is employed in this work is persistent homology, a technique of topological data analysis (TDA). TDA has recently become more accessible for scientists in a variety of disciplines due to the development of software packages that compute the homology of data [37, 38]. Persistent homology detects and quantifies the unique topological structures of a data set. Furthermore, it provides an organization of those structures according to their prominence [39, 40] by tracking the persistence of unique structures across a family of filtrations. Structures that persist throughout a large range of filtrations are the most prominent features of the data set, while structures that persist through only a narrow range of filtrations are commonly associated with noise. Because of its flexibility and ability to distinguish noisy structures from prominent ones, persistent homology has been employed to characterize a wide variety of data sets in neuroscience [41], molecular biology [42], network science [43], and granular media [44]. Persistent homology has multiple applications in fluid systems, specifically [45, 35, 46]. It can be used to characterize the state space representation of a fluid flow [28] and to characterize image data of the flow itself [47, 36] in a variety of systems from models such as the Kuramoto-Sivashinsky equation to flows

realizable in a laboratory such as Rayleigh-Bénard convection and quasi-2D Kolmogorov flow.

Persistent homology allows for the representation of data in a variety of formats, such as persistence landscapes [48], persistence images [49], and, particularly, persistence diagrams [47]. Persistence diagrams organize the topological structures of a data set, such as a time snapshot of an image of fluid flow, into a representation of reduced dimensions that still captures features of interest. Moreover, persistence diagrams can be quantitatively compared using a variety of metrics, such as the family of Wasserstein distance metrics [50]. It is shown in this work that representing fluid flow time snapshots as persistence diagrams can provide new insights into those flows.

In this work, two different spatiotemporally complex fluid systems are characterized with persistent homology. In the case of the 1-D Kuramoto-Sivashinsky equation, the persistence diagrams of image data of both chaotic trajectories and ECS (particularly, relative periodic orbits) are computed and used as a symmetry-reduced alternative to a state space representation. In this representation, the Wasserstein distance metric is used to quantify close passes to and shadowing of the ECS by the chaotic trajectory. In the case of Rayleigh-Bénard convection in the spiral defect chaos regime, persistent homology is used to detect and quantify convective plumes, a dynamically relevant observable feature in both experiments and simulations.

1.4 Thesis overview

This thesis is organized as follows. In chapter 2, the necessary background for exact coherent structures, computation of the Lyapunov spectrum, spiral defect chaos in Rayleigh-Bénard convection, and persistent homology is reviewed. In chapter 3, the study of ECS in the 1-D Kuramoto-Sivashinsky system is discussed, and the persistent homology approach to detecting shadowing is compared and contrasted to the traditional approach that uses a state space representation. In chapter 4, the quantification of plumes in Rayleigh-Bénard

convection experiments and simulations using persistent homology is discussed. In chapter 5, conclusions and next steps are offered.

CHAPTER 2

BACKGROUND

2.1 Exact coherent structures: a deterministic view of spatiotemporal chaos

2.1.1 Overview of ECS

Due to the limitations of statistical theories, a deterministic, dynamical theory of fluid turbulence has long been sought. Despite the unpredictable and chaotic nature of turbulence, scientists have long noticed that there are recurrent patterns that appear in turbulent flows, commonly referred to as coherent structures [22]. We now know that many such structures correspond to unstable solutions of the Navier-Stokes equation called exact coherent structures (ECS) that have simple time dependence. We seek to understand the effect ECS have on spatiotemporally chaotic dynamics.

To understand ECS, let's start by considering a state space picture of dynamical systems. In principle, partial differential equations that produce spatiotemporally chaotic flow, like the Navier-Stokes equation, are infinite-dimensional. However, empirical studies of such equations often discretize the field into a finite number of point. For example, the state of the 1-D Kuramoto-Sivashinsky (KS) equation, a spatiotemporally chaotic nonlinear PDE, is given by a field $u(x)$ where $x \in [0, L]$. In practice, it is common to discretize the field into $N = 32$ grid points in the case where $L = 22$ (Figure 2.1). The state of this field can then be represented as a $N \times 1$ vector $\vec{s} = (u_1, ..., u_N)$. This vector, and consequently the dynamics of the system, live in an N-dimensional state space. Each state of KS can be represented as a single point in this space. As the equation evolves, that point moves continuously in time through the state space, tracing out a trajectory.

Although the dimension of our state space is N-dimensional, in practice the dynamics that take place in state space can be described by a much lower dimensional *inertial man-*

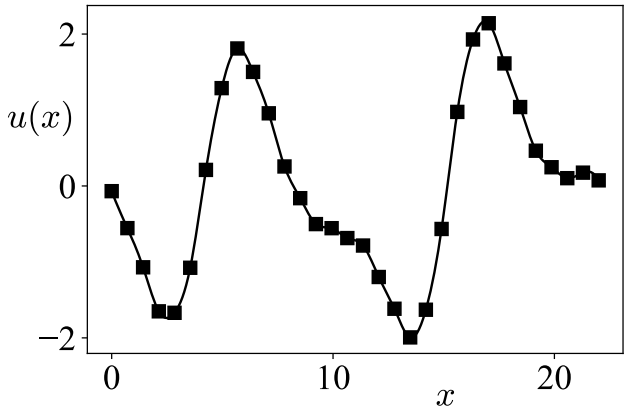

Figure 2.1: A state of the Kuramoto-Sivashinsky equation, discretized into $N = 32$ grid points. The state of $u(x)$ be represented as a $N \times 1$ vector $\vec{s} = (u_1, ..., u_N)$ (black squares).

ifold (or chaotic or turbulent attractor). For KS with $L = 22$, the dimension of that inertial manifold is about $M = 9$ [51, 52], which is much lower than the number of grid points $N = 32$. ECS are solutions of the governing equations that are embedded in or near this inertial manifold. These solutions shape the geometry of the state space through their stable and unstable manifolds. Chaotic trajectories approach ECS from directions on their stable manifolds and leave ECS along directions on their unstable manifolds, thus they provide structure to the dynamics in the inertial manifold. In practice, ECS often have few unstable directions, so turbulence tends to leave the vicinity of the ECS in predictable ways [53].

There are a few types of ECS to keep in mind. Equilibria are static solutions of the governing equations. Unless they are perturbed, they will not evolve in time from their starting point in state space. Relative equilibria, or traveling waves, are static solutions that move with a fixed speed along a direction of continuous symmetry. For example, in pipe flows, relative equilibria are static solutions that flow along the pipe. The continuous symmetries in this case would be streamwise translations and azimuthal rotations, i.e. the dynamical system is equivariant to infinitesimal translations in the streamwise direction and infinitesimal rotations in the azimuthal direction of the flow. Periodic orbits are solutions of the governing equations that, unlike equilibria, evolve in time. This time dependence has a characteristic period T, after which a periodic orbit return to its starting point in state space.

Relative periodic orbits are similar to periodic orbits, except instead of returning precisely to the starting point after a period T, they return to a continuous-symmetry transformed copy of that point.

2.1.2 Previous work in flows with continuous symmetries

Relative equilibria, or traveling waves (TWs), and relative periodic orbits (RPOs) have been the focus of much study in turbulent dynamics that have continuous symmetries. Investigation of TWs and RPOs is more complicated than that of their counterparts with no continuous symmetry because each TW and RPO is a member of symmetry-related family of solutions. TWs, instead of occupying a static, zero dimensional point in state space, move with a fixed speed in state space along the directions of continuous symmetry. Each point along the trajectory is a member of this symmetry-related family of solutions. In a system with N continuous symmetries, TWs form an N-torus in state space. Likewise, RPOs do not comprise a single loop in state space like periodic orbits. They form a $N + 1$-torus in state space in which a family of symmetry-related periodic solutions live. This property of TWs and RPOs must be accounted for in systematic studies of their effect on dynamics and the geometry of state space. Schemes to "reduce" the continuous symmetry of state space, by collapsing all symmetry-transformed points as single, symmetry-invariant point are commonly sought. For specific applications, other ways of minimizing a quantity of interest over the continuous symmetry directions can be done. One such application is the detection and quantification of *shadowing* of ECS, in which a turbulent trajectory in state space closely approaches an ECS and, in cases such as RPOs where the ECS possesses simple time dependence, co-evolves with it. Specifics of techniques used to quotient or reduce continuous symmetries will be discussed shortly.

TWs and RPOs have been detected in fluid flows, mostly in numerical simulations. In plane Couette flow, a flow that is sheared between two planar boundaries, equilibria, TWs, periodic orbits, and RPOs have been computed and studied [54, 55]. In plane Poiseuille

flow, TWs are computed and visualized by Park *et al.* [56]. They detected and quantified shadowing of those TWs, as well. TWs and RPOs have been investigated in pipe flows [57, 58, 31]. TWs have been experimentally observed in pipe flows [27, 59]. An example of a continuous symmetry in the previously-mentioned flows is a continuous translational symmetry in the streamwise direction. Another fluid flow with a continuous symmetry of interest is Taylor-Couette flow, which is invariant to rotations about the cylindrical axis. TWs and RPOs have been computed and shadowing events of those ECS have been detected and quantified [13].

Simple model equations with continuous symmetries serve as a testing ground for study of TWs and RPOs. The main example that will be examined is the Kuramoto-Sivashinsky equation (KS), given by

$$u_t + u_{xx} + u_{xxxx} + uu_x = 0, \ x \in [0, L] \tag{2.1}$$

in which relative equilibria [26, 60] and relative periodic orbits have been studied [61, 52, 28]. The continuous symmetry in KS is translational, i.e. the governing equations are invariant to transformations of the form $u(x,t) \to u(x+s,t)$, where s is a constant shift. In the case where the length parameter $L = 22$, the effect of TWs and RPOs on the geometry of the full state space has been studied, at first without a scheme to reduce the continuous symmetry in state space [60]. TWs and RPOs were visualized by Cvitanović *et al.* [60] using low-dimensional projections in state space and plotting energy dissipation of solutions. They find TWs by substituting the first time derivative with the relative equilibrium condition $u_t = -cu_x$, where c is a constant speed, and solving the ODE

$$cu_x = u_{xx} + u_{xxxx} + uu_x, \ x \in [0, L]. \tag{2.2}$$

In principle, equilibria and periodic orbits can appear in systems with continuous symmetries as well. Cvitanović *et al.* find equilibria in [60] by taking $c = 0$. They find RPOs using

multiple shooting and a Levenberg–Marquardt algorithm. Budanur *et al.* [26] advanced the field by developing a technique of continuous symmetry reduction, applicable to KS, that involves "slicing" trajectories in state space by a Fourier mode. This symmetry-reduced space can be difficult to work with, however, due to discontinuities that occur when trajectories cross the "slice border". Ding *et al.* used RPOs to perform a statistical estimation of the dimension of the KS inertial manifold, known to take the approximation value $M = 9$ in the $L = 22$ case [51]. The fact that an ensemble of RPOs can replicate essential statistical properties of the inertial manifold, such as its dimension, demonstrates their dynamical significance. Budanur *et al.* [28] apply persistent homology analysis to a Fourier-sliced, symmetry-reduced state space of KS to detect signatures of RPOs. Their analysis is quite different from the one we discuss, since they perform persistent homology on point clouds in state space using Vietoris-Rips complexes, while we apply persistent homology directly to KS states using cubical complexes. Importantly, Budanur *et al.* detect shadowing of RPOs in their symmetry-reduced state space using "state space persistent analysis" and visualize the state space trajectories using 3-D projections obtained from principle component analysis. The detection and quantification of shadowing events provide direct evidence that ECS are structuring the turbulent dynamics of the flow. With a catalog of dynamically relevant ECS and quantification of shadowing events to those ECS, it is in theory possible to develop a picture of symbolic dynamics, a coarse-grained depiction of the dynamics that is described by visits between specific ECS. Such a depiction would be the culmination of the deterministic, dynamical theory of spatiotemporal chaos that has long been pursued. To my knowledge, the work by Budanur *et al.* [28] is the deepest investigation of turbulent shadowing of RPOs in the KS equation to date, and is the latest step toward a dynamical understanding of KS.

2.2 Spiral defect chaos in Rayleigh-Bénard convection

A rich, experimentally accessible paradigm to probe underlying dynamics of spatiotemporal chaos is the spiral defect chaos (SDC) regime of Rayleigh-Bénard convection (RBC) [62]. Unlike a fully turbulent flow, SDC is composed of convection rolls, which are well-organized in the vertical direction. A convection roll is characterized by hot fluid rising on one side and cold fluid falling on the other side. The lateral orientation of these rolls and the topological defects they form are spatiotemporally chaotic. There is a relatively small range of spatial scales to consider in SDC compared to fully developed turbulence, making it a tractable setting for study of spatiotemporal chaos in fluid flows. SDC is the setting for the experimental and numerical study of spatiotemporal chaos undertaken in chapter 4. Key aspects of SDC and RBC are covered in detail in chapter 4, but here some essential background and previous work in characterizing spatiotemporal chaos in SDC will be discussed.

2.2.1 Overview of spiral defect chaos

RBC is a widely studied problem in nonlinear pattern formation. It consists of two horizontal plates with a fluid in between (Figure 2.2). The bottom plate is heated relative to the top plate and for a large enough temperature difference, buoyancy differences in the fluid drives a convective flow. The strength of this driving is quantified by the Rayleigh number $\mathrm{Ra} = \frac{\beta g \Delta T d^3}{\nu \kappa}$ where β is the thermal expansion coefficient, g is acceleration due to gravity, ΔT is the temperature difference, d is the separation between the bottom and top plate, or depth, ν is the kinematic viscosity, and κ is the thermal diffusivity. Other important nondimensional parameters include the Prandtl number $\mathrm{Pr} = \frac{\nu}{\kappa}$ and the aspect ratio $\Gamma = \frac{r_0}{d}$ where r_0 quantifies the lateral size of the cell. In a cylindrical geometry, which is used in this work, r_0 is the radius of the cell. In the laterally infinite case, when Ra is increased past the critical value $\mathrm{Ra} \approx 1708$, a striped pattern of parallel convection rolls form [1, 63]. Experimental conditions can modify this critical behavior, however. In a finite experimental

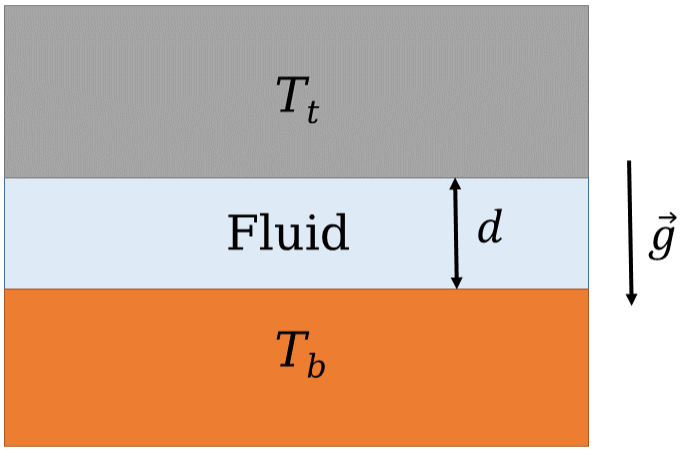

Figure 2.2: Sketch of the Rayleigh-Bénard convection setup. The temperature difference $\Delta T = T_b - T_t$ opposes the direction of gravity $\vec{g}$. The separation of the plates, or thickness of the fluid layer, is denoted d

cell, when the sidewalls are heat conducting, the first convection rolls tend to form parallel to them. The striped pattern then propagates inward. This results in an onset pattern that reflects the geometry of the cell. In the cylindrical case, a "target" pattern is formed (Figure 2.3). As the Rayleigh number is increased further from the critical value, the center of target can become unstable and begin to move from the center, eventually forming a topological defect with another convection roll. That defect is itself unstable, and the formation and motion of new defects gives rise to spiral defect chaos, a spatiotemporally chaotic state made up of convection rolls that form topological defects with one another (Figure 2.4).

SDC is typically studied under the assumption of the Boussinesq equations

$$\mathrm{Pr}^{-1}\left(\frac{\partial \boldsymbol{u}}{\partial t}+\boldsymbol{u}\cdot\nabla\boldsymbol{u}\right) = -\nabla p+\nabla^2\boldsymbol{u}+\mathrm{Ra}T\hat{z} \tag{2.3}$$

$$\frac{\partial T}{\partial t}+\boldsymbol{u}\cdot\nabla T = \nabla^2 T \tag{2.4}$$

$$\nabla\cdot\boldsymbol{u}=0 \tag{2.5}$$

which represent conservation of momentum, energy, and mass for an incompressible fluid, respectively. In these non-dimensional equations, $\boldsymbol{u}=\boldsymbol{u}(x,y,z,t)$ is the velocity field, $p=p(x,y,z,t)$ is the pressure field, and $T=T(x,y,z,t)$ is the temperature field where t is time, x and y are the lateral spatial coordinates, and z is the spatial coordinate opposing the direction of gravity. The crucial assumption of the Boussinesq equations is that the only fluid parameter that is dependent on the local temperature is the density. This dependence is linear and appears in the last term of Equation 2.3. The extent that a fluid deviates from the Boussinesq approximation can be quantified by the Busse parameter Q (see chapter 4 for more information about Q). Specifics on how the simulations and experiments of SDC are conducted are given in chapter 4.

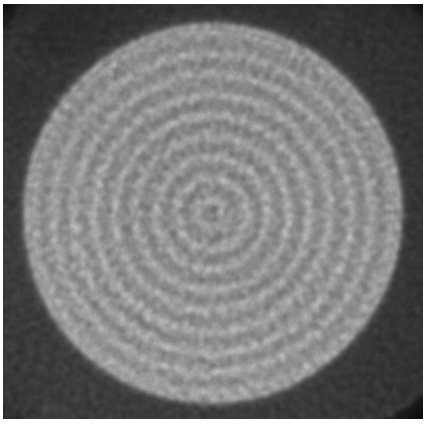

Figure 2.3: Raw shadowgraph image of onset "target" pattern in a cylindrical RBC experiment. This image was taken from above the top plate, looking down into the fluid layer. Bright fluid represents cool downflow while dark fluid represents hot upflow.

Figure 2.4: Shadowgraph image of convection roll patterns in Rayleigh-Bénard convection undergoing spiral defect chaos. A large spiral can be seen on the left side of the image.

2.2.2 Previous work in characterizing spiral defect chaos

Clever and Busse [64, 65] provide a detailed nonlinear stability analysis for RBC that is driven just past equilibrium. Investigation of the "Busse balloon", a parameter regime in which stable, straight convection rolls are predicted, led to discovery of SDC [62]. SDC was later observed in the Swift-Hohenberg equation [66], a model of Rayleigh-Bénard convection, and the Boussinesq equations [67]. A thorough review of the work in SDC in the 1990's was done by Bodenschatz *et al.* [68]. The study of SDC using experimental cells imaged by shadowgraph has become standard [11] because it is minimally invasive, easily controlled, allows for study of large aspect ratios, and provides a visualization of convection roll patterns. Our experimental apparatus is very similar to the type de Bruyn *et al.* describe [11].

The characterization of SDC by its pattern features has been an ongoing study in the decades since. Cakmur *et al.* [69] found that SDC and the parallel straight roll pattern are both stable in the same parameter regime in which SDC was originally discovered, but

that different initial conditions give rise to one state or the other. Egolf *et al.* [34] found that stability boundaries for the idealized, infinite case can be applied to spatially localized patches of SDC and observed the relevance of multiple length scales in SDC. Riecke and Madruga [70] studied the characteristics of spirals by measuring their winding number and arc length. Becker and Ahlers [71] characterized the patterns of SDC using local wave director analysis. They found that higher values of Pr produce larger spirals with higher winding number. Krishan *et al.* [45] used computational homology make comparisons between experiments and numerical simulations of SDC and found that they exhibit differences in the number of hot and cold components of the convection roll patterns. Kurtuldu *et al.* [46] deepened the analysis of experimental SDC to demonstrate that non-Boussinesq effects in experiments that are not present in simulations of the Boussinesq equations cause this difference. Kurtuldu *et al.* quantified the degrees of freedom in SDC using both Karhunen-Loève decomposition dimension and the computational homology dimension and showed that both metrics scale with system size. Kramár *et al.* employed persistent homology to reduce the dimensionality of SDC dynamics. This reduced description allowed for easy identification of important flow features. Levanger *et al.* studied myriad topological pattern features in SDC to see if they correspond to high leading Lyapunov vector magnitude, a fundamental dynamical quantity of interest. We continue their work by focusing on a particular SDC pattern feature that was identified to be most correlated with high leading Lyapunov vector magnitude, convective plumes. In chapter 4 plumes will be discussed in detail, but shortly the Lyapunov spectrum and its importance to the underlying dynamical structure of SDC will be reviewed in section 2.3.

2.3 Lyapunov spectrum

The Lyapunov exponents and vectors characterize spatiotemporally chaotic dynamics by quantifying their sensitive dependence on initial conditions [72, 73]. For a discretized spatiotemporally chaotic system, like the ones studied in this thesis, the Lyapunov exponents

are the eigenvalues of the Jacobian matrix and the Lyapunov vectors are their eigenvector partners. The Lyapunov exponents contain information about the number of degrees of freedom of chaotic dynamics, and are used to compute the Kaplan-Yorke dimension [74, 75]. This approach for estimating the inertial manifold dimension is useful in large, extensive spatiotemporally chaotic systems because it is not dependent on the geometry of the inertial manifold, which increases in complexity as the system increases in size [76]. For an estimation of the inertial manifold dimension, it is common to consider Lyapunov exponents averaged over a long time series, but one can also consider the *finite time*, or instantaneous Lyapunov exponents that are relevant at a specific time in a chaotic state's evolution.

The standard method for computation of finite time Lyapunov exponents from time-series of data [23] makes it possible to find the finite time leading Lyapunov vector, the eigenvector partner of the largest unstable Lyapunov exponent at that time. The leading Lyapunov vectors contain spatial information associated with the growth rates quantified by the corresponding largest Lyapunov exponent. The directions of the other Lyapunov vectors are unfortunately lost due to Gram-Schmidt reorthonormalizations used for efficient computation of the exponents. The leading Lyapunov vector indicates the local directions in state space that are most unstable. Put another way, it indicates the most unstable spatial regions of the flow at a given time. In these sensitive regions, perturbations grow more quickly than in other spatial regions of the flow. Thus, regions of high leading Lyapunov vector magnitude are of specific interest for characterizing spatiotemporally chaotic dynamics. The leading Lyapunov vector has been found to be correlated with significant topological events in numerical simulations of Rayleigh-Bénard convection undergoing spiral defect chaos [77, 78, 76, 36]. Egolf *et al.* [77] found, in a spatially periodic simulation, that spatial regions of high Lyapunov vector magnitude were correlated with significant dynamical events in spiral defect chaos: roll pinch-offs and reconnections. Scheel *et al.* [78] extended the analysis to simulations with no-slip boundary conditions and found

that only events that are not repeated make a significant contribution to the long time Lyapunov spectrum. Karimi *et al.* [76] also found correlations between dynamical events and the leading Lyapunov vector magnitude, this time in an experimentally realistic simulation over a range of Rayleigh and Prandtl numbers.

These topological features are observable in experimental shadowgraph images of convection roll patterns, but it remains an open question whether or not features observed in experiments can be connected to the Lyapunov spectrum. The full significance of the finite time and time-averaged Lyapunov spectrum on spiral defect chaos is an area of active investigation. For a connection to be made, it must first be ensured that numerical simulations in which the Lyapunov spectrum is computed are experimentally realistic. That is, simulations must produce similar observable features with similar statistics as those in experiments. If this is not the case, then simulations must be modified to more successfully capture experimental conditions. Steps are taken toward the goal of connecting the Lyapunov spectrum computed in simulations to observable features in experiments in chapter 4 by developing a method to quantify and track observable features in time in both experimental and simulated data. The features in question are convective plumes, which Levanger *et al.* [36] found to be most indicative of spatially-localized high Lyapunov vector magnitude out of many features that can be detected using topological data analysis. The long-term goal is to able to use features observable in both experiments and simulations as a proxy for the leading Lyapunov vector, which can be computed in simulations only. The dynamical insights provided by the Lyapunov spectrum could then be applied to experiments.

2.4 Computational Homology and Persistent Homology

Topological data analysis is becoming an increasingly useful toolkit for understanding complex data. Essential information can often be obtained from just the topology of a data set. Persistent homology is among the most useful tools in the toolkit of topological data analysis that can provide insight into myriad spatiotemporally complex systems. One of the

most common application of persistent homology makes use of Vietoris-Rips complexes to examine the topological structure of point clouds. Situations where persistent homology can be computed using cubical complexes, such as 1-D functions and 2-D image data, are considered here instead. First, the method used to compute the homology and, by extension, the persistent homology of 1-D functions will be discussed. Then, the discussion will be extended further to persistent homology of 2-D image data and then useful distance metrics in the space of persistence diagrams will be discussed.

2.4.1 Computational homology of 1-D functions

Computational homology allows for the efficient computation of topological structures in a binary data set. While the homology of many types of data sets, one of the most popular being point clouds, can be computed, 1-D functions will first be discussed. An excellent resource on the mathematical theory underlying the computation of homology of *cubical complexes* obtained from image data is given by [37].

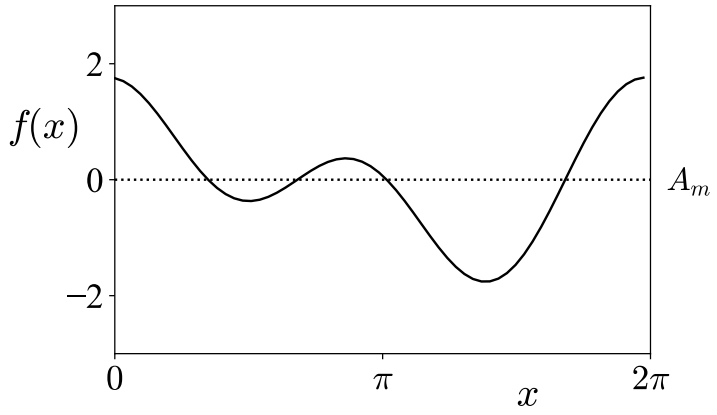

Figure 2.5: A periodic function $f(x)$ (solid line) with a filtering threshold $A_m = 0$ shown (dotted line).

As a simple example, define the function $f(x) = \sin(x+1) + \sin(2x+2), x \in [0, 2\pi]$ Figure 2.5. Note that this function is periodic on the domain of x, i.e. $f(0) = f(2\pi)$. To compute the homology of this function, the first thing we must do is *filter* it into two sets. One set will consist of the surface on which we compute homology, and the other set will

be empty space. The outcome of the homology that we compute depends fundamentally on the choice of this filtering, so the physical properties of a data set that one wishes to extract with homology must be carefully considered. For this example, let's consider a *sublevel filter* with a threshold A equal to some value of the function over the interval $[0, 2\pi]$. In sublevel filtering, we take the surface on which we will compute homology to be

$$S = \{x | f(x) \leq A\}. \tag{2.6}$$

In principle, A could be any real number, but let's consider specifically the case where $A = A_m = 0$, the mean value of f over the interval $[0, 2\pi]$ Figure 2.5. Then, the surface S would consist of two intervals in x, $x \in [1.10, 2.14]$ and $x \in [3.19, 5.28]$, approximately (Figure 2.6). These are the intervals of x such that $f(x) \leq A_m = 0$.

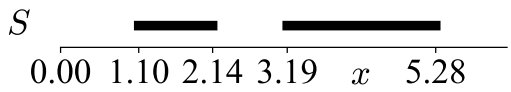

Figure 2.6: The surface $S = \{x | f(x) \leq A_m\}$ is denoted as the solid black lines while values of x that are not a member of S are empty.

The surface S can then be represented as a set of *elementary cubes*. A set of elementary cubes of the dimension of the data set being considered (in this case, 1-D) is commonly referred to as a *cubical complex*. A detailed definition of elementary cubes and cubical complexes (referred to as cubical sets) is offered in [37]. Now, the *homology groups* of S can be computed using software such as GUDHI [38] or CHomP [79]. The details of how this is done won't be discussed in this setting, but the useful information that can be obtained from this computation will be explored. Homology groups are objects in algebraic topology that encode the topological structure of the surface S. The dimension of the nth homology group, referred to as the nth Betti number, is the number of distinct topological structures of dimension n that make up that surface. The zeroth through $D-1$th homology groups exist for a given surface, where D is the dimension of that surface. The surface

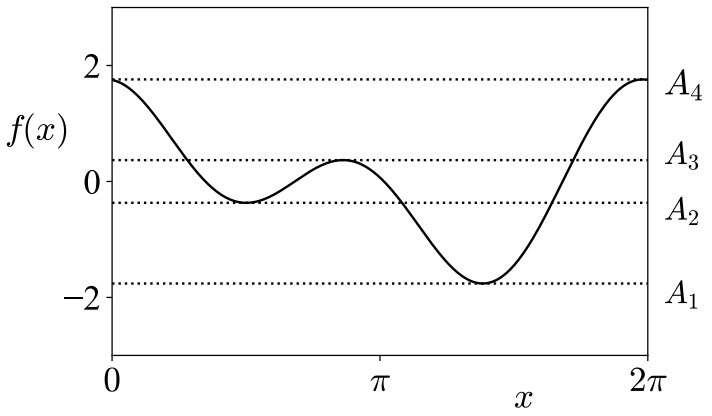

Figure 2.7: A periodic function $f(x)$ (solid line) with a filtering thresholds A_1, A_2, A_3, and A_4 shown (dotted lines).

we are considering here, S, is 1-D, therefore it has only a zeroth homology group with a dimension of $\beta_0 = 2$, since there are two *distinct connected components* that make up the surface.

If we chose a different filtering threshold value A, the Betti numbers given by the homology computation could be very different. For example, we could choose $A = 10$, then every value of $f(x)$ would be counted as a member of the surface S, which would consist of one distinct connected component. In that case, we would have $\beta_0 = 1$. We could also choose $A = -10$, in which case our surface S would be completely empty with $\beta_0 = 0$.

We can also see what would happen if we chose a filtering scheme that is different than sublevel filtering. In *superlevel filtering* we take the surface to be

$$S = \{x | f(x) \geq A\}. \tag{2.7}$$

Both filtering schemes can yield different, but equally useful information about the topological structure of the data set. In this work, we make use of both sublevel and superlevel filtering in computational homology and its extension, persistent homology.

2.4.2 Persistent Homology of 1-D functions

In persistent homology, an extension of computational homology, the threshold A is varied over all values of $f(x)$. Imagine, in the case of sublevel filtering, the threshold A starts at $-\infty$ and is swept up through all values to ∞. At each value of A, elementary cubes will join the surface S. Those cubes could form distinct connected components, be attached to previously existing distinct connected components, or even act as the link that causes two distinct connected components to merge. We compute the homology using sub- and superlevel filtering and keep track of the distinct connected components at every threshold. This technique allows us to not just identify topological structure of $f(x)$ at a specific A, it also quantifies the prominence of topological structures. Returning to the example in Figure 2.5, we can see that there are some values A we could choose for sublevel filtering that are special (Figure 2.7). $A = A_1$, shown in Figure 2.7, is the lowest threshold for which a distinct connected component will appear. We thus say that distinct connected component is *born* at $A = A_1$ as the threshold is increased from $-\infty$. This distinct connected component grows as A is increased past A_1, but this growth does not affect the homology of the surface. At $A = A_2$, we can see that a second distinct connected component is born. At $A = A_3$, the two distinct connected components merge into one distinct connected component. Topologically, this is a significant event. One of the distinct connected components has *died* by merging with the other one. Specifically, we say that the component that was born at a higher value of A is the one that died, while the one that was born "first" remains. One can think of this is a "first-in-last-out" scheme for assigning deaths of topological structures. At $A = A_4$, one distinct connected component remains. For all $A \geq A_4$ the homology will remain unchanged, so this distinct connected component will never die. In Figure 2.8, the surfaces S_i where $i = 1, 2, 3, 4$ corresponding to filtering thresholds A_i where $i = 1, 2, 3, 4$ are shown. S_1 consists of one distinct connected component, S_2 consists of two distinct connected components, and that in S_3 those two distinct connected components have merged. The surface S_4 consists of the entire domain of $f(x)$.

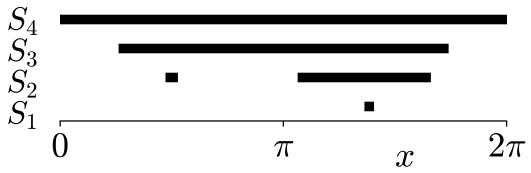

Figure 2.8: The surfaces $S_i = \{x | f(x) \leq A_i\}$ where $i = 1, 2, 3, 4$ are denoted as the solid black lines in ascending order while values of x that are not a member of S are empty.

The information given by persistent homology can be intuitively visualized in many ways, such as in persistence barcodes [40] and persistent images [49]. In this work, the information from persistent homology is organized into *persistence diagrams* (Figure 2.9), 2-D scatter plots where each point represents a topological structure. A point's position on the x-axis gives the value of its birth A_b and its position on the y-axis gives the value of its death A_d. Since one distinct connected component never dies, its death value is plotted at $A_d = \infty$. For analysis of persistence diagrams, it is common to throw out this "infinite" persistence point. In Figure 2.9, the sub- and superlevel persistence diagrams are plotted. Every persistence point in the sublevel persistence diagrams must lie on or above the dashed diagonal, while every persistence point in the superlevel persistence diagrams must lie on or below the dashed diagonal. Thus, throughout the rest of the work, it will be common for me to plot both sub- and superlevel persistence on the same diagram.

In Appendices subsection D.2.1 and subsection D.2.2, a sample code for the persistent homology calculation of $f(x)$ using GUDHI in Python is provided. By default, GUDHI performs sublevel filtering to compute persistent homology. To perform superlevel filtering, the input is transformed from $f(x)$ to $f_r(x)$ such that $f_r(x) = \max(f(x)) + \min(f(x)) - f(x)$. In this simple case, $\min(f(x)) = -\max(f(x))$, so $f_r(x) = -f(x)$. The transformation must be applied again to the persistence points that GUDHI outputs to get the superlevel filtering. The superlevel persistence points of $f(x)$ are simply the sublevel persistence points reflected about the diagonal, yielding identical topological information, but that is not generally true. In both the KS and RBC systems that are discussed later on, the

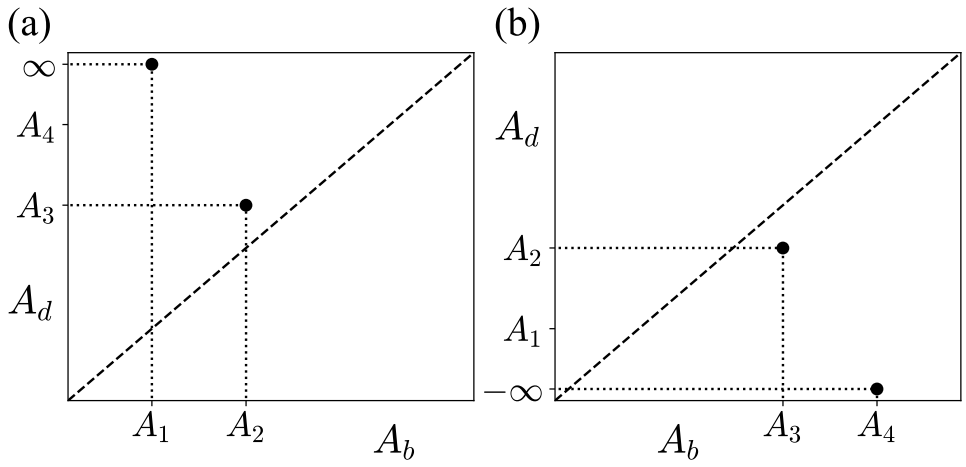

Figure 2.9: A persistence diagram for the (a) sublevel and (b) superlevel filterings of $f(x)$ in Figure 2.7. The points both represent a distinct connected component. The dashed line is the diagonal representing $A_d = A_b$ Their x and y positions denote the value of their births A_b and deaths A_d, respectively. Dotted lines show their position in the diagram more clearly.

superlevel persistence points do not have such a trivial relationship to the sublevel persistence points, so it is important to include both.

2.4.3 Computational homology and persistent homology of 2-D image data

Now that we understand how to compute the persistent homology for a 1-D function, we will extend the discussion to 2-D image data. For example, a shadowgraph snapshot from Rayleigh-Bénard convection in the spiral defect chaos regime is considered (Figure 2.4). This data is a collection of pixels 8-bit gray scale intensity values from 0 to 255 that act as a proxy for the temperature of the fluid in that region.

To get an intuition on the topological structures present in the convection roll pattern in Figure 2.4, the image is sublevel filtered at the median gray scale value $A = 127$ to generate the surface S on which we can compute homology (Figure 2.10. We can see that in a 2-D surface, the familiar distinct connected components exist, just as they did in the 1-D surface. Note that the boundary conditions, unlike in the 1-D case, are not periodic. In a 2-D surface, we also have *holes*, the first dimensional topological structures counted

Figure 2.10: Convection roll pattern from Figure 2.4 sublevel filtered at $A = 127$.

by the first Betti number β_1. Holes are empty regions that are completely enclosed by a distinct connected component. Just as we computed the persistent homology of the 1-D function and identified its distinct connected components, we can compute the persistent homology of a 2-D image and identify its distinct connected components and holes. Both structures can be plotted in persistence diagrams.

In Figure 2.11, the sub- and superlevel persistence diagrams of Figure 2.4 are plotted for both distinct connected components (Figure 2.11(a)) and holes (Figure 2.11(b)). Samples of the Python code used to generate these persistence diagrams are available in subsection D.2.3 and subsection D.2.4. These sample codes also show how to detect plumes in the persistence data, which will be discussed in chapter 4. Access to the sample code IPython notebooks is available upon request, which contain more helpful functions than just those shown in subsection D.2.3 and subsection D.2.4 .

When we compute the persistent homology of a data set, we can think of the output as a representation of that data in the space of persistence diagrams. This representation greatly reduces the dimension of the data set. Geometric factors, such as the shapes or orientation of distinct connected components do not matter in this space. All that matters are the

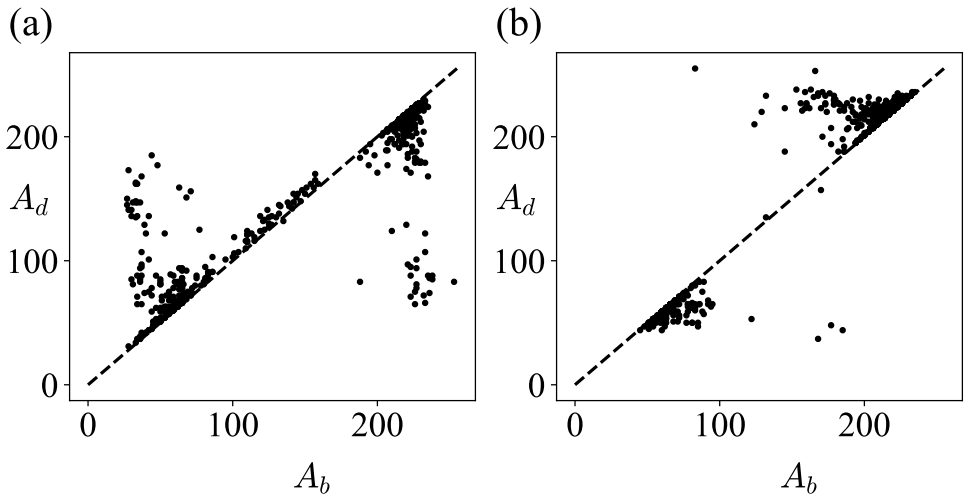

Figure 2.11: Sublevel and superlevel persistence diagrams of Figure 2.4 for (a) distinct connected components and (b) holes.

number of identifiable topological structures and their prominence, or range of thresholds A over which they persist. We note that this representation is invariant to myriad symmetries, such as translations, reflections, stretching, rotations, etc. Depending on what properties of a data set one wants to look at, this large reduction of information and invariance to symmetries can be a feature, not a bug.

2.4.4 Distance metrics in the space of persistence diagrams

If we want to analyze a time-series of data in the space of persistence diagrams, it is necessary to have metrics that make quantitative comparisons between persistence diagrams. A couple of such metrics are worthy of specific consideration. Say there are two persistence diagrams h_1 and h_2. They could be either sub- or superlevel persistence diagrams of either distinct connected components or holes. First, we can define the *degree-p Wasserstein distance* with a L_2 norm ground metric as

$$d_{W^p}(h_1, h_2) = \left(\inf_{\gamma : h_1 \to h_2} \sum_{(p_1, p_2) \in (h_1, h_2)} ||p_2 - \gamma(p_1)||_2^p \right)^{1/p} \tag{2.8}$$

where $||||_2^p$ is the L_2 norm raised to the pth power. p_1 and p_2 are points in the persistence diagrams h_1 and h_2, respectively. The Wasserstein distance finds a mapping γ of the points in h_1 to h_2 that minimizes sum of the distances between matched points $\sum_{(p_1,p_2)\in(h_1,h_2)} ||p_2 - \gamma(p_1)||_2^p$. Note that the diagonals of the persistence diagrams h_1 and h_2 act as a reservoir of infinite persistence points that can be used for matching. So if, say, h_1 and h_2 have differing numbers of persistence points, the Wasserstein distance between them can still be computed. It is also common to use the infinity norm $||||_\infty$ as a ground metric, but this norm is not used in this work. The value of p provides a weighting of pairs for the distance calculation. With $p=1$, all pairwise distances are weighted equally in the calculation. As p is increased, pairs that are further apart are given greater weight in the calculation. In the limit $p \to \infty$, the only pairwise distance that matters is the largest one. This metric is referred to as the *bottleneck distance*. We can define the bottleneck distance more succinctly as

$$d_B(h_1, h_2) = \inf_{\gamma: h_1 \to h_2} \max_{(p_1,p_2)\in(h_1,h_2)} ||p_2 - \gamma(p_1)||_2 \tag{2.9}$$

where in this case the ground metric is the L_2 norm.

These metrics provide quantitative information about the difference between persistence diagrams. Depending on the properties of the persistence diagrams in question, one might choose different values for p. Say you want to compare persistence diagrams that both have a very prominent structure but the other structures are associated with noise or otherwise uninteresting. Then, the best metric for the job would be the bottleneck distance. On the other hand, say you are interested in the case where the comparison between all the points in the persistence diagrams are important, then you might consider using the Wasserstein distance with $p=1$. Finding the optimal mapping γ, in either case, is a nontrivial task. The algorithm for computing the Wasserstein distance exactly has efficiency $O(n^3)$, where n is the number of persistence points per diagram. Persistent homology software packages,

like GUDHI, provide efficient functions for approximating both Wasserstein and bottleneck distances. The details of these algorithms are beyond the scope of this work, but in many cases these metrics can be approximated at near-linear efficiency. A simple example of a Wasserstein distance pairing can be found in chapter 3 in Figure 3.6(b).

CHAPTER 3
SHADOWING OF RELATIVE PERIODIC ORBITS IN THE KURAMOTO-SIVASHINSKY EQUATION

3.1 Relative periodic orbits in spatiotemporally complex systems with continuous symmetries

Fluid turbulence remains difficult to characterize, despite hundreds of years of scientific study. The governing field equation of fluid turbulence, the Navier-Stokes equation, is a strongly nonlinear partial differential equation for which analytical solutions are intractable. Statistical theories of turbulence that do not rely on analytical solutions of Navier-Stokes have thus been explored in great detail [80, 20]. While such statistical theories have provided insights into fluid turbulence, they fail to predict many phenomena directly related to the governing equations. A particular example of such phenomena are coherent structures, large-scale structures that recur in turbulent fluid flows [22, 81]. The existence of such structures with regular spatial and temporal dependence that appear in turbulent flows challenges the assumption of spatial and temporal uniformity at the core of statistical descriptions of turbulence.

It is clear that a dynamical description of turbulence that is based on the governing fluid flow equations is necessary, even if the equations are difficult to work with. Fortunately, numerical methods for finding solutions to strongly nonlinear PDEs have advanced, and it has been demonstrated that coherent structures correspond to dynamically unstable solutions of Navier-Stokes with simple time-dependence that have been termed *exact coherent structures* (ECS) [25]. ECS are found in a variety of turbulent fluid flow systems such as pipe flow [27, 57], plane Couette flow [82], Taylor-Couette flow [13], channel flow [83], and quasi-2D Kolmogorov-like flow [53, 84]. Classes of ECS include equilibria [82, 53],

relative equilibria (or traveling waves) [83, 27], periodic orbits [84], and relative periodic orbits [26, 13] (RPOs).

Turbulent dynamics take place on a finite-dimensional attractor in state space which governs the temporal trajectory of flows [17, 51, 52]. Eberhard Hopf, in his 1948 paper [85], envisioned turbulence as a deterministic walk through the attractor in which characteristic patterns are visited. We now understand that those characteristic patterns are ECS, embedded in state space and structuring the dynamics and statistical properties of turbulence [53, 86]. Turbulence often "shadows" ECS, moving toward them along their stable manifolds in state space and leaving their vicinity along unstable manifolds [53, 26, 28, 13, 31]. Shadowing events allow for short term prediction of turbulent flow and, in the long-term, are a key dynamical event for a deterministic understanding of turbulence as described by Hopf's vision.

In this work, we characterize shadowing of RPOs by turbulence, which are closely related to periodic orbits. RPOs are solutions of the governing equations that are time-periodic up to a transformation along a direction of continuous symmetry. Equilibria are static points in state space and periodic orbits are loops, while RPOs are tori on which families of symmetry-transformed related solutions reside. RPOs have been observed in both simulations and experiments of Taylor-Couette flow [13] and simulations of pipe flow [58, 31]. In both systems, shadowing of RPOs by turbulence has been detected in simulated data [13, 31], but it remains difficult to deal with continuous symmetries in both systems when characterizing shadowing. In principle, turbulence could shadow any symmetry-related copy of the RPO in state space, so it is not straightforward to compute the distance in state space between turbulence and a RPO. Krygier *et al.* [13] address this issue by minimizing the distance over all possible symmetry transformations in Taylor-Couette flow. Budanur *et al.* [31] reduce the continuous symmetry for the RPOs and turbulence by first-Fourier-mode slicing, representing every symmetry-related state as a single state in a symmetry-reduced state space in which they detect shadowing.

We seek a simple and computationally efficient method for dealing with continuous symmetries for detection of shadowing in the Kuramoto-Sivashinsky (KS) equation [18], a partial differential equation with one periodic spatial dimension and one temporal dimension. This equation serves as a simple model system that mimics some of the properties of fluid turbulence, such as spatiotemporal chaos, in a more accessible setting. Therefore, we will refer to the chaotic evolution of KS as "turbulence" in this work. RPOs have been studied extensively in this model system [26, 28] because it has a relatively low-dimensional inertial manifold [51, 52] while featuring a continuous translational symmetry along its spatial dimension. It is possible to minimize the distance between turbulence and RPOs over all symmetry transformations in KS as Krygier *et al.* [13] do in Taylor-Couette flow. Even when this minimization is done cleverly, it is computationally expensive. As in pipe flows, Budanur *et al.* [26, 28] reduce the continuous symmetry in KS using Fourier-mode slicing. While this method has yields insights into the RPOs of KS, it can be difficult to deal with discontinuities when trajectories cross the "slice border" (i.e. when the slicing Fourier mode vanishes). Thus, we seek a symmetry-reduced representation of KS that is suitable for detecting shadowing of RPOs by turbulence and is simple to implement and computationally inexpensive for large data sets.

A symmetry-reduced representation of KS can be obtained using persistent homology [39], which encodes information about the topological structure of a KS state. This topological information is invariant to continuous symmetries such as those present in KS, Taylor-Couette flow, and pipe flow. Time-series of data in dynamical systems can be represented in the space of persistence diagrams [47]. Moreover, the distance between persistence diagrams representing states can be quantified using metrics such as the Wasserstein distance [87]. These prerequisites allow us to detect shadowing in the space of persistence diagrams instead of a state space that is costly to symmetry-reduce.

In this work, we detect and characterize shadowing of a catalog of RPOs by turbulence in KS using two methods, comparing and contrasting the computation time and shadowing

statistics yielded by both. In the first method, we minimize the distance over the continuous symmetry in state space between a turbulent state and RPOs, similar to the procedure described for Taylor-Couette flow by Krygier *et al.* [13]. In the second method, we compute time-series of persistence diagrams for turbulent and RPO states, then compute the Wasserstein distance between states in the space of persistence diagrams.

This article is organized as follows. In section 3.2, we describe the procedure for detecting shadowing in KS in state space. In section 3.3, we describe the procedure for detecting shadowing in the space of persistence diagrams. In section 3.4, we discuss the computational complexity of both approaches. In section 3.5, we show statistics of shadowing events detected by both the state space and space of persistence diagrams approaches and discuss the results. In section 3.6, we discuss shadowing of an equilibrium solution. In section 3.7 we offer conclusing remarks.

3.2 Shadowing in state space

The KS equation in one spatial dimension is a fourth-order nonlinear partial differential equation of a scalar field $u(x,t)$ expressed as

$$u_t + u_{xx} + u_{xxxx} + uu_x = 0,\ x \in [0, L] \tag{3.1}$$

where x is the spatial coordinate, t is time, and L is the parameter that sets the width. We consider periodic boundary conditions such that $u(0,t) = u(L,t)$ for all t. The field u, the parameter L and the coordinates x and t are nondimensional. In this work we consider the case where $L = 22$. Due to the periodic spatial boundary conditions, the KS equation has a continuous translational symmetry, meaning that Equation 3.1 is invariant to transformations of the form $u(x,t) \to u(x + \tilde{s}, t)$ where $\tilde{s} \in [0, L]$ is some constant spatial shift. Trajectories of $u(x,t)$ in time are obtained by integrating an initial condition with an exponential time-differencing fourth-order Runge-Kutta method [88] with a time step of 0.02

time units and grid spacing $dx = L/N$, where $N = 32$ grid points. After integrating a trajectory of Equation 3.1, we take subsamples of every tenth time step for analysis. Thus, for simplicity in this work we will refer to $dt = 0.2$.

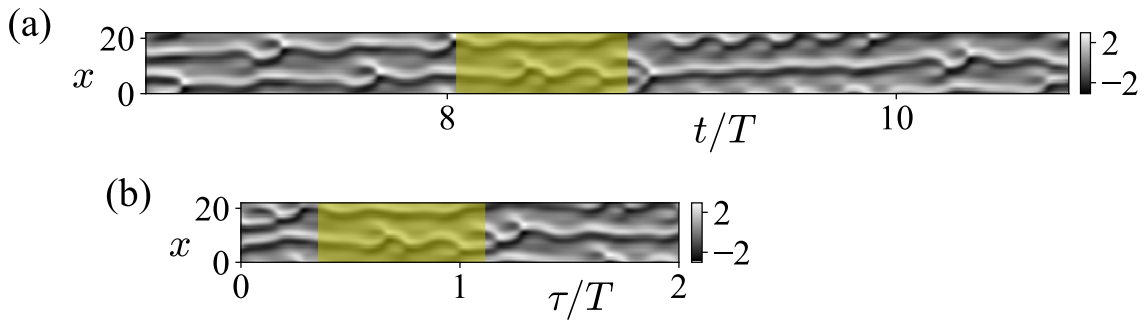

Figure 3.1: A long time series of (a) turbulent KS often shadows exact coherent structures such as (b) RPOs. In (a), we show a spacetime segment of a long turbulent trajectory $u_2(x,t)$ of (Equation 3.1). In (b), we show a spacetime segment of a RPO $u_1(x,\tau)$ over two periods $T = 97.3$ (Table 3.1). In both spacetime blocks, we highlight two segments that are visually similar. We seek to characterize this shadowing behavior of the RPO by the turbulent trajectory.

A turbulent trajectory of the KS equation will traverse a turbulent attractor in a high-dimensional state space in which exact coherent structures are embedded. Turbulence often appear visually similar, or "close" to exact coherent structures over finite stretches of time and appears to be influenced by them. Two prominent examples of exact coherent structures are equilibria and periodic orbits. Equilibria are static states, or single points in state space, so turbulence can pass close to them in state space and then depart. Periodic orbits, however, are one-dimensional trajectories that evolve through state space in time, so turbulent trajectories can not only pass close to regions of the periodic orbit, but they can *co-evolve* alongside such an orbit before departing. Such behavior is commonly referred to as *shadowing* (Figure 3.1) and we seek to characterize this behavior in Equation 3.1. Continuous symmetries, such as the continuous translational symmetry in KS, allow for the existence of relative equilibria and RPOs, which are families of equilibria and periodic orbits. The members of these families are identical up to a shift in the continuous symmetry direction. While a periodic orbit takes the form of a one-dimensional trajectory in state

space, a RPO is a torus on which a family of trajectories evolves. RPOs in Equation 3.1 are periodic trajectories that satisfy

$$u(x,\tau) = u(x+\Phi, \tau+T) \tag{3.2}$$

where τ is the temporal coordinate along a RPO, T is the period, and Φ is the spatial shift of the RPO over a period. In this work, we examine 16 RPOs (Table 3.1) in the $L = 22$ regime that have been computed using a multi-shooting method described in Appendix A of Budanur *et al.* [26]. We seek to characterize and quantify shadowing by turbulence of RPOs, which are plentiful in Equation 3.1 even at the relatively small size $L = 22$ (Figure 3.1(b)).

Table 3.1: Summary of RPOs used in this study. T is the period, λ^{-1} is the inverse of the largest Floquet exponent, or escape time, Φ is the total spatial shift of the RPO over one period, and $L = 22$ is the size of the spatial domain.

RPO	T	λ^{-1}	λ^{-1}/T	Φ/L
1	32.8	53.7	1.64	0.498
2	36.0	20.5	0.569	0.179
3	47.9	19.8	0.415	0.258
4	59.9	26.8	0.447	0.247
5	69.3	11.5	0.166	0.437
6	80.5	24.2	0.301	0.171
7	81.5	6.05	0.074	0.130
8	82.4	12.0	0.145	0.137
9	83.3	6.10	0.073	0.456
10	87.9	5.66	0.064	0.218
11	88.9	8.42	0.095	0.166
12	93.1	6.21	0.067	0.215
13	94.2	8.66	0.092	0.199
14	94.2	9.63	0.102	0.159
15	96.9	13.3	0.137	0.007
16	97.3	10.8	0.111	0.38

A prerequisite to the detection of shadowing is the detection of close passes to a RPO by turbulence. To detect close passes, some notion of distance in state space between a turbulent trajectory and the family of trajectories that make up a RPO must be employed.

For such a metric, it is common to employ the Euclidean norm [13, 84]

$$D(t,\tau) = \min_{\tilde{s}\in[0,L]} D_{\tilde{s}}, \tag{3.3}$$

where

$$D_{\tilde{s}} = D_r^{-1} ||u_2(x+\tilde{s},t) - u_1(x,\tau)||_2 \tag{3.4}$$

is the difference between two states u_1 (RPO) and u_2 (turbulence) at times t and τ and $\tilde{s}\in[0,L]$ is a constant spatial shift. The scale factor D_r is an estimate of the attractor radius that will be discussed shortly. Because the u_1 exists on a torus in state space, we seek to find the closest point of that entire torus to u_2, thus we have to find a shift in the continuous symmetry direction (in this case, continuous translational symmetry) that minimizes the distance between u_1 and u_2. The shift that minimizes Equation 3.3 $s = \arg\min_{\tilde{s}\in[0,L]} D_{\tilde{s}}$ is found using a Fourier method described in Appendix A. (Figure 3.2).

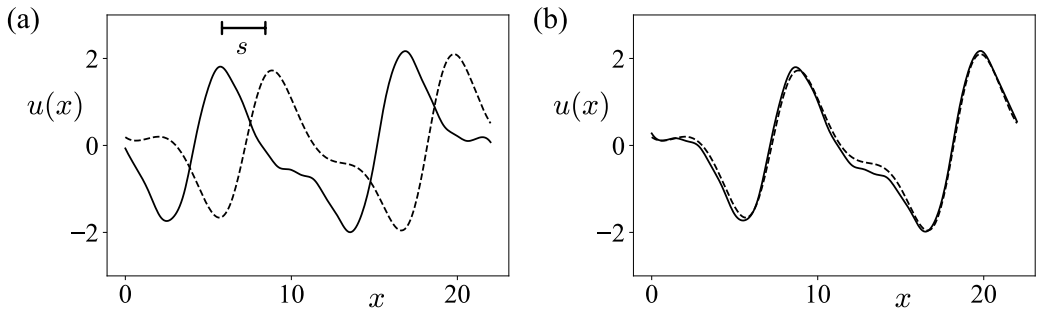

Figure 3.2: Continuous translational symmetry must be addressed to use Equation 3.3 to compare a time step of a turbulent trajectory (solid line) to a time step of a RPO (dashed line). In (a), we plot the value of u for a single time step as a function of x for the turbulent trajectory and RPO. In (b), the turbulent point is shifted in space by s, the amount that minimizes Equation 3.3 for the turbulent point and the RPO. We find such a shift for all t and τ to compute $D(t,\tau)$ in Equation 3.5.

In principle, the RPO could be visited by the turbulence at any point in time along its trajectory. So, to find the closest approach to the RPO by the turbulence, we define the metric

$$D_1(t) = \min_{\tau\in[0,T]} D(t,\tau) \tag{3.5}$$

to detect the closest distance between the turbulent trajectory at a given time t and any time on the RPO τ (Figure 3.3).

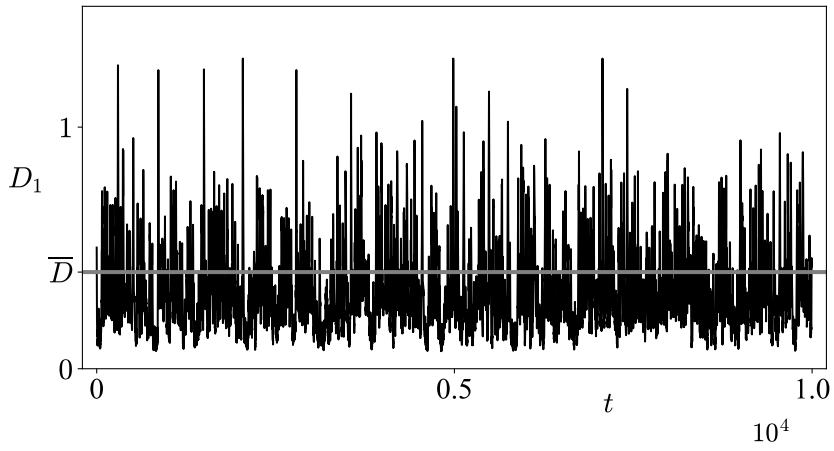

Figure 3.3: Close passes of the turbulent trajectory to a RPO are detected at minima of $D_1(t)$ (Equation 3.5), the closest distance of the turbulent trajectory $u_2(x,t)$ at t to RPO $u_1(x,\tau)$ for all τ. Shown is $D_1(t)$ for a long turbulent trajectory and RPO 16 in units of the estimated radius of the turbulent attractor D_r. When the turbulent trajectory passes below the threshold $\overline{D} = 0.4$, it is considered a close pass. The minimum value of D_1 for each close pass is then checked for shadowing.

We scale $D(t,\tau)$ (and thus $D_1(t)$ as well) by an estimate of the turbulent attractor radius $D_r = \langle ||u_2(x,t) - u_m(x)||_2 \rangle_t$, where $u_m(x)$ is the mean point of the attractor and $\langle\rangle_t$ denotes the average over t, to obtain a physically intuitive scale of distance in state space. Because of the continuous translational symmetry of the turbulent trajectory, obtaining $u_m(x)$ is nontrivial. If we were to naively time-average a long turbulent trajectory, all of the dynamics would be averaged out as the trajectory slowly drifts in space, and the result would converge to $u_m(x) = 0$, which is not a dynamically relevant state. Instead we slice the turbulent trajectory by its second Fourier mode, similar to the first Fourier mode slice procedure described by Budanur *et al.* [26] to obtain a symmetry-reduced representation of the turbulent trajectory $u_2^*(x,t)$. The mean state $u_m(x) = \langle u_2^*(x,t) \rangle_t$ resembles a realizable state of Equation 3.1.

We define a close pass to occur when $D_1(t) \leq \overline{D}$ where $\overline{D} = 0.4$, i.e. when the turbulence passes closer to the RPO than 40% of D_r (Figure 3.3). This threshold was chosen by

visual inspection of the states $u_2(x+s,t)$ and $u_1(x,\tau)$ for similarity. While this method of choosing a threshold value is subjective, it has been used successfully in prior work [59, 13]. Also, while this threshold may seem quite large relative to the attractor radius, we note that the inertial manifold dimension $M = 9$ [51, 52]. Thus the volume of state space surrounding a point in time on a RPO that is within $\overline{D}$ scales as $\overline{D}^M \approx 10^{-4}$ or less than 0.01% of the attractor volume. Minima in $D_1(t)$ at t_c and a corresponding τ_c are chosen as close passes if they are the minimum point within a 160 time unit interval. This is done to avoid double counting close pass events that occur very close to each other in time.

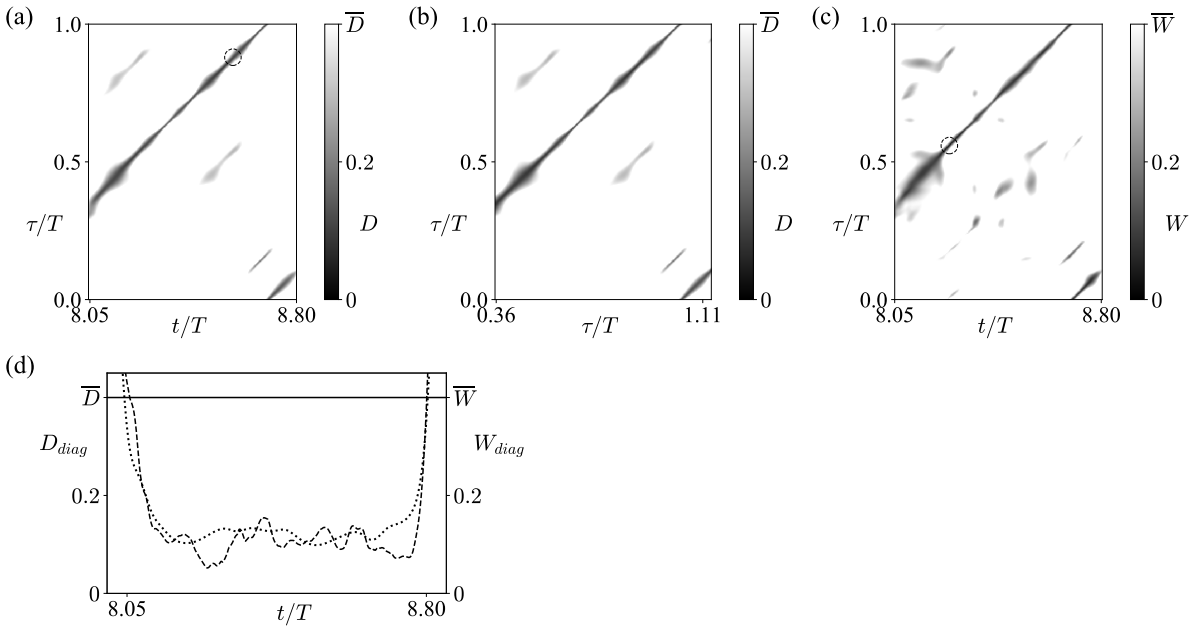

Figure 3.4: Shadowing can be detected by drawing a diagonal line in the heat map of (a) $D(t,\tau)$ or (c) $W(t,\tau)$ through a close pass at (t_c,τ_c) and checking that $D(t_c+t_s,\tau_c+t_s) \leq \overline{D}$ (same for W) for some interval of t_s that is longer than t_{corr}. In (a) and (c), the dashed circle highlights the time of the close pass (t_c,τ_c). For comparison to (a), we show in (b) the auto-recurrence of the RPO, i.e. the heat map of $D(\tau,\tau)$ comparing the RPO $u_1(x,\tau)$ with itself. In (d), we denote $D(t_c+t_s,\tau_c+t_s)$ (dotted line) and $W(t_c+t_s,\tau_c+t_s)$ (dashed line) as D_{diag} and W_{diag}, respectively. The solid horizontal line is the threshold distance $\overline{D} = \overline{W} = 0.4$. The shadowing event depicted corresponds to highlighted portions of the turbulent trajectory and the RPO in Figure 3.1.

Once we have detected a close pass in $D_1(t)$ at points in time t_c and τ_c, we search for shadowing in the form of a diagonal line running through the heat map $D(t_c,\tau_c)$

parametrized by shadowing time t_s, such that $D(t_c + t_s, \tau_c + t_s) \leq \overline{D}$ for all $t_s \in [S_a, S_b]$ (Figure 3.4(b) and (d)). $S_a \leq 0$ is the length of time before the close pass where there is shadowing and $S_b \geq 0$ is the length of time after the close pass where there is shadowing, thus the total time of a shadowing event $T_s = S_b - S_a$. We require that shadowing events have length of $T_s \geq t_{corr} = 8.64$, where t_{corr} is the correlation time, to ensure that the turbulence and RPO are co-evolving. t_{corr} is defined as the value of t' when the temporal auto-correlation function

$$C(t') = \frac{\langle u_2(x,t) \cdot u_2(x,t+t') \rangle_t}{\langle u_2(x,t) \cdot u_2(x,t) \rangle_t} \tag{3.6}$$

first drops below $1/e$. t_{corr} measures how long it takes the turbulent trajectory to significantly change.

The shift s that minimizes $D(t,\tau)$ can change throughout a potential shadowing event due to the continuous translational symmetry. To avoid detection of spurious shadowing events in which s changes abruptly, we require that

$$|s(t_s) - s(t_s - dt)| \leq dx \tag{3.7}$$

throughout a shadowing event when using $D(t,\tau)$ as the metric. If a close pass is detected, but $s(t_s)$ violates this condition, shifts $\tilde{s}$ that satisfy Equation 3.7 are tested and accepted if the distance between the turbulence shifted by $\tilde{s}$ and the RPO is below $\overline{D}$, i.e. if the condition $D_r^{-1} ||u_2(x + \tilde{s}, t_c + t_s) - u_1(x, \tau_c + t_s)||_2 \leq \overline{D}$ and $T_s \geq t_{corr}$ is met.

To smooth out fluctuations at small time scales over the course of a shadowing event, we average the metric $D(t,\tau)$ over diagonals in t and τ. Specifically, we take

$$D_w(t,\tau) = \frac{dt}{w} \sum_{t'=0,dt,2dt,\ldots}^{w-dt} D(t+t', \tau+t') \tag{3.8}$$

where D_w is the average of D over diagonals of length w and the time step $dt = 0.2$. In this

work, we consider the cases $w = dt$ (no averaging) and $w = 20dt = 4.0$. From here on, D will refer to the averaged $D_{4.0}$ unless otherwise specified.

3.3 Shadowing in the space of persistence diagrams

The Euclidean norm applied in state space is a common metric for determining the distance between two states, however we present here an alternative method that employs persistent homology, an extension of computational homology. Both computational homology and persistent homology identify unique topological structures in a data set, however persistent homology also provides a natural organization of those structures according to their prominence into a "persistence diagram". Persistence diagrams present a symmetry-reduced alternative to the state space representation for the KS system in which a family of trajectories that make up a RPO are collapsed to a single trajectory. This representation simplifies the detection of close passes of turbulence to RPOs. Just as the distance between two states in state space can be computed using Equation 3.3, the distance between two states in the space of persistence diagrams can be computed using a family of metrics known as the Wasserstein distance. In this section, we will describe how close passes, and subsequently shadowing, can be detected in the space of persistence diagrams in much the same way as the traditional method described in section 3.2.

States of Equation 3.1, such as those shown in Figure 3.1, can be projected into the space of persistence diagrams by filtering the points in the field $u(x,t)$ into a binary set, with one set constituting a surface to be analyzed and the other set representing empty space. We filter the state based on a threshold value A. Considering a single time step of KS $u(x)$, in "sublevel filtering", all points of $x \in [0, L]$ that satisfy $u(x) \leq A$ are considered members of the surface and all others are considered empty. Figure 3.5(a) shows a cartoon example of a periodic function $u(x)$ sublevel filtered at four different thresholds A_1, A_2, A_3, and A_4. In Figure 3.5(b), the corresponding filterings are shown for each filtering threshold. We also consider "superlevel filtering", though it is not shown

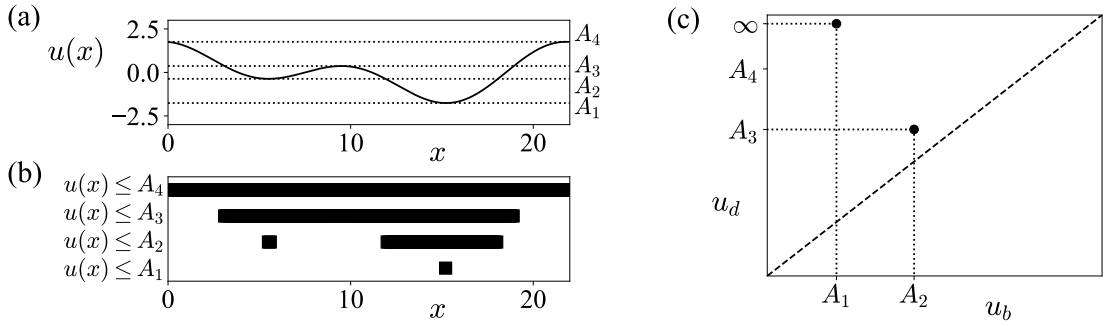

Figure 3.5: States of Equation 3.1, such as the cartoon in (a) can be represented by persistent homology, which is invariant to the continuous translational symmetry. In (a), there is a cartoon of a KS trajectory $u(x)$ (solid curve) that is periodic in x and filtering thresholds $A_1 = -1.76$, $A_2 = -0.37$, $A_3 = 0.37$, and $A_4 = 1.76$ (dotted lines). In (b) we show binary sublevel filterings of $u(x)$ corresponding to each A. Values of x that satisfy $u(x) \leq A_i$ with $i = 1, 2, 3, 4$ are plotted as black components in each row from bottom to top, while values of x that do not satisfy this condition are plotted as white. Homology then detects and quantifies the distinct connected components of these surfaces. At the threshold A_1 there is one distinct connected component. At A_2 a second has been born. At A_3, the two distinct connect components have merged into one, so one of the components has died. At A_4, all points of $u(x)$ lie under threshold and there remains one distinct connected component that will remain for all greater threshold values. (c) shows the persistence diagram for $u(x)$ in (a) in which each component is plotted at the location of its birth (u_b) and death (u_d) thresholds. Both distinct connected components of $u(x)$ that appear in (b) are plotted at their birth thresholds A_1 and A_2. The component that was born at A_2 dies at A_3, where it is plotted. The component born at A_1 does not die at any threshold value so its death threshold u_d is plotted at infinity. In further analysis, we throw out this "infinite" persistence point. We have used an artificial periodic function $u(x)$ in this figure for demonstration purposes because states of Equation 3.1 often feature minima and maxima at very similar values (see Figure 3.2).

in Figure 3.2, in which all points that satisfy $u(x,t) \geq A$ are considered members of the surface and all others are considered empty. We can then detect the number of distinct connected components of these filterings using 1-dimensional computational homology. A distinct connected component is *born* at A_1 and A_2 and one of those components *dies* at A_3 when the two distinct connected components merge. A_4 is the maximum of $u(x)$, so every value of $u(x)$ now makes up one distinct connected component that will not die at any value higher than A_4, so it is said to die at infinity. The birth and death thresholds of these distinct connected components are plotted in persistence diagrams (Figure 3.5(c)).

We use a cartoon state $u(x)$ for the purpose of clarity in Figure 3.5 because real states of Equation 3.1 often feature minima and maxima at very similar values (see Figure 3.2). That being said, the persistence diagrams still accurately capture the topological structure of real states. Note that the number of distinct connected components yielded by a given filtering threshold is invariant to spatial shifts s like the one in Figure 3.2(a), therefore the persistence diagrams are invariant to such shifts. All possible spatial shifts along the continuous symmetry direction are represented as a single point in the space of persistence diagrams.

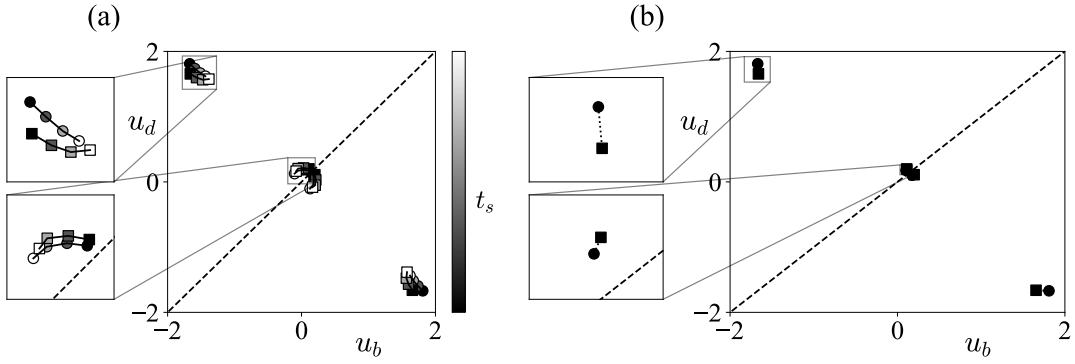

Figure 3.6: During shadowing events, the turbulence and RPO co-evolve through the space of persistence diagrams and the difference between their persistence diagrams can be quantified by $W(t, \tau)$. In (a), the persistence points of both the turbulence (circles) and the RPO (squares) are represented in both the sublevel (above the diagonal) and superlevel (below the diagonal) filterings for multiple time steps during the shadowing event in Figure 3.1. Each time step shown is coded by gray scale intensity such that trajectory and RPO points that are the same intensity occur at the same time step of the shadowing event t_s. In (b), persistence points that are paired by the mapping γ to compute the Wasserstein distance $W(t, \tau)$ in (Equation 3.9) are connected with dotted lines.

During shadowing events, turbulence and RPOs exhibit close passes and co-evolution in the space of persistence diagrams in addition to state space (Figure 3.6(a)). Thus, we seek to quantify shadowing in the space of persistence diagrams using a metric similar to D. The RPO $u_1(x, \tau)$ and turbulence $u_2(x, t)$ of Equation 3.1 can be represented as persistence diagrams that reduce the continuous translational symmetry and then compared at every t and τ. We will refer to these persistence diagrams as $h_1(x, \tau)$ and $h_2(x, t)$, respectively. These persistence diagrams include both sublevel and superlevel filtered persistence

points. We can detect close passes between $h_1(x,\tau)$ and $h_2(x,t)$ in the space of persistence diagrams in a similar way to how we detect close passes in state space. We first require a metric that is analogous to $D(t,\tau)$ that quantifies a distance between $h_1(x,\tau)$ and $h_2(x,t)$. We use the degree-1 *Wasserstein distance* with a L_2 norm ground metric given by

$$W(t,\tau) = {W_r}^{-1} \inf_{\gamma:h_1(x,\tau)\to h_2(x,t)} \sum_{(p_1,p_2)\in(h_1,h_2)} ||p_2 - \gamma(p_1)||_2 \tag{3.9}$$

where γ is a one-to-one mapping of persistence points p_1 in h_1 to persistence points p_2 in h_2 and W_r is the estimate of the attractor radius in the Wasserstein metric. The mapping γ that minimizes the sum in Equation 3.9 is used for $W(t,\tau)$. γ cannot match sublevel points to superlevel points. An example of how the optimal γ would pair points in two persistence diagrams to minimize the sum of distances between points is shown in Figure 3.6(b). It is possible that h_1 and h_2 might have a different number of off-diagonal points. In that case, any leftover points are paired with the diagonal, which acts as a reservoir of persistence points. The estimate of the attractor radius W_r is given by

$$W_r = \langle \inf_{\gamma:h_1(x)\to h_2(x,t)} \sum_{(p_1,p_2)\in(h_1,h_2)} ||p_2 - \gamma(p_1)||_2 \rangle_t \tag{3.10}$$

where $h_1(x)$ is the persistence diagram of the estimate of the attractor's mean point $u_m(x)$ discussed in the Section section 3.2 and $h_2(x,t)$ is the persistence diagram of the turbulent trajectory at t. When D and W are expressed in units of the attractor radius in the respective metric, the close pass threshold $\overline{W} = \overline{D} = 0.4$ in both cases. Similar to $D_1(t)$, we define

$$W_1(t) = \min_{\tau\in[0,T]} W(t,\tau) \tag{3.11}$$

to detect close passes of the turbulence to a RPO in the Wasserstein metric. Once close passes are detected in $W_1(t)$ we detect shadowing events using a similar method to the one described for $D(t,\tau)$ (Figure 3.4(c) and (d)). We also consider diagonally averaged $W_{4.0}$

and $W_{0.2}$ (the latter case being no averaging) just as we do with D. Identically to that case, W denotes $W_{4.0}$ unless otherwise specified. The only difference between the computation of shadowing in D and W is that in W there is no need to check that the relative shift s between the RPO and the turbulence is changing slowly because the space of persistence diagrams are invariant to such shifts. The same shadowing event shown in Figure 3.4(a) is detected in W in Figure 3.4(c) and the distance in D and W along the diagonal are compared in Figure 3.4(d).

To compare the shadowing events detected by D and W, we are interested in the conditional probability of shadowing in W given D $P(W|D)$, i.e. the probability that a shadowing event is detected in W when a shadowing event is detected in D, and the opposite case $P(D|W)$. To compute, for example, $P(W|D)$, we first find shadowing events in D throughout a long turbulent time series (10^4 time units). Then, for a given shadowing event in D centered around close pass (t_c, τ_c) we check if there is a valid shadowing event that runs through the same (t_c, τ_c) in W. However, we allow that the corresponding shadowing event in W could begin within t_{corr} of (t_c, τ_c). $P(W|D)$ is thus the percentage of shadowing events in D that yield a corresponding event in W according to these criteria.

3.4 Benchmarking

To assess the operational advantages and disadvantages of the persistence space shadowing detection over the minimization method, the time to compute D and W is benchmarked on the same machine and programming environment. In this section, the theoretical computational complexity of both approaches are discussed and the results of a benchmarking analysis are presented. It is demonstrated that for larger data sets of the size needed to gather meaningful shadowing statistics, W takes less time to compute than D.

The computational cost of detecting shadowing events is dominated by the computation the distances between every time step of a RPO of length m and a turbulent trajectory of length n. Thus, we compare the complexity of computing D and W. When calculating

D, the spatial shifts s that minimize Equation 3.3 must be found and then the Euclidean distance between the shifted trajectory and the RPO must be found. Via the Fourier method, finding all needed spatial shifts has complexity $\Theta(mnN \log N)$, where N is the number of spatial points of the RPO and trajectory. Performing all needed computations of Euclidean distance after shifting appropriately has complexity $\Theta(mnN)$. So, the time it takes to find the optimal spatial shift dominates the complexity of D.

In contrast, finding $W(t, \tau)$ does not involve finding any spatial shifts, but incurs a cost in computing persistence points, and requires a potentially more costly distance calculation. In the worst case finding the persistence diagram for N data points has complexity $\Theta(N^3)$. However, near-linear performance is common and we observed near-linear performance as well when varying the number of grid points for the trajectories. So, the average complexity of calculating all needed persistence diagrams is $\Theta((m+n)N)$. Next, finding the Wasserstein distance between two persistence diagrams has complexity $\Theta(P^3)$, where P is the total number of persistence points. P is dependent on the spacing and magnitude of the extrema of the RPO and chaotic trajectory: more complicated flows will have persistence diagrams with more points. But, it is also true that more complicated flows will have more spatial points (increased N), increasing the complexity of the computation of D. Persistence diagrams we obtained generally did not have many points, so on average the complexity to find all needed Wasserstein distances, $\Theta(mnP^3)$, did not grow as rapidly as the cubic term might suggest. In practice for larger trajectories, it was faster to compute $W(t, \tau)$ than $D(t, \tau)$ (Figure 3.7). When computing D and W for a long turbulent trajectory (about 10^{-4} time units) and a single RPO, the computation of D takes an order of magnitude longer than W (Figure 3.7).

3.5 Statistics of relative periodic orbit shadowing

When examining shadowing of a long time series (10^4 time units) of turbulence to a single RPO, such as RPO 14 in Figure 3.9, dozens of shadowing events are detected, and the

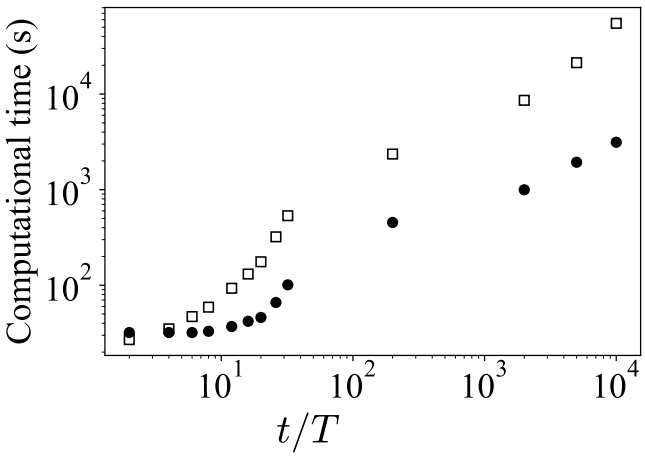

Figure 3.7: The computation of the heat map of $W(t, \tau)$ (closed circles) is faster than $D(t, \tau)$ (open squares) for large data sets of the size needed to detect multiple shadowing events over a long turbulent trajectory. Here, the time in seconds it takes to compute both D and W for a full RPO and a varying length of time of turbulence is shown. The computation time for W includes the time it takes to generate the persistence diagrams for the turbulence and RPO.

lengths of shadowing events T_s can vary over a wide range. The mean shadowing time $\langle T_s \rangle_t / T$ is positively correlated with the time associated with the largest unstable Floquet exponent λ^{-1}/T (Figure 3.8). In other words, RPOs that are more linearly stable have longer shadowing events as a portion of their period, on average. This occurs despite the fact that shadowing occurs well away from the regime where linear stability can be assumed. $\langle T_s \rangle_t / T$ increases linearly with λ^{-1}/T for this set of RPOs (Table 3.1) in the case of D and W (Figure 3.8). For W, $\langle T_s \rangle_t / T$ tends to be slightly larger than for D. Most RPOs in the set are clustered together with small values of $\langle T_s \rangle_t / T$ and λ^{-1}/T in Figure 3.8. These tend to be RPOs with longer periods and similar statistics of shadowing events. RPO 14 was chosen as a representative of this cluster in Figure 3.9. We can see that the length of shadowing events of RPO 14 fall off at the rate of the largest unstable Floquet exponent λ. The shadowing event lengths detected by W (Figure 3.9(b)) tend to be longer than those detected by D (Figure 3.9(a)). RPOs with $\lambda^{-1}/T \geq 0.4$ have longer T_s that sometimes exceed a period of the RPO.

The probability that a given time step t of turbulence is shadowing a given RPO $P(S)$

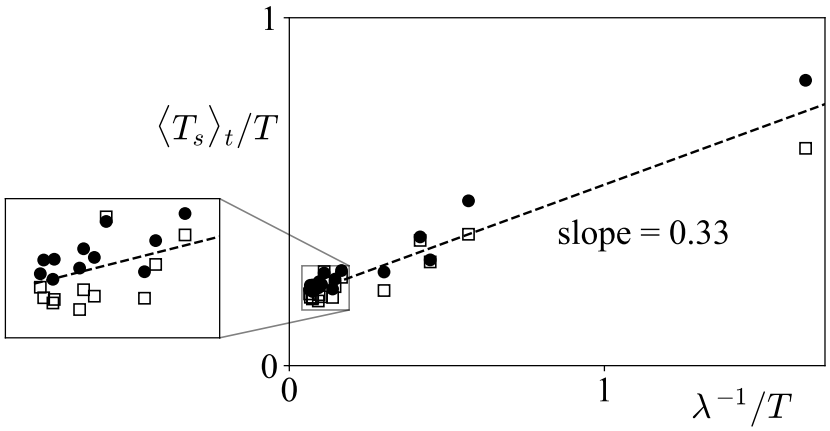

Figure 3.8: The mean time of all shadowing events $\langle T_s \rangle_t / T$ for a RPO and the inverse of the largest Floquet exponent λ^{-1}/T are positively correlated for shadowing events in D (open squares) and W (closed circles). Each point in the scatter plot represents a RPO. A dotted line fit to the points in D and W is shown for reference.

is shown in Figure 3.10 for $D_{4.0}$, $D_{0.2}$, $W_{4.0}$, and $W_{0.2}$. $P(W_{4.0})$ far exceeds $P(W_{0.2})$ for all RPOs, but there is no such difference between $P(D_{4.0})$ and $P(D_{0.2})$. Additionally, $P(W_{4.0})$, $P(D_{4.0})$, and $P(D_{0.2})$ are comparable for all RPOs. By those metrics, between 10% and 20% of the time, the turbulent trajectory is shadowing a given RPO.

The averaged metric $W_{4.0}$ detect shadowing for a higher portion of the turbulent trajectory than the non-averaged metric $W_{0.2}$ (Figure 3.10). However, $D_{4.0}$ and $D_{0.2}$ detect shadowing for a similar portion of the turbulent trajectory (Figure 3.10). The averaged metrics $D_{4.0}$ and $W_{4.0}$ indicate shadowing for a similar portion of time. Note that a turbulent trajectory can shadow multiple RPOs at the same time. When averaging in time is done (i.e. $w = 4.0$), D and W detect shadowing events at similar times very often (Figure 3.11). The outlier point near the middle of the figure is RPO 2. The mean value of $P(W|D)$ in Figure 3.11 is 0.95 and the mean value of $P(D|W)$ is 0.82.

While metrics D and W similarly quantify shadowing behavior in KS, W offer a major operational advantage. Specifically, W is invariant to the continuous translational symmetry of Equation 3.1 and thus computes more quickly than D. Our benchmarking calculations (Figure 3.7) indicate that computation of $W(t, \tau)$ is 4 times faster for a given RPO

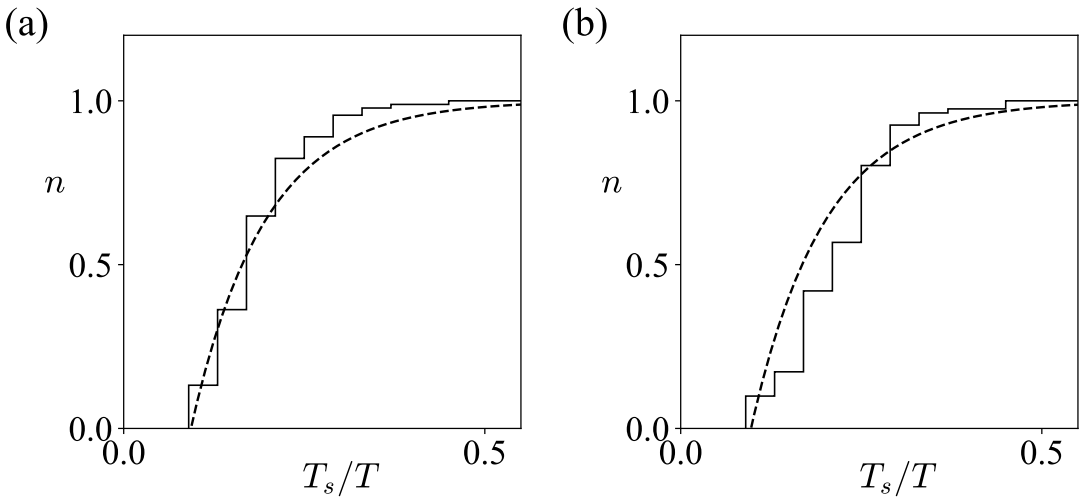

Figure 3.9: Cumulative histograms of shadowing events detected in (a) D and (b) W show that shadowing time falls off at an exponential rate similar to λ determined by sample RPO's largest unstable Floquet exponent and that shadowing events in W are often longer than D. Shown are histograms of shadowing events of RPO 14 binned by the duration of shadowing (solid lines). We also show cumulative exponential functions $1 - e^{\lambda(T_s - \min T_s)}$ (dashed lines) for reference.

when computing for a long time series of turbulence (10^4 time units). This computational advantage comes from the speed of the Wasserstein distance calculation compared to the speed of computation of the norm of the difference between a shifted turbulent trajectory time step and RPO. Finding the shift that minimizes $D(t, \tau)$ for each t and τ is computationally expensive. However, the computation of $W(t, \tau)$ simply requires the (continuous translational symmetry-invariant) persistence diagrams at the times t and τ, which can be computed beforehand.

D and W quantify similar statistics for shadowing events. Figure 3.8 shows the mean shadowing time $\langle T_s \rangle_t / T$ of the turbulence to each RPO in Table 3.1 compared to their escape time λ^{-1}/T. In both the D and W metrics, $\langle T_s \rangle_t / T$ is positively correlated with λ^{-1}/T. The escape time λ^{-1}/T of each RPO is the inverse of the largest unstable Floquet exponent, which quantifies how much a trajectory that is slightly perturbed from the RPO will diverge from it after one period T. RPOs with larger λ^{-1}/T are less linearly unstable than RPOs with smaller λ^{-1}/T. It is notable that $\langle T_s \rangle_t / T$ is positively correlated

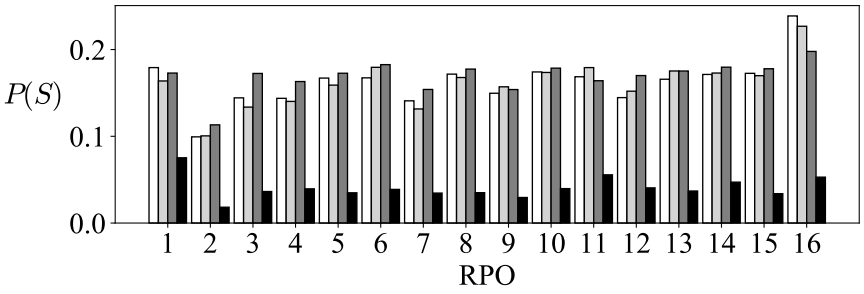

Figure 3.10: D and W metrics detect shadowing at similar rates across multiple RPOs when averaged over a window of $t_w = 4.0$. Here we show the probability that there is shadowing $P(S)$ of given RPO at any t of the turbulent trajectory for $D_{4.0}(t, \tau)$ (white bars) $D_{0.2}(t, \tau)$ (light gray bars), $W_{4.0}(t, \tau)$ (dark gray bars), and $W_{0.2}(t, \tau)$ (black bars). $P(W_{0.2})$ is much lower than $P(W_{4.0})$ for all RPOs, while $P(D_{0.2})$ and $P(D_{4.0})$ are similar to each other and $P(W_{4.0})$.

with λ^{-1}/T because shadowing is commonly detected well outside of the "linear regime", i.e. the region around the RPO in state space where the dynamics are governed by its linear manifold. Where solutions are governed by the linear manifold, we expect turbulence to approach the RPO exponentially along a stable direction of the manifold before leaving exponentially along an unstable direction of the manifold. In shadowing events that are detected, we rarely observe this linear behavior. Instead, we observe shadowing that is nonlinear in nature which we would not, in principle, expect to have any relationship to λ^{-1}/T. We also observe in Figure 3.8 that, for most RPOs, $\langle T_s \rangle_t / T$ is on average larger for W than for D. This could occur because turbulence tends to approach or leave a shadowing event for a short time along a direction of state space to which the space of persistence diagrams is invariant. This difference would be quantified by D but not W. There is a cluster of RPOs in the zoomed inset of Figure 3.8 that have relatively low λ^{-1}/T and $\langle T_s \rangle_t / T$ and high T. These RPOs rarely show shadowing events longer than half a period (Figure 3.9). There is another subset with $\lambda^{-1}/T \geq 0.4$ and low T which sometimes show shadowing events longer than an entire period (Figure 3.8).

Figure 3.10 demonstrates the large effect of temporal averaging along the diagonal (Equation 3.8) on the probability of detecting shadowing in W ($P(W)$), while the averaging

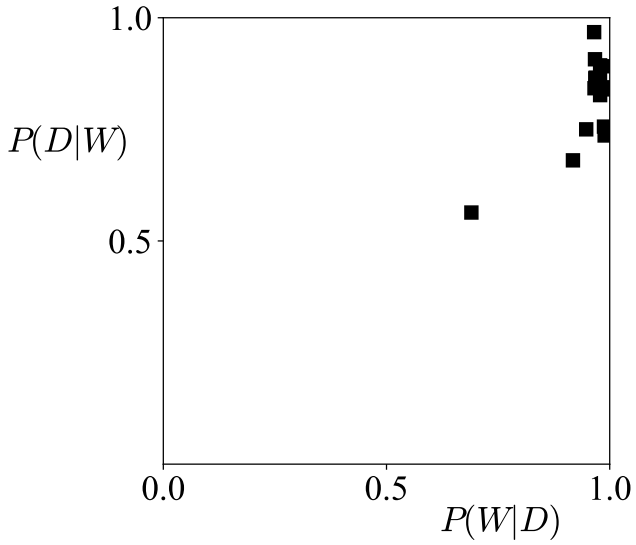

Figure 3.11: D and W detect shadowing at similar times when $w = 4.0$. $P(W|D)$ represents the probability that shadowing will be detected in W at a given t_c and τ_c given that shadowing has been detected in D, and vice versa.

has little effect on $P(D)$. We believe that averaging has a large effect on W because of the amount of information that is removed when representing a KS state in the space of persistence diagrams. For a given state $u(x,t)$, there tend to be only two to four persistence points across both sublevel and superlevel persistence diagrams, resulting in relatively low-dimensional dynamics (Figure 3.6). Averaging in time allows temporal information to feature more prominently in the W metric. This is less of an issue in D because the full state $u(x,t)$ is used for analysis at a given t. However, for the sake of comparison of D and W in this work we have applied the same averaging scheme (Equation 3.8 with $w = 4.0$) to both metrics. We chose the value $w = 4.0$ by trial and error because it produces good agreement between the shadowing events detected in D and W (Figure 3.11). This value is consistent with other work on the subject of delay embedding which found that averaging can be done over times on the order of the correlation time ($t_{corr} = 8.6$ in our case) to capture the dynamics of trajectories through a turbulent attractor [89]. Figure 3.11 demonstrates D and W detect similar shadowing events when averaging is done with $w = 4.0$. The outlier seen in Figure 3.11 corresponds to RPO 2, which has a relatively small period.

Given the relatively large size of $P(S)$ shown in Figure 3.10, one might notice that it is possible, even common, for turbulence to shadow multiple RPOs at once. At any point in time on the turbulent trajectory, between 2 and 3 RPOs are being shadowed, on average. It is common for many RPOs to exist in state space near relative equilibria. Such regions tend to be dense with RPOs moving in a similar direction in the neighborhood of equilibria. Thus, when turbulence is shadowing a RPO in such regions, it tends to shadow multiple RPOs simultaneously. When turbulent trajectories approach an equilibrium, their evolution through state space tends to slow. Accordingly, we observe in our data set that the rate of evolution of turbulence through state space is negatively correlated with the number of RPOs being shadowed at that time. This will be discussed further in section 3.6.

The fact that the persistence space metric W can quantify similar shadowing events to D while maintaining a computational advantage in KS demonstrates its utility. We anticipate that it will maintain its advantage in higher dimensional systems with continuous symmetries due to its scaling with the complexity of the patterns involved, not the total number of grid points that make up the data set.

3.6 Shadowing of equilibrium

There are certain regions of state space that are visited often by turbulence and RPOs. Often, there is a particularly dynamically relevant ECS in this area structuring the dynamics. Many of the RPOs in our set are visited simultaneously by turbulence in an area of state space in which an unstable equilibrium solution resides. This equilibrium is dominated by the second Fourier mode (this solution, E_2 is plotted in Figure 3(b) of Cvitanovic *et al.* [60]). In this section, I will discuss the relationship between shadowing of RPOs and close passes to this equilibrium solution.

In Figure 3.10, it is demonstrated that every RPO in the analysis set is shadowed for between 10%-20% of the turbulent trajectory's duration. Given that there are 16 RPOs in the set, it is clear that multiple RPOs must be shadowed at once for at least some portion

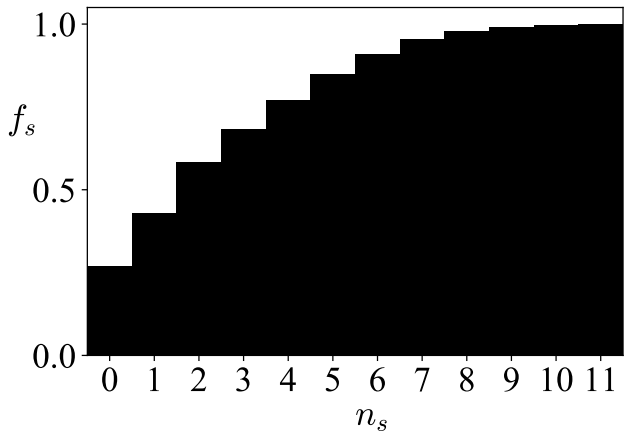

Figure 3.12: Cumulative distribution function f_s of the number of RPOs being shadowed simultaneously n_s shows that it is common for RPOs to be shadowed simultaneously. A time step t of the turbulent trajectory is counted in a given n_s bin if the number of RPOs being shadowed in the metric D at t is less than or equal to n_s.

of the turbulent trajectory. Figure 3.12 shows this is indeed the case. About 50% of the duration of the turbulent trajectory, two or more RPOs are being shadowed ($n_s \geq 2$) simultaneously while about 25% of the time, no RPOs are being shadowed at all ($n_s = 0$). There are instances where up to 11 RPOs are being shadowed simultaneously ($n_s = 11$). From this, we conclude that there are some regions of state space in which the RPOs, or at least portions of those RPOs, are densely packed.

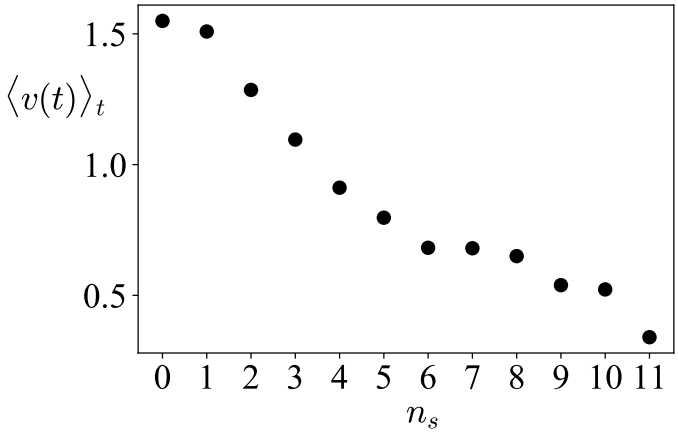

Figure 3.13: The average state space speed $\langle v(t) \rangle_t$ is slower when more RPOs are being shadowed n_s.

When the turbulent trajectory is in the densely packed region of state space, its rate of

evolution slows down (Figure 3.13) and it is close to E_2 on average. The *state space speed* of the turbulent trajectory at time t is defined as

$$v(t) = \frac{||u(t+dt) - u(t)||_2}{dt}, \tag{3.12}$$

where $dt = 0.2$ is the time step. This metric quantifies the rate at which the turbulent state is changing at a given time. Figure 3.13 shows that the average state space speed $\langle v(t) \rangle_t$ is negatively correlated with n_s. The distance to the equilibrium solution E_2 in the D metric can be defined as

$$D_{E_2}(t) = \min_{\tilde{s} \in [0,L]} D_r^{-1} ||u_2(x + \tilde{s}, t) - E_2(x)||_2, \tag{3.13}$$

which is similar to Equation 3.3. D_r is the estimate of the attractor radius given in section 3.2 and $\tilde{s}$ are possible shifts that minimize the distance between the turbulence at a given time $u_2(x,t)$ and $E_2(x)$. The average distance to the equilibrium $\langle D_{E_2}(t) \rangle_t$ is negatively correlated to n_s (Figure 3.14).

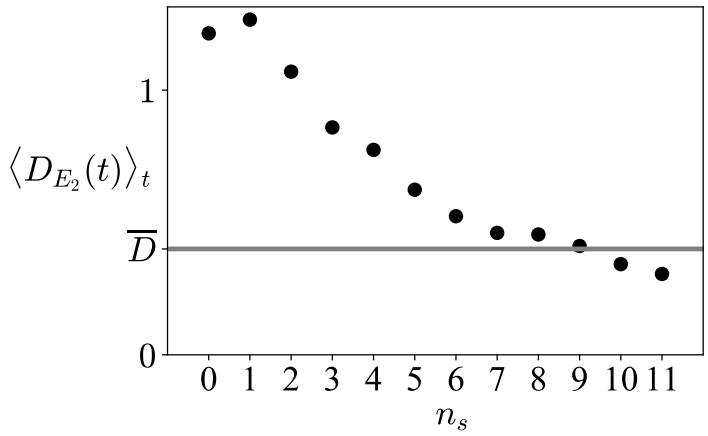

Figure 3.14: The average distance of the turbulent trajectory to the equilibrium $\langle D_{E_2}(t) \rangle_t$ is lower when there are more RPOs being shadowed n_s. When $n_s = 10$ and $n_s = 11$, $\langle D_{E_2}(t) \rangle_t$ is below $\overline{D}$, and is considered "close", on average.

It is typical for turbulent trajectories, as they approach an unstable equilibrium, to "slow down", i.e. change at a slower rate. The negative correlation of both $\langle v(t) \rangle_t$ and $\langle D_{E_2}(t) \rangle_t$ to n_s is thus expected. Taken together, these metrics demonstrate that a region of state space

consists of an unstable equilibrium and many RPOs pass close to this equilibrium. When a turbulent trajectory is in this region of state space, its evolution slows down as it shadows multiple RPOs passing through the region and passes close to the unstable equilibrium.

3.7 Summary of persistence space shadowing detection and next steps

We offered statistics of shadowing events in both state space and the space of persistence diagrams. We have found that the mean shadowing time is positively correlated to the linear stability of the RPOs, which is not expected in principle because turbulence shadows RPOs well away from the neighborhood where the linear manifold dominates the dynamics. We also found that shadowing of this set of RPOs is very frequent. For a given RPO in the set, turbulence shadows it almost 20% of the time, on average. The ubiquity of shadowing in KS, even outside of regions in state space where linear behavior is dominant, provides supporting evidence of Hopf's vision of turbulence as a deterministic walk among characteristic patterns in state space.

In this work, we sought to determine if shadowing of RPOs in KS could be detected in the space of persistence diagrams that is invariant to the system's continuous translational symmetry. We also sought to determine if such detection was more computationally efficient than detection of shadowing in state space. We answer both of these questions in the affirmative. We show that the vast majority of shadowing events that are detected in state space can be detected with similar statistics in the space of persistence diagrams, and vice versa. We demonstrated that the computation in the space of persistence diagrams is faster for large data sets of the size needed to gather statistics of shadowing events. We anticipate that the ability to detect shadowing of ECS in the space of persistence diagrams is not particular to the KS equation and that this method generalizes to higher dimensional systems. In particular, we anticipate that detection of shadowing systems with continuous symmetries, such as laboratory experiments like Taylor-Couette flow and pipe flow, will see gains in computational efficiency and operational simplicity using persistent homology.

CHAPTER 4

QUANTIFYING PLUME STATISTICS IN RAYLEIGH-BÉNARD CONVECTION

4.1 Connecting plumes to dynamics in Rayleigh-Bénard convection experiments and simulations

Complex spatiotemporal behavior is omnipresent in nature, so it is helpful to focus on relatively simple paradigm problems to gain insights into such behavior [1]. Spatiotemporal complexity occurs in biological [2, 90], chemical [91], optical [4] and a variety of fluid systems [5, 6, 8], but in this work we focus on spatiotemporal complexity in convective phenomena. Convection is important in many physical systems that feature a temperature gradient in a fluid, such as the internal structure of stars and planets, the atmosphere, and large bodies of water like the oceans [92, 93]. Because of the complexity of natural convective phenomena, many scientists study select paradigm problems, such as Rayleigh-Bénard convection (RBC) [32, 33], that are relatively simple to describe but still display features of interest. However, RBC can exhibit fully turbulent behavior which can be difficult to analyze [94]. Thus, we focus on spiral defect chaos (SDC), a spatiotemporally complex, moderate Rayleigh number (Ra $\sim 4\times10^3$) regime in which a time-evolving pattern of convection rolls pinch-off, merge, and form topological defects [62]. The Rayleigh-Bénard problem in SDC is amenable to study in both high-resolution experiments and realistic numerical simulations [11, 14].

Even in paradigm problems such as RBC undergoing SDC, it remains challenging to connect simulations and experiments. Theoretical descriptions, such as the Boussinesq equations in RBC [14], provide fundamental insights into the physics of convective phenomena. Numerical simulations of the Boussinesq equations allow for the computation of physical quantities that are often inaccessible to experiments because they provide a com-

plete physical description of the flow field [95, 77, 96, 97]. Experiments can be connected to simulations by comparing observable features that are common to both settings. Many analytical tools have been developed that quantify observable features in SDC, such as computations of the local wave number [34], local wave directors [71], the fractal dimension [76], the Karhunen-Loève dimension [98, 35], and other quantities of interest [68]. However, when descriptions are approximate, like in the case of the Boussinesq equations, their numerical simulations can produce observable features that differ from those of experiments because they are not capturing key experimental behaviors [99, 100]. The Boussinesq equations are based on the assumption temperature-independent fluid parameters, which we do not expect to be satisfied in our experiments. Also, numerical simulations require imposed temperature boundary conditions at the fluid layer that may not match experiments. We seek to overcome these difficulties and develop a class of descriptions that capture key experimental behaviors and produce observable features that are similar to experiments.

Persistent homology is a novel tool of topological data analysis that can be used to quantify observable features in both simulations and experiments [39]. Topological structures, such as those detected by computational homology [37], observed in convection roll patterns of SDC simulations and experiments have provided insights into convective phenomena [45, 46]. Computational homology can detect and quantify basic topological structures in convection roll patterns, however persistent homology also provides a natural organization to those structures based on their prominence [47, 36]. This allows persistent homology to differentiate between distinct observable features in convection roll patterns.

Plumes, which can be detected by persistent homology, are a notable feature of convective flows with direct connections to dynamics that drive spatiotemporally complex behavior. They are spatially-localized regions of rising fluid that are hotter than the surrounding fluid, or spatially-localized regions of falling fluid that are cooler than the surrounding fluid. In turbulent RBC ($\mathrm{Ra} \gtrsim 10^4$), the regime in which plumes are mostly studied, they appear

as mushroom-like shapes when imaged from the side [101, 102] and have a sheet-like form when viewed from above [103, 104]. When viewed from above in SDC, both cold and hot plumes appear as small circles at the center of an emerging target pattern. Plumes play a significant role in driving large-scale flow in turbulent RBC [94, 105, 106], and recent research by Levanger *et al.* [36] has demonstrated their connection to the leading Lyapunov vector in SDC. The leading Lyapunov vector is the eigenvector partner of the largest Lyapunov exponent and thus indicates the direction in state space of most rapid growth for a perturbation. Lyapunov exponents quantify the linear stability of dynamical trajectories and have been studied extensively in model chaotic systems [73, 23, 78, 95]. Perturbations grow rapidly in spatial regions of high Lyapunov vector magnitude, so such regions may have a significant impact on the evolution of the overall pattern dynamics in convection simulations [77]. Levanger *et al.* [36] used persistent homology to identify plumes as emerging targets (see Fig. 8 of [36]) located in spatial regions of Boussinesq simulations where the leading Lyapunov vector magnitude at a given time was significant on average.

Our long-term vision is to use plumes as an indicator of dynamical activity in experiments, where the leading Lyapunov vector is inaccessible. We take the first step toward achieving this vision in this work by using plume statistics collected by persistent homology to assess whether observable behaviors of plumes in experiments are quantitatively captured by a Boussinesq description with spatially uniform time-independent temperature boundary conditions. We identify that this description produces plume statistics that are significantly different from experiments because it does not accurately describe the thermal boundary conditions of the fluid layer and ignores deviations from the Boussinesq approximation that are significant in our experiments. We conduct numerical simulations of a Boussinesq description with radially-varying time-independent temperature boundary conditions that more closely matches experimental measurements.

This chapter is structured as follows. In section 4.2, we discuss RBC, the Boussinesq equations, the relevant parameters that characterize the dynamics of our experiments and

simulations, procedures of how experiments and simulations were conducted, and how convection roll patterns were prepared for analysis. In section 4.3, computational homology (subsection 4.3.1) and persistent homology (subsection 4.3.2) are introduced and we describe how they were used to collect plume statistics in experiments and simulations (subsection 4.3.3). In section 4.4, a conduction model of the experimental cell geometry that yields radially-varying time-independent temperature boundary conditions for use in Boussinesq simulations is described and details of specific simulations are given. In section 4.5, plume statistics of various simulations are compared and contrasted to experiments at similar parameter values and these results are discussed. In section 4.6, we offer concluding remarks.

4.2 Rayleigh-Bénard convection experiments and simulations

RBC is the buoyancy-driven flow of a fluid between two horizontal plates [1]. The bottom plate is heated relative to the top plate. When the temperature difference between the plates is large enough, convective flow occurs. For a Boussinesq fluid, RBC is described by the nondimensional equations

$$\mathrm{Pr}^{-1}\left(\frac{\partial \boldsymbol{u}}{\partial t}+\boldsymbol{u}\cdot\nabla\boldsymbol{u}\right) = -\nabla p+\nabla^2\boldsymbol{u}+\mathrm{Ra}T\hat{z} \tag{4.1}$$

$$\frac{\partial T}{\partial t}+\boldsymbol{u}\cdot\nabla T = \nabla^2 T \tag{4.2}$$

$$\nabla\cdot\boldsymbol{u}=0 \tag{4.3}$$

which represent conservation of momentum, energy, and mass for an incompressible fluid, respectively. In these non-dimensional equations, $\boldsymbol{u}=\boldsymbol{u}(x,y,z,t)$ is the velocity field, $p=p(x,y,z,t)$ is the pressure field, and $T=T(x,y,z,t)$ is the temperature field where t is time, x and y are the lateral spatial coordinates, and z is the spatial coordinate opposing the direction of gravity. The strength of the thermal driving is quantified by the Rayleigh number $\mathrm{Ra}=\frac{\beta g\Delta T d^3}{\nu\kappa}$ where β is the thermal expansion coefficient, g is accel-

eration due to gravity, ΔT is the temperature difference between the heating and cooling heat baths, d is the separation between the bottom and top plate, or depth, ν is the kinematic viscosity, and κ is the thermal diffusivity. The Prandtl number is given by $\mathrm{Pr} = \frac{\nu}{\kappa}$. Equation 4.1-Equation 4.3 are nondimensionalized using d as the length scale, ΔT as the temperature scale, and the vertical diffusion time for heat $t_v = d^2/\kappa$ as the time scale. In the case where the temperatures at the interfaces between the convecting fluid and the bottom and top plates are constant and spatially uniform this yields $T(x, y, z = 0, t) = 1$ and $T(x, y, z = 1, t) = 0$. Equation 4.1-Equation 4.3 are an approximation that assumes the fluid properties contained in Ra and Pr are constant and do not depend on the temperature except for a buoyancy term in the momentum equation, Equation 4.1, where the temperature dependence of the density is included as $\mathrm{Ra}T\hat{z}$.

Physically, t_v corresponds to the time it takes for heat to diffuse from the bottom of the cell to the top and serves as an intuitive time scale for analyzing convection. We will also find it useful to refer to the horizontal diffusion time for heat which is defined as $t_h = \Gamma^2 t_v = r_0^2/\kappa$ where r_0 is the radius of the cylindrical convection domain whose axis $\hat{z}$ is opposite to the direction of gravity. The aspect ratio $\Gamma = \frac{r_0}{d}$ describes the domain's lateral extent relative to d. t_h quantifies how long in time it takes for heat to diffuse from the center of the domain to a lateral sidewall and serves as a good time scale for estimating the duration of time required for initial transients to decay significantly.

In the laterally infinite case with time-independent temperature boundary conditions at the bottom and top plates, the onset of convective flow occurs at the critical Rayleigh number $\mathrm{Ra} \approx 1708$ [63]. As Ra is increased well past onset, the convection rolls become unstable and move relative to each other, allowing topological defects to form where rolls combine and new rolls are formed. At sufficient Ra, the flow will be in the regime of SDC, a spatiotemporally chaotic state that consists of convection roll stripes composing a time-evolving pattern of topological defects (Figure 4.1) [62, 107]. We perform convection simulations and experiments in this SDC regime, with parameter values of $t_v \approx 2$ s, $\mathrm{Pr} =$

0.84, $\Gamma \approx 20$, and $3500 \leq \mathrm{Ra} \leq 4800$ (Table 4.1 and Table 4.3). Below, we describe how simulations are performed, discuss experimental procedures, and detail how images from both experiments and simulations are prepared for analysis.

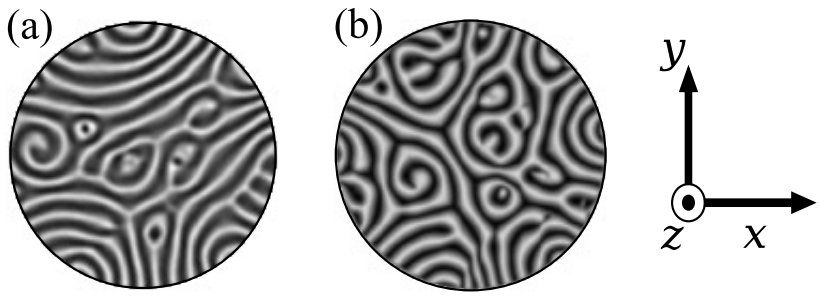

Figure 4.1: RBC produces spatiotemporally complex convection roll patterns in SDC that are difficult to characterize. (a) A processed shadowgraph image from experiment at $\mathrm{Ra} = 4000$ (Table 4.1) with aspect ratio $\Gamma = 19.6$. (b) The mid-depth temperature field $T(x, y, z = 1/2)$ from numerical simulation of Equation 4.1-Equation 4.3 with spatially uniform temperature boundary conditions at $\mathrm{Ra} = 4000$ with aspect ratio $\Gamma = 20$ (Table 4.3). Dark regions represent hot upflows and bright regions represent cool downflows. A single two-dimensional image in the x-y plane is analyzed at each time step for both experiments and simulations. The z-axis, displayed on the right, points out of the page, opposite of gravity.

We numerically integrate Equation 4.1-Equation 4.3 using a parallel spectral element approach [108, 109] that has been used extensively to study open questions regarding fluid convection. We use the no-slip velocity boundary condition for the fluid flow on all material surfaces and we use a perfectly conducting sidewall temperature boundary condition. When conducting simulations with spatially uniform temperature boundary conditions at the lower and upper plates, we integrate Equation 4.1-Equation 4.3 using Ra based upon ΔT. In the course of our investigation we will also numerically simulate RBC for the situation where the temperatures of the fluid-solid interfaces at the bottom and top plates, in contact with the convecting fluid layer, are not constant, and are instead radially-varying functions that we will specify. In this case, we numerically simulate the Boussinesq equa-

tions in dimensional form, which are not shown here, where it is straightforward to implement the radially-varying boundary conditions. The initial condition is a quiescent flow field with zero fluid velocity and small random perturbations to the temperature field. We conduct long-time numerical simulations for 370 time units to allow for all initial transients to decay. Simulations were then continued for another 330 time units to generate a long time series of data for analysis and for comparison with experiment. More details about the application of the numerical approach to the study of SDC can be found in Refs. [36, 76, 110].

In order to clearly distinguish between our use of dimensional or nondimensional independent and dependent variables we will use a superscript * to indicate a dimensional variable. Parameters and constants may be dimensional or nondimensional, this will be specified when they are defined.

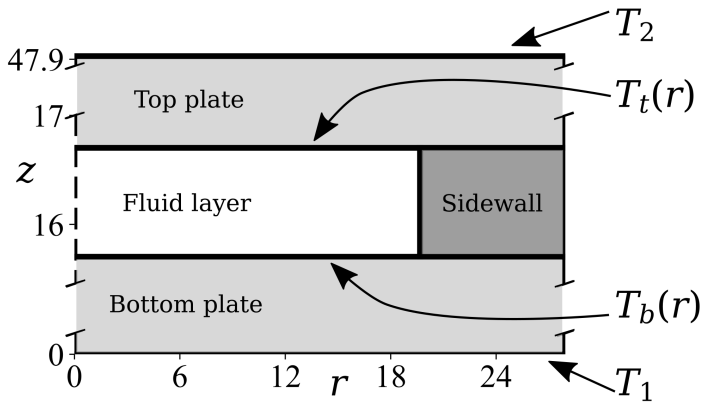

Figure 4.2: An axisymmetric diagram of the experimental domain. This geometry is used to model heat conduction through the experiment to obtain radially-varying temperature profiles at the fluid layer boundaries to use in simulations (section 4.4). The radial coordinate r and height z are nondimensionalized by the separation between the bottom and top plate, d. From the z-axis one can observe that the experimental geometry is not shown to scale. We solve Equation 4.5 with the constant temperature of the heat baths at the top and bottom boundaries (T_2 and T_1, respectively) on this domain to obtain radial temperature profiles $T_t(r)$ and $T_b(r)$, which are then used as temperature boundary conditions in separate convection simulations (Table 4.3). The boundary conditions at the $r = 0$ and $r \approx 28$ sides are zero-heat flux. The fluid layer lies in between the top and bottom plate from $r = 0$ to $r = \Gamma \approx 20$. The sidewall, which has greater thermal conductivity than the fluid, surrounds the fluid layer for $r > \Gamma$.

The top plate of our experimental convection cell is a transparent sapphire window and the bottom plate is a zinc selenide window (Figure 4.2). Between these plates is a polyethersulfone annulus with an inner radius $r_0 \approx 20d$ and outer radius $r_1 \approx 28d$ where $d \approx 600\mu$m. The center of this annulus is where the convection of sulfur hexafluoride gas is imaged. The cell is pressurized and the pressure is monitored using an in-line pressure sensor. The temperature is controlled by running water from heat baths through heating and cooling chambers positioned below the bottom plate and above the top plate, respectively. The heating and cooling chambers each contain a thermistor probe that monitors their temperature.

Table 4.1: Rayleigh-Bénard experiments were performed for a range of Ra and Q values in SDC. Ra is determined using the temperatures of the bottom and top heat baths $\Delta T = T_1 - T_2$. Q is defined in Equation 4.4. The aspect ratio is given by $\Gamma = \frac{r_0}{d}$ where r_0 is the cell radius and d is the depth. p^* is the dimensional pressure of the fluid layer. The effective Rayleigh number at the center of the cell Ra_c is determined using the radially-varying temperature at the center of the cell $\Delta T_c = T_b(r=0) - T_t(r=0)$ that was computed by the conduction model described in section 4.4. The effective Rayleigh number at the edge of the cell Ra_e is determined using the radially-varying temperature near the edge of the cell is $\Delta T_e = T_b(r=19.5) - T_t(r=19.5)$ also computed using the conduction model (Figure 4.2). $T_0 = (T_1 + T_2)/2$ is the mean temperature of the warm and cool heat baths. $\mathrm{Pr} = 0.84$ for all experiments.

Exp.	Ra	Q	Γ	p^* (MPa)	T_0 (K)	ΔT (K)	Ra_c	Ra_e
1	3500	0.83	19.5	1.51	294.27	5.01	3200	3000
2	3800	0.90	19.5	1.50	294.54	5.54	3400	3100
3	4000	0.96	19.6	1.51	294.64	5.82	3600	3300
4	4100	0.97	19.6	1.51	294.55	5.90	3600	3400
5	4200	1.01	19.5	1.50	294.89	6.22	3800	3500
6	4300	1.02	19.5	1.51	294.89	6.22	3900	3600
7	4400	1.05	19.6	1.51	294.79	6.38	4000	3700
8	4800	1.17	19.5	1.50	295.38	7.19	4300	4000
9	4800	1.19	19.6	1.50	295.28	7.34	4400	4100

We characterize the degree to which our experiments are expected to satisfy the Boussinesq approximation. The extent of the deviation from the Boussinesq approximation by a fluid, or its "non-Boussinesqness", can be quantified using the Busse parameter Q which is

defined as [111]

$$Q = \sum_{i=0}^{4} \gamma_i B_i \tag{4.4}$$

where $\gamma_0 = -\frac{\rho_b - \rho_t}{\rho_m}$, $\gamma_1 = \frac{\beta_b \rho_b - \beta_t \rho_t}{\beta_m \rho_m}$, $\gamma_2 = \frac{\nu_b - \nu_t}{\nu_m}$, $\gamma_3 = \frac{\lambda_b - \lambda_t}{\lambda_m}$, $\gamma_4 = \frac{c_{p_b} - c_{p_t}}{c_{p_m}}$, λ is the thermal conductivity, and c_p is the specific heat at constant pressure. The subscripts b, t, and m, denote that the given fluid property is evaluated at the bottom, top, or mean temperature of the cell, respectively. The heating and cooling bath temperatures, T_1 and T_2, are used to determine the fluid parameters at the bottom and top of the cell, respectively. The mean temperature is then defined as $T_0 = (T_1 + T_2)/2$. The B_i terms are linear functions of Pr^{-1} and are available in Table 2 of [111]. Larger difference between the fluid parameters at the top and bottom of the cell will result in larger Q, which implies that there are larger deviations from the Boussinesq approximation.

Nine experiments with different Ra and Q are analyzed (see Table 4.1). Experiments over the parameter range $3500 \leq \mathrm{Ra} \leq 4800$ were chosen because at the lower end of this range, hardly any plumes are produced in experiments and Q is low (Figure 4.10). At the upper end of this range, significant deviations from the Boussinesq approximation (high Q) can be observed in the plume statistics. Fluid parameters used to determine Ra and Q were computed using the Santa Barbara Pattern Formation and Turbulence Group software package for convection fluid properties [68]. We increased Q in experiments by increasing the temperature of the bottom heat bath T_1 while holding other parameters constant. This increased both the mean temperature T_0 and the temperature difference ΔT, thus increasing the size of non-Boussinesq effects (Table 4.1). Further information on physical properties of materials used in the experiments can be found in Appendix B.

Experiments were imaged with a shadowgraph technique [11], where a beam of collimated light is passed through the flow and then collected by a camera. Temperature variations in the fluid cause changes in the refractive index, so the intensity of light captured by the camera will vary depending on the local temperature. This allows us to differentiate between hot upflows and cold downflows in the fluid. For our cylindrical convection cell,

the collected images are two-dimensional circles with convection patterns represented by intensities in 8-bit gray scale values (see Figure 4.1(a)).

Ten frames per t_v are extracted from the video data. Images are pre-processed in several steps before they are analyzed using computational homology. In our experiments, heat baths pump water through the heating and cooling chambers below and above the cell for temperature control. This leads to slight mechanical vibrations in our captured images due to the flow of temperature control fluid. To account for this effect, images are spatially shifted to line up against a background image using a Fourier method. Bubbles in the temperature control fluid, refractive index fluctuations, and mechanical vibrations can all cause temporal noise. So, each image is time-averaged with the three preceding and successive frames in order to smooth out the images in time. The time over which the images are averaged is much smaller than the time scale on which the pattern changes. Then, the images are normalized in gray scale value with respect to the global maximum gray scale value measured in the data set in space and time. The background image is then subtracted off to account for static visual imperfections in the cell. The gray scale values of the images are shifted and linearly scaled such that the global minimum is set to 0 and the global maximum is set to 255. Finally, a Gaussian smoothing kernel with a standard deviation of 3.5 is applied to the image to smooth out spatial noise at small scales.

The analysis of experiments is performed by examining time-series of convection patterns captured in shadowgraph images. The temperature field from convection simulations is taken at mid-depth in 8-bit gray scale intensity values so that the same methods of analysis can be applied to both simulations and experiments (see Figure 4.1(b)). In images of both experiments and simulations, darker regions represent hot upflows while brighter regions represent cool downflows. In large aspect ratio SDC, convection rolls, when viewed from above, form a pattern containing many defects such as spirals, targets, dislocations, and others [62]. We analyze these patterns in experiments and simulations using computational homology.

4.3 Detection of plumes using persistent homology

4.3.1 Computational homology of Rayleigh-Bénard image data

We are interested in characterizing the geometry of deformable time dependent structures within two-dimensional images, e.g. Figure 4.1, which suggests that an appropriate mathematical language is that of topology. To provide quantification based on large sets of images we make use of techniques from computational homology [37]. Homology is the most computationally accessible subdomain of algebraic topology and, as employed in this paper, identifies geometric features of a space in the language of linear algebra. While homology can be applied to spaces, such as point clouds, of arbitrary dimension we are interested in planar regions observed in image data. Hence, without loss of generality, we restrict our attention to the zeroth and first homology groups, H_0 and H_1, and more specifically their dimensions β_0 and β_1, which are referred to as the zeroth and first $Betti$ numbers. In particular, given a subset of the plane X, $\beta_0(X)$ equals the number of distinct connected components of X and $\beta_1(X)$ equals the number of holes in X.

We employ a two step procedure to extract the desired homological information from an image such as that of Figure 4.1. First, we arbitrarily threshold the data to compute homology and second, we use persistent homology (see the next section) to organize the homological data over all thresholds. Thus, beginning with an image of a convection roll pattern such as Figure 4.1 we pick a threshold intensity value A and separate all the pixels in the image based on whether they are above or below the threshold value. For example, if A is chosen to be the mean intensity value of Figure 4.1 and all pixels with threshold value less than or equal to A are drawn in black and all pixels with threshold value greater than A are drawn in white, then one obtains the image shown in Figure 4.3 (this is the convention adopted in [46] to separate the cold and hot regions of the flow). We magnify a region of the filtered field in the inset of Figure 4.3 to emphasize a plume, however in the course of our analysis we do not divide the field into sub-regions. We use homology to analyze

the entire temperature field at once. Declaring X to be the set of black pixels allows us to compute $\beta_0(X)$ and $\beta_1(X)$ [37]. More generally, this is an example of what is referred to as *sublevel filtering*, given a threshold value the space of interest is defined by the pixels with values less than or equal to the threshold value. *Superlevel filtering* is the opposite: pixels greater than or equal to the threshold define the region on which homology is computed. This filtering of image data by its gray scale value is different than many other applications of homology in which point cloud data is filtered by the distance between points.

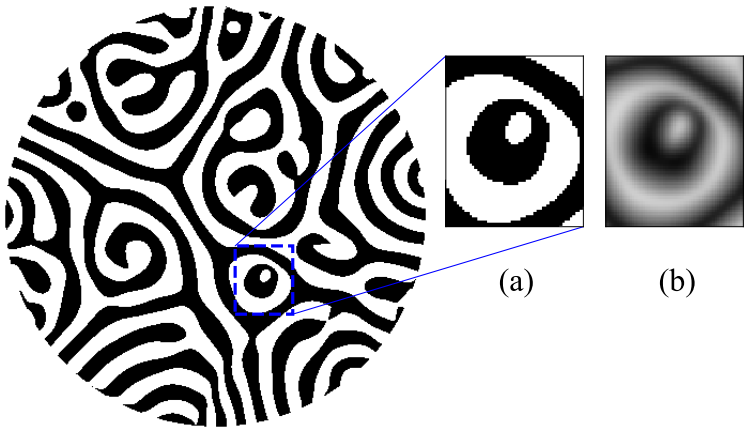

Figure 4.3: Topological structures can be extracted from convection roll pattern image data from simulations and experiments that is filtered into two binary sets. (a) shows a sublevel filtration of the simulated mid-depth temperature field in Figure 4.1(b) at the mean intensity value $A = 127$. The inset of (a) zooms in on a specific plume in the convection roll pattern. (b) shows the same zoomed-in area unfiltered, in gray scale. Black pixels are included as members of the surface while white pixels are treated as empty space. Each distinct black shape in the image is a distinct connected component. In the inset, we zoom-in on a plume, which appears as a hole in the filtered surface. We further analyze this zoomed-in region at different filtering thresholds in Figure 4.4 and Figure 4.5 for illustrative purposes. The data analysis is performed over the entire temperature field, not in specific spatial regions.

In summary, once an image is filtered, the Betti numbers provide us with the number of components and holes. In Figure 4.3, $\beta_0 = 28$ and $\beta_1 = 4$.We mention the Betti numbers β_0 and β_1 in this case because they provide useful topological information about the convective state explored in past work [28] and help illustrate how computational homology works. However, we use persistence as opposed to the Betti numbers for analysis in this work. An example of a distinct connected component is the black annulus shape in the middle of

the inset of Figure 4.3(a) and an example of a hole is the patch of white embedded in the previously mentioned black annulus shape in the inset.

If, when computing homology of convection rolls in images from experiments or simulations, one wishes to restrict to a single threshold, then it is intuitive to select the mean intensity value of the pattern as a threshold value. Sublevel filtering at the mean value picks out the regions of hot, rising upflow, while superlevel filtering picks out the regions of cool, falling downflow [45, 112]. In Figure 4.3, sublevel filtering has picked out the dark regions of hot upflow, which are shown as black pixels. However, as is discussed in the next section it is possible to vary the computations over all possible thresholds and organize this information in a meaningful manner.

4.3.2 Extension to persistent homology

At each possible intensity threshold A, a different surface of black pixels is generated and computational homology detects the distinct connected components and holes of that surface. As A is varied, components and holes may appear or disappear. The value of A at which a structure first appears is the value at which it is born A_b and a structure dies at the value at which it disappears A_d. Note that if two components merge, then one component must disappear. The (necessary) convention is that the component that first appeared later disappears [39]. In this study, we vary the filtering threshold over the range of possible intensity 8-bit integer values, $0 \leq A \leq 255$ (Figure 4.4) [47]. For an example involving holes, see the white hole at the center of Figure 4.4(b). In Figure 4.4(a), this hole has not yet been born since the surrounding black region is not closed. In Figure 4.4(c), this hole has died because it has been completely filled in with black pixels. Figure 4.4 contains a similar example for distinct connected components. In Figure 4.4(a), four distinct connected components, a partial black annulus in the center and three surrounding stripes, have already been born at some lower values of $A < 80$. In Figure 4.4(b), the three surrounding stripes have merged and there are two black distinct connected components, the inner

annulus shape and a surrounding black ring. Thus, at some values of A between $A = 80$ and $A = 127$, two distinct connected components died when they merged with another. In Figure 4.4(c), at a higher value of $A = 208$, the remaining two distinct connected components are now connected, so they have merged. The birth and death intensity value (A_b and A_d, respectively) of each topological structure of the image can be plotted in persistence diagrams (Figure 4.5).

Figure 4.4: Topological structures from convection roll pattern image data are born and die at different filtering thresholds. Here, we show the zoomed-in region from Figure 4.3(b) sublevel filtered at threshold A values of (a) 80, (b) 127, and (c) 208. In (a) there are no holes. There is one prominent hole in (b) because the surrounding black ring has been partially cropped out for this example. In (c), there are three holes associated with features of the temperature field that are not very prominent. The prominent hole in (b) has been filled in (c) and has thus died.

A structure for which the difference between A_d and A_b is large, i.e. a point that lies far from the diagonal in the persistence diagram, persists over a large range of filtrations and thus can be interpreted as being a prominent topological feature. In contrast, a feature whose persistence point lies close to the diagonal is often associated with image noise. In the example of Figure 4.4(a), there are four prominent distinct connected components: the central black shape and three black stripes surrounding it. These four components give rise to the points far from the diagonal in the 0-dimensional persistence diagram in Figure 4.5(a). There is one prominent hole in the image enclosed by the central black shape. The persistence point for this hole is shown as an open circle in Figure 4.5(b). The holes that appear in Figure 4.4(c) give rise to the persistence points that are close to the diagonal.

We apply persistent homology to the intensity fields obtained from simulated and pro-

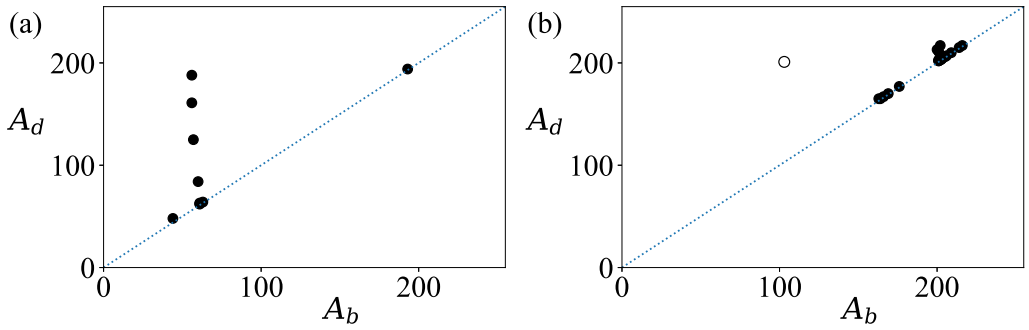

Figure 4.5: Persistence diagrams quantify the birth and death filtering thresholds of topological structures in Figure 4.4 and indicate which structures are most prominent at a single time step. In (a), the persistence diagram of distinct connected components for the zoomed-in pattern shown in the inset of Figure 4.3(b) is shown. In (b), the persistence diagram of holes is shown for the same pattern. Each point represents a distinct topological structure and its position reflects the threshold values A at which it is born and dies. The most prominent topological structures of both types lie far from the dotted diagonal. In (b), the open circle corresponds to the large hole in the inset of Figure 4.3 and shown in Figure 4.4(b).

cessed experimental images using topological image data analysis software developed by Levanger and Harker [113] to generate the sub- and superlevel persistence diagrams of every time snapshot of image data.

4.3.3 Detection of plumes in persistence diagrams

Cold and hot plumes appear as the center of a target pattern in convection roll patterns (Figure 4.6(a)). A cold plume, for example, is a small region of cold fluid completely surrounded by hot fluid. Thus, a cold plume can be detected by persistent homology as a hole of a certain prominence in a hot fluid region. Hot plumes can be detected in a similar way: as a hole of a certain prominence in a cold fluid region. Though the persistence diagrams for distinct connected components contain valuable information on the topological structure of the temperature field [45, 46], we only need to consider the persistence diagrams for holes to detect plumes. Sublevel filtering includes hot regions of the flow as part of the surface of which homology is computed, so cold plumes are detected as holes in the sublevel persistence diagrams. Likewise, superlevel filtering includes cold regions as parts

of the surface, so hot plumes are detected as holes in the superlevel persistence diagrams. There are other image processing methods that could detect plumes, but we note that the essential information for plume identification is conveniently given in the dimension-1 persistence diagrams. Holes are, by definition, completely surrounded by fluid of the opposite type (hot or cold) and the birth and death thresholds in the persistence diagrams allow us to pick out structures of a certain prominence with ease. We show an example of a sublevel persistence diagram for holes that is used to detect cold plumes in Figure 4.6(b). Note that the persistence diagrams for distinct connected components are not necessary for the detection of plumes, so we do not consider them in Figure 4.6. The superlevel persistence diagram for holes that is used to detect hot plumes is not shown. The plumes are found in the "plume region" of the persistence diagrams diagrams for holes. In the sublevel example for detecting cold plumes in Figure 4.6(b), The plume region is bounded from above to cut out holes that are too prominent, i.e. they die at high intensity values. It is bounded from the right at the median intensity value to exclude holes bounded by regions of hot upflow. Finally, it is bounded by a line that is parallel to the diagonal but offset by a small amount to exclude holes close to the diagonal that are associated with noise. These three bounds can be observed in Figure 4.6(b). Similar bounds are used in the superlevel persistence diagram for holes for the detection of hot plumes. The optimal values of these boundaries have to be chosen by inspecting how well they capture visible plumes in the image. The plume region has to be chosen separately for each experiment, since the exact distribution of intensity values in the pattern will not always be the same because of variability in optical conditions when preparing the shadowgraph imaging.

When a plume is identified in the plume region of a dimension-1 persistence diagram for a time snapshot of the temperature field, it is tracked in time as it moves continuously through the space of persistence diagrams so that plumes that exist throughout multiple frames are not double-counted. When a plume is identified, previous and future persistence diagrams are checked for components that are close to the plume in birth value, death value,

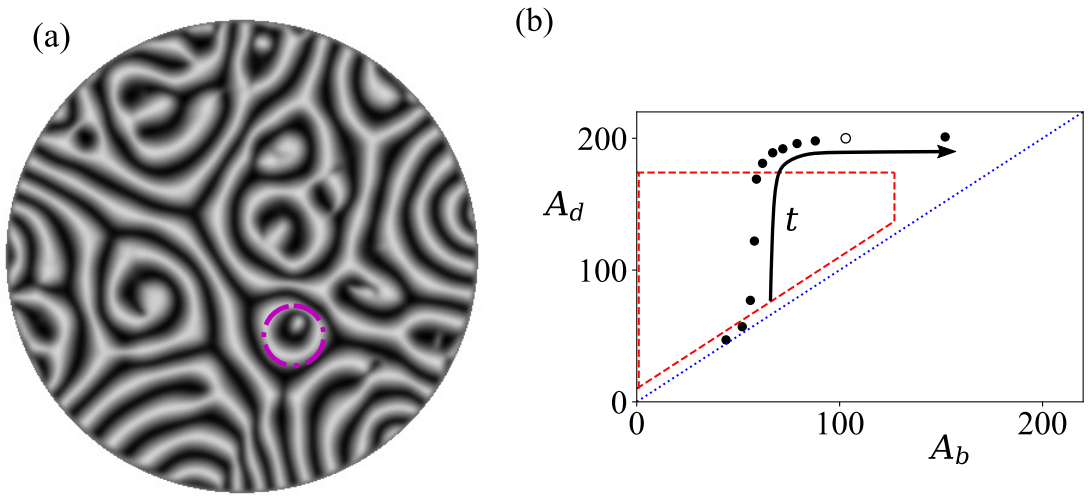

Figure 4.6: Convective plumes move continuously in time through physical space and the space of persistence diagrams, where they can be detected if they pass through the "plume region". In (a), a cold plume, highlighted with a dashed-dotted circle, is identified in a simulated temperature field. In (b), the same cold plume is plotted in the persistence diagram for holes at multiple time steps over an interval of approximately one time unit t_v in length. The open circle in (b) corresponds to the time step in (a) where the plume is highlighted. The plume begins close to the dotted diagonal with a birth value $A_b \approx 50$ and moves in time through the dashed plume region in (b), away from the diagonal, before it eventually returns to the diagonal.

and spatial coordinates. A plume can be tracked in time even before or after it falls within the plume region of the persistence diagram (Figure 4.6).

We consider plumes that appear in the range $0 \leq r \leq r_s$ where $r_s = 0.92\Gamma$ and the aspect ratio $\Gamma \lesssim 20$ in experiments and $\Gamma = 20$ in simulations (Appendix B). We throw out plumes that appear too close to the boundary because false positives often appear there, due to the presence of significantly damped convection rolls. A damped convection roll can appear as a plume because its decreased intensity can lead to lower birth and death values than other convection rolls, causing it to appear in the plume region of a persistence diagram.

After accounting for every individual plume in space and time, we can compute plume statistics. In this work, we compute plumes per time per unit area in bins with a duration of 30 time units. We refer to this quantity as the areal plume rate P. We also consider areal plume rates of cold plumes P_{cold} and hot plumes P_{hot} separately.

4.4 Heat conduction model of thermal boundary conditions

It is typical to determine Ra in experiment using the value of ΔT that is imposed on the system with temperature control devices, such as heating and cooling heat baths [35]. In Figure 4.2, the warm heat bath at T_1 is below the bottom plate and the cold heat bath at T_2 is above the top plate. By using the parameter $\Delta T = T_1 - T_2$, we assume that the bottom and top plates have an infinite thermal conductivity such that T_1 and T_2 can be used as the lower and upper boundary conditions for temperature of the fluid layer. In practice, the top and bottom plates have thermal conductivity values that are three orders of magnitude higher than the fluid layer (Table 4.2). In addition, the choice of $\Delta T = T_1 - T_2$ also assumes that there is no significant heat conduction between the bottom and top plates and the lateral sidewalls of the domain. As a result, there is no radial dependence for the bottom and top plates which are assumed to be of constant temperature. Both assumptions are not accurate in our experimental cell, and here we describe a heat conduction model that was introduced to more closely approximate the thermal conditions inside the experimental cell. Past work in experimental Rayleigh-Bénard convection has taken into account the finite thermal conductivity of the heating and cooling plates to compare observations to theoretical predictions [114]. Other experiments have extensively studied the effect of thermally conducting sidewalls on convective flows [115, 116, 96].

In order to obtain thermal boundary conditions of the fluid layer that account for the finite thermal conductivity of the top and bottom plates and sidewall, we determine the temperature variations in the experimental apparatus by solving the axisymmetric steady heat conduction problem given by

$$\nabla^2 T^*(r, z) = 0 \tag{4.5}$$

where $T^*(r, z)$ is the dimensional temperature field. The experimental cell has four regions with different values of thermal conductivity λ: the top plate, the bottom plate, the fluid,

and the sidewall (see Figure 4.2 and Table 4.2). We set the boundary conditions at the upper surface of the top plate and the lower surface of the bottom plate to the constant values of the temperature measured by probes submerged in the heat baths. At $r = 0$ we impose a zero-heat flux boundary condition because of the symmetry of the domain. At $r \approx 28$ we impose a zero-heat flux boundary condition because the outer edge of the sidewall is essentially an insulated boundary due to the air surrounding the experimental apparatus.

Equation 4.5 describes heat conduction in a material with fixed thermal conductivity and does not include heat transport due to convection. The Nusselt number, however, quantifies the ratio of convective to conductive heat transport and is defined as $\mathrm{Nu} = \langle -\frac{\partial T}{\partial z}|_{z=0,1} \rangle_{x,y}$ where z is expressed in units of the depth d and $\frac{\partial T}{\partial z}|_{z=0,1}$ is the partial derivative of the non-dimensional temperature field with respect to z, evaluated at either the bottom or top fluid layer boundary. The brackets $\langle \rangle_{x,y}$ denote that the quantity is spatially averaged over the lateral directions x and y. We use an effective thermal conductivity in the fluid layer that is approximated by multiplying the thermal conductivity λ of the fluid layer by Nu obtained from simulations of Equation 4.1-Equation 4.3 that use spatially uniform temperature boundary conditions $T(z=0)=1$ and $T(z=1)=0$.

We use a finite element method in MATLAB's PDE Toolbox to solve Equation 4.5 numerically. In the finite element method, the partial differential equation is linearly approximated on a triangular mesh. At the boundary conditions between materials with differing thermal conductivity, the temperature fields are matched to ensure continuity. Simple test cases are used to verify the model's accuracy in Appendix C.

From this model we obtain radially-varying dimensional temperature distributions at the top and bottom boundaries which we define as $T_t(r)$ and $T_b(r)$, respectively, where r is the nondimensional radial coordinate expressed in units of the depth d. We then fit these distributions with fourth-order polynomial functions (the expressions are given in Appendix B). These functions are then used as boundary conditions in dimensional spectral element simulations of RBC which we compare with experiment. Overall, by including

Table 4.2: Values of the thermal conductivity λ for the different materials used in the experimental apparatus shown in Figure 4.2 vary by up to three orders of magnitude. In conduction simulations, the thermal conductivity λ of the fluid layer is multiplied by $\mathrm{Nu} = 1.65$ in the $\mathrm{Ra} = 4000$ case and $\mathrm{Nu} = 1.76$ in the $\mathrm{Ra} = 4800$ case in order to approximate convective effects in the heat transfer through the fluid layer.

Region	$\lambda\ (\frac{\mathrm{W}}{\mathrm{mK}})$
Fluid layer	0.014
Top plate	35
Bottom plate	18
Sidewall	0.20

the effects of finite thermal conductivity, the temperature difference between the surfaces of the bottom and top plates is reduced when compared with the infinite thermal conductivity approximation. This results in a reduction of the effective Ra (Figure 4.7). At the center of the convection cell, the radially-varying $\Delta T_c = T_b(r=0) - T_t(r=0)$ is about 10% less than $\Delta T = T_1 - T_2$. Near the edge of the convection domain, the difference between ΔT_e and ΔT becomes as much as 16%. Thus, the effective Ra throughout the cell is 10% to 16% less than the Ra computed from $\Delta T = T_1 - T_2$ (Table 4.1 and Figure 4.7). Values of Ra_c computed at the center of the cell ($r=0$) and Ra_e computed near the edge ($r=19.5$) are listed in Table 4.1.

We have performed numerical simulations of the Boussinesq equations, using different boundary conditions at the bottom and top of the fluid layer at $\mathrm{Ra} = 4000$ and $\mathrm{Ra} = 4800$ to compare to experiments 3 and 8, respectively. Experiment 3 was chosen for comparison because at $\mathrm{Ra} = 4000$ experiments, plumes begin to be produced in significant numbers (Figure 4.10). Experiment 3 produces more cold than hot plumes, but their difference is relatively small. Experiment 8 was chosen because at this higher $\mathrm{Ra} = 4800$, deviations from the Boussinesq approximation become very prominent in the plume statistics and it produces a great deal more cold than hot plumes (Figure 4.10). In simulation 1, we use spatially uniform temperature boundary conditions for T_1 and T_2 at the bottom and top of the convection domain, respectively, with $\mathrm{Ra} = 4000$. In simulation 2, we use spatially uniform temperature boundary conditions for T_1 and T_2 with $\mathrm{Ra} = 3500$. In simulation

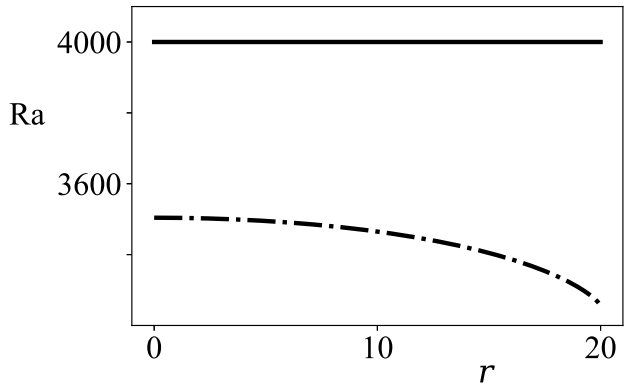

Figure 4.7: The radially-varying temperature profiles at the fluid layer boundaries computed using the cell conduction model (Figure 4.2 and Equation 4.5) lead to smaller effective driving throughout the cell that becomes even smaller near the sidewall. The solid line is Ra of experiment 3 computed using the temperatures of the heat bath $\Delta T = T_1 - T_2$. The dashed-dotted line is the effective Ra computed using the temperature computed by the conduction model at the fluid layer boundaries $\Delta T(r) = T_b(r) - T_t(r)$ as a function of the radial coordinate r. The value $r = 0$ corresponds to the center of the cell and $r = 20$ is the sidewall. The conduction model demonstrates that the effective driving in the experimental flow is lower than the heat bath settings would indicate and that the effective driving varies in the radial coordinate.

3, we use the radially-varying temperature boundary conditions $T_b(r)$ and $T_t(r)$ obtained from the conduction calculation with Ra $= 4000$. Simulations 1-3 correspond to experiment 3. In simulation 4, we use the radially-varying temperature boundary condition at Ra $= 4800$ (this corresponds to experiment 8). For simulations 3 and 4, the temperature boundary conditions are implemented using the fourth-order polynomial fits of the temperature profiles of the conduction simulation of Equation 4.5 (Figure 4.2). Simulations 3 and 4 are integrated using a dimensional form of the Boussinesq equations not shown here in order to incorporate the radially-varying thermal boundary conditions. The coefficients of the polynomial fits are given in Appendix B. Simulations 1-4 are summarized in Table 4.3.

4.5 Plume statistics

With the ability to track convective plumes in time using persistence diagrams, we looked at the areal plume rate P throughout the experiment at Ra $= 4000$ (Exp. 3) for a duration

Table 4.3: Simulations are conducted for various conditions for comparison with experiments 3 and 8. Ra_c is the Rayleigh number of the simulation at the center of the cell. The temperature boundary condition column indicates whether the simulation uses the temperature of the heat baths in the corresponding experiment T_1 and T_2 (spatially uniform) or the radially-varying temperature profiles $T_b(r)$ and $T_t(r)$ (radially-varying) produced by the conduction model (Figure 4.2) as temperature boundary conditions. The experiment column indicates the experiment that a given simulation is compared to. For simulations 3 and 4, the corresponding experiment supplied the heat bath temperatures T_1 and T_2 used by the conduction model to solve for the radially-varying temperature profiles at the fluid layer.

Simulation	Ra_c	Temperature boundary condition	Experiment
1	4000	Spatially uniform	3
2	3500	Spatially uniform	3
3	3600	Radially-varying	3
4	4400	Radially-varying	8

of more than $4t_h$ (one hour lab time or about 1600 time units) and compared it to the areal plume rate throughout simulation 3 (Table 4.3) of the Boussinesq equations at $\mathrm{Ra} = 4000$ for a duration of about t_h (Figure 4.8).

We see in Figure 4.8 that the simulation with radially-varying temperature boundary conditions (Sim. 3) and the corresponding experiment at $\mathrm{Ra} = 4000$ had similar average total areal plume rates. On average, one plume occurred every 1000 time units in one unit of area. The spatially uniform temperature boundary conditions $\mathrm{Ra} = 4000$ simulation (Sim. 1) produced more plumes on average than experiment at $\mathrm{Ra} = 4000$, while the spatially uniform thermal boundaries $\mathrm{Ra} = 3500$ simulation (Sim. 2) produced fewer plumes on average. The average areal plume rates indicate that the plume statistics of the $\mathrm{Ra} = 4000$ experiment are most closely matched by the simulation with radially-varying temperature boundary conditions (Sim. 3). However, in order to rigorously assess the significance of plume statistics, we examine cumulative distributions of areal plume rates in bins of 30 time units.

For all simulations and experiments we divided the time-series into bins with a duration

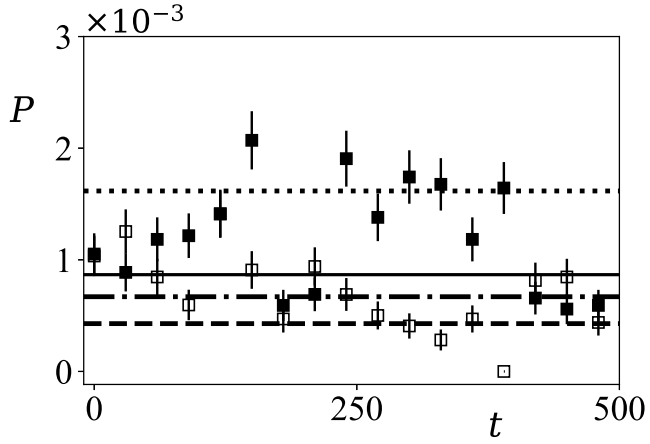

Figure 4.8: Areal plume rates measured in simulation and experiment vary throughout a time-series, but are similar on average when boundary conditions that resembled experimental conditions are used in simulations. Shown are areal plume rates in bins of 30 time units for the experiment at $\mathrm{Ra} = 4000$ (Exp. 3, closed squares) and the simulation which used radially-varying thermal boundary conditions (Sim. 3, open squares) for a portion of the whole time-series. The average areal plume rate throughout the whole time-series for experiment 3 and simulation 3 are represented as a solid and dashed-dotted line, respectively. The average total plume rates for spatially uniform temperature boundary simulations at $\mathrm{Ra} = 4000$ (Sim. 1, dotted line) and $\mathrm{Ra} = 3500$ (Sim. 2, dashed line) are also shown, but their values in each bin are not shown. Vertical bars are obtained by estimating the Poisson fluctuations $\sqrt{N}$, where N is the number of events counted in a given bin, and scaling the estimations by the bin size and area.

of 30 time units and compared the cumulative distributions of the areal plume rates in those bins (Figure 4.9). We performed a two-sample Kolmogorov-Smirnoff test (K-S test) to compare two given areal plume rate distributions. The K-S test provides a p-value p_{KS} that quantifies the probability that the two given samples were drawn from the same distribution. Thus, if the p-value was high ($p_{KS} \sim 10^{-1}$), then the two samples were similar [117]. If the p-value was low, then the two samples were significantly different ($p_{KS} \lesssim 10^{-2}$). For a given experiment or simulation, when we performed the K-S test on two halves of the whole time-series available, the p-values were high ($p_{KS} \sim 10^{-1}$). This indicates that the statistics of a given experiment and simulation are consistent throughout the time-series.

The simulation with radially-varying boundary conditions (Sim. 3) and the experiment with $\mathrm{Ra} = 4000$ (Exp. 3) had very similar total areal plume rate distributions overall, with

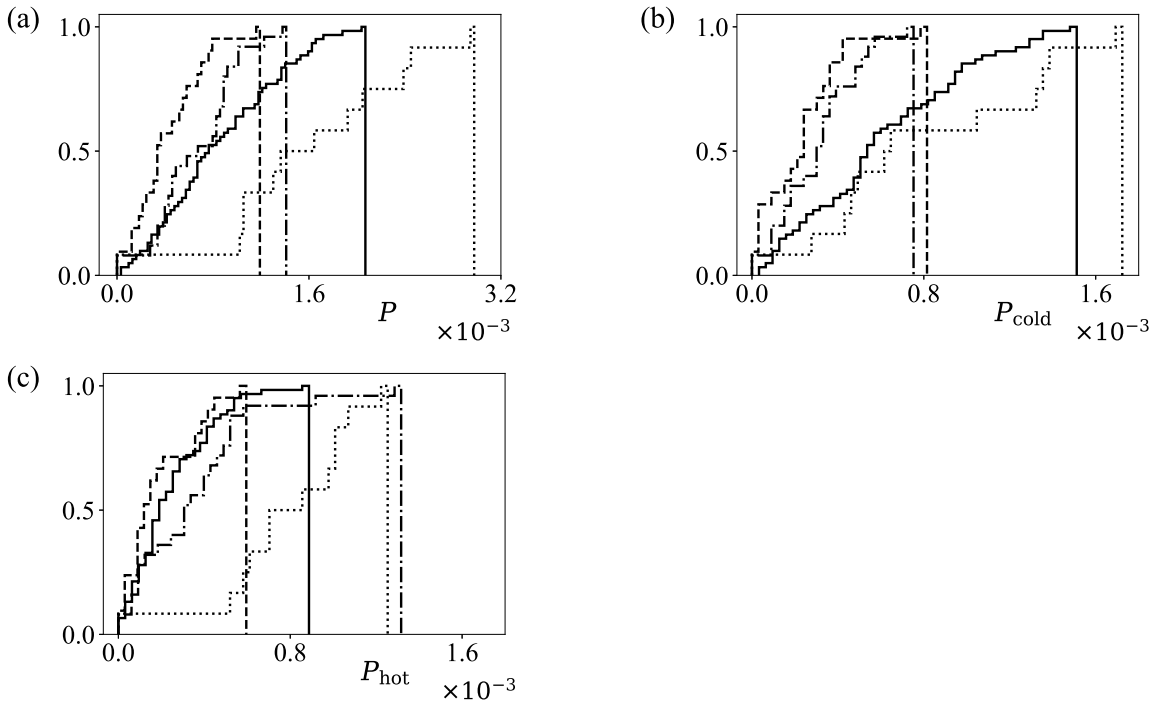

Figure 4.9: Cumulative histograms for total (a), cold (b), and hot (c) areal plume rates in 30 time unit bins reveal the difference between experiments (solid line) and simulations, even in simulation 3 with radially-varying boundary conditions. Simulation 1 is the dotted line, simulation 2 is the dashed line, and simulation 3 is the dashed-dotted line (Table 4.3). Cumulative histograms represent the fraction of bins that have an areal plume rate less than or equal to values on the x-axis. The experiment with $\mathrm{Ra} = 4000$ (Exp. 3, Table 4.1) and simulation 1 ($\mathrm{Ra} = 4000$ with spatially uniform temperature boundary conditions) show significant disagreement in their total areal plume rate distributions (a). Experiment 3 and simulation 2 ($\mathrm{Ra} = 3500$ with spatially uniform temperature boundary conditions) also show significant disagreement in their total areal plume rate distributions (a). Experiment 3 and simulation 3 (radially-varying thermal boundary conditions as shown in Figure 4.7) appear to have similar total plume rate distributions ($p_{KS} \sim 10^{-1}$), just as their average values would suggest (Figure 4.8). However, they show disagreement in both cold (b) and hot (c) areal plume rate distributions ($p_{KS} \sim 10^{-3}$ and $p_{KS} \sim 10^{-2}$, respectively).

a K-S test p-value of $p_{KS} \sim 10^{-1}$ (Figure 4.9(a)). Simulation 1, however, with spatially uniform boundary conditions at $\mathrm{Ra} = 4000$, had a significantly higher areal plume rate distribution than the same experiment ($p_{KS} \sim 10^{-4}$). Simulation 2, with spatially uniform boundary conditions at $\mathrm{Ra} = 3500$, had a significantly lower areal plume rate distribution than the same experiment ($p_{KS} \sim 10^{-3}$).

While the simulation with radially-varying boundary conditions (Sim. 3) produced a

total areal plume rate distribution that is similar to the experiment with $\mathrm{Ra} = 4000$, it does not capture the cold ($p_{KS} \sim 10^{-3}$) and hot ($p_{KS} \sim 10^{-2}$) areal plume rate distributions separately (Figure 4.9(b) and (c)). Deviations from the Boussinesq approximation in experiments caused them to produce more cold plumes than hot plumes.

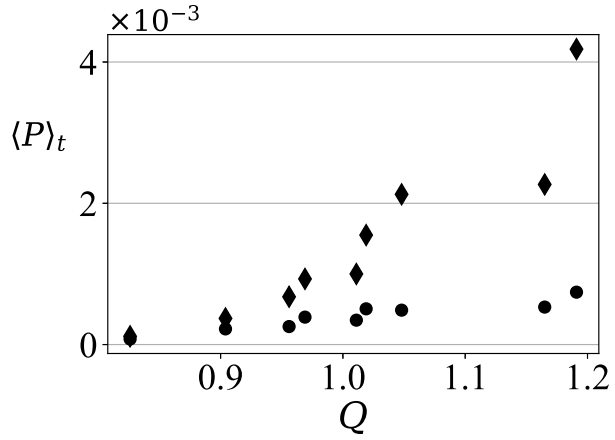

Figure 4.10: Cold and hot average areal plume rate difference increases as experiments stray further from the Boussinesq approximation. Here, we show cold (diamonds) and hot (circles) average areal plume rates ($\langle P_{\text{cold}} \rangle_t$ and $\langle P_{\text{hot}} \rangle_t$, respectively) versus Q, which quantifies the extent of departure from the Boussinesq approximation for experiments 1-9.

The difference in average cold and hot plume rates was correlated to Q, which quantifies the non-Boussinesqness of the flow in experiments (Figure 4.10). For $Q < 0.90$, there was no statistically significant difference in average areal cold and hot plume rates according to the K-S test p-value. As Q was increased to $Q = 0.90$ and greater, the difference in areal cold and hot plume rate distributions increased, as shown by their averages represented by diverging diamonds (average areal cold plume rate) and circles (average hot plume rate) in Figure 4.10. All K-S test p-values for experimental areal cold and hot plume rate distributions with $Q \geq 0.90$ were $p_{KS} \sim 10^{-6}$ or smaller. All simulations were Boussinesq, with $Q = 0$. Thus, all areal cold and hot plume rate distributions in simulations were not significantly different ($p_{KS} \sim 10^{-1}$).

For the experiment with $\mathrm{Ra} = 4000$ (Exp. 3), there was a significant difference between areal cold and hot plume rate distributions (Figure 4.9(b) and (c)). The experiment with

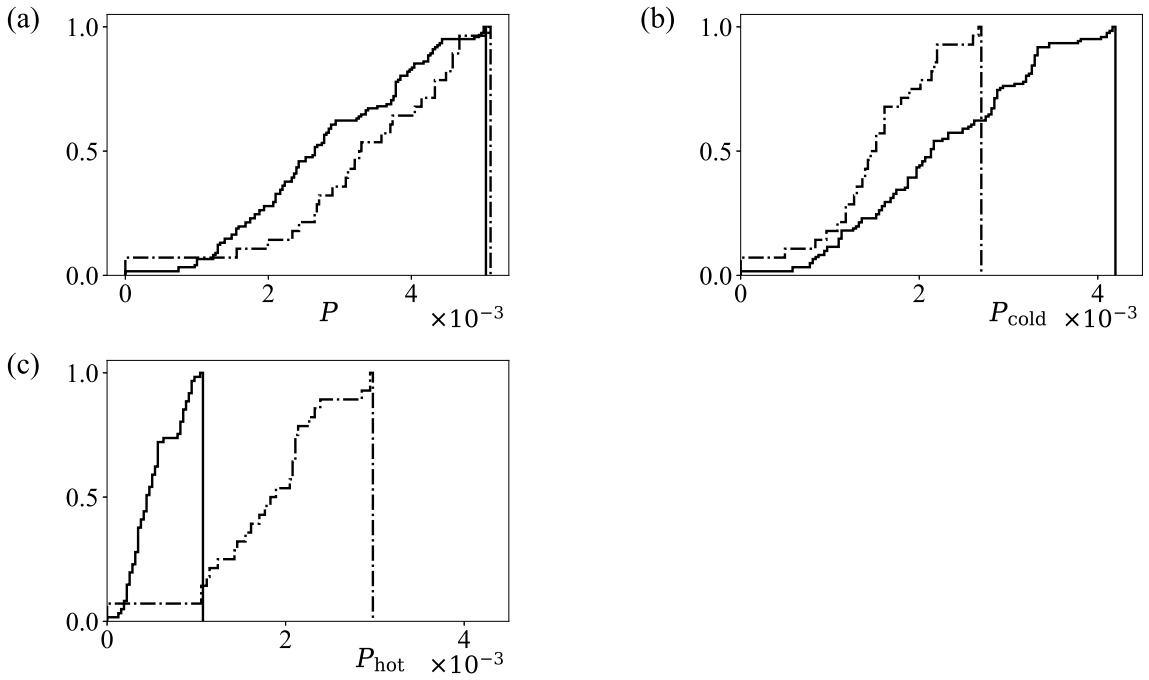

Figure 4.11: Cumulative histograms for Ra $= 4800$ total (a), cold (b), and hot (c) areal plume rates in 30 time unit bins show disagreement between experiment 8 (solid line) and simulation 4 (dashed-dotted line) due to increased non-Boussinesq effects (Figure 4.10). Cumulative histograms represent the fraction of bins that have an areal plume rate less than or equal to values on the x-axis. Experiments at higher Ra show a much more significant difference between cold and hot plume rates than those at lower Ra, while simulations show little difference. Experiment 8 and simulation 4 show some disagreement in their cold areal plume rate distributions (b), but significantly disagree in their hot areal plume rate distributions (c). The total (a) areal plume rates also show disagreement ($p_{KS} \sim 10^{-2}$).

Ra $= 4800$ (Exp. 8), however, had a much larger areal cold and hot plume rate difference (Figure 4.11(b) and (c)). Simulation 4, with radially-constant boundary conditions obtained using the heat bath settings from the Ra $= 4800$ experiment (Table 4.3), had an areal cold plume rate distribution that shows some disagreement with the corresponding experiment, however it had an areal hot plume rate distribution that differs from the experimental distribution substantially (Figure 4.11). The total areal plume rate distributions of simulation 4 and experiment 8 were significantly different, with a p-value of $p_{KS} \sim 10^{-2}$. The simulation 4 and experiment 8 (Ra $= 4800$) showed a larger difference in their total areal plume rate distributions than simulation 3 and experiment 3 (Ra $= 4000$) did because the discrepancy between hot areal plume rate distributions is much higher in the former pair than in the

latter.

The plume statistics detected in this work demonstrate the power of persistent homology to analyze pattern features. Our method allowed us to collect the statistics of unique plumes for an analysis that revealed key differences between descriptions and experiments inaccessible to other methods.

Persistent homology demonstrated that simulations that employ temperature boundary conditions that account for heat transfer in the finite size and thermal conductivity heating and cooling plates and sidewall, as described in section 4.4, produce total areal plume rates that are closer to experiments than those with temperature boundary conditions that match the temperatures imposed by heat baths in experiments (information on experiments and simulations can be found in Table 4.1 and Table 4.3). Simulation 3, with the radially-varying boundary conditions, has a total areal plume rate distribution that is close to the experiment with $\mathrm{Ra} = 4000$ (Exp. 3), while simulation 1, with the spatially uniform boundary conditions, has a total areal plume rate distribution that is much higher (Figure 4.7 and Figure 4.9(a)). The K-S test reveals that the total areal plume rate distributions for simulations with spatially uniform boundary conditions at $\mathrm{Ra} = 4000$ and $\mathrm{Ra} = 3500$ and the experiment with $\mathrm{Ra} = 4000$ are significantly different ($p_{KS} \sim 10^{-4}$ and $p_{KS} \sim 10^{-3}$, respectively) while the same distributions for the simulation with radially-varying boundary conditions (Sim. 3) and the same experiment are similar ($p_{KS} \sim 10^{-1}$).

The simulation with radially-varying boundary conditions describe the total areal plume rate distribution of the experiment more accurately than those with spatially uniform boundary conditions because the effective temperature difference at the fluid layer boundaries of experiments $\Delta T(r)$ is everywhere smaller than the temperature difference imposed by the heat baths $\Delta T = T_1 - T_2$. Additionally, near the thermally conducting sidewalls, the effective temperature difference becomes even smaller (Figure 4.7). The total reduced effective driving and radial dependence of the simulations that used radially-varying temperature boundary conditions produced fewer detectable plumes overall, which agrees more with

experimental measurements. That being said, the radially-varying temperature boundary conditions do not fully describe experimental conditions. The temperature at the bottom of the bottom plate and top of the top plate are fixed by the heat baths in our experiments. Heat flux is then distributed throughout the experimental cell according to the thermal conductivity of various materials in order to satisfy these boundary conditions. Thus, the thermal boundaries of the fluid layer adjust accordingly and they could be considered neither constant-temperature nor constant-heat flux. The approach to the thermal boundaries outlined in section 4.4 more closely approximates the experimental thermal boundary conditions in simulations. However, since there is still a time-independent temperature at the boundaries of the fluid layer in simulations, these conditions still do not perfectly match experiments. In future work, convection simulations that incorporate the heat transfer in the top and bottom cooling and heating plates and the sidewalls could provide more insight.

The simulation with radially-varying boundary conditions (Sim. 3) and the experiment with Ra $= 4000$ (Exp. 3) have very similar total areal plume rate distributions, however when we look at their cold and hot areal plume rate distributions separately, we see disagreement (Figure 4.9(b) and (c)). The simulation produces fewer cold plumes than the experiment and more hot plumes. Further investigation reveals that non-Boussinesq effects cause experiments to produce more cold than hot plumes, and this difference increases as non-Boussinesq effects become stronger.

By comparing the radially-varying boundary condition simulations (Sims. 3 and 4) and their corresponding experiments at Ra $= 4000$ and Ra $= 4800$, respectively, we observe that the total areal plume rates increased with Ra in both Boussinesq simulations and experiments (compare Figure 4.9(a) and Figure 4.11(a)). In experiments, as Q increased alongside Ra, the deviations from Boussinesq behavior became more significant. In our results we find that the difference between average areal cold and hot plume rates increased with increasing Q as shown in Figure 4.10. This is consistent with the more general difference in cold and hot 1-dimensional Betti numbers in past work [35]. In that work, it

was demonstrated that the difference in the 1-dimensional Betti numbers β_1 of cold and hot regions of experimental RBC was correlated with Q, not Ra. Plumes are a subset of the dimension-1 homology group that quantifies holes, so it is expected that cold and hot plumes occur at different rates and that the difference depends on Q.

The simulation with radially-varying boundary conditions (Sim. 4) and the corresponding experiment with $\mathrm{Ra} = 4800$ (Exp. 8) have total areal plume rate distributions that disagreed (Figure 4.11(a)) due to the increased deviations from the Boussinesq approximation relative to the experiment with $\mathrm{Ra} = 4000$. Figure 4.11(b) and (c) show that simulation 4 produced a cold areal plume distribution that was smaller than experiment 8 and a hot areal plume rate distribution that was much larger than experiment 8.

Simulations of the Boussinesq equations contain no non-Boussinesq effects, by definition. Thus, the Boussinesq simulations examined in this work demonstrate no significant difference between cold and hot plumes (Figure 4.9 and Figure 4.11). The K-S test confirms that all cold and hot areal plume rate distributions in simulations were not significantly different ($p_{KS} \sim 10^{-1}$). When Ra was increased in simulations, both cold and hot areal plume rates rose equally on average. In our experiments, Q increases with Ra. As Q and Ra increase, non-Boussinesq effects cause experimental cold and hot areal plume rates to increase at different rates (Fig. Figure 4.10). Cold areal plume rates increase steadily with Q and Ra in experiments while hot areal plume rates increase hardly at all. Because the cold and hot plume rate both increase steadily as Ra is increased in simulations, the difference between cold areal plume rate distributions in experiments and radially-varying boundary condition simulations is steady (Figure 4.9(b) and Figure 4.11(b)), while the difference between hot areal plume rates widens (Figure 4.9(c) and Figure 4.11(c)).

4.6 Next steps in plume statistics analysis

Direct comparisons between pattern features that appear in both numerical simulations and experiments of spatiotemporally complex systems continue to be sought, even in well-

understood paradigm problems such as RBC. Such comparisons allow us to both assess the accuracy of theoretical descriptions and gain insights into experimental observations. Persistent homology, a technique of topological data analysis, gives us the opportunity to make such comparisons between the topological structures of both numerical simulations of descriptions and experiments in RBC undergoing SDC. In this work, we established a method to compare statistics of plumes, pattern features detected by persistent homology, observed in simulations of the Boussinesq equations to those observed in experiments.

This method allowed us to assess the accuracy of the Boussinesq description through comparison of plume statistics in numerical simulations to those of experiments. We found that the description that incorporates radially-varying time-independent temperature boundary conditions obtained from a conduction model of the experiment that account for the finite thermal conductivity of the bottom and top heating and cooling plates and sidewall of our experimental cell produce total plume statistics more similar to experiments than the description that uses spatially uniform time-independent temperature boundary conditions obtained from our heat bath temperature settings. However, we argue that the thermal boundary conditions of the fluid layer in our experiments are not truly time-independent temperature boundary conditions. The comparison of plume statistics also revealed that deviations from the Boussinesq approximation in our experiments cause significantly more cold plumes to be detected than hot plumes. In numerical simulations of the Boussinesq equations, cold and hot plumes were observed at similar rates, on average. Thus, we anticipate that a non-Boussinesq description that includes the heat transfer in the finite size, finite thermal conductivity bottom and top plates as well as the sidewall will provide a better description of total, cold, and hot plume statistics of our experiments.

We aim to quantitatively compare plume statistics in numerical simulations and experiments because plumes are associated with high leading Lyapunov vector magnitude in simulations of the Boussinesq equations [36]. The leading Lyapunov vector indicates spatially-localized areas of a convective flow in which perturbations grow rapidly [77].

This fundamental dynamical quantity can currently only be obtained in numerical simulations. In future work, we anticipate that theoretical descriptions can be tailored further to reproduce the dynamics of topological structures for specific experimental conditions. This includes topological structures, such as plumes, where it may be possible to gain insight into experimental flows where fundamental dynamical quantities, such as the leading Lyapunov vector, cannot be observed directly. The leading Lyapunov vector is not the only fundamental quantity of interest of a convective flow. A future step that could build upon this work is investigation of the local heat transport of the convective state and its correlation with plumes. Plumes may serve as a diagnostic for local changes in heat transport because it has been shown in the past that other significant topological activity in the convective state correlates with drops in the heat transport [118].

In the future, we seek to bring plume statistics in experiments and simulations into alignment by performing simulations with experimentally realistic boundary conditions and deviations from the Boussinesq approximation. Then, we must reassess the spatiotemporal correlation between plumes and the leading Lyapunov vectors in the improved simulations to determine if their relationship is robust to perturbations in the boundary conditions and fluid parameters. Only then can we determine if the leading Lyapunov vectors are consistently indicated by plumes. Our method of quantifying plume statistics in convection roll patterns of simulations and experiments undergoing SDC, made possible by persistent homology, enabled us to assess areal plume rates in both settings, allowing us to take the first step toward this goal.

CHAPTER 5
CONCLUSIONS

In this dissertation, the dynamics underlying spatiotemporal chaos were analyzed using persistent homology in different settings. In the Kuramoto-Sivashinsky equation, a model system that exhibits spatiotemporal chaos used to study turbulence, persistent homology was employed to detect shadowing of turbulence to relative periodic orbits. In RBC experiments and simulations, persistent homology was used to detect plumes, an observable feature associated with dynamical features. These efforts represent significant progress in the studies of nonlinear dynamics and spatiotemporal chaos. The scientific contributions of this work are summarized in section 5.1. Then, in section 5.2, the next steps that should be taken to advance these projects are outlined.

5.1 Summary of scientific contributions

5.1.1 Exact coherent structures in the Kuramoto-Sivashinsky equation

In chapter 3, it was demonstrated that shadowing of ECS can be detected and quantified in the space of persistence diagrams in a spatiotemporally chaotic dynamical system with a continuous symmetry, the KS equation. KS is a "toy model" in which spatiotemporally chaotic behavior can be explored in a simplified setting. Importantly, this model features a continuous symmetry, a feature which has historically frustrated efforts to compute ECS and detect shadowing. When a system has a continuous symmetry, TWs and RPOs, families of symmetry-related solutions to the governing equations, are possible. In principle, turbulence could be shadowing any member of the symmetry-related family of solutions that make up a TW or RPO. So, it is necessary to reduce this continuous symmetry in the data itself, minimize the desired metric over the continuous symmetry, or search for shad-

owing in a symmetry-invariant space. In this work, it was demonstrated that shadowing can be detected in a symmetry-invariant space given by persistent homology that is simple to implement, detects similar shadowing events compared to those detected by a minimization approach, and is less computationally expensive.

In previous studies of ECS in KS, Budanur *et al.* [26, 28] reduced the continuous symmetry using a Fourier mode slicing method. Though this method yields insights into ECS in the KS equation, it is difficult to work with because of discontinuities that occur when trajectories cross the "slice border" in state space. The method described in this work represents KS trajectories in a symmetry-invariant space without such difficulties. Persistence diagrams of KS trajectories evolve in time continuously, without having to traverse a discontinuous, symmetry-reduced state space. KS states that are represented in the space of persistence diagrams have much lower dimension than a symmetry-reduced state space. We demonstrated that even with so much information about the flow filtered out in the persistence space representation, it is still possible to detect shadowing. Furthermore, an appeal of this method is that it is straightforward to compute persistence diagrams of data of an arbitrary dimension, so it is applicable to three-dimensional systems with continuous symmetries, such as pipe flow or Taylor-Couette flow. Gaining an understanding of ECS in these systems is especially important because they are experimentally realizable paradigms for the study of fluid turbulence. In other previous studies of ECS in Taylor-Couette flow, Krygier *et al.* [13] detected shadowing by minimizing the difference between a turbulent trajectory and ECS over the continuous symmetry direction. This approach yielded insights into ECS in Taylor-Couette flow, but it is computationally expensive and difficult to implement. In this method, the distance between any two states in state space must be minimized over the continuous symmetry direction to be compared. This can be most efficiently done with a Fourier method. This method was implemented in KS and its performance was compared to that of the persistence space shadowing detection method. Shadowing events detected by the two methods were shown to be similar in frequency and statistics. Cru-

cially, the persistence space shadowing detection was much more computationally efficient because no minimization over a continuous symmetry is required.

In addition to serving as a proof-of-concept for persistence space shadowing detection, this work provides new insight into shadowing behavior in KS. Shadowing of RPOs was found to be frequent, even when analyzing a relatively small set of ECS. The time scales associated with shadowing are correlated to time scales associated with the linear stability of the ECS. This is an intriguing finding, because shadowing occurs well away from the ECS itself, far outside the regime where turbulence could be thought to be following the linear manifold. This frequent, nonlinear shadowing behavior indicates that ECS structure the dynamics of turbulence as it traverses the inertial manifold, not just in extreme situations when turbulence is very close to a particular ECS.

In sum, this work found that persistence space shadowing detection, a novel method of analyzing ECS, could quantify shadowing events in a similar manner to other existing methods. The statistics of these shadowing events offer insights into the ECS that structure dynamics in the KS equation. Crucially, this method is more computationally efficient than a traditional method when applied in this setting due to the system's continuous symmetry. These findings represent significant advances in the study of turbulence as a dynamical system.

5.1.2 Quantification of plumes in Rayleigh-Bénard convection

In chapter 4, a method to detect and track plumes in time in convection roll patterns in the SDC regime of RBC was developed. This convective flow exhibits spatiotemporal chaos in an experimentally realizable setting and simulation techniques are well-developed, making it an appealing setting for comparison of dynamics in experiments and simulations. In simulations under "ideal conditions" (constant temperature boundary conditions and the Boussinesq approximation), Levanger *et al.* [36] showed that plumes are correlated with high leading Lyapunov vector magnitude, suggesting that plumes indicate regions of the

flow in which perturbations grow rapidly. This makes them a particularly interesting pattern feature that can be observed in both experimental and simulated SDC. By detecting and tracking plumes in time in both experimental and simulated data, the extent to which simulation conditions captured the conditions of experiments was assessed. It was found that simulations under ideal conditions do not capture the plume statistics observed in experiments. Instead, plume statistics indicated that simulations need to incorporate boundary conditions that capture the heat transport through the experimental cell and fluid parameters that do not obey the Boussinesq approximation. Using temperature boundary conditions in simulations that were obtained through a heat transport model of the experimental cell, total plume statistics that agree with experiments more than those in the constant temperature boundary conditions studied by Levanger *et al.* [36] were observed. Thus, it was demonstrated that temperature boundary conditions are significantly correlated to the plume statistics observed in simulations. This correlation indicates that temperature boundary conditions play a significant role in the dynamics of simulated RBC. This is a significant contribution to the study of SDC, especially in cases where researchers seek to compare observable features in experimental and simulated data.

The significance of non-Boussinesq effects on the observation of cold and hot plumes in experiments was also demonstrated. As the Busse parameter Q that quantifies the extent of deviation from the Boussinesq approximation increased in experiments, so did the difference between cold and hot plume statistics. No such difference between cold and hot plume statistics was detected in simulated data, which was generated under the assumption of the Boussinesq approximation ($Q = 0$). This was the case in simulations that employed both types of temperature boundary conditions, the "ideal" case and boundaries that were obtained from the heat transport model of the experimental cell. This indicates that temperature boundary conditions only affect the total plume statistics observed, not the difference between cold and hot plumes.

In sum, it was demonstrated that both experimentally realistic boundary conditions and

deviations from the Boussinesq approximation affect plume statistics observed in SDC. These are significant findings for the ongoing effort to use observable pattern features in experiments and simulations to study the underlying dynamics of spatiotemporal chaos.

5.2 Next steps

5.2.1 Further study of ECS in the Kuramoto-Sivashinsky equation

A sample of ECS in the KS equation with $L = 22$ were studied in chapter 3. There remains a great deal of work to be done to fully characterize the dynamics of KS using ECS. In total, sixteen RPOs and one equilibrium were considered in this work. This represents a tiny fraction of the relevant ECS embedded in the turbulent attractor of the KS system. Cvitanović *et al.* [60], for example, computed thousands of ECS in the KS equation. There are certainly many ECS that we do not account for that are visited by turbulent KS trajectories. In our data set, about 25% of the time, a turbulent trajectory is not shadowing any of our sixteen RPOs at all. Ideally, in a deterministic walk in state space where one ECS after another is visited by turbulence, one should be able to completely account for the dynamics of a turbulent KS trajectory at all moments using ECS. Once dynamically relevant shadowing events completely account for the turbulent trajectory at all moments, one should be able to construct symbolic dynamics [28]. In such a theory, the KS state space (or space of persistence diagrams) would be coarse-grained into regions governed by shadowing of local ECS. Transitions between these regions would then be tracked by a Markov chain. Such a theory would provide a detailed description of turbulent dynamics in which future states could be predicted based on the current location of the trajectory in state space. This is the ultimate goal of the study of ECS in fluid flows, and a simple, relatively low-dimensional model system like KS is an amenable setting for achieving this vision of a deterministic, dynamical theory of spatiotemporal chaos.

5.2.2 Persistence space shadowing detection in more complex fluid flows

It has been demonstrated that persistence space shadowing detection is a viable and computationally efficient method for detecting shadowing in KS, a model system with a continuous symmetry. A natural next step is to apply this method in higher dimensional systems with continuous symmetries, where gains in computational efficiency over traditional methods of shadowing detection are anticipated to become more pronounced.

One candidate system could be 2-D Kolmogorov flow with periodic boundary conditions, which has a continuous translational symmetry in one spatial direction similar to KS. In this system, persistent homology would be applied to 2-D periodic image data representing a scalar vorticity field instead of 1-D functions like in the KS case. Many more spatial points are needed to described a 2-D image than 1-D KS states, so this system exists in a state space of much higher dimension. The computation of the distance in state space scales with the number of points, however the computation in space of persistence diagrams scales with the complexity of the pattern, so it is anticipated that the persistence space shadowing detection would be much more computationally efficient in higher dimensions. Further, persistence space shadowing detection could be applied in 3-D, experimentally realizable systems like pipe flow and Taylor-Couette flow. Pipe flow has a streamwise continuous translational symmetry and a continuous rotational symmetry about the central axis of the pipe. Traditional methods of shadowing detection must account for these symmetries by, for example, symmetry-reducing the state space [61]. This is unnecesary in persistence space shadowing detection, so it's expected that it would be simpler to implement. Persistence space shadowing detection requires a scalar field for analysis, so a suitable scalar diagnostic of a flow state would have to be considered in pipe flow, such as the downstream velocity component relative to the parabolic laminar flow profile [27]. In principle, however, one could choose a set of scalar fields, such as one set for each velocity component, to analyze. In this case, the distance between three sets of persistence diagrams would have to be combined to form a single metric of distance between two states. Taylor-Couette

flow has a continuous rotational symmetry, which was dealt with by Krygier *et al.* [13] by minimizing over the possible rotations of the flow field. This minimization would be unnecessary using persistence space shadowing detection, so gains in computational efficiency are anticipated.

5.2.3 Aligning plume statistics in spiral defect chaos simulations and experiments

In chapter 4, it was demonstrated that plume statistics in simulations under "ideal conditions" and experiments of SDC are different. Those ideal conditions consist of constant temperature boundary conditions and the assumption of the Boussinesq approximation, both of which are violated in experiments. Using radially-varying temperature boundary conditions obtained from a heat transport model of the experimental convection cell improved agreement between total plume statistics of simulations and experiments, but it failed to bring them into complete alignment. The next step should therefore be conducting improved simulations in which we can observe plume statistics that align with experiments.

To align the plume statistics in experiments and simulations, simulations must be modified to feature experimentally realistic boundary conditions and fluid parameters. In experiments, the temperature profile in the heating and cooling plates at the boundary of the fluid is dynamically coupled to the convective flow in the fluid layer itself [100]. It is not simply a constant temperature or constant heat flux boundary as it is often modeled. Therefore, it is necessary to simulate not just convection in the fluid layer, as is most commonly done; we must also simulate the heat transport through the heating and cooling plates below and above the fluid layer, as well as the thermally conducting sidewall. Only then can the heat flow through the fluid layer be accurately captured by simulations. Additionally, it was demonstrated that deviations from the Boussinesq approximation in experimental fluid yield a significant difference in the observation of cold and hot plumes. Future simulations of SDC must allow for fluid that deviates from the Boussinesq approximation to the extent that is observed in experiments.

Aligning plume statistics in simulations and experiments is important because plumes in "ideal" simulations indicate spatial regions of the convective flow in which the leading Lyapunov vector magnitude is high. This fundamental dynamical quantity indicates spatial regions of the flow in which perturbations grow rapidly, making them important for characterizing the spatiotemporal dynamics of SDC. We thus seek to use plumes, an observable pattern feature in experiments and simulations, as an indicator of underlying dynamics. To establish that the underlying dynamics are similar in experiments and simulations, similar plume statistics must be observed in both. Once this is established, the relationship between plumes and the leading Lyapunov vector magnitude in experimentally realistic simulations must be re-established. At this point, it will be possible to characterize the dynamics of both experiments and simulations with plumes, a feature that is readily observable in convection roll patterns using persistent homology.

Appendices

APPENDIX A

FOURIER SHIFTING TO MINIMIZE OVER CONTINUOUS TRANSLATIONAL SYMMETRY

To use the D to compare two states of the KS equation, we must find the shift $s = \arg\min_{\tilde{s}\in[0,L]} D_{\tilde{s}}$ that minimizes Equation 3.3. We can find such a shift with subpixel resolution using a Fourier method. For simplicity, we will consider two states $v(x)$ and $u(x)$ that have no time dependence and their unit vectors $\hat{v}(x) = v(x)/|v(x)|$ and $\hat{u}(x) = u(x)/|u(x)|$. We seek $s = \arg\min_{\tilde{s}\in[0,L]} ||v(x+\tilde{s}) - u(x)||_2$. We can express $v(x)$ as a linear combination of the shifted copy of u and the orthogonal state $u_\perp$ such that $v(x) = c_1(\tilde{s})u(x+\tilde{s}) + c_2(\tilde{s})u_\perp(x+\tilde{s})$. We seek the value of $\tilde{s}$ where $c_1(\tilde{s})$ is a maximum and $c_2(\tilde{s})$ is as close to 0 as possible. We can write $u(x)$ and $v(x)$ in terms of their Fourier transforms

$$v(x) = \int V(k)e^{ikx}dk = \int [c_1(\tilde{s})U(k) + c_2(\tilde{s})U_\perp(k)]e^{ik(x+\tilde{s})}dk \tag{A.1}$$

. Further, we can construct a function

$$F(\tilde{s}) = \int \bar{\hat{U}}\hat{V}dk = \int \frac{\bar{U}(k)}{|u(x)|}\frac{[c_1(\tilde{s})U(k) + c_2(\tilde{s})U_\perp(k)]}{|v(x)|}e^{ik(\tilde{s})}e^{-ik(\tilde{s})}dk \tag{A.2}$$

where the overbar denotes the complex conjugate. By Plancherel's theorem the second term of the sum is zero and the first term gives the modulus of u such that

$$F(\tilde{s}) = \frac{c_1(\tilde{s})}{|v(x)|}. \tag{A.3}$$

So, $s = \arg\max_{\tilde{s}\in[0,L]} F(\tilde{s})$ where $F(\tilde{s})$ is a function that can be constructed easily from the fast Fourier transforms of $v(x)$ and $u(x)$. We integrate trajectories of Equation 3.1 that have $N = 32$ spatial grid points, but we compute s to subpixel resolution by interpolating $F(\tilde{s})$

to a higher resolution of $N = 256$. This increases the computational cost of computing D, but is necessary to detect shadowing because subpixel shifts can cause large differences in $u(x+s)$ and $v(x)$. The trajectory $u(x)$ can then be shifted by rotating its Fourier transform by applying $U_s = e^{ik\tilde{s}}U$ where $\tilde{s} = \frac{2\pi s}{N}$ where k is the wavenumber for each Fourier mode.

APPENDIX B

RAYLEIGH-BÉNARD CONVECTION EXPERIMENTAL SPECIFICATIONS

The working fluid in experiments is compressed sulfur hexafluoride gas SF_6. The Prandtl number is Pr$=0.84$. The aspect ratio of the experimental cell is $\Gamma=19.63$ for experiments 3, 4, 7, and the aspect ratio is $\Gamma = 19.53$ for experiments 1, 2, 5, 6, and 8. The aspect ratio for all the Boussinesq simulations is $\Gamma = 20$. The height of the fluid layer, or depth $d = 6.08 \times 10^{-4}$m for experiments 3, 4, 7, and 9. The depth $d = 6.05 \times 10^{-4}$m for experiments 1, 2, 5, 6, and 8.

The following fluid properties are given at the mean temperature $T_0 = 294.64$ of experiment 3. The density $\rho=114\frac{\text{kg}}{\text{m}^3}$. The thermal expansion coefficient $\beta=0.0074\frac{1}{\text{K}}$. The thermal conductivity of fluid $\lambda = 0.014\frac{\text{W}}{\text{mK}}$. The heat capacity $c_p = 7.5 \times 10^2 \frac{\text{J}}{\text{kgK}}$. The kinematic viscosity $\nu = 1.4 \times 10^{-7}\frac{\text{m}^2}{\text{s}}$. The thermal diffusivity $\kappa = 1.7 \times 10^{-7}\frac{\text{m}^2}{\text{s}}$. The Nusselt number computed in simulation 1 is Nu $= 1.65$. The vertical diffusion time unit $t_v=d^2/\kappa=2.20$ s.

The polyethersulfone sidewalls have thermal conductivity $0.18 \leq \lambda \leq 0.24\frac{\text{W}}{\text{mK}}$ [119]. The sapphire single crystal top plate has thermal conductivity $35 \leq \lambda \leq 40\frac{\text{W}}{\text{mK}}$ [120]. The zinc selenide crystal bottom plate has $\lambda=18\frac{\text{W}}{\text{mK}}$ [121].

Dimensional radially-varying temperature profiles at the bottom and top boundary of the fluid layer in experiments are obtained by solving Equation 4.5 on the model cell geometry (Figure 4.2) and then fitting a fourth-order polynomial to obtain a functional form $T^*(r)=C_1r^4+C_2r^3+C_3r^2+C_4r+C_5$. These functions are then used as the radially-varying thermal boundary conditions for convection simulations 3 (corresponding to the experiment with Ra$=4000$) and 4 (corresponding to the experiment with Ra$=4800$) (Table B.1). All dimensional temperature values have units of Kelvin (K).

Table B.1: Coefficients for fourth-order polynomial temperature functions used as temperature boundary conditions in simulations 3 (Ra $= 4000$) and 4 (Ra $= 4800$). The functions take the form $T^*(r) = C_1 r^4 + C_2 r^3 + C_3 r^2 + C_4 r + C_5$. The units of the temperature coefficients are expressed in Kelvin (K).

Ra $= 4000$	C_1 (K)	C_2 (K)	C_3 (K)	C_4 (K)	C_5 (K)
$T_b(r)$	-3.3×10^{-6}	9.2×10^{-5}	-1.3×10^{-3}	3.2×10^{-3}	297.40
$T_t(r)$	1.5×10^{-6}	-4.2×10^{-5}	6.2×10^{-4}	-1.5×10^{-3}	292.17
Ra $= 4800$					
$T_b(r)$	-4.0×10^{-6}	1.1×10^{-4}	-1.5×10^{-3}	3.9×10^{-3}	298.77
$T_t(r)$	1.9×10^{-6}	-5.3×10^{-5}	7.6×10^{-4}	-1.8×10^{-3}	292.33

APPENDIX C

RAYLEIGH-BÉNARD CONVECTION CONDUCTION MODEL VERIFICATION

The model convection cell conduction is performed using the MATLAB PDE Toolbox. The toolbox uses a finite element method for approximating the solution to PDEs as linear functions on a triangular mesh. The PDE we use is the steady-state heat equation in two dimensions

$$\nabla^2 T^*(r, z) = 0. \tag{C.1}$$

In cylindrical coordinates, the steady-state heat equation is expressed as

$$\frac{1}{r}\frac{\partial}{\partial r} r \frac{\partial T}{\partial r} + \frac{\partial^2 T}{\partial z^2} = 0 \tag{C.2}$$

In MATLAB this is written as -div(c*grad(u))+a*u=f where u is the field variable of interest (T), λ is the thermal conductivity of the material, div is the divergence, grad is the gradient, and a=f=0. The convection cell is modeled as a radial cross-section of a cylinder, broken up into four distinct regions, each with different thermal conductivity (Table C.1).

The top and bottom boundaries of the domain are fixed to the temperature of the heat baths for a given experiment. The boundaries at the sides of the domain are zero-flux. The boundaries between the different materials have the conditions $T_1(r_1, z_1) = T_2(r_1, z_1)$ and $\lambda_1 * \nabla T_1(r_1, z_1) = \lambda_2 * \nabla T_2(r_1, z_1)$ where r_1 and z_1 are coordinates of a boundary between

Table C.1: Spatial coordinates left, bottom, width, and height of the four regions in the conduction model displayed in units of centimeters

Region	λ ($\frac{\mathrm{W}}{\mathrm{mK}}$)	Left	Bottom	Width	Height
Fluid layer	0.014	0	0.95	1.1875	0.0605
Top plate	35	0	1.0105	1.6875	1.91
Bottom plate	18	0	0	1.6875	0.95
Side wall	0.20	1.1875	0.95	0.5	0.0605

materials ”a” and ”b”.

The thermal conductivity of the fluid layer is also multiplied by the Nusselt number computed from non-dimensional Boussinesq simulations to approximate convective conditions in the fluid. In the Ra = 4000 case this is Nu = 1.65. In the Ra = 4800 this is Nu = 1.76.

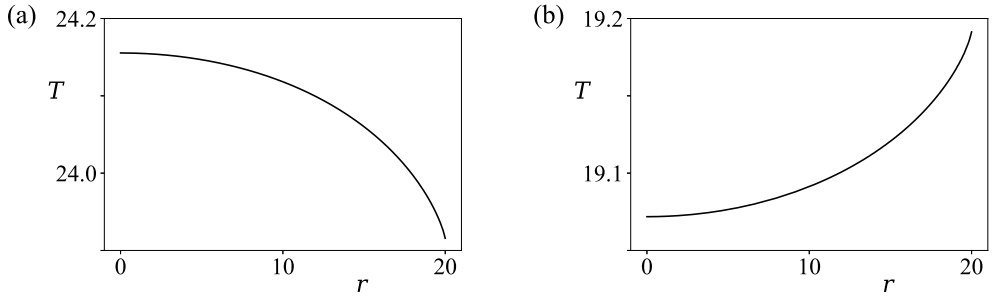

Figure C.1: Temperature profiles as a function of the radial coordinate r at the bottom (a) and top (b) fluid layer boundaries computed by the conduction model.

We confirmed that the PDE Toolbox approximation scheme was producing results consistent with analytic solutions of the heat equation in circumstances where the result is known. In particular, we considered a model with two distinct regions of different thermal conductivities λ_1 and λ_2 (Figure C.2) with temperature varying in the radial direction. The left region (region ”a” for $1 \leq r \leq 2$) has thermal conductivity $\lambda_1 = 1$ and the right region (region ”b” for $2 \leq r \leq 3$) has thermal conductivity $\lambda_2 = 2$. The boundary at $r = 1$ has a fixed temperature of $T_1(r = 1) = 1$ while the boundary at $r = 3$ has a fixed temperature of $T_2(r = 3) = 0$. All quantities in this validation are given in arbitrary units for simplicity. The boundaries at the top ($z = 1$) and the bottom ($z = -1$) are zero-flux ($\frac{\partial T}{\partial z} = 0$), thus the temperature has no z-dependence and $T(r, z) = T(r)$. Because there are two regions with two different values of thermal conductivity, there will be separate solutions for the left region $T_1(r)$ and the right region $T_2(r)$ At the boundary between the two regions at $r = 2$ there must be continuous temperature $T_1(r = 2) = T_2(r = 2)$ and continuous heat flux $\lambda_1 \frac{\partial T_1}{\partial r}(r = 2) = \lambda_2 \frac{\partial T_2}{\partial r}(r = 2)$.

If r and z are incorrectly treated as Cartesian coordinates x and y, then MATLAB will

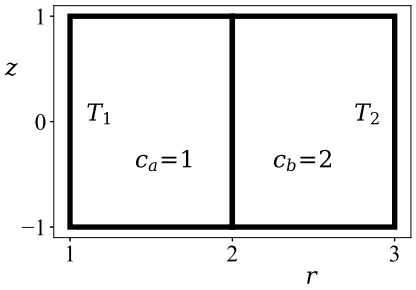

Figure C.2: Sample geometry used to validate MATLAB PDE toolbox. The left region ($1 \leq r \leq 2$) has thermal conductivity $\lambda_1 = 1$ and the right region ($2 \leq r \leq 3$) has thermal conductivity $\lambda_2 = 2$. The boundary at $r = 1$ has a fixed temperature of $T_1 = 1$ while the boundary at $r = 3$ has a fixed temperature of $T_2 = 0$.

solve Equation 4.5 as

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = 0. \tag{C.3}$$

Given the zero-flux boundary condition at the top and bottom, this simplifies to

$$\frac{\partial^2 T}{\partial x^2} = 0, \tag{C.4}$$

which has a general solution of $T(x) = C_1 x + C_2$. Applying boundary conditions, the analytic solution for both regions is

$$T_a(x) = \begin{cases} T_1(x), \text{if } 1 \leq x \leq 2 \\ T_2(x), \text{if } 2 < x \leq 3, \end{cases} \tag{C.5}$$

where

$$T_1(x) = -\frac{2}{3}x + \frac{5}{3} \tag{C.6}$$

$$T_2(x) = -\frac{1}{3}x + 1. \tag{C.7}$$

The temperature at the boundary is thus $T_1(x = 2) = T_2(x = 2) = \frac{1}{3}$ in the Cartesian case. The MATLAB PDE toolbox finds the same result using the finite element method and when

treating the model in Cartesian coordinates (the dashed curve in Figure C.3 (a)).

If the geometry is handled using radial coordinates, then Equation 4.5 will simplify to

$$\frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial T}{\partial r}=0, \tag{C.8}$$

which has the general solution $T(r) = C_1 \ln r + C_2$. Applying boundary conditions, the analytic solution for both regions is

$$T_a(r)=\begin{cases} T_1(r), \text{if } 1 \leq r \leq 2 \\ T_2(r), \text{if } 2 < r \leq 3, \end{cases} \tag{C.9}$$

$$T_1(r)=\frac{1}{\ln \frac{1}{2}\sqrt{\frac{2}{3}}}\ln r+1 \tag{C.10}$$

$$T_2(r)=\frac{1}{2\ln \frac{1}{2}\sqrt{\frac{2}{3}}}\ln r-\frac{\ln 3}{2\ln \frac{1}{2}\sqrt{\frac{2}{3}}}. \tag{C.11}$$

The temperature at the boundary is thus $T_1(r=2)=T_2(r=2)\approx 0.2263$ in the cylindrical case.

Figure C.3 (b) shows the absolute value of the difference between the analytic (T_a) and MATLAB finite element method (T_m) solution for both the Cartesian and cylindrical cases. The MATLAB PDE toolbox finds the same result using the finite element method and when treating the model in cylindrical coordinates (the solid curve in Figure C.3 (a)). The MATLAB PDE Toolbox's finite element approximation method thus provides the right answer in known cases within accuracy of about 10^{-5}.

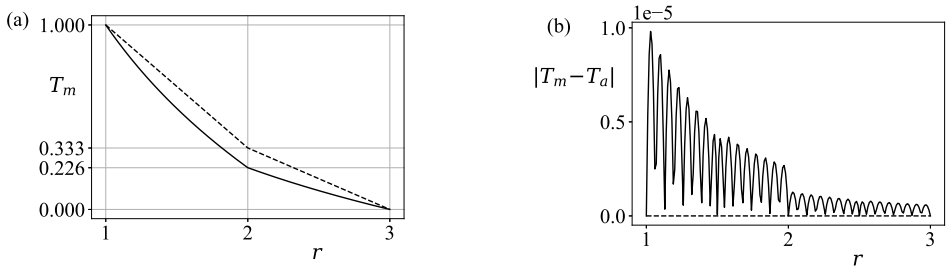

Figure C.3: Temperature profiles (a) and residuals (b) as a function of the radial coordinate r for the Cartesian version of Equation 4.5 (dashed line) and the cylindrical version (solid line) computed by the conduction validation model. The analytically computed values for the temperature field at $r = 2$ ($T(x = 2) = 0.333$ for the Cartesian case and $T(r = 2) = 0.226$ for the cylindrical case) are marked on the y-axis of (a). In (b), the absolute value of the difference between the analytic solution (T_a) and the finite element method model (T_m) are shown. For the Cartesian case, this difference is zero for all values of r because the finite element method linear approximation is exact when the analytic solution itself is linear.

APPENDIX D

RAYLEIGH-BÉNARD CONVECTION PLUME ANALYSIS TUTORIAL

In this chapter, I provide a tutorial for analyzing plumes in RBC in the Schatz lab. This tutorial will be very specific, written for the person who will continue this work. It will cover where to find raw experimental and simulation data that has already been generated, where to find metadata for these data sets such as fluid parameters and temperature settings, where to find tutorials for running the RBC experimental cell, how to pre-process experimental and simulated image data, how to compute the persistent homology of simulated and experimental image data, and how to detect plumes from persistent homology. If a file is referenced in this tutorial, assume it is located in the `RBC_tutorial_files` folder in `F:/RBC_data` on the RBC Windows 10 computer or in the RBC tutorials Teams folder.

D.1 Experimental and simulated data

D.1.1 Location of existing experimental data and pre-processing

The raw experimental data used in this thesis are located in the external hard drive on the RBC Windows 10 computer in the Schatz lab (W303). At the time of writing, this drive is labeled `F:` on the RBC computer and the relevant folder is `RBC_data`. Alternatively, they are backed up in the Schatz / Grigoriev Teams group in the Files tab in the Experimentalists channel. The relevant folder is `Rayleigh-Benard convection tutorial files`. A third backup is located on the drive `D:` in the folder `RBC` on the Windows 7 RBC machine. The login info to this machine can be found in Brett Tregoning's small red lab notebook in the Schatz lab. This last location is the master location for raw experimental data and contains many experimental runs. The metadata for the experimental data is located in a spreadsheet called `data_summary.xlsx` located in `RBC_tutorial_files` and

the master location for raw experimental data. This spreadsheet contains information like the date of data runs and the fluid properties and temperature settings during those runs.

For experimental data (Table 4.1), the folder locations in `RBC_data` are as follows:

```
Experiment 1 is located in
"2016_06_03_turb/run_01n02/0603180728_5fps/filt_sigma_03"
and
"2016_06_03_turb/run_01n02/0603190946_5fps/filt_sigma_03".
These represent two data runs taken
in close proximity and with similar
parameter values that are
subsequently combined. The raw
videos and raw temperature data for
both runs are found in
"2016_06_03_turb/run_01n02".
There is a similar pattern of file
locations for every other
experimental run.  Raw videos are
located in the "run_..." parent
folder and then raw images and
subsequently processed images are
located in the subfolders.
Experiment 2 is located in
"2016_06_03_turb/run_03n04/0603205324_5fps/filt_sigma_03"
and
"2016_06_03_turb/run_03n04/0603221015_5fps/
processed/filt_sigma_03".
Experiment 3 is located in
```

"2016_05_17_turb/run_03n04/0517200707_5fps/filt_sigma_03"
and
"2016_05_17_turb/run_03n04/0517211056_5fps/filt_sigma_03".
Experiment 4 is located in
"2016_05_20_turb/run_07n08n09/0520220719_5fps/filt_sigma_03",
"2016_05_20_turb/run_07n08n09/0520232612_5fps/filt_sigma_03",
and
"2016_05_20_turb/run_07n08n09/0521011911_5fps/filt_sigma_03"
(this experiment is a combination of 3 runs).
Experiment 5 is located in
"2016_06_03_turb/run_05n06/0603234837_5fps/
processed/filt_sigma_03"
and
"2016_06_03_turb/run_05n06/0604004838_5fps/filt_sigma_03".
Experiment 6 is located in
"2016_06_04_turb/run_01/0604151614_5fps/filt_sigma_03".
Experiment 7 is located in
"2016_05_21_turb/run_01n02/0521172914_5fps/filt_sigma_03"
and
"2016_05_21_turb/run_01n02/0521182916_5fps/filt_sigma_03".
Experiment 8 is located in
"2016_06_04_turb/run_02n03/0604164444_5fps/filt_sigma_03"
and
"2016_06_04_turb/run_02n03/0604174453_5fps/filt_sigma_03".
Experiment 9 is located in
"2016_05_22_turb/run_01n02/0522070903_5fps/
processed/filt_sigma_03"

and

"2016_05_22_turb/run_01n02/0522081248_5fps/filt_sigma_03."

Images are extracted from the videos at 5 frames per second (roughly 10 frames per vertical diffusion time using the program "Video to Picture Image Converter 3", but any program that extracts frames from a video at constant framerate will do.

The raw images are pre-processed after being extracted from the videos. The pre-processing is explained in videos in the "Data processing (Skype videos)" slide of the "RBC manual" powerpoint presentation. MATLAB scripts that largely automate this process are available in the "MATLAB pre-processing code" subfolder of `RBC_tutorial_files`. The master file that calls the other relevant pre-processing steps is

`shift_through_fourier_filter_master_old`. To run this, you first need to have an image of quiescent RBC to use as a background and then use it to create a mask. The procedure for this is covered in the previously mentioned tutorial videos.

The data pre-processing tutorial videos also go over how the fluid parameters of the RBC experiments are obtained using C code from the Santa Barbara Pattern Formation group. This code can be found on the RBC Windows 10 machine in the folder

`F:/RBC_data/parameters/ahlers_c_code`. From there, Bala's instructions in the video can be followed to obtain fluid parameters or make changes if necessary. To simply run the code as it currently exists, use the command `./RBC` from the previously specified folder. Then, enter the required fluid parameters.

Once you have obtained pre-processed experimental image data, you are ready to process it using persistent homology. The procedure for doing this is discussed in a later section.

D.1.2 Location of existing simulated data and pre-processing

The raw simulated data used in this thesis are located in the external hard drive on the RBC Windows 10 computer in the Schatz lab (W303). At the time of writing, this drive is labeled

”F:” on the RBC computer and the relevant folder is `RBC_data`. Alternatively, they are backed up in the Schatz / Grigoriev Teams group in the Files tab in the Experimentalists channel. The relevant folder is ”Rayleigh-Benard convection tutorial files”.

For simulated data (Table 4.3), the folder locations in `RBC_data` are as follows:

```
Simulation 1 is located in
"Dimensional_Bous/dimBouss_data/ims/".
Simulation 2 is located in
"simulation_3500/nondim_r3500txt.tar/nondim_r3500txt/ims/"
and
"simulation_3500_2/set2_nondim_Ra3500/ims/".
Simulation 3 is located in
"Bouss_corrected_Tr_Ra_4000/new_prof_r4000/ims/"
and
"Ra4000_radial_restart/ra4000_restart1.tar/
ra4000_restart1/ra4000_restart1/ims/".
Simulation 4 is located in
"Bouss_corrected_Tr_Ra_4800/newprof_r4800/ims/"
and
"Ra4800_radial_restart/Ra4800_restart1.tar/
Ra4800_restart1/Ra4800_restart1/ims/".
```

The only pre-processing of simulated data that needs to be done is reflection of the gray scale intensity values about the mean point (in typical 8-bit gray scale, this is done by taking the negative of the gray scale intensity values, then adding 255). This is to ensure that dark regions represent hot upflows and bright regions represent cold downflows in both experiments and simulations.

D.2 Sample code for computing persistent homology

The following subsections contain code for computing persistent homology in 1-D functions and 2-D image data. A complete IPython notebook containing this code and other useful tutorial code called "Persistent Homology Tutorial.ipynb" is available in

`RBC_tutorial_files`. The sample data can also be found there in the subfolder `data`. This IPython notebook demonstrates how to start with a pre-processed experiment or simulation 2-D image data file and detect plumes in the image. A program like Jupyter Notebook is required to run it.

The Python libraries required to run this code are mostly standard, with the exception of the GUDHI library for computing persistent homology. The GUDHI documentation and installation instructions can be found at `https://gudhi.inria.fr/`.

It is most convenient to install the latest version of Anaconda, following the instructions at

`https://docs.anaconda.com/anaconda/install/index.html`.

This will automatically install Jupyter Notebook as well. Then, GUDHI can be installed using conda by following the instructions at

`https://gudhi.inria.fr/python/latest/installation.html`.

Once installed, GUDHI can be imported like any other Python library when running Jupyter Notebook. This is demonstrated in the sample code.

D.2.1 Code for computing sublevel persistence points of a 1-D function

```
#sample Python code for the computation of
#persistent homology for a 1-D function
import numpy as np
from gudhi import PeriodicCubicalComplex
```

```
#definition of the domain for the function f(x)
x = np.linspace(0,2*np.pi,100)
 #definition of the function f(x)
f = np.sin(x+1)+np.sin(2*x+2)

def create_gudhi_input(array):
    #this helper function formats the
    #input in a way GUDHI will accept
    #it works with 1D functions and
    #2D image arrays
    #note that it essentially "flattens"
    #the input array with a Fortran-like
    #index order, i.e.
    #the first dimension changes fastest
    #and the last dimension changes slowest
    try:
        N = array.shape[0]*array.shape[1]
        gudhi_input = np.zeros((N))
        gudhi_input[:] = np.reshape(array,
                                    N,
                                    order='F')
    except IndexError:
        N = array.shape[0]
        gudhi_input = np.zeros((N))
        gudhi_input[:] = np.reshape(array,
                                    N,
                                    order='F')
```

```
    return gudhi_input

#compute the sublevel persistent homology of f(x)

#format f into a GDUHI-friendly input
f_gudhi = create_gudhi_input(f)
#define the periodic cubical complex
#of which the persistent homology will
#be computed
#periodic_dimensions accepts a list of
#Booleans corresponding to which dimensions
#given to the "dimensions" argument should
#be treated as periodic
#top_dimensional_cells takes the GUDHI-friendly
#input of f(x)
f_complex = PeriodicCubicalComplex(dimensions= f.shape,
                   top_dimensional_cells =f_gudhi,
                   periodic_dimensions=[True])
#compute the persistence of the periodic cubical
#complex, starting at the minimum value of f(x)
f_pers = f_complex.persistence(min_persistence=np.amin(f))
#return the birth and death values
#0-dimensional topological structures
#(distinct connected components) as a N row,
#2 column numpy array
#where there are N persistence points
f_pers_0 = f_complex.persistence_intervals_in_dimension(0)
```

```
#remove trivial persistence points that lie on the diagonal
f_pers_0 = f_pers_0[f_pers_0[:,0]!=f_pers_0[:,1]]
#print the persistence points (-0.369,0.369) and (-1.758, inf)
print(f_persistence_0)
```

D.2.2 Code for computing superlevel persistence points of a 1-D function

```
#sample Python code for the computation of
#persistent homology for a 1-D function
import numpy as np
from gudhi import PeriodicCubicalComplex

#definition of the domain for the function f(x)
x = np.linspace(0,2*np.pi,100)
 #definition of the function f(x)
f = np.sin(x+1)+np.sin(2*x+2)
#reflect f for computation of superlevel
#persistent homology
f = -f

def create_gudhi_input(array):
    #this helper function formats the
    #input in a way GUDHI will accept
    #it works with 1D functions and
    #2D image arrays
    #note that it essentially "flattens"
    #the input array with a Fortran-like
    #index order, i.e.
```

```
    #the first dimension changes fastest
    #and the last dimension changes slowest
    try:
        N = array.shape[0]*array.shape[1]
        gudhi_input = np.zeros((N))
        gudhi_input[:] = np.reshape(array,
                                    N,
                                    order='F')
    except IndexError:
        N = array.shape[0]
        gudhi_input = np.zeros((N))
        gudhi_input[:] = np.reshape(array,
                                    N,
                                    order='F')
    return gudhi_input

#compute the sublevel persistent homology of f(x)

#format f into a GDUHI-friendly input
f_gudhi = create_gudhi_input(f)
#define the periodic cubical complex
#of which the persistent homology will
#be computed
#periodic_dimensions accepts a list of
#Booleans corresponding to which dimensions
#given to the "dimensions" argument should
#be treated as periodic
```

```
#top_dimensional_cells takes the GUDHI-friendly
#input of f(x)
f_complex = PeriodicCubicalComplex(dimensions= f.shape,
                    top_dimensional_cells =f_gudhi,
                    periodic_dimensions=[True])
#compute the persistence of the periodic cubical
#complex, starting at the minimum value of f(x)
f_pers = f_complex.persistence(min_persistence=np.amin(f))
#return the birth and death values
#0-dimensional topological structures
#(distinct connected components) as a N row,
#2 column numpy array
#where there are N persistence points
f_pers_0 = f_complex.persistence_intervals_in_dimension(0)
#remove trivial persistence points that lie on the diagonal
f_pers_0 = f_pers_0[f_pers_0[:,0]!=f_pers_0[:,1]]
#print the reflected persistence points
#(0.369,-0.369) and (1.758, -inf)
print(-f_persistence_0)
```

D.2.3 Code for computing sublevel persistence points of 2-D RBC image data and detecting plumes

```
import numpy as np
from gudhi import CubicalComplex
from PIL import Image
import matplotlib.pyplot as plt
data_files_loc = 'D:\Research_Writing/Thesis/figures/data/'
```

```
#read the image
exp_img = Image.open(data_files_loc
                     + '000131_exp.bmp')
#represent image as numpy array
exp_arr = np.array(exp_img)

#compute the sublevel persistent homology of the image

#define the cubical complex
#of which the persistent homology will
#be computed
#top_dimensional_cells takes the GUDHI-friendly
#input of the image
#dimensions shape of the image in number of pixels
#img_complex = CubicalComplex(dimensions= exp_arr.shape,
#                     top_dimensional_cells =img_gudhi)
img_complex = CubicalComplex(top_dimensional_cells=exp_arr)
#compute the persistence of the cubical
#complex, starting at the minimum value of the image
img_pers = img_complex.persistence()
#return the birth and death values of
#0-dimensional topological structures
#(distinct connected components) as a N row,
#2 column numpy array
#where there are N persistence points
img_pers_0 = img_complex.persistence_intervals_in_dimension(0)
#return the birth and death values  of
```

```
#1-dimensional topological structures
#(holes) as a N row,
#2 column numpy array
#where there are N persistence points
img_pers_1 = img_complex.persistence_intervals_in_dimension(1)
#now, the persistence points can be plotted
#plots are displayed in a cell beneath
img_sub_0 = img_pers_0
img_sub_1 = img_pers_1
#now we will detect a plume in the persistence data
#first, a GUDHI function retrieves the birth and
#death locations for all persistence points in
#the image
img_sub_locs = img_complex.cofaces_of_persistence_pairs()
#now we limit the relevant points to holes
img_sub_locs = img_sub_locs[0][1]
#define the plume region with a list of the form:
#[lower cutoff, upper cutoff, distance from diagonal,
#mean gray scale value]
plume_reg_params = [55,135,10,87]
#take only persistence points that are in the plume region
#for sublevel plume, the lower cutoff is not relevant
#the death value must be <= plume_reg_params[1]
#the point must be >= plume_reg_params[2]
#from the diagonal
#the birth value must be <= plume_reg_params[3]
plumes_sub = img_sub_1[
```

```
    np.logical_and(
    np.logical_and(
    img_sub_1[:,1]<=plume_reg_params[1],
    img_sub_1[:,1]-img_sub_1[:,0]>=plume_reg_params[2]),
    img_sub_1[:,0]<=plume_reg_params[3]),:]
#the locations of persistence points that we kept by
#the above critera are also kept
plumes_sub_locs = img_sub_locs[
    np.logical_and(
    np.logical_and(
    img_sub_1[:,1]<=plume_reg_params[1],
    img_sub_1[:,1]-img_sub_1[:,0]>=plume_reg_params[2]),
    img_sub_1[:,0]<=plume_reg_params[3]),:]
#get the death locations of the sublevel plumes
#we take the death locations because the birth
#locations are at some location around the plume,
#not on the plume itself
death_sub_locs = np.zeros(plumes_sub_locs.shape)
for i in range(plumes_sub_locs.shape[0]):
    #the indicies returned by GUDHI are
    #indices of the flattened array
    #so we use a numpy function to transform
    #back to the 2-D array
    death_sub_locs[i,0] = np.unravel_index(
    plumes_sub_locs[i,:],exp_arr.shape,order='F')[1][1]
    death_sub_locs[i,1] = np.unravel_index(
    plumes_sub_locs[i,:],exp_arr.shape,order='F')[0][1]
```

```
#throw out any points that are too close to the
#boundaries due to the presence of damped
#convection rolls in this region that are
#easily confused for plumes
ctr = 189
rad_mask= 172
death_sub_locs = death_sub_locs[
(death_sub_locs[:,0]-ctr)**2+(death_sub_locs[:,1]-ctr)**2<=rad_mask**2
    ,:]
#highlight the detected plume's location
#in the temperature field
fig,ax=plt.subplots()
im = plt.imshow(exp_arr,cmap='gray')
plt.scatter(death_sub_locs[:,0],death_sub_locs[:,1],
            s=250,marker='o',facecolors='None',
            edgecolors='m',linewidths=3)
ax.axis('off')
plt.show()
```

D.2.4 Code for computing superlevel persistence points of 2-D RBC image data and detecting plumes

```
import numpy as np
from gudhi import CubicalComplex
from PIL import Image
#read the image
exp_img = Image.open(data_files_loc
                     + '000082_exp.bmp')
```

```
#find max and min of image
exp_img = np.array(exp_img)
exp_img_max = 255
exp_img_min = 0
#reflect and translate the image for superlevel
#filtering represent it as numpy array
exp_arr = exp_img_max+exp_img_min-exp_img
#mask the image so that only the temperature
#field and not the boundaries are reflected
ctr = 189
rad = 182
for i in range(exp_arr.shape[0]):
    for j in range(exp_arr.shape[1]):
        if (i-ctr)**2+(j-ctr)**2>rad**2:
            exp_arr[i,j]=255

#compute the sublevel persistent homology of the image

#define the cubical complex
#of which the persistent homology will
#be computed
img_complex_sup = CubicalComplex(
                top_dimensional_cells =exp_arr)
#compute the persistence of the cubical
#complex
img_pers_sup = img_complex_sup.persistence()
#return the birth and death values of
```

```
#0-dimensional topological structures
#(distinct connected components) as a N row,
#2 column numpy array
#where there are N persistence points
img_pers_0_sup = img_complex_sup.persistence_intervals_in_dimension(0)
#return the birth and death values  of
#1-dimensional topological structures
#(holes) as a N row,
#2 column numpy array
#where there are N persistence points
img_pers_1_sup = img_complex_sup.persistence_intervals_in_dimension(1)
img_sup_locs = img_complex_sup.cofaces_of_persistence_pairs()
img_sup_locs = img_sup_locs[0][1]
#reflect and transform points again to represent
#superlevel persistence points
img_pers_0_sup = 255-img_pers_0_sup
img_pers_1_sup = 255-img_pers_1_sup
#now the persistence points can be plotted
img_sup_0 = img_pers_0_sup
img_sup_1 = img_pers_1_sup

#get the locations of the persistence points
#on the temperature field
img_sup_locs = img_complex_sup.cofaces_of_persistence_pairs()
#take only holes
img_sup_locs = img_sup_locs[0][1]
```

```
#define the plume region with a list of the form:
#[lower cutoff, upper cutoff, distance from diagonal,
#mean gray scale value]
plume_reg_params = [55,135,10,87]
#take only persistence points that are in the plume region
#for superlevel plume, the upper cutoff is not relevant
#the death value must be >= plume_reg_params[0]
#the point must be >= plume_reg_params[2]
#from the diagonal
#the birth value must be >= plume_reg_params[3]
plumes_sup = img_sup_1[
    np.logical_and(
    np.logical_and(
    img_sup_1[:,1]>=plume_reg_params[0],
    img_sup_1[:,0]-img_sup_1[:,1]>=plume_reg_params[2]),
    img_sup_1[:,0]>=plume_reg_params[3]),:]
#take only the locations of holes in the plume region
plumes_sup_locs = img_sup_locs[
    np.logical_and(
    np.logical_and(
    img_sup_1[:,1]>=plume_reg_params[0],
    img_sup_1[:,0]-img_sup_1[:,1]>=plume_reg_params[2]),
    img_sup_1[:,0]>=plume_reg_params[3]),:]
#get the death locations of the superlevel plumes
#we take the death locations because the birth
#locations are at some location around the plume,
#not on the plume itself
```

```
death_sup_locs = np.zeros(plumes_sup_locs.shape)
for i in range(plumes_sup_locs.shape[0]):
    #the indicies returned by GUDHI are
    #indices of the flattened array
    #so we use a numpy function to transform
    #back to the 2-D array
    death_sup_locs[i,0] = np.unravel_index(
    plumes_sup_locs[i,:],exp_arr.shape,order='F')[1][1]
    death_sup_locs[i,1] = np.unravel_index(
    plumes_sup_locs[i,:],exp_arr.shape,order='F')[0][1]
#throw out any points that are too close to the
#boundaries due to the presence of damped
#convection rolls in this region that are
#easily confused for plumes
ctr = 189
rad_mask= 172
death_sup_locs = death_sup_locs[
(death_sup_locs[:,0]-ctr)**2+(death_sup_locs[:,1]-ctr)**2<=rad_mask**2
    ,:]
#highlight the detected plume's location
#in the temperature field
fig,ax=plt.subplots()
im = plt.imshow(exp_img,cmap='gray')
plt.scatter(death_sup_locs[:,0],death_sup_locs[:,1],
            s=250,marker='o',facecolors='None',
            edgecolors='m',linewidths=3)
ax.axis('off')
```

```
plt.show()
```

D.3 Assigning the plume region

In subsection D.2.3 and subsection D.2.4, four parameters were used to classify points in the persistence diagrams as plumes. These include the upper cutoff, the lower cutoff, the mean gray scale value, and a minimum distance from the diagonal ("delta"). The mean gray scale value can be computed from a set of image data and the minimum distance from the diagonal is routinely set to 10 (in 8-bit gray scale units). However, the selection of the upper and lower cutoff often proves to be more difficult. The upper cutoff defines the bound where sublevel holes are too prominent to be classified as plumes, while the lower cutoff defines the bound where superlevel holes are too prominent to be classified as plumes. These values must be carefully selected to include plumes and exclude fully formed convection rolls. However, one advantage is that plumes can be tracked throughout a time series, so a plume only needs to pass through the plume region once to be classified as a plume and then tracked. Thus, the plume region can be as tight as possible to exclude very prominent features while capturing plumes right as they become visible in the image.

Additionally, holes that appear close to the boundaries in experiments, specifically within 15 pixels, are thrown out due to the presence of convection features near the boundaries with damped intensity. This causes features to be spuriously identified as plumes.

Plume regions in the simulated data are usually easily found. Up to now, we have used the same parameters in every simulation run. Those are a delta of 10, a mean integer value of 127, an upper cutoff of 174 and a lower cutoff of 80. Plume regions in the experiments are more difficult. Delta is taken to be 10 and the mean integer is most often 87 and the lower cutoff is most often 55. These are good starting values to form a preliminary plume region. From there, the search can be narrowed. The upper cutoff in experiments shows quite a bit of variability given the differences in optical conditions that can occur, from 119 to 135. It is typical to begin at 135 and the lower the upper cutoff as necessary to crop

out spurious features classified as plumes. To do this, it is best to start with a preliminary plume region and then use it to analyze a time-series of data. Plot the persistence diagrams for holes at each time step and show the plume region in these plots. Simultaneously, produce images of the convection roll patterns with plumes labeled. It will be obvious, when scrolling through a time-series of these images, when spurious plumes are detected and labeled on the temperature field. The birth and death values of these spurious holes can then be used to decrease the size of the plume region.

REFERENCES

[1] M. C. Cross and P. C. Hohenberg, "Pattern formation outside of equilibrium", *Rev. Mod. Phys.*, vol. 65, no. 3, pp. 851–1112, 1993.

[2] A. Turing, "The chemical basis of morphogenesis", *Bull. Math. Biol.*, vol. 52, no. 1-2, pp. 153–197, 1990.

[3] R. J. Field and M. Burger, *Oscillations and traveling waves in chemical systems*. Wiley, 1985.

[4] F. Arecchi, "Space-time complexity in nonlinear optics", *Physica D*, vol. 51, no. 1, pp. 450–464, 1991.

[5] J. R. A. Pearson, "On convection cells induced by surface tension", *J. Fluid Mech.*, vol. 4, no. 5, pp. 489–500, 1958.

[6] E. Dubois-Violette, "Determination of thermal instabilities thresholds for homeotropic and planar nematic liquid crystal samples", *Solid State Commun.*, vol. 14, no. 8, pp. 767–771, 1974.

[7] A. V. Getling, *Rayleigh-Bénard convection*, ser. Advanced series in nonlinear dynamics ; v. 11. 1998, ISBN: 9810226578.

[8] R. J. Donnelly, "Taylor-Couette flow: The early days", *Phys. Today*, vol. 44, no. 11, pp. 32–39, 1991.

[9] I. Marusic and S. Broomhall, "Leonardo Da Vinci and fluid mechanics", *Annu. Rev. Fluid Mech.*, vol. 53, no. 1, pp. 1–25, 2021. eprint: https://doi.org/10.1146/annurev-fluid-022620-122816.

[10] S. T. Wereley and R. M. Lueptow, "Velocity field for Taylor–Couette flow with an axial flow", *Phys. Fluids*, vol. 11, no. 12, pp. 3637–3649, 1999. eprint: https://doi.org/10.1063/1.870228.

[11] J. R. de Bruyn, E. Bodenschatz, S. W. Morris, S. P. Trainoff, Y. Hu, D. S. Cannell, and G. Ahlers, "Apparatus for the study of Rayleigh–Bénard convection in gases under pressure", *Rev. Sci. Instrum.*, vol. 67, no. 6, pp. 2043–2067, 1996.

[12] J. Tithof, B. Suri, R. K. Pallantla, R. O. Grigoriev, and M. F. Schatz, "Bifurcations in a quasi-two-dimensional Kolmogorov-like flow", *J. Fluid Mech.*, vol. 828, pp. 837–866, 2017.

[13] M. C. Krygier, J. L. Pughe-Sanford, and R. O. Grigoriev, "Exact coherent structures and shadowing in turbulent Taylor–Couette flow", *J. Fluid Mech.*, vol. 923, Jul. 2021.

[14] C. Normand, Y. Pomeau, and M. G. Velarde, "Convective instability: A physicist's approach", *Rev. Mod. Phys.*, vol. 49, pp. 581–624, 3 Jul. 1977.

[15] B. Suri, J. Tithof, R. Mitchell, R. O. Grigoriev, and M. F. Schatz, "Velocity profile in a two-layer Kolmogorov-like flow", *Phys. Fluids*, vol. 26, no. 5, p. 053 601, 2014. eprint: https://doi.org/10.1063/1.4873417.

[16] J. Swift and P. C. Hohenberg, "Hydrodynamic fluctuations at the convective instability", *Phys. Rev. A*, vol. 15, pp. 319–328, 1 Jan. 1977.

[17] E. N. Lorenz, "Deterministic nonperiodic flow", *J. Atmos. Sci.*, vol. 20, no. 2, pp. 130–141, 1963.

[18] Y. Kuramoto, "Phase dynamics of weakly unstable periodic structures", *Prog. Theor. Phys.*, vol. 71, no. 6, pp. 1182–1196, 1984.

[19] G. K. Batchelor, *An Introduction to Fluid Dynamics*. Cambridge University Press, 2000.

[20] A. N. Kolmogorov, V. Levin, J. C. R. Hunt, O. M. Phillips, and D. Williams, "The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers", *Proc. R. Soc. A*, vol. 434, no. 1890, pp. 9–13, 1991. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1991.0075.

[21] T. Von Kármán, "Mechanische ahnlichkeit und turbulenz", *Math.-Phys. Klasse*, 1930.

[22] A. K. . F. Hussain, "Coherent structures—reality and myth", *Phys. Fluids*, vol. 26, no. 10, pp. 2816–2850, 1983. eprint: https://aip.scitation.org/doi/pdf/10.1063/1.864048.

[23] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov exponents from a time series", *Physica D*, vol. 16, no. 3, pp. 285–317, 1985.

[24] M. Nagata, "Three-dimensional finite-amplitude solutions in plane Couette flow: Bifurcation from infinity", *J. Fluid Mech.*, vol. 217, pp. 519–527, 1990.

[25] G. Kawahara, M. Uhlmann, and L. van Veen, "The significance of simple invariant solutions in turbulent flows", *Annu. Rev. Fluid Mech.*, vol. 44, no. 1, pp. 203–225, 2012. eprint: https://doi.org/10.1146/annurev-fluid-120710-101228.

[26] N. B. Budanur, D. Borrero-Echeverry, and P. Cvitanović, "Periodic orbit analysis of a system with continuous symmetry—a tutorial", *Chaos*, vol. 25, no. 7, p. 073 112, 2015. eprint: https://doi.org/10.1063/1.4923742.

[27] B. Hof, C. W. H. van Doorne, J. Westerweel, F. T. M. Nieuwstadt, H. Faisst, B. Eckhardt, H. Wedin, R. R. Kerswell, and F. Waleffe, "Experimental observation of nonlinear traveling waves in turbulent pipe flow", *Science*, vol. 305, no. 5690, pp. 1594–1598, 2004. eprint: https://www.science.org/doi/pdf/10.1126/science.1100393.

[28] G. Yalnız and N. B. Budanur, "Inferring symbolic dynamics of chaotic flows from persistence", *Chaos*, vol. 30, no. 3, p. 033 109, 2020. eprint: https://doi.org/10.1063/1.5122969.

[29] J. F. Gibson, J. Halcrow, and P. Cvitanović, "Equilibrium and travelling-wave solutions of plane Couette flow", *J. Fluid Mech.*, vol. 638, pp. 243–266, 2009.

[30] U. Ehrenstein and W. Koch, "Three-dimensional wavelike equilibrium states in plane Poiseuille flow", *J. Fluid Mech.*, vol. 228, pp. 111–148, 1991.

[31] N. B. Budanur, K. Y. Short, M. Farazmand, A. P. Willis, and P. Cvitanović, "Relative periodic orbits form the backbone of turbulent pipe flow", *J. Fluid Mech.*, vol. 833, pp. 274–301, 2017.

[32] F. Busse, "Transition to turbulence in thermal convection with and without rotation", in *Transition and Turbulence*, Elsevier Inc, 1981, pp. 43–61, ISBN: 9781483264592.

[33] P. Manneville, "Chapter 4 - Thermal Convection", in *Dissipative Structures and Weak Turbulence*, P. Manneville, Ed., Boston: Academic Press, 1990, pp. 95–136, ISBN: 978-0-08-092445-8.

[34] D. A. Egolf, I. V. Melnikov, and E. Bodenschatz, "Importance of local pattern properties in spiral defect chaos", *Phys. Rev. Lett.*, vol. 80, no. 15, pp. 3228–3231, 1998.

[35] H. Kurtuldu, K. Mischaikow, and M. F. Schatz, "Extensive scaling from computational homology and Karhunen-Loève decomposition analysis of Rayleigh-Bénard convection experiments", *Phys. Rev. Lett.*, vol. 107, no. 3, pp. 034 503–034 503, 2011.

[36] R. Levanger, M. Xu, J. Cyranka, M. F. Schatz, K. Mischaikow, and M. R. Paul, "Correlations between the leading Lyapunov vector and pattern defects for chaotic Rayleigh-Bénard convection", *Chaos*, vol. 29, no. 5, p. 053 103, 2019. eprint: https://doi.org/10.1063/1.5071468.

[37] T. Kaczynski, K. Mischaikow, and M. Marian, *Computational Homology*, ser. Applied Mathematical Sciences (Springer-Verlag New York Inc.) ; v. 157. New York: Springer, 2004, ISBN: 0387215972.

[38] *Gudhi library*, 2021.

[39] A. Zomorodian and G. Carlsson, "Computing persistent homology", *Discrete Comput. Geom.*, vol. 33, no. 2, pp. 249–274, 2005.

[40] R. Ghrist, "Barcodes: The persistent topology of data", *Bull. Amer. Math. Soc. (N.S.)*, vol. 45, no. 1, pp. 61–75, 2008.

[41] H. Lee, H. Kang, M. K. Chung, B.-N. Kim, and D. S. Lee, "Persistent brain network homology from the perspective of dendrogram", *IEEE Trans. Med. Imaging*, vol. 31, no. 12, pp. 2267–2277, 2012.

[42] K. Xia and G.-W. Wei, "Persistent homology analysis of protein structure, flexibility, and folding", *Int. J. Numer. Methods Biomed. Eng.*, vol. 30, no. 8, pp. 814–844, 2014.

[43] D. Hernández Serrano, J. Hernández-Serrano, and D. Sánchez Gómez, "Simplicial degree in complex networks. Applications of topological data analysis to network science", *Chaos Solitons Fractals*, vol. 137, p. 109 839, 2020.

[44] M. Kramar, A. Goullet, L. Kondic, and K. Mischaikow, "Persistence of force networks in compressed granular media", *Phys. Rev. E*, vol. 87, p. 042 207, 4 Apr. 2013.

[45] K. Krishan, H. Kurtuldu, M. F. Schatz, M. Gameiro, K. Mischaikow, and S. Madruga, "Homology and symmetry breaking in Rayleigh-Bénard convection: Experiments and simulations", *Phys. Fluids (1994)*, vol. 19, no. 11, pp. 117 105–117105-6, 2007.

[46] H. Kurtuldu, K. Mischaikow, and M. F. Schatz, "Measuring the departures from the Boussinesq approximation in Rayleigh–Bénard convection experiments", *J. Fluid Mech.*, vol. 682, pp. 543–557, 2011.

[47] M. Kramár, R. Levanger, J. Tithof, B. Suri, M. Xu, M. Paul, M. F. Schatz, and K. Mischaikow, "Analysis of Kolmogorov flow and Rayleigh-Bénard convection using persistent homology", *Physica D*, vol. 334, pp. 82–98, 2016.

[48] P. Bubenik *et al.*, "Statistical topological data analysis using persistence landscapes.", *J. Mach. Learn. Res.*, vol. 16, no. 1, pp. 77–102, 2015.

[49] H. Adams, T. Emerson, M. Kirby, R. Neville, C. Peterson, P. Shipman, S. Chepushtanova, E. Hanson, F. Motta, and L. Ziegelmeier, "Persistence images: A stable vector representation of persistent homology", *J. Mach. Learn. Res.*, vol. 18, no. 8, pp. 1–35, 2017.

[50] D. Cohen-Steiner, H. Edelsbrunner, J. Harer, and Y. Mileyko, "Lipschitz functions have l p-stable persistence", *Found. Comut. Math.*, vol. 10, no. 2, pp. 127–139, 2010.

[51] H.-l. Yang and G. Radons, "Geometry of inertial manifolds probed via a Lyapunov projection method", *Phys. Rev. Lett.*, vol. 108, p. 154 101, 15 Apr. 2012.

[52] X. Ding, H. Chaté, P. Cvitanovi ć, E. Siminos, and K. A. Takeuchi, "Estimating the dimension of an inertial manifold from unstable periodic orbits", *Phys. Rev. Lett.*, vol. 117, p. 024 101, 2 Jul. 2016.

[53] B. Suri, J. Tithof, R. O. Grigoriev, and M. F. Schatz, "Forecasting fluid flows using the geometry of turbulence", *Phys. Rev. Lett.*, vol. 118, p. 114 501, 11 Mar. 2017.

[54] D. Viswanath, "Recurrent motions within plane Couette turbulence", *J. Fluid Mech.*, vol. 580, pp. 339–358, 2007.

[55] P. Cvitanović and J. Gibson, "Geometry of the turbulence in wall-bounded shear flows: Periodic orbits", *Physica Scripta*, vol. 2010, no. T142, p. 014 007, 2010.

[56] J. S. Park and M. D. Graham, "Exact coherent states and connections to turbulent dynamics in minimal channel flow", *J. Fluid Mech.*, vol. 782, pp. 430–454, 2015.

[57] R. R. Kerswell and O. R. Tutty, "Recurrence of travelling waves in transitional pipe flow", *J. Fluid Mech.*, vol. 584, pp. 69–102, 2007.

[58] A. P. Willis, P. Cvitanović, and M. Avila, "Revealing the state space of turbulent pipe flow by symmetry reduction", *J. Fluid Mech.*, vol. 721, pp. 514–540, 2013.

[59] A. de Lozar, F. Mellibovsky, M. Avila, and B. Hof, "Edge state in pipe flow experiments", *Phys. Rev. Lett.*, vol. 108, p. 214 502, 21 May 2012.

[60] P. Cvitanović, R. L. Davidchack, and E. Siminos, "On the state space geometry of the Kuramoto–Sivashinsky flow in a periodic domain", *SIAM J. Appl. Dyn. Syst.*, vol. 9, no. 1, pp. 1–33, 2010. eprint: https://doi.org/10.1137/070705623.

[61] N. B. Budanur and P. Cvitanović, "Unstable manifolds of relative periodic orbits in the symmetry-reduced state space of the Kuramoto–Sivashinsky system", *J. Stat. Phys.*, vol. 167, no. 3-4, pp. 636–655, 2017.

[62] S. W. Morris, E. Bodenschatz, D. S. Cannell, and G. Ahlers, "Spiral defect chaos in large aspect ratio Rayleigh-Bénard convection", *Phys. Rev. Lett.*, vol. 71, pp. 2026–2029, 13 Sep. 1993.

[63] A. Pellew and R. V. Southwell, "On maintained convective motion in a fluid heated from below", *Proc. R. Soc. A*, vol. 176, no. 966, pp. 312–343, 1940.

[64] R. Clever and F. Busse, "Transition to time-dependent convection", *J. Fluid Mech.*, vol. 65, no. 4, pp. 625–645, 1974.

[65] F. H. Busse and R. M. Clever, "Instabilities of convection rolls in a fluid of moderate prandtl number", *Journal of Fluid Mechanics*, vol. 91, no. 2, pp. 319–335, 1979.

[66] H.-W. Xi, J. D. Gunton, and J. Vinals, "Spiral defect chaos in a model of Rayleigh-Bénard convection", *Phys. Rev. Lett.*, vol. 71, no. 13, p. 2030, 1993.

[67] W. Decker, W. Pesch, and A. Weber, "Spiral defect chaos in Rayleigh-Bénard convection", *Phys. Rev. Lett.*, vol. 73, no. 5, p. 648, 1994.

[68] E. Bodenschatz, W. Pesch, and G. Ahlers, "Recent developments in Rayleigh-Bénard convection", *Annu. Rev. Fluid Mech.*, vol. 32, no. 1, pp. 709–778, 2000.

[69] R. V. Cakmur, D. A. Egolf, B. B. Plapp, and E. Bodenschatz, "Bistability and competition of spatiotemporal chaotic and fixed point attractors in Rayleigh-Bénard convection", *Phys. Rev. Lett.*, vol. 79, no. 10, p. 1853, 1997.

[70] H. Riecke and S. Madruga, "Geometric diagnostics of complex patterns: Spiral defect chaos", *Chaos*, vol. 16, no. 1, p. 013 125, 2006.

[71] N. Becker and G. Ahlers, "Local wave director analysis of domain chaos in Rayleigh–Bénard convection", *J. Stat. Mech.*, vol. 2006, no. 12, P12002–P12002, 2006.

[72] D. Ruelle, "Large volume limit of the distribution of characteristic exponents in turbulence", *Commun. Math. Phys.*, vol. 87, no. 2, pp. 287–302, 1982.

[73] J.-P. Eckmann, "Ergodic theory of chaos and strange attractors", *Rev. Mod. Phys.*, vol. 57, no. 3, pp. 617–656, 1985.

[74] J. L. Kaplan and J. A. Yorke, "Chaotic behavior of multidimensional difference equations", in *Functional Differential Equations and Approximation of Fixed Points*, Springer, 1979, pp. 204–227.

[75] J. D. Farmer, E. Ott, and J. A. Yorke, "The dimension of chaotic attractors", *Physica D*, vol. 7, no. 1-3, pp. 153–180, 1983.

[76] A. Karimi and M. R. Paul, "Quantifying spatiotemporal chaos in Rayleigh-Bénard convection", *Phys. Rev. E*, vol. 85, no. 4 Pt 2, pp. 046 201–046 201, 2012.

[77] D. A. Egolf, R. E. Ecke, W. Pesch, and I. V. Melnikov, "Mechanisms of extensive spatiotemporal chaos in Rayleigh-Bénard convection", *Nature (London)*, vol. 404, no. 6779, pp. 733–736, 2000.

[78] J. D. Scheel and M. C. Cross, "Lyapunov exponents for small aspect ratio Rayleigh-Bénard convection", *Phys. Rev. E*, vol. 74, no. 6 Pt 2, pp. 066 301–066 301, 2006.

[79] *Computational homology project*, 2004.

[80] G. I. Taylor, "Statistical theory of turbulence", *Proc. R. Soc. A*, vol. 151, no. 873, pp. 421–444, 1935.

[81] S. K. Robinson, "Coherent motions in the turbulent boundary layer", *Annu. Rev. Fluid Mech.*, vol. 23, no. 1, pp. 601–639, 1991.

[82] J. F. Gibson, J. Halcrow, and P. Cvitanović, "Visualizing the geometry of state space in plane Couette flow", *J. Fluid Mech.*, vol. 611, pp. 107–130, 2008.

[83] F. Waleffe, "Exact coherent structures in channel flow", *J. Fluid Mech.*, vol. 435, pp. 93–102, 2001.

[84] B. Suri, L. Kageorge, R. O. Grigoriev, and M. F. Schatz, "Capturing turbulent dynamics and statistics in experiments with unstable periodic orbits", *Phys. Rev. Lett.*, vol. 125, p. 064 501, 6 Aug. 2020.

[85] E. Hopf, "A mathematical example displaying features of turbulence", *Commun. Pure Appl. Math.*, vol. 1, no. 4, pp. 303–322, 1948. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpa.3160010401.

[86] L. van Veen, A. Vela-Martín, and G. Kawahara, "Time-periodic inertial range dynamics", *Phys. Rev. Lett.*, vol. 123, p. 134 502, 13 Sep. 2019.

[87] H. Edelsbrunner and J. Harer, *Computational Topology: An Introduction*. Am. Math. Soc., 2010.

[88] A. K. Kassam and L. N. Trefethen, "Fourth-order time stepping for stiff PDEs", *SIAM J. Sci. Comput.*, vol. 26, pp. 1214–1233, 2005.

[89] H. D. Abarbanel, R. Brown, J. J. Sidorowich, and L. S. Tsimring, "The analysis of observed chaotic data in physical systems", *Rev. Mod. Phys.*, vol. 65, no. 4, p. 1331, 1993.

[90] A. V. Pertsov, R. Salomonsz, J. Jalife, W. Baxter, and J. M. Davidenko, "Stationary and drifting spiral waves of excitation in isolated cardiac muscle", *Nature (London)*, vol. 355, no. 6358, pp. 349–351, 1992.

[91] C. Vidal, A. Pagola, J. M. Bodet, P. Hanusse, and E. Bastardie, "Etude expérimentale statistique des structures cibles de la réaction de Belousov-Zhabotinsky en régime oscillant", *J. Phys. France*, vol. 47, no. 11, pp. 1999–2009, 1986.

[92] F. H. Busse, "Non-linear properties of thermal convection", *Rep. Prog. Phys.*, vol. 41, no. 12, pp. 1929–1967, 1978.

[93] A. V. Getling, "Hydrodynamic instabilities and photospheric structures", *Astron. Rep.*, vol. 44, no. 1, pp. 56–63, 2000.

[94] X. L. Qiu and P. Tong, "Onset of coherent oscillations in turbulent Rayleigh-Bénard convection", *Phys. Rev. Lett.*, vol. 87, no. 9, pp. 094 501–094 501, 2001.

[95] M. R. Paul, M. I. Einarsson, P. F. Fischer, and M. C. Cross, "Extensive chaos in Rayleigh-Bénard convection", *Phys. Rev. E*, vol. 75, no. 4 Pt 2, pp. 045 203–045 203, 2007.

[96] M. Xu and M. R. Paul, "Spatiotemporal dynamics of the covariant Lyapunov vectors of chaotic convection", *Phys. Rev. E*, vol. 97, no. 3-1, pp. 032 216–032 216, 2018.

[97] ——, "Covariant Lyapunov vectors of chaotic Rayleigh-Bénard convection", *Phys. Rev. E*, vol. 93, no. 6, pp. 062 208–062 208, 2016.

[98] A. Duggleby and M. R. Paul, "Computing the Karhunen–Loève dimension of an extensively chaotic flow field given a finite amount of data", *Comput. & Fluids*, vol. 39, no. 9, pp. 1704–1710, 2010.

[99] R. W. Walden and G. Ahlers, "Non-Boussinesq and penetrative convection in a cylindrical cell", *J. Fluid Mech.*, vol. 109, pp. 89–114, 1981.

[100] M. Xu and M. R. Paul, "Chaotic Rayleigh-Bénard convection with finite sidewalls", *Phys. Rev. E*, vol. 98, p. 012 201, 1 Jul. 2018.

[101] E. M. Sparrow, R. B. Husar, and R. J. Goldstein, "Observations and other characteristics of thermals", *J. Fluid Mech.*, vol. 41, no. 4, pp. 793–800, 1970.

[102] H.-D. Xi, S. Lam, and K.-Q. Xia, "From laminar plumes to organized flows: The onset of large-scale circulation in turbulent thermal convection", *J. Fluid Mech.*, vol. 503, pp. 47–56, 2004.

[103] B. A. Puthenveettil and J. H. Arakeri, "Plume structure in high-Rayleigh-number convection", *J. Fluid Mech.*, vol. 542, no. 1, pp. 217–249, 2005.

[104] O. Shishkina and C. Wagner, "Analysis of sheet-like thermal plumes in turbulent Rayleigh–Bénard convection", *J. Fluid Mech.*, vol. 599, pp. 383–404, 2008.

[105] J. Bosbach, S. Weiss, and G. Ahlers, "Plume fragmentation by bulk interactions in turbulent Rayleigh-Bénard convection", *Phys. Rev. Lett.*, vol. 108, no. 5, pp. 054 501–054 501, 2012.

[106] A. Pandey, J. D. Scheel, and J. Schumacher, "Turbulent superstructures in Rayleigh-Bénard convection", *Nature Commun.*, vol. 9, no. 1, pp. 2118–11, 2018.

[107] E. Vitral, S. Mukherjee, P. H. Leo, J. Viñals, M. R. Paul, and Z.-F. Huang, "Spiral defect chaos in Rayleigh-Bénard convection: Asymptotic and numerical studies of azimuthal flows induced by rotating spirals", *Phys. Rev. Fluids*, vol. 5, p. 093 501, 9 Sep. 2020.

[108] *NEK: A fast and scalable high-order solver for computational fluid dynamics*. https://nek5000.mcs.anl.gov/, 2021.

[109] M. O. Deville, P. F. Fischer, P. F. Fischer, E. Mund, *et al.*, *High-Order Methods for Incompressible Fluid Flow*. Cambridge University Press, 2002, vol. 9.

[110] K.-H. Chiam, M. R. Paul, M. C. Cross, and H. S. Greenside, "Mean flow and spiral defect chaos in Rayleigh-Bénard convection", *Phys. Rev. E*, vol. 67, p. 056 206, 5 May 2003.

[111] F. H. Busse, "The stability of finite amplitude cellular convection and its relation to an extremum principle", *J. Fluid Mech.*, vol. 30, no. 4, pp. 625–649, 1967.

[112] M. Allili, K. Mischaikow, and A. Tannenbaum, "Cubical homology and the topological classification of 2d and 3d imagery", in *Proceedings 2001 International Conference on Image Processing (Cat. No.01CH37205)*, vol. 2, IEEE, 2001, 173–176 vol.2, ISBN: 0780367251.

[113] *TDA Image Analysis Project*, https://github.com/rachellevanger/tda-image-analysis, 2018.

[114] F. H. Busse and J. A. Whitehead, "Instabilities of convection rolls in a high Prandtl number fluid", *J. Fluid Mech.*, vol. 47, no. 2, pp. 305–320, 1971.

[115] V. Steinberg, G. Ahlers, and D. S. Cannell, "Pattern formation and wave-number selection by Rayleigh-Bénard convection in a cylindrical container", *Physica Scripta*, vol. T9, pp. 97–110, Jan. 1985.

[116] G. Ahlers, M. C. Cross, P. C. Hohenberg, and S. Safran, "The amplitude equation near the convective threshold: Application to time-dependent heating experiments", *J. Fluid Mech.*, vol. 110, pp. 297–334, 1981.

[117] G. W. Corder and D. I. Foreman, *Nonparametric statistics: a step-by-step approach*, 2nd ed. Somerset: Wiley, 2014, ISBN: 1118840313.

[118] M. R. Paul, M. C. Cross, P. F. Fischer, and H. S. Greenside, "Power-law behavior of power spectra in low Prandtl number Rayleigh-Bénard convection", *Phys. Rev. Lett.*, vol. 87, p. 154 501, 15 Sep. 2001.

[119] http://www.matweb.com/search/datasheet.aspx?matguid=6be926d8eb0842abbfb80b5658ade95a&ckck=1, 1999.

[120] https://www.azom.com/properties.aspx?ArticleID=1721, 2002.

[121] https://www.crystran.co.uk/optical-materials/zinc-selenide-znse, 2012.