University of California
Santa Barbara

Model Reduction for Treatment of Neurological Diseases and Cardiac Arrhythmias

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

Doctor of Philosophy
in
Mechanical Engineering

by

Daniel David Wilson

Committee in charge:

Professor Jeff Moehlis, Chair
Professor Frederic Gibou
Professor Igor Mezić
Professor João Hespanha

June 2016

The Dissertation of Daniel David Wilson is approved.

Professor Frederic Gibou

Professor Igor Mezić

Professor João Hespanha

Professor Jeff Moehlis, Committee Chair

April 2016

Model Reduction for Treatment of Neurological Diseases and Cardiac Arrhythmias

by

Daniel David Wilson

To my wife, Jackie.

Acknowledgements

Foremost, I would like to thank my advisor, Jeff. This dissertation would not have been possible without his support and mentorship. Jeff's mastery of all aspects of academia have made the years of work this document represents thoroughly enjoyable. I also appreciate the suggestions and advice of committee members Igor Mezić and Frederic Gibou during my candidacy.

I would like to thank the collaborators who I have had the opportunity to work with during this time. I extend gratitude to Tay Netoff and Abbey Holt at the University of Minnesota and I appreciate their openness in sharing their expertise in experimental collaborations in the field of Neuroscience. Likewise, Ken Showalter and Mark Tinsley at the West Virginia University have been more than generous in sharing their time and resources in coupled oscillator experiments. I am also grateful for discussions with Flavio Fenton at Georgia Tech as well as Elizabeth Cherry and Niels Otani at Rochester Institute of Technology which have all inspired research directions with applications to cardiology.

Perhaps most importantly, I thank my wife Jackie, my strongest supporter throughout this endeavor. I also appreciate the support and encouragement received from our respective parents as well as from friends we have met during our time in California.

Curriculum Vitæ

Daniel David Wilson

Education

2016	Ph.D. in Mechanical Engineering, University of California, Santa Barbara.
2011	B.S. in Mechanical Engineering, University of Washington.
2011	B.A. in Natural and Mathematical Sciences, Whitman College.

Publications

- A. B. Holt, D. Wilson, M. Shinn, J. Moehlis, and T. I. Netoff. Closed-loop approach to tuning deep brain stimulation parameters for Parkinson's disease. In review at *PLoS Computational Biology*.
- D. Wilson and J. Moehlis. Clustered desynchronization from high-frequency deep brain stimulation. *PLoS Computational Biology*. 11(12): Art. No. e1004673, 2015.
- R. Snari, M. Tinsley, D. Wilson, S. Faramarzi, T. I. Netoff, J. Moehlis, and K. Showalter. Desynchronization of stochastically synchronized chemical oscillators. *Chaos*. 25(12): Art. No. 123116, 2015.
- D. Wilson and J. Moehlis. Analytical bounds on the critical coupling strength in a population of heterogeneous biological oscillators. To appear: *Proceedings of the 2016 American Control Conference*.
- D. Wilson and J. Moehlis. Determining individual phase response curves from aggregate population data. *Physical Review E*. 92: Art. No. 022902, 2015.
- D. Wilson, A.B. Holt, T. Netoff, and J. Moehlis. Optimal entrainment of heterogeneous noisy neurons. *Frontiers in Neuroscience*. 9: Art. No. 192, 2015.
- D. Wilson and J. Moehlis. Towards a more efficient implementation of low energy anti-fibrillation pacing. In review at *PLoS One*.
- D. Wilson and J. Moehlis. Extending phase reduction to excitable media: theory and applications. *SIAM Review*. 57(2): 201-222, 2015.
- D. Wilson and J. Moehlis. Isostable reduction with applications to partial differential equations. In review at *Physical Review E*.
- D. Wilson and J. Moehlis. An energy-optimal approach for entrainment of uncertain circadian oscillators. *Biophysical Journal*. 107(7): 1744-1755, 2014.
- D. Wilson and J. Moehlis. A Hamilton-Jacobi-Bellman approach for termination of seizure-like bursting. *Journal of Computational Neuroscience*. 37(2): 345-355, 2014.

- G. S. Schmidt, D. Wilson, F. Allgöwer, and J. Moehlis. Selective averaging with application to phase reduction and neural control. *Nonlinear Theory and Its Applications. IEICE* 5(4): 424-435, 2014.
- D. Wilson and J. Moehlis. An energy-optimal methodology for synchronization of excitable media. *SIAM Journal on Applied Dynamical Systems.* 13(2): 994-957, 2014.
- D. Wilson and J. Moehlis. Locally optimal extracellular stimulation for chaotic desynchronization of neural populations. *Journal of Computational Neuroscience.* 37(2): 243-257, 2014.
- D. Wilson and J. Moehlis. Optimal chaotic desynchronization for neural populations. *SIAM Journal on Applied Dynamical Systems.* 13(1): 276-305, 2014.

Abstract

Model Reduction for Treatment of Neurological Diseases and Cardiac Arrhythmias

by

Daniel David Wilson

For more than half a century, phase reduction techniques have been a remarkably powerful tool to aid in the understanding of oscillatory nonlinear dynamical systems, with a wide array of examples spanning the biological, physical, and chemical sciences. In the first part of this dissertation, we draw on the utility of phase reduction to investigate three problems with neuroscientific applications. Chapter 2 details a strategy to efficiently desynchronize a population of pathologically synchronized limit cycle oscillators which could represent a low energy treatment for Parkinson's disease. Chapter 3 gives theoretical details and experimental validation of a strategy to entrain a noisy and heterogeneous population of limit cycle oscillators with potential applications to the treatment of pathological conditions such as hearing loss and circadian dysfunction. Also, Chapter 4 details a general methodology which could make the implementation of control strategies from Chapters 2 and 3 more feasible in an *in vivo* setting.

While phase reduction has been exceedingly useful for systems with limit cycle solutions, analogous methodologies are not similarly developed for systems with excitable dynamics. The second part of this dissertation (Chapters 5 and 6) remedies this situation by developing a reduction methodology for excitable systems which approach a stable stationary solution. This reduction method relies on the notion of isostables, which are defined as sets of points in phase space that approach a stationary solution together, in a well-defined sense. An adjoint method is derived for calculating infinitesimal isostable response curves and this strategy is used to devise and implement sophisticated control

strategies in complex dynamical systems of neurological and cardiological behavior. Emphasis is given to the problem of eliminating cardiac alternans which has been implicated as a precursor to fibrillation, a leading cause of death in the industrialized world. It is envisioned that this new reduction strategy could be as useful as a phase reduction has been in the past decades as it shows tremendous promise as a tool for the understanding and control of nonlinear systems.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This dissertation investigates model reduction and optimal control techniques in the context of both the single cell and network behavior of neurons and cardiac cells. Motivation for the neuroscience aspects of this work come from Parkinson's disese, a neurological disease characterized by the progressive loss of motor function. Motivation for the cardiological aspects of this work come from the treatment of alternans, the beat-to-beat alternation of electrochemical cardiac dynamics which has been implicated as a precursor to cardiac arrest.

Phase reduction methods have fruitfully applied to many physical, chemical and biological systems [1–12]. Such methods are useful for understanding the dynamics of perturbed nonlinear oscillators because they allow for the reduction of potentially complicated and high dimensional systems of ordinary differential equations (ODEs) and partial differential equations (PDEs) to a system with a single phase variable. Phase reduction is particularly powerful in conjunction with Hamilton-Jacobi-Bellman, calculus of variations, or dynamic programming [13] control frameworks, since the computational effort for finding the desired control input grows exponentially with the number of state variables.

As a concrete example of the power of phase reduction, the first part of this dissertation examines a control strategy with relevance to neurological networks. Pathological synchronization among bursting neurons in the basal ganglia-cortical loop within the brain has been hypothesized to play a contributing role in the tremors seen in patients with Parkinson's disease [14–19]. Deep Brain Stimulation (DBS) is a well-established technique for alleviating these tremors and has been hypothesized to desynchronize these neurons through the injection of a high-frequency, pulsatile input into an appropriate region of the brain [14], [20], [21]. As shown in Chapter 2, phase reduction is an essential step in formulating a numerically tractable control strategy to desynchronize a population of pathologically synchronized neural oscillators in an energy-efficient manner.

The merits of phase reduction extend beyond the theoretical and computational realm, and have also been useful for systems for which the state dynamics are not fully known. For example, *in vitro* experiments on biological tissue have successfully controlled spike timing in periodically firing neurons using inputs computed from phase models [22], [23]. The full neural models are not required in this context. Instead these control strategies only require a measurement of the phase response curve which can be obtained experimentally using the so-called "direct method" [24], [25] (see Appendix B).

Given the wide applicability of phase reduction, it is quite surprising that analogous methodologies are not similarly developed for the study of systems with dynamics that asymptotically approach a fixed point (for ODEs) or a stationary solution (for PDEs). In Chapters 5 and 6, we describe a general method of model reduction for excitable systems based on the notion of isostables [26], defined as sets of points in phase space that approach a fixed point together, in a well-defined sense. Here, a system is said to be excitable if it contains a stable fixed point for which perturbations above a given threshold yield a response that is well beyond that threshold.

One motivation for isostable reduction comes from cardiology, where the muscular

tissue of the heart, the myocardium, is modeled by electrically coupled excitable cells. Under certain conditions, the myocardium can become susceptible to cardiac fibrillation, the uncoordinated contraction of the cardiac muscle, which can be deadly in an otherwise healthy heart if not treated within minutes of onset. Most researchers agree that atrial fibrillation is caused by the presence of unwanted spiral waves [2] within the myocardium which interfere with the normal sinoatrial rhythm [27], [28], [29]. Cardiac fibrillation can be the final step in an increasingly complex series of events that begins with a phenomenon known as cardiac alternans. Alternans is the beat to beat alternation of electrochemical cardiac dynamics at a constant rate of pacing and has been implicated as a precursor to fibrillation [30]. In recent decades, researchers have looked into ways of suppressing alternans as a way of preventing cardiac fibrillation, obviating the need for painful and damaging defibrillating shocks. Many of these methods, such as those described in [31] and [32], attempt to suppress alternans by monitoring the action potential duration (APD), or the length of time that the tissue remains depolarized (i.e., at elevated voltage), and appropriately modifying the time at which the next action potential is elicited. These methods have been shown to be successful in real cardiac tissue [33], [34], [35]. In this dissertation, we will show how isostable reduction can aid in the development of control strategies to eliminate cardiac alternans. While the applications considered in this dissertation mostly pertain to cardiological applications, this reduction strategy could have wide applicability to other physical, chemical and biological systems. We envision that this new reduction strategy could prove to be as useful as phase reduction has been in the past decades, as it shows tremendous promise as a tool for the understanding and control of nonlinear systems.

The outline of this dissertation is as follows. In Chapter 2, we consider the problem of desynchronizing a pathologically synchronized population of neurons. A procedure is developed for finding an energy-optimal stimulus which gives a positive Lyapunov

exponent, and hence desynchronization, for a neural population. Not only does this strategy achieve desynchronization for each model we consider, but it also does so using less energy than recently proposed methods, suggesting a powerful alternative to pulsatile stimuli for deep brain stimulation. Furthermore, we calculate error bounds on the optimal stimulus which will guarantee a minimum Lyapunov exponent and develop a related strategy for desynchronizing neurons based on the population's phase distribution.

In Chapter 3, in collaboration with the Netoff laboratory at the University of Minnesota, we develop a methodology to design an efficient stimulus to entrain nonlinear, noisy limit cycle oscillators with uncertain properties. Conditions are derived which guarantee that the stimulus will entrain the oscillators despite uncertainty in the individual neural parameters. Using these conditions, we devise an energy optimal control strategy for entrainment and apply it to numerical models of noisy phase oscillators and to *in vitro* hippocampal neurons. In both instances, the optimal stimuli outperform other similar but suboptimal entraining stimuli. Because this control strategy explicitly accounts for both noise and inherent uncertainty of model parameters, it could have experimental relevance to neural circuits where robust spike timing plays an important role.

In Chapter 4, we present a simple yet novel methodology to calculate PRCs of individual oscillators using an aggregate signal from a larger population. Current experimental techniques for inferring PRCs require data from individual oscillators, which can be impractical to obtain when the oscillator is part of a much larger population. The proposed methodology is shown to be accurate in the presence of inter-oscillator coupling and noise and can also provide a good estimate of an average PRC of a heterogeneous population. We also find that standard experimental techniques for PRC measurement can produce misleading results when applied to aggregate population data.

In Chapter 5, using the notion of isostables we present a general method for isostable reduction for excitable systems. We also devise an adjoint method for calculating in-

finitesimal isostable response curves, which are analogous to infinitesimal phase response curves for oscillatory systems. Through isostable reduction, we are able to implement sophisticated control strategies in a high-dimensional model of cardiac activity for the termination of alternans. Results based on numerical simulation indicate that this control strategy requires many orders of magnitude less energy to implement than other recently proposed alternans elimination strategies.

In Chapter 6, we present a general method for isostable reduction of partial differential equations, with the potential to reduce the dimensionality of a nonlinear system from infinity to one. We illustrate the utility of this reduction by applying it to two different models with biological relevance. In the first example, isostable reduction of the Fokker-Planck equation provides the necessary framework to design a simple control strategy to desynchronize a population of pathologically synchronized oscillatory neurons. In the second example, the reduction of a partial differential equation describing a ring of cardiac tissue allows one to develop an energy optimal control strategy to eliminate cardiac alternans. We also propose a strategy for measuring isostable response curves *in vitro* which could pave the way for experimental validation of these results.

Supporting Publications This work is supported by results originally documented in the following peer reviewed publications:

- [10] D. Wilson and J. Moehlis. Optimal chaotic desynchronization for neural populations. *SIAM Journal on Applied Dynamical Systems*. 13(1): 276-305, 2014.
- [36] D. Wilson and J. Moehlis. Determining individual phase response curves from aggregate population data. *Physical Review E*. 92: Art. No. 022902, 2015.
- [37] D. Wilson, A.B. Holt, T. Netoff, and J. Moehlis. Optimal entrainment of heterogeneous noisy neurons. *Frontiers in Neuroscience*. 9: Art. No. 192, 2015.

- [38] D. Wilson and J. Moehlis. Extending phase reduction to excitable media: theory and applications. *SIAM Review*. 57(2): 201-222, 2015.

- [39] D. Wilson and J. Moehlis. Isostable reduction with applications to partial differential equations. In review at *Physical Review E*.

Chapter 2

Optimal Chaotic Desynchronization for Neural Populations

2.1 Introduction

Pathological synchronization among spiking neurons in the basal ganglia-cortical loop within the brain is thought to be one factor contributing to the tremors exhibited by patients with Parkinson's disease [40]. Deep Brain Stimulation (DBS), a well-established technique for mitigating these tremors, has been hypothesized to desynchronize these neurons through the injection of a high-frequency, pulsatile input into an appropriate region of the brain [14], [20], [21]. Typically, DBS is implemented with a permanent, high frequency, pulsatile signal which is administered in an open-loop fashion. This has motivated researchers to search for alternative stimuli that consume less energy in order to prolong battery life and to mitigate side effects of DBS such as aggregate tissue damage. Control methods that employ feedback control are of particular interest because they can be used only when needed. For example, in [41], double-pulse stimulation was shown to desynchronize a population of noisy phase oscillators and prevent resynchronization.

Nonlinear, time-delayed feedback is used in [42] to experimentally desynchronize a system of electro-chemical oscillators. In [43], a minimum time desynchronizing control based on phase resetting for a coupled neural network was established using a Hamilton-Jacobi-Bellman approach, which was later extended by [44] to desynchronize neurons using an energy-optimal criterion. In [45], an energy-optimal, charge-balanced stimulus was used to control neural spike timing. More recently, [46] developed a model to control neural networks using a light sensitive protein, instead of electrical stimuli.

In this chapter, we present a procedure for finding an energy-optimal DBS stimulus which exponentially desynchronizes a population of coupled neurons. Unlike the methods in [43] and [44], it does not need the full model for the dynamics, and unlike [47] and [45], it only requires a single input. Methods presented in [48–53] use delayed-feedback stimulation to counter the effects mean field coupling on an ensemble of heterogeneous oscillators, and the method presented in this chapter notably differs from these methods because it is applicable to networks without mean-field coupling, and does not require simultaneous stimulation and measurement. We compare this method to other recently proposed methods, not only to show its desynchronizing capabilities for many types of neural models, but also its ability to do so with comparatively small control inputs. Furthermore, we compute error bounds on the optimal stimulus that will guarantee a resulting signal that will have the required properties to desynchronize a network of neurons, and propose a control strategy for desynchronizing a population based on its phase distribution.

The organization of the chapter is as follows. In Section 2.2, we derive an expression for the Lyapunov exponent for two neurons with similar initial phases and use the result to develop a control method. Section 2.3 uses the optimal stimulus calculated from methods in Section 2.2 to develop sufficient conditions to yield a stimulus with a predetermined guarantee on its Lyapunov exponent. In Section 2.4 we develop a control methodology

based on a population's phase distribution. Section 2.5 examines the control methodology applied to three neural models in order to make comparisons with previous work. Section 2.6 gives concluding comments. This chapter is based on work originally appearing in [10].

2.2 Derivation of the Lyapunov Exponent and Optimal Control

We present a procedure for finding an energy-optimal DBS stimulus which exponentially desynchronizes a population of neurons. An advantage of this approach is that it only requires knowledge of a neuron's phase response curve (PRC), which is experimentally measurable by perturbing an oscillatory neuron at different phases, and determining the change in spike timing (see, for example, Appendix B or [25]). The PRC can also be calculated numerically if all equations and parameters in the neural model are known, e.g. [54]. Through phase reduction, as illustrated, for example, in [54], we can obtain a reduced model for a single neuron of the form

$$\frac{d\theta}{dt} = \omega + Z(\theta)u(t), \tag{2.1}$$

where θ is the phase of the neuron and is 2π-periodic on $[0, 2\pi)$ and, by convention, $\theta = 0$ corresponds to the spiking of the neuron. Here, ω gives the neuron's baseline dynamics which are determined from its natural period T as $\omega = 2\pi/T$, $Z(\theta)$ is the PRC, and $u(t) = I(t)/C$ with $I(t)$ being the control input and $C = 1\mu\text{F/cm}^2$ is the constant neural membrane capacitance.

Lyapunov exponents are commonly used to describe the rate at which nearby trajectories diverge, and have proven useful in other problems, for example, by serving as

biomarkers for seizure prediction and control [55–57]. We now derive an expression for the Lyapunov exponent for (2.1) by considering two identical neurons subject to a common stimulus:

$$\frac{d\theta_i}{dt} = \omega + Z(\theta_i)u(t), \quad i = 1, 2. \tag{2.2}$$

Here we assume that the neurons are nearly in-phase, so that $\theta_1 \approx \theta_2$. Letting $\phi \equiv |\theta_2 - \theta_1|$, we obtain

$$\frac{d\phi}{dt} = Z'(\theta)u(t)\phi + \mathcal{O}(\phi^2). \tag{2.3}$$

Since we have linearized the equation, we assume that solutions are of the form $\phi \sim e^{\Lambda t}$, which yields the finite time Lyapunov exponent (cf. [58])

$$\Lambda(\tau) = \frac{\log(\phi(\tau))}{\tau} = \frac{\int_a^{a+\tau} Z'(\theta(s))u(s)ds}{\tau}. \tag{2.4}$$

We now consider a population of neurons, each described by an equation of the form (2.1). Suppose for some time t_1, for all stimuli $u(t)$ that advance θ from $\theta(0) = 0$ to $\theta(t_1) = \omega t_1$ (that is, stimuli that do not create any net change of phase), we want to find the stimulus that minimizes the cost function $G[u(t)] = \int_0^{t_1}[u(t)^2 - \beta Z'(\theta(t))u(t)]dt$. Here, $\int_0^{t_1}[u(t)^2]dt$ corresponds to the power associated with the stimulus, and $\beta > 0$ is a scaling parameter that determines the relative importance of minimizing the energy versus maximizing the Lyapunov exponent, $\Lambda(t_1)$. We apply calculus of variations to minimize [59]

$$\mathcal{C}[u(t)] = \int_0^{t_1} \left\{ u(t)^2 - \beta Z'(\theta)u(t) + \lambda\left(\frac{d\theta}{dt} - \omega - Z(\theta)u(t)\right)\right\} dt, \tag{2.5}$$

where the Lagrange multiplier λ forces the neural dynamics to obey equation (2.1). The

resulting Euler-Lagrange equations are

$$u(t) = [\beta Z'(\theta) + \lambda Z(\theta)]/2, \tag{2.6}$$

$$\dot{\theta} = Z(\theta)\left[\beta Z'(\theta) + \lambda Z(\theta)\right]/2 + \omega, \tag{2.7}$$

$$\dot{\lambda} = -\left[\beta Z'(\theta) + \lambda Z(\theta)\right]\left[\beta Z''(\theta) + \lambda Z'(\theta)\right]/2, \tag{2.8}$$

where $' = d/d\theta$. To find the optimal control, $u(t)$, equations (2.7) and (2.8) must be solved subject to the boundary conditions $\theta(0) = 0$, $\theta(t_1) = \omega t_1$. This can be done by numerically finding the initial condition $\lambda(0) \equiv \lambda_0$ that satisfies the boundary conditions, for example, by using a shooting method. We note that the above conditions are only necessary and not sufficient for global optimality. However, the phase reduction (2.1) requires inputs of small amplitude so that solutions remain close to the periodic orbit. Since $u(t)$ is directly proportional to λ in equation (2.6), we can limit our search to include solutions obtained with reasonably small values of $\lambda(0)$ in order to find a feasible solution.

While the previous procedure will produce an energy-optimal stimulus, it will not necessarily be charge-balanced (CB). The importance of CB stimuli in DBS has been known for many years. Over time, non-charge-balanced (NCB) stimuli can create an accumulation of charge and cause harmful Faradaic reactions that may damage surrounding neural tissue or the DBS electrode [60]. If we consider the total charge q imparted to a neuron to be the integral of the current, then $\dot{q}(t) = Cu(t)$, and we can derive an optimal CB stimulus by optimizing the same cost function as in the NCB case, $G[u(t)]$, subject to the additional constraints $q(0) = 0$ and $q(t_1) = 0$, the latter ensuring that the stimulus will be charge neutral at t_1. We again apply calculus of variations [59] to minimize $\mathcal{C}[\mathbf{\Phi}(t), \dot{\mathbf{\Phi}}(t), u(t)] = \int_0^{t_1} \mathcal{M}[u(t)]dt$ where

$$\mathcal{M}[u(t)] = u(t)^2 - \beta Z'(\theta)u(t) + \begin{bmatrix}\lambda_1(t) & \lambda_2(t)\end{bmatrix}\begin{bmatrix}\frac{d\theta}{dt} - \omega - Z(\theta)u(t) \\ \frac{dq}{dt} - u(t)\end{bmatrix}, \tag{2.9}$$

and $\mathbf{\Phi}(t) = [\theta(t), q(t), \lambda_1(t), \lambda_2(t)]^T$. The Lagrange multipliers λ_1 and λ_2 force the dynamics to satisfy the evolution equations for θ and q given above. The associated Euler-Lagrange equations are

$$\frac{\partial \mathcal{M}}{du} = \frac{\partial}{dt}\left(\frac{\partial \mathcal{M}}{d\dot{u}}\right), \qquad \frac{\partial \mathcal{M}}{d\mathbf{\Phi}} = \frac{\partial}{dt}\left(\frac{\partial \mathcal{M}}{d\dot{\mathbf{\Phi}}}\right). \tag{2.10}$$

With the above boundary conditions, this is a two-point boundary value problem for $u(t)$ which is solved using a double bisection algorithm described in [45]. As with the NCB formulation, (2.10) is only necessary and not sufficient for global optimality, but we can limit our search for $\lambda_1(0)$ and $\lambda_2(0)$ so that the resulting solutions yield optimal stimuli that are small enough so that the phase reduction (2.1) is still valid.

2.3 Guaranteed Lyapunov Exponents

In an experimental setting, errors in measuring the PRC will induce errors in the calculated optimal stimulus. An important question to ask is whether or not these errors will stifle the desynchronizing efforts of the electrical signal, and how large these errors need to be before they completely degrade its desynchronizing capabilities. To answer these questions, we consider an NCB optimal stimulus, $I_{opt}(t)$, found using methods from Section 2.2. Consider another stimulus $I_c(t)$ that is different from $I_{opt}(t)$. We define the

error signal as $I_e(t) = I_c(t) - I_{opt}(t)$ and the infinity norm of $I_e(t)$ as

$$||I_e(t)||_\infty = \sup_{0 \le t \le t_1} |I_e(t)|, \tag{2.11}$$

where t_1 is the duration of the optimal signal. As we will demonstrate in Section 2.5, a signal with a larger Lyapunov exponent will desynchronize two neurons with similar initial phase more quickly. For this reason, we use the Lyapunov exponent from (2.4) as a measure of the desynchronizing capabilities of a signal. In order to identify a bound on $||I_e(t)||_\infty$ which can guarantee desynchronization, we must find the worst possible Lyapunov exponent for any signal with $||I_e(t)||_\infty \le E$, where E is a constant. To do so, we define the cost function $\mathcal{L}[I_e(t)] = \int_0^{t_1} [I_{opt}(t) + I_e(t)] Z'(\theta(t)) dt$ subject to (2.1), $\dot{\theta} = \omega + Z(\theta)[I_{opt}(t) + I_e(t)]$, with the additional constraint $-E \le I_e(t) \le E$ for all $t \in [0, t_1]$. Here, $\mathcal{L}[I_e(t)]$ corresponds to the Lyapunov exponent for a stimulus $I_c(t)$. The inequality constraint ensures that $||I_e(t)||_\infty \le E$. Using Pontryagin's Minimum Principle [13], a necessary condition for the control that minimizes $\mathcal{L}[I_e(t)]$ is that the control minimizes the Hamiltonian

$$\mathcal{H}(\theta(t), I_e(t), p_1(t)) = [I_{opt}(t)+I_e(t)]Z'(\theta(t))+\omega p_1(t)+[I_{opt}(t)+I_e(t)]p_1(t)Z(\theta(t)), \tag{2.12}$$

and is given by $I_e(t) = -E\text{sign}(Z'(\theta) + p_1(Z(\theta))$ where $' = d/d\theta$, and p_1 is a Lagrange multiplier. Furthermore,

$$\dot{\theta} = \omega + Z(\theta)[I_{opt}(t) + I_e(t)], \tag{2.13}$$

$$\dot{p_1} = -[I_{opt}(t) + I_e(t)]Z''(\theta) - p_1 Z'(\theta)[I_{opt}(t) + I_e(t)], \tag{2.14}$$

with the boundary condition $\theta(0) = 0$. From the problem formulation, we only have

one boundary condition, and in order to find the global minimum of $\mathcal{L}[I_e(t)]$, we must simulate equations (2.13) and (2.14) for all plausible values for $p_1(0)$ to determine the minimum (worst case) Lyapunov exponent of the associated signals. Using this approach, we can find boundaries on a stimulus that are sufficient, but not necessary, to yield a given Lyapunov exponent.

2.4 Optimal Control of the Phase Distribution

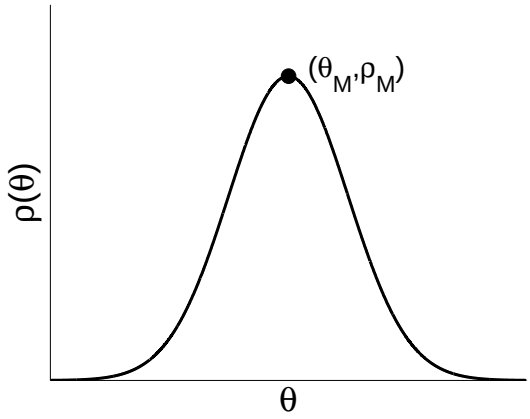

Figure 2.1: The maximum value, ρ_M for a large distribution of neurons.

The methods from Section 2.2 are optimal for desynchronizing two neurons with close initial phase. However, brain regions can contain on the order of billions of neurons [61]. For large populations of neurons, rather than examining individual neurons, which can be too large to simulate *in silico*, it is appropriate to monitor the probability density of neurons with phase θ at a given time, $\rho(\theta, t)$. For an uncoupled population that obeys equation (2.1), the probability density evolves according to the advection equation as

described in [20]:

$$\frac{\partial \rho(\theta,t)}{\partial t} = -\frac{\partial}{\partial \theta}\left[\{\omega + Z(\theta)u(t)\}\,\rho(\theta,t)\right] = -\rho\frac{\partial \nu}{\partial \theta} - \nu\frac{\partial \rho}{\partial \theta}, \tag{2.15}$$

where $\nu(\theta,t) = \omega + Z(\theta)u(t)$. In Section 2.5, we show that the CB and NCB optimal signals obtained using methods from Section 2.2 can effectively desynchronize a network of 100 coupled neurons for many different neural models. Therefore, it is no surprise that the same signal is effective at desynchronizing a population evolving according to (2.15) (not shown).

However, we wish to approach this problem from a neural population perspective, to see if we can find an effective control strategy to desynchronize large neural populations. Suppose we are interested in the evolution of the maximum of the distribution, ρ_M, at $\theta \equiv \theta_M$. Noting that the total time derivative of the distribution is $\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \frac{\partial \rho}{\partial \theta}\frac{d\theta}{dt}$, and that $\frac{\partial \rho}{\partial \theta} = 0$ at the local maximum, we find

$$\frac{d\rho_M}{dt} = -\rho_M \frac{\partial \nu}{\partial \theta}(\theta_M, t) = -Z'(\theta_M)u(t)\rho_M. \tag{2.16}$$

Note the similarity of (2.16) to equation (3.4). From the two-neuron optimal control formulation, $Z'(\theta)u(t) > 0$ corresponds increasing the phase difference, whereas here, it corresponds to a decreasing value of ρ_M. In order to find an equation for the evolution of θ_M, we again make use of the relation $\frac{\partial \rho}{\partial \theta}|_{\theta_M} = 0$. Taking the total time derivative yields

$$\begin{aligned} 0 =& \frac{\partial}{\partial\theta}\frac{d\rho}{dt}, \\ =& \frac{\partial}{\partial\theta}\left[\rho_\theta \frac{d\theta_M}{dt} + \rho_t\right], \\ =& \frac{\partial}{\partial\theta}\left[\rho_\theta \frac{d\theta_M}{dt} - \nu_\theta\rho - \rho_\theta\nu\right], \\ \rho_{\theta\theta}\frac{d\theta_M}{dt} =& \nu_{\theta\theta}\rho + 2\nu_\theta\rho_\theta + \rho_{\theta\theta}\nu. \end{aligned} \tag{2.17}$$

All expressions in the above equation are evaluated at ρ_M and θ_M, thus $\rho_\theta = 0$. Also, since ρ_M is a local maximum, $\rho_{\theta\theta} < 0$, and equation (2.17) becomes

$$\frac{d\theta_M}{dt} = \omega + Z(\theta)u(t) + \frac{\rho_M}{\rho_{\theta\theta}}Z''(\theta)u(t). \tag{2.18}$$

Equations (2.16) and (2.18) are special cases of the equations derived in [62]. In order to effectively use (2.18), we must find some function $\kappa(\theta_M, \rho_M) \approx \rho_{\theta\theta}$. One such κ, which works well in practice, can be obtained by starting with a Gaussian distribution and assuming that the distribution remains Gaussian for all time. We note that while the phase reduction is valid for $\theta \in [0, 2\pi)$, a Gaussian distribution is defined for all θ. However, because we are considering synchronized systems with a small variance, the values of the Gaussian distribution that we are ignoring are insignificant. Using this strategy, one can easily show that $\kappa(\rho_M) = -2\pi\rho_M^3$.

Equation (2.16) is separable, and we can solve explicitly to determine the change in the value of ρ over some time interval of length T

$$\begin{aligned} \int_a^{a+T}\frac{1}{\rho_M}d\rho_M &= -\int_a^{a+T} Z'(\theta_M)u(t)dt, \\ \implies \log\left(\frac{\rho_M(a+T)}{\rho_M(a)}\right) &= -\int_a^{a+T} Z'(\theta_M)u(t)dt. \end{aligned} \tag{2.19}$$

If we want to minimize ρ_M while taking into account the energy expended, we can use calculus of variations [59] to minimize $\mathcal{C}[\mathbf{\Phi}(t), \dot{\mathbf{\Phi}}(t), u(t)] = \int_0^{t_1} \mathcal{G}[u(t)]dt$ where

$$\mathcal{G}[u(t)] = u(t)^2 - \beta Z'(\theta)u(t) + \begin{bmatrix}\lambda_1(t) & \lambda_2(t)\end{bmatrix} \begin{bmatrix} \frac{d\theta}{dt} - \omega - Z(\theta)u(t) - \frac{\rho_M}{\kappa(\theta_M,\rho_M)}Z''(\theta)u(t) \\ \frac{d\rho_M}{dt} + \rho_M Z'(\theta)u(t) \end{bmatrix}, \tag{2.20}$$

and $\mathbf{\Phi}(t) = [\theta_M(t), \rho_M(t), \lambda_1(t), \lambda_2(t)]^T$. Lagrange multipliers λ_1 and λ_2 force the dynamics to satisfy (2.16) and (2.18). Note the similarity of (2.20) to the NCB cost function (2.5). When $\kappa(\theta_M, \rho_M) \gg \rho_M Z''(\theta)u(t)$ (as might be the case for a highly synchronized network) the function that minimizes (2.20) is approximately the same function that minimizes (2.5). Taking $\kappa(\rho_M) = -2\pi\rho_M^3$ as previously mentioned, the associated Euler-Lagrange equations are

$$u(t) = \left[\beta Z'(\theta_M) + \lambda_1\left(Z(\theta_M) - \frac{1}{2\pi\rho_M^2}Z''(\theta_M)\right) - \lambda_2\rho_M Z(\theta_M)\right]/2, \tag{2.21}$$

$$\dot{\rho_M} = -\rho_M Z'(\theta_M)u(t), \tag{2.22}$$

$$\dot{\theta_M} = \omega + Z(\theta_M)u(t) - \frac{1}{2\pi\rho_M^2}Z''(\theta_M)u(t), \tag{2.23}$$

$$\dot{\lambda_1} = \lambda_1 u(t)\left[\frac{1}{2\pi\rho_M^2}Z'''(\theta_M) - Z'(\theta_M)\right] + Z''(\theta_M)u(t)\left[\rho_M\lambda_2 - \beta\right], \tag{2.24}$$

$$\dot{\lambda_2} = u(t)\left[-\frac{\lambda_1 Z''(\theta_M)}{\pi\rho_M^3} + \lambda_2 Z'(\theta_m)\right]. \tag{2.25}$$

Unlike the problem for finding the maximum Lyapunov exponent, the distribution minimization problem depends on the initial height of the distribution, $\rho_M(0)$. We will take $\theta_M(0) = 0$ which leaves two initial conditions to be determined later.

2.5 Results and Discussion

2.5.1 Maximizing Lyapunov Exponents for the Thalamus Model

Because pathological neural synchronization in the thalamus is thought to play an important role in Parkinson's disease [40], we consider a three-dimensional model to describe thalamic neural activity [63]:

$$\begin{aligned}
\dot{V_i} &= (-I_L(V_i) - I_{Na}(V_i, h_i) - I_K(V_i, h_i) - I_T(V_i, r_i) \\
&\quad + I_{SM} + \frac{1}{N}\sum_{i=1}^{N} \alpha_{ij}(V_j - V_i) + u(t) + \eta_i(t))/C, \\
\dot{h_i} &= (h_\infty(V_i) - h_i)/\tau_h(V_i), \\
\dot{r_i} &= (r_\infty(V_i) - r_i)/\tau_r(V_i), \quad i = 1, \ldots, N.
\end{aligned} \tag{2.26}$$

We have augmented the voltage equation by additively including electrotonic coupling [64], DBS input, and Gaussian white noise. Here, N is the total number of neurons, V_i, h_i, and r_i are membrane voltage and gating variables for neuron i, α_{ij} characterizes the coupling strength between electrotonically coupled neurons i and j, with $\alpha_{ij} = \alpha_{ji}$ and $\alpha_{ii} = 0$ for all i, $\eta_i(t) = \sqrt{2D}\mathcal{N}(0, 1)$ is the i.i.d. noise associated with each neuron, assumed to be zero-mean Gaussian white noise with variance $2D$, and $u(t) = I(t)/C$ represents a common control input. In this equation I_{SM} represents the baseline current which we take to be $5\mu\text{A/cm}^2$. For a full explanation of the functions $I_L, I_{Na}, I_K, I_t, h_\infty, \tau_h, r_\infty$ and τ_r, we refer the reader to [63].

The baseline current causes the neuron to fire with period $T = 8.395$ms. To obtain the optimal control, we take $t_1 = 7.35$ms (corresponding to $\theta = 5.5$ on the limit cycle) and $\beta = 40$. Note that t_1 sets the duration of the stimulus, and could be chosen differently

provided that it is sufficiently smaller than T, and β has been chosen so that a positive Lyapunov exponent is favored, but that the magnitude of the control input is small enough so that the phase reduction from equation (2.1) is still valid. We solve equations (2.7) and (2.8) numerically with a fourth order Runge-Kutta solver and the optimal control is found from (2.6). We do the same for the Euler-Lagrange equations from (2.10).

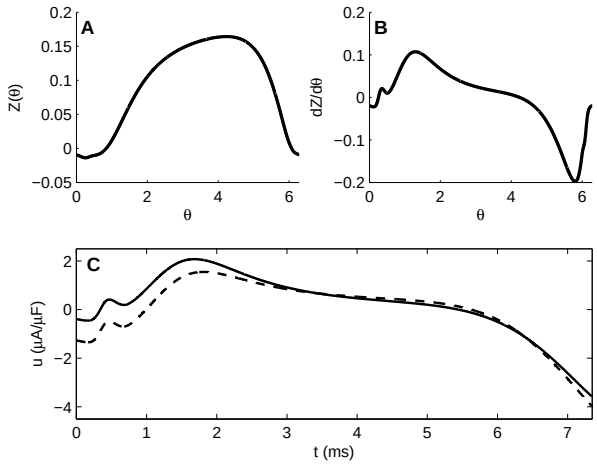

Figure 2.2: (A-B) Numerically computed PRC and its derivative for the thalamus model (2.26). (C) The CB and NCB optimal stimuli are shown as a dashed and solid line, respectively.

Panels A and B of Figure 2.2 show the PRC and its first derivative for (2.26), numerically obtained using the software X-Windows Phase Plane (XPP) [65]. Panel C of Figure 2.2 shows the optimal CB and NCB stimuli. Both stimuli are similar because the NCB stimulus is nearly charge-balanced. We note that the optimal stimulus is strikingly similar to $Z'(\theta)$ for $\theta \leq 5.5$ and explain this occurrence by noting that the equation for the optimal stimulus from (2.6) depends directly on the sum of $\beta Z'(\theta(t))$ and $\lambda(t)Z(\theta(t))$. The optimal stimulus has a relatively small magnitude, so $\theta(t) \approx \omega t$ and, numerically, we find that $\lambda(t)$ is relatively small compared to β.

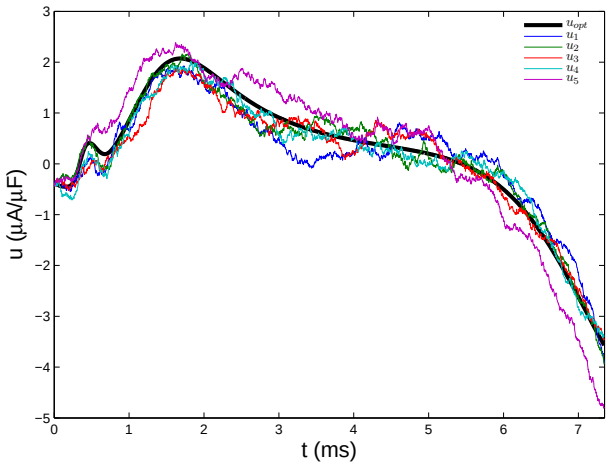

Figure 2.3: The NCB optimal stimulus u_{opt} and 5 instances of u_{opt} that have been corrupted by noise. Indeed, out of the 5 stimuli, u_{opt} yields the smallest value of $\mathcal{C}[u(t)]$ (2.5), but the other stimuli give values that are close to optimal.

In order to numerically verify that the NCB stimulus is optimal for minimizing the cost function, we analyze five other stimuli given by $u_i(t) = u_{opt}(t) + 0.5W_i(t)$, $i = 1 \ldots 5$ where u_{opt} is the optimal NCB stimulus shown in Figure 2.2, and W_i is a Wiener process, added to corrupt the optimal stimulus. We note that each of these stimuli are subject to the same end point conditions described in Section 2.2. Figure 2.3 shows these stimuli as well as the NCB optimal stimulus for reference. We find that u_{opt} does have the best performance in terms of overall cost, but the other stimuli still yield comparable Lyapunov exponents. This prompts the question of how robust this procedure is to errors, which will be addressed later.

The Lyapunov exponents calculated using equation (2.4) for NCB and CB stimuli are 0.066 and 0.060 respectively. We find that requiring the CB constraint decreases the Lyapunov exponent, and hence, the rate of desynchronization. Figure 2.4 shows the phase separation of two neurons which obey equation (2.2) for the PRC found in Figure 2.2 subject to repeated iterations of both the NCB (Panels A and B) and CB (Panels C and

Table 2.1: Stimulus properties from Figure 2.3

stimulus	$\Lambda(T)$	Energy	$\mathcal{C}[u(t)]$
u_{opt}	0.066	11.66	-10.43
u_1	0.055	8.80	-9.68
u_2	0.062	10.7	-10.03
u_3	0.058	9.66	-9.89
u_4	0.061	10.44	-10.06
u_5	0.086	19.98	-8.95

D) stimuli. We find that the neurons exponentially desynchronize at a rate determined by their Lyapunov exponent until they are nearly π radians out of phase. At this point, the assumption that the neurons are close in phase is no longer valid, and no further desynchronization occurs.

Results from Figure 2.4 only apply to neurons obeying the phase reduction (2.1). In order to determine the validity of the phase reduction for (2.26), we simulate the deterministic version of equations (2.26), i.e. with $D = 0$, using a fourth order Runge-Kutta solver. The top panel of Figure 2.5 shows time histories of three neurons with initial conditions that correspond to $\theta = -0.1$, 0, and $+0.1$. on the periodic orbit. The control is applied every time the reference neuron, with initial phase $\theta = 0$, fires. We find that after 70 ms, the neurons no longer show any evidence of their initial synchronization.

We now apply the NCB optimal control found above to a network of $N = 100$ coupled, initially synchronized, noisy neurons, with an identical coupling strength of $\alpha_{ij} = 0.1$ and i.i.d. noise with $D = 1$, in order to test the desynchronizing effects on the full neural model. We define the mean voltage as our system observable, $\bar{V}(t) = \frac{1}{N}\sum_{i=1}^{N} V_i(t)$. The controller has two states: *active* and *inactive*. When the controller is *active*, a new stimulus is triggered when $\bar{V} > -45$mV and $\dot{\bar{V}} < 0$ with the caveat that a new stimulus cannot occur until the previous stimulus is either finished, or within 0.3ms of finishing. This last condition is included to give the controller flexibility in when to start the next

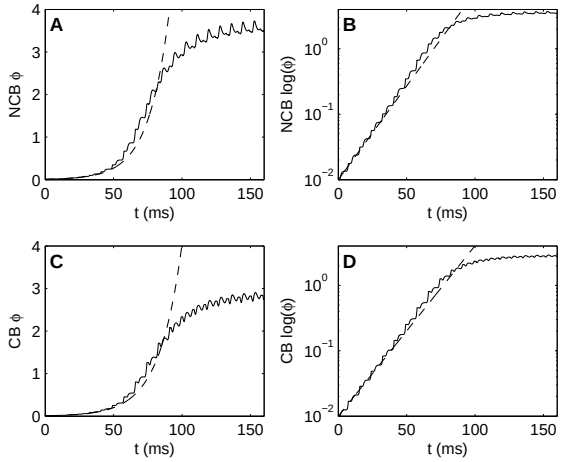

Figure 2.4: A-B show phase difference between two neurons over time for the NCB optimal stimulus (giving $\Lambda = 0.066$) applied repeatedly to two neurons with nearly identical initial phases. (C-D) give the same plots for the CB stimulus ($\Lambda = 0.060$). In both cases the neurons desynchronize at a rate determined by Λ, calculated from equation (2.4), until $\phi \approx 2$ at which point the solution begins to asymptote to $\phi \approx \pi$, i.e. the anti-phase state. Dashed lines show exponential functions based on the Lyapunov exponents.

stimulus because the system is sensitive to the time when the stimulus is presented. Once $\bar{V}$ no longer spikes above -45mV, the controller changes to an *inactive* state. It changes back to the *active* state if $\bar{V}$ registers above -40mV. We use the algorithm presented in [66] to simulate the noisy system (see Figure 2.6). The desynchronizing effect of the stimulus can clearly be seen in the raster plot. We call this event-based control because the controller is only triggered when the mean voltage crosses a certain threshold. It is worth noting that the optimal stimulus works equally well for a network of neurons that are synchronized by a common pulsatile input, which [21] suggests is the mechanism that yields synchronization. Results for such a system are qualitatively similar to the results presented in this chapter.

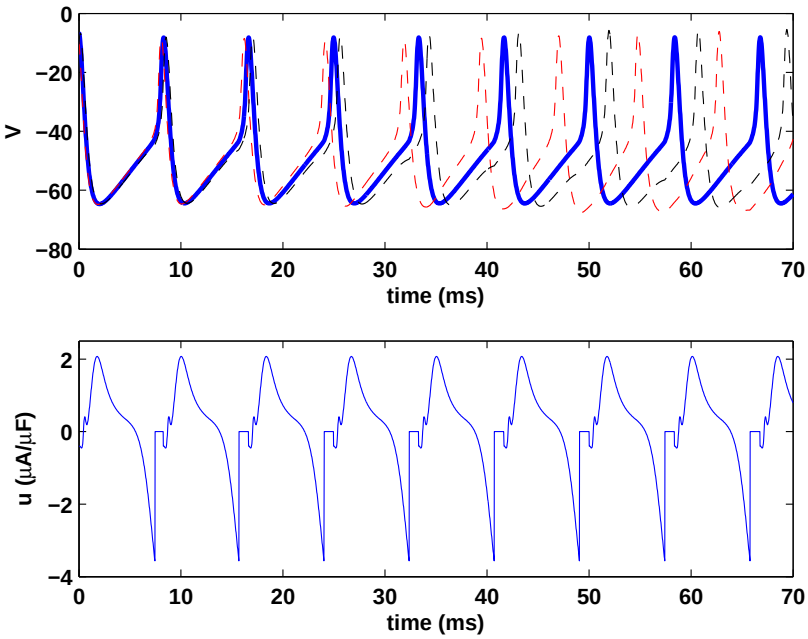

Figure 2.5: Time histories of three neurons with initial conditions $\theta = -0.1, 0, \text{and} +0.1$ on the periodic orbit. A new stimulus is applied each time the reference neuron with initial phase $\theta = 0$ (shown as the black curve) fires.

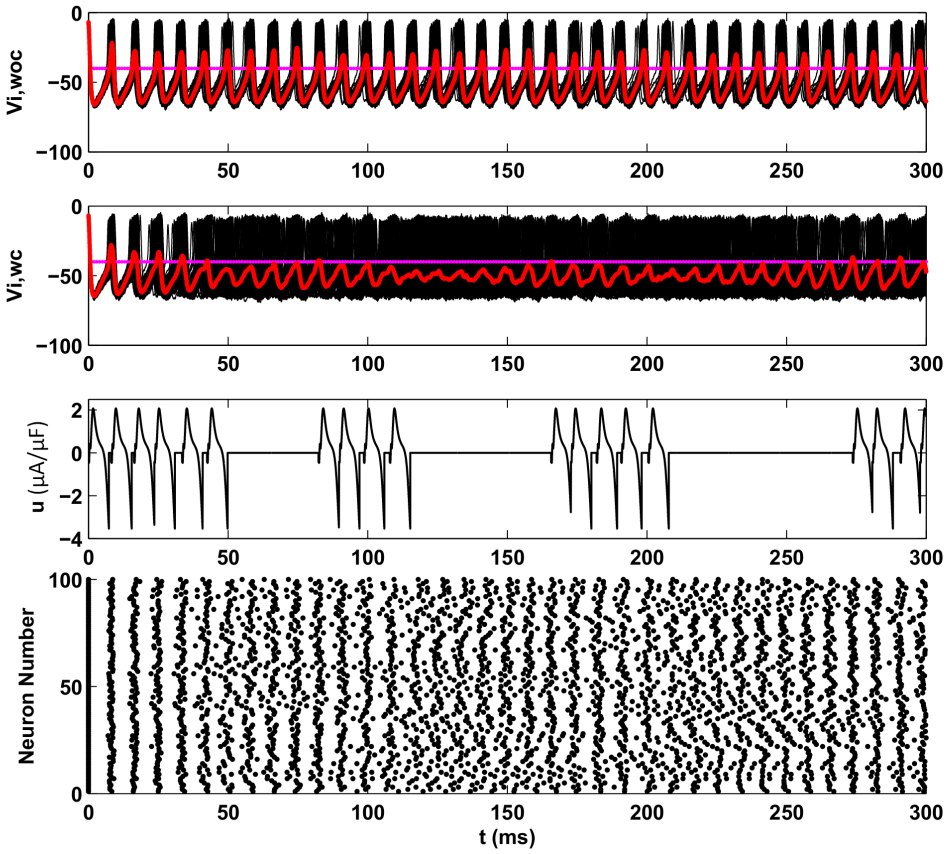

Figure 2.6: Results for a population of $N = 100$ noisy, coupled neurons. The first panel shows the network in the absence of control. The second and third panels show results for the same network with the event-based control applied. The traces give the mean voltages for the system and the horizontal dotted line shows the control activation threshold. Substantial desynchronization can be seen from the raster plot.

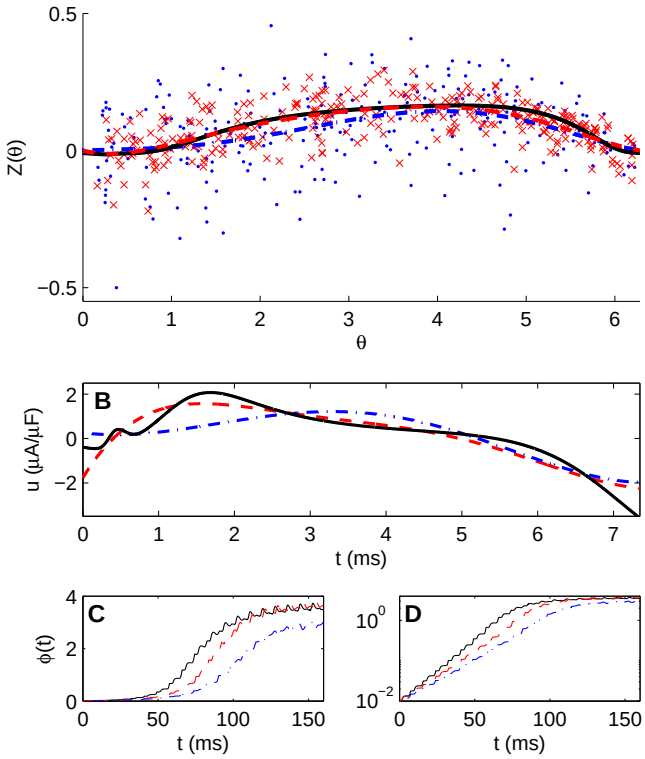

Figure 2.7: (A) Obtaining the PRC from a noisy system with the direct method. Data points associated with noise level $D = 1$ and 0.25 are shown as dots and x's, respectively, with 6^{th} order polynomial fits given as a dashed and dot-dashed line, respectively. Pearson Correlation coefficients of the polynomial fits for $D = 1$ and 0.25 are 0.3617 and 0.6638, respectively. For comparison, the numerically computed (exact) PRC is shown as a solid line. (B) NCB optimal stimuli computed with experimentally determined PRC's with $D = 1$ and 0.25 are shown as dot-dashed and dashed lines, respectively. The true NCB optimal stimulus from Figure 2.2 is shown as a solid line for reference. (C-D) Phase differences between two neurons over time for the optimal stimulus obtained from systems with noise level $D = 1, 0.25$, and 0 applied repeatedly to two neurons with dynamics governed by the exact PRC and with close initial phase differences are shown as dot-dashed, dashed, and solid lines, respectively.

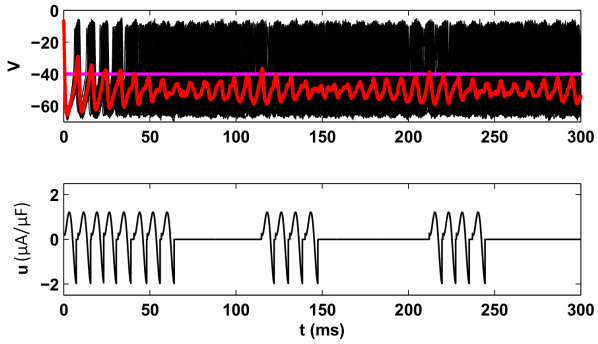

Figure 2.8: Results for a population of $N = 100$ noisy, coupled neurons with optimal stimulus found using the direct method for a noisy ($D = 1$) neuron. The top and bottom panels show results for the network with the event-based control applied. The traces give the mean voltages for the system and the horizontal dashed line shows the control activation threshold.

The critical advantage of using this control strategy is that it only requires knowledge of the PRC for the system. Methods for controlling neurons that require precise knowledge of the neural model (see [44], [43]) are difficult to implement because, despite recent progress [67], it is challenging to obtain accurate full scale models of real neurons for control purposes. Methods that rely on the PRC are attractive because it is experimentally measurable. To illustrate the effectiveness of this method, we employ the following method which we will refer to as the direct method [24] (see also Appendix B)to obtain a PRC for a single neuron obeying equation (2.26). In order to obtain one data point, a short-duration pulse of current is applied to a neuron at a random phase θ, and the resulting phase change is measured by observing the change in spike time. The resulting value $Z(\theta)$ is $\frac{\Delta\theta}{Q_p/C}$ where Q_p is the total charge imparted to the neuron from the pulse, and $\Delta\theta$ is the change in phase. An experimentally reasonable sampling size of 300 data points were obtained for noise levels of both $D = 1$ and 0.25, and the data was fit to a 6^{th} order polynomial constrained to be zero at both θ equals 0 and 2π, as in [25]. The results are shown in panel A of Figure 2.7. As expected, a larger noise value yields a

larger spread in the data, and a less reliable PRC. The optimal stimuli obtained using the fit PRC are shown in panel B. Finding the PRC with a D value of 0, 0.25, and 1 yields an optimal stimulus with $\Lambda = 0.066$, 0.055, and 0.046, respectively. When we do not know the PRC exactly, the performance of the desynchronizing stimulus is somewhat degraded as evidenced in panels C, and D, but the stimulus still performs remarkably well. We note that for a noise level of $D = 1$, the spread in the collected data for the PRC is similar to data previously collected for *in vitro* neurons [25].

Figure 2.8 shows results from a simulation using the optimal stimulus obtained from this data applied to 100 coupled neurons obeying equation (2.26) with the same noise and coupling parameters as the test shown in Figure 2.6. The stimulus still shows desynchronizing capabilities similar to the stimulus obtained from the true PRC.

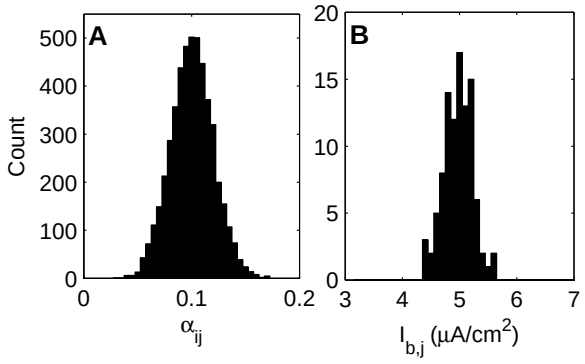

Figure 2.9: (A) Normal distributions of the coupling strengths with $\alpha_{ij} \in \mathcal{N}(0.1, 0.02)$. (B) Normal distributions of the baseline currents with $I_{b,j} \in \mathcal{N}(5, 0.25)$.

To model more realistic neural networks, we include network heterogeneities. For comparison, our homogeneous network simulations will use $N = 100$, $\alpha_{ij} = 0.1$ and $I_b = 5$. Other simulations will consider network heterogeneities in coupling strength by drawing α_{ij} values from a normal distribution with mean $\bar{\alpha} = 0.1$ and a standard deviation $\sigma_\alpha = 0.02$; panel A of Figure 2.9 shows the specific distribution used for each simulation. Network heterogeneities in I_b are also considered by simulating a system with

baseline currents drawn from a normal distribution (shown in panel B of Figure 2.9). In each $N = 100$ neuron simulation, we use the same control logic as the simulation with results shown in Figure 2.7 to determine when to apply each new stimulus. Again, we use the algorithm presented in [66] to simulate the noisy system.

Results of each simulation are shown in Figure 2.10. The top panel shows the result for the network with homogeneous baseline current and coupling. The second panel shows results for a neural network with homogeneous baseline currents $I_b = 5$ and coupling strengths drawn from a random distribution (shown in panel A of Figure 2.9) and the third panel shows results for a network with homogeneous coupling $\alpha_{ij} = 0.1$ and baseline currents drawn from a random distribution (shown in panel B of Figure 2.9). We find that when we include heterogeneity the network requires fewer applications of the optimal stimulus to desynchronize. The bottom panel shows results for a network with both heterogeneous coupling and baseline currents. Overall, we find that heterogeneity in a network decreases its tendency to resynchronize, which increases the effectiveness of the optimal control.

Clearly the optimal stimulus works well for desynchronizing a neural network, even when only an approximation to the true PRC is known. This prompts the search for error bounds on the optimal stimulus that can still guarantee desynchronization. In an *ad hoc* manner using methods from Section 2.3, we fix a particular value of E for the signal I_e and minimize (2.12) to find the minimal (worst case) Lyapunov exponent. We then simulate (2.26) to with the associated signal to determine if it can affect network desynchronization. Using this strategy with different E values, we find that for a network of 100 neurons with homogeneous coupling $\alpha_{ij} = 0.1$ and baseline current, $I_b = 5\mu\text{A/cm}^2$ we require a Lyapunov exponent of at least, $\Lambda = 0.024$, corresponding to $E = 0.8$, in order to achieve sufficient desynchronization so that spikes of $\bar{V}$ remain below -40mV. Thus, $||I_e(t)||_\infty \leq 0.8$ will guarantee $\Lambda \geq 0.024$. To illustrate the utility of this measure, we

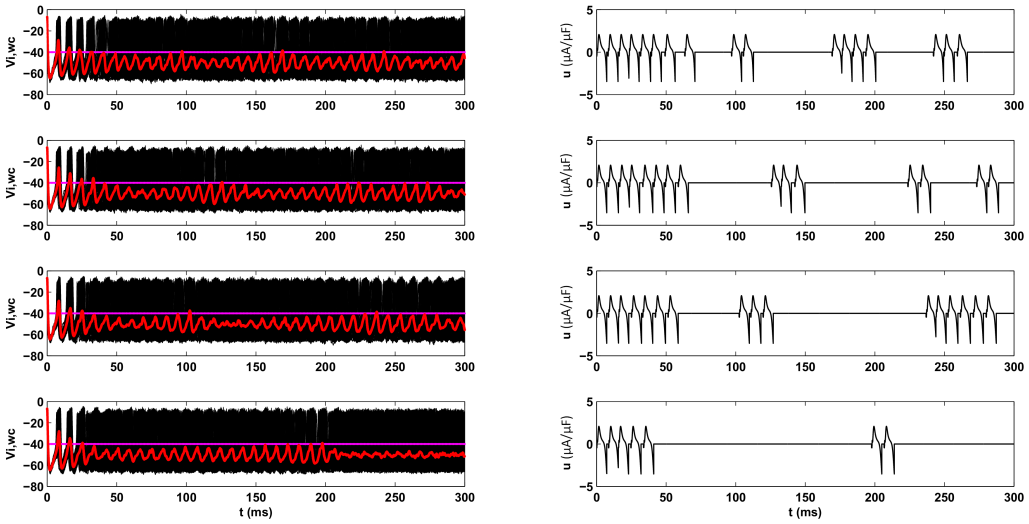

Figure 2.10: Results for a population of N = 100 coupled neurons with i.i.d. noise strength $D = 1$. The top panel shows a network with homogeneous coupling strength $\alpha_{ij} = 0.1$ and baseline current $I_b = 5\mu A/cm^2$. In the second panel, coupling strengths drawn from a distribution as shown in panel A of Figure 2.9 but baseline current is the same for all neurons, $I_b = 5\mu A/cm^2$. In the third panel baseline currents are drawn from a distribution as shown in panel B of Figure 2.9, but coupling is the same for all neurons, $\alpha_{ij} = 0.1$. The bottom panel shows a network with both heterogeneous baseline currents and coupling strengths drawn from the corresponding distributions in Figure 2.9. Once the network is sufficiently desynchronized, the controller will only activate once the voltage has crossed the threshold, shown as a horizontal line.

choose a simple, piecewise linear signal which is nearly contained entirely in the boundary $||I_e(t)||_\infty \leq 0.8$.

$$u(t) = \begin{cases} t, & 0 \leq t \leq 2.2 \\ -4.4 - t, & 2.2 < t \leq 7.35. \end{cases} \tag{2.27}$$

The signal in (2.27) is used in a simulation of the same homogeneous network with the same control logic described previously. The results are shown in Figure 2.11. Numerically, we find the $\Lambda = 0.0577$ for this stimulus. We note that even though the stimulus

is not entirely contained within the shaded region shown in the top panel of Figure 2.11, and therefore, not guaranteed to desynchronize the system, it still produces a sufficiently high Lyapunov exponent to achieve desynchronization.

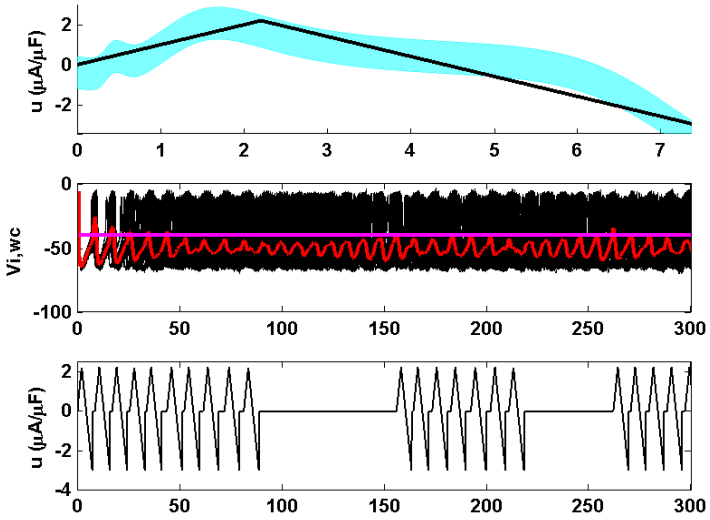

Figure 2.11: Results for the population of $N = 100$ coupled neurons from Figure 4 with the simple, piecewise linear stimulus from (2.27). The top panel shows the stimulus as well as a shaded boundary that will ensure sufficient desynchronization. The traces in the middle panel give the mean voltages for the system and the horizontal line gives the control activation threshold. We see in the third panel that desynchronization occurs, but requires more applications of the stimulus than the optimal stimulus, which is expected due to a smaller Lyapunov exponent.

2.5.2 Optimally Decreasing the Peak of a Distribution for the Thalamus Model

We now turn our attention to controlling populations of neurons using the methods described in Section 2.4 to optimally decrease the peak of their phase distribution. Here, we attempt to provide reasonable values for the still undetermined parameters in the Euler-Lagrange Equations (2.21)-(2.25). Because we have undetermined parameters, we

cannot claim to have found an optimal stimulus to minimize our cost function, but we can still gather powerful insight about stimuli that will affect network desynchronization. In order to make explicit comparisons with the optimal stimulus which maximizes the Lyapunov exponent, we take $t_1 = 7.35$ms and $\beta = 40$. We take our initially synchronized distribution to be a normal distribution with standard deviation $\sigma_d = 0.2$ centered at $\theta = 0$, which corresponds to a spiking event. These conditions were chosen as a reasonable approximation of the observed distribution for simulations of network (2.26) just before the control threshold $\bar{V} = -40$mV is reached. This gives $\theta_M(0) = 0$ and $\rho_M(0) = 1.995$ as initial conditions to (2.22) and (2.23); however, we still need to determine $\lambda_1(0)$ and $\lambda_2(0)$. In order to find the best choice of the remaining initial conditions, we minimize the cost function for reasonable choices of $\lambda_1(0)$ and $\lambda_2(0)$. Using this approach we find $\lambda_1(0) = 18$ and $\lambda_2(0) = 2$ approximately minimizes the cost function, and gives the stimulus shown in Figure 2.12, which will be referred to as u_D.

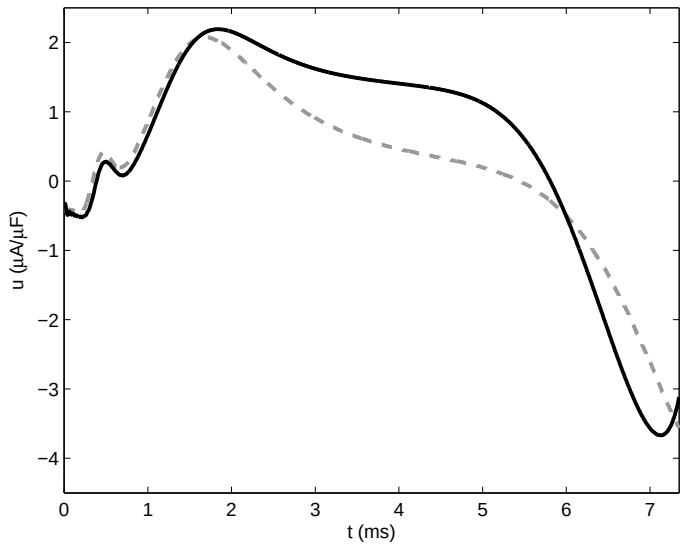

Figure 2.12: The optimal stimulus u_D for decreasing the peak of the phase distribution is shown as a solid, black line. For reference, the NCB optimal stimulus for maximizing the Lyapunov exponent is shown as a grey, dashed line.

We apply u_D to (2.15) and (2.26) with $N = 250$, $D = 0$ and $\alpha_{ij} = 0$ to determine the validity of the results found using the phase reduction. Throughout the simulation, we infer the phase of each of neuron in (2.26) by simulating each neuron separately in the absence of DBS input and noise to determine when its next spiking event occurs. Figure 2.13 shows three snapshots from this simulation. In order to clearly present the results, the mapping $\theta \in (\pi, 2\pi) \to (-\pi, 0)$ is used, e.g., $\theta = \frac{3\pi}{2}$ is plotted as $\theta = -\frac{\pi}{2}$. Theoretical and numerical evolutions are shown in black and blue, respectively. The red curve gives the theoretical evolution of the distribution in the absence of stimulus and is shown for reference. Throughout the simulation, the dynamics of the 250 neuron system is well approximated by (2.15), and the optimal stimulus brings ρ_M from 1.995 to 0.776. We now see utility of the approach which minimizes the peak of the distribution compared to an approach which maximizes the NCB Lyapunov exponent (producing a stimulus we will refer to as u_Λ). From (2.22), we see that a stimulus has the potential to greatly decrease a distribution when $|Z'(\theta_M)|$ is relatively large. For the Thalamus model, this can be done by applying a large negative stimulus when $\theta_M \approx 5.8$. As shown in Figure 2.12, u_D and u_Λ are nearly identical for $0 \leq t < 2$ but from about $2 \leq t < 6$, $u_D > u_\Lambda$ at a time when $Z(\theta_M) > 0$ and $Z'(\theta_M) \approx 0$. This has the effect of speeding up the distribution, but not decreasing the peak height. The extra effort used in speeding up the peak is repaid when $\theta_M \approx 5.8$, when a large negative stimulus decreases the peak rapidly at a time when $Z'(\theta) < 0$. The distribution deforms in such a way that ρ_M remains at $\theta_M \approx 5.8$ longer than we would expect for a system of only two neurons, which is reflected in the inequality $u_D < u_\Lambda$ for $t \approx 6$.

We have found stimuli for desynchronizing neural networks using two different approaches. Both approaches produce similar results, but we would like to know if one is better than the other. To answer this question, we simulate (2.26) with $N = 250$, $D = 1$, and $\alpha_{ij} = 0.1$ for 1000ms using the same control strategy as the simulations from Figure

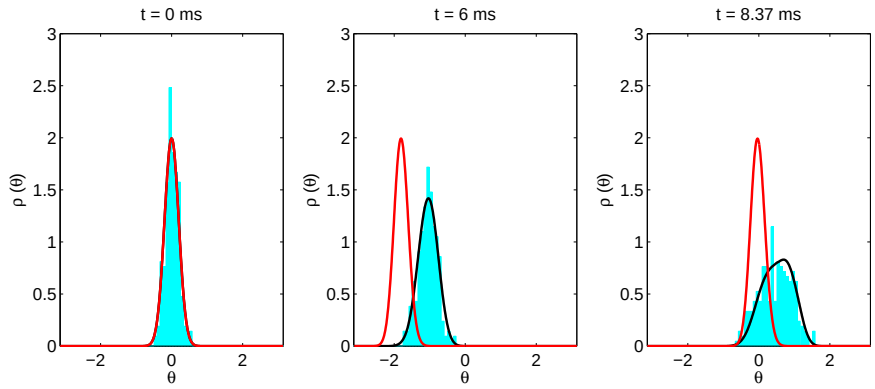

Figure 2.13: Three snapshots from simulations of (2.26) and the associated model (2.15). In order to clearly present the results, the mapping $\theta \in (\pi, 2\pi) \to (-\pi, 0)$ is used, e.g., $\theta = \frac{3\pi}{2}$ is plotted as $\theta = -\frac{\pi}{2}$. Theoretical and numerical evolutions are shown in black and blue, respectively. The red curve gives the theoretical evolution of the distribution in the absence of stimulus and is shown for reference.

2.10. The initial phase distribution is drawn from a normal distribution with $\sigma_d = 0.2$ centered at $\theta = 0$. A representative portion of the comparison is shown in Figure 2.14. When we use u_D as the control, the overall energy use is 533 units, while for u_Λ, the energy consumption is 460 units. In terms of energy, u_D performs slightly worse than u_Λ, most likely because the neuron distribution is not quite Gaussian throughout the simulation as we had assumed to derive u_D, but both strategies yield an effective control input using a comparable amount of energy. We also note that u_D desynchronizes the network more quickly than u_Λ, which is expected because it has a higher Lyapunov exponent than u_D as calculated from (2.4), $\Lambda = 0.107$ and 0.066, respectively, but requires more energy per application of the stimulus.

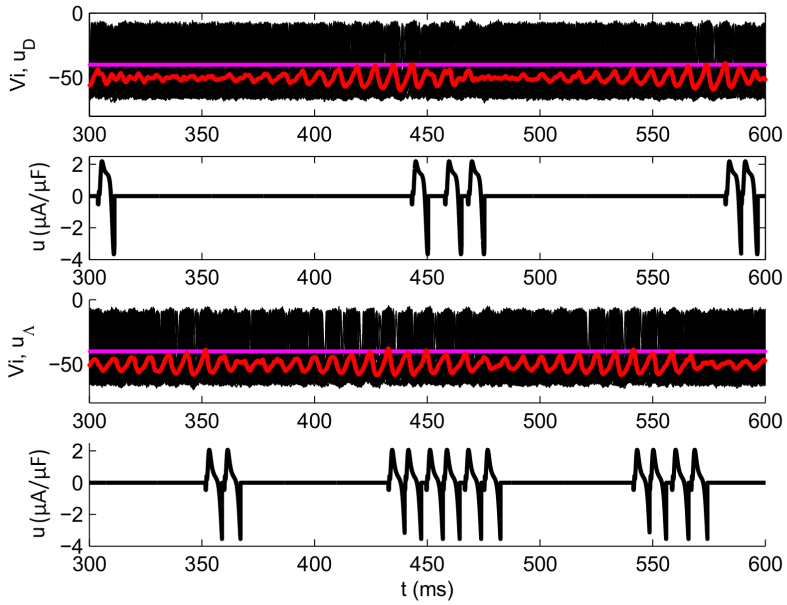

Figure 2.14: Comparing u_D to u_Λ. Representative excerpts from a 1000 ms simulation using u_D to u_Λ as the control are shown in the top and bottom panels, respectively.

2.5.3 Maximizing Lyapunov Exponents for the Reduced Hodgkin-Huxley Model

We next apply our optimization method to a population of neurons, each described by a two-dimensional reduction of the renowned four-dimensional Hodgkin Huxley (HH) model [68] that reproduces the essential characteristics of a neuron's dynamical behavior (cf., [69], [9]). An alternative strategy for desynchronizing a populations of such neurons was investigated in [44]; we will make comparisons with that method later in the present chapter. The population of neurons is modeled as follows:

$$\dot{V}_i = f_V(V_i, n_i) + \frac{1}{N}\sum_{i=1}^{n} \alpha_{ij}(V_j - V_i) + u(t) + \eta_i(t),$$
$$\dot{n}_i = f_n(V_i, n_i). \tag{2.28}$$

Here, $i = 1, \ldots, N$ where N is the total number of neurons, V_i and n_i are membrane voltage and gating variables for neuron i, α_{ij} characterizes the coupling strength between electrotonically coupled neurons i and j [64], with $\alpha_{ij} = \alpha_{ji}$ and $\alpha_{ii} = 0$ for all i, $\eta_i(t) \in \sqrt{2D}\mathcal{N}(0, 1)$ is the noise associated with each neuron, assumed to be zero-mean Gaussian white noise with variance $2D$, $u(t) = I(t)/C$ represents a common control input where $I(t)$ is a DBS input current in $\mu A/\mu F$, and $C = 1\mu F/cm^2$ is the membrane capacitance. Furthermore,

$$f_V = (I_b - \bar{g}_{Na}[m_\infty(V)]^3(0.8 - n)(V - V_{Na}) - \bar{g}_K n^4(V - V_k) - \bar{g}_L(V - V_L))/C,$$
$$f_n = a_n(v)(1 - n) - b_n(V)n.$$

Other functions and parameters for the reduced model are

$$
\begin{aligned}
m(V) &= \frac{a_m(V)}{a_m(V) + b_m(V)}, \\
a_m(V) &= 0.1(V+40)/(1-\exp(-(V+40)/10)), \\
b_m(V) &= 4\exp(-(V+65)/18, \\
a_n(V) &= 0.01(V+55)/(1-\exp(-(V+55)/10)), \\
b_n(V) &= 0.125\exp(-(V+65)/80,
\end{aligned}
$$

$$
\begin{aligned}
V_{Na} &= 50mV, V_K = -77mV, V_L = -54.4mV, \\
\bar{g}_{Na} &= 120mS/cm^2, \bar{g}_K = 36ms/cm^2, \\
\bar{g}_L &= 0.3mS/cm^2, c = 1\mu F/cm^2, \\
I_b &= 10\mu A/cm^2.
\end{aligned}
$$

Here, $\bar{g}_{Na}$, $\bar{g}_K$, and $\bar{g}_L$ represent the conductances of the sodium, potassium and leakage channels, respectively, and V_{Na}, V_K, and V_L are their respective reversal potentials. Note that I_b, the neuron's baseline current, represents the effect of the surrounding brain regions on the neuron. This is a bifurcation parameter, with the value $I_b = 10\mu A/cm^2$ chosen to ensure that the neuron is in an oscillatory (periodically spiking) regime. The natural period, T of oscillation is 11.81ms.

The PRC, $Z(\theta)$, for this system is computed numerically with the software X-Windows Phase Plane (XPP) [65] and is shown in Figure 2.15. To perform computations with the PRC, we approximate it as a Fourier series with 1200 terms. We take this many terms to get a reasonably non-oscillatory estimate of $Z'''(\theta)$ when solving (2.24).

The bottom panel of Figure 2.15 shows the result of the optimization process with and without the charge-balanced constraint for the choice of parameters $t_1 = 10.34$ ms

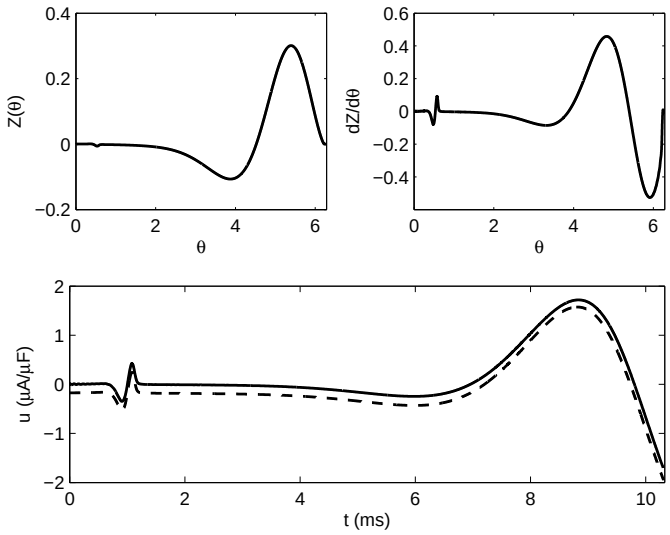

Figure 2.15: Top panels show the numerically computed PRC and its derivative for the reduced Hodgkin-Huxley model. The bottom panel shows The CB and NCB optimal stimuli as a dashed and solid line, respectively.

(corresponding to $\theta = 5.5$ on the limit cycle) and $\beta = 9$. Note that t_1 sets the duration of the stimulus and could be chosen differently provided that it is sufficiently smaller than the natural period, T, of the neuron, and β is chosen so that a positive Lyapunov exponent is favored, but the magnitude of the control input will be small enough so that the phase reduction is still valid. We find that the CB stimulus looks nearly identical to the NCB stimulus except for a downward shift. This is to be expected since, as is the case in the Thalamus model, the NCB stimulus is nearly charge balanced. Also, we find that the optimal control looks similar to the derivative of the PRC as shown in Figure 2.15. An explanation for this phenomenon is given in Section 2.5.1

Panels A and B in Figure 2.16 show the phase difference between two neurons with nearly identical initial phases, subject to the NCB optimal stimulus as shown in Figure 2.15. The optimal stimulus is event-based, and is applied every time the phase of the trailing neuron crosses $\theta = 0$. Panels C and D show the results of a similar test with the

CB stimulus. The Lyapunov exponents, Λ, for the NCB and CB stimuli are found to be 0.0823, and 0.0782, respectively. Dashed lines show exponential functions based on the Lyapunov exponents. We find that all of the plots match closely until $\phi \approx 1$, with the NCB stimulus performing slightly better than the CB stimulus. In this case, balancing charge comes at the cost of degrading performance.

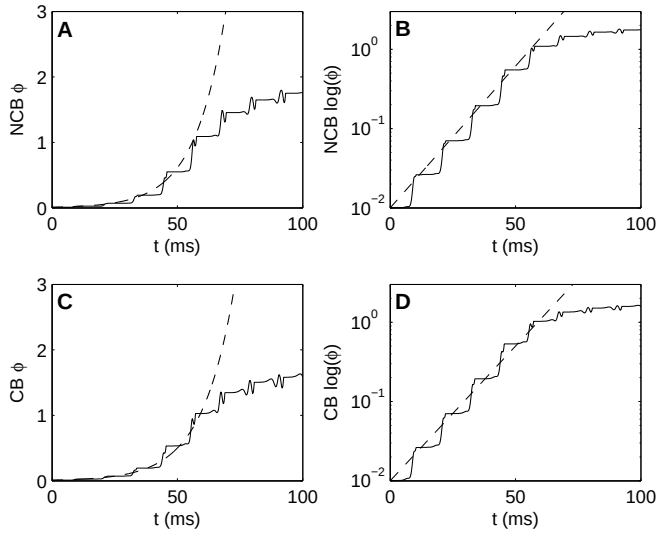

Figure 2.16: A-B show the phase difference between two neurons over time for the NCB optimal stimulus (giving $\Lambda = 0.0823$) applied repeatedly to two neurons with nearly identical initial phases. (C-D) show the same plots for the CB stimulus ($\Lambda = 0.0782$). In both cases the neurons desynchronize at a rate determined by Λ, calculated from (2.4) until $\phi \approx 1$. Dashed lines show exponential functions based on the Lyapunov exponents.

It is natural to wonder to what degree the optimal control found from the phase model will desynchronize neurons that obey the set of equations with which we started. To this end, we simulate the deterministic version of equations (2.28), i.e., with $D = 0$, using a fourth order Runge-Kutta solver. The top panel in Figure 2.17 shows the time histories of three neurons with initial conditions that correspond to $\theta = -0.1, 0, \text{and} + 0.1$ on the periodic orbit. The control is applied every time the reference neuron, with initial

phase $\theta = 0$, fires. Figure 2.17 also shows the input for reference. We find that after 90 milliseconds, the neuron that started at $\theta = +0.1$ now fires approximately 6 milliseconds after the reference neuron, while the neuron that starts at $\theta = -0.1$ fires approximately 2 milliseconds before the reference neuron. The reason for this discrepancy can be explained by noting shape of the PRC. We see in Figure 2.15, the PRC has a local minimum at $\theta \approx 4$. After approximately 9 ms after the optimal stimulus is first presented, the remaining stimulus has negative sign, the reference neuron has a phase corresponding to a positive value on the PRC, and the neuron which starts behind has a phase of $\theta \approx 4$. Because of these conditions near the end of the cycle, the optimal stimulus to brings the phases closer together, negating the desynchronization achieved earlier in the cycle.

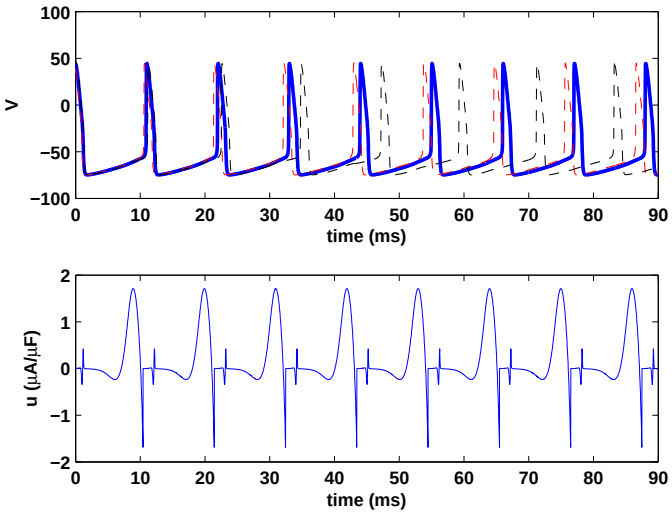

Figure 2.17: Time histories of three neurons with initial conditions $\theta = -0.1, 0, \text{and} + 0.1$ on the periodic orbit. A new stimulus is applied each time the reference neuron with initial phase $\theta = 0$ (shown as a black line) fires. Once the neurons are no longer close in phase, the neuron which started ahead of the reference neuron desynchronizes faster than the neuron that starts behind the reference neuron which can be explained by the shape of the PRC, as described in the main text.

We now apply this optimal control to a network of $N = 100$ coupled, initially syn-

chronized, noisy neurons, with an identical coupling strength of $\alpha_{ij} = 0.05$ and i.i.d. noise with $D = 0.7$. The control logic used is similar to the logic presented in Section 2.5.1, with the controller switching to the *inactive* state if spikes remain below -40mV and switching back to the *active* state if a spike registers above -30mV.

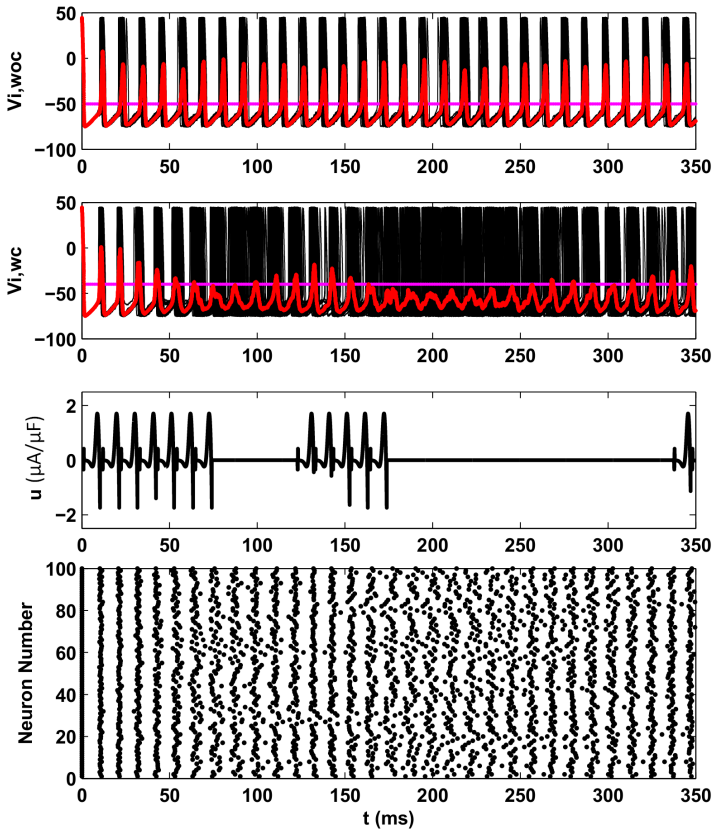

Figure 2.18: Results for a population of $N = 100$ noisy, coupled neurons obeying the reduced Hodgkin-Huxley model. The first panel shows results in the absence of control. The second and third panels show results for the same network with the event-based control applied. The traces give the mean voltages for the system and the horizontal line shows the control activation threshold. Substantial desynchronization can be seen from the raster plot.

Figure 2.18 shows the result of this simulation. The top panel shows voltages of each neuron as well as the average voltage of the coupled, noisy system without control. We find that $\bar{V}$ peaks near 0 mV throughout the simulation, and the neurons stay

synchronized. The second panel shows the individual neuron voltages and mean voltage for the same coupled system with both noise and control input. The horizontal line in this panel represents the threshold voltage of $-30mV$. The control input is shown in the third panel. Clearly the control input desynchronizes the network of neurons, as seen from the raster plot, and since the control is event-based, it only needs to be turned on once $\bar{V}$ crosses the threshold line.

In an experimental setting, it is unlikely that the optimal desynchronizing stimulus $I_{opt}(t)$ can be found exactly. For this reason, we would like to calculate bounds on $I_e(t)$, where $I(t) = I_{opt}(t) + I_e(t)$ with $I(t)$ found using an experimentally obtained PRC such that we can guarantee a given Lyapunov exponent. Using the strategy developed in Section 2.3, we can determine a worst case Λ for a particular $||I_e||_\infty$. Here, we use a shooting method to determine that in order to guarantee stimuli with $\Lambda \geq 0.06, 0.04, 0.02$, and 0, we require $||I_e(t)|| \leq 0.242, 0.44, 0.615$, and 0.797, respectively, for all t. Graphical representations of these error bounds are shown in Figure 2.19. We emphasize that these are only bounds that will *guarantee* a certain value of Λ. As in Section 2.5.1, even if a stimulus falls outside of a shaded region, the associated Λ may still be larger than the value guaranteed for that shaded region. For a network of neurons without coupling or noise, any stimuli with $\Lambda > 0$ should be able to desynchronize an initially coupled neural network. However, for real networks, the value of Λ required for desynchronization will depend on the strength of the coupling and noise.

We now compare the energy used by our NCB, event-based control shown in Figure 2.15 (which we will refer to as the exponential control) to the energy used by a control that will desynchronize a system of neurons by optimally driving them to a phaseless set as developed by Nabi et al. in [44] (which we will refer to as the phaseless control). For a system obeying Ohm's law, the power P applied by an input is given by $P \sim u^2$. A representative portion of a comparison over 1500ms is shown in Figure 2.20. The top two

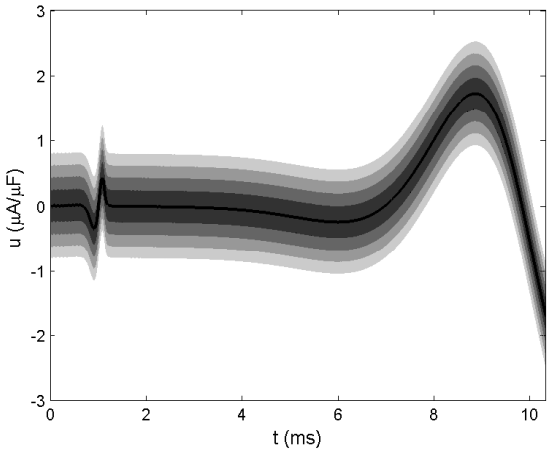

Figure 2.19: The optimal NCB stimulus, shown as a solid line, yields $\Lambda = 0.0823$. In order to guarantee a stimulus with $\Lambda \geq 0.06, 0.04, 0.02$, and 0, we require $||I_e(t)|| \leq 0.242, 0.440, 0.615$, and 0.797, respectively, for all t. Darker shaded regions on the plot correspond to regions with larger guaranteed Lyapunov exponents.

panels show a noisy system of coupled neurons stimulated by the exponential control, while the bottom two panels show the same system stimulated by the phaseless control.

For a single application of the exponential control, $\int_0^{t_1} u^2(t) \approx 2.32$ while for one application of the phaseless control, $\int_0^{T_{end}} u^2(t) \approx 194$. Throughout the simulation, the exponential control is active for more time than the phaseless control; however, one cycle of the exponential control costs much less than one cycle of the phaseless control. Over the entire 1500ms simulation, we find that the total power used is proportional to 236 and 1904 for the exponential control and the phaseless control, respectively. Not only is the maximum applied control much less for the exponential control, but it also uses nearly an order of magnitude less energy.

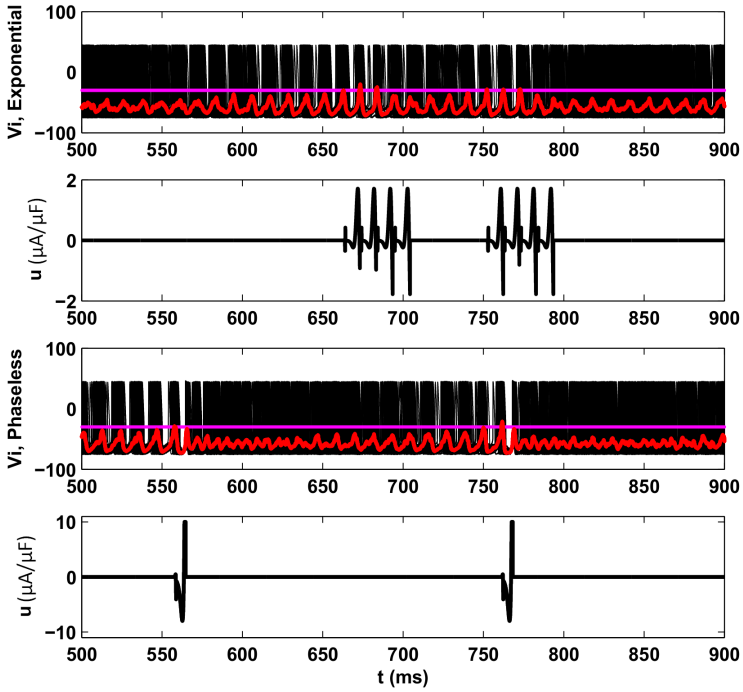

Figure 2.20: Comparing the NCB optimal (exponential) control to the control presented in [44] (phaseless). Representative excerpts from a 1500 ms test of the exponential and phaseless control are shown in the top and bottom panels, respectively. The exponential control fires more often than the phaseless control, but uses less energy because the magnitude of the stimulus is much smaller.

2.5.4 Optimally Decreasing Distribution Peak Height for the Reduced Hodgkin-Huxley Model

Finally, we look from the perspective of controlling distributions by using the strategy described in Section 2.4 to see if we can find a more effective control to desynchronize (2.28) than the control that maximizes the Lyapunov exponent. Here, we attempt to provide reasonable values for the still undetermined parameters in the Euler-Lagrange Equations (2.21)-(2.25). Because we have undetermined parameters, we cannot claim to have found an optimal stimulus to minimize our cost function, but we can still gather

powerful insight about stimuli that will affect network desynchronization. In order to make explicit comparisons to the optimal stimulus which maximizes the Lyapunov exponent, we take $t_1 = 10.34$ms. We choose $\beta = 8$ this time because $\beta = 9$ gives a solution which is too large in magnitude, and invalidates the phase reduction. As in Section 2.5.2, we take our initially synchronized distribution to be a normal distribution with standard deviation $\sigma_d = 0.2$ centered at $\theta = 0$. These conditions were chosen as a reasonable approximation to observed distributions for simulations of the network (2.28) just before the control threshold $\bar{V} = -30$mv is reached. This gives $\theta_M(0) = 0$ and $\rho_M = 1.995$ as initial conditions to (2.22) and (2.23). We optimize the cost function over reasonable values of $\lambda_1(0)$, $\lambda_2(0)$ in order to find the best choice for the remaining initial conditions. Using this approach we find $\lambda_1(0) = 6$ and $\lambda_2(0) = 0$ approximately minimizes the cost function, and gives the stimulus shown in Figure 2.21, which will be refered to as u_D.

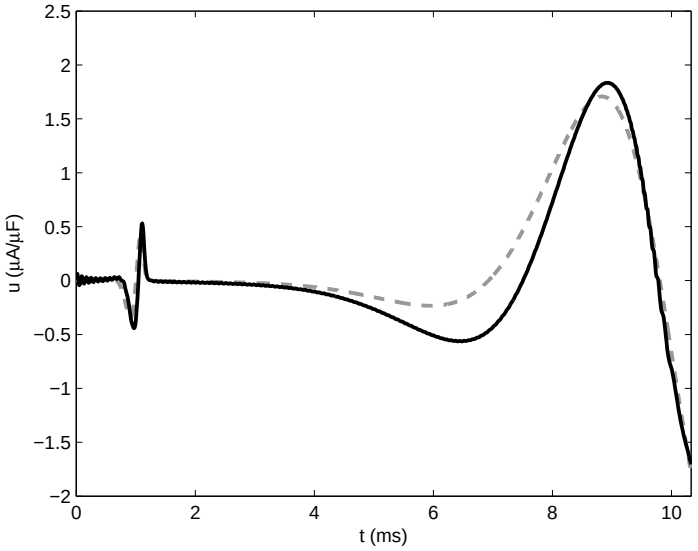

Figure 2.21: The optimal stimuli for decreasing the peak of a distribution is shown as a solid, black line. For reference, the NCB optimal stimulus for maximizing the Lyapunov exponent is shown as a grey, dashed line.

We apply u_D to (2.15) and (2.28) with $N = 250$, $D = 0$ and $\alpha_{ij} = 0$ to determine the

validity of the results found using the phase reduction. For a given neuron, we can deduce its phase at a given time by integrating the reduced Hodgkin-Huxley equations without input or noise until the neuron's next spike. Figure 2.22 shows three snapshots from this simulation. In order to clearly present the results, the mapping $\theta \in (\pi, 2\pi) \to (-\pi, 0)$ is used, e.g., $\theta = \frac{3\pi}{2}$ is plotted as $\theta = -\frac{\pi}{2}$. Theoretical and experimental evolutions are shown in black and blue, respectively. The red curve gives the theoretical evolution of the distribution in the absence of stimulus and noise and is shown for reference. Throughout the simulation, the phase reduction well approximates the dynamics of the 250 neuron system, and the optimal stimulus takes ρ_M from 1.995 to approximately 0.7. The stimuli u_D and u_Λ are nearly identical for the reduced Hodgkin-Huxley mode, with discrepancies in the two answers most likely due to the more restrictive constraints on u_Λ. The similar answers are of interest because we approached the desynchronization problem from two different perspectives.

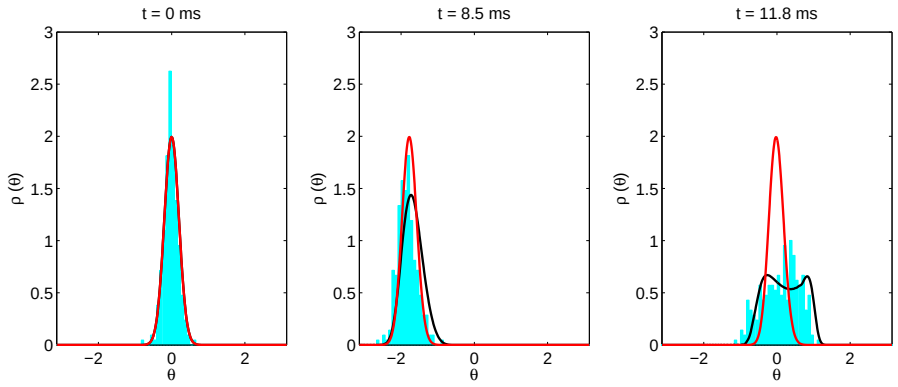

Figure 2.22: Three snapshots from simulations of (2.28) and the associated model (2.15). In order to clearly present the results, the mapping $\theta \in (\pi, 2\pi) \to (-\pi, 0)$ is used, e.g., $\theta = \frac{3\pi}{2}$ is plotted as $\theta = -\frac{\pi}{2}$. Theoretical and numerical evolutions are shown in black and blue, respectively. The red curve gives the theoretical evolution of the distribution in the absence of stimulus and is shown for reference.

To gauge whether u_D is an improvement over u_Λ, we simulate (2.28) with the same parameters and control strategy as the simulations from Figure 2.18. When we use u_D

as the control, the overall energy use is 270 units, while for I_Λ, the energy consumption is 214 units. The cause of this discrepancy is most likely due to the assumption that u_D is Gaussion is no longer valid near the end of the cycle, which wastes a small amount of energy. Results for this simulation do not differ significantly from the results in Figure 2.18, and are not shown.

2.5.5 Comparison to Pulsatile Input

As mentioned in Section 2.1, clinical DBS is currently implemented with a high frequency pulsatile input. While the exact mechanism by which this waveform mitigates the symptoms of Parkinson's disease is unknown, Wilson *et al* postulated that DBS may be effective because it chaotically desynchronizes neurons in the Thalamus region of the brain [21]. They used phase reduction to show that for certain stimulus intensities with frequencies that are approximately twice the natural frequency ($2 \times \frac{1}{T}$) of a neuron, pulsatile stimuli can effectively desynchronize a population of neurons. Using the same conventions as [21], we take $\theta \in [0, 1)$ with $\theta = 0$ corresponding to the spiking of a neuron and scaled to 1 time unit. We use the same PRC as [21], and apply our NCB and CB optimization process with $\beta = 1.5$ and $t_1 = 0.85$. The PRC, its first derivative, and the resulting optimal stimuli are shown in Figure 2.23.

The respective Lyapunov exponents for the NCB and CB optimal stimuli from (2.4) are 1.429 and 1.937, with respective power consumptions ($\int_0^{t_1} u^2 dt$) 1.11 and 1.99. We note that the CB stimulus outperforms the NCB stimulus at the expense of using almost twice as much energy. Figure 2.24 shows the phase separation of two neurons which obey equation (2.2) for the PRC shown in Figure 2.23 subject to repeated iterations of both the NCB (Panels A and B) and CB (Panels C and D) stimuli. As seen in previous sections, the neurons exponentially desynchronize at a rate determined by their Lyapunov

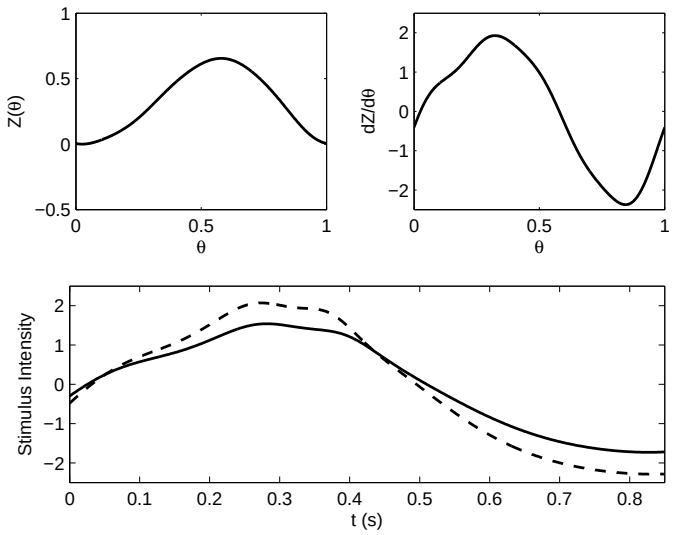

Figure 2.23: Top panels show the PRC and its derivative for the model used in [21]. The bottom panel shows the CB and NCB optimal stimuli as a dashed and solid line, respectively.

exponents until the neurons are nearly antiphase.

We now apply the NCB optimal stimulus to a population of 100 noisy neurons:

$$\dot{\theta}_i = \omega + Z(\theta_i)\left[u(t) + A\gamma_i(t) + B\zeta(t)\right] + \frac{A^2+B^2}{2} Z(\theta_i)Z'(\theta_i), \quad i = 1, \ldots, 100, \quad (2.29)$$

where $\omega = 1$ gives the neural baseline dynamics (recall that $\theta \in [0, 1)$), γ and ζ are individual and common white noise processes with strength $A = \sqrt{0.2}$ and $B = \sqrt{0.3}$ respectively, and $u(t)$ is the common DBS input. The final term in equation (2.29) corresponds to the Ito term for the phase reduction [70]. In order to determine when the optimal stimulus should be presented, we need to know when the average phase of the system of neurons is $\theta = 0$. In real neurons, a spiking event is defined to be $\theta = 0$, and can be observed as a sudden increase in voltage. For the model under consideration, we are simulating a phase reduced model, with no observable voltage spikes. In lieu of $\bar{V}$

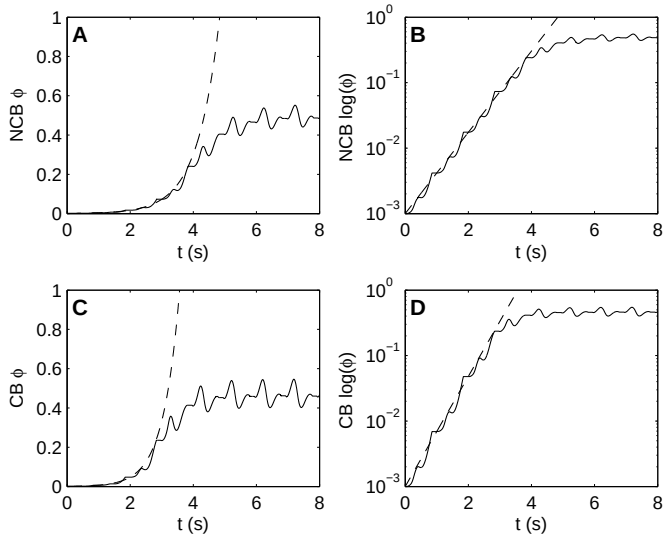

Figure 2.24: A-B show the phase difference between two neurons over time for the NCB optimal stimulus (giving $\Lambda = 1.429$) applied repeatedly to two neurons with nearly identical initial phases. (C-D) give the same plots for the CB stimulus ($\Lambda = 1.937$). In both cases the neurons desynchronize at a rate determined by Λ, calculated from (2.4) until $\phi \approx 0.25$ at which point the solution begins to asymptote to $\phi \approx 0.5$, i.e., the anti-phase state. Dashed lines show exponential functions based on the Lyapunov exponents.

for the system of neurons, we monitor the average phase of the system, $\bar{\theta} = \frac{1}{N}\sum_{i=1}^{N}\theta_i$. Note that $\bar{\theta}$ is equivalent to ψ in the first order Kuramoto parameter, $re^{i\psi} = \frac{1}{N}\sum_{j=1}^{N}e^{i\theta_j}$. For a completely synchronized network, $\bar{\theta}$ varies between 0 and 1, while for a network in the splay state, $\bar{\theta}$ remains constant at 0.5. This gives a continuum from which we can qualitatively gauge the level of synchronization by noting the maximum value of $\bar{\theta}$ on a particular cycle. To determine when to apply a new stimulus, a flag is set when $\bar{\theta} > 0.65$, indicating that the network is still sufficiently synchronized, and a new stimulus begins if the flag is set and $\bar{\theta} < 0.5$, indicating that a majority of the neurons have phase $\theta \approx 0$.

To characterize the desynchronization of the neural network, following [21] we use the

entropy

$$\text{Ent} = \sum_{j=1}^{B} p(\psi_j) \log(p(\psi_j)),$$

where $p(\psi_j)$ is the probability of being in bin j of B total bins. The splay state has the highest entropy for a population of 100 neurons at 4.6.

We note that the results from [21] were computed using a mapping based on the PRC to determine the effect of a DBS pulse. In our trials, we found that adding pulse width somewhat degraded the effectiveness of the pulsatile stimulus. In order to replicate Wilson's results and make comparisons, we use equation (2.29) with $u(t) = 0$ for all time, and instead iterate when a pulse stimulus is presented as follows:

$$\theta_{i+1} = \theta_i + Z(\theta_i)\delta,$$

where δ is the amplitude of the stimulus. For the simulation, we choose $\delta = 0.63$ at a frequency of 1.92 Hz. These are in the range with the best desynchronization capabilities for this model [21]. We use an Euler-Maruyama method [71] to solve (2.29) for both the optimal and pulsatile stimulus, with results shown in Figure 2.25. We find that both stimuli are able to desynchronize the population to similar levels as characterized by the entropy of the system. However, the pulsatile input takes approximately 15 seconds and 29 pulses for the entropy to reach a reasonably steady entropy value of 3.5, while it only takes 5 seconds and three applications of the optimal stimulus to increase the entropy to 3.5. Assuming the power usage $P \sim u^2$, we cannot directly approximate the energy required by the pulsatile input since it has no pulse-width (PW). However, in human trials, [72] and [73] used pulsatile inputs with a PW to period ratio of 7.8×10^{-3} and 9.5×10^{-3} respectively. We take the average of these two ratios to estimate a pulse-width and require $\text{PW} \cdot \text{u} = \delta$, and find that $\text{PW} \sim 0.0045$ and $u \sim 155$ in order to

be therapeutically effective. Letting the power consumption, $P = u^2 \cdot \mathrm{PW}$, we find that $P \sim 108$ units. Conversely, for the optimal stimulus, $\int_0^{t_1} u^2 dt \sim 1.11$. As a very rough estimate of total energy used to achieve steady desynchronization, three applications of the optimal stimulus use 3.33 units of energy while 29 applications of the pulsatile stimulus use 3132 units of energy. Both stimuli are able to desynchronize the neural network, but the optimal stimulus is able to do so using three orders of magnitude less energy.

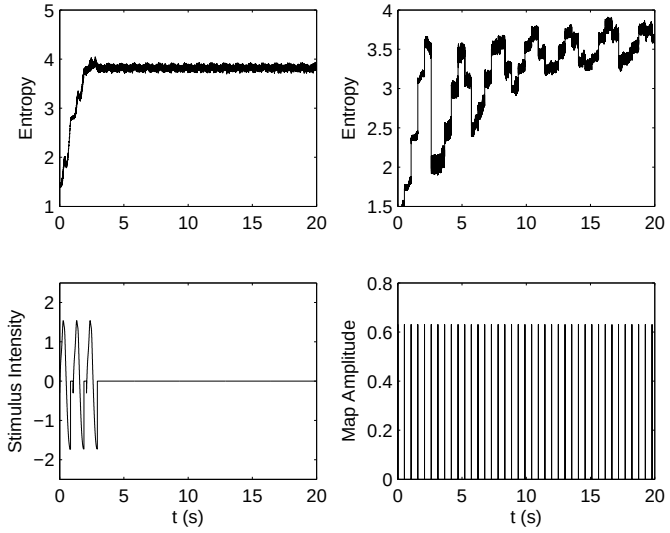

Figure 2.25: Left panels show the entropy of 100 neurons and the applied stimulus for the exponentially desynchronizing control. The right panel shows entropy of 100 neurons for the pulsatile input applied at 1.92 Hz. Note that the bottom-right panel shows the map amplitude, δ, and not the stimulus intensity to emphasize that the stimuli are simulated as delta functions. Each pulse uses 108 units of energy.

2.6 Conclusion

We have described two methods for desynchronizing neural networks: by optimally maximizing the Lyapunov exponent (2.4) and optimally decreasing the peak height of

the phase distribution. Most notably, while both of these methods are based on different perspectives, they produce answers that are quite similar. We find that this method uses three orders of magnitude less energy than a method that uses pulsatile stimuli to achieve desynchronization, which represents an enormous potential savings in battery life of a pacemaker and could also mitigate some of the negative side-effects of DBS. We have also shown that the approach for maximizing the Lyapunov exponent is robust to inaccuracies in finding the optimal stimulus and found bounds for a stimulus derived from a network without coupling that will guarantee a minimum Lyapunov exponent required to give desynchronization for a given network with coupling. Because this method is robust to inaccuracies, it has potential to work well in an *in vitro* setting, and could realistically provide an effective treatment for Parkinson's disease. As a final note, the results presented in this chapter were later extended in [11] to investigate exponential desynchronization using extracellular stimulation.

Chapter 3

Optimal Entrainment of Heterogeneous Noisy Neurons

3.1 Introduction

Precise timing of synchronized oscillators is an important aspect of many biological functions. For example, entrainment of circadian oscillators to a 24-hour light-dark cycle is necessary in nearly all organisms for the maintenance of rhythmic physiological function [2], [74]; irregularities in circadian regulation can contribute to a wide variety of diseases [75–77]. Furthermore, pancreatic cells can be entrained to periodic variations in blood glucose levels, synchronizing the activity of the insulin secreting cells [78], [79]. Also, synchronized patterns of firing neurons give rise to macro scale brain rhythms which are thought to be relevant to cognition and perception [80–83], and in specific examples lack of synchrony can contribute to hearing loss in animals [84], [85].

Promoting synchrony by means of entrainment to an external stimulus could facilitate physiological processes where synchronization is important. Optimal control frameworks can be used to achieve specific control objectives where timing is the control variable

and entrainment is the goal [42], [86], [87]. However, these approaches fall short when heterogeneity in the oscillator properties is large with respect to intensity of the entraining stimulus. Furthermore, most optimal control techniques cannot explicitly account for strong noise in the system, which is often inherent in biological systems (especially in neurons [88], [89]) degrading the efficacy of an optimal stimulus.

We use standard phase reduction techniques to model the response of an oscillator to external perturbations. A phase response curve (PRC) is fit to the phase advance as a function of the phase at which the stimulus is applied [4], [24], [2]. Phase reduction techniques are advantageous because they characterize much of the system's input-output function without the full nonlinear dynamical equations. Using the PRC, we are able to derive sufficient conditions for a stimulus to entrain a noisy, heterogeneous ensemble of phase oscillators. Furthermore, once the sufficient conditions are identified, we can then design efficient external stimuli for entrainment. This strategy does not require the explicit properties of any single oscillator, but only requires the bounds within which all the oscillators' PRCs and natural frequencies must be contained. Using stochastic averaging techniques [90], we can design a stimulus which creates a potential well of minimum depth near the in-phase solution between a nominal oscillator and any other oscillator within the heterogeneous ensemble. Deeper potential wells will be harder to escape from when the noise is present in the system, ensuring entrainment. Unlike other approaches we have used [91], the innovation proposed in this strategy is that the control design explicitly accounts for noise and heterogeneity present in the biological system.

We test the efficacy of the optimal synchronizing stimulus on coupled phase oscillator models compared to other entraining stimuli. We then design optimal stimulus waveforms from previously collected PRCs and test the resulting optimized stimulus *in vitro* on pyramidal neurons from the CA1 region of the hippocampus. The organization of this chapter is as follows. In Section 3.2 we derive the necessary framework for designing

stimuli to entrain a heterogeneous population of oscillators. In Section 3.3, we apply this control strategy to a numerical model of a population of heterogeneous phase oscillators. In Sections 3.4 and 3.5 we provide experimental methods and results, respectively, for entrainment of *in vitro* neurons, and finally in Section 3.6 we discuss our findings and make concluding remarks. This chapter is based on work origionally appearing in [37]. The experiments in this chapter were devised and implemented in collaboration with the Netoff laboratory at the University of Minnesota.

3.2 Efficiently Maximizing the Depth of the Potential Well

Consider the following deterministic phase oscillator

$$\dot{\theta}_1 = \omega_0 + Z(\theta_1)\epsilon u(t). \tag{3.1}$$

Here, $\theta_1 \in [0, 2\pi)$ is the phase of a nominal reference oscillator with natural frequency ω_0 and period $T = 2\pi/\omega_0$, $Z(\theta)$ is the phase response curve, $u(t)$ is an external input, and $0 < \epsilon \ll 1$. Note that we assume that ϵ is small enough so that higher order noise terms are negligible (c.f. [92], [93]). A second noisy oscillator, θ_2, whose parameters are not fully known, can be represented as follows:

$$\dot{\theta}_2 = \omega_0 + \Delta\omega + \left[Z(\theta_2) + \Delta Z(\theta_2)\right]\epsilon u(t) + \left[Z(\theta_2) + \Delta Z(\theta_2)\right]\epsilon\eta(t). \tag{3.2}$$

This function is illustrated in panel A of Figure 3.1. The variable $\Delta\omega \in [-\epsilon\omega_-, \epsilon\omega_+]$ represents some uncertainty in the natural frequency, ΔZ represents uncertainty in the phase response curve, and $\eta(t) = \mathcal{N}(0, 1)$ is i.i.d zero mean white noise with vari-

ance 1. However, for the population, we can determine bounds for the range of the PRCs $-E_-(\theta) \leq \Delta Z(\theta) \leq E_+(\theta)$ with strictly nonnegative functions $E_-(\theta)$ and $E_+(\theta)$. Intuitively, θ_1 in equation (3.1) represents the nominal parameters of an oscillator while θ_2 in equation (3.2) accounts for the uncertain terms, which may not be fully known, as well as noise that might be present in the system. Alternatively, θ_2 can represent a range of properties for a heterogeneous population of oscillators.

We assume that the reference oscillator is entrained to the external stimulus so that

$$\int_0^T \left[\omega_0 + Z(\theta_1)\epsilon u(t)\right] dt = 2\pi. \tag{3.3}$$

Therefore, to guarantee that the noisy, uncertain oscillator is also entrained, our goal is to design $u(t)$ such that the phase difference between the two oscillators is small. Defining $\phi = \theta_2 - \theta_1$, we may write

$$\dot{\phi} = \Delta\omega + \left[Z(\theta_1 + \phi) - Z(\theta_1) + \Delta Z(\theta_1 + \phi)\right] \epsilon u(t) + \left[Z(\theta_1 + \phi) + \Delta Z(\theta_1 + \phi)\right] \epsilon\eta(t). \tag{3.4}$$

Here, ϕ gives the phase difference between the nominal oscillator and the noisy, unknown oscillator so that when $\phi = 0$, both oscillators are in phase. Asymptotically expanding θ_1 in powers of ϵ yields

$$\theta_1 = \theta_1^{(0)}(t) + \epsilon\theta_1^{(1)}(t) + \epsilon^2\theta_1^{(2)}(t) + \ldots \tag{3.5}$$

Note that all terms of equation (3.4) are $\mathcal{O}(\epsilon)$, which implies that $\theta_1^{(0)}(t) = \theta_1(0) + \omega_0 t$ so that $\theta_1(t) = \theta_1(0) + \omega_0 t + \mathcal{O}(\epsilon)$. For convenience, we take $\theta_1(0) = 0$, but note that the analysis to follow could still be performed for $\theta_1(0) \neq 0$. Substituting (3.5) into (3.4)

and Taylor expanding terms of the form $Z(\cdot)$ in powers of ϵ yields

$$\begin{aligned}\dot{\phi} &= \Delta\omega + \left[Z(\omega_0 t + \phi) - Z(\omega_0 t) + \Delta Z(\omega_0 t + \phi)\right] \epsilon u(t) \\ &\quad + \left[Z(\omega_0 t + \phi) + \Delta Z(\omega_0 t + \phi)\right] \epsilon \eta(t) + \mathcal{O}(\epsilon^2). \end{aligned} \tag{3.6}$$

Through stochastic averaging [94], [90] in the limit of small ϵ, we can approximate ϕ in equation (3.4) by φ where

$$\begin{aligned}\dot{\varphi} &= \Delta\omega + \underbrace{\frac{1}{T}\int_0^T \left[Z(\omega_0 t + \varphi) - Z(\omega_0 t)\right] \epsilon u(t) dt}_{f(\varphi)} \\ &\quad + \underbrace{\frac{1}{T}\int_0^T \left[\Delta Z(\omega_0 t + \varphi)\right] \epsilon u(t) dt}_{e(\varphi)} + \epsilon\sigma\eta(t) + \mathcal{O}(\epsilon^2) \\ &= \Delta\omega + f(\varphi) + e(\varphi) + \epsilon\sigma\eta(t) + \mathcal{O}(\epsilon^2), \end{aligned} \tag{3.7}$$

where $f(\psi)$ represents the known part of (3.4), $e(\psi)$ represents the uncertain part of (3.4), and

$$\begin{aligned}\sigma^2 &= \frac{1}{T}\int_0^T \left[Z(\omega_0 t + \varphi) + \Delta Z(\omega_0 t + \varphi)\right]^2 dt \\ &= \frac{1}{T}\int_0^T \left[Z(\omega_0 t) + \Delta Z(\omega_0 t)\right]^2 dt \end{aligned} \tag{3.8}$$

determines the strength of the noise. Note that equivalence in equation (3.8) comes from periodicity in the PRC. The equation for the probability distribution function associated with equation (3.7) is given by the Fokker-Planck equation [70]

$$\frac{\partial \rho}{\partial t} = -\frac{\partial}{\partial \varphi}\left[A(\varphi)\rho(t,\varphi)\right] + \frac{1}{2}\frac{\partial^2}{\partial \varphi^2}\left[B\rho(t,\varphi)\right], \tag{3.9}$$

where $\rho(t,\varphi)$ is a probability density function, $A(\varphi) = \Delta\omega + f(\varphi) + e(\varphi)$ and $B = \epsilon^2\sigma^2$. From this perspective, it is not possible to maintain indefinite entrainment of the noisy neuron, because there is always a chance that noise in the system could push the neuron arbitrarily far from $\varphi \approx 0$. However, to reduce the likelihood of this event, the problem of entraining a noisy neuron to a periodic stimulus can be viewed as maximizing the average escape time from $\varphi = 0$ over a potential barrier. From this perspective, for an oscillator whose probability density obeys (3.9), $\int -A(\varphi)d\varphi$ can be viewed as a potential function. Therefore, our goal is to design a stimulus so that there is a potential barrier with a minimum near $\varphi \approx 0$. The escape time from this barrier can be expected to be proportional to the exponential of the height of the potential barrier [70].

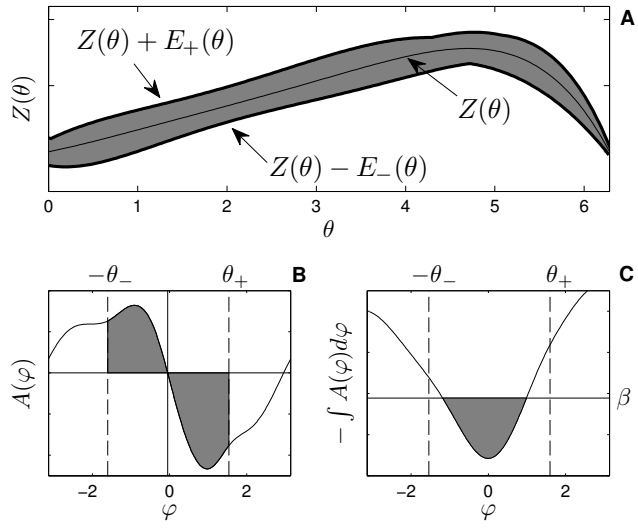

Figure 3.1: Panel A gives a visual representation of the uncertainty allowed in the phase response curve from equation (3.1), i.e. any PRC that can be drawn inside the shaded region is allowable. Panel B shows an example of $A(\psi)$. In order to meet the requirements (3.10) and (3.11), for any oscillator with any allowable PRC, the area of each shaded region in panel B must be at least β. If this is the case, the resulting potential well in panel C will be at least β high by the time ψ is smaller than $-\theta_-$ or larger than θ_+.

To maximize the escape time we want to design a stimulus $u(t)$ such that the change in a potential trough near $\varphi = 0$ and a potential peak located at either $\varphi = \theta_+ > 0$ or

$\varphi = -\theta_- < 0$ is greater than or equal to β. This requirement can be stated as:

$$\int_0^{\theta_+} -A(\varphi)d\varphi \geq \beta, \tag{3.10}$$

$$\int_0^{-\theta_-} -A(\varphi)d\varphi \geq \beta. \tag{3.11}$$

Here β can be thought of as the minimum height required for the potential well near $\varphi = 0$. Alternatively, β in panel B of Figure 3.1 represents the minimum area in each shaded region required to produce a potential well with a size of at least β between $-\theta_-$ and θ_+, as in panel C.

Recall that to leading order in ϵ, $A(\varphi) = \Delta\omega + f(\varphi) + e(\varphi)$, and substituting these terms from equation (3.7) into (3.10) gives

$$\begin{aligned}
&\int_0^{\theta_+} \left[\Delta\omega + \frac{1}{T}\int_0^T \left[Z(\omega_0 t + \varphi) - Z(\omega_0 t)\right]\epsilon u(t)dt \right.\\
&\qquad\qquad \left. + \frac{1}{T}\int_0^T \left[\Delta Z(\omega_0 t + \varphi)\right]\epsilon u(t)dt\right] d\varphi \leq -\beta,\\
&\frac{1}{T}\int_0^T \left[\Delta\omega\theta_+ + \left(\int_0^{\theta_+} \left[Z(\omega_0 t + \varphi) - Z(\omega_0 t)\right]d\varphi\right)\epsilon u(t)\right.\\
&\qquad\qquad \left. + \left(\int_0^{\theta_+} \left[\Delta Z(\omega_0 t + \varphi)\right]d\varphi\right)\epsilon u(t)\right] dt \leq -\beta.
\end{aligned} \tag{3.12}$$

By noting that

$$\int_0^{\theta_+} -E_-(\omega_0 t + \varphi)d\varphi \leq \int_0^{\theta_+} \Delta Z(\omega_0 t + \varphi)d\varphi \leq \int_0^{\theta_+} E_+(\omega_0 t + \varphi)d\varphi, \tag{3.13}$$

we can use this inequality in (3.12) to conclude that if

$$\frac{1}{T}\int_0^T \left[\epsilon\omega_+\theta_+ + \left[g_+(t) + E_p(t,u)\right]\epsilon u(t)\right] dt \leq -\beta, \tag{3.14}$$

where

$$E_p(t,u) = \begin{cases} \int_0^{\theta_+} E_+(\omega_0 t + \varphi) d\varphi & \text{if } u \geq 0, \\ \int_0^{\theta_+} -E_-(\omega_0 t + \varphi) d\varphi & \text{if } u < 0, \end{cases} \tag{3.15}$$

$$g_+(t) = \int_0^{\theta_+} \left[Z(\omega_0 t + \varphi) - Z(\omega_0 t) \right] d\varphi, \tag{3.16}$$

then (3.12) and hence (3.10) must also hold. Using similar logic, (i.e. manipulating (3.11) so it is in the same form as (3.12) then using the inequality (3.13)) we can conclude that

$$\frac{1}{T} \int_0^T \left[-\epsilon \omega_- \theta_- + [g_-(t) + E_m(t,u)] \epsilon u(t) \right] \geq \beta, \tag{3.17}$$

where

$$E_m(t,u) = \begin{cases} \int_{-\theta_-}^0 -E_-(\omega_0 t + \varphi) d\varphi & \text{if } u \geq 0, \\ \int_{-\theta_-}^0 E_+(\omega_0 t + \varphi) d\varphi & \text{if } u < 0, \end{cases} \tag{3.18}$$

$$g_-(t) = \int_{-\theta_-}^0 \left[Z(\omega_0 t + \varphi) - Z(\omega_0 t) \right] d\varphi, \tag{3.19}$$

is a sufficient condition for (3.11) to be true.

Thus, the control objective of creating a potential well that is at least β deep can be accomplished by designing a stimulus such that (3.14) and (3.17) are satisfied. We can solve for an energy-optimal stimulus which accomplishes this goal with a Hamilton-Jacobi-Bellman (HJB) approach [13] by defining the auxiliary state vector z such that

$$\dot{z} = \begin{bmatrix} \dot{a} \\ \dot{b} \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} \frac{1}{T}(\epsilon \omega_+ \theta_+ + [g_+(t) + E_p(t,u)] \epsilon u(t)) \\ \frac{1}{T}(-\epsilon \omega_- \theta_- + [g_-(t) + E_m(t,u)] \epsilon u(t)) \\ \omega_0 + Z(\theta) \epsilon u(t) \end{bmatrix}, \tag{3.20}$$

where a and b are auxiliary variables which come from the constraints (3.14) and (3.17), respectively. The variable θ is included so that we can specify the end point conditions

$$\theta(0) = 0 \text{ and } \theta(T) = 2\pi, \tag{3.21}$$

requiring the noiseless oscillator with nominal properties, θ_1, to be perfectly entrained to the external stimulus. Hence, the true noisy uncertain oscillator θ_2 will also be entrained to the external stimulus when it is inside the potential well at $\varphi \approx 0$.

For the initial state $z = [0, 0, 0]^T$, the energy optimal stimulus will minimize

$$J(z, u(t)) = \int_0^T u^2 dt + q(z(T)), \tag{3.22}$$

where $\int_0^T u^2 dt$ represents the energy consumed by the stimulus, and $q(z(T))$ is an endpoint cost function where $q(z([a(T), b(T), \theta(T)]^T))$ is small for final states states where $a(T) \leq -\beta$, $b(T) \geq \beta$ and $\theta(T) = 2\pi$, and large otherwise. This endpoint cost is chosen to give a prohibitive penalty if the stimulus $u(t)$ does not satisfy the required constraints (3.14), (3.17), and (3.21). The energy optimal stimulus, $u^*(t)$, can be found with standard HJB techniques, and a related example is given in greater detail in [91].

As a final note, we examine how the optimal stimulus changes when the natural frequency ω_0 changes and both θ_- and $\theta_+ = 0$. To this end, suppose that we have already solved (3.22) for the optimal stimulus $u^*(t)$ when the natural frequency is $\omega_0 = 2\pi/T_o$. Suppose that $u^*(t) = \epsilon_1 u_o(t)$, to leading order, the requirement (3.3) is

$$\int_0^{T_o} [\omega_0 + Z(\omega_0 t)\epsilon_1 u_o(t) + \mathcal{O}(\epsilon_1^2)]dt = 2\pi. \tag{3.23}$$

Now consider a different natural frequency $\omega_1 = 2\pi/T_1$. The requirement (3.3) is

$$\int_0^{T_1} [\omega_1 + Z(\omega_1 t)\epsilon_1 u_1(t) + \mathcal{O}(\epsilon_1^2)]dt = 2\pi. \tag{3.24}$$

Changing variables so that $\tau = \frac{T_o}{T_1}t$ we can rewrite (3.24) as

$$\begin{aligned}
\int_0^{T_o} \left[\omega_0 + \frac{T_1}{T_o} Z(\omega_0\tau)\epsilon_1 u_1(\tau) + \mathcal{O}(\epsilon_1^2)\right] d\tau &= 2\pi \\
\int_0^{T_o} \left[\frac{T_1}{T_o} Z(\omega_0\tau)\epsilon_1 u_1(\tau) + \mathcal{O}(\epsilon_1^2)\right] d\tau &= 0 \\
\int_0^{T_o} \left[\omega_0 + Z(\omega_0\tau)\epsilon_1 u_1(\tau) + \mathcal{O}(\epsilon_1^2)\right] d\tau &= 2\pi.
\end{aligned} \tag{3.25}$$

Notice that (3.25) is equivalent to (3.23). One can verify that using the same change of variables, the constraints (3.14) and (3.17) can be made identical for the two natural frequencies. Thus the auxiliary state dynamics $\dot{z} = [\dot{a}, \dot{b}, \dot{\theta}]^T$ will be the same, which implies $u_0(t) = u_1(\tau) = u_1\left(\frac{T_o}{T_1}t\right)$, which is a useful property from an experimental perspective.

3.3 Numerical Results

For numerical validation of the theory, we apply the methods from Section 3.2 to a large population of $N = 1000$ noisy phase oscillators,

$$\theta_j = \omega_j + Z_j(\theta_j)u(t) + \eta_j(t), \qquad j = 1, \ldots, N. \tag{3.26}$$

Here, $Z_j(\theta)$ is constrained to be within the envelope from the left panel of Figure 3.2, which is determined from experimental calculations of pyramidal neurons from the CA1 region of the hippocampus. We choose the envelope in this way to mimic the *in vitro*

experiments performed in the sections to follow. To determine the PRC for each oscillator, 11 control points are randomly chosen at equally spaced intervals within the envelope, and $Z_j(\theta)$ is linearly interpolated between the control points. Example PRCs are shown in Figure 3.2. We note that this envelope is relatively large, so the variance in ω_j is taken to be relatively small with $\omega_j = \omega_0 + \Delta\omega$, with $\Delta\omega \in [-0.004\omega_0, 0.004\omega_0]$ chosen from a uniform distribution. A larger variance in the natural frequencies could be chosen if the envelope of possible PRCs is smaller. We also take $n_j(t) = \sqrt{0.05\nu_j}\mathcal{N}(0,1)$ to be i.i.d. zero mean white noise with variance $0.05\ \nu_j$, where $\nu_j = \frac{1}{2\pi}\int_0^{2\pi} Z_j^2(\theta)d\theta$.

For calculation of the optimal stimulus, we take the nominal PRC to be the average of the PRCs taken from multiple CA1 pyramidal cells, which is close to the average between the top and bottom curves in panel A of Figure 3.2. We take $\theta_+ = \theta_- = 0.94$, $\beta = 10^{-4}$, and $T = 24$ ms. The optimal control is shown as $u^*(t)$ in panel B of Figure 3.2 with the other applied stimuli shown in panels C-E. Generally, the optimal control seeks to apply a positive (resp. negative) stimulus when the slope of the PRC is negative (resp. positive) and when the derivative is large in magnitude relative to the size of the envelope. For example, a large positive stimulus is given near the end of the cycle when the derivative is very negative and the uncertainty is relatively small; conversely, no stimulus is given near the beginning of the cycle when the slope is small in magnitude and the uncertainty is high. We also compare the resulting entrainment when using the optimal stimulus to the entrainment using a sine wave, square wave, and square pulse, u_{sin}, u_{sq}, and u_{pul}, respectively, each using an equivalent amount of power. We simulate the system (3.26) for $60T$ with $u(t)$ taken to be one of these four stimuli applied periodically and report the time at which the cells spike (i.e. cross $\theta = 2\pi$) as a probability distribution ρ relative to the phase of the periodic stimulus. Results are shown in panel F of Figure 3.2. We find that the optimal stimulus gives the sharpest distribution of spike times. The coefficient of variation (CV) from a sample of 60000 spike times are reported in Panel G of Figure

3.2 with error bars representing a 95 percent confidence interval assuming that the spike time distribution is well approximated by a normal distribution.

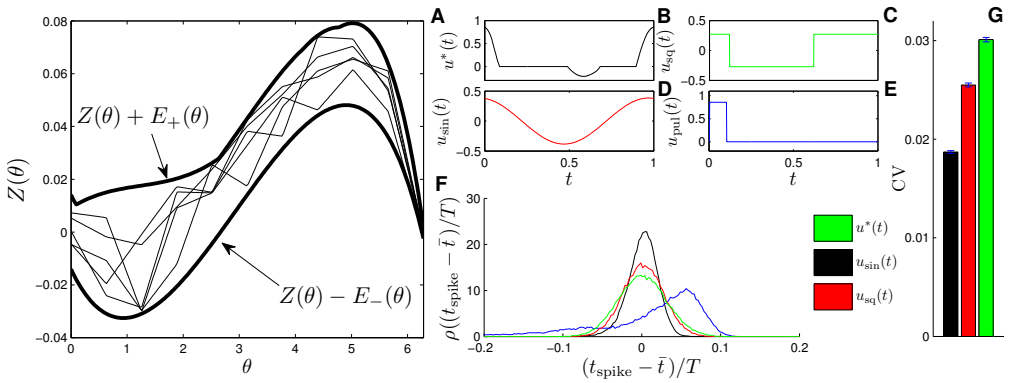

Figure 3.2: Panel (A) shows the envelope in which the PRCs of each of the phase oscillators fit, shown as thick lines. Examples of individual PRCs are shown as thin lines. The optimal entraining stimulus u^* and three other stimuli $u_{\rm sin}$, $u_{\rm sq}$, and $u_{\rm pul}$ (shown in panels (B-E)) are applied periodically to test their entrainment of the noisy ensemble (3.26). Panel (F) shows a probability density of spike times relative to the phase of the entraining stimulus, $\rho(t_{\rm spike}-\bar{t}\,)$, where $\bar{t}$ is the average spike time. We find that the optimal stimulus yields a significantly tighter distribution of spike times, as reflected in the coefficient of variation shown in panel (F) calculated from their distribution of spike times. We note that the coefficient of variation for the pulsed stimulus is 0.091, and do not show it on the graph because it is much larger than the other values.

3.4 Experimental Methods

To test the efficacy of the optimized stimulus waveform in a biological system, we designed stimulus waveforms to entrain hippocampal CA1 pyramidal neurons in a brain slice preparation. PRCs were first measured from several pyramidal neurons to estimate the variability in the PRC waveform. Then, optimized stimulus waveforms were designed and applied to neurons using patch clamp recording techniques. Experimental protocol was developed in collaboration with the Netoff laboratory at the University of Minnesota and patch clamp experiments were performed by the Netoff laboratory. All experimen-

tal procedures were performed following guidelines from Research Animal Resources of the University of Minnesota and approved by the Institutional Animal Care and Use Committee.

3.4.1 Electrophysiology Recordings

Hippocampal brain slices were prepared from Sprague Dawley rats aged 14-21 days old. Rats were deeply anesthetized using isoflurane before decapitation and extraction of the brain. Following extraction, the brain was chilled in artificial cerebral spinal fluid (aCSF) composed of (in mM): 125 NaCl, 25 $NaHCO_3$, 11 D-glucose, 3 KCl, 1.25 NaH_2PO_4, 2 $CaCl_2$, and 1 $MgCl_2$. Transverse slices of the hippocampus were sectioned 400 μm thick using a Vibratome. Slices were oxygenated with 95% O_2 and 5% CO_2 and incubated at 37 °C for at least one hour. Slices were visualized using differential interference contrast optics (Olympus, Center Valley, PA) while in a chamber with circulating aCSF. Patch-clamp electrodes (3-6 MΩ) were pulled from borosilicate glass (P-97 micropipette puller; Sutter Instrument) and filled with intracellular recording fluid composed of (in mM): 120 K-gluconate, 10 HEPES, 1 EGTA, 20 KCl, 2 $MgCl_2$, 2 Na_2ATP, and 0.25 Na_3GTP. Recordings from whole-cell patch clamped CA1 pyramidal neurons in the hippocampus were made using a current-clamp amplifier (MultiClamp 700B; Axon Instruments, Molecular Devices, Sunnyvale, CA). Data were collected using the Real-Time eXperimental Interface (RTXI) software publicly available (www.rtxi.org) and sampled at 5 kHz.

3.4.2 Estimating PRCs from neurons

To estimate PRCs from CA1 pyramidal neuron neurons, stimuli were applied at different phases of the neuron's interspike interval and deviations from the unperturbed

period was measured. PRCs were estimated as previously described in [23], c.f. [25] (see also Appendix B). Briefly, short-duration (0.02-0.4ms) current pulses (300-400pA) were injected into the periodically firing neuron through the patch clamp electrode to elicit a significant phase change without inducing an action potential. Each data point was obtained using a direct method (see Appendix B) by stimulating at a random phase θ, and measuring the change in spike time with the resulting value $Z(\theta)$ equal to $\Delta\theta/Q$, where $\Delta\theta$ is the change in phase and Q is the charge injected by the pulse. Constant drive or an oscillatory input to these neurons causes them to fire periodically. To compensate for drift in the neuron's natural firing rate over the experiment, a proportional-integral (PI) controller was used to adjust the current applied to the neuron slowly to maintain the neuron at a target firing rate [95]. Spike advance as a function of the stimulus phase was fit with a low order polynomial constrained to zero at the beginning and end of the phase by minimizing least squares error (Matlab's `fminsearch`). Examples of PRCs are shown in Figure 3.3. We note that the waveform optimization presented in Section 3.2 requires the mean phase advance estimated by the PRC to be within the envelope, but the phase advance on any particular cycle can be outside the envelope, due to noise.

3.4.3 Stimulus Waveform

PRCs from ten CA1 pyramidal cells, collected for previous experiments under similar conditions described here [23, 96], were used to design optimal stimulus waveforms. The PRCs used for optimization had slightly different shapes, so four envelopes and their corresponding stimulus waveforms were calculated; shown in Figure 3.3. The parameters used in this optimization were $\theta_+ = \theta_- = 1.89$, and $\beta = 0.0024$. We assume that we have direct control over the natural frequency of the neuron with the PI controller and set $\Delta\omega = 0$. Each optimal waveform was defined by an envelope containing all PRCs

within that group of cells. For each envelope, one optimal and two suboptimal stimulus waveforms with equal energy were generated. The suboptimal waveforms were created by either inverting and time-shifting the optimal waveform, or by stretching out the positive portion of the optimal waveform and renormalizing to preserve the total energy. For the *in vitro* experiments, whole-cell patch clamp recordings were made from CA1 pyramidal neurons. For each cell, Matlab was used to determine which envelopes the measured PRC fit within. The stimuli for the envelope with the best margins for each cell were applied as current through the patch clamp electrode. For some cells, PRCs fit within multiple envelopes, and all were tested if possible. Each of the three waveforms were applied continuously for at least 30 seconds to a few minutes. The stimulus waveform was applied at the target frequency of the neuron, set at 10 Hz using the PI controller, for the duration of the experiment. The peak-to-peak amplitude of the waveform was less than $1nA$. The sequence in which the waveforms were applied was selected at random. In most cases the PI controller to hold the neuron at the target firing rate was on while the stimuli were being applied, however in a few cases the PI controller was turned off to ensure it was not affecting the synchrony. The amplitude of stimulation was chosen so that the stimulus waveform could be seen in the baseline membrane potential without eliciting a spike. The experimenter was blinded to which stimulus was optimal until after completion and analysis of all experiments.

3.4.4 Entropy Estimation

Entropy values calculated from spike density histograms (Figure 3.5) were used to compare how well a stimulus entrained the neuron. Data were analyzed using Matlab. For entropy calculations, we subdivide phases into B equally spaced bins and denote $P(i)$ as the probability that a spike occurs in bin i. An entropy bias term was used to correct

for the different number of spikes in each trial [97]:

$$\text{Entropy}_{\text{bias}} = \frac{B-1}{2N}, \tag{3.27}$$

where N is the total number of spikes. To calculate the unbiased normalized entropy measure from each spike density histogram, the entropy, accounting for the bias, was normalized by the maximum possible entropy:

$$\text{Entropy} = \frac{\sum_{i=1}^{B} P(i) \log P(i) - \text{Entropy}_{\text{bias}}}{B \ln \frac{1}{B}}. \tag{3.28}$$

The standard error of the entropy was estimated as follows [97]:

$$\text{SEM} = \sqrt{\frac{1}{N} \sum_{i=1}^{B} (\log(1 - P(i)) + \text{Entropy})^2 P(i)(1 - P(i))}. \tag{3.29}$$

Statistical comparisons between entropy values were made using the Student's t-test, and p values < 0.05 were considered significant.

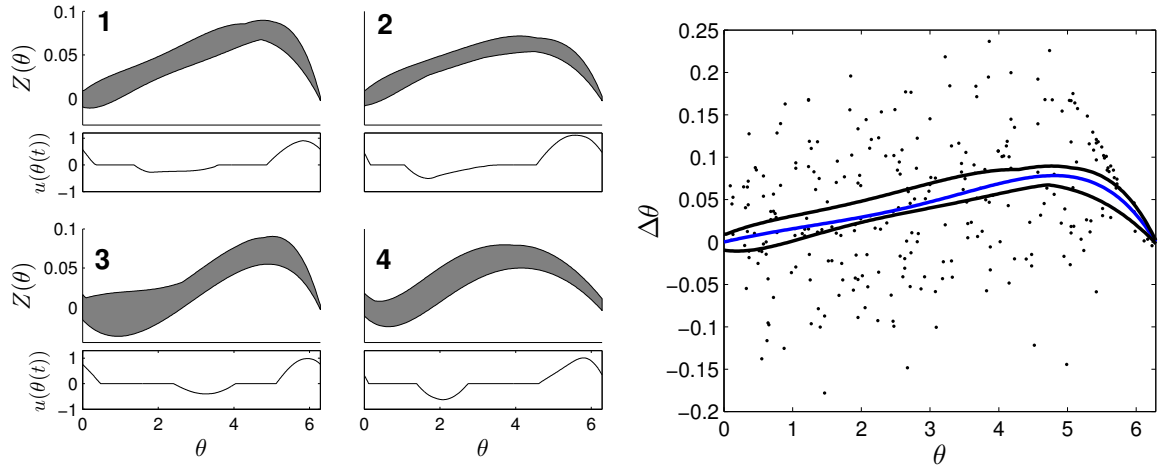

Figure 3.3: Four envelopes with separate optimal waveforms were used to determine which stimuli to use on a given cell. The left panels show envelopes in gray with corresponding optimal stimulus waveforms directly below. On the right panel, a PRC (blue) calculated from individual measurements of $\Delta\theta/Q$ (dots) from a CA1 pyramidal neuron fits within the black curves of envelope 1.

3.5 Experimental Results

Stimulus waveforms were applied to ten CA1 pyramidal neurons. An example cell can be seen in Figure 3.4. The PRC from this example neuron fit within envelope 3 best. For each stimulus the coefficient of variation of the interspike intervals, and the entropy of the spike times with respect to the phase of the stimulus waveform was measured. In this cell, the optimal stimulus waveform resulted in the lowest coefficient of variation in the interspike intervals, indicating that the cell fires more periodically than with the suboptimal waveforms. Furthermore, the optimal stimulus waveform had the lowest entropy of spike times with respect to the stimulus phase, indicating that the neurons phase locked to the optimal stimulus better than the suboptimal stimulus waveforms.

The optimal stimulus was the best, compared to the suboptimal waveforms with the same amount of power, at entraining across all recorded neurons where the stimulus had a significant effect on the entropy. Figure 3.5 shows the entropy values for stimuli across

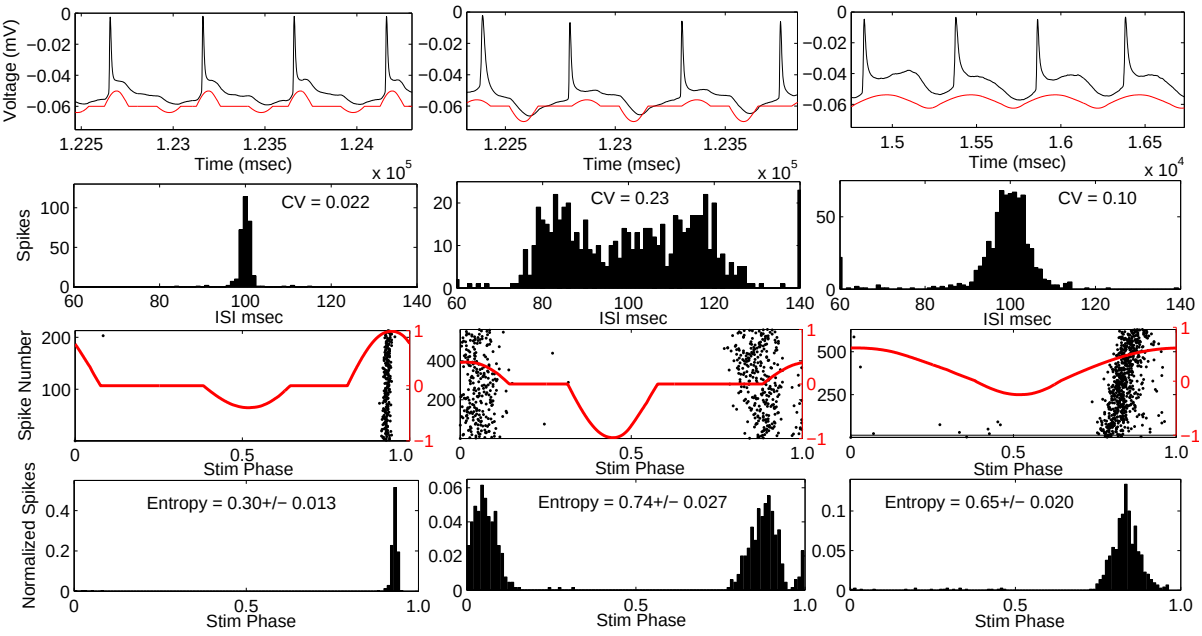

Figure 3.4: Example cell using envelope 3. Response to optimal stimulus is plotted in left column and two suboptimal stimuli applied in right columns. Top Row: voltage trace (black) and applied stimulus waveform (red). Second row: Histograms of inter-spike-intervals. Coefficient of variation (CV) values are indicated. Third row: phase of the stimulus at each action potential (black dots) with stimulus waveform (red). Bottom row: spike density histogram with respect to stimulus phase. Entropy values ± SEM are indicated.

all cells. Stimuli from envelope 1 were applied to seven cells. For six out of the seven cells, the entropy values for the optimal waveform were significantly lower ($p < 0.05$) than the non-optimal waveforms, as tested with a Student's T-test. For cell number 1, the entropy remained high across all stimuli without any noticeable effect from any of the waveforms, perhaps because the stimulus amplitude was too low. Stimuli from envelope 2 were applied to one cell, from envelope 3 were applied to four cells, and from envelope 4 were applied to one cell. For each of these cells, the entropy values were significantly lower for the optimal waveform than the suboptimal waveforms. We conclude that the optimal stimulus waveform was the most effective at entraining the neurons to the stimulus. In three cells experiments were done without the PI controller to control the firing rate to confirm that the PI controller was not affecting the findings; the results in these cells were consistent with the experiments done with the PI controller.

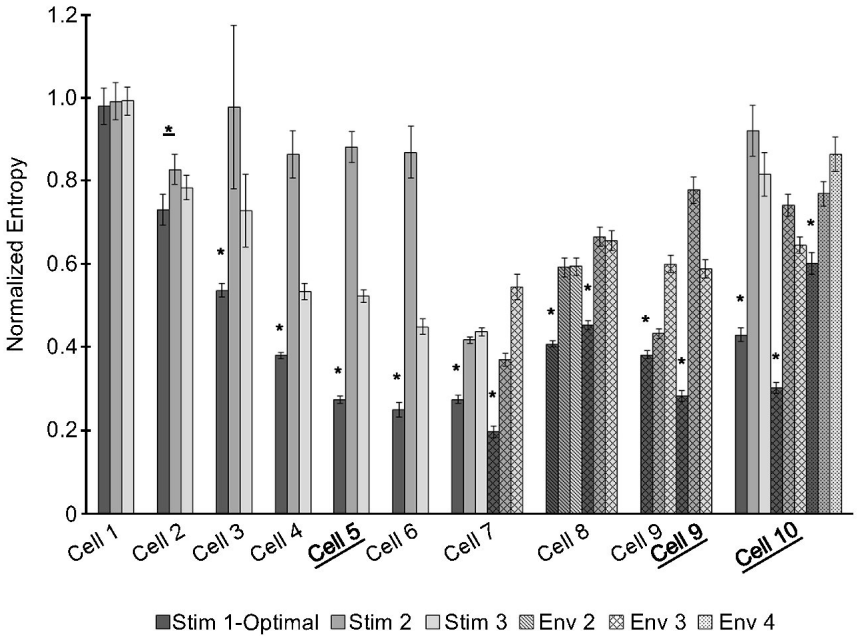

Figure 3.5: The optimal stimulus waveform is significantly better at entraining neurons to the stimulus across cells. Entropy values for each stimulus applied are shown for 10 cells. Some cells had stimuli from more than one envelope applied. Envelopes are indicated by different patterns, with envelope 1 being the solid fill. For each envelope, three stimuli were applied: the waveform optimized for entraining the neuron (dark gray) and two sub-optimal waveforms (gray and white). Certain cells did not have the PI controller on to control the firing rate of the neuron (underlined). Significant differences between the optimal stimulus waveform and the other waveforms at p values<0.05 are indicated by $*$.

3.6 Discussion

In this chapter we have developed an approach to generate an optimal stimulus waveform to achieve entrainment of a noisy, heterogeneous group of phase oscillators. The waveforms were tested in computational models and *in vitro* neurons. The optimal waveform for entrainment was determined by maximizing the average escape time from a potential well near the entrained solution for any oscillator within the ensemble. We note that this optimal control methodology only requires bounds, (3.14) and (3.17), in which all oscillators in the population must be contained. While a stimulus waveform may be designed to entrain a particular oscillator in the population, the resulting stimulus may not entrain another oscillator within the population resulting in poor entrainment overall. Our method uses a worst case scenario approach to optimization which guarantees that each individual oscillator will be well entrained by the resulting stimulus leading to better entrainment at an ensemble level rather than optimizing the waveform for a single representative cell within the population. Allowing for uncertainty in PRCs could be particularly useful in neurons because there is often a great degree of variability in PRCs between samples, as seen in this study as well as in [98] and [99].

While the numerical methodology developed in this chapter generates energy-optimal stimuli to achieve entrainment of an ensemble, it is difficult to experimentally prove that a given stimulus is truly optimal. However, the experimental evidence suggests that the resulting stimuli are probably at least close to optimal. Experimental results *in silico* in Section 3.3 demonstrate that our optimized stimulus resulted in better entrainment in a heterogeneous population of oscillators than other waveforms. We also tested this method *in vitro* using CA1 pyramidal neurons from the hippocampus. In neurons the optimal stimulus performed better than the other suboptimal stimuli with the same power for every cell recorded. This is reassuring because the optimal stimulus is only guaran-

teed to be optimal for heterogeneity of neurons within an ensemble, but not necessarily for heterogeneity of dynamics within an individual neuron over time. To reconcile the differences between intracellular vs intercellular heterogeneity, we postulate that the intrinsic properties of the CA1 hippocampal cells, and hence their PRCs, might be slowly changing throughout each trial. Another recent study [100] suggested that the natural frequencies of periodically firing neurons can drift over time. A methodology that explicitly accounts for uncertainty in the PRCs across time can help ensure entrainment over the entire duration of the experiment.

In this study, we have performed optimization just for energy minimization but note that strategy could be modified to account for other important constraints by adding terms to the cost function (3.22). For example, as we have done in other work, a cost function could be modified to require charge balanced stimuli [23] or to limit harmful Faradaic reactions [11]. These and other considerations could be handled on an application specific basis.

A major benefit of this optimization is that it only requires knowledge the phase response properties of an oscillatory system, which can be measured experimentally. The full dynamics of the CA1 hippocampal neuron results from a complex interaction of ionic currents flowing across a cell membrane [88]. While it may one day be possible to estimate the full dynamics of neurons *in vitro* to design an better controller, for example with a Kalman filter [101], [102], phase reduction provides a convenient and experimentally obtainable means of understanding an oscillatory system, even when underlying system dynamics are inaccessible. Furthermore, because the methodology developed in this chapter explicitly accounts for heterogeneity in model parameters, different stimuli could be designed accounting for variability across the a small sampling of PRCs.

While this method was tested using electrical stimulation of neurons, the approach could be generalized to many different kinds of stimulation modalities and oscillators.

With an optimal stimulation waveform tailored to the dynamics of the system's response to the stimulus, entrainment of the oscillators may be done with greater reliability and less energy than other stimulus waveforms, such as periodic pulsing or sine wave stimulation.

Chapter 4

Determining Individual Phase Response Curves from Aggregate Population Data

4.1 Introduction

Collective oscillation in populations of limit cycle oscillators is a widely observed phenomenon in nonlinear biological sciences [2], [4], [24], [6]. To understand the collective behavior of these systems, it is often useful to reduce the dynamics of the constituent oscillators through phase reduction [88], [6], [54], so that each oscillator obeys an equation of the form

$$\dot{\theta} = \omega + Z(\theta)u(t), \tag{4.1}$$

where the phase $\theta \in [0, 2\pi)$ describes an oscillator's position in the basin of attraction of a limit cycle, ω is the natural frequency so that the natural period $T = 2\pi/\omega$, and $Z(\theta)$ is the infinitesimal phase response curve (PRC) which captures the oscillator's response

to a small perturbation, $u(t)$. We note that (4.1) can be appended to include additional terms such as noise and coupling in a population. Phase reduction has been applied fruitfully to many applications to both understand and control populations of phase oscillators [103], [21], [88], [37], [104], [105].

Essential to the understanding of these oscillatory group dynamics is the ability to accurately compute PRCs, which for systems *in silico* has been rendered nearly trivial with modern computing algorithms and software [65], [106], [107]. For living systems, however, the model equations are not usually known, and calculating PRCs is more difficult. For example, accurately measuring PRCs in neurons using the "direct method" [24], [108], [25] requires current to be injected through a dynamic clamp, piercing the cell membrane and ultimately killing the neuron in the process. While recordings from individual neurons can be difficult to measure, readings from populations of neurons are readily available in experimental neurology, for example, in the form of the local field potential, which represents a filtered sum of current traveling across the cell membranes of a population of nearby cells [109].

When it is difficult or impractical to obtain data from an individual oscillator for calculating its PRC, it may be more convenient to study the macroscopic behavior of the population. To this end, [110–113] investigate the relationship between the phase sensitivity of a individual limit cycle oscillators and the phase sensitivity of their collective oscillation which arises due to coupling. Also, [12] derived a method to calculate phase response curves for the collective oscillations in excitable systems. These methods, however, require that the population oscillation approaches a limit cycle, which can be a relatively strict assumption.

In this chapter, we propose a methodology that can calculate individual PRCs using only an aggregate signal produced by the collective oscillation of a population of homogeneous oscillators which does not require the collective oscillation itself to approach a limit

cycle. While this methodology is developed for a homogeneous uncoupled population, we find that it is robust to both heterogeneity in the individual oscillators, uncertainty in the signal being measured, and terms which are unaccounted for such as noise and coupling. Furthermore, we find that the standard methods used to calculate PRCs in individual oscillators can produce misleading results when directly applied to populations of oscillators. This methodology could make control strategies such as [42], [87], and [10] more feasible for *in vivo* testing when the individual elements in the population are difficult to observe. This chapter is based on work originally appearing in [36].

4.2 Numerical Procedure

To begin, consider a large group of N identical, uncoupled phase oscillators [2], [24]:

$$\dot{\theta}_i = \omega + Z(\theta_i)\psi\delta(t-\tau) + \mathcal{O}(\epsilon). \tag{4.2}$$

Here, $\theta_i \in [0, 2\pi)$ is the phase of oscillator $i = 1, 2, \ldots, N$, $Z(\theta)$ is the PRC, and $\psi\delta(t-\tau)$ is a δ-function impulse of strength $\psi \in \mathbb{R}$ applied identically to each oscillator. In (4.2), we allow for unknown but small $\mathcal{O}(\epsilon)$ perturbations. Suppose we have no information about any individual oscillators, but that each cell emits a phase dependent signal $s(\theta)$ so that the aggregate signal

$$\overline{s}(t) = \frac{1}{N}\sum_{i=1}^{N} s(\theta_i(t)) \tag{4.3}$$

can be measured from the distribution.

When N is very large, we can characterize the distribution of phase oscillators with

a probability density $\rho(t,\theta)$ [20]:

$$\begin{aligned}\frac{\partial \rho}{\partial t} &= -\frac{\partial}{\partial \theta}[(\omega + Z(\theta)\psi\delta(t-\tau) + \mathcal{O}(\epsilon))\rho(t,\theta)] \\ &= -\omega\rho_\theta - \psi[Z(\theta)\rho_\theta + Z_\theta(\theta)\rho]\delta(t-\tau) + \mathcal{O}(\epsilon). \end{aligned} \tag{4.4}$$

Equation (4.4) implicitly assumes that the first derivative of ρ is $\mathcal{O}(1)$ and the $\mathcal{O}(\epsilon)$ terms are small. In the absence of δ-function forcing and $\mathcal{O}(\epsilon)$ terms, equation (4.4) admits periodic traveling wave solutions. This knowledge can be exploited to calculate the probability density immediately prior to and after a δ-function pulse, which will be necessary for calculating phase response curves.

4.2.1 Estimating the Population Distribution from the Aggregate Signal

Consider the one dimensional Fokker-Planck equation (4.4). For all times $t \neq \tau$, we asymptotically expand the solution of (4.4) in orders of ϵ as $\rho(t,\theta) = \rho^{(0)}(t,\theta) + \epsilon\rho^{(1)}(t,\theta) + \epsilon^2\rho^{(2)}(t,\theta)\ldots$, and find that, $\frac{\partial}{\partial t}\rho^{(0)}(t,\theta) = -\omega\rho_\theta$ so that

$$\rho(t,\theta) = \rho_o(\theta - \omega t) + \epsilon\rho^{(1)}(t,\theta) + \mathcal{O}(\epsilon^2), \tag{4.5}$$

where $\rho_o(\theta) = \rho(0,\theta)$. For the moment, we will neglect the $\mathcal{O}(\epsilon)$ and $\mathcal{O}(\epsilon^2)$ terms so that

$$\bar{s}(t) = \frac{1}{2\pi}\int_0^{2\pi} \rho_o(\theta - \omega t)s(\theta)d\theta. \tag{4.6}$$

If we take $\Delta\theta$ small enough, errors in the following approximation of (4.6) are negligible:

$$\bar{s}(\Delta t m) = \frac{1}{M}\sum_{j=1}^{M} \rho_o(\Delta\theta j - \omega m\Delta t)s(\Delta\theta j). \tag{4.7}$$

Assuming that we have carefully chosen $\Delta\theta$ so that the sampling rate $\Delta t = \Delta\theta/\omega$, letting $\bar{s}(\Delta t m) \to \bar{s}_m$, $\rho_o(\Delta\theta m) \to \rho_{o_m}$, and $s(\Delta\theta m) \to s_m$, we may write

$$\bar{s}_m = \frac{1}{M}\sum_{j=1}^{M} \rho_{o_{j-m}} s_j, \tag{4.8}$$

where $\bar{s}$, s, and ρ_o now represent vectors in $\mathbb{R}^M$ and $M = 2\pi/\Delta\theta = T/\Delta t$. Defining $\bar{s}^c_x = \bar{s}_{M-x}$ and $s^c_x = s_{M-x}$, and letting $m = M-k$ and $j = g+M-k$, we can manipulate the previous equation to give

$$\begin{aligned}
\bar{s}^c_{M-m} &= \frac{1}{M}\sum_{j=1}^{M} \rho_{o_{j-m}} s^c_{M-j}, \\
\bar{s}^c_k &= \frac{1}{M}\sum_{j=1}^{M} \rho_{o_{j-M+k}} s^c_{M-j}, \\
\bar{s}^c_k &= \frac{1}{M}\sum_{g=1}^{M} \rho_{o_g} s^c_{k-g}, \\
\implies \bar{s}^c &= \frac{1}{M}(\rho_o * s^c)
\end{aligned} \tag{4.9}$$

where $*$ is the convolution operator. Note here that we are using periodicity in s so that, for example, if $k-g \leq 0$, $s^c_{k-g} = s^c_{M+k-g}$. Thus, letting $\mathcal{F}$ represent the discrete Fourier transform, ρ_o can be found using the relation

$$\frac{\rho_o}{M} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(\bar{s}^c)}{\mathcal{F}(s^c)}\right). \tag{4.10}$$

If we account for the leading order ϵ terms in equation (4.5) we can rewrite (4.8) as

$$\begin{aligned}\bar{s}_m &= \frac{1}{M}\sum_{j=1}^{M}(\rho_{o_{j-m}} + \epsilon\rho^{(1)}_{j,m})s_j \\ &= \frac{1}{M}\sum_{j=1}^{M}\rho_{o_{j-m}}s_j + \frac{1}{M}\sum_{j=1}^{M}\epsilon\rho^{(1)}_{j,m}s_j,\end{aligned} \tag{4.11}$$

where $\rho^{(1)}_{j,m} \in \mathbb{R}^M$ is a vector representing the discretized distribution $\rho^{(1)}(m\Delta t, j\Delta\theta)$. With the same manipulations we used in equation (4.9), we arrive at the relation

$$\bar{s}^c_k = \frac{1}{M}\sum_{g=1}^{M}\rho_{o_g}s^c_{k-g} + \frac{\epsilon}{M}\sum_{j=1}^{M}(\rho^{(1)}_{j,M-k}s^c_{M-j}), \tag{4.12}$$

and therefore,

$$\frac{\mathcal{F}(\rho_o)}{M} = \frac{\mathcal{F}(\bar{s}^c)}{\mathcal{F}(s^c)} - \frac{\epsilon\mathcal{F}(W)}{M\mathcal{F}(s^c)}, \tag{4.13}$$

where $W_k = \frac{1}{M}\sum_{j=1}^{M}(\rho^{(1)}_{j,M-k}s^c_{M-j})$. Thus, the term $\frac{\epsilon\mathcal{F}(W)}{M\mathcal{F}(s^c)}$ adds order ϵ error to each of the Fourier coefficients. Generally, $\mathcal{F}(s^c)$ will be close to zero for contributions from higher order Fourier coefficients, which can amplify error in calculating higher order modes of the distribution ρ_o. For this reason, we truncate the calculated distribution in the main text to include only the lowest order Fourier modes.

4.2.2 Phase Response Functionals of the Advection Equation

Consider the advection equation, which might describe the probability density $\rho(t,\theta)$ of a large group of identical phase oscillators, each with phase $\theta \in [0, 2\pi)$, on a one dimensional ring, c.f [20]:

$$\frac{\partial\rho}{\partial t} = -\omega\frac{\partial\rho}{\partial\theta} + P(t,\theta). \tag{4.14}$$

Here, ω is the natural frequency of each oscillator so that the natural period $T = 2\pi/\omega$ and $P(t, \theta)$ is a time and phase dependent perturbation. When $P \equiv 0$, equation (4.14) has a T-periodic traveling wave solution

$$\rho(t, \theta) = \rho_o(\theta - \omega t), \tag{4.15}$$

where $\rho_o = \rho(0, \theta)$. In the analysis to follow, we will define this periodic solution as γ.

It will be useful to define a group phase, Θ, such that when $P \equiv 0$, $d\Theta/dt = \omega$. To this end, we define $\Theta \in [0, 2\pi)$ as

$$\begin{aligned} \Theta(\rho(t, \theta)) &= \text{arctan2}(a, b), \\ a(t) &= \frac{1}{\pi} \int_0^{2\pi} \rho(t, \theta) \sin\theta d\theta, \\ b(t) &= \frac{1}{\pi} \int_0^{2\pi} \rho(t, \theta) \cos\theta d\theta, \end{aligned} \tag{4.16}$$

where arctan2 is the signed arctangent function, so that the first Fourier mode of the distribution $\rho(t, \theta)$ is given by $\sqrt{a^2 + b^2}\cos(\theta - \Theta)$. Notice that this definition of the group phase allows us to not only define phase in relation to the traveling wave solution, γ, but also for any perturbed solution of (4.14).

Changing to group phase coordinates using the chain rule, we find

$$\begin{aligned} \frac{d\Theta}{dt} &= \left\langle \nabla\Theta(\rho), -\omega\frac{\partial\rho}{\partial\theta} + P(t, \theta) \right\rangle \\ &= \omega + \langle \nabla\Theta(\rho), P(t, \theta) \rangle. \end{aligned} \tag{4.17}$$

Here, $\nabla\Theta$ is the group phase response functional (GPRF) which represents the gradient of the group phase field and $\langle \cdot, \cdot \rangle$ is the L^2 inner product. Note that equivalence in (4.17) comes from the fact that $d\Theta/dt = \omega$ when $P \equiv 0$. In order to use (4.17) we need an

explicit expression for the GPRF. Following a similar derivation of an adjoint equation for the calculation of GPRFs for limit cycle oscillators [12], evaluating the vector field at $\rho^\gamma(\Theta)$ which we define as the intersection of the Θ level set and the trajectory γ, we have

$$\frac{d\Theta}{dt} = \omega + \langle \nabla\Theta(\rho^\gamma(\Theta)), P(t,\theta)\rangle. \tag{4.18}$$

To proceed, we assume that $P \equiv 0$ for $t > 0$ and give a small perturbation $\Delta\rho$ at time $t = 0$ to the trajectory $\rho(t,\theta) \in \gamma$. Letting $\rho_\epsilon(t,\theta) = \rho^\gamma(t,\theta) + \Delta\rho(t,\theta)$ be the perturbed initial condition, we have

$$\begin{aligned}\frac{\partial\Delta\rho(t,\theta)}{\partial t} &= J(\rho(t,\theta)) \cdot \Delta\rho(t,\theta) + \mathcal{O}(||\Delta\rho(t,\theta)||^2) \\ &= -\omega\frac{\partial}{\partial\theta} \cdot \Delta\rho(t,\theta) + \mathcal{O}(||\Delta\rho(t,\theta)||^2),\end{aligned} \tag{4.19}$$

where $J \equiv -\omega\frac{\partial}{\partial\theta}$. We also define the phase shift associated with the perturbation $\Delta\rho(t,\theta)$ as $\Delta\Theta = \Theta(\rho_\epsilon(t,\theta)) - \Theta(\rho(t,\theta))$ and write

$$\Delta\Theta = \langle \nabla_{\rho(t,\theta)}\Theta, \Delta\rho(t,\theta)\rangle + \mathcal{O}(||\Delta\rho(t,\theta)||^2), \tag{4.20}$$

where $\nabla_{\rho(t,\theta)}\Theta$ is the gradient of Θ evaluated at $\rho(t,\theta)$. After the initial perturbation at $t = 0$, $\Delta\Theta$ is independent of time, and taking time derivatives of (4.20) and neglecting $\mathcal{O}(||\Delta\rho(t,\theta)||^2)$ terms gives

$$\begin{aligned}\left\langle \frac{\partial\nabla_{\rho(t,\theta)}\Theta}{\partial t}, \Delta\rho(t,\theta)\right\rangle &= -\left\langle \nabla_{\rho(t,\theta)}\Theta, \frac{\partial\Delta\rho(t,\theta)}{\partial t}\right\rangle \\ &= -\left\langle \nabla_{\rho(t,\theta)}\Theta, -\omega\frac{\partial}{\partial\theta} \cdot \Delta\rho(t,\theta)\right\rangle \\ &= -\left\langle \omega\frac{\partial}{\partial\theta} \cdot \nabla_{\rho(t,\theta)}\Theta, \Delta\rho(t,\theta)\right\rangle.\end{aligned} \tag{4.21}$$

Equivalence in the last line comes from the fact that $\omega\frac{\partial}{\partial\theta}$ is the adjoint of $-\omega\frac{\partial}{\partial\theta}$ on the periodic domain. Equation (6.9) holds for arbitrary perturbation $\Delta\rho(t,\theta)$ and therefore gives the relation

$$\frac{\partial\nabla_{\rho(t,\theta)}\Theta}{\partial t} = -\omega\frac{\partial}{\partial\theta}\cdot\nabla_{\rho(t,\theta)}\Theta. \tag{4.22}$$

Equations of the form (4.22) are sometimes referred to as "adjoint equations" for calculating phase response functionals (or phase response curves) [12], [54]. The GPRF for this system will be a T-periodic solution to (4.22) which also satisfies $\langle\nabla_{\rho(t,\theta)}\Theta, -\omega\partial\rho/\partial\theta\rangle = \omega$, as was required by (4.17). Furthermore, because the phase Θ was defined to be a function of the first Fourier mode of the distribution ρ, any perturbations to higher modes will not effect the group phase. Therefore,

$$\langle\nabla_{\rho(t,\theta)}\Theta, p\sin(n\theta) + q\cos(n\theta)\rangle = 0, \quad \forall\, p,q\in\mathbb{R}, \quad \forall\, n = 2,3,4,\ldots, \tag{4.23}$$

which means that the GPRF must be of the form

$$\nabla_{\rho(t,\theta)}\Theta = \alpha\sin(\theta-\vartheta), \tag{4.24}$$

with $\alpha\in\mathbb{R}$ and $\vartheta\in[0,2\pi)$. Recalling that when $P\equiv 0$, $d\Theta/dt = \omega$, one can verify that for the group phase defined in (4.16), the GPRF is given by

$$\nabla_{\rho(t,\theta)}\Theta = \frac{\sin(\theta-\Theta)}{\pi\sqrt{a^2+b^2}}. \tag{4.25}$$

4.2.3 Connecting the Fokker-Planck Phase Response Functional to the Individual Neuron Phase Response Curve

Again, neglecting $\mathcal{O}(\epsilon)$ terms, (4.4) admits traveling wave solutions allowing us to study its solution in terms of a group phase Θ, defined here so that $\Theta \in [0, 2\pi)$ and $\kappa\cos(\theta - \Theta)$ is the first Fourier mode of the distribution $\rho(t, \theta)$. With this definition of group phase, we may view the network (4.4) itself as a phase oscillator which evolves according to

$$\dot{\Theta} = \omega + \left\langle G(\theta, \Theta), -\psi[Z(\theta)\rho_\theta + Z_\theta(\theta)\rho]\delta(t - \tau) + \mathcal{O}(\epsilon) \right\rangle, \tag{4.26}$$

where $G(\theta, \Theta)$ is the group phase response functional, and $\langle \cdot, \cdot \rangle$ is the L^2 inner product. As we show in Section 4.2.2 by using techniques similar to those in [12], $G(\theta, \Theta) = \frac{\sin(\theta - \Theta)}{\kappa\pi}$, where κ is the magnitude of the first Fourier mode of $\rho(t, \theta)$. Therefore, the change in group phase $\Delta\Theta$ due to the δ-function impulse is

$$\Delta\Theta = -\frac{\psi}{\kappa\pi}\langle \sin(\theta - \Theta), Z(\theta)\rho_\theta + Z_\theta(\theta)\rho \rangle. \tag{4.27}$$

Because $\rho(t, \theta)$ can be determined by through measurements of $\bar{s}(t)$ using (4.10), and $\Delta\theta$ is defined so that it can be found with knowledge of $\rho(t, \theta)$, equation (4.27) allows us to use these observable values in order to infer the phase response curve of the individual oscillators. It should be noted that if $\rho(t, \theta) = \delta(\theta - \theta_o)$ equation (4.27) reduces to $\Delta\Theta(\theta_o)/\psi = Z(\theta_o)$, i.e. the PRC can be measured precisely with a direct method. However, when ρ is not a δ-function, we will find that simply using the direct method to calculate PRCs can yield misleading and incorrect results. To proceed, we define two

new functions

$$\rho_f(t,\theta) = \rho(t,\theta) - \sum_{k=q+1}^{\infty} B_{\rho,k}\sin(k(\phi_{\rho,k}-\theta)),$$
$$Z_f(\theta) = Z(\theta) - \sum_{k=q+1}^{\infty} B_{Z,k}\sin(k(\phi_{Z,k}-\theta)), \tag{4.28}$$

where $B_{\rho,k}\sin(k(\phi_\rho-\theta))$ and $B_{Z,k}\sin(k(\phi_Z-\theta))$ represent the k^{th} Fourier modes of $\rho(t,\theta)$ and $Z(\theta)$, respectively. If we take q large enough so that $|\rho-\rho_f|$, $|Z-Z_f|$ and their first derivatives with respect to θ are small, it is reasonable to assume that

$$\Delta\Theta = -\frac{\psi}{\kappa\pi}\langle\sin(\theta-\Theta), Z_f\rho_f' + Z_f'\rho_f\rangle + \mathcal{O}(\epsilon), \tag{4.29}$$

where $' \equiv \partial/\partial\theta$. Here we have assumed that using the truncated terms for ρ_f and Z_f in (4.27) only lead to $\mathcal{O}(\epsilon)$ errors in the inner product. The Fourier coefficients of the phase distributions immediately preceding and following a δ-function pulse, $\rho(\tau^-,\theta)$, and $\rho(\tau^+,\theta)$, respectively, can be determined from (4.10). For instance, $\rho(\tau^-,\theta)$ can be found taking $\bar{s}^c = [\bar{s}(\tau-T/M), \bar{s}(\tau-2T/M), \ldots, \bar{s}(\tau-T)]$, with $s^c = [s(2\pi(1-\frac{1}{M})), s(2\pi(1-\frac{2}{M})), \ldots, s(0)]$. We can then use the recordings after the δ-function stimulus to calculate $\rho(\tau^+,t)$ and hence calculate $\Delta\Theta$. Finally, we are in a position to present a strategy to obtain $Z(\theta)$:

1. Record $\bar{s}(t)$ on the interval $t\in[\tau-T,\tau+T]$, where a δ-function pulse is given at $t=\tau$.

2. Repeat p times where $p\geq 2q+1$. Recall that q determines the number of Fourier modes of Z that we wish to estimate.

3. Using the data from steps 1 and 2, construct the matrix $A\in\mathbb{R}^{p\times(2q+1)}$ and vector $b\in\mathbb{R}^p$, as defined below.

4. Solve for the best fit of the Fourier coefficients $c = A^{\dagger}b$, where † denotes the pseudoinverse [114].

Here b_i corresponds to the i^{th} measurement of $\Delta\Theta$. For the k^{th} set of recordings, we use the first q modes of $\rho(\tau^-, \theta)$ to construct $\rho_{f,k}$. Using the notation $Z_f(\theta) = c_1 + \sum_{v=1}^{q} c_{2v}\sin(v\theta) + \sum_{w=1}^{q} c_{2w+1}\cos(w\theta)$, we can rewrite equation (4.29) as:

$$\begin{aligned}
[A_{k,1}, A_{k,2}, \ldots, A_{k,2q+1}][c_1, c_2, \cdots, c_{2q+1}]^T &= b_k + \mathcal{O}(\epsilon) \\
A_{k,1} &= -\frac{\psi}{\kappa\pi}\langle \sin(\theta - \Theta), \rho'_{f,k}\rangle \\
A_{k,2v} &= -\frac{\psi}{\kappa\pi}\langle \sin(\theta - \Theta), \sin(v\theta)\rho'_{f,k} + v\cos(v\theta)\rho_{f,k}\rangle \\
A_{k,2w+1} &= -\frac{\psi}{\kappa\pi}\langle \sin(\theta - \Theta), \cos(w\theta)\rho'_{f,k} - w\sin(w\theta)\rho_{f,k}\rangle. \qquad (4.30)
\end{aligned}$$

In the equation above, we calculate Θ from $\rho(\tau^-, \theta)$. An estimate of the Fourier coefficients of Z can then be determined by taking $A^{\dagger}b$. We note that the $\mathcal{O}(\epsilon)$ terms from (4.30) will not cause the estimate of the true Fourier coefficients of Z to deviate by more than $\mathcal{O}(\epsilon)$ because A does not depend on ϵ, and $||A^{\dagger}||$, while potentially large, will be bounded if we take enough independent measurements so that A has full rank.

4.2.4 Measuring Phase Response Curves in an Almost Advective Equation

Here, we will show that using this methodology to estimate phase response curves, $Z(\theta)$, will yield results that are accurate to leading order ϵ regardless of whether the true dynamics evolve according to

$$\frac{\partial \rho}{\partial t} = -\omega\rho_\theta - [Z(\theta)\rho_\theta + Z_\theta\rho]\psi\delta(t - \tau) \qquad (4.31)$$

or

$$\frac{\partial \rho}{\partial t} = -\omega \rho_\theta - [Z(\theta)\rho_\theta + Z_\theta \rho]\psi\delta(t-\tau) + \mathcal{O}(\epsilon), \tag{4.32}$$

where the $\mathcal{O}(\epsilon)$ terms represent small but unknown perturbations. In the main text, all results are obtained using (4.31), and those results will be used for comparison here.

To begin, in the presence of the $\mathcal{O}(\epsilon)$ terms, suppose that at $t = \tau - T$ the probability distribution is $\rho_o(\theta)$. Using the asymptotic expansion (4.5) from Section 4.2.1 of the main text, at time $t = \tau^-$ the solution of (4.32) will be

$$\rho(\tau^-, \theta) = \rho_o(\theta) + \epsilon \int_{\tau - T}^{\tau} [\rho^{(1)}(t, \theta) + \mathcal{O}(\epsilon)]dt. \tag{4.33}$$

Therefore, $\rho(\tau^-, \theta) = \rho_o(\theta) + \mathcal{O}(\epsilon)$. To calculate $G(\theta, \Theta_1)$, where $\Theta_1 \equiv \Theta(\tau^-)$, we need to know $\Theta(\tau^-)$, which can be found by calculating Fourier coefficients of (4.33):

$$\begin{aligned} a &= \frac{1}{\pi}\int_0^{2\pi} \rho_o(\theta)\sin(\theta)d\theta + \frac{\epsilon}{\pi}\int_0^{2\pi}\left\{\int_{\tau-T}^{\tau}[\rho^{(1)}(t,\theta)+\mathcal{O}(\epsilon)]dt\right\}\sin(\theta)d\theta = a^* + \mathcal{O}(\epsilon), \\ b &= \frac{1}{\pi}\int_0^{2\pi} \rho_o(\theta)\cos(\theta)d\theta + \frac{\epsilon}{\pi}\int_0^{2\pi}\left\{\int_{\tau-T}^{\tau}[\rho^{(1)}(t,\theta)+\mathcal{O}(\epsilon)]dt\right\}\cos(\theta)d\theta = b^* + \mathcal{O}(\epsilon), \end{aligned} \tag{4.34}$$

where a^* and b^* are the Fourier coefficients if we were using (4.31). Therefore, the group phase is $\Theta_1 = \text{arctan2}(a^* + \mathcal{O}(\epsilon), b^* + \mathcal{O}(\epsilon))$ which through Taylor expansion can be shown to be equal to $\text{arctan2}(a^*, b^*) + \mathcal{O}(\epsilon) = \Theta_1^* + \mathcal{O}(\epsilon)$, where Θ_1^* would be the group phase if we were using (4.31). Then from (4.25),

$$G(\theta, \Theta_1) = \frac{\sin(\theta - \Theta_1^* + \mathcal{O}(\epsilon))}{\pi\sqrt{a^{*^2} + b^{*^2} + \mathcal{O}(\epsilon)}} = \frac{\sin(\theta - \Theta_1^*)}{\pi\sqrt{a^{*^2} + b^{*^2}}} + \mathcal{O}(\epsilon). \tag{4.35}$$

Note that equivalence in (4.35) comes from Taylor expansion and assumes that $\sqrt{a^{*^2} + b^{*^2}}$ is large compared to ϵ. Using the phase reduction for (4.32), the effect on the δ-function

pulse on the group phase will be

$$\Delta\Theta = -\frac{\psi}{\pi\sqrt{a^{*2}+b^{*2}}}\Big\langle \sin(\theta-\Theta_1^*)+\mathcal{O}(\epsilon), \frac{\partial}{\partial\theta}\big(\rho_o(\theta)+\epsilon\rho^{(1)}(\tau^-,\theta)+ \mathcal{O}(\epsilon^2)\big)Z(\theta)+\frac{\partial}{\partial\theta}Z(\theta)\big(\rho_o(\theta)+\epsilon\rho^{(1)}(\tau^-,\theta)+\mathcal{O}(\epsilon^2)\big)\Big\rangle$$

$$\implies \Delta\Theta+\mathcal{O}(\epsilon) = -\frac{\psi}{\pi\sqrt{a^{*2}+b^{*2}}}\left\langle \sin(\theta-\Theta_1^*), \rho_o'(\theta)Z(\theta)+Z'(\theta)\rho_o(\theta)\right\rangle, \tag{4.36}$$

where $' \equiv \partial/\partial\theta$. Note that equivalence in the last line of (4.36) requires that the derivatives of the $\mathcal{O}(\epsilon)$ terms are still $\mathcal{O}(\epsilon)$. We observe that the right hand side of (4.36) is of the same form as (4.27). Therefore, including the $\mathcal{O}(\epsilon)$ terms from (4.32) will cause the effect of the δ-function pulse to differ by $\mathcal{O}(\epsilon)$.

In Section 4.2.1, we show that to leading order ϵ, we can measure the phase of a distribution that evolves according to (4.32) by measuring $\bar{s}(t)$ for one period. Therefore, for an initial distribution $\rho(\tau-T,\theta)$, if we apply the methodology to measure the phase response curve of the system,

$$\Theta_1 = \Theta_1^* + \mathcal{O}(\epsilon). \tag{4.37}$$

Using the coefficients from (4.34) the true value of the group phase at τ^-, $\widetilde{\Theta}(\tau^-)$, is equal to $\Theta_1^*+\mathcal{O}(\epsilon)$, and using (4.36) we can say that after the pulse $\widetilde{\Theta}(\tau^+) = \Theta_1^*+\Delta\Theta^*+\mathcal{O}(\epsilon)$, where $\Delta\Theta^*$ is the change in group phase if the distribution evolved according to (4.31). Again, we will be able to measure the group phase of $\rho(\tau^+,\theta)$ to leading order ϵ, so that

$$\Theta_2 = \Theta_1^* + \Delta\Theta^* + \mathcal{O}(\epsilon) \tag{4.38}$$

and

$$\Delta\Theta = \Theta_2 - \Theta_1 = \Delta\Theta^* + \mathcal{O}(\epsilon). \tag{4.39}$$

Therefore, $b = b^* + \mathcal{O}(\epsilon)$, where b is a vector of readings of $\Delta\Theta$, and b^* is what would

have been measured if the distribution evolved according to (4.31).

We now turn our attention to the values in the A matrix given by (4.30). In equation (4.13) from Section 4.2.3, we show that we can calculate the Fourier coefficients of ρ_o to leading order in ϵ. Therefore, the truncated distribution from equation (4.28) of the main text can be written as

$$\begin{aligned}\rho_f(\theta) &= \rho_o(\theta) - \sum_{k=q+1}^{\infty} B_{\rho,k}\sin(k(\phi_\rho - \theta)) + \mathcal{O}(\epsilon) \\ &= \rho_f^*(\theta) + \mathcal{O}(\epsilon),\end{aligned} \tag{4.40}$$

where $B_{\rho,k}\sin(k(\phi_\rho - \theta))$ is the k^{th} Fourier mode of $\rho_o(\theta)$ and $\rho_f^*(\theta)$ is the function $\rho_f(\theta)$ that would have been calculated if the probability distribution evolved according to (4.31). Therefore, the values in A calculated from equation (4.30) become

$$\begin{aligned}A_{k,1} &= -\psi\langle G(\theta,\Theta^*) + \mathcal{O}(\epsilon), (\rho_{f,k}^* + \mathcal{O}(\epsilon))'\rangle = A_{k,1}^* + \mathcal{O}(\epsilon) \\ A_{k,2v} &= -\psi\langle G(\theta,\Theta^*) + \mathcal{O}(\epsilon), \sin(v\theta)(\rho_{f,k}^* + \mathcal{O}(\epsilon))' + v\cos(v\theta)(\rho_{f,k}^* + \mathcal{O}(\epsilon))\rangle \\ &= A_{k,2v}^* + \mathcal{O}(\epsilon) \\ A_{k,2w+1} &= -\psi\langle G(\theta,\Theta^*) + \mathcal{O}(\epsilon), \cos(w\theta)(\rho_{f,k}^* + \mathcal{O}(\epsilon))' - w\sin(w\theta)(\rho_{f,k}^* + \mathcal{O}(\epsilon))\rangle \\ &= A_{k,2w}^* + \mathcal{O}(\epsilon),\end{aligned} \tag{4.41}$$

where $A_{i,j}^*$ is the value of $A_{i,j}$ that would have been calculated if the probability distribution evolved according to (4.31). Therefore, to calculate the Fourier coefficients c of $Z(\theta)$ we must solve

$$(A^* + \epsilon E)c = b^* + \epsilon\beta \tag{4.42}$$

so that $c = (A^* + \epsilon E)^\dagger b^* + \epsilon(A^* + \epsilon E)^\dagger\beta$. Here, ϵE represents the $\mathcal{O}(\epsilon)$ terms from (4.41) and $\epsilon\beta$ represents the $\mathcal{O}(\epsilon)$ terms from (4.39).

Consider the difference between c and $c^* = A^{*\dagger}b^*$, and let $D = (A^* + \epsilon E)$. Then

$$\begin{aligned} c - c^* &= D^\dagger b^* + \epsilon D^\dagger \beta - A^{*\dagger} b^* \\ ||c - c^*|| &= ||(D^\dagger - A^{*\dagger})b^* + \epsilon D^\dagger \beta - \epsilon A^{*\dagger}\beta + \epsilon A^{*\dagger}\beta|| \\ &= ||(D^\dagger - A^{*\dagger})(b^* + \epsilon\beta) + \epsilon A^{*\dagger}\beta|| \\ &\leq ||D^\dagger - A^{*\dagger}|| \cdot ||b^* + \epsilon\beta|| + \epsilon||A^{*\dagger}|| \cdot ||\beta||. \end{aligned} \tag{4.43}$$

Because the difference between the true Fourier coefficients and c^* is $\mathcal{O}(\epsilon)$, if we can show that $||D^\dagger - A^{*\dagger}||$ is an order ϵ term, then the difference between c and the true Fourier coefficients will also be $\mathcal{O}(\epsilon)$.

To proceed, we assume that we have taken enough independent measurements of $\Delta\Theta$ so that A^* has full rank. Then, let $U\Sigma V^T$ be the singular value decomposition of A so that

$$\Sigma = \begin{bmatrix} \Sigma_{11} \\ 0 \end{bmatrix}, \qquad U^T E V = \begin{bmatrix} E_{11} \\ E_{21} \end{bmatrix}, \qquad U^T D V = \begin{bmatrix} D_{11} \\ D_{21} \end{bmatrix} = \begin{bmatrix} \Sigma_{11} + \epsilon E_{11} \\ \epsilon E_{21} \end{bmatrix}. \tag{4.44}$$

Using a combination of Theorem 2.2 and Theorem 3.8 from [115], if we assume ϵE is small enough so that it is an acute perturbation of A^*, as defined in [115], we can write

$$||D^\dagger - A^{*\dagger}|| \leq \frac{\epsilon||\Sigma_{11}^{-1}||_2||A^{*^\dagger}||(||E_{11}|| + ||E_{21}||)}{1 - \epsilon||E_{11}||_2||\Sigma_{11}^{-1}||_2}. \tag{4.45}$$

Therefore, $||D^\dagger - A^{*\dagger}||$ can be bounded as an $\mathcal{O}(\epsilon)$ term, and from (4.43), c is at most $\mathcal{O}(\epsilon)$ away from c^* which is in turn $\mathcal{O}(\epsilon)$ away from the true Fourier coefficients of $Z(\theta)$.

4.3 Numerical Results for a Network of Oscillatory Neurons

We now test the utility of this method on a model which exhibits periodic neural spiking behavior. First consider a model of $N = 1000$ periodically spiking thalamic neurons [63]:

$$\begin{aligned} C\dot{V}_i &= -I_L(V_i) - I_{Na}(V_i, h_i) - I_K(V_i, h_i) \\ &\quad - I_T(V_i, r_i) + I_{i,SM} + \zeta_i u(t) + D\eta_i(t) + \frac{\alpha_c}{N}\sum_{j=1}^{N}(V_j - V_i), \\ \dot{h_i} &= (h_\infty(V_i) - h_i)/\tau_h(V_i), \\ \dot{r_i} &= (r_\infty(V_i) - h_i)/\tau_r(V_i), \quad i = 1, \ldots, N. \end{aligned} \tag{4.46}$$

Here, V_i, h_i, and r_i are transmembrane voltage and gating variables for neuron i, $u(t) = I(t)/C$ represents a control input common to all neurons, ζ_i is a constant representing the proximity to the stimulus electrode, $I_{i,SM}$ represents a baseline current chosen so that each neuron fires periodically, $\eta_i(t) = \mathcal{N}(0, 1)$ is i.i.d. noise with zero mean and variance 1, D and α_c are constants determining the strength of the noise and electrotonic coupling [64], respectively, and all other functions and parameters are given in [63]. We note that this network could be generalized to include, for instance, chemical synaptic coupling with more complicated coupling structures. In our first example, we take the network to be homogeneous with $I_{i,SM} = 5\mu\text{A/cm}^2$ and $\zeta_i = 1$ for all i. We also set $D = \alpha_c = 0$ so that the distribution evolves according to (4.4) with no $\mathcal{O}(\epsilon)$ terms. When $u(t) \equiv 0$, each neuron settles to the same stable limit cycle, and we take $s(\theta)$ to be the transmembrane voltage along this cycle, shown in panel A of Figure 4.1. We take $p = 70$ measurements taking $u(t)$ to approximate a δ-function with $\psi = 0.025$ using

the methodology described above. Panel B shows an example of $\bar{s}(t)$ centered at $t = \tau$, the time at which the perturbation is applied. Panel C shows raw measurements of $\Delta\Theta/\psi$ plotted against the phase that the stimulus was applied, similar to how PRCs are typically measured in single neuron recordings [108], and panel D gives the estimated PRC using the methodology detailed above with $M = 1000$ and $q = 4$ (black line) with the exact PRC (grey line) measured using XPPAUT [65]. Note that while there is a strong, seemingly sinusoidal, correlation between Θ and $\Delta\Theta$, this does not capture the phase response properties of the individual neurons.

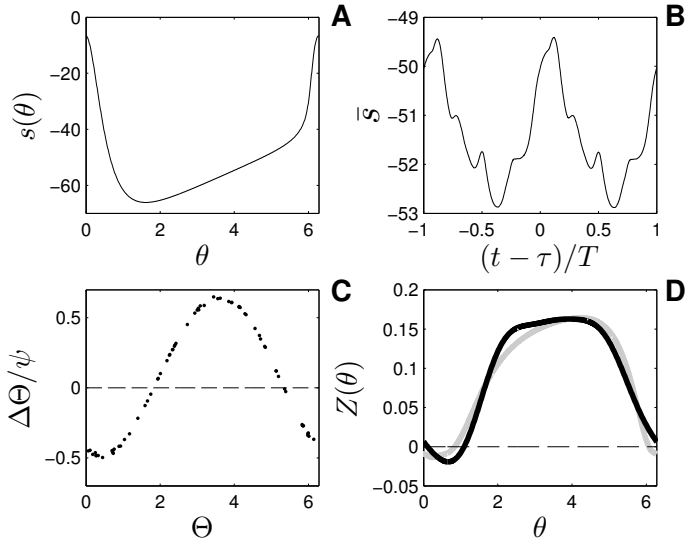

Figure 4.1: Panel A shows the transmembrane voltage as a function of θ on the limit cycle, which we take to be $s(\theta)$. Panel B shows an example measurement of $\bar{s}$ for a perturbation at $t = \tau$. Panel C shows a strong, potentially misleading correlation between the initial phase and $\Delta\Theta$ which does not accurately reflect the true PRC in Panel D. The PRC calculated from the data using the methods of Section 4.2 is shown in black in Panel D with the true PRC in grey.

4.3.1 Including Heterogeneity, Coupling and Noise

In the derivation of the phase response calculation methodology from Section 4.2, we assume that the phase response curves and natural frequencies of each oscillator are

identical. In all but the simplest applications, this is an overly restrictive assumption. Here, we provide numerical evidence that in a heterogeneous network, the proposed methodology can accurately estimate the average phase response curve of the system. Furthermore, while we show in Section 4.2.4 that this methodology is guaranteed to be accurate when each oscillator is subject to small unknown external perturbations (e.g. noise and coupling), we find that this methodology can still yield accurate results when external perturbations are large.

To include heterogeneity in the neural network (4.46), we draw the parameters I_{SM}, ζ and the leak current conductance, (g_L from [63]) in (4.46) from normal distributions, with histograms for the chosen values shown in Figure 4.2. For this choice of parameters, both the PRCs and the natural periods of each neuron in the population are no longer identical. We take $s(\theta)$ to be voltage along the limit cycle, averaged over each neuron which gives a similar $s(\theta)$ to what was used in the homogeneous population example (we note that this calculation of $s(\theta)$ would not be feasible in a real experiment, and provide a discussion about robustness with respect to the choice of $s(\theta)$ in the next section).

In order to apply the numerical procedure from Section 4.2.3 to the heterogeneous population, we need to determine an appropriate value of T, the natural period of oscillation for our system. When the network is homogeneous, T from steps 1 and 2 can simply be taken as the natural frequency of each oscillator. In this case, however, we assume that we do not have *a priori* knowledge of T for the heterogeneous population, and implement steps 1 and 2 by continuously recording $\bar{s}(\theta)$, and intermittently perturb the system with δ-function pulses. We then take T to be the period corresponding to the largest mode of the Fourier transformed data $\bar{s}(\theta)$, taken over the entire duration of the simulation. By recording the time at which the pulses were presented, we can then extract the portions of the measurement $\bar{s}(t)$ necessary to implement the numerical procedure. We illustrate this strategy for both a noiseless and uncoupled ($D = 0$, $\alpha_c = 0$) and a noisy and coupled

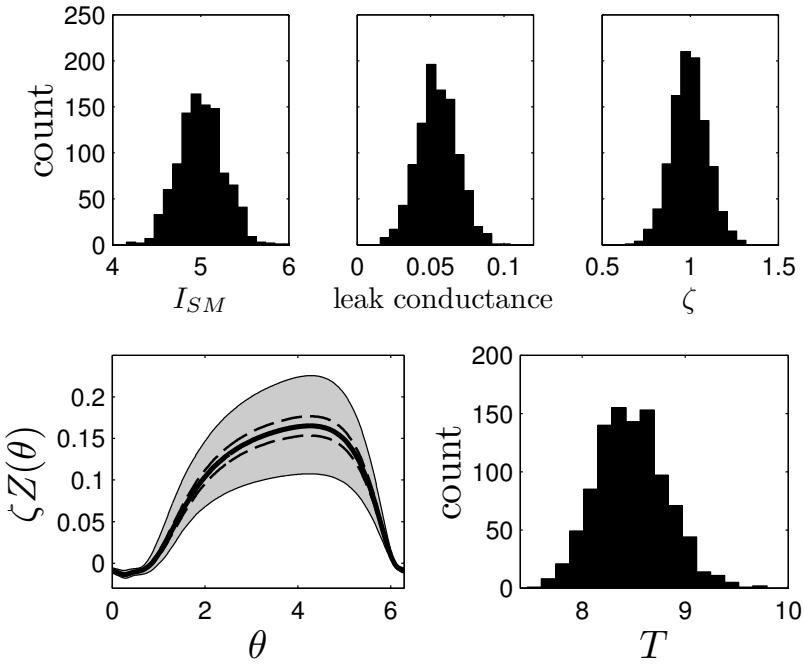

Figure 4.2: The top panels show histograms representing the number of neurons with each value of baseline current (I_{SM}), leak current conductance, and relative stimulus magnitude, ζ, applied to each neuron. The bottom left panel shows the spread of $\zeta_i Z_i(\theta)$ which represents the effective PRC for each neuron. The boundaries of the shaded region represent maximum and minimum resulting values, the dashed lines represent boundaries of the 25^{th} and 75^{th} percentiles, and the thick black line gives the average value for the 1000 neuron population. A histogram representing the natural periods of oscillation for each neuron is shown in the bottom-right panel.

network ($D = 2$, $\alpha_c = 0.1$), with results shown in panels A-C and D-F of Figure 4.3, respectively. In the noiseless, uncoupled simulations, we take $q = 1$ to estimate the first Fourier mode of the individual phase response curves, and in Panel C we see that the result agrees well with the average effective PRC of the population. For the network with both noise and coupling, we find in panel F that the magnitude and phase of the first Fourier mode are slightly worst than the result from the noiseless, uncoupled network, but the estimate is still quite good. We note that in both examples, the raw phase data from panels B and E do not come close to matching the shape or the magnitude of the individual phase response curves.

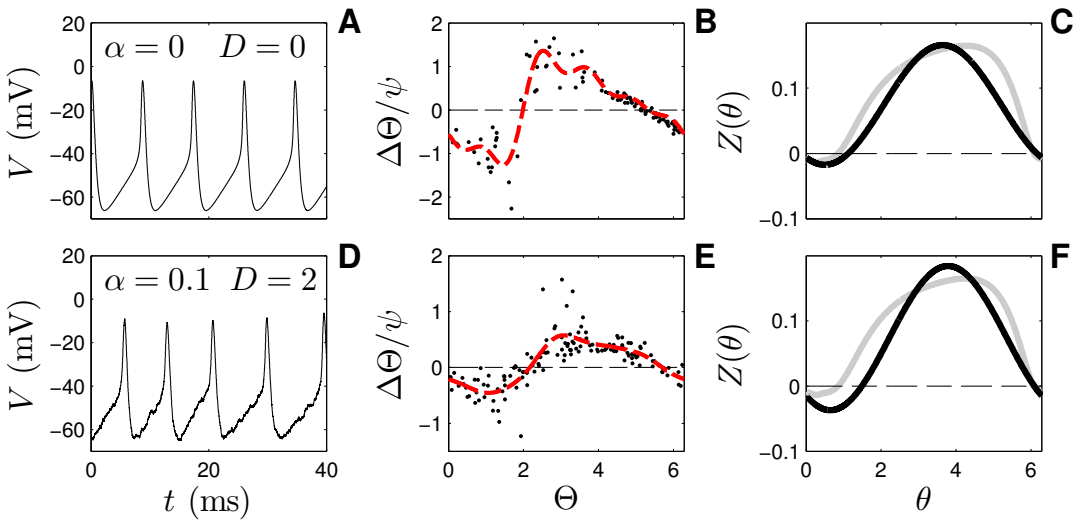

Figure 4.3: Panels A-C show results for a population of heterogeneous, uncoupled, noiseless neurons. Panels D-F show results for a population of heterogeneous, coupled, noisy neurons. Panels A and D show an individual neural trace from each population. Panels B and E show the raw data using $\psi = 1$ as black dots ($p = 105$ and $p = 137$ measurements, respectively) while the red line gives a five mode Fourier fit of the raw data. Panels C and F show the resulting PRCs (black lines) and the averaged effective PRC (grey line) for reference.

4.3.2 Choosing a Signal for Robust Measurement of PRCs

The signal $s(\theta)$ which each oscillator contributes to the population observation is necessary for determining the distribution before and after the application of the pulsatile stimulus and hence, for finding the PRC. For experimental applications, it is likely that $s(\theta)$ cannot be obtained with absolute certainty. In this section, we show that PRCs can still be obtained rather robustly if an approximation to the true value of $s(\theta)$ is known. To begin, consider the infinite time average of (4.3)

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \overline{s}(t)dt = \lim_{T\to\infty} \frac{1}{T} \int_0^T \left[\frac{1}{N} \sum_{i=1}^{N} s(\theta_i(t)) \right] dt. \tag{4.47}$$

We assume that external perturbations are small so that $\theta_i(t)$ is well approximated by $\theta_i(0) + \omega_i t$. We may then manipulate (4.47) as follows:

$$\begin{aligned}
\lim_{T\to\infty} \frac{1}{T}\int_0^T \overline{s}(t)dt &= \lim_{T\to\infty}\left(\frac{1}{T}\int_0^T \left[\frac{1}{N}\sum_{i=1}^{N} s(\theta_i(0)+\omega_i t)\right] dt\right), \\
&= \frac{1}{N}\sum_{i=1}^{N} \lim_{T\to\infty}\left(\frac{1}{T}\int_0^T \left[s(\theta_i(0)+\omega_i t)\right] dt\right), \\
&= \frac{1}{2\pi}\int_0^{2\pi} s(\theta)d\theta\left(\frac{1}{N}\sum_{i=1}^{N} 1\right), \\
&= \frac{1}{2\pi}\int_0^{2\pi} s(\theta)d\theta. \qquad (4.48)
\end{aligned}$$

Therefore, when choosing the signal $s(\theta)$, its mean is well approximated by $\overline{s}(t)$, provided a long enough measurement is taken. Therefore, it is only necessary to estimate the shape and magnitude of the signal $s(t)$, as the mean can be determined from the experimental data. Figure 4.4 replicates the results from Section 4.3 using the homogeneous network and the heterogeneous, noisy, coupled network for two different choices of $s(\theta)$. For the first choice, we take $s(\theta)$ to be the true transmembrane voltage, to which we add a Wiener process. The second choice is a simple piecewise linear approximation to an action potential. For both signals, we vertically shift the resulting $s(\theta)$ so that (4.48) is satisfied. As long as $s(\theta)$ is reasonably close to the true signal, $s^*(\theta)$, the PRC calculation results are not significantly degraded.

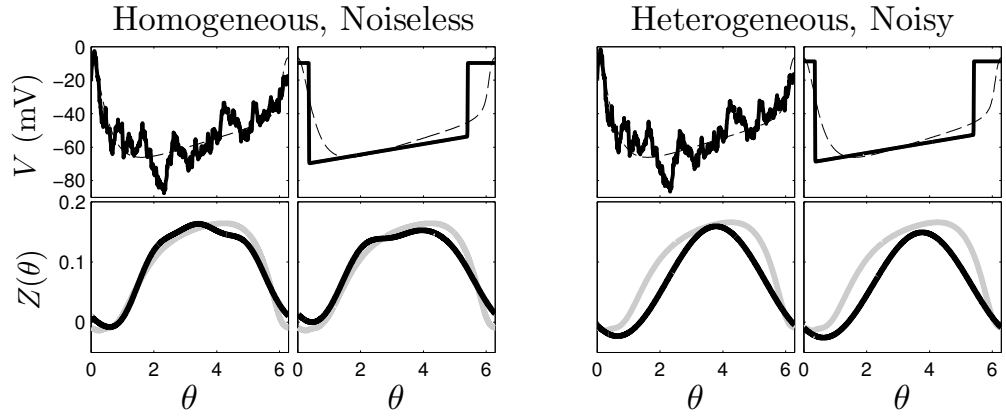

Figure 4.4: Left (resp. Right): Replicating the PRC calculation results using the noiseless, homogeneous, uncoupled, (resp. noisy, heterogeneous, coupled) network from Section 4.3 using different values of $s(\theta)$. Top panels show two different choices of $s(\theta)$ as bold lines with the true (resp. population average) $s^*(\theta)$ shown as a dashed line. Bottom panels show the calculated PRCs in black using the $s(\theta)$ function directly above. Grey lines represent the true (resp. population average) PRC.

4.4 Bursting Neuron PRCs

For a second test, we consider a more complex network of 400 bursting Hindmarsh-Rose model [116] neurons which was modified in [117] to include a synaptic current:

$$\begin{aligned}
\dot{V}_i &= n_i - aV_i^3 + bV_i^2 - h_i + I \\
&\quad + \epsilon_{\text{syn}} \sum_{j=1}^{N} \xi_j (V_i - V_{\text{syn}}) + \epsilon_\eta \eta_i(t) + u(t), \\
\dot{n}_i &= c - dV_i^2 - n_i(t), \\
\dot{h}_i &= r(\sigma(V_i - V_0) - h_i), \\
\dot{\xi}_i &= \alpha_{\text{syn}} T_\infty(V_i)(1 - \xi_i) - \beta_{\text{syn}} \xi_i, \quad i = 1, \ldots, 400.
\end{aligned} \tag{4.49}$$

Here, V_i, n_i, and h_i represent transmembrane voltage and gating variables, ξ_i is a synaptic variable which could represent a neurotransmitter, $\eta_i(t) = \mathcal{N}(0,1)$ is i.i.d. noise with zero

mean and variance 1, and $u(t)$ is an external input. We take $\epsilon_{\text{syn}} = 0.03$, $\epsilon_\eta = 0.01$ and all other parameters identical to those in [117] except for $\beta_{\text{syn}} = 0.0304$ and $\alpha_{\text{syn}} = 1.304$ which were modified so that the synaptic variable $\xi(\theta)$ changed on a slower time scale as compared to the transmembrane voltage $V(\theta)$. For this choice of parameters, in the absence of synaptic coupling and noise, each neuron settles to a limit cycle with a period of T =430 ms consisting of nine-spike bursts followed by a period of quiescence. Here, we use pulses of $u(t) = 0.4$ for 5 ms so that $\psi = 2$. We take $s(\theta)$ to be the synaptic variable, which is shown in Panel A of Figure 4.5, but note that similar results can also be achieved by using the transmembrane voltage. Panel B shows $\bar{s}(t)$ for one of $p = 106$ measurements, centered about $t = \tau$. We note that because of the small noise and coupling terms, $\bar{s}(t)$ is not perfectly T-periodic for $t \neq \tau$. Panel C shows a fit of five Fourier modes to raw data of $\Delta\Theta/\psi$ plotted as a function of Θ. In panel D, the true PRC calculated using the direct method [24] on a single neuron is shown in grey, and the PRC estimated from the methodology above with $M = 1000$ and $q = 5$ is shown in grey. The structure of the true PRC is much more complex than in the previous model, but the estimated PRC accurately captures the slowly varying part. To capture the rapid fluctuations in the earlier part of the cycle, we would need to include more Fourier modes in the calculation, but because of noise and network coupling structure, it is not possible to accurately calculate these modes for this model. In panel E, we show that the calculated PRC (black) is very close to the first five Fourier modes of the true PRC (grey). The fit from panel C (red) is also shown for comparison and is not a good approximation of the true PRC. While we cannot calculate the higher order modes which give rise to the rapidly varying part of the PRC, these are not necessary in many applications. In one such example, we use the calculated PRC to predict entrainment of individual neurons from (4.49) to the external stimulus $u(t) = 0.0025\cos(\omega_o t) + 0.005\sin(\omega_o t)$, where $\omega_o = 2\pi/T$, shown in panel F. If we assume that $u(t)$ is small enough so that $\theta(t) \approx \theta(0) + \omega_o t$, using

standard averaging techniques [118], we can reduce the dynamics of individual cells from (4.49) to

$$\dot{\varphi} = \Gamma(\varphi), \tag{4.50}$$

where $\varphi \equiv \theta/\omega_o - t \pmod{T}$ and $\Gamma(\varphi) = \frac{1}{T}\int_0^T Z(\omega_o t + \varphi)u(t)dt$. Panel G shows $\Gamma(\varphi)$ calculated using the PRC obtained from the methodology described in Section 4.2 as a black line, which predicts three stable fixed points of (4.50) marked by $*$'s. These fixed points are verified in panel H from numerical simulations of individual, noiseless neurons from (4.49). For comparison, using the red curve from panel C to calculate $\Gamma(\varphi)$ predicts only one stable fixed point.

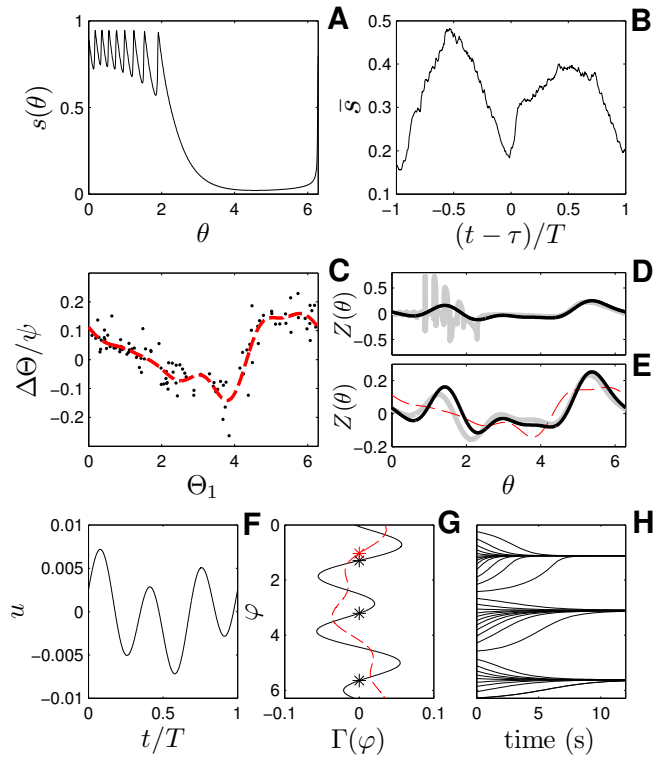

Figure 4.5: Panel A shows the synaptic variable, ξ, as a function of θ, which we take to be $s(\theta)$ in (4.49). Panel B shows an example measurement of $\bar{s}$ for a perturbation at $t = \tau$. Panel C shows raw data of $\Delta\Theta/\psi$ plotted against Θ with a five mode Fourier fit to the data in red. In panel D, The true PRC is shown in grey with the PRC calculated from the data using $q = 5$ in black. Panel E shows the first five Fourier modes of the true PRC, the calculated PRC, and the fit from the raw data in grey, black, and red, respectively. We predict the entrainment of (4.49) to the signal $u(t)$ shown in panel F. In panel G, $\Gamma(\varphi)$ is calculated using the red and black curves from panel E, with resulting functions shown in red and black, respectively. Predicted stable fixed points are denoted with $*$'s for each curve. The stable fixed points predicted from the black curve accurately predict entrainment as verified from simulations of individual cells from (4.49) in panel H.

4.5 Conclusions

This study provides an experimentally feasible methodology for calculating PRCs of individual components from aggregate population data. We have applied this methodology to a model of both periodically spiking and bursting neurons and show that it can accurately calculate the slowly varying modes of the PRCs in the constituent elements. We explicitly show in Section 4.2.4 that in the limit as the truncated order ϵ terms from (4.4) are small, the Fourier coefficients are also accurate to leading order ϵ. Nevertheless, when noise and coupling in the system is relatively large, we can still obtain accurate results. Furthermore, using techniques based solely on the direct method [24] (see also Appendix B), as is typically used to measure PRCs, can yield potentially misleading results in the examples presented here.

In each of the examples given here, we are only able to calculate the first few modes of the individual PRCs before either noise, heterogeneity or truncated order ϵ effects from the phase reduction begin to significantly degrade the calculation. We do not provide analytical limits on the number of modes we can take, but heuristically we find that as we continue to take more and more modes, the solution becomes dominated by the higher order modes (i.e. the methodology produces PRCs that oscillate rapidly). In a setting where the individual phase PRCs are unknown, the procedure could be repeated for increasing values of q (yielding solutions with different numbers of Fourier modes) until the higher order Fourier begin to dominate the solution, indicating that the results are no longer valid.

The proposed methodology is shown to work well for the systems tested here, but modifications could improve the accuracy of the calculation. For example, we have chosen a Fourier basis functions to calculate the probability distribution and PRC in (4.28) because of the intrinsic periodicity of the solutions, and the effect of using different bases

has not been investigated here. Furthermore, for simplicity of implementation, we have used a least-square fitting technique to determine the Fourier coefficients, and have not investigated the effect of using different curve fitting techniques. When we considered heterogeneity, coupling and noise in the periodically spiking population of neurons, we were still able to obtain a reasonably accurate estimate of the average phase response curve of the system. It may be possible to improve this estimate by explicitly accounting for these effects in the underlying partial differential equation (4.4). Most likely this would require specific estimates of the underlying coupling structure and noise strength.

It may be interesting to adapt the proposed methodology for use in excitable systems which through coupling may admit stable periodic oscillations [12]. The constituent elements of these populations are excitable, not periodic, so that perturbations to the individual elements can be understood in terms of isostable response curves [38]. Such systems have relevance to problems in cardiology [29], [28], [119], [34], systems of chemical oscillators [120], waves of spreading depression in the brain [121], [122]. It is possible that isostable response curves could be calculated using a similar strategies for these excitable systems.

Phase reduction has a rich history in the nonlinear sciences, and has led to a greater understanding of many physical, chemical, and biological systems. The methodology presented in this chapter could allow for the use of phase reduction in large systems where it is not feasible to directly observe the individual elements, allowing for their study in a more useful coordinate system. In addition, the algorithm presented here is relatively simple and can be readily implemented with modern mathematical software.

Chapter 5

Extending Phase Reduction to Excitable Systems

5.1 Introduction

Phase reduction methods have fruitfully applied to many physical, chemical and biological systems [1–11]. Such methods are useful for understanding the dynamics of perturbed nonlinear oscillators because they allow for the reduction from a system

$$\dot{\mathbf{x}} = F(\mathbf{x}) + G(\mathbf{x}, t), \quad \mathbf{x} \in \mathbb{R}^N, \tag{5.1}$$

where $\mathbf{x} \in \mathbb{R}^N$ is a vector of states and $G \in \mathbb{R}^N$ is the effect from an external stimulus, to a system with a single variable

$$\dot{\theta} = \omega + Z(\theta)^T G(\mathbf{x}(\theta), t), \quad \theta \in \mathbb{S}^1. \tag{5.2}$$

Here $\theta \in [0, 2\pi)$ describes the oscillator's phase, ω is the oscillator's natural frequency, and $Z(\theta) \in \mathbb{R}^N$ is the oscillator's infinitesimal phase response curve (iPRC) (also called the PRC in previous chapters) which describes the change of the oscillator's phase from a small external stimulus. Phase reduction is particularly powerful with control frameworks such as Hamilton-Jacobi-Bellman, calculus of variations, or dynamic programming [13], since the computational effort for finding the desired control input grows exponentially with the number of state variables. Phase reduction methods have also been useful for systems for which the state dynamics are not fully known. For example, *in vitro* experiments on biological tissue have successfully controlled spike timing in periodically firing neurons using inputs computed from phase models [22], [23]. The calculation of iPRCs is central to the utility of phase reduction. iPRCs can be calculated using a direct method [123] (see also Appendix B), which entails perturbing the system at a given phase and measuring the resulting phase change, or numerically using a so-called "adjoint method" [5], c.f. [54]. Previously, the notions of phase and iPRCs have been extended to excitable systems with a stable fixed point [124], [125], but those methods relied on the existence of slow-fast dynamics, and required response curves to be calculated with a grid-based approach.

In this chapter, we propose a more general method of phase reduction for excitable systems based on the notion of isostables, which as shown in [26], can be defined in relation to level sets of the eigenfunctions of the Koopman operator. Here, a system is said to be excitable if all initial conditions within a small neighborhood of a stable fixed point give trajectories that decay directly to it, but some initial conditions further away from the stable fixed point give trajectories that undergo large excursions before approaching the fixed point. Isostables of excitable systems are analogous to isochrons for asymptotically periodic systems. Isochrons can be understood as an extension of the phase of oscillation to the basin of attraction of the limit cycle [2]; they represent surfaces

of constant phase. In this way, two locations in phase space which start on the same isochron will asymptotically approach the same location on a periodic orbit as time tends toward infinity. For excitable systems, it is not possible define a phase because there is no periodic orbit. Instead, isostables can be defined as sets of points in phase space that approach a fixed point together, in a well-defined sense described below. Figure 5.1 highlights the similarities and differences between isochrons for a two dimensional system with a periodic orbit (left) and isostables for an excitable system with a stable fixed point (right). The left panel shows a system with equations found in [69], [126], and two initial conditions which start on the same isochron asymptotically converge to the same phase on the limit cycle. The right panel shows an excitable system (5.18); initial conditions which start on the same isostable approach the stable fixed point together.

The calculation of an isostable field, $\mathcal{I}(\mathbf{x})$, exploits the linear nature of nonlinear dynamics near a fixed point $\mathbf{x}_o$. For a linear system

$$\dot{\mathbf{x}} = A(\mathbf{x} - \mathbf{x_o}), \tag{5.3}$$

and solutions $\phi(t, \mathbf{x}(0))$ (also known as the flow) approach the fixed point as

$$\phi(t, \mathbf{x}(0)) - \mathbf{x}_o = \sum_{j=1}^{n} s_j(\mathbf{x}(0))\mathbf{v}_j e^{\lambda_j t}, \tag{5.4}$$

where $s_j(\mathbf{x})$ are the coordinates of the vector $\mathbf{x}$ in the basis $\{\mathbf{v}_j,\ j = 1, \ldots, n\}$ of unit eigenvectors of A, with associated eigenvalues $\{\lambda_j,\ j = 1, \ldots, n\}$, sorted so that λ_1 corresponds to a unique slowest direction of the stable manifold, i.e., $\mathrm{Re}(\lambda_j) < \lambda_1 < 0,\ \forall\ j > 1$. Here, we assume that λ_1 is real and unique, and as shown in [26], the magnitude of $s_1(x)$ determines the infinite time approach to the origin. In other words, hyperplanes of constant isostables, $\mathcal{I}_\tau \equiv \{\mathbf{x} \in \mathbb{R}^n | \mathcal{I}(\mathbf{x}) = \tau\}$, *near a fixed point* are

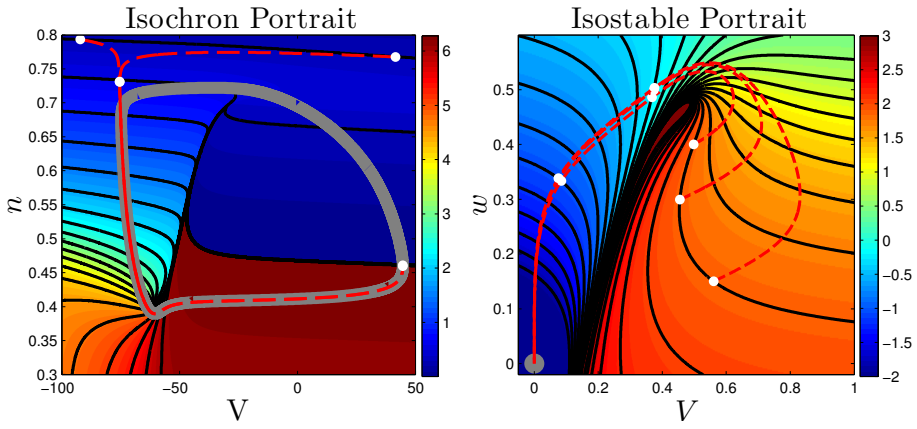

Figure 5.1: Isochrons and isostables in two *different* systems. Isochrons can be used to analyze a system with a periodic orbit, while isostables are appropriate to analyze systems with a stable fixed point. The left panel shows isochrons for a system with a stable periodic orbit shown as the gray curve, with the phase ranging from 0 to 2π. Black lines show isochrons equally spaced in time in the sense that the time required to go from one of the displayed isochrons to the next is always the same. Two trajectories shown as red, dashed lines, start on the same isochron and approach the periodic orbit with the same phase. Snapshots are shown as white dots. The right panel shows an isostable field for a system with a stable fixed point shown as the gray dot, and black lines show isostables which are equally spaced in time. Three trajectories, shown as red, dashed lines start on the same isostable will approach the fixed point together. Snapshots are shown as white dots. Isostables are calculated on a grid of initial conditions using (5.5). Isochrons are calculated using a similar grid-based approach but can also be calculated in finer detail using methods detailed in [126].

parallel to the faster directions $\mathbf{v}_2, \ldots, \mathbf{v}_n$.

For nonlinear systems, the calculation of isostable field within the fixed point's entire basin of attraction, $\mathcal{I}(\mathbf{x})$, can be calculated by monitoring the infinite time approach of $\phi(t, \mathbf{x})$ to the fixed point, $\mathbf{x}_o$, by computing

$$\mathcal{I}(\mathbf{x}) = \lim_{t \to \infty} e^{-\lambda_1 t} ||\phi(t, \mathbf{x}) - \mathbf{x}_o||, \tag{5.5}$$

where $||\cdot||$ can be any norm, but for this chapter, we will be working with the 1-norm on $\mathbb{R}^n$. Intuitively, Equation (5.5) compares the asymptotic approach to the fixed point

along the slowest direction of the stable manifold, $\mathbf{v}_1$, to an exponential function governed by the associated eigenvalue, λ_1. We emphasize that Equation (5.5) is valid for systems with a stable fixed point where λ_1 is real and unique. In other cases, such as when λ_1 is complex or the fixed point is unstable, isostables can still be calculated, and we refer the interested reader to [26] for a more complete discussion.

As mentioned previously, phase reduction is particularly useful to make complicated problems in nonlinear dynamics analytically and computationally tractable. In this chapter, we choose a problem from cardiology for which we will apply the aforementioned isostable reduction, but first we give a brief overview of the heart from a dynamical systems point of view.

The heart beats roughly three billion times during a typical human lifetime to transport blood containing oxygen, nutrients, immune cells, regulatory molecules, and waste products to keep the body functioning. The heart is driven by a small area of specialized cardiac muscle cells known as the sinoatrial node, which acts as pacemaker, producing a propagating wave of electrical activity which spreads through the four chambers of the heart (right atrium, left atrium, right ventricle, and left ventricle) in a coordinated fashion. From a modeling perspective, the propagation of electrical activity throughout the heart is often viewed as a reaction-diffusion equation on a one, two, or three dimensional representation of the *myocardium*, the muscular tissue of the heart. This cardiological medium is made up of individual cells, or *myocytes*, which are specialized muscle tissue which coordinate the action of a heart beat.

Myocytes are excitable cells: a large enough positive perturbation to the transmembrane voltage of a myocyte will elicit a depolarizing action potential and a contraction of the muscle tissue [2]. Cardiac action potentials last for a significant amount of time, allowing the elevated transmembrane voltage to spread to nearby cells, and creating a wave of action potentials. After this wave propagates through the myocardium, under

normal conditions, the cells repolarize and remain quiescent until the pacemaker elicits another pulse.

Ideally, this healthy pattern of coordinated activity occurs throughout the host's lifespan. However, under certain conditions, the myocardium can become susceptible to *cardiac fibrillation*, the uncoordinated contraction of the cardiac muscle, which can be deadly in an otherwise healthy heart if not treated within minutes of onset. Most researchers agree that atrial fibrillation is caused by the presence of unwanted spiral waves [2] within the myocardium which interfere with the normal sinoatrial rhythm [27], [28], [29].

Cardiac fibrillation can be the final step in an increasingly complex series of events that begins with a phenomenon known as cardiac *alternans*. Alternans is the beat to beat alternation of electrochemical cardiac dynamics at a constant rate of pacing and has been implicated as a precursor to fibrillation [30]. These alternations include the strength of cardiac muscle contraction, action potential duration, and intracellular calcium concentration dynamics. In particular, the phenomenon known as *discordant alternans*, where cardiac tissue in different spatial locations exhibit alternans of action potential duration of opposite phase, can produce particularly favorable conditions for initiating spiral wave reentry required for fibrillation [119,127–130]. Spiral waves are more likely to be initiated, for instance, with a premature beat when there is a large degree of spatial heterogeneity in the refractory cells in the heart. For a review of alternans from a dynamical systems perspective, we refer the reader to [131].

In recent decades, researchers have looked into ways of suppressing alternans as a way of preventing cardiac fibrillation, bypassing the need for painful and damaging defibrillating shocks. Many of these methods, such as those described in [31] and [32], attempt to suppress alternans by monitoring the action potential duration (APD), or the length of time that the tissue remains depolarized (i.e., at elevated voltage), and appropriately

modifying the time at which the next action potential is elicited. These methods have been shown to be successful in real cardiac tissue [33–35].

Using isostable reduction techniques we are able to work in a more useful coordinate system to devise a novel and energy-efficient approach for terminating alternans. In Section 5.2 we develop the mathematical framework required for an isostable reduction for excitable systems (analogous to phase reduction for periodic systems), and the corresponding infinitesimal isostable response curve (iIRC) (analogous to the iPRC for periodic systems). We also develop an "adjoint method" to calculate an iIRC, requiring only the knowledge of the trajectory along which the system approaches the fixed point. Section 5.3 illustrates the isostable reduction for a relatively simple, two-dimensional model of excitability. We devise an energy optimal control technique to eliminate cardiac alternans in Section 5.4, and illustrate this technique for the 13 dimensional Fox model of canine cardiac activity [132] in Section 5.5. We find numerically in Section 5.6 that we can eliminate alternans using many orders of magnitude less energy than with a pulsatile strategy, which may be important from a clinical perspective to maximize the battery life of an implantable device and limit tissue damage. We give concluding remarks in Section 5.7. This chapter is based on work which originally appeared in [38].

5.2 Infinitesimal Isostable Response Curves

Following the derivation of the adjoint method for calculating infinitesimal phase response curves (iPRC's) set forth in, for example, [54], we consider a general, N-dimensional differential equation

$$\dot{\mathbf{x}} = F(\mathbf{x}) + G(\mathbf{x}, t); \quad \mathbf{x} \in \mathbb{R}^N, \tag{5.6}$$

where $F(\mathbf{x})$ is the vector field, $G(\mathbf{x}, t)$ is an external stimulus, and $x_1, x_2, \ldots, x_N$ are the state variables. For a given set of initial conditions, suppose that the system follows the known trajectory γ to the stable fixed point $\mathbf{x}_o$.

Our objective is to simplify (5.6) to a one-dimensional equation by defining scalar isostable coordinates, $\psi(\mathbf{x}) \in (-\infty, \infty]$, for all $\mathbf{x}$ in some neighborhood U of $\mathbf{x}_o$ within its basin of attraction. It will be convenient to take $\psi(\mathbf{x}) = -\log(\mathcal{I}(\mathbf{x}))$, where $\mathcal{I}(\mathbf{x})$ is defined in (5.5). Changing to isostable coordinates, from the chain rule we find

$$\frac{d\psi(\mathbf{x})}{dt} = \nabla\psi(\mathbf{x}) \cdot (F(\mathbf{x}) + G(\mathbf{x}, t)) \,. \tag{5.7}$$

In order to simplify equation (5.7), first consider any trajectory $\phi(t, \mathbf{x})$ in the basin of attraction of the fixed point for which the external stimulus G is set to zero. Using (5.5), we have

$$\begin{aligned}
\frac{d\psi(\mathbf{x})}{dt} &= \lim_{\Delta t \to 0} \frac{\psi(\phi(t + \Delta t, \mathbf{x})) - \psi(\phi(t, \mathbf{x}))}{\Delta t} \\
&= \lim_{\Delta t \to 0} \left(\frac{-\log\left[\lim_{t\to\infty}\left(e^{-\lambda_1 t}(\phi(t + \Delta t, \mathbf{x}) - \mathbf{x}_o)\right)\right]}{\Delta t} \right. \\
&\qquad \left. + \frac{\log\left[\lim_{t\to\infty}\left(e^{-\lambda_1 t}(\phi(t, \mathbf{x}) - \mathbf{x}_o)\right)\right]}{\Delta t} \right) \\
&= \lim_{\Delta t \to 0} \frac{-\log\left[\vartheta e^{-\lambda_1 t} e^{\lambda_1 (t+\Delta t)}\right] + \log\left[\vartheta e^{-\lambda_1 t} e^{\lambda_1 t}\right]}{\Delta t} \\
&= -\lambda_1,
\end{aligned} \tag{5.8}$$

where ϑ is a positive constant determined by the direction of $\mathbf{v}_1$. In the second to last line, we use the fact that the trajectory will eventually approach the fixed point along slowest direction of the stable manifold, and the last line is obtained solely through algebraic manipulation. Recall that Equation (5.8) was derived for $G = \mathbf{0}$, which, using (5.7),

implies that

$$\nabla\psi(\mathbf{x}) \cdot F(\mathbf{x}) = \omega, \tag{5.9}$$

and thus

$$\frac{d\psi(\mathbf{x})}{dt} = \omega + \nabla\psi(\mathbf{x}) \cdot G(\mathbf{x}, t), \tag{5.10}$$

where $\omega = -\lambda_1$ (recall that λ_1 is negative), and is similar to the natural frequency term in the classic phase reduction (5.2). We note that the isostable field can be scaled by a constant if desired, which will yield a different value of ω. In the absence of external stimuli, $\frac{d\psi}{dt} = \omega$, i.e., $\psi(\mathbf{x})$ increases at a constant rate. By definition, $\psi(\mathbf{x}) = \infty$ corresponds to $\mathbf{x} = \mathbf{x}_o$, meaning that in the absence of external control, all trajectories in the domain of attraction of $\mathbf{x}_o$ approach the fixed point in infinite time.

Evaluating the vector field at $\mathbf{x}^\gamma(\psi)$, which we define as the intersection of the trajectory γ and the $\psi(\mathbf{x})$ level set (i.e. isostable), we have

$$\frac{d\psi(\mathbf{x})}{dt} = \omega + \nabla\psi(\mathbf{x}^\gamma(\psi)) \cdot G(\mathbf{x}^\gamma(\psi, t)), \tag{5.11}$$

where we have dropped an error term of order $(|G|^2)$, c.f. [7], so that (5.11) is valid for perturbations with small $|G|$. For phase reductions of systems with limit cycles, the analog of the term $\nabla\psi(\mathbf{x}^\gamma(\psi))$ can be calculated experimentally using the "direct method" [123] as well as the numerically using the "adjoint method" [5], c.f. [54]. Here we detail an extension of the adjoint method for calculation of $\nabla\psi(\mathbf{x}^\gamma(\psi))$ for systems with a stable fixed point.

Suppose for the following analysis that $G = \mathbf{0}$ for $t > 0$. Consider an infinitesimal perturbation $\Delta\mathbf{x}$ to the trajectory $\mathbf{x}(t) \in \gamma$ at time $t = 0$. Let $\mathbf{x}_\epsilon(t) = \mathbf{x}^\gamma(t) + \Delta\mathbf{x}(t)$ be

the trajectory starting at the perturbed initial condition. Then

$$\frac{d\Delta\mathbf{x}(t)}{dt} = DF(\mathbf{x}(t))\Delta\mathbf{x}(t) + \mathcal{O}(||\Delta\mathbf{x}||^2), \tag{5.12}$$

where DF is the Jacobian matrix . Furthermore, for the associated isostable shift defined as $\Delta\psi = \psi(\mathbf{x}_\epsilon(t)) - \psi(\mathbf{x}(t))$, we have

$$\Delta\psi = (\nabla_{\mathbf{x}(t)}\psi)^T \cdot \Delta\mathbf{x}(t) + \mathcal{O}(||\Delta\mathbf{x}||^2), \tag{5.13}$$

where $\nabla_{\mathbf{x}(t)}\psi$ is the gradient of ψ evaluated at $\mathbf{x}(t)$. Let $\langle\cdot,\cdot\rangle$ denote the standard Euclidean inner product on $\mathbb{R}^n$. Following the derivation presented in [54], after the initial perturbation at $t = 0$, $\Delta\psi$ is independent of time. Therefore, taking the time derivative of (5.13) yields, to lowest order in $||\Delta\mathbf{x}||$,

$$\begin{aligned}\left\langle \frac{d\nabla_{\mathbf{x}(t)}\psi}{dt}, \Delta\mathbf{x}(t) \right\rangle &= -\left\langle \nabla_{\mathbf{x}(t)}\psi, \frac{d\Delta\mathbf{x}(t)}{dt} \right\rangle, \\ &= -\langle \nabla_{\mathbf{x}(t)}\psi, DF(\mathbf{x}(t))\Delta\mathbf{x}(t)\rangle, \\ &= -\langle DF^T(\mathbf{x}(t))\nabla_{\mathbf{x}(t)}\psi, \Delta\mathbf{x}(t)\rangle. \end{aligned} \tag{5.14}$$

The matrix $DF^T(\mathbf{x}(t))$ is the transpose, or adjoint, of the real-valued matrix $DF(\mathbf{x}(t))$. Equation (A.9) holds for arbitrary perturbations $\Delta\mathbf{x}(t)$, which gives

$$\frac{d\nabla_{\mathbf{x}(t)}\psi}{dt} = -DF^T(\mathbf{x}(t))\nabla_{\mathbf{x}(t)}\psi. \tag{5.15}$$

In order to use (5.11) we need the isostable gradient at the intersection of γ and the $\psi(\mathbf{x})$ level set, but (5.15) gives the isostable gradient along γ. However, one can readily be inferred from the other with knowledge of the map $\mathbf{x}(t) \to \psi(t)$.

The solution of (5.15) requires N initial conditions. Recalling that $G = \mathbf{0}$ for $t > 0$ in

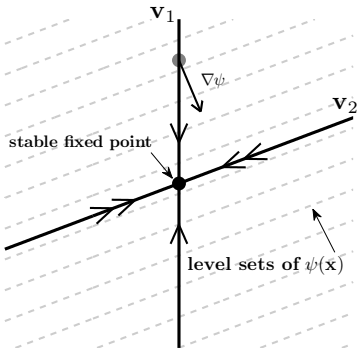

Figure 5.2: Level sets of $\psi(\mathbf{x})$ for a two dimensional system are shown as dashed lines. The faster and slower directions of the stable manifold are with two and one arrows, respectively.

the present analysis, we note that close to the fixed point, we can rewrite equation (5.6) as

$$\dot{\mathbf{x}} = A(\mathbf{x} - \mathbf{x}_o) + \mathcal{O}((\mathbf{x} - \mathbf{x}_o)^2), \tag{5.16}$$

where $A = DF(\mathbf{x}_o)$. Neglecting higher order terms, equation (5.16) is linear. We assume that each eigenvalue $\lambda_i = \sigma_i + i\kappa_i$ of A has negative real part, geometric multiplicity of one, and corresponds to the right eigenvector $\mathbf{v}_i$. For convenience, we sort the eigenvalues so λ_1 has the smallest magnitude of its real part, i.e. $|\sigma_1| < |\sigma_2| \leq \cdots \leq |\sigma_N|$. Note that $|\sigma_1|$ must be strictly less than $|\sigma_2|$ and that $\frac{dx(t)}{dt} \cdot \mathbf{v}_1 \neq 0$ near the fixed point for the following analysis. From linear systems theory, we know that trajectories $\phi(t, \mathbf{x}(0))$ will approach the fixed point in a manner given previously by equation (5.4), with $s_j(\mathbf{x}(0)) = \langle \mathbf{x}(0), \mathbf{v}_j \rangle$. The infinite time approach to the fixed point will be governed by the value of $s_1(\mathbf{x}(0))$, and any perturbation to $s_j(\mathbf{x}(0))$ for $j \geq 2$ will not have any effect on the trajectory's infinite time approach to the origin. Thus,

$$\left\langle \nabla_{\mathbf{x}(t)}\psi, \mathbf{v}_j \right\rangle = 0 \qquad \forall\, j \geq 2. \tag{5.17}$$

Equation (5.17) mandates that close to the fixed point, $\nabla_{\mathbf{x}(t)}\psi$ must be orthogonal to the subspace spanned by $\{\mathbf{v}_2, \ldots, \mathbf{v}_n\}$; in other words, $\nabla_{\mathbf{x}(t)}\psi$ is orthogonal to hyperplanes of constant isostables. Because the geometric multiplicity for each eigenvector is one, this defines a unique direction for $\nabla_{\mathbf{x}(t)}\psi$ near the fixed point. The isostable portrait is summarized in Figure 5.2 for a two-dimensional system. The magnitude can be obtained by noting that (5.10) must hold at all points along $\mathbf{x}(t)$. Using (5.17) and (5.10), we can choose any point $\mathbf{x}_1$ along γ, close enough to the fixed point so that the system is described accurately by its linearization, to serve as an appropriate end point condition to equation (5.15) and solve by integrating backwards in time to determine $\nabla\psi$ at all points close to γ. For convenience, we will scale the isostable coordinates $\psi(\mathbf{x})$ so that $\psi(\mathbf{x}_1) = 1$, which means that ω in equation (5.11) is determined by the time it takes to reach $\psi = 1$ from $\psi = 0$, where $\psi = 0$ is chosen to be another location on γ as convenient. Intuitively, $\nabla\psi$ does not exist for $\psi = \infty$ since any perturbation from the fixed point will take an infinitely long time to reach the fixed point again. We can, however, evaluate $\nabla\psi$ arbitrarily close to the fixed point.

We note that the preceding analysis is valid for stable fixed points with strictly real λ_1. If λ_1 is complex, the iIRC can still be calculated by a similar analysis, but instead of $\nabla\psi$ being orthogonal to the slower directions of the stable manifold near the fixed point, as in (5.17), $\nabla\psi$ will be perpendicular to a cylindrical hypersurface of constant isostable. We refer the interested reader to [26] for a discussion of isostables for the case when λ_1 is complex.

5.3 Illustration For a Simple FitzHugh-Nagumo Based Model

We first illustrate the calculation of $\nabla\psi(\mathbf{x}^\gamma(\psi))$, which we will refer to as the infinitesimal isostable response curve (iIRC), for FitzHugh-Nagumo based equations as a prototypical model for excitability [133]:

$$\begin{aligned} \dot{V} &= c_1 V(V-a)(1-V) - c_2 V w, \\ \dot{w} &= b(V - dw). \end{aligned} \tag{5.18}$$

Non-dimensionalized variables V and w represent, for example, voltage and gating variables, respectively, of an excitable cell membrane. We will take $a = 0.13$, $b = 0.013$, $c_1 = 0.26$, $c_2 = 0.1$, and $d = 1$. We can see from Figure 5.3 that trajectories follow a similar path as they approach the stable fixed point $(V, w) = (0, 0)$; [124] and [125] would call this a transient attractor. When the system is linearized around this fixed point, we find $|\lambda_1| = 0.0130$ and $|\lambda_2| = 0.0338$. The top-right panel of Figure 5.3 shows nullclines for this system. The bottom-left panel shows the global isostable portrait for (5.18), and the bottom-right panel gives a zoomed-in isostable portrait near the fixed point.

We choose $(V, w) \approx (0.91, 0.24)$ to correspond to $\psi = 0$, and integrate forward until the final condition is close enough to the origin for the linearization to be a good model. The initial condition is chosen because most trajectories that start around $(V, w) \approx (1, 0.1)$ pass close to here, as shown in the top-left panel of Figure 5.3. For this model, integrating for 1000 time units is sufficient, and we take the resulting trajectory to be γ, and the end point condition to correspond to $\psi = 1$. We use γ to calculate the iIRC for (5.18) using the adjoint method presented in Section 5.2. For this choice of γ, $\omega = \frac{1}{1000}$. Results are shown in Figure 5.4. The top-right panel shows γ, with the bottom-

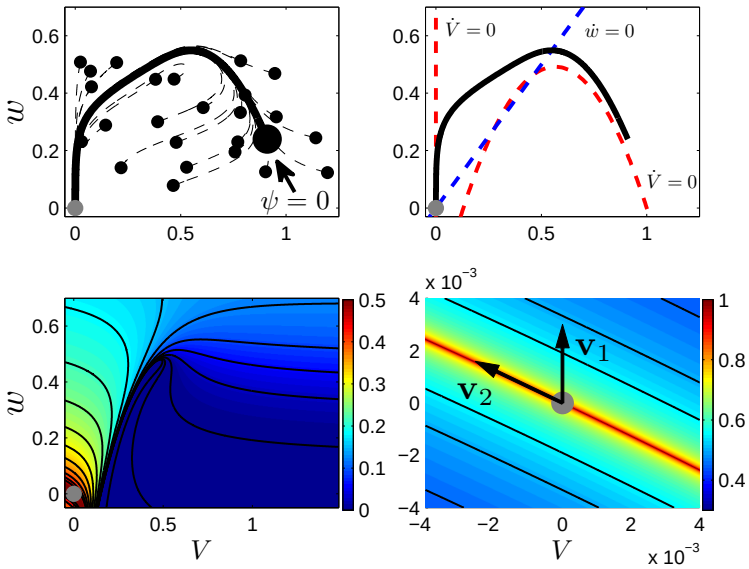

Figure 5.3: In the top-left panel, trajectories of (5.18) follow a similar path toward the fixed point; this path is shown as a bold line. The top-right panel shows the V and w nullclines of (5.18), as dashed lines. The bottom-left panel gives a global isostable portrait for (5.18) with black lines showing constant isostables, equally spaced in time. The bottom-right panel shows a close-up near the fixed point where the system is well-approximated by a linear system, and the isostables are parallel to $\mathbf{v}_2$, the fastest direction of the stable manifold. In all panels, the stable fixed point is shown as a grey dot.

right panels showing $V(\psi)$ and $w(\psi)$. The left panels show iIRCs for each variable at different levels of zoom. As might be expected, as the trajectory approaches the fixed point, the magnitude of the iIRC becomes very large, because even small perturbations will significantly affect the asymptotic approach to the fixed point. Bottom-left panels show zoomed in plots of the iIRC as well as numerical verification at various points, for example, by perturbing by ΔV at a given value of ψ, and measuring the resulting difference $\Delta\psi$. The iIRC can then be approximated by $\frac{\partial \psi}{\partial V} \approx \frac{\Delta \psi}{\Delta V}$.

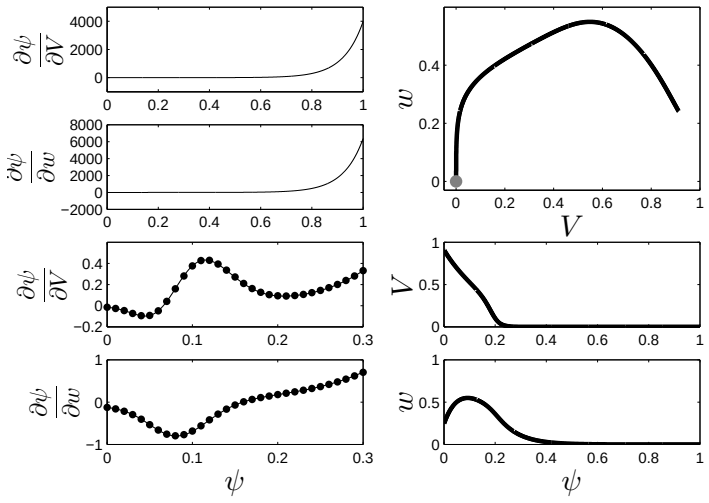

Figure 5.4: Calculation of the iIRC for the FitzHugh-Nagumo based model using the adjoint method. The top-right panel shows γ, with the bottom-right panels showing $V(\psi)$ and $w(\psi)$. iIRCs calculated using the adjoint method are given in the left panels, with dots calculated according to $\frac{\Delta\psi}{\Delta V}$ and $\frac{\Delta\psi}{\Delta w}$ for a small but finite ΔV and Δw.

5.4 An Energy-Optimal Control Strategy For Terminating Cardiac Alternans

As mentioned in Section 5.1, cardiac alternans, the beat-to-beat alternation between long and short action potentials, have been implicated as a precursor to cardiac fibrillation. In order to understand how alternans develop in cardiac tissue, one must examine the relationship between action potential duration (APD), diastolic interval (DI), and basic cycle length (BCL) in cardiac cells. The APD is defined as the amount of time an action potential lasts, i.e. the amount of time the cell remains depolarized (i.e. at elevated voltage) after it is excited, and the DI is the amount of time the cell remains quiescent before the next action potential. The BCL is the amount of time between successive action potentials and can be thought of as a constant "heart rate", for example as generated by pacemaker cells. The APD of a cardiac cell is typically a monotoni-

cally increasing function of DI, and the left panel of Figure 5.5 shows the APD plotted against the DI (commonly referred to as the APD restitution curve) for the Fox model of cardiac activity [132], which we will consider further in Sections 5.5 and 5.6; the solid curve uses nominal parameters, and the dashed curve is obtained when the value of the L-type Ca^{2+} channel permeability to Ca^{2+} is reduced by 20 percent. As originally shown in [134], when the APD restitution curve has a slope greater than 1, alternans can develop in cardiac tissue. To understand how alternans may develop, we first examine the dashed APD restitution curve, which has a slope that is strictly less than one. The plot in the bottom-right panel of Figure 5.5 uses this curve and can be interpreted as follows. Suppose for a constant BCL of 185 ms, an action potential is elicited when the DI is around 42 ms. From the APD restitution curve we can see that the following APD will be around 115 ms. Because $\mathrm{BCL} = \mathrm{APD} + \mathrm{DI}$, the next DI will be about 70 ms. The arrows show the result when this procedure over multiple iterations, and the successive iterations spiral towards an APD of about 125 ms. Conversely, the solid APD curve in the left panel of Figure 5.5 has a slope larger than one in places, and the top-right panel shows that for the same BCL of 185 ms, a stable orbit develops with APDs alternating between 118 and 156 ms; this is called a period-2 orbit because it takes two BCLs to repeat. Note that in the top-right panel, every fifth transition is shown for clarity of the figure; also note that there is an unstable period-1 orbit, meaning that it repeats every BCL, at the intersection of APD restitution curve and the dashed line, represented by an open circle.

As might be expected, the APD restitution curve is not strictly a function of the preceding DI. The APD can be affected by other things such as pacing history or interactions between neighboring cells [135–137]. Nevertheless, the goal of "flattening" the APD restitution curve by reducing its slope provides a basis for developing drug interventions that make it less likely for alternans and cardiac fibrillation to develop [138–140].

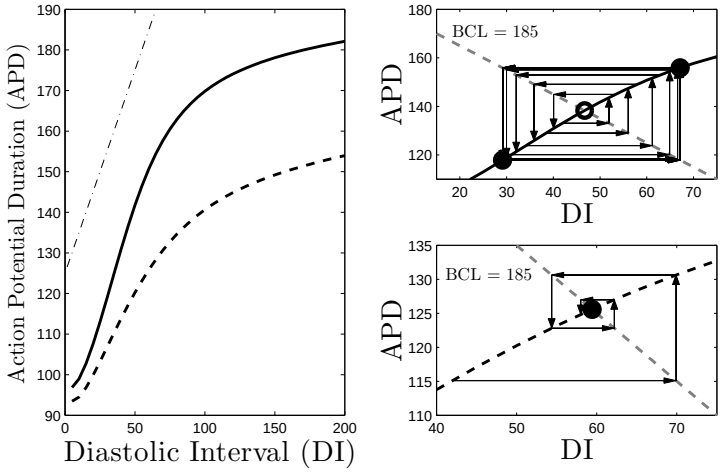

Figure 5.5: The left panel shows APD restitution curves for two different parameter sets for the Fox model and a dot-dashed line with a slope of 1 for reference. The solid APD restitution curve curve has slope greater than one in places, and the associated top-right panel shows how a stable period-2 orbit develops when the BCL is a constant 185 ms (represented by the grey dashed line plotting 185 = DI + APD). For the dashed APD restitution curve, with slope strictly less than one, the period-1 orbit is stable for a constant BCL of 185 ms as shown in the bottom-right panel.

Other strategies to terminate alternans have focused on stabilizing the unstable period-1 rhythm through the application of external stimuli [32], [31], and have been useful in designing experimentally successful, model-independent control techniques [33], [34], [35]. Some other control techniques have attempted to control alternans using model-based, feedback control [141], [142]. While model-independent control techniques are certainly more directly applicable to *in vivo* experiments, they must sacrifice knowledge of the underlying dynamics which makes them, most likely, not optimal from an energy perspective. On the other hand, model-based control strategies which use feedback control make use of the underlying dynamics of the system, but can be hard to implement accurately in real time experiments, especially when the underlying models for the experiment are of high dimension. Here, we propose an energy optimal, model-based control approach that only requires knowledge of a system's iIRC. Control techniques

based on response curve reduction are useful in high order models because they do not require the full dynamics, only the one-dimensional dynamics near a trajectory. Furthermore, infinitesimal phase response curves have been successfully measured for neurons for control purposes [23], and we expect these methods to be readily adaptable to measuring infinitesimal isostable response curves for cardiac systems.

In a common experiment to test control algorithms for alternans, *in vitro* cardiac tissue is excited by an artificial pacemaker at a prescribed BCL which is known to produce action potential duration alternans. The action potential duration of each beat is measured as the length of its action potential. Previous, model-independent methods to suppress alternans have operated by adjusting the pacemaker's cycle length according to the following control algorithm (c.f. [34], [35]):

$$T_n = \begin{cases} \text{BCL} + \Delta T_n & \text{if } \Delta T_n < 0, \\ \text{BCL} & \text{if } \Delta T_n \geq 0, \end{cases} \tag{5.19}$$

where

$$\Delta T_n = (\beta/2)(\text{APD}_n - \text{APD}_{n-1}), \tag{5.20}$$

T_n is the control-perturbed (external) pacemaker cycle length, β is the feedback gain, APD is the action potential duration, and n is the beat number. In this chapter, we will use $\beta = 1$. Typically, as the DI increases, so will the following APD, and intuitively, this control strategy works by giving a premature pulse after the short APD, reducing the DI of the upcoming long APD, and subsequently increasing the DI of the of the next short APD. See panels C and E of Figure 5.11 for a numerical example of the control strategy from equations (5.19) and (5.20) applied to a single cardiac cell.

We devise a new alternans suppression methodology in the context of the 13-dimensional Fox model for canine cardiac activity [132]. We emphasize that the following results are

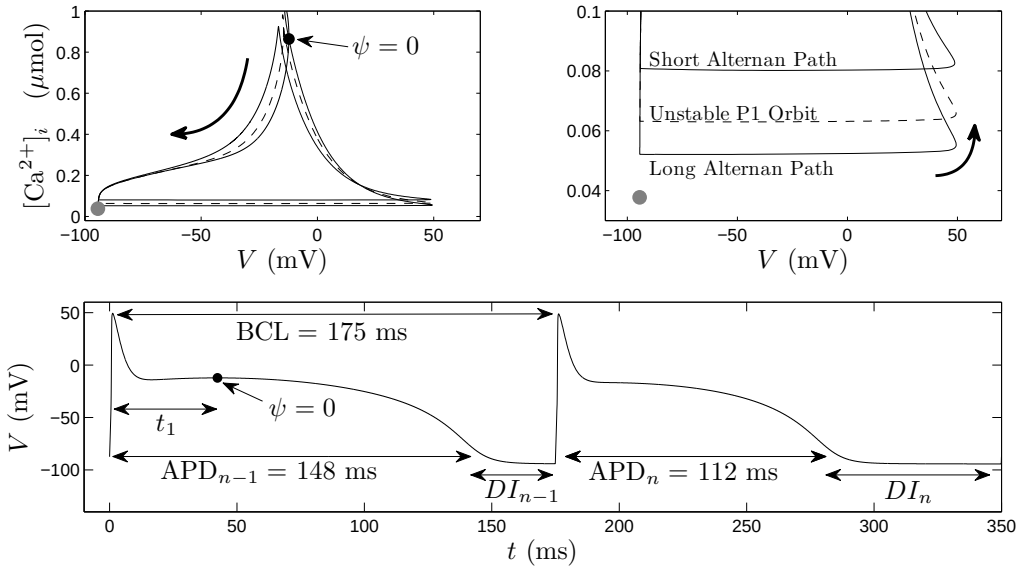

Figure 5.6: Unstable period one orbit (dashed line) and stable period two orbit (solid line) in the $V - [Ca^{2+}]_i$ plane for the Fox model. In the top panels, curved arrows show the direction of the orbits, and grey dots show the location of the stable fixed point.

not limited to this model, and we expect that this control strategy could be applied successfully to other models which display alternans. For a BCL of 175 milliseconds, cardiac cells from the Fox model will display beat-to-beat alternans with successive APD differences of approximately 36 ms. Figure 5.6 shows the associated stable (solid line) and unstable (dashed line) periodic orbits in the $V - [Ca^{2+}]_i$ plane, where V is the transmembrane voltage and $[Ca^{2+}]_i$ is the intracellular Ca^{2+} concentration. We note that the control strategy from (5.19) is sufficient to suppress alternans in this model, but our goal is to devise a control strategy that accomplishes the same objective while using less energy and without artificially modifying T_n. To this end, following [9] we adapt the energy-optimal methodology proposed for spike-timing control in neurons for use in cardiac cells.

Consider an isostable model for a cardiac cell

$$\dot{\psi} = \omega + \nabla\psi \cdot G(t). \tag{5.21}$$

We assume that we only have control over the voltage variable, which we take to be the first of the 13 components of $\mathbf{x}$, so that $G(t) = [u(t), 0, \cdots, 0]^T$, where $u(t) = i(t)/C$ represents the control input, $i(t)$ is the applied current, and C is the cell capacitance. Equation (5.21) becomes

$$\dot{\psi} = \omega + \frac{\partial\psi}{\partial V}(\psi)u(t) = \omega + I(\psi)u(t), \tag{5.22}$$

where $I \equiv \frac{\partial\psi}{\partial V}$. Note that explicit $\mathbf{x}$ dependence has been dropped for notational simplicity. While cardiac alternans are caused by a period-2 rhythm in the presence of periodic forcing, the iIRC must be calculated by removing the forcing and letting the state approach a fixed point. Recall from Section 5.2 that the values $\psi = 0$ and $\psi = 1$ can be chosen as is convenient, to define the iIRC. In a numerical setting, we can take $\psi = 1$ as sufficiently close to the fixed point, but in an experimental setting, this may not be possible. Therefore, we take $I(\psi) = 0 \; \forall \; \psi > 1$.

Now that the iIRC has been found, we return to the case of periodic forcing. We define $t = 0$ to be the time that the present action potential started and t_1 as the time when $\psi(\mathbf{x}(t_1))$ next reaches 0. Suppose that, at a time determined by the BCL (when the next action potential will start), for all stimuli $u(t)$ which evolve $\psi(t_1) = 0$ to $\psi(\text{BCL}) = \alpha > 0$ for some phase α which will be chosen later, we want to find the stimulus $u(t)$ which minimizes the cost function $\mathcal{M}[u(t)]^2 = \int_{t_1}^{\text{BCL}} u^2(t)\, dt$, that is, the power associated with the stimulus. (Other costs including derivative of the current, and Faradaic charge transfer [60] can also be handled by other suitable cost functions.) This

is now in the form of an optimal control problem. We apply calculus of variations to minimize [59]

$$\mathcal{C}[u(t)] = \int_{t_1}^{\text{BCL}} \underbrace{\left[u^2(t) + \lambda_1 \left(\frac{d\psi}{dt} - \omega - I(\psi)u(t)\right)\right]}_{\mathcal{M}[u(t)]} dt, \tag{5.23}$$

where λ_1 is a Lagrange multiplier, forcing the dynamics to satisfy (5.22). Associated Euler-Lagrange equations are

$$u(t) = \lambda_1(t)I(\psi)/2, \tag{5.24}$$

$$\dot{\psi} = \omega + (\lambda_1 I^2(\psi))/2, \tag{5.25}$$

$$\dot{\lambda_1} = -\lambda_1^2 I(\psi)I'(\psi)/2, \tag{5.26}$$

where $' = d/d\psi$. The optimal control $u(t)$ can be found by solving (5.25) and (5.26) subject to $\psi(t_1) = 0$, $\psi(\text{BCL}) = \alpha$. This requires finding the initial condition $\lambda_1(t_1)$ which satisfies the boundary conditions, e.g. numerically with a shooting method. To examine the uniqueness of such an optimal control, we first note that the Hamiltonian for this system,

$$H_o = \lambda_1^2 I^2(\psi)/4 + \lambda_1\omega, \tag{5.27}$$

is conserved along trajectories of (5.25) and (5.26). With this in mind, we find that

$$\begin{aligned}\psi(\text{BCL}) &= \int_{t_1}^{\text{BCL}} \dot{\psi}dt = \int_{t_1}^{\text{BCL}} \frac{\omega + \lambda_1 I^2(\psi(t))}{2} dt = \int_{t_1}^{\text{BCL}} \sqrt{\omega^2 + I^2(\psi)H_o}dt, \\ \frac{\partial\psi(\text{BCL})}{\partial H_o} &= \int_{t_1}^{\text{BCL}} \frac{I^2(\psi(t))}{2\sqrt{\omega^2 + I^2(\psi)H_o}} dt.\end{aligned} \tag{5.28}$$

Note that in order for $\psi(\text{BCL})$ to be real-valued, $H_o > -\omega^2/\sup[I^2(\psi)]$. Therefore, for all valid choices of H_o, $\frac{\partial\psi(\text{BCL})}{\partial H_o} > 0$. Thus $\psi(\text{BCL})$ increases monotonically with H_o and

implies that if a solution to (5.25) and (5.26) exists, it must be unique.

To implement the control strategy, we choose α to be

$$\alpha^* \equiv \omega(\mathrm{BCL} - t_1) - \omega(\mathrm{APD}_n - \mathrm{APD}_{n-1})/2. \tag{5.29}$$

This particular choice for α^* is explained in Figure 5.7, which shows zoomed-in orbits from Figure 5.6. We note that our proposed control strategy is functionally identical to the strategy in equations (5.19) and (5.20), but the means of achieving the goal are different. In (5.19) and (5.20), the time of the next pulse is modulated, which requires a large amount of energy to elicit a premature pulse, while our new control strategy modifies the cell's phase when the next pulse arrives. Assume the next action potential will be a long APD; in the absence of control, the trajectory will reach the location corresponding to the short APD at precisely $t = \mathrm{BCL}$, and a short APD will occur after. If we wanted to drive the trajectory to another long APD, we would need to advance ψ by an extra $-\omega(\mathrm{APD}_n - \mathrm{APD}_{n-1})$ with an external control (recall that before a long action potential, $\mathrm{APD}_n - \mathrm{APD}_{n-1} < 0$). Note that the above control strategy is only valid until the cell's next action potential at $t = \mathrm{BCL}$. At this point, the time can be reset to $t = 0$, and if alternans persist, the control can be reapplied in order to gradually drive the system to the unstable period-1 orbit. We emphasize that it is implicit in the preceding analysis that the state of the system remains close to the trajectory γ from which the iIRC was derived. If the applied control creates a large excursions from γ, it will degrade the performance of the controller, and in extreme cases, can elicit a premature excitation.

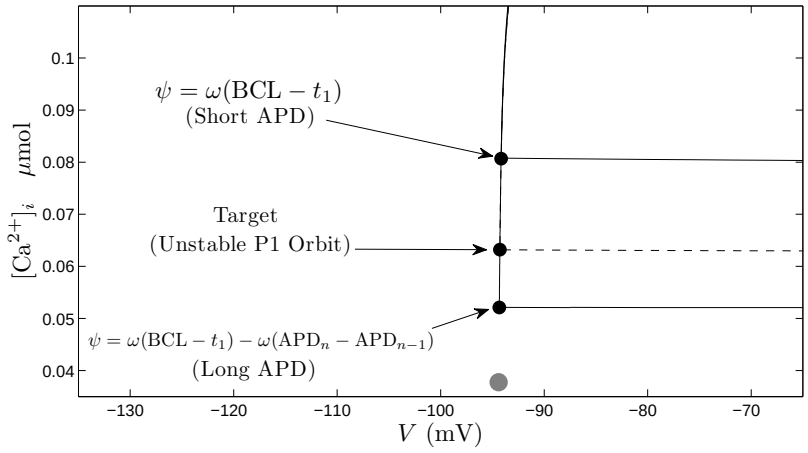

Figure 5.7: Illustration of the alternans termination control strategy showing zoomed in orbits from Figure 5.6. Assuming the next action potential will be a long APD, the trajectory will reach the short APD curve at exactly $t = \mathrm{BCL}$, and the ensuing APD will be short. Advancing the phase by an extra $-\omega(\mathrm{APD}_n - \mathrm{APD}_{n-1})$ will cause the ensuing APD to be long again. The control strategy is only valid until the cell's next action potential at $t = \mathrm{BCL}$. At this point, the time can be reset to $t = 0$, and if alternans persist, the control can be reapplied in order to gradually drive the system to the unstable period-1 orbit. For reference, the stable fixed point is shown as a grey dot.

5.5 Optimal Control Illustrated for the Fox Model of Cardiac Activity

We now illustrate the optimal control strategy for terminating cardiac alternans for the 13 dimensional Fox model [132]. The Fox model is particularly attractive because it considers ionic currents within canine cardiac cells and produces action potentials that are consistent with experimental observations. Furthermore, it produces stable electrical alternans at fast BCLs, making it ideal to test our optimal control strategy. This model has one slowly changing variable, the sarcoplasmic reticulum Ca^{2+} concentration, $[Ca^{2+}]_{SR}$, which does not change much on a beat-to-beat basis for BCLs in which the model exhibits alternans. This allows for the reduction of the dimensionality through

averaging [143], assuming that $[\mathrm{Ca}^{2+}]_{\mathrm{SR}} = 318\,\mu$mol. For reference, when the reduced system is linearized around its stable fixed point, $|\lambda_1| = 0.0015$, and $|\lambda_2| = 0.0043$.

The Fox model exhibits alternans for BCLs between 150 and 210 ms. Using a Poincaré map, shown in Figure 5.8, we visualize the approach γ toward the fixed point for many action potentials. We stimulate the reduced model for BCLs of 150, 160, 170, 180, 190, 200, and 210 and record the state during repolarization when V crosses -40, and -80, as well as when $\frac{dV}{dt} = 0$. Variables $(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12})$ correspond to $(V, [\mathrm{Ca}^{2+}]_i, f, d, m, h, j, f_{Ca}, X_{\mathrm{Kr}}, X_{\mathrm{Ks}}, X_{\mathrm{to}}, Y_{\mathrm{to}})$ from the reduced model. Note that each variable has been normalized as the fraction of its maximum value during stimulation; see [132] for the definitions of these variables and equations for their dynamics.

In order to implement the control strategy from Section 5.4, we restrict our attention to the relaxation to the fixed point. The reason for this is two-fold: first, as shown in Figure 5.7, the state variables follow a similar trajectory towards the fixed point during repolarization so that the iIRCs will be expected to be similar on a beat-to-beat basis. Second, the initial upstroke of the action potential occurs on a rapid time scale, which would make it difficult to apply the control at the precise time it is required. During the last moments of repolarization as the transmembrane potential approaches its resting potential of about -94 mV, the beat-to-beat trajectories tend to follow a similar path, but it is not as clearly defined as it is for the FitzHugh-Nagumo based model (5.18). There is some natural variance in the preferred approach in the Fox model, and for this reason, we calculate the iIRC for multiple action potentials, which in the absence of an external depolarizing stimulus, take different paths toward the fixed point. We define the state at which V is largest during its final approach to the fixed point on the long alternan as $\psi = 0$ (i.e. $\mathbf{x}^\gamma(0) \equiv \mathbf{x} \mid \frac{dV}{dt} = 0$ and $\frac{dV}{dt} < 0 \;\forall\; \psi \geq 0$), and simulate for 2000 ms to obtain $\mathbf{x}^\gamma$ for each approach toward the fixed point. We emphasize that there are multiple possible trajectories toward the stable fixed point, and the curve $\mathbf{x}^\gamma$ between

trials are different. Results are shown in Figure 5.9. The black line give the mean iIRC for

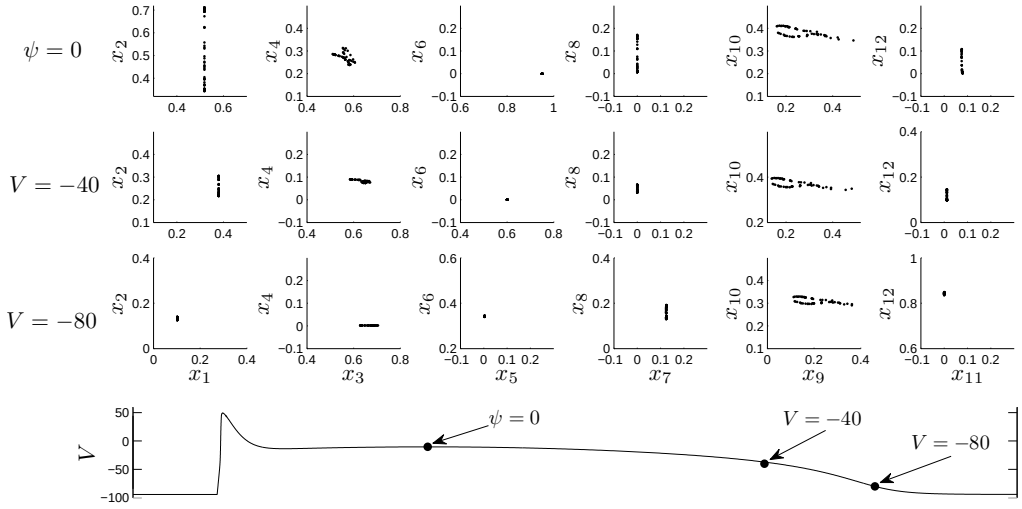

Figure 5.8: State variables in the Fox model at various points during the long alternan action potential, stimulated using multiple BCLs that produce alternans. Trajectories become closer as the transmembrane voltage approaches the resting potential of about -94 mV.

the voltage variable, and the grey band gives an interval in which 90 percent of calculated iIRCs reside. We find that perturbing later in the cycle will have a much smaller effect than perturbing earlier. Note that the iIRCs for the other 11 variables were calculated but are not shown.

Using average values of $\frac{\partial \psi}{\partial V}$ from Figure 5.9, we employ the control strategy on the phase reduction (5.22) and illustrate for a BCL of 175 ms, the cycle length at which alternans are most pronounced. Also, we only apply control on the long alternan because the iIRC from Figure 5.9 was calculated for long alternan trajectories. Results are shown in Figure 5.10. The top panel shows the voltage trace of the cell during a time when the controller is turned on. We see that the control eliminates alternans by the next action potential. We note that once the alternans are eliminated, it takes relatively little control effort compared to the first control application to keep the cell in an alternan-free

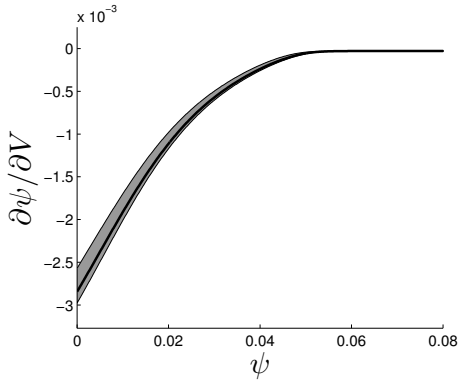

Figure 5.9: iIRC for the transmembrane voltage in the Fox model reduction using the adjoint method. Ninety percent of values lie within the grey band, with the mean value plotted as a thick solid line. The transmembrane voltage is within 95 percent of its resting potential at $\psi \approx 0.025$.

trajectory. The bottom two panels give plots of the beat-to-beat action potentials, as well as $|\mathrm{APD}_n - \mathrm{APD}_{n-1}|$. Times when the controller is turned on are denoted by shaded regions marked with a "C". When the controller is on, the cell remains alternan-free. When the controller is turned off, the cell gradually displays alternans that become more and more pronounced, which are quickly alleviated once the controller is turned back on.

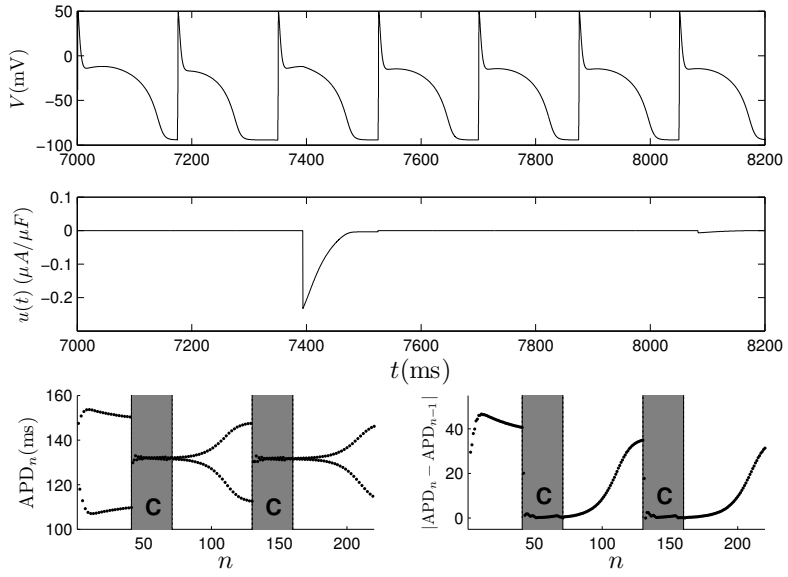

Figure 5.10: Control of alternans in the Fox model. The top and middle panel shows the voltage traces for the cell and the applied control, respectively. The bottom two panels show plots of APD and $|\text{APD}_n - \text{APD}_{n-1}|$, with the shaded sections showing when the controller is on. The control strategy is quickly able to eliminate alternans in the model. Note that the control applied at $t \approx 8100$ is much smaller in magnitude than the first control application. Also, the action potential that begins at $t \approx 7350$ corresponds to APD_{42}.

5.6 Alternan Control with iIRCs Determined by the Direct Method

A marked advantage of this control strategy over other model-based strategies is that it only requires knowledge of the iIRC of the system. In neuroscience, methods that rely on analogous iPRCs are attractive because the iPRC can be approximated experimentally using the direct method [123], [25], [23]. We note that the direct method can be used to measure the iPRC along the entire periodic orbit in neuroscience applications, because a neural spike, which can be measured as a sudden rise in the transmembrane voltage, is a clear marker of a particular phase. In cardiological applications, the primary observable (transmembrane voltage) only changes during repolarization, after which, an-

other observable would be necessary to obtain a full iIRC. Nevertheless, here, we devise a feasible means of using the direct method to obtain at least part of a cardiac cell's iIRC for control purposes.

First, $\psi = 0$ is defined to be some marker voltage during repolarization; we will illustrate this using the Fox model, and as in the above, we take the marker $\psi = 0$ to be the final local maximum during repolarization, i.e. at $\mathbf{x}$ for which $\frac{dV}{dt} = 0$ and $\frac{dV}{dt} < 0 \; \forall \; \psi \geq 0$. Note that the phase resets upon each APD. Further, we define $\psi = 1$ to be when the transmembrane voltage reaches 95 percent of its resting potential. The value ω can be obtained by determining $t_{0\to1}$, the average time it takes a cell to travel from $\psi = 0$ to $\psi = 1$. For the Fox model, we average over BCLs ranging from 150 to 210 to find $t_{0\to1} = 105.06$ ms, and calculate $\omega = 1/t_{0\to1}$. In order to obtain one data point for the iIRC, a short-duration pulse of current can be applied to a cardiac cell at a random isostable ψ, and the resulting isostable change is measured by observing the change in the time it takes to reach $\psi = 1$. The resulting value $I(\psi)$ is $\frac{\Delta\psi}{Q_p/C}$ where Q_p is the total charge injected through the cell membrane, and $\Delta\psi$ is the change in isostable. An experimentally reasonable sampling size of 300 data points were calculated according to $I(\psi) = \frac{\Delta\psi}{Q_p/C} + 0.025\mathcal{N}(0,1)$ where uncertainty is included to mimic the undesired, but unavoidable noise in experimental measurements. The data was fit to a 6^{th} order polynomial, as is customary in calculation of iPRCs for neurons [25], [23], and constrained so that $I(1) = 0$ and $\frac{dI}{d\psi}(0) = 0$. Note that these constraints are consistent with the qualities of the iIRC shown in Figure 5.9. Recall that we assume $I(\psi) = 0$ for $\psi > 1$, and we enforce these extra constraints so that the iIRC is smooth in at least the first derivative at $\psi = 1$.

In an experimental setting, an additional concern is that a control strategy that must calculate a new optimal control on every application would be difficult to implement because by the time the optimal control is calculated, the window to implement it would

have already passed. Indeed, a Matlab routine on a good workstation takes on the order of one second to solve the boundary value problem given by (5.25) and (5.26), and the processing power of an implantable device will certainly be less than that of a computer. For this reason, we seek an approximate solution to the boundary value problem when $\alpha \approx \omega(\text{BCL} - t_1)$. First, notice that the line $\lambda_1 = 0$ is invariant for (5.25)-(5.26), and corresponds to $\frac{d\psi}{dt} = \omega$, thus $\alpha = \omega(\text{BCL} - t_1)$. For $\alpha \approx \omega(\text{BCL} - t_1)$, we Taylor expand $\psi(\text{BCL})$ with respect to initial condition $\lambda_1(0)$, yielding $\alpha = \omega(\text{BCL} - t_1) + \left(\frac{\partial \alpha}{\partial \lambda_1(0)}\Big|_{\lambda_1(0)=0} \right) \lambda_1(0)$ to lowest order in $\alpha - \omega(\text{BCL} - t_1)$. Recalling that $\alpha = \psi(\text{BCL})$ and using α^* from (5.29) for α, we have

$$\lambda_1(0) = \frac{\alpha - \omega(\text{BCL} - t_1)}{\frac{\partial \alpha}{\partial \lambda_1(0)}\Big|_{\lambda_1(0)=0}} = \frac{-\omega(APD_n - APD_{n-1})}{2\,\frac{\partial \psi(\text{BCL})}{\partial \lambda_1(0)}\Big|_{\lambda_1(0)=0}}. \tag{5.30}$$

Let $\text{APD}_n - \text{APD}_{n-1} = \mathcal{O}(\epsilon)$. Asymptotically expanding $\lambda_1(t)$ and $\psi(t)$ yields

$$\lambda_1(t) = \lambda_1^{(0)}(t) + \epsilon\lambda_1^{(1)}(t) + \epsilon^2\lambda_1^{(2)}(t) + \ldots, \tag{5.31}$$

$$\psi(t) = \psi^{(0)}(t) + \epsilon\psi^{(1)}(t) + \epsilon^2\psi^{(2)}(t) + \ldots. \tag{5.32}$$

From equations (5.25) and (5.26), we see that $\lambda_1^{(0)}(t) = 0$, $\lambda_1^{(1)}(t) = \lambda_1(0)/\epsilon + \mathcal{O}(\epsilon)$, and $\psi^{(0)}(t) = \omega(t - t_1)$. In order to further simplify equation (5.30), we use the asymptotic expansion from (5.31) and recall that $\lambda_1(0)$ is of order ϵ to yield,

$$\frac{\partial \psi(\text{BCL})}{\partial \lambda_1(0)}\Big|_{\lambda_1(0)=0} = \lim_{\lambda_1(0)\to 0} \frac{\int_{t_1}^{\text{BCL}} [\lambda_1(0) + \mathcal{O}(\epsilon^2)](I^2(\psi)/2)dt}{\lambda_1(0)} = \int_{t_1}^{\text{BCL}} \frac{I^2(\psi)}{2} dt. \tag{5.33}$$

Using (5.30), this implies that $\lambda_1(0) = \frac{-\omega(APD_n - APD_{n-1})}{\int_{t_1}^{\text{BCL}} I^2(\psi)dt}$. Finally, substituting (5.31) and

(5.32) into (5.24), we find

$$\begin{aligned} u(t) &= \frac{1}{2}\lambda_1(0)I(\psi^{(0)}) + \mathcal{O}((\mathrm{APD}_n - \mathrm{APD}_{n-1})^2) \\ &= \frac{-\omega(\mathrm{APD}_n - \mathrm{APD}_{n-1})I(\omega(t-t_1))}{2\int_{t_1}^{\mathrm{BCL}} I^2(\omega(\tau - t_1))d\tau} + \mathcal{O}((\mathrm{APD}_n - \mathrm{APD}_{n-1})^2). \end{aligned} \tag{5.34}$$

Using the iIRC computed using the direct method on noisy data, we illustrate the use of the *approximation* to the optimal control strategy (5.34), and compare to the method from (5.19). For the pulsed control strategy, as in [132] we use a -80 $\mu A/\mu F$ pulse of duration 1 ms to elicit a premature action potential, and note that for pulse intensities much below this, the pulse will not always elicit an action potential. Further, we take the threshold for the control applications to be $|\mathrm{APD}_n - \mathrm{APD}_{n-1}| \geq 5$. We note that both control strategies work equally well with a smaller threshold, but we choose 5 in order to better illustrate small differences between strategies. We choose a BCL of 175 ms and begin applying control after $n = 30$ action potentials for a total of 150 action potentials. Results are shown in Figure 5.11. Panel A shows each data point in the numerical calculation of the iIRC, as well as the 6^{th} order polynomial fit (black line), and the iIRC calculated from the adjoint method (grey dashed line), scaled by $\omega_{dir}/\omega_{adj}$, where ω_{dir} and ω_{adj} are the natural frequency for the iIRC calculated for the direct method and adjoint method, respectively. Panels B and D show a portion of the applied optimal control and the associated APD history, respectively. In panel B, the first applied control is the largest, with the following control applications being smaller in magnitude, because required alterations to the action potential are much smaller. While the approximately optimal control strategy does not perform quite as well as before, because we are using both a polynomial fit for $\frac{\partial \psi}{\partial V}$ and an approximation of the optimal control, it is still sufficient to terminate alternans. Panels C and E show a portion of the pulsed control and the associated APD history, respectively. The most notable difference

between these control strategies is the energy consumption as measured by $\int u(t)^2 dt$. The energy-optimal control strategy uses approximately 0.7097 units of energy while the pulsed control strategy uses 44800 units. We emphasize that while this is not actual experimental data, and that we are neglecting the power consumption of the processor required to implement our control logic. Nevertheless, our numerical simulations suggest that the optimal control strategy may be able terminate alternans at a cellular level using *several orders of magnitude less energy* than a pulsed method.

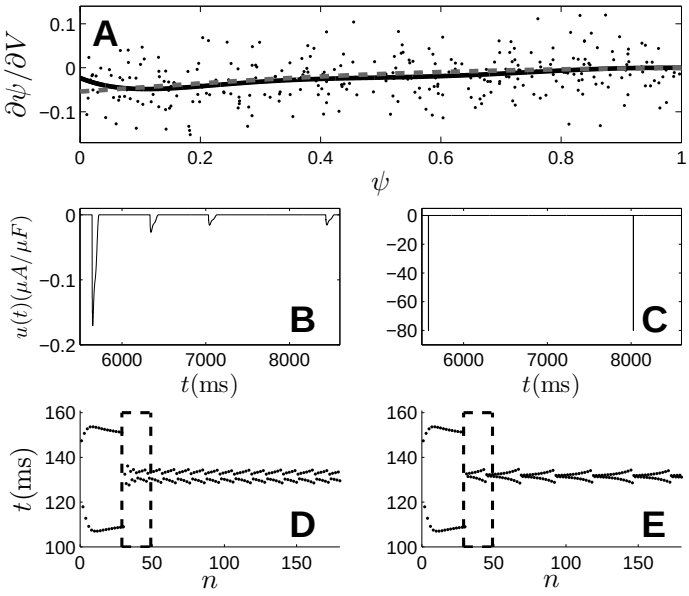

Figure 5.11: Comparison of a pulsed versus approximately optimal control strategy. Panel A shows the polynomial fit iIRC (black line) along with an appropriately scaled iIRC calculated from the adjoint method (grey dashed line). Panel B (resp., C) gives a portion of the applied optimal (resp., pulsatile) control associated with the APD history in the dashed rectangular box from Panel D (resp., E). Panel D (resp., E) gives the APD history using the optimal (resp., pulsatile) control strategy.

5.7 Conclusion

This chapter extends the notion of phase reduction to systems with a stable fixed point through the development of an adjoint method for the efficient calculation of infinitesimal isostable response curves (iIRCs). This method of reduction is particularly useful for excitable systems that take similar trajectories towards their fixed point, so that the iIRCs between trajectories are similar. This novel method of phase reduction has been illustrated for both a simple, two-dimensional model of excitability, and for a more complicated, higher dimensional model. Sections 5.4-5.6 of this chapter display the strength of this new reduction method, as it allows us to work with a complicated, high-dimensional model in a more useful coordinate system.

With only the knowledge of a system's iIRC, one can design a control strategy to eliminate cardiac alternans, which has been implicated as a precursor to fibrillation, that outperforms a pulsatile, model-independent strategy by several orders of magnitude in terms of energy. Even though the processor for a complicated control strategy will likely use more energy than a processor for a simple controller, there is still the potential for a significant energy savings compared to a pulsatile strategy. While model independent control strategies are attractive from an experimental standpoint, forfeiture of the state dynamics almost certainly comes at the price of using a non-optimal control strategy in terms of overall energy consumption, as we have shown here. Furthermore, the proposed control strategy does not require premature excitation of cardiac cells to eliminate alternans, but rather gently perturbs cells during an action potential to eliminate alternans.

Certainly there are other considerations that need to be addressed before this control strategy can be used in real experiments. For instance, it takes a non-trivial amount of time to calculate the optimal control for a given sequence of action potentials, and the processor must be reasonably quick at calculating the control and determining the precise

time at which it should be applied. However, some of these challenges could be addressed by using the approximation to the optimal control derived in Section 5.6. Also, it may be challenging to both stimulate cardiac tissue and record APDs simultaneously. Nevertheless, we believe that this control strategy holds promise for successful implementation in real cardiac systems.

Phase reduction is particularly attractive when the full dynamics of a system are unknown, as is the case for biological tissue in a experimental setting. Control strategies based on phase reduction have been successful in the control of *in vitro* neurons [23], [22], and we expect that similar techniques to experimentally measure iIRCs will carry over to excitable biological systems. We imagine that this control strategy could ultimately be applied to hearts *in vivo* with a grid of independent pairs of sensors and actuators, each using this control strategy to locally eliminate alternans. While not as elegant as a strategy that eliminates alternans from a single point source control, a control strategy that employs reduction using iIRCs to view cardiac action potentials in a more natural coordinate system allows for lower-amplitude, lower-energy perturbations, and thus may be desirable from a clinical standpoint in terms of maximizing the battery life of an implantable device and mitigating the risk of tissue damage.

Phase reduction has proven to be a powerful tool for the analysis of complex systems with periodic orbits and has led to many interesting and important results. This new method of isostable reduction could be of similar use by allowing for the study of complicated excitable systems in a more natural and convenient coordinate system.

Chapter 6

Isostable Reduction with Applications to Time Dependent Partial Differential Equations

6.1 Introduction

Phase reduction provides a convenient means of changing variables in the neighborhood of a stable limit cycle, distilling the essence of a nonlinear system into a phase in relation to the location on the limit cycle, and a phase response curve which describes the effect of an external input on the phase. However, it can only be applied to dynamical systems with periodic orbits, excluding a large number of nonlinear systems with dynamics that approach an equilibrium solution. For such systems, the notion of isostables, which are sets of initial conditions that share the same asymptotic convergence towards a fixed point [26] (similar to the notion of stable foliations of an attracting manifold, which can also be used when the system dynamics do not approach a fixed point [144], [145]), provides an analogous set of coordinates from which the intrinsic dynamics of the system

can be more conveniently understood.

For example Chapter 5, (c.f. [146]) develops the necessary framework for an isostable reduction (analogous to a phase reduction) for finite dimensional ordinary differential equations (ODEs) with dynamics that approach a fixed point. Here, we extend this framework to systems described by time dependent partial differential equations (PDEs) with dynamics that approach stationary solutions by deriving a methodology to calculate isostable response curves, which allows for the understanding of the effect of a perturbation on a system's approach to the stationary solution. Using this reduction strategy allows us to understand a complicated nonlinear PDE in terms of its isostables, reducing the dimensionality of the system from infinity to one and making complicated problems in nonlinear science more computationally and analytically tractable.

Because the isostable reduction proposed in this chapter is valid near a single trajectory that approaches a stationary solution, it is perhaps most similar to PDE reduction methods that involve the use of inertial manifolds [147–150], which represent exponentially attracting, finite dimensional manifolds that can be used to understand an infinite dimensional PDE. However, inertial manifolds can be difficult to identify, and further analysis must be performed in order to understand the response to an external perturbation. Also, as we find in Section 6.4, isostable reduction does not necessarily require the dynamics to collapse to a lower dimensional manifold, as multiple trajectories can display similar reduced dynamics.

This isostable reduction methodology looks to be a promising tool for the understanding and control of PDEs, particularly when the rate and timing of the system's approach to a stationary solution are of interest. We showcase its utility in two different systems with biological applications. The first is in the reduction of a Fokker-Planck equation with the goal of desynchronizing a large population of pathologically synchronized neurons, as might be relevant in the treatment of Parkinson's disease. The second is in the

reduction of a reaction-diffusion system that describes the electric activity of a ring of cardiac cells, with the goal of eliminating and suppressing cardiac alternans which are widely regarded as a possible precursor to cardiac arrest. Some parts of this chapter also appear in [39].

6.2 Isostable Reduction and Numerical Calculation of Infinitesimal Isostable Response Curves in Infinite Dimensions

In this section we detail a method for calculating infinitesimal isostable response curves (iIRCs) in time dependent partial differential equations which asymptotically approach a stationary solution. To begin, consider a weakly perturbed system on the domain Ω

$$\frac{\partial}{\partial t}\mathbf{X}(\mathbf{r},t) = \mathbf{F}(\mathbf{X}(\mathbf{r},t),\mathbf{r}) + \mathbf{G}(\mathbf{X}(\mathbf{r},t)) + \mathbf{p}(\mathbf{r},t). \tag{6.1}$$

Here, $\mathbf{X}(\mathbf{r},t) \in \mathbb{R}^m$ represents the local state in the medium at point $\mathbf{r}$ and time t, $\mathbf{F}(\mathbf{X}(\mathbf{r},t),\mathbf{r})$ describes the local dynamics at point $\mathbf{r}$, $\mathbf{G}(\mathbf{X}(\mathbf{r},t))$ represents spatial coupling (e.g. advection or diffusion) throughout the medium, and $\mathbf{p}(\mathbf{r},t)$ represents a weak spatiotemporal perturbation. For given boundary conditions, we assume that both J and $J^\dagger$ exist, where $J \equiv \nabla[\mathbf{F}(\mathbf{X}(\mathbf{r},t),\mathbf{r}) + \mathbf{G}(\mathbf{X}(\mathbf{r},t))]$ with $\nabla \equiv \partial/\partial\mathbf{X}$ and † denotes the adjoint associated with $\langle\cdot,\cdot\rangle$, the L^2 inner product. For a given initial condition, in the absence of perturbation (i.e. $\mathbf{p}(\mathbf{r},t) = 0$), suppose that the system follows a known trajectory γ towards a stationary solution $\mathbf{X_o}(\mathbf{r})$ which satisfies $\mathbf{F}(\mathbf{X_o}(\mathbf{r}),\mathbf{r}) + \mathbf{G}(\mathbf{X_o}(\mathbf{r})) = 0$ with all eigenvalues having negative real part and bounded away from both the imaginary axis and the continuous part of the spectrum if it exists.

For nonlinear systems near a stationary solution, the equations can be linearized so that eigenfunctions and eigenvalues can be calculated. Suppose ϕ_j are the eigenfunctions of the linearized system, ordered so that ϕ_1 is associated with the eigenvalue λ_1 whose real part has smallest magnitude. In the simplest case λ_1 is real and unique; then near the stationary solution, functions of constant isostable, I_τ, are given by

$$I_\tau = \left\{ \mathbf{X}(\mathbf{r}) \in \mathbb{R}^m | \, \mathbf{X}(\mathbf{r}) = \pm e^{\lambda_1 \tau} \phi_1 + \sum_{j=2}^{\infty} \alpha_j \phi_j, \;\; \forall \, \alpha_j \in \mathbb{R} \right\}. \tag{6.2}$$

Intuitively the magnitude of ϕ_1 determines the isostable, as all other eigenfunctions will decay much faster. As we will show, in the limit of weak forcing isostables provide a convenient means of understanding a system described by a PDE in terms of a single ODE through means of an isostable reduction. Following the adjoint method for calculating infinitesimal phase response curves (iPRCs) as used in [54], (c.f. [38], [12]), our goal is to simplify (6.1) to a one-dimensional equation by defining scalar isostable coordinates, $\psi(\mathbf{X}) \in (-\infty, \infty]$, for all $\mathbf{X}$ in a neighborhood U of $\mathbf{X_o}$ within its basin of attraction. Because $\mathbf{X}$ takes values over a spatial domain, this introduces a phase functional, $\psi = \Psi\{\mathbf{X}\}$ that maps the state $\mathbf{X}$ to a scalar isostable. In general, the isostable field will have a more complicated structure than (6.2), because (6.1) is not guaranteed to be linear.

It will be convenient to define the isostable field $\Psi\{\mathbf{X}\}$ so that

$$\left\langle \nabla \psi(\mathbf{X}), \mathbf{F}(\mathbf{X}(\mathbf{r}), \mathbf{r}) + \mathbf{G}(\mathbf{X}) \right\rangle = \kappa. \tag{6.3}$$

Explicit time dependence has been dropped for convenience of notation. Equation (6.3) mandates that in the absence of external control, the value of the isostable coordinate associated with the state of the system changes at a constant rate κ. Such an isostable field can be obtained along γ, for instance, by integrating the initial condition forward in

time with $\mathbf{p}(\mathbf{r},t)=0$ and choosing $\psi(\mathbf{X}(t))=\kappa t$. By definition $\psi(\mathbf{X})=\infty$ corresponds to $\mathbf{X}=\mathbf{X_o}$, meaning that in the absence of external control, all trajectories in U approach the stable stationary solution in infinite time. We assume that near the trajectory γ, we can take the gradient of the isostable field so that changing to isostable coordinates using the chain rule we find

$$\begin{aligned}\frac{d\psi(\mathbf{X})}{dt} &= \big\langle \nabla\psi(\mathbf{X}), \mathbf{F}(\mathbf{X}(\mathbf{r},t),\mathbf{r}) + \mathbf{G}(\mathbf{X}(t)) + \mathbf{p}(\mathbf{r},t)\big\rangle \\ &= \kappa + \big\langle \nabla\psi(\mathbf{X}), \mathbf{p}(\mathbf{r},t)\big\rangle. \end{aligned} \tag{6.4}$$

Here we have used (6.3) in the last line of (6.4). Evaluating the vector field at $\mathbf{X}^\gamma(\psi)$, which we define as the intersection of the trajectory γ and the $\psi(\mathbf{X})$ level set (i.e. isostable), we have

$$\frac{d\psi(\mathbf{X})}{dt} = \kappa + \big\langle \nabla\psi(\mathbf{X}^\gamma(\psi)), \mathbf{p}(\mathbf{r},t)\big\rangle. \tag{6.5}$$

To proceed, we assume that $\mathbf{p}(\mathbf{r},t)\equiv 0$ for $t>0$. Consider a small perturbation $\Delta\mathbf{X}$ at time $t=0$ to the trajectory $\mathbf{X}(\mathbf{r},t)\in\gamma$. Letting $\mathbf{X}_\epsilon(\mathbf{r},t)=\mathbf{X}^\gamma(\mathbf{r},t)+\Delta\mathbf{X}(\mathbf{r},t)$ be the perturbed initial condition, we have

$$\frac{\partial\Delta\mathbf{X}(\mathbf{r},t)}{\partial t} = J(\mathbf{X}(\mathbf{r},t))\cdot\Delta\mathbf{X}(\mathbf{r},t) + \mathcal{O}(||\Delta\mathbf{X}(\mathbf{r},t)||^2), \tag{6.6}$$

where J is the Jacobian and $(\mathbf{Y}\cdot\mathbf{X})(\mathbf{r},t)\equiv\mathbf{Y}(\mathbf{r},t)\mathbf{X}(\mathbf{r},t)$ for any $\mathbf{X}$ and $\mathbf{Y}$ in Ω. We define the isostable shift associated with the perturbation $\Delta\mathbf{X}(\mathbf{r},t)$ as $\Delta\psi=\psi(\mathbf{X}_\epsilon(\mathbf{r},t))-\psi(\mathbf{X}(\mathbf{r},t))$ and can also write

$$\Delta\psi = \big\langle \nabla_{\mathbf{X}(\mathbf{r},t)}\psi, \Delta\mathbf{X}(\mathbf{r},t)\big\rangle + \mathcal{O}(||\Delta\mathbf{X}(\mathbf{r},t)||^2), \tag{6.7}$$

where $\nabla_{\mathbf{X}(\mathbf{r},t)}\psi$ is the gradient of ψ evaluated at $\mathbf{X}(\mathbf{r},t)$. Both Equations (6.6) and (6.7)

are accurate to leading order in $\Delta\mathbf{X}(\mathbf{r},t)$. After the initial perturbation at $t=0$, $\Delta\psi$ is independent of time. Taking the time derivative of (6.7) yields

$$\begin{aligned}\left\langle \frac{\partial\nabla_{\mathbf{X}(\mathbf{r},t)}\psi}{\partial t}, \Delta\mathbf{X}(\mathbf{r},t)\right\rangle &= -\left\langle \nabla_{\mathbf{X}(\mathbf{r},t)}\psi, \frac{\partial\Delta\mathbf{X}(\mathbf{r},t)}{\partial t}\right\rangle \\ &= -\left\langle \nabla_{\mathbf{X}(\mathbf{r},t)}\psi, J(\mathbf{X}(\mathbf{r},t))\cdot\Delta\mathbf{X}(\mathbf{r},t)\right\rangle \\ &= -\left\langle J(\mathbf{X}(\mathbf{r},t))^{\dagger}\nabla_{\mathbf{X}(\mathbf{r},t)}\psi, \Delta\mathbf{X}(\mathbf{r},t)\right\rangle. \end{aligned} \tag{6.8}$$

Here, $J(\mathbf{X}(\mathbf{r},t))^{\dagger}$ is the adjoint of the real-valued operator $J(\mathbf{X}(\mathbf{r},t))$. Equation (6.8) holds for arbitrary perturbations $\Delta\mathbf{X}(\mathbf{r},t)$, which gives the relation

$$\frac{\partial\nabla_{\mathbf{X}(\mathbf{r},t)}\psi}{\partial t} = -J(\mathbf{X}(\mathbf{r},t))^{\dagger}\nabla_{\mathbf{X}(\mathbf{r},t)}\psi. \tag{6.9}$$

In practice when $J^{\dagger}$ is difficult to derive analytically, and it can be estimated numerically by first discretizing (6.1) and solving for the adjoint of the resulting system of ODEs with appropriate boundary conditions [151]. In this case, the Hermitian transpose of the Jacobian of the discretized ODE system provides an estimate of the adjoint. In order to use (6.9) one must have knowledge of $\nabla_{\mathbf{X}(\mathbf{r},t)}\psi$ (i.e. the iIRC) at some initial time. In general, it is possible to determine the iIRC exactly near the stationary solution where (6.1) can be well approximated by linearization. Full details of this process are given in Section 6.2.2. An initial condition can be taken from this linear regime, and (6.9) can be solved by integrating backwards in time to determine $\nabla\psi$ for the iIRC in the fully nonlinear regime at all points close to γ. The resulting iIRC will vary depending on the specific trajectory γ which must be chosen appropriately for a given application. If the underlying system has fast-slow dynamcs that quickly converges to, for instance, an inertial manifold [147], [148], this manifold could be taken to be γ. However, fast convergence to a known trajectory is not required, a point which is illustrated in Section

6.4, where the iIRC for a cardiac system is similar for many reasonable choices of γ.

6.2.1 Higher Dimensional Reductions

While isostables and the associated iIRC describe the infinite time approach towards the stationary solution, in practice, most engineering applications can only consider finite time. Therefore, for the one dimensional reduction (6.4) to be of practical use, $|\lambda_1| - |\lambda_2|$ multiplied by the time scale of interest must be large enough so that higher order modes becomes negligible. If the time scales are too short for a one dimensional reduction to be effective, one can also use (6.9) to calculate response curves which characterize the infinite time decay of higher order modes in response to the perturbation $\mathbf{p}(\mathbf{r}, t)$. Here, we illustrate a reduction to N dimensions for the simplest case where $|\lambda_1| < |\lambda_2| < \cdots < |\lambda_N|$ and $\lambda_1, \ldots, \lambda_N$ are real and negative, but situations with repeated or complex eigenvalues could be considered.

For the PDE (6.1) on the domain Ω with given boundary conditions, isostables describe the asymptotic convergence of a solution, $\mathbf{X}(\mathbf{r}, t)$ to $\mathbf{X_o}$. We can define an isostable field in terms of a functional Ψ_1 that maps the state $\mathbf{X}(\mathbf{r}, 0)$ to a scalar isostable (c.f. [38], [26])

$$\Psi_1\{\mathbf{X}(\mathbf{r}, 0)\} \equiv \lim_{t\to\infty} e^{-\lambda_1 t} \int_\Omega \big|\mathbf{X}(\mathbf{r}, t) - \mathbf{X_o}\big| d\Omega. \tag{6.10}$$

Here $\int_\Omega \cdot\, d\Omega$ denotes integration over the entire domain Ω. In infinite time, projections of the solution $\mathbf{X}(\mathbf{r}, t)$ onto the eigenfunctions $\phi_2, \phi_3, \ldots$ become infinitesimally small compared to the projection of $\mathbf{X}(\mathbf{r}, t)$ onto ϕ_1, so that we can give an equivalent definition of (6.10)

$$\Psi_1\{\mathbf{X}(\mathbf{r}, 0)\} = \lim_{t\to\infty} e^{-\lambda_1 t} \int_\Omega \big|w_1(\mathbf{X}(\mathbf{r}, t)) - \mathbf{X_o}\big| d\Omega, \tag{6.11}$$

where w_1 gives projection of $\mathbf{X}(\mathbf{r}, t)$ onto ϕ_1. We can characterize the asymptotic behavior

of the projection onto any other eigenfunction ϕ_j, by defining additional functionals

$$\Psi_j\{\mathbf{X}(\mathbf{r},0)\} = \lim_{t\to\infty} e^{-\lambda_j t}\int_\Omega \left|w_j(\mathbf{X}(\mathbf{r},t)) - \mathbf{X_o}\right| d\Omega, \tag{6.12}$$

where w_j gives the projection of $\mathbf{X}(\mathbf{r},t)$ onto ϕ_j. Because $w_j(\mathbf{X}(\mathbf{r},t)) - \mathbf{X_o} \propto \phi_j e^{\lambda_j t}$ as $t\to\infty$, one can verify (c.f. equation (2.3) from [38]) that functionals (6.11) and (6.12) are defined so that $\psi_j \equiv \log(\Psi_j)/\lambda_j$ defines fields for which

$$\left\langle \nabla\psi_j(\mathbf{X}), \mathbf{F}(\mathbf{X}(\mathbf{r}),\mathbf{r}) + \mathbf{G}(\mathbf{X})\right\rangle = 1. \tag{6.13}$$

The analysis from Section 6.2 can be applied to any of these fields so that for a given trajectory γ, an infinitesimal response curve $\nabla_{\mathbf{X}(\mathbf{r},t)}\psi_j$, the gradient of ψ_j evaluated at $\mathbf{X}(\mathbf{r},t)$, can be calculated using (6.9). Near the stationary solution, any perturbation in the direction ϕ_k for $k\neq j$ will not affect the infinite time behavior of the projection of $\mathbf{X}(\mathbf{r},t)$ onto the eigenfunction ϕ_j. Therefore, near the stationary solution $\nabla_{\mathbf{X}(\mathbf{r},t)}\psi_j$ satisfies

$$\langle \nabla_{\mathbf{X}(\mathbf{r},t)}\psi_j, \phi_k\rangle = 0 \qquad \text{for } k = 1,2,\ldots,j-1,j+1,j+2\ldots. \tag{6.14}$$

Equations (6.13) and (6.14) determine an initial condition for each response curve, and (6.9) can be used to calculate N response curves for the N eigenfunctions of interest. Even though eigenfunctions do not carry any meaning for the fully nonlinear equations, these response curves still describe how a perturbation will affect the infinite time approach to the stationary solution when the perturbation is applied in the nonlinear regime. This yields the N-dimensional reduction:

$$\frac{d\psi_j}{dt} = 1 + \left\langle \nabla\psi_j(\mathbf{X}^\gamma(\psi_j)), \mathbf{p}(\mathbf{r},t)\right\rangle, \qquad j = 1,\ldots,N. \tag{6.15}$$

For $N = 1$, (6.15) is equivalent to the one-dimensional isostable reduction considered in Section 6.2. When $N > 1$, the reduction characterizes the decay of the projection of the solution $\mathbf{X}(\mathbf{r}, t)$ onto the eigenfunctions $\phi_1, \ldots, \phi_N$ and ignores the higher order modes. Thus, for an N-dimensional reduction we require $|\lambda_1| - |\lambda_N|$ multiplied by the time scale of interest to be large enough so that the truncated modes become negligible. This is in direct contrast to inertial manifold reductions, which require the gap between consecutive eigenvalues to be large [148], [152].

6.2.2 Isostable Fields Near a Stationary Solution

In order to solve (6.9) in the calculation of iIRCs we must know $\nabla_{\mathbf{X}(\mathbf{r},t)}\psi$ at some initial time. This can be accomplished by linearizing around the stable stationary solution, $\mathbf{X_o}$ where in many cases we can determine the isostable field exactly for the linearized equation. We emphasize that the linearization is only required to determine an initial condition for the solution to (6.9). Upon solving (6.9), the isostable response curve is valid for the reduction of the full nonlinear equations.

Consider a linearization of (6.1) near a stationary solution $\mathbf{X_o}$:

$$\frac{\partial}{\partial t}\mathbf{X} = J(\mathbf{X_o})(\mathbf{X} - \mathbf{X_o}) + \mathcal{O}\left((\mathbf{X} - \mathbf{X_o})^2\right). \tag{6.16}$$

Neglecting higher order terms, we can write its solution as an infinite sum of eigenfunctions:

$$\mathbf{X} - \mathbf{X_o} = \sum_{j=1}^{\infty} s_j(\mathbf{X}(0))\boldsymbol{\phi}_j e^{\lambda_j t}. \tag{6.17}$$

Here, $\boldsymbol{\phi}_j$ corresponds to an eigenfunction of J with the associated eigenvalue λ_j, and $s_j(\mathbf{X})$ gives the coordinates of function $\mathbf{X}$ in the basis of eigenfunctions $\{\boldsymbol{\phi}_j, j = 1, \ldots, \infty\}$. Furthermore, we sort the associated eigenvalues $\{\lambda_j, j = 1, \ldots, \infty\}$ so that $|\mathrm{Re}(\lambda_j)| \leq$

$|\mathrm{Re}(\lambda_{j+1})|$.

In the simplest case, λ_1 is real and unique, then the magnitude of $s_1(\mathbf{X})$ will determine the infinite time approach to the stationary solution (c.f. [26]). Using this perspective, a set of isostables, i.e. $\Psi_\tau \equiv \{\mathbf{X}|\Psi\{\mathbf{X}\} = \tau\}$, can be uniquely determined by finding the magnitude of $s_1(\mathbf{X})$. In other words, sets of constant isostables are spanned by the faster eigenfunctions $\boldsymbol{\phi}_2, \boldsymbol{\phi}_3, \boldsymbol{\phi}_4, \ldots$, and any perturbations to $s_j(\mathbf{X}(0))$ for $j \geq 2$ will not have any effect on the infinite time approach to the origin. Therefore,

$$\langle \nabla_{\mathbf{X}(\mathbf{r},t)}\psi, \phi_j \rangle = 0 \qquad \forall j \geq 2. \tag{6.18}$$

In the case that λ_1 and $\lambda_2 = \overline{\lambda_1}$ are a complex conjugate pair (with corresponding eigenfunctions ϕ_1 and $\phi_2 = \overline{\phi_1}$, equation (6.9) is still valid, but the strategy for finding the initial condition changes slightly. Consider an infinite time approach to the linearized stationary solution, given in (6.17). Assuming that $\mathrm{Re}(\lambda_1) > \mathrm{Re}(\lambda_3)$, as time becomes large, the solution is dominated by the complex eigenfunction pair

$$\mathbf{X} - \mathbf{X_o} = s_1\phi_1 e^{\lambda_1 t} + \overline{s_1}\overline{\phi_1} e^{\overline{\lambda_1} t} + \mathcal{O}(e^{\lambda_3 t}). \tag{6.19}$$

Using Euler's formula and algebraic manipulation, equation (6.19) can be rewritten as

$$\mathbf{X} - \mathbf{X_o} = Ae^{\mathrm{Re}(\lambda_1)t}\left[\cos(\theta_o + \mathrm{Im}(\lambda_1)t)\mathrm{Re}(\phi_1) - \sin(\theta_o + \mathrm{Im}(\lambda_1)t)\mathrm{Im}(\phi_1)\right] + \mathcal{O}(e^{\lambda_3 t}), \tag{6.20}$$

where A and θ_o are constants which can be determined from s_1 and $\overline{s_1}$. Therefore, near

the stationary solution, functions of constant isostable can be defined as follows (c.f. [26]):

$$I_\tau = \Bigg\{ \mathbf{X}(\mathbf{r}) \in \mathbb{R}^m | \, \mathbf{X}(\mathbf{r}) = e^{\mathrm{Re}(\lambda_1)\tau} \left[\cos(\theta)\mathrm{Re}(\phi_1) - \sin(\theta)\mathrm{Im}(\phi_1)\right] + \sum_{j=3}^{\infty} \alpha_j \phi_j, \\ \forall\, \alpha_j \in \mathbb{R}, \quad \forall\, \theta \in [0, 2\pi) \Bigg\}. \tag{6.21}$$

Thus, similar to (6.18),

$$\begin{aligned} \langle \nabla_{\mathbf{X}(\mathbf{r},t)} \psi, \phi_j \rangle = 0, \;\; \forall j \geq 3 \qquad \text{and} \\ \langle \nabla_{\mathbf{X}(\mathbf{r},t)} \psi, w(t) \rangle = 0 \end{aligned} \tag{6.22}$$

where $w(t) = \frac{\partial}{\partial t}\left[\cos(\theta_o + \mathrm{Im}(\lambda_1)t)\mathrm{Re}(\phi_1) - \sin(\theta_o + \mathrm{Im}(\lambda_1)t)\mathrm{Im}(\phi_1)\right]$. The second orthogonality condition in (6.22) comes from the fact that the isostable response curve must be orthogonal to I_τ near the stationary solution. A perturbation in the direction $w(t)$ will simply speed up or slow down oscillations about the stationary solution, but will not affect their magnitude (and hence will not affect the isostable).

When λ_1 is real and unique (resp. part of a complex conjugate pair) equation (6.18) (resp. (6.22)) defines a unique shape of the isostable response curve near the stationary solution, and the magnitude can be obtained from (6.3). Thus, in order to determine an initial condition for (6.9), we can choose a point $\mathbf{X_1}$ along γ, close enough to the stationary solution so that the $\mathcal{O}\left((\mathbf{X} - \mathbf{X_o})^2\right)$ terms in equation (6.16) are negligible. With this initial condition, (6.9) can be solved by integrating backwards in time to determine $\nabla\psi$ for the full nonlinear equations at all points close to γ.

6.3 Application to Parkinson's Disease and Neural Networks

We consider a model of pathological neural synchronization, thought to contribute to the motor symptoms associated with Parkinson's disease [17–19]. For patients whose symptoms are drug resistant, Deep Brain Stimulation (DBS) offers an alternative treatment to mitigate these motor symptoms through the injection of a high-frequency, pulsatile input into an appropriate brain region. It has been hypothesized that the functional mechanism of DBS is the desynchronization of these pathologically synchronized brain regions [40], [20]. In recent years, researchers have developed alternative desynchronization strategies [41], [44], [48] that could make DBS more efficient in order to prolong battery life of the implantable device as well as to mitigate potential side effects of DBS [153]. In previous work [10], we developed a methodology for chaotically desynchronizing a population of synchronized oscillators which only required knowledge of the neurons' phase response curves (PRCs). However, this strategy only guarantees desynchronization when the neurons are close in phase, and as a consequence cannot guarantee that a population of will be driven to a splay, or completely desynchronized, state.

Using isostable reduction techniques derived in Section 6.2, we can design an external stimulus to desynchronize a population of neurons by driving them towards the splay state. To begin, we assume that each individual neuron from the larger population can be represented as a noisy limit cycle oscillator:

$$\dot{\theta}_j = \omega + Z(\theta_j)\left(u(t) + \frac{1}{N}\sum_{i=1}^{N}\sigma(V(\theta_i) - V(\theta_j)) + \sqrt{2S}\eta_j(t)\right), \qquad j = 1, \ldots, N. \quad (6.23)$$

Equation (6.23) is an example of a phase reduction. Here, $\theta_j \in [0, 2\pi)$ is the 2π-periodic phase of the j^{th} neuron with $\theta = 0$ corresponding to the neuron firing an action potential,

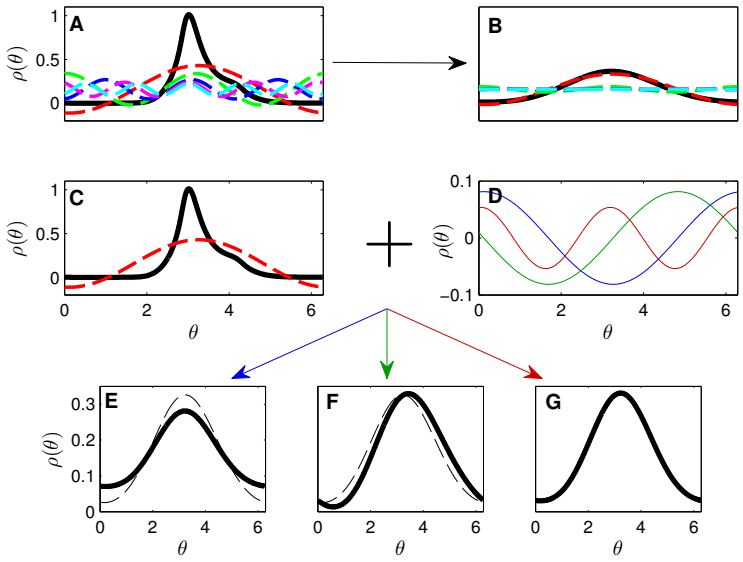

Figure 6.1: Illustration of isostables in the advection-diffusion equation. Panel A shows an initial distribution in black and several associated Fourier modes as dashed lines. After some time passes, panel B shows that all other modes except for the first have decayed substantially. Panel C shows the same initial distribution with the first mode shown in red for reference, to which we add one of three perturbations given in panel D. After $4T$ has elapsed, the distribution resulting from the blue, green, and red perturbation is shown as a solid line in panels E, F, and G, respectively, with the unperturbed distribution after $4T$ has elapsed shown as a dashed line. The perturbed and unperturbed distributions are indistinguishable in panel G.

$\omega = 2\pi/T$ represents the neuron's baseline frequency and is determined from its natural period T, $Z(\theta)$ is the phase response curve, which gives the change in phase due to an infinitesimal impulsive current, σ is the coupling strength in an all-to-all network, $\sqrt{2S}\eta_j(t)$ is i.i.d. zero mean white noise with intensity $2S$, $u(t) = I(t)/C$ with $I(t)$ being a common external current control input and $C = 1\mu F/\text{cm}^2$ the constant neural membrane capacitance, N is the total number or neurons, and $V(\theta)$ gives the transmembrane voltage as a function of the phase. We assume S is small enough that higher order noise terms are negligible (c.f. [92]). Also, we assume that the coupling in this network is electrotonic, but this could be generalized to include, for example, chemical synaptic coupling. Assuming that the population is large and noise perturbations are small, with stochastic averaging

[90] we can approximate (6.23) according to its probability density $\rho(\theta, t)$ with the Fokker-Planck equation in one dimension [70]

$$\begin{aligned}\frac{\partial \rho(\theta,t)}{\partial t} &= -\frac{\partial}{\partial \theta}[(\omega + Z(\theta)(u(t) + \sigma(\overline{V} - V(\theta)))\rho(\theta,t)] + \frac{1}{2}\frac{\partial^2}{\partial \theta^2}[B^2\rho(\theta,t)] \\ &= -\omega\rho_\theta + \frac{B^2}{2}\rho_{\theta\theta} \quad \underbrace{-[Z(\theta)\rho_\theta + Z'(\theta)\rho]u(t)}_{\text{external input}} \quad \underbrace{-\frac{\partial}{\partial \theta}[Z(\theta)\sigma(\overline{V} - V(\theta))\rho]}_{\text{intrinsic coupling}}, \quad (6.24)\end{aligned}$$

where $\overline{V} = \int_0^{2\pi} V(\theta)\rho(\theta)d\theta$ and $B^2 = \frac{2S}{2\pi}\int_0^{2\pi} Z^2(\theta)d\theta$ represent averaged transmembrane voltage and noise, respectively. For the moment, we will consider (6.24) in the absence of external input and intrinsic coupling. Without these terms, we have a linear advection-diffusion equation on a ring, whose eigenfunctions are related to the Fourier modes of the distribution. Panel A of Figure 6.1 shows an example distribution in black along with several corresponding Fourier modes, shown as dashed lines. After some time passes, all but the first mode has significantly decayed, as shown in panel B. The behavior of this system can be understood from the perspective of isostables, which as outlined in [26] are sets of initial conditions which share the same asymptotic convergence towards a stationary solution, in this case the uniform distribution $\rho = 1/(2\pi)$. As shown for this advective-diffusive system in Appendix C, the magnitude of the first mode determines the isostable, and thereby the asymptotic approach of the system to the stationary solution, because all other modes die out at a faster rate in the presence of noise.

Solutions to (6.24) exist in infinite dimensional space, and for this reason it can be difficult to work with for control purposes. To reduce the complexity of this equation, one can instead track the change in isostables in response to an external perturbation, with examples given in panels C-G of Figure 6.1. Panel C shows the probability distribution along with the first mode, for reference, to which we add one of the three perturbations shown in panel D. Panels E, F, and G show the resulting distribution after $4T$ elapses.

We find that adding the blue perturbation, which is antiphase to the first mode, results in a faster approach to the stationary solution, as shown in panel E. Adding a perturbation orthogonal to the first mode (shown in green) does not affect the isostable, but shifts the distribution to the right slightly, shown in panel F. Adding a perturbation corresponding to a higher order mode (shown in red) does not effect the system's approach to the stationary solution, shown in panel G.

Using the intuition gleaned from Figure 6.1, one can understand the essence of the following isostable reduction: in isostable coordinates, (6.24) becomes

$$\begin{aligned}\dot{\psi} &= \kappa + \left\langle \mathcal{I}(\theta,\psi), -[Z(\theta)\rho_\theta + Z'(\theta)\rho]u(t) - \frac{\partial}{\partial\theta}[Z(\theta)\sigma(\overline{V}-V(\theta))\rho] \right\rangle, \\ \mathcal{I}(\theta,\psi) &= \frac{-2\kappa}{A_1^* B^2 \pi}\cos\left(\theta - \varphi_1^* - \frac{\omega\psi}{\kappa}\right) e^{\frac{B^2\psi}{2\kappa}}. \end{aligned} \tag{6.25}$$

Here, ψ is the isostable of the system with $\psi \equiv 0$ corresponding to an arbitrary time t^*, $\kappa \equiv 1$ represents the system's natural rate of approach to the stationary solution, $\mathcal{I}(\theta,\psi)$ is the infinitesimal isostable response curve (iIRC), which represents the gradient of the isostable field, $\langle\cdot,\cdot\rangle$ is the L^2 inner product, and $A_1^*\cos(\theta-\varphi_1^*)$ is the first Fourier mode of $\rho(\theta,t^*)$. Complete details of the derivation of this reduction are given in Appendix C. The iIRC for this system is shown in panel B of Figure 6.2. The magnitude of the iIRC grows exponentially with increasing ψ so that it requires smaller perturbations to influence the isostable of the system as it approaches the stationary solution (c.f. [38]).

Equation (6.25) is a one dimensional ODE, and is much easier to work with than the infinite dimensional PDE from (6.24), yet still retains essential characteristics of original equation. Equation (6.25) can be understood as follows: if the inner product is zero, noise will be responsible for driving the system to larger isostables, which are closer to the stationary, desynchronized distribution. However, the term responsible for internal coupling, $-\frac{\partial}{\partial\theta}[Z(\theta)\sigma(\overline{V}-V(\theta))\rho]$, can counteract the effects of noise, maintaining some

level of synchrony in the network. Therefore, we can use the term associated with the external control, $-[Z(\theta)\rho_\theta + Z'(\theta)\rho]u(t)$, to reverse the effects of internal coupling, driving the distribution closer to the stationary solution, thereby desynchronizing the network. Viewing the network from this perspective leads to a novel control strategy, which would not be apparent from the full equations (6.24). Assuming each neuron receives the same input $u(t)$, and assuming that the maximum and minimum input that can be applied is $u_{\max} \geq 0$ and $u_{\min} \leq 0$, respectively, which may be relevant for hardware or tissue limitations, we can apply the control

$$u(t) = \begin{cases} u_{\max} & \text{if } \langle \mathcal{I}(\theta, \varphi_1(\rho)), -[Z(\theta)\rho_\theta + Z'(\theta)\rho] \rangle > 0, \\ u_{\min} & \text{otherwise.} \end{cases} \tag{6.26}$$

When this strategy is implemented, the external control serves to maximize the right hand side of (6.25), driving the system towards higher values of ψ, which correspond to less synchronous population distributions.

We apply the control strategy to a population of neurons, each with intrinsic dynamics characterized by a two dimensional reduction of the Hodgkin-Huxley (HH) equations [68] that reproduces the essential characteristics of the neuron's dynamical behavior, [154], c.f. [155]:

$$\begin{aligned} \dot{V}_j &= f_V(V_j, n_j) + I + \frac{\sigma}{N}\sum_{i=1}^{N}(V_i - V_j) + u(t) + \eta_j(t), \\ \dot{n}_j &= f_n(V_j, n_j). \end{aligned} \tag{6.27}$$

Here, f_V and f_n are functions which describe the intrinsic neural dynamics, with all terms and parameters given in [155], $I = 10\ \mu\text{A/cm}^2$ is a baseline current given so that each neuron is in a periodically spiking regime, $j = 1, \ldots, N$ where N is the total number of

neurons, and V_j and n_j are the transmembrane voltage and gating variable for neuron j, respectively. Note that (6.23) can be obtained from (6.27) through phase reduction. The phase response curve, $Z(\theta)$, which gives the change in phase due to an infinitesimal impulsive current stimulus is calculated using the software XPP [65] and is shown in panel A of Figure 6.2. To test our control strategy (6.26), we take the coupling strength σ to be 0.06 and a white noise intensity $2S = 2$, (with the resulting value $B^2 = 0.0251$), and $u_{\max} = -u_{\min} = 0.7\mu\text{A}/\mu\text{F}$. The specific values of $u_{\max}$ and $u_{\min}$ can be changed without reducing the effectiveness of the control strategy provided they are large enough to overcome the synchronizing effects of intercellular coupling, but small enough that the phase reduction (6.23) is valid [4].

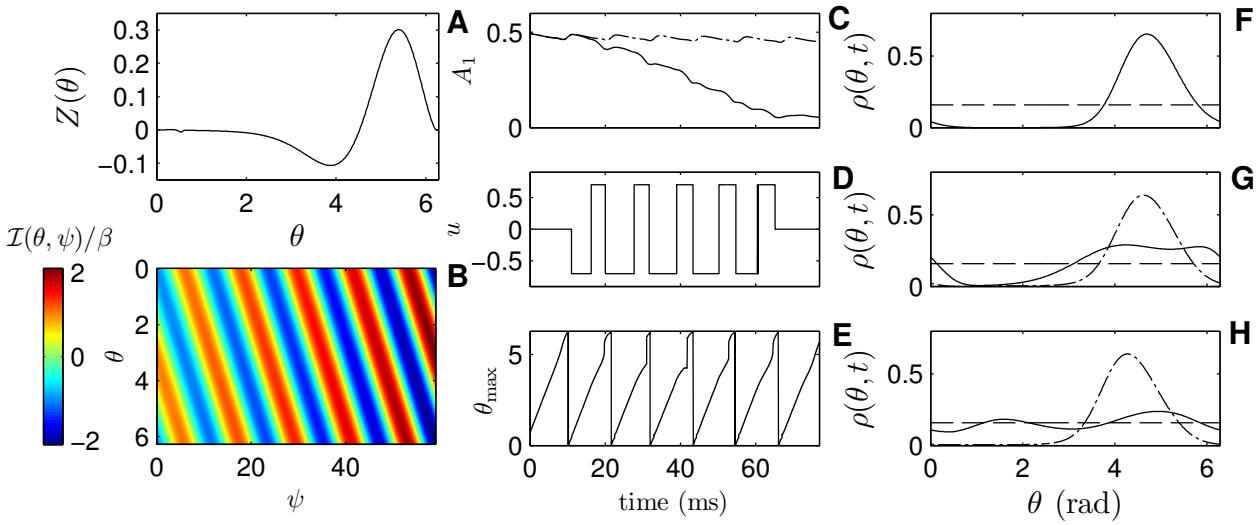

Figure 6.2: Isostable reduction applied to the Fokker-Planck equation. Panel A shows $Z(\theta)$, used in the desynchronizing control strategy. Panel B shows the iIRC, $\mathcal{I}(\theta,\psi)/\beta$ where $\beta \equiv 2\kappa/A_1^* B^2 \pi$. Panels C-H show the result of applying the control strategy to the PDE (6.24). The control input (in μA/μF) in D drives the first mode of the distribution from 0.5 to 0.05. Panel C gives the amplitude of this mode as a function of time for the equations with control applied as well as the system without control, shown as a solid and dashed-dot line, respectively. Panel E gives the location of the maximum of the distribution, $\theta_{\max}$, as a function of time. Panels F-H show a comparison of the controlled distribution to the uncontrolled distribution at $t = 7.4, 41.8$ and 75.8 ms. The horizontal dashed line shows the ideal, uniform distribution.

In the absence of external control, the internal coupling will cause a large population of neurons to approach a distribution that is periodic in time, shown at a particular snapshot in panel F of Figure 6.2. Starting with this distribution, we simulate the reduced PDE (6.24) with our control strategy. The resulting control is shown in panel D of Figure 6.2. Panels F-H show snapshots of the controlled distribution as solid lines with the uncontrolled distribution shown as a dot-dashed line for reference. We find that the applied control drives the system close to a uniform distribution with a stimulus that is close to periodic. This is useful because the control strategy (6.26) requires knowledge of the underlying distribution, which is generally unavailable when simulating the full neural network (6.27). Instead we can make the following observations about the desynchronizing control to design a simple closed loop control strategy: 1) The average voltage of the system achieves a maximum value approximately when $\theta_{\max} = 2\pi$. 2) Each time $\theta_{\max} = 2\pi$, the external control is at $u_{\min}$. 3) The control transitions from $u_{\min}$ to $u_{\max}$ at approximately 6.1 ms after $\theta_{\max} = 2\pi$. 4) The control remains at $u_{\max}$ for approximately 3.7 ms before transitioning back to $u_{\min}$.

These observations allow for the design of the following control strategy for use in the full neural equations. Monitor the average voltage, $\overline{V} = \sum_{j=1}^{N} V_j$, of the system. Let t_{spike} denote the most recent time $\overline{V}$ reaches a local maximum which is above -50 mV, denoting that $\theta_{\max} \approx 2\pi$, and apply the control

$$u = \begin{cases} u_{\max} & \text{if } 3.7 < t - t_{\text{spike}} < 9.8, \\ 0 & \text{if } t - t_{\text{spike}} > 15, \\ u_{\min} & \text{otherwise.} \end{cases} \tag{6.28}$$

Notice that if the average voltage remains low enough, the controller is simply turned off. We can apply this control when the average voltage spikes above -50 mV, and turn

the controller off until the population has resynchronized.

We apply (6.28) to a population of $N = 300$ neurons, shown in the top panel of Figure 6.3 where the black lines give individual neural traces for 30 representative neurons from the population, and the red line gives the average voltage for the system. We find that the applied control (middle panel), quickly desynchronizes the system, as can be seen from the Kuramoto order parameter [4], shown in the bottom panel (for the splay state, $R = 0$). Turning off the controller allows the neurons to resynchronize, at which point the controller can be turned on again to desynchronize the system. We note that the control strategy which produces chaotic desynchronization from [10] cannot desynchronize the population with this parameter set. This happens because individual cells tend to break off from the synchronized population, quickly travel around the limit cycle, and rejoin the synchronized population before it can desynchronize.

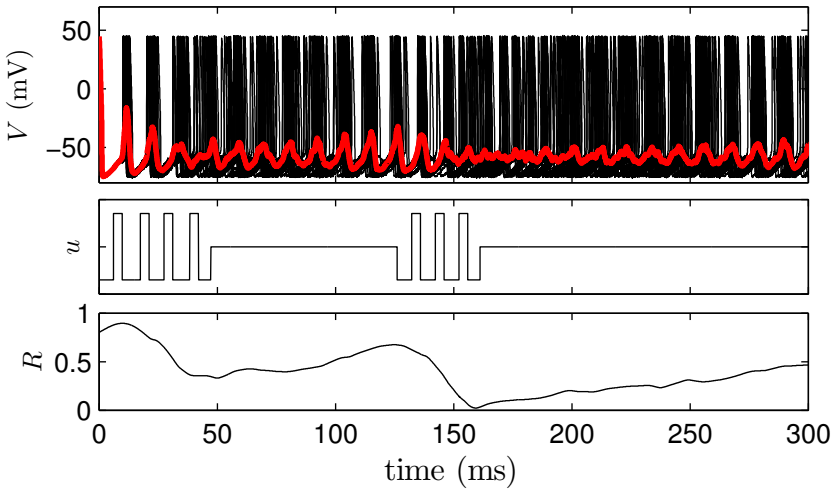

Figure 6.3: The top panel shows the approximated control strategy (6.28) applied to a population of Hodgkin-Huxley neurons, with the middle and bottom panels showing the applied control and the Kuramoto order parameter, R, for the network, respectively.

6.4 Application to Cardiac Alternans

A single cardiac cell can be viewed as an excitable system: a large enough perturbation to the transmembrane voltage of an individual cell will result in an action potential and cause a contraction of the muscle tissue [2]. Cardiac tissue is made up of many individual cells, electrically coupled through gap junctions, allowing waves of depolarization to spread through the medium in a coordinated manner, resulting in a heartbeat. Cardiac fibrillation, also known as cardiac arrest, represents an interruption of this coordinated activity, and is a leading cause of death in the industrialized world. Fibrillation is caused by unwanted and self-sustaining spiral waves within the heart which interfere with the normal rhythm [28], [29].

Cardiac fibrillation can be the final step in a complex series of events beginning with cardiac alternans, the pathological beat to beat alternation of electrochemical cardiac dynamics at a constant rate of pacing (see e.g. the bottom panel of Figure 6.4). At a cellular level, alternans often manifest in alternating long and short action potential durations (APDs), defined as the length of time the transmembrane voltage remains above resting potential after the cell fires. While alternans can be present during physiological cardiac conditions, they have been the subject of much attention in recent years because they produce favorable conditions for initiating fibrillation [30, 119, 128, 129, 131], and it has been postulated that eliminating alternans in tissue could serve as a method of prevention for those who are susceptible to cardiac fibrillation.

In [38], using a newly developed methodology of isostable reduction in ODEs, we were able to eliminate alternans in a model of a single excitable cardiac cell using significant less energy than other control strategies [34], [35]. However, the control strategy previously developed in [38] lacked the framework necessary for application in connected cardiac tissue, which is more relevant to the true problem. Here, we apply our isostable reduction

method to connected tissue modeled as a PDE.

We consider a ring of connected cardiac tissue

$$C_m \frac{\partial V(r,t)}{\partial t} = D \frac{\partial^2}{\partial r^2} V(r,t) + (-I_{\text{ion}}(r) + I_{\text{stim}}(r,t)), \tag{6.29}$$

where V represents the transmembrane potential, $D = 1.5$ cm^2/s is the diffusion constant, $I_{\text{ion}}(r)$ gives the membrane current density, $I_{\text{stim}}(r,t)$ is an external current controlled input, $C_m = 1\mu F/\text{cm}^2$ is the membrane capacitance, and r gives the position around the ring. While the diffusive term in (6.29) is linear, the underlying ionic currents are, in general, nonlinear functions of the underlying cellular dynamics. In this study, we consider the Fox model of canine cardiac tissue [132] for membrane current density and other gating variables. We take all parameters to be nominal. Equation (6.29) has periodic boundary conditions, $V(r + l) = V(r)$, and for a long enough ring the tissue admits traveling wave solutions. If we let the length of the ring be $l = 10$ cm, and create a traveling wave in the medium, we find alternans develop in the system, as shown in Figure 6.4.

6.4.1 Isostable Reduction of the Cardiac System

Our goal is to eliminate alternans in the system using an external control $I_{\text{stim}}(r,t)$. At a cellular level, alternans is caused by the existence of a stable period-2 orbit, and in Chapter 5 we found that it was possible to use a control algorithm based on an isostable reduction to eliminate alternans by stabilizing the underlying unstable period-1 orbit (c.f. [31], [34], [35]). Ultimately, terminating alternans is a problem of controlling the timing of cell repolarization. By appropriately speeding up the repolarization of the long action potentials, the unstable period-1 orbit can be stabilized to eliminate alternans in the system; with knowledge of the system's iIRC, one can know exactly how an external

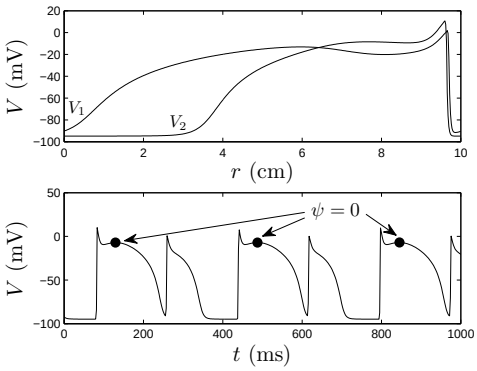

Figure 6.4: The top panel shows spatially concordant alternans throughout the ring. The snapshots V_1 and V_2 are 175 ms apart, approximately the time it takes for the wave to travel once throughout the medium. In the snapshot V_1, the action potentials last much longer, while in the snapshot V_2, the cells repolarize quickly. The bottom panel shows the voltage profile of a single cell in the medium at $r = 5$ cm, showing that alternans are also present at the cellular level. The bottom panel also shows the location at which we define $\psi \equiv 0$ on the long alternan.

perturbation will affect a cell's repolarization. In this example the stationary solution occurs when all cells within the ring are completely quiescent (with no traveling pulse). In the control strategy to follow, alternans will be eliminated by perturbing the system on its way towards the stationary solution, but the traveling pulse will never die out.

To cast (6.29) in the form of equation (6.1) so that we may calculate isostable response curves, let $\mathbf{X} = \left[V, [\mathrm{Ca}^{2+}]_i, [\mathrm{Ca}^{2+}]_{sr}, f, d, m, h, j, f_{\mathrm{Ca}}, X_{\mathrm{Kr}}, X_{\mathrm{Ks}}, X_{\mathrm{to}}, Y_{\mathrm{to}}\right]^T$, consisting of all cellular state variables. Let $F(\mathbf{X}(r,t))$ correspond to the differential equations for the cellular dynamics with a nominal parameter set given in [132]. Let G($\mathbf{X}$(r,t)) be the diffusive voltage coupling between cells, leaving $p(r,t) = [I_{\mathrm{stim}}/C_m \ 0 \ \ldots \ 0]^T$. Consider an isostable model of this cardiac system

$$\begin{aligned}\dot{\psi} &= \kappa + \langle \mathcal{I}_V(r,\psi), I_{\mathrm{stim}}(r,\psi)\rangle + \langle \mathcal{I}_{[\mathrm{Ca}^{2+}]_i}(r,\psi), \mathbf{0}\rangle + \cdots + \langle \mathcal{I}_{Y_{\mathrm{to}}}(r,\psi), \mathbf{0}\rangle \\ &= \kappa + \langle \mathcal{I}(r,\psi), I_{\mathrm{stim}}(r,\psi)\rangle. \end{aligned} \tag{6.30}$$

Because we can only perturb in the voltage direction, we define $\mathcal{I}(r,\psi) \equiv \mathcal{I}_V(r,\psi)$. For convenience, we scale the isostable field so that $\kappa = 1$. When γ is close to the stationary solution, we can determine the iIRC for use as an initial condition for (6.9). Solely for the purposes of calculating the system's iIRC, we reduce the dimensionality of the system by averaging [156], taking the slowly varying sarcoplasmic reticulum Ca^{2+} to be equal to a constant 318 μmol and modify the ionic current I_{Kr} (given in [132]) so that $I_{\mathrm{Kr}} = 0$ when the variable $X_{\mathrm{Kr}} < 10^{-4}$. These modifications make the eigenfunctions near the stationary solution easier to calculate without perceptibly altering the dynamics. With these modifications, when the system is close to the stationary solution, the value of $X_{\mathrm{Kr}}(r)$ (the variable associated with the gating variable for one of the potassium currents) approaches the stationary solution according to $X_{\mathrm{Kr}}(r) = \lambda_1(X_{\mathrm{Kr}}(r) - X_{\mathrm{Kr},\infty})$ with $X_{\mathrm{Kr},\infty}$ denoting a steady state value of X_{Kr} and λ_1 denoting the slowest eigenvalue of the linearized system. Because the rate of approach of $X_{\mathrm{Kr}}(r)$ is the same at all locations r, there are an infinite number of eigenfunctions corresponding to the slowest direction of the stable manifold which span the function space $X_{\mathrm{Kr}}(r)$). These eigenfunctions span $X_{\mathrm{Kr}}(r)$ and are orthogonal to all other eigenfunctions with no component in $X_{\mathrm{Kr}}(r)$. Consider a cell at $r = r_0$. When the system is close to the stationary solution, any perturbation to $X_{\mathrm{Kr}}(r)$ for $r \neq r_0$ will have no effect on $X_{\mathrm{Kr}}(r_0)$ as it approaches the steady state. With this information, and using (6.3), near the stationary solution, the iIRC associated with a cell at r_0, which describes its infinite time approach towards the stationary solution, is given by

$$\mathcal{I}_{\mathrm{X_{Kr}}}(r,\psi) = \begin{cases} 1/\dot{X}_{\mathrm{Kr}} & \text{if } r = r_0, \\ 0 & \text{otherwise} \end{cases}. \tag{6.31}$$

With this definition, the iIRC describes the change in isostable for a cell at a given

location r_0 caused by a small perturbation. Note that (6.31) is only valid near the stationary solution.

To numerically calculate the iIRC using (6.9), the trajectory γ must approach the stationary solution so that (6.31) can be used to obtain an initial condition for (6.9). This can be accomplished, for instance, by taking $r_0 = 5$ cm, and locally setting $D = 0$ between $r = 0$ cm and $r = 10$ cm at a moment immediately after the cell at r_0 fires. This allows the traveling waves to die out so that the system approaches the stationary solution. With an initial condition determined from (6.31), the iIRC at all other values in time is determined by integrating (6.9) backwards in time. As represented in Figure 6.5, the resulting iIRC is valid until t_D, the time that the cell at r_0 would have fired its next action potential.

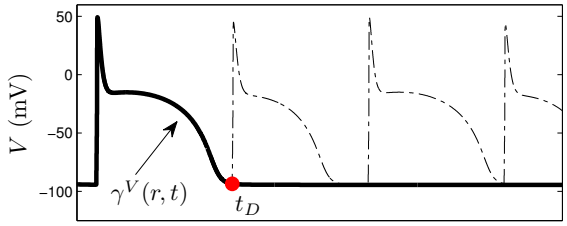

Figure 6.5: In order to calculate the iIRC, γ must come arbitrarily close to the stationary solution. By modifying the boundary conditions to allow the traveling wave to die out, we can numerically calculate an iIRC, which is accurate for the unmodified system until t_D. Here $\gamma^V(r,t)$ represents the voltage component of γ at location r along the ring.

The system does not take the same trajectory γ towards the stationary solution every time, and we calculate the iIRC 96 times using different initial conditions to determine an average iIRC, $\overline{\mathcal{I}}(r,\psi)$ shown in the top panel of Figure 6.6, with the associated standard deviation $\sigma_{\mathcal{I}}$ shown in the bottom panel. For the problem of eliminating alternans, each cell can be thought of as having its own isostable value and iIRC, but because the system is circularly symmetric, we simply report the iIRC in terms of the effect of a perturbation

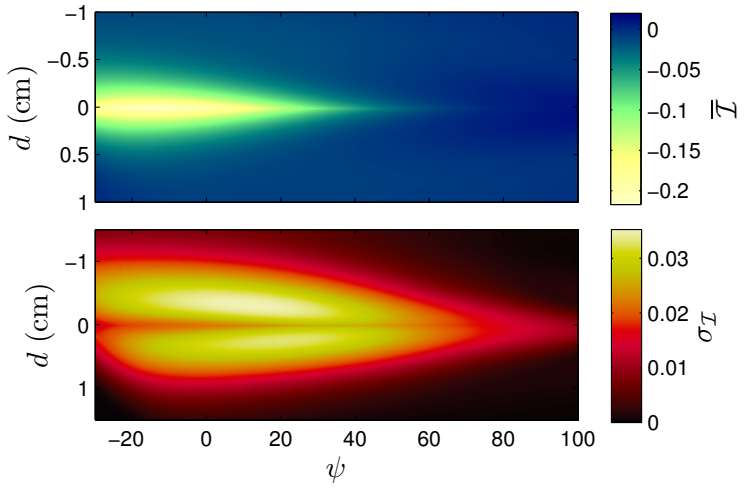

Figure 6.6: Top and bottom panels show the mean and standard deviation of the iIRC, respectively, taken over 96 trials. Here d is defined so that $\mathcal{I}(d, \psi)$ represents the effect of a perturbation on a cell d cm away.

given at a signed distance d. Also, we define $\psi = 0$ to correspond to the time at which $\frac{\partial V}{\partial t}\big|_{r=r_0} = 0$, during a long alternan. The effect of an external stimulus is largest when d, is zero, corresponding to direct stimulation of that cell, and is near zero when $|d| > 0.5$ cm. The standard deviation of the iIRC at $d \approx 0$ is small compared to the magnitude of the iIRC itself, allowing us to use $\overline{\mathcal{I}}(r, \psi)$ as an approximation of the true iIRC in the control strategy to follow.

6.4.2 Developing a Control Objective to Eliminate Cardiac Alternans

Recall that if control is applied to a single cell, alternans can be eliminated by driving the cell dynamics to the unstable period-1 orbit. However, in tissue we must eliminate alternans in the connected tissue with control nodes that are sparsely distributed throughout the tissue. For the following analysis, we consider a single cell isostable reduction at each cell on the ring, using methods similar to those of [38]. We assume the

existence of an alternans free state, $\psi_0(r,t)$, throughout the ring which evolves so that $\psi_0(r,t) = \psi_0(r - ct, 0)$, with $c > 0$. Furthermore, we assume that the state of each cell is close to the alternans-free trajectory, denoted by $\gamma_a + \mathcal{O}(\epsilon||\Delta x||)$, where x represents the state variables of the system. This means that to leading order ϵ, there is a 1 to 1 relationship between ψ and the state variables.

Panel A of Figure 6.7 shows such a distribution, which can be found by applying the control strategy (6.54) to the dynamic equations (6.29) until the control effort disappears [31]. This distribution is an unstable solution to (6.29) which is periodic in both space and time; the wavefront (located at $r \approx 4.6$ cm in the figure) travels at approximately 57 cm/s so that the period of oscillation is approximately 175 ms. Isostables for each cell are scaled so that at all locations except the wavefront $\dot{\psi} = 1$ in the absence of external stimuli. Here, $\psi = 0$ is defined to be the time at which a cell has just repolarized (i.e. reached 95 percent of its resting potential). At the depolarizing wavefront, quiescent cells are reexcited. The end goal of our control strategy is to guide the system to this unstable, alternans-free periodic orbit.

If we consider a single cell at location $r = x$, the dynamics evolve according to (6.29):

$$C_m \left.\frac{\partial V(\psi_0)}{\partial t}\right|_{r=x} = D \left.\frac{\partial^2 V(\psi_0)}{\partial r^2}\right|_{r=x} - \left. I_{\text{ion}}(\psi_0)\right|_{r=x} + I_{\text{stim}}(x,t). \tag{6.32}$$

Through single cell isostable reduction, provided the state dynamics are close to γ_a, we may write (6.32) as

$$\left.\dot{\psi}\right|_{r=x} = \kappa + \mathcal{I}_s(\left.\psi\right|_{r=x}) I_{\text{stim}}(x,t). \tag{6.33}$$

Here, $\left.\psi\right|_{r=x}$ gives the isostable of the single cell dynamics at the location $r = x$, and $\mathcal{I}_s$ is the single cell isostable response curve. Note that in this reduction, both $\left. I_{\text{ion}}(\psi)\right|_{r=x}$ and $D \left.\frac{\partial^2 V(\psi_0)}{\partial r^2}\right|_{r=x}$ guide the cell dynamics along the trajectory γ_a, and therefore disappear in the reduction (6.33).

As derived in Chapter 5, we can calculate $\mathcal{I}_s(\psi)$ by solving an adjoint equation

$$\frac{d\mathcal{I}_s}{dt} = -\nabla F^T(x(\psi))\mathcal{I}_s. \tag{6.34}$$

Here, ∇F is the Jacobian, and $\nabla F^T(x(\psi))$ is the transpose, or adjoint, of the real-valued matrix $\nabla F(x(\psi))$, and $F(x)$ represents state dynamics of the transmembrane voltage and all auxiliary variables associated with I_{ion}.

Suppose now that the distribution is not exactly at the ideal, ψ_0, but rather can be written as $\psi(r,t) = \psi_0(r,t) + \epsilon\psi_\epsilon(r,t)$, with $0 < \epsilon \ll 1$, so that ψ_ϵ represents a small perturbation from the ideal distribution. For clarity of notation in the following equation, we let $P \equiv \psi_\epsilon|_{r=x}$, which is a constant with respect to derivatives in space. Suppose also that there are no external perturbations so that $I_{\text{stim}} = 0$. To leading order, the dynamics at a given location x, starting with (6.32) can be written as

$$\begin{aligned} C_m \left.\frac{\partial V(\psi_0+\epsilon\psi_\epsilon)}{\partial t}\right|_{r=x} &= D \left.\frac{\partial^2 V(\psi_0+\epsilon\psi_\epsilon+\epsilon P-\epsilon P)}{\partial r^2}\right|_{r=x} - \left. I_{\text{ion}}(\psi_0+\epsilon\psi_\epsilon)\right|_{r=x} \\ &= D \left.\frac{\partial^2 \left(V\left(\psi_0+\epsilon P\right)+\epsilon\frac{\partial V}{\partial\psi}\cdot(\psi_\epsilon-P)\right)}{\partial r^2}\right|_{r=x} - \left. I_{\text{ion}}(\psi_0+\epsilon\psi_\epsilon)\right|_{r=x} + \mathcal{O}(\epsilon^2) \\ &= D \left.\frac{\partial^2 V\left(\psi_0+\epsilon P\right)}{\partial r^2}\right|_{r=x} - \left. I_{\text{ion}}(\psi_0+\epsilon\psi_\epsilon)\right|_{r=x} + \epsilon D\frac{\partial V}{\partial\psi}\cdot\left.\frac{\partial^2\psi_\epsilon}{\partial r^2}\right|_{r=x} + \mathcal{O}(\epsilon^2). \end{aligned} \tag{6.35}$$

Finally, equation (6.35) is in the same form as (6.32), with the term $\epsilon D\frac{\partial V}{\partial\psi}\cdot\left.\frac{\partial^2\psi_\epsilon}{\partial x^2}\right|_{r=x}$ taking the place of $I_{\text{stim}}(x,t)$. We may therefore rewrite (6.35) in the form of (6.33):

$$\left.\left(\dot{\psi}_0+\epsilon\dot{\psi}_\epsilon\right)\right|_{r=x} = \kappa + \left.\left(\mathcal{I}_s(\psi_0+\epsilon\psi_\epsilon)\epsilon D\frac{\partial V}{\partial\psi}\cdot\frac{\partial^2\psi_\epsilon}{\partial r^2}\right)\right|_{r=x} + \mathcal{O}(\epsilon^2). \tag{6.36}$$

Finally, subtracting (6.33) with $I_{\text{stim}} = 0$ from (6.36) and Taylor expanding the $\mathcal{I}_s$ in

orders of ϵ, we have

$$\epsilon\dot{\psi}_\epsilon\Big|_{r=x} = \left(\epsilon D\frac{\partial V}{\partial \psi}\cdot\mathcal{I}_s(\psi_0)\frac{\partial^2\psi_\epsilon}{\partial r^2}\right)\Big|_{r=x} + \mathcal{O}(\epsilon^2). \tag{6.37}$$

Equation (6.37) mandates that for all points at which the isostables are continuous in space, provided that $\left(\epsilon D\frac{\partial V}{\partial \psi}\cdot\mathcal{I}_s(\psi_0)\right)\Big|_{r=x} > 0$, any local isostable gradient will spread through the system diffusively. Panels C and D of Figure 6.7 shows the term $\frac{\partial V}{\partial \psi}(\psi)\cdot\mathcal{I}_s(\psi_0)$ plotted as a function of ψ for this system at two different scales, with the transmembrane voltage shown in panel B for reference. We find that this term is strictly positive, and is particularly large in magnitude soon after an action potential.

Recall that (6.37) is only valid for points at which the isostable distribution is continuous in space. For the point at which this distribution is not continuous (i.e. at the wavefront) the dynamics can be reduced using the action potential duration (APD) restitution curve which gives the next action potential duration, defined to be the length of time the transmembrane voltage of the cell remains above resting potential after an action potential is fired, as a function of the diastolic interval (DI), defined to be the amount of time the cell stays quiescent before the next action potential is fired:

$$\mathrm{APD}_{i+1} = f(\mathrm{DI}_i). \tag{6.38}$$

Here, because we have a series of action potentials, we refer to them with an index i. The APD restitution curve and its first derivative are shown in panels E and F of Figure 6.7, respectively. In this case, the basic cycle length (BCL), defined as the time between succesive action potentials, is fixed, and determined by the wave speed c and the length of the ring, and therefore BCL = DI + APD. One can show that (6.38) has a fixed point DI_0 when $\mathrm{BCL} = f(\mathrm{DI}_0) + \mathrm{DI}_0$. As originally shown in [134], this fixed point will be unstable

if $\left|\frac{df}{d\mathrm{DI}}(\mathrm{DI}_0)\right| > 0$, and this instability gives rise to the stable period-2 orbit responsible for alternans. We have strategically defined the isostables in this system so that positive and negative isostable values can be used interchangeably with DI and APD in equation (6.38), respectively. Therefore, when a cell is reexcited, we can write (c.f. [157])

$$\psi_{\epsilon_{i+1}}(x) = -\Lambda\psi_{\epsilon_i}(x), \tag{6.39}$$

where $\Lambda = \frac{df}{d\mathrm{DI}}(\mathrm{DI}_0)$, and $\psi_{\epsilon_i}(x)$ and $\psi_{\epsilon_{i+1}}(x)$ denote the value of ψ_ϵ at location $r = x$ before and after the cell fires its next action potential. Equation (6.37) and (6.39) specify a reduced system. The diffusive element of the system (6.37) will tend to bring it towards $\psi_\epsilon(r) = \overline{\psi_\epsilon}$, the average value of the initial distribution, and equation (6.39) describes the inherent instability of each cell. Therefore, our control strategy will be to actively drive the system to $\psi_0(r) + \epsilon\psi_\epsilon(r)$ where $\overline{\psi_\epsilon} = 0$. Provided that the diffusion acts much faster than the inherent instability at a cellular level, the system will settle close to ψ_0. If the diffusion does not act quickly enough, we can apply the external control a few cycles later in order to gradually drive the system to ψ_0.

Our objective for the elimination of alternans is now to drive the system to a state $\psi(r) = \psi_0(r) + \epsilon\psi_\epsilon(r)$. Suppose that we have $2M + 1$ control nodes, equally spaced at a distance L. Let $u_N(t)$ represent the control applied at node N which is of order ϵ. Using the coordinate system presented in Figure 6.8, the local reduced cell dynamics at position r, which we denote by ψ_r, are

$$\dot{\psi}_r = \kappa + \sum_{j=-M}^{M} u_{N+j}(t)\mathcal{I}(jL - r, \psi_r) + \mathcal{O}(\epsilon^2). \tag{6.40}$$

Here, the summation represents the effect of input from each control node. Suppose that throughout the tissue, severity of alternans is similar, (i.e. $\mathrm{APD}_n - \mathrm{APD}_{n-1} = \alpha + \mathcal{O}(\epsilon)$).

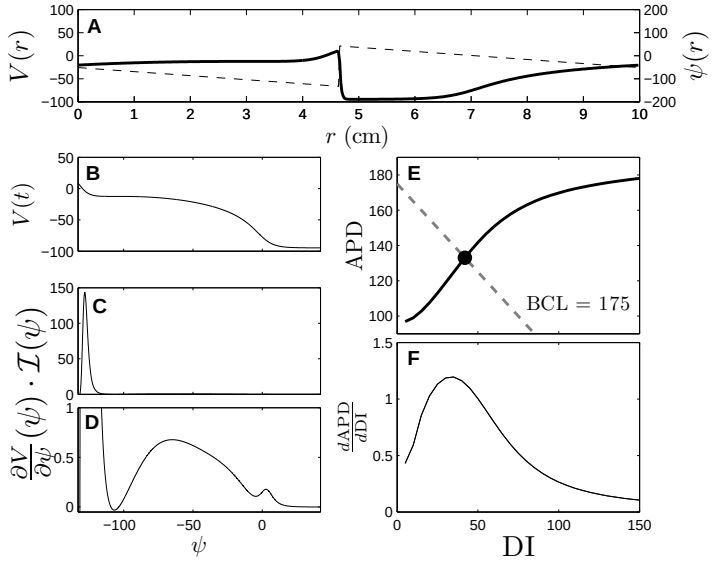

Figure 6.7: Panel A shows an alternans-free solution of (6.29), and an underlying isostable distribution for each individual location throughout the network. The scales on the left and right correspond to the solid and dashed curve, respectively. Panels C and D show the term from equation (6.37) as a function of isostable, with the transmembrane voltage shown in panel B for reference. Because $\frac{\partial V}{\partial \psi}(\psi) \cdot \mathcal{I}_s(\psi)$ is predominantly greater than zero, (6.37) shows that the isostable distribution ψ_ϵ should spread diffusively through the network. Panels E and F show the APD restitution curve as well as $d\text{APD}/d\text{DI}$. The BCL in this system is about 175 ms, which gives an unstable fixed point at $\text{DI}_0 \approx 42$ ms.

Suppose also that the wave speed in this region of tissue is equal to $c + \mathcal{O}(\epsilon)$. This means that if $\psi_r = 0$ at $t = 0$, then ψ_{r+d} will reach zero at $t = d/c + \mathcal{O}(\epsilon)$.

We assume that each node acts independently, and require that an $\mathcal{O}(\epsilon)$ change in the severity of alternans will lead to an $\mathcal{O}(\epsilon)$ change in the applied control. We also assume that each node will apply control which lasts for a duration of k ms, starting when the isostable at that node reaches $\psi = 0$. Under these assumptions, to leading order ϵ the external control applied at each node will be identical in shape, but time shifted by the time it takes for the wave to travel between control nodes. This allows us to rewrite

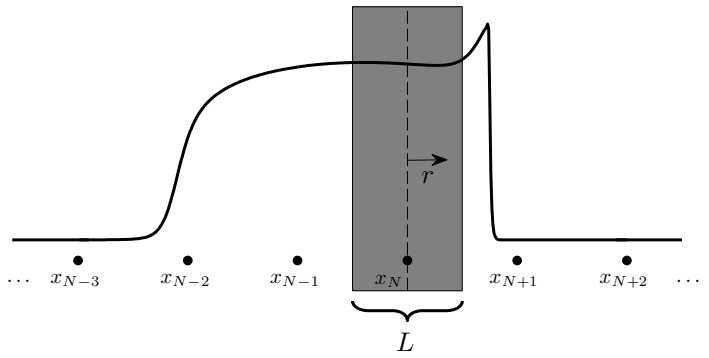

Figure 6.8: The reduced cell dynamics at node x_N are described by equation (6.40).

(6.40) as

$$\dot{\psi}_r = \kappa + \sum_{j=-M}^{M} u(t - jL/c)\mathcal{I}(jL - r, \psi_r) + \mathcal{O}(\epsilon^2), \tag{6.41}$$

where $u(t) = 0$ when either $t < 0$ or $t > k$. Asymptotically expanding ψ_r in powers of ϵ,

$$\psi_r(t) = \psi_r^{(0)}(t) + \epsilon\psi_r^{(1)}(t) + \epsilon^2\psi_r^{(2)}(t), \tag{6.42}$$

we find that $\psi_r^{(0)}(t) = \psi_r(0) + \kappa t$. We will define $t = 0$ to be the time at which $\psi_0 = 0$, which can be written more concisely as $\psi_0(0) = 0$. Therefore $\psi_r(0) = -\kappa r/c$. Substituting this result into (6.41) yields

$$\dot{\psi}_r = \kappa + \sum_{j=-M}^{M} u(t - jL/c)\mathcal{I}(jL - r, \kappa(t - r/c)) + \mathcal{O}(\epsilon^2). \tag{6.43}$$

As shown in Section 6.4.2, if the average distribution of isostables is on the unstable period-1 orbit, the system will relax to that orbit through a process similar to diffusion. Therefore, our goal is drive the average of the isostable distribution in the shaded region of interest in Figure 6.8 to the unstable period-1 orbit. Taking the spatial average of

(6.43) yields

$$\dot{\overline{\psi}} = \frac{1}{L}\int_{-L/2}^{L/2} \dot{\psi}_r dr = \kappa + \frac{1}{L}\int_{-L/2}^{L/2}\left[\sum_{j=-M}^{M} u(t-jL/c)\mathcal{I}(jL-r,\kappa(t-r/c))\right]dr + \mathcal{O}(\epsilon^2), \tag{6.44}$$

where $\dot{\overline{\psi}}$ is the average rate of change in isostables in the shaded region.

We can use (6.44) as part of a strategy to eliminate alternans. Suppose that the next action potentials will be fired T ms after the present action potentials. Then, we may guide the system from the stable period-2 orbit to the unstable period-1 orbit by requiring (c.f. [38])

$$\int_0^T \dot{\overline{\psi}} dt = \kappa T - \zeta\kappa(\mathrm{APD}_n - \mathrm{APD}_{n-1})/2, \tag{6.45}$$

which implies the following boundary condition

$$\int_0^T \left(\dot{\overline{\psi}} - \kappa\right) dt = -\zeta\kappa(\mathrm{APD}_n - \mathrm{APD}_{n-1})/2, \tag{6.46}$$

where $0 < \zeta < 2\pi$ is a positive constant which reflects the relative distance between the long and short alternans pathways to the alternans free distribution [34]. Recall that we apply control when the cells are currently firing long action potentials, so that $\mathrm{APD}_n - \mathrm{APD}_{n-1} < 0$. In equation (6.46), setting $\zeta = 0$ would mean that at the next action potential, no control would need to be applied, and the next action potentials will follow a short alternan trajectory. If $\zeta = 2$, the result of the control would be to speed up the average rate of change of the isostables, and on average, the cells would take another long alternan path for the next action potential. Intuitively, this control requirement assumes that the unstable period-1 orbit lies in between these two paths. Recall that for $t < 0$ and $t > k$, $u(t) = 0$. This allows us to conveniently time shift the terms in the

integrand of (6.46) after substituting (6.44):

$$\int_0^T \left(\dot{\overline{\psi}} - \kappa\right) dt = \int_0^k \left[\frac{1}{L}\int_{-L/2}^{L/2}\left[\sum_{j=-M}^{M} u(t)\mathcal{I}(jL - r, \kappa(t - r/c + jL/c))\right] dr\right] dt. \quad (6.47)$$

For convenience of notation, let $f(t) = \frac{1}{L}\int_{-L/2}^{L/2}\left[\sum_{j=-M}^{M}\mathcal{I}(jL - r, \kappa(t - r/c + jL/c))\right] dr$. By defining an auxiliary equation $\dot{\Psi} \equiv u(t)f(t)$, we can cast this as an optimal control problem by minimizing the cost functional $\mathcal{M}[u(t)] = \int_0^k u^2(t)dt$, which gives the power associated with the stimulus, and can apply calculus of variations to minimize [13]

$$C[u(t)] = \int_0^k [u^2(t) + \lambda(\dot{\Psi} - u(t)f(t))]dt, \quad (6.48)$$

where λ is a Lagrange multiplier. The resulting Euler-Lagrange equations are

$$u(t) = \lambda f(t)/2, \quad (6.49)$$

$$\dot{\Psi} = \lambda f^2(t)/2, \quad (6.50)$$

$$\dot{\lambda} = 0. \quad (6.51)$$

The optimal control $u^*(t)$ can be found by solving (6.50) and (6.51) subject to $\Psi(0) = 0$ and $\Psi(k) = -\zeta\kappa(\mathrm{APD}_n - \mathrm{APD}_{n-1})/2$. By noting that (6.51) requires λ to be a constant, and integrating (6.50) directly, we can explicitly solve for the required control as a function of time:

$$u^*(t) = \frac{-\zeta\kappa(\mathrm{APD}_n - \mathrm{APD}_{n-1})f(t)}{2\int_0^k f^2(\tau)d\tau}. \quad (6.52)$$

Thus, to leading order ϵ, the optimal control is proportional to a superposition of appropriately time shifted iIRCs.

6.4.3 Numerical Results

We assume that we can place $2M+1$ stimulators equally spaced L cm apart throughout the ring. With this framework, not all cells receive the same external input, but as we show in Section 6.4.2, to eliminate alternans it is sufficient to drive system to a distribution $\psi_0(r) + \epsilon\psi_\epsilon(r)$, where $\psi_0(r)$ represents the alternans free isostable distribution and $\epsilon\psi_\epsilon(r)$ is any small perturbation such that $\int \psi_\epsilon(r)dr = 0$. Assuming that the wave travels at a constant speed c throughout the ring, as shown in Section 6.4.2, if each stimulator acts independently and applies a control that is proportional to the severity of alternans (i.e. $u(t) \propto \text{APD}_n - \text{APD}_{n-1}$ where APD_n is the duration of the n^{th} action potential), then an approximation of the energy optimal control to eliminate alternans is

$$u(t) = \frac{-\zeta\kappa(\text{APD}_n - \text{APD}_{n-1})f(t)}{2\int_0^k f^2(\tau)d\tau}, \tag{6.53}$$

where $t = 0$ corresponds to the time at which the controller detects that $\psi = 0$, k is the duration of the applied control, ζ is a positive constant which reflects the relative distance between the long and short alternans pathways to the alternans free distribution (c.f. [34]), and $f(t)$ can be calculated from the iIRC which effectively gives a summation of the average control effort applied by each control node. The PDE (6.29) is rather unwieldy and directly calculating an optimal control for elimination of alternans would not be possible. However, isostable reduction allows for the derivation of the relatively simple control scheme (6.53). Panel A of Figure 6.9 shows the calculated optimal control $u(t)$ when $k = 70$, and $\zeta = 1.3$ scaled by the severity of alternans. Perhaps not suprisingly, when the nodes are spaced closer together, the required control at each node is smaller in magnitude. For comparison, we also test the non-optimal control strategy presented

in [31]:

$$u(t) = \begin{cases} \nu(V(t) - V(t-\delta)) & \text{if } V(t) - V(t-\delta) < 0, \\ 0 & \text{otherwise,} \end{cases} \tag{6.54}$$

where ν is a positive constant and δ is the time between successive depolarizations of a single cell. Intuitively, (6.54) works similarly to (6.53) by providing a hyperpolarizing stimulus when a region of tissue has a long alternan, but also requires simultaneous accurate monitoring of the transmembrane potential.

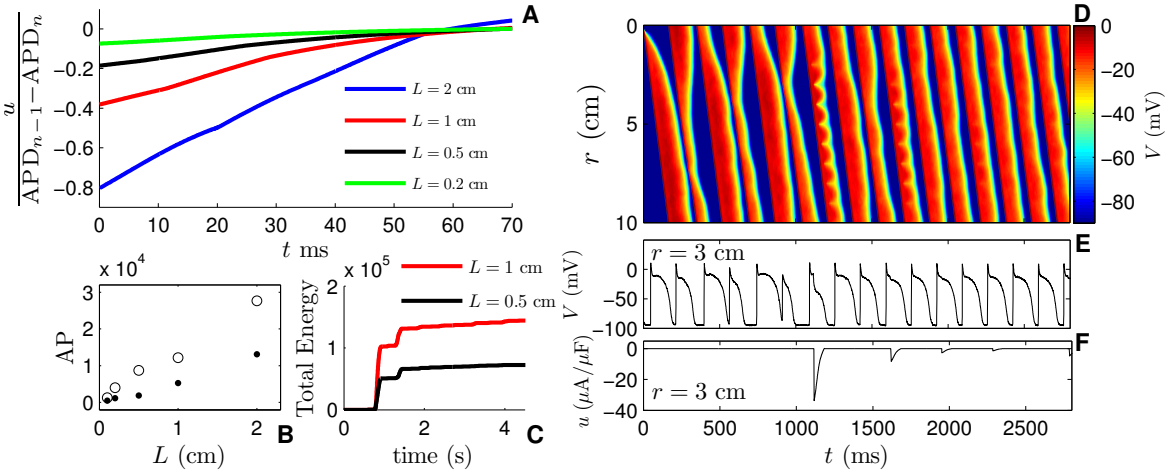

Figure 6.9: Using the iIRC, optimal stimuli to eliminate alternans calculated from (6.53) for and $\zeta = 1.3$ and different node spacings L, are shown in panel A with u given in units of μA/μF and APDs in ms. Panel B compares the average power (AP) once the controller reaches a steady state, for the optimal strategy (dots) and the non-optimal strategy from (6.54) (open circles). At each spacing, the approximately optimal strategy uses between 2 and 4.5 times less energy than the non-optimal strategy. Panel C shows the total cumulative energy consumption as a function of time for two different node spacings. Panel D shows the spatial transmembrane voltage as a function of time with $L = 1$ cm. Alternans can be seen in the tissue when $t < 1000$ ms, and are quickly eliminated when the controller is turned on. Panels E and F show the transmembrane voltage and control effort at $r = 3$ cm.

We test each control strategy on (6.29) with i.i.d. zero mean, unit intensity white noise added. For each control strategy, more power is initially consumed because the controller is driving the system to the period-1, alternans-free orbit (see $t = 1$ to $t = 2$ in

panel C). After this initial period of time, the controller requires considerably less energy $\left(E = \int\int I_{\text{stim}}^2(r,t)dtdr\right)$ to maintain the alternans-free state, making small adjustments to combat the effects of noise. The approximately optimal control strategy uses 2 to 4.5 times less energy than the non-optimal strategy. Results for choices of $\zeta \in [1.1, 1.5]$ are not significantly different.

6.4.4 Possible Experimental Applications

In the previous example, the optimal control strategy requires a reasonably accurate estimate of the iIRC. If a good approximation of the model is known, an iIRC could be calculated using the adjoint method given in Section 6.2. However, given the complicated nature of (6.29) it may be more accurate to measure the iIRC using strategies akin to the direct method [24], [25], a well established technique for experimentally estimating phase response curves in periodically oscillating systems. Here, we detail how a direct method could be used to experimentally implement the control strategy from Section 6.4.1 in e.g. Purkinje fibers [34], [158] which have been used to validate alternans control strategies in one dimensional tissue. Our approach follows from the interpretation of the iIRC, $\mathcal{I}(d,\psi)$, as the change in the isostable coordinate resulting from a perturbation at a distance d, with ψ being the isostable coordinate when the stimulus was applied.

One can gauge how a perturbation affects the isostable using the APD restitution curve [28], which gives the duration of the next action potential as a function of the time since the cell has been quiescent (also known as the recovery time or diastolic interval). Generally, longer recovery times will lead longer action potentials. It will be convenient to define $\psi = 0$ as the time at which a cell repolarizes (i.e. reaches 95 percent of its resting potential) with $\dot{\psi} = 1$ in the absence of forcing so that

$$\text{APD} = f(\psi). \tag{6.55}$$

Note that if we instead define $\psi = 0$ the same way as in Section 6.4.1 we would ultimately require measurements that are more difficult to take. The APD restitution curve for this system is shown in panel B of Figure 6.10. For a constant pacing rate P that does not produce alternans, the system will settle to a state for which the action potentials are constant on a beat to beat basis which can be found by solving the equation $P = \psi + f(\psi) = \psi^* + \text{APD}^*$ [28] where $\text{APD}^* = f(\psi^*)$. For a given voltage perturbation, using Taylor expansion, the effective change in isostable can be determined by measuring the next action potential duration according to

$$\frac{\Delta\text{APD}}{f'(\psi^*)} = \Delta\psi + \mathcal{O}(\Delta\psi^2), \tag{6.56}$$

where $\Delta\text{APD} = \text{APD}^+ - \text{APD}^*$, with APD^+ being the duration of the action potential following the perturbation (see panel A of Figure 6.10).

Suppose instead of a ring, we have a fiber (6.29) with no-flux boundary conditions. With multiple recording nodes and a single stimulating electrode the following protocol can be used to experimentally determine the iIRC: 1) At one end of the fiber, pace at a constant rate P chosen so that alternans do not develop in the system. 2) After the APDs are close to the steady state value APD^*, apply a short pulsatile perturbation at time t_p with strength u and duration Δt with the stimulating electrode. 3) For each recording node at a distance d from the perturbing electrode, let $\psi_p \equiv t_p - t_s - \text{APD}^*$ with t_s defined to be the time at which the recorded cell spiked before the perturbation was applied. The value of $\Delta\psi$ can be determined from (6.56) by measuring the duration of the next action potential. 4) at each recording electrode obtain a datapoint $\mathcal{I}_V(d, \psi_p) = \frac{\Delta\psi}{u\Delta t}$. 5) Repeat steps 2-4 until a $\mathcal{I}_V(r, \psi)$ can be obtained with a fit to the data. Panel C of Figure 6.10 shows a polynomial fit to the data obtained using the above procedure with white noise of intensity 0.2 added to the transmembrane voltage equation (6.29) to model

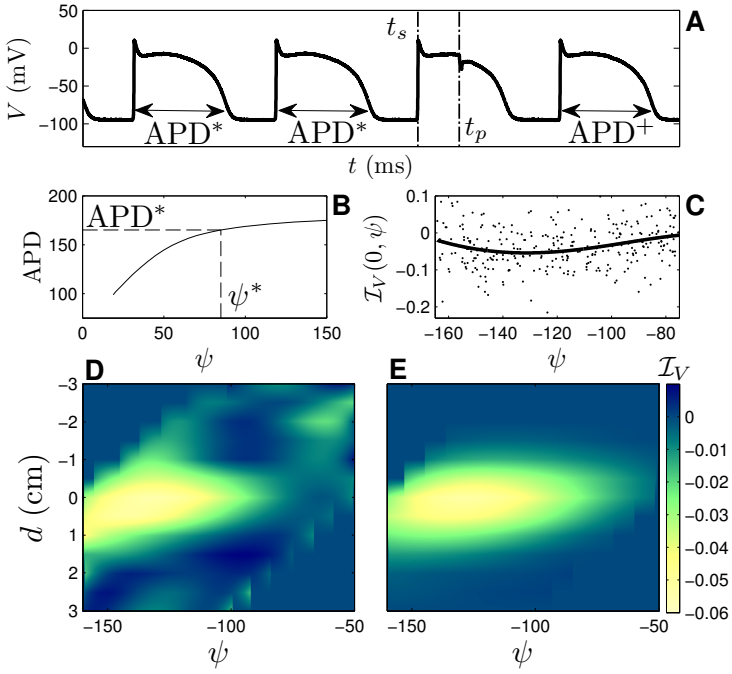

Figure 6.10: In panel A, once the APDs reach their steady state value, a perturbation is given. APD$^+$ is measured as on the following action potential. Panel B shows the APD restitution curve for this system using an S1-S2 pacing protocol [159]. Panel C shows individual datapoints for the iIRC obtained from (6.56) fitted to a polynomial (black line). Panels D and E show the resulting iIRCs using data from a noisy and noiseless system, respectively.

system noise and measurement error. Panel D shows an iIRC interpolated using data from recording nodes spaced 0.5 cm apart. Panel E shows an iIRC determined without noise in the system. In both cases the obtained iIRC is similar to the one obtained with the adjoint equation (6.9), but is less concentrated around $d = 0$. Because we assumed that we only have one stimulating electrode, perturbation timing was chosen to give the best measurement of the system near $d = 0$ and no data points in the top-left and bottom-right corners were obtained.

Using the iIRCs obtained from the noisy and noiseless data, we can apply the control strategy (6.53) to eliminate alternans. In this case, because the definition of $\psi = 0$ has changed, we define $t = 0$ at each stimulator as the time at which $\psi \approx -130$ during a long

action potential, estimated by assuming that when a cell fires a long action potential, $\psi = -\text{APD}_L^-$, where APD_L^- is the duration of the previous long action potential. The stimuli obtained from the noiseless (resp. noisy) data are shown in the top left (resp. bottom left) panel of Figure 6.11 with $\zeta = 1.3$. They are qualitatively similar to the stimuli obtained using the numerically obtained iIRC from Figure 6.6. In an identical numerical test as the one shown in Figure 6.9, each applied control strategy quickly eliminates alternans in the system. Using the stimuli associated with the iIRC obtained using the a direct method requires slightly less power than the stimuli associated with the iIRC calculated numerically using (6.9). An iIRC obtained from a noisy system using the direct method only leads to a slight increase in overall power usage.

In an experimental setting iIRCs obtained from the direct method give the change to APDs themselves in response to a perturbation. When the iIRCs are obtained numerically using (6.9), they give the change to the approach along the slowest direction of the stable manifold from a perturbation. Because the control strategy itself uses the APDs as an indicator of the severity of alternans, this may explain why experimental iIRCs produce stimuli which are slightly more efficient. As a final note, measurement of the iIRC using the direct method requires knowledge of the APD restitution curve, which may not be available in an experimental setting. If this is the case, simply assuming $f'(\psi^*) = 1$ in equation (6.56) will yield stimuli for which the shape is correct but the magnitude is potentially wrong. The appropriate stimulus could be found, for instance, by applying the resulting control and modifying the magnitude until alternans are eliminated.

6.5 Discussion

State reduction can be a powerful strategy for both understanding and controlling systems with complicated and high dimensional dynamics. Here, we have developed a

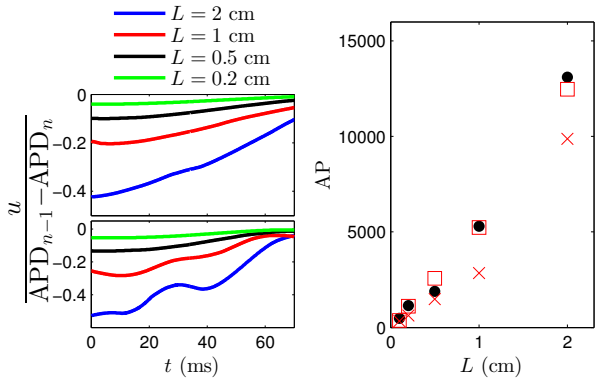

Figure 6.11: Top-left and bottom-left panels show the optimal stimuli obtained from iIRCs using the direct method from a noiseless and noisy system (6.29), respectively. The average power required to maintain the alternans free state for the stimuli obtained from an iIRC with noisy data (red squares), noiseless data (red $\times$), and using (6.9) (black dot) are shown in the right panel of Figure 6.11 for different node spacings L.

methodology of state reduction for PDEs which tracks the time it takes for an initial condition to reach a stable stationary solution in the presence of external inputs. To illustrate both the power and flexibility of this method, we have applied it to problems from two different fields of study.

In the first application, isostable reduction allows us to design an input to desynchronize a large population of phase oscillators. This control strategy could have potential applications to Parkinson's disease, where excessive synchrony in a network of neurons is thought to contribute to the motor symptoms of the disease. While this control strategy represents a significant improvement over previous control strategies, it will certainly need to be developed further before it can be applied *in vivo*. For instance, the control strategy (6.26) requires simultaneous monitoring and stimulation of the network dynamics, an issue that will need to be addressed in practice. While increased synchronization is not the only pathological change associated with Parkinson's disease, and desynchronization may not alleviate all symptoms, control strategies based on mitigating the excessive synchrony associated with Parkinson's disease have shown promise in both human and

primate studies [160], [161].

In a second application, we were able to use isostable reduction to design a control strategy to eliminate alternans in model of cardiac tissue. Because alternans can be a precursor to cardiac fibrillation, eliminating them could prevent fibrillation for those who are at risk for this life threatening arrhythmia. Here we have found that this new control strategy eliminates alternans using less energy than a previously proposed methodology [31]. While we applied this control strategy to a model of a one-dimensional ring of cardiac tissue, the theory presented in Section 6.2 readily extends to two and three dimensional models.

The ability to determine iIRCs is essential to the control strategy for eliminating alternans, a calculation which we have demonstrated by using the adjoint equation (6.9) in a numerical model. However, in live tissue experiments it is likely that we would not have access the full dynamic equations required to use the methods from Section 6.2. We present a protocol akin to the direct method for calculating phase response curves [24], [25] by which an iIRC could be measured *in vitro*. In this chapter, we have assumed the ability to directly inject current at a few control nodes, but the iIRC measurement strategy would be equally applicable for other types of perturbations. Numerically, we find that iIRCs measured directly are similar to those calculated from (6.9), yielding control stimuli which terminate alternans effectively. This direct method could be adapted on a per application basis to measure iIRCs in other biological systems.

Isostable reduction is not the only reduction strategy that can be applied to PDEs. This method of isostable reduction is perhaps most similar to reduction through the use of inertial manifolds [147–150] which attract solutions of PDEs exponentially quickly. This reduction strategy may be of use for these examples (particularly for the cardiac ring example in Section 6.4), however, inertial manifolds can be difficult to approximate, and it is not guaranteed that a resulting inertial manifold would be low dimensional. Fur-

thermore, once an inertial manifold is found, further analysis would need to be performed in order to understand how small perturbations affect the dynamics. The isostable reduction strategy described in this chapter does not require the state dynamics to collapse to a one-dimensional manifold. This point is illustrated in the cardiac ring example, because when we calculate the iIRC shown in Figure 6.6, the state variables differ on a beat-to-beat basis, but the resulting isostable response curves show little variability between beats.

Phase reduction has played an important role in the understanding of nonlinear oscillatory dynamics in the past half century, and we imagine that isostable reduction could play a similarly important role in the understanding of nonlinear PDEs by allowing for the study of complicated systems in a more useful coordinate system. Isostable reduction allowed us to make progress on two unrelated but important problems in the fields of neuroscience and cardiology, and we imagine that it could be a useful reduction strategy in other contexts, especially when the rate and timing of a system's approach towards a stationary solution are of interest.

Chapter 7

Conclusions

Phase reduction strategies have been indispensable in the study of nonlinear limit cycle oscillators. In this dissertation we have investigated applications of phase reduction to periodically spiking neural models in order to suggest new strategies for deep brain stimulation in the treatment of Parkinson's disease and to derive optimal control techniques for entraining a noisy population of neurons. We also develop a related strategy that can be applied to dynamical systems with stable fixed points in order to understand their dynamics in terms of their approach to a fixed point. We showed that this strategy was particularly useful in the study of cardiological alternans, allowing us to devise energy optimal control strategies to eliminate this arrhythmia.

In Chapter 2 we investigated a strategy to desynchronize a population of pathologically synchronized neural oscillators. In this chapter, we designed external stimuli which balance the trade off between overall energy usage and the magnitude of a positive Lyapunov exponent. For many different neural models, this strategy was shown to use orders of magnitude less energy than other strategies for desynchronizing a population of coupled neurons [44], [21]. This work was later extended in [11] to include extracellular inputs. The control strategies considered here are most effective when the oscillators are

close in phase, and do not guarantee that a population of oscillators will approach a uniformly distributed state. One approach for improving these strategies could be to design external perturbations to stabilize the steady state distribution of the underlying Fokker-Planck equation [70] describing the probability density (c.f. [88], [48]) of a population of neurons. Recent work in collabroation with researchers at the University of Minnesota and West Virginia University [162] has shown that finite time Lyapunov exponents can be used to accurately predict desynchronization in a large network of chemical oscillators. Recent experimental work [163] suggests that phase reduction could be applied to the entire Parkinsonian network, and it would be interesting to investigate whether designing stimuli which optimize the trade-off between power usage and magnitude of Lyapunov exponents could be used in experimental trials to mitigate the symptoms of Parkinson's disease.

In Chapter 3, we used phase reduction to design external stimuli which could efficiently entrain a noisy population of limit cycle oscillators. *In vitro* experiments were performed in collaboration with the Netoff lab at the University of Minnesota. The work from this chapter is an extension of previous work [91] which investigated the problem of entraining a population of noiseless limit cycle oscillators with applications to circadian rhythms. The strategies presented in this chapter could be adapted for use with models of glucose secretion in pancreatic cells [78], [79] as well as in animal models of hearing loss [84], [85]. Future work will be devoted to understanding the emergence of synchronization in heterogeneous networks of coupled oscillators.

Chapter 4 considers the problem of inferring PRCs in a population of limit cycle oscillators when only aggregate population data is available. We propose a methodology which works well for measuring the PRCs of both oscillatory and bursting neurons. Current measurement strategies (e.g. patch clamp and dynamic clamp techniques [25]) are certainly of great practical use, aiding in the understanding of the nonlinear dynamics

of neural networks, but ultimately kill the neurons being measured in the process. It would be of interest to extend the ideas presented in this chapter to measure PRC data from local field potential data from neural recordings, an advance which could make control methodologies based on phase reduction techniques more amenable for experimental testing.

Chapter 5 develops a method for isostable reduction for systems of ODEs which approach a fixed point. This method of reduction is analogous analogous to phase reduction for systems which posses a limit cycle orbit. Chapter 6 extends this framework for use in PDEs. This strategy is particularly useful for control problems where a dynamical system's timing on its approach to a stationary solution is of importance. In these chapters, we considered the problem of eliminating alternans in ODE and PDE models of cardiac function and developed control strategies which do so using signiflcantly less energy than other recently proposed control strategies. Isostable reduction has many other potential applications and could be used to study waves of spreading depression which have been associated with aura in migranes [121], [122]. Such reduction strategies could also be applied to models of traffic flow with the goal of better understanding how congestion can be reduced.

While phase reduction was originally developed within the context of ODE models of dynamical systems, recent theoretical work has examined the possibility of phase reduction in PDEs describing excitable systems [12] which could be useful in the context of problems in cardiology. When compared to ODE models, PDE models of biological dynamics are more complicated with regard to both the underlying theoretical considerations as well to the overall computational effort for the associated simulations. Simulations on graphics processors are a relatively inexpensive way to speed up these calculations [164] and will be a useful tool as computational complexity of problems continues to escalate.

Finally, this dissertation spends much effort on the on the problem of alternans, but isostable reduction strategies could also be useful in the development of low energy methods to eliminate the spiral waves which are associated with cardiac arrest [2], [28], a leading cause of death in industrialized nations. Viewing this problem in terms of isostables has proven to be effective in solving related control problems [165], [166], and continued research from this perspective could lead to future insight on new and better control strategies for low energy defibrillation.

Appendix A

Phase Models

Consider the ordinary differential equation

$$\dot{\mathbf{x}} = F(\mathbf{x}) + \epsilon G(\mathbf{x}, t), \quad \mathbf{x} \in \mathbb{R}^N, \tag{A.1}$$

where $\mathbf{x} \in \mathbb{R}^N$ is a state vector and $G \in \mathbb{R}^N$ captures the effect from an external perturbation. We will assume that the system (A.1) has a stable limit cycle γ with basin of attraction $\mathcal{B}_\gamma$. In certain instances, it is be useful to analyze (A.1) in reference to its location along the limit cycle γ. To this end, we define a scalar phase variable $\theta \in [0, 2\pi)$ such that when $G(\mathbf{x}, t) \equiv \mathbf{0}$ and $\mathbf{x}$ is on the limit cycle,

$$\dot{\theta} = \omega. \tag{A.2}$$

With this definition, $\omega = 2\pi/T$, where T is the unforced period of oscillation of (A.1). We can extend the notion of phase to $\mathcal{B}_\gamma$ with the concept of isochrons. We define isochrons in the following way: Let $\mathbf{x}(0)$ correspond to a location on γ. Then, the isochron associated

with $\mathbf{x}(0)$ is the set of all initial conditions $\mathbf{y}(0)$ such that

$$\lim_{t\to\infty} ||\mathbf{x}(t) - \mathbf{y}(t)|| = 0. \tag{A.3}$$

By definition, all initial conditions on the same isochron approach the periodic orbit with the same phase. Here, $\theta(\mathbf{x})$ gives the scalar isostable associated with the location $\mathbf{x} \in \mathcal{B}_\gamma$.

As a concrete example, we consider a two dimensional reduction of the Hodgkin-Huxley (HH) equations [68] that reproduces the essential characteristics of the neuron's dynamical behavior, [154], c.f. [155]:

$$\begin{aligned} \dot{V}_j &= f_V(V_j, n_j) + I + u(t), \\ \dot{n}_j &= f_n(V_j, n_j). \end{aligned} \tag{A.4}$$

Here, f_V and f_n are functions which describe the intrinsic neural dynamics, with all terms and parameters given in [155], $I = 10\ \mu\text{A/cm}^2$ is a baseline current given so that the neuron is in a periodically spiking regime, and $u(t)$ is an external perturbation. When we set $u(t) \equiv 0$, the system approaches the limit cycle shown in grey in Figure A.1. Black lines show isochrons for this system equally spaced in time, i.e. the time required to go from one of the displayed isochrons to the next is always the same. Two trajectories shown as red, dashed lines start on the same isochron and approach the periodic orbit with the same phase. Snapshots are shown as white dots. Generally, it is computationally intensive to calculate isochrons within the entire basin of attraction of the limit cycle (see [126]).

Starting with (A.1), it can be useful to change variables to understand the phase

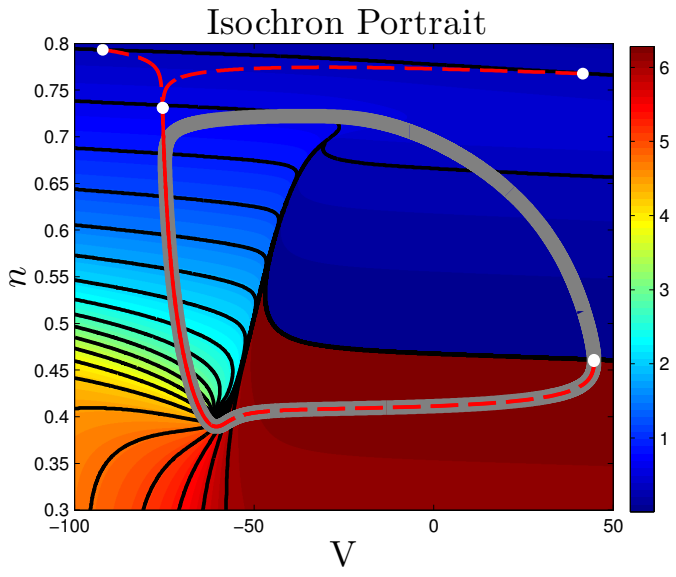

Figure A.1: Isochrons for (A.4). Two trajectories with initial conditions that start on the same isochron approach the periodic orbit (grey line) with the same phase.

evolution of the system. Using the chain rule, we write

$$\begin{aligned}\frac{d\theta(\mathbf{x})}{dt} &= \nabla\theta(\mathbf{x}) \cdot \frac{d\mathbf{x}}{dt} \\ &= \nabla\theta(\mathbf{x}) \cdot (F(\mathbf{x}) + \epsilon G(\mathbf{x}, t)) \\ &= \omega + \epsilon \nabla\theta(\mathbf{x}) G(\mathbf{x}, t). \end{aligned} \tag{A.5}$$

Here, $\nabla \equiv d/d\mathbf{x}$. In the last line, we use the fact that by construction $\frac{d\theta}{d\mathbf{x}} \cdot F(\mathbf{x}) = \omega$ when $G(\mathbf{x}, t)) \equiv 0$. Here, $\nabla\theta(\mathbf{x})$ is the spatial derivative of the isochron field, also called that phase response curve (PRC).

If we restrict our attention locations near the orbit γ, we can efficiently calculate the PRC using an "adjoint method" (c.f. [5], [54]). Evaluating the vector field at $\mathbf{x}^\gamma(\theta)$, which we define as the intersection of the trajectory γ and the $\theta(\mathbf{x})$ level set (i.e. isochron), we have

$$\frac{d\theta(\mathbf{x})}{dt} = \omega + \epsilon \nabla\theta(\mathbf{x}^\gamma(\theta)) \cdot G(\mathbf{x}^\gamma(\theta), t), \tag{A.6}$$

where we have dropped an error term of order $(|\epsilon G|^2)$ (c.f. [7]). Suppose that $G = \mathbf{0}$ for all $t > 0$ and consider an infinitesimal perturbation $\Delta\mathbf{x}$ to the trajectory $\mathbf{x}(t) \in \gamma$ at time $t = 0$. Let $\mathbf{x}_\epsilon(t) = \mathbf{x}^\gamma(t) + \Delta\mathbf{x}(t)$ be the trajectory starting at the perturbed initial condition. Then

$$\frac{d\Delta\mathbf{x}(t)}{dt} = DF(\mathbf{x}(t))\Delta\mathbf{x}(t) + \mathcal{O}(||\Delta\mathbf{x}||^2), \tag{A.7}$$

where DF is the Jacobian matrix. Furthermore, for the associated isochron shift defined as $\Delta\theta = \theta(\mathbf{x}_\epsilon(t)) - \theta(\mathbf{x}(t))$, we have

$$\Delta\theta = (\nabla_{\mathbf{x}(t)}\theta)^T \cdot \Delta\mathbf{x}(t) + \mathcal{O}(||\Delta\mathbf{x}||^2), \tag{A.8}$$

where $\nabla_{\mathbf{x}(t)}\theta$ is the gradient of θ evaluated at $\mathbf{x}(t)$. Let $\langle\cdot,\cdot\rangle$ denote the standard Euclidean inner product on $\mathbb{R}^n$. Following the derivation presented in [54], after the initial perturbation at $t = 0$, $\Delta\theta$ is independent of time. Therefore, taking the time derivative of (A.8) yields, to lowest order in $||\Delta\mathbf{x}||$,

$$\begin{aligned}\left\langle \frac{d\nabla_{\mathbf{x}(t)}\theta}{dt}, \Delta\mathbf{x}(t) \right\rangle &= -\left\langle \nabla_{\mathbf{x}(t)}\theta, \frac{d\Delta\mathbf{x}(t)}{dt} \right\rangle, \\ &= -\langle \nabla_{\mathbf{x}(t)}\theta, DF(\mathbf{x}(t))\Delta\mathbf{x}(t)\rangle, \\ &= -\langle DF^T(\mathbf{x}(t))\nabla_{\mathbf{x}(t)}\theta, \Delta\mathbf{x}(t)\rangle. \end{aligned} \tag{A.9}$$

The matrix $DF^T(\mathbf{x}(t))$ is the transpose, or adjoint, of the real-valued matrix $DF(\mathbf{x}(t))$. Equation (A.9) holds for arbitrary perturbations $\Delta\mathbf{x}(t)$, which gives

$$\frac{d\nabla_{\mathbf{x}(t)}\theta}{dt} = -DF^T(\mathbf{x}(t))\nabla_{\mathbf{x}(t)}\theta. \tag{A.10}$$

The PRC is the periodic solution of (A.10) subject to the constraint

$$\frac{d\theta}{d\mathbf{x}} \cdot F(\mathbf{x}) = \omega. \tag{A.11}$$

Practically, this solution can be obtained numerically by integrating (A.10) backwards in time with an arbitrary initial condition and periodically renormalizing so that the solution satisfies (A.11). Using (A.10), we can calculate the phase response curves of the neural oscillator (A.4).

Appendix B

The Direct Method for Measuring Phase Response Curves in Neurons

Consider the dynamical equations of a periodically spiking neuron

$$\begin{aligned} \dot{V} &= I_{\text{ion}}(V, \mathbf{n}) + \epsilon u(t) \\ \dot{\mathbf{n}} &= F(V, \mathbf{n}) \end{aligned} \tag{B.1}$$

Here, V represents the transmembrane voltage of the neuron, $\mathbf{n} \in \mathbb{R}^m$ denotes a collection of ancillary variables (e.g. chemical concentrations, gating variables) with dynamics given by F which determine the ionic current I_{ion} flowing into and out of the cell membrane, $u(t)$ represents an external voltage perturbation, and $0 < \epsilon \ll 1$. Following the phase reduction methodology from Appendix A, equation (B.1) can be phase reduced to the following form

$$\dot{\theta} = \omega + \epsilon Z(\theta) u(t), \tag{B.2}$$

where $\theta \in [0, 2\pi)$ and $\omega = 2\pi/T$ represent the phase and natural frequency of the oscillation, respectively, T is the period of oscillation, and $Z(\theta)$ represents the neuron's

PRC to voltage perturbations. We define $\theta = 0$ to represent the moment the neuron fires an action potential. If the dynamics (B.1) are known *a priori*, the phase response curve can be calculated numerically with the adjoint equation (A.10), a task which has been rendered nearly trivial with modern computing packages [65], [106], [107].

When the full dynamical equations are not known, a direct method [24], [108], [25] can be used to determine the PRC. To use the direct method, at a given instant in time, $t = 0$ is defined as the time at which the neuron spiked last. A measurement of the PRC can be obtained by using $\epsilon u(t)$ to approximate a δ-function pulse $\epsilon\beta_p\delta(T-\tau)$ where β_p is a positive constant. By comparing the expected time of the next spike $T_{\text{ex}} \equiv T$ with the actual measured time of spiking, T_{sp}, a measurement of the PRC can be obtained using

$$Z(\theta(\tau)) = Z(\omega\tau) = \frac{2\pi(T - T_{\text{sp}})}{T\beta_p}. \tag{B.3}$$

Equation (B.3) provides a point estimate of the PRC, and this process must be repeated multiple times over many phases to estimate the full PRC.

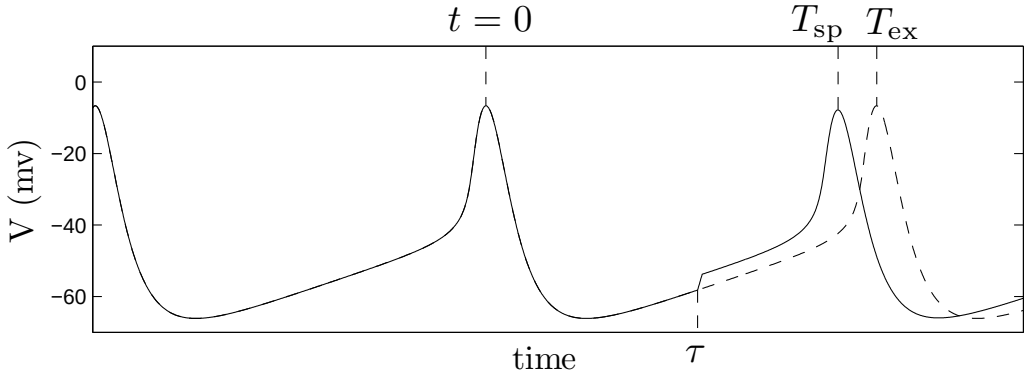

Figure B.1: An illustration of the direct method: a pulsatile perturbation is given at a time τ. By comparing the expected time of the next spike, T_{ex}, to the measured spike time, T_{sp}, a measurement of $Z(\omega\tau)$ can be obtained using (B.3)

In most practical applications, a nonnegligible amount of noise will be present in the system. With noise, T_{sp} can be viewed as a random variable which depends on the

particular realization of the noise process, and hence, the measurement $Z(\omega\tau)$ will also be a random variable. In the case that the noise Gaussian with relatively small intensity, one can show that (B.3) is an unbiased estimator of the true PRC [93]. For this reason, in experimental systems many measurements of the PRC are usually obtained and the PRC can be approximated by fitting a curve to the data.

Appendix C

Neural Networks and the Isostable Reduction of Their Associated Fokker-Planck Equation

Consider a population of neurons represented by a set of noisy limit cycle oscillators

$$\dot{\theta}_j = \omega + Z(\theta_j)\left(u(t) + \frac{1}{N}\sum_{i=1}^{N}\sigma(V(\theta_i) - V(\theta_j)) + \sqrt{2S}\eta_j(t)\right), \qquad j = 1, \ldots, N. \quad \text{(C.1)}$$

Here, $\theta \in [0, 2\pi)$ is the 2π-periodic phase of the neuron, $\omega = 2\pi/T$ represents the neuron's baseline frequency and is determined from its natural period T, $\sqrt{2S}\eta_j(t)$ is i.i.d zero mean white noise with intensity $2S$, $u(t) = I(t)/C$ with $I(t)$ being a common external current control input and $C = 1\mu F/\text{cm}^2$ the constant neural membrane capacitance, and $V(\theta)$ gives the transmembrane voltage as a function of the phase. We assume S is small enough that higher order noise terms are negligible [92]. Also, we assume that the coupling in this network is electrotonic, but this could be generalized to include, for example, chemical synaptic coupling in the network. The coupling is all-to-all with a

strength determined by σ. For a large population, it is useful to track the probability density of neurons with phase θ at a given time, $\rho(\theta, t)$. In this case, the coupling in (C.1) can be written as $\sigma(\overline{V} - V(\theta_j))$, where $\overline{V} = \int_0^{2\pi} V(\theta)\rho(\theta, t)d\theta$, rather than as a finite sum. In the limit that noise perturbations are small, through stochastic averaging, [90], we can approximate (C.1) as

$$\dot{\theta}_j = \omega + Z(\theta_j)(u(t) + \sigma(\overline{V} - V(\theta_j))) + B\eta_j(t), \tag{C.2}$$

where $B^2 = \frac{2S}{2\pi}\int_0^{2\pi} Z^2(\theta)d\theta$. For a population of neurons described by (C.2), and each receiving an identical input $u(t)$, the probability density evolves according to [70]:

$$\begin{aligned}\frac{\partial \rho(\theta,t)}{\partial t} &= -\frac{\partial}{\partial \theta}[(\omega + Z(\theta)(u(t) + \sigma(\overline{V} - V(\theta)))\rho(\theta,t)] + \frac{1}{2}\frac{\partial^2}{\partial \theta^2}[B^2\rho(\theta,t)] \\ &= -\omega\rho_\theta + \frac{B^2}{2}\rho_{\theta\theta} \quad \underbrace{-[Z(\theta)\rho_\theta + Z'(\theta)\rho]u(t)}_{\text{external input}} \quad \underbrace{-\frac{\partial}{\partial\theta}[Z(\theta)\sigma(\overline{V} - V(\theta))\rho]}_{\text{intrinsic coupling}}.\end{aligned} \tag{C.3}$$

For the moment, we will analyze (C.3) in the absence of external input or intrinsic coupling in order to perform an isostable reduction on the system, where the stable stationary solution is $\rho(\theta) = 1/(2\pi) = \overline{\rho}$. Without these terms, we have a linear advection-diffusion equation on a ring:

$$\rho_t = M\rho, \tag{C.4}$$

where $M \equiv -\omega\frac{\partial}{\partial\theta} + \frac{B^2}{2}\frac{\partial^2}{\partial\theta^2}$. The eigenfunctions and eigenvalues of (C.4) are

$$\phi_{2n-1,2n} = e^{\pm in\theta}, \qquad \lambda_{2n-1,2n} = -\frac{B^2n^2}{2} \mp \omega ni. \tag{C.5}$$

The infinite time approach to the origin will be governed by the eigenfunctions associated with eigenvalues with the smallest magnitude real part,

$$\rho(\theta,t)-\bar{\rho}=\sum_{n=1}^{\infty}\left[s_{2n-1}e^{in\theta}e^{(-\frac{B^2n^2}{2}-\omega ni)t}+s_{2n}e^{-in\theta}e^{(-\frac{B^2n^2}{2}+\omega ni)t}\right], \tag{C.6}$$

where the constants $\{s_1, s_2, \dots\}$ are determined by the initial distribution $\rho(\theta, 0)$. Equation (C.6) can be rewritten using Euler's formula and the knowledge that the solution is strictly real as

$$\begin{aligned}\rho(\theta,t)-\bar{\rho}&=\sum_{n=1}^{\infty}\left[(a_n\sin(n(\theta-\omega t))+b_n\cos(n(\theta-\omega t)))e^{-\frac{B^2n^2}{2}t}\right]\\&=\sum_{n=1}^{\infty}\left[A_n\cos(n(\theta-\omega t)-\varphi_n)e^{-\frac{B^2n^2}{2}t}\right],\end{aligned} \tag{C.7}$$

where a_n and $b_n \in \mathbb{R}$, $\tan(\varphi_n) = a_n/b_n$ and $A_n = \sqrt{a_n^2 + b_n^2}$, and the second line of (C.7) is obtained through trigonometric identities. Equation (C.7) mandates that the infinite time approach to the stationary solution will be governed by the magnitude of the first mode of the Fourier expansion of the initial distribution $\rho(\theta, 0)$. Here, because λ_1 is complex, the infinite time approach to the stationary solution is governed by two linearly independent eigenfunctions, $\sin(\theta)$ and $\cos(\theta)$ which spiral towards the stationary solution. A perturbation in the direction $-\sin(\theta - \omega t - \varphi_1)$ perturbs in the direction of the spiral and will not change the isostable on which the function lies. Conversely, perturbations in the orthogonal direction, $\cos(\theta - \omega t - \varphi_1)$, represent the gradient of the isostable field. Examples of the effect of these types of perturbations are shown in Panels C-G of Figure 6.1. With this information and using equation (6.3), by arbitrarily

defining $\psi = 0$ to correspond to time t^*, we have the initial condition

$$\mathcal{I}(\theta, 0) = -\frac{2\kappa}{A_1^* B^2 \pi} \cos(\theta - \varphi_1^*), \tag{C.8}$$

where $\mathcal{I}(\theta, \psi) \equiv \nabla\psi$ is the isostable response curve and $A_1^* \cos(\theta - \varphi_1^*)$ is the first Fourier mode of the distribution $\rho(\theta, t^*)$. The Jacobian is identical to the operator M for the linear system. Therefore, according to (6.9), the isostable response curve along an unperturbed trajectory, γ, will change in time according to

$$\frac{d\mathcal{I}}{dt} = -M^\dagger \mathcal{I} = -\omega \mathcal{I}_\theta - \frac{B^2}{2} \mathcal{I}_{\theta\theta}, \tag{C.9}$$

where $M^\dagger = \omega \frac{\partial}{\partial\theta} + \frac{B}{2} \frac{\partial^2}{\partial\theta^2}$ is the adjoint of the operator M. Using (C.9) and (C.8) we can explicitly solve for the iIRC

$$\mathcal{I}(\theta, \psi) = \frac{-2\kappa}{A_1^* B^2 \pi} \cos\left(\theta - \varphi_1^* - \frac{\omega\psi}{\kappa}\right) e^{\frac{B^2 \psi}{2\kappa}}. \tag{C.10}$$

Equation (C.10) is valid provided that an external input is small enough so that it does not drive the distribution far from γ, the expected trajectory towards the stationary solution, which in this example is given by equation (C.7). If the input does drive the system far from γ, the isostable response curve at a given time can be found from the distribution $\rho(\theta, t)$ subject to (C.8).

Using the isostable response curve, we can write an isostable reduction of (C.3) as

$$\dot{\psi} = \kappa + \left\langle \mathcal{I}(\theta, \psi), -[Z(\theta)\rho_\theta + Z'(\theta)\rho]u(t) - \frac{\partial}{\partial\theta}[Z(\theta)\sigma(\overline{V} - V(\theta))\rho] \right\rangle. \tag{C.11}$$

Bibliography

[1] A. T. Winfree. Patterns of phase compromise in biological cycles. *Journal of Mathematical Biology*, 1(1):73–93, 1974.

[2] A. Winfree. *The Geometry of Biological Time*. Springer Verlag, New York, second edition, 2001.

[3] J. Guckenheimer. Isochrons and phaseless sets. *Journal of Mathematical Biology*, 1(3):259–273, 1975.

[4] Y. Kuramoto. *Chemical Oscillations, Waves, and Turbulence*. Springer-Verlag, Berlin, 1984.

[5] G. B. Ermentrout and N. Kopell. Multiple pulse interactions and averaging in systems of coupled neural oscillators. *Journal of Mathematical Biology*, 29(3):195–217, 1991.

[6] F. C. Hoppensteadt and E. M. Izhikevich. *Weakly Connected Neural Networks*. Springer, 1997.

[7] Y. Kuramoto. Phase-and center-manifold reductions for large populations of coupled oscillators with application to non-locally coupled systems. *International Journal of Bifurcation and Chaos*, 7(04):789–805, 1997.

[8] E. Brown, P. Holmes, and J. Moehlis. Globally coupled oscillator networks. In *Perspectives and Problems in Nonlinear Science*, pages 183–215. Springer, 2003.

[9] J. Moehlis, E. Shea-Brown, and H. Rabitz. Optimal inputs for phase models of spiking neurons. *ASME Journal of Computational and Nonlinear Dynamics*, 1(4):358–367, 2006.

[10] D. Wilson and J. Moehlis. Optimal chaotic desynchronization for neural populations. *SIAM Journal on Applied Dynamical Systems*, 13(1):276–305, 2014.

[11] D. Wilson and J. Moehlis. Locally optimal extracellular stimulation for chaotic desynchronization of neural populations. *Journal of Computational Neuroscience*, 37(2):243–257, 2014.

[12] H. Nakao, T. Yanagita, and Y. Kawamura. Phase-reduction approach to synchronization of spatiotemporal rhythms in reaction-diffusion systems. *Physical Review X*, 4(2):021032, 2014.

[13] D. Kirk. *Optimal Control Theory*. Dover Publications, New York, 1998.

[14] A. Nini, A. Feingold, H. Slovin, and H. Bergman. Neurons in the global pallidus do not show correlated activity in the normal monkey, but phased locked oscillations appear in the MPTP model of Parkinsonism. *Journal of Neurophysiology*, 74(4):1800–1805, 1995.

[15] P. Brown and C. D. Marsden. What do the basal ganglia do? *The Lancet*, 351(9118):1801–1804, 1998.

[16] R. Levy, W. Hutchison, A. Lozano, and J. Dostrovsky. High-frequency synchronization of neuronal activity in the subthalamic nucleus of Parkinsonian patients with limb tremor. *The Journal of Neuroscience*, 20(20):7766–7775, 2000.

[17] C. Chen, V. Litvak, T. Gilbertson, A. Kühn, C. S. Lu, S. T. Lee, C. H. Tsai, S. Tisch, P. Limousin, M. Hariz, and P. Brown. Excessive synchronization of basal ganglia neurons at 20 Hz slows movement in Parkinson's disease. *Experimental Neurology*, 205(1):214–221, 2007.

[18] A. Pogosyan, F. Yoshida, C. C. Chen, I. Martinez-Torres, T. Foltynie, P. Limousin, L. Zrinzo, M. I. Hariz, and P. Brown. Parkinsonian impairment correlates with spatially extensive subthalamic oscillatory synchronization. *Neuroscience*, 171(1):245–257, 2010.

[19] T. Wichmann, M. R. DeLong, J. Guridi, and J. A. Obeso. Milestones in research on the pathophysiology of Parkinson's disease. *Movement Disorders*, 26(6):1032–1041, 2011.

[20] P. Tass. *Phase Resetting in Medicine and Biology: Stochastic Modelling and Data Analysis*. Springer, New York, 2007.

[21] C. Wilson, B. Beverlin II, and T. Netoff. Chaotic desynchronization as the therapeutic mechanism of deep brain stimulation. *Front. Syst. Neurosci.*, 5:Art. No. 50, 2011.

[22] T. Stigen, P. Danzl, J. Moehlis, and T. Netoff. Controlling spike timing and synchrony in oscillatory neurons. *J. Neurophysiol*, 105:2074–82, 2011.

[23] A. Nabi, T. Stigen, J. Moehlis, and T. Netoff. Minimum energy control for *in vitro* neurons. *Journal of Neural Engineering*, 10(3):036005, 2013.

[24] E. M. Izhikevich. *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. MIT Press, London, 2007.

[25] T. Netoff, M. A. Schwemmer, and T. J. Lewis. Experimentally estimating phase response curves of neurons: theoretical and practical issues. In *Phase Response Curves in Neuroscience*, pages 95–129. Springer, 2012.

[26] A. Mauroy, I. Mezić, and J. Moehlis. Isostables, isochrons, and Koopman spectrum for the action–angle representation of stable fixed point dynamics. *Physica D: Nonlinear Phenomena*, 261:19–30, 2013.

[27] R. A. Gray, A. M. Pertsov, and J. Jalife. Spatial and temporal organization during cardiac fibrillation. *Nature*, 392(6671):75–78, 1998.

[28] F. Fenton, E. Cherry, H. Hastings, and S. Evans. Multiple mechanisms of spiral wave breakup in a model of cardiac electrical activity. *Chaos*, 12:852–892, 2002.

[29] E. M. Cherry and F. H. Fenton. Visualization of spiral and scroll waves in simulated and experimental cardiac tissue. *New Journal of Physics*, 10(12):125016, 2008.

[30] D. S. Rosenbaum, L. E. Jackson, J. M. Smith, H. Garan, J. N. Ruskin, and R. J. Cohen. Electrical alternans and vulnerability to ventricular arrhythmias. *New England Journal of Medicine*, 330(4):235–241, 1994.

[31] W.J. Rappel, F. Fenton, and A. Karma. Spatiotemporal control of wave instabilities in cardiac tissue. *Physical Review Letters*, 83(2):456, 1999.

[32] D. J. Christini and J. J. Collins. Using chaos control and tracking to suppress a pathological nonchaotic rhythm in a cardiac model. *Physical Review E*, 53(1):R49, 1996.

[33] K. Hall, D. J. Christini, M. Tremblay, J. J. Collins, L. Glass, and J. Billette. Dynamic control of cardiac alternans. *Physical Review Letters*, 78(23):4518–4521, 1997.

[34] D. J. Christini, M. L. Riccio, C. A. Culianu, J. J. Fox, A. Karma, and R. F. Gilmour Jr. Control of electrical alternans in canine cardiac Purkinje fibers. *Physical Review Letters*, 96(10):104101, 2006.

[35] U. B. Kanu, S. Iravanian, R. F. Gilmour, and E. J. Christini. Control of action potential duration alternans in canine cardiac ventricular tissue. *Biomedical Engineering, IEEE Transactions on*, 58(4):894–904, 2011.

[36] D. Wilson and J. Moehlis. Determining individual phase response curves from aggregate population data. *Physical Review E*, 92(2):022902, 2015.

[37] D. Wilson, A. B. Holt, T. I. Netoff, and J. Moehlis. Optimal entrainment of heterogeneous noisy neurons. *Frontiers in Neuroscience*, 9:Art. No. 192, 2015.

[38] D. Wilson and J. Moehlis. Extending phase reduction to excitable media: Theory and applications. *SIAM Review*, 57(2):201–222, 2015.

[39] D. Wilson and J. Moehlis. Isostable reduction with applications to time dependent partial differential equations. In review at *Physical Review E*.

[40] C. Hammond, H. Bergman, and P. Brown. Pathological synchronization in Parkinson's disease: networks, models and treatments. *Trends Neurosci.*, 30:357–64, 2007.

[41] P. Tass. Desynchronizing double-pulse phase resetting and application to deep brain stimulation. *Biol. Cybern.*, 85:343–354, 2001.

[42] I. Z. Kiss, C. G. Rusin, H. Kori, and J. L. Hudson. Engineering complex dynamical structures: sequential patterns and desynchronization. *Science*, 316:1886–1889, 2007.

[43] P. Danzl, J. Hespanha, and J. Moehlis. Event-based minimum-time control of oscillatory neuron models. *Biol. Cybern.*, 101:387–399, 2009.

[44] A. Nabi, M. Mirzadeh, F. Gibou, and J. Moehlis. Minimum energy desynchronizing control for coupled neurons. *Journal of Computational Neuroscience*, 34:259–271, 2013.

[45] P. Danzl, A. Nabi, and J. Moehlis. Charge-balanced spike timing control for phase models of spiking neurons. *Discrete and Continuous Dynamical Systems*, 28:1413–1435, 2010.

[46] S. Talathi, P. Carney, and P. Khargonekar. Control of neural synchrony using channelrhodopsin-2: a computational study. *J. Comput. Neurosci.*, 31:87–103, 2011.

[47] C. Hauptmann, O. Popovych, and P. Tass. Effectively desynchronizing deep brain stimulation based on a coordinated delayed feedback stimulation via several sites: a computational study. *Biological Cybernetics*, 93(6):463–470, 2005.

[48] M. Rosenblum and A. Pikovsky. Controlling synchronization in an ensemble of globally coupled oscillators. *Physical Review Letters*, 92(11):114102, 2004.

[49] O. Popovych, C. Hauptmann, and P. Tass. Effective desynchronization by nonlinear delayed feedback. *Physical Review Letters*, 94(16):164102, 2005.

[50] N. Tukhlina, M. Rosenblum, A. Pikovsky, and J. Kurths. Feedback suppression of neural synchrony by vanishing stimulation. *Physical Review E*, 75(1):011918, 2007.

[51] K. Pyragas, O. Popovych, and P. Tass. Controlling synchrony in oscillatory networks with a separate stimulation-registration setup. *Europhysics Letters*, 80(4):40002, 2007.

[52] M. Luo, Y. Wu, and J. Peng. Washout filter aided mean field feedback desynchronization in an ensemble of globally coupled neural oscillators. *Biological Cybernetics*, 101(3):241–246, 2009.

[53] O. Popovych and P. Tass. Synchronization control of interacting oscillatory ensembles by mixed nonlinear delayed feedback. *Physical Review E*, 82(2):026204, 2010.

[54] E. Brown, J. Moehlis, and P. Holmes. On the phase reduction and response dynamics of neural oscillator populations. *Neural Computation*, 16(4):673–715, 2004.

[55] L. Iasemidis, D. Shiau, Wanpracha W. Chaovalitwongse, J. Sackellares, P. Pardalos, J. Principe, P. Carney, A. Prasad, B Veeramani, and K. Tsakalis. Adaptive epileptic seizure prediction system. *Biomedical Engineering, IEEE Transactions on*, 50(5):616–627, 2003.

[56] L. Iasemidis, D. Shiau, J. Sackellares, P. Pardalos, and A. Prasad. Dynamical resetting of the human brain at epileptic seizures: application of nonlinear dynamics and global optimization techniques. *Biomedical Engineering, IEEE Transactions on*, 51(3):493–506, 2004.

[57] P. Pardalos and V. Yatsenko. Optimization approach to the estimation and control of Lyapunov exponents. *Journal of Optimization Theory and Applications*, 128(1):29–48, 2006.

[58] A. Abouzeid and B. Ermentrout. Type-II phase resetting curve is optimal for stochastic synchrony. *Phys. Rev. E*, 80:011911, 2009.

[59] D. B. Forger and D. Paydarfar. Starting, stopping, and resetting biological oscillators: In search of optimal perturbations. *J. Theor. Biol.*, 230:521–32, 2004.

[60] D. Merill, M. Bikson, and J. Jefferys. Electrical stimulation of excitable tissue: design of efficacious and safe protocols. *J. Neurosci Methods*, 141(2):171–198, 2005.

[61] B. Pakkenberg and H. Gundersen. Neocortical neuron number in humans: Effect of sex and age. *Journal of Comparative Neurobiology*, 384:312–320, 1997.

[62] J. Ritt. *A Probabilistic Analysis of Forced Oscillators With Application to Neuronal Response Reliability*. PhD thesis, Boston University, 2003.

[63] J. Rubin and D. Terman. High frequency stimulation of the subthalamic nucleus eliminates pathological thalamic rhythmicity in a computational model. *Journal of Computational Neuroscience*, 16:211–235, 2004.

[64] D. Johnston and S. M.-S. Wu. *Foundations of Cellular Neurophysiology*. MIT Press, Cambridge, MA, 1995.

[65] G. B. Ermentrout. *Simulating, Analyzing and Animating Dynamical Systems: A Guide to XPPAUT for Researchers and Students*. SIAM, Philadelphia, 2002.

[66] R. Honeycutt. Stochastic Runge-Kutta algorithms. I. white noise. *Physical Review A*, 45:600–603, 1992.

[67] S. Schiff. *Neural Control Engineering*. MIT Press, Cambridge, MA, 2011.

[68] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.*, 117:500–44, 1952.

[69] J. Keener and J. Sneyd. *Mathematical Physiology*. Springer-Verlag, New York, 1998.

[70] C. W. Gardiner. *Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences*. Springer, Berlin, 2004.

[71] D. Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, 43:525–546, 2001.

[72] G. Deuschl et. al. A randomized trial of deep-brain stimulation for Parkinson's disease. *The New England Journal of Medicine*, 355:896–908, 2006.

[73] S. Thobois, P. Mertens, M. Guenot, M. Hermier, H. Mollion, M. Bouvard, G. Chazot, E. Broussolle, and M. Sindou. Subthalamic nucleus stimulation in Parkinson's disease clinical evaluation of 18 patients. *Journal of Neurology*, 249:529–543, 2002.

[74] D. A. Golombek and R. E. Rosenstein. Physiology of circadian entrainment. *Physiological Reviews*, 90(3):1063–1102, 2010.

[75] L. Canaple, T. Kakizawa, and V. Laudet. The days and nights of cancer cells. *Cancer Research*, 63(22):7545–7552, 2003.

[76] K. B. Klerman. Clinical aspects of human circadian rhythms. *Journal of Biological Rhythms*, 20(4):375–386, 2005.

[77] N. Takeda and K. Maemura. Circadian clock and cardiovascular disease. *Journal of Cardiology*, 57(3):249–256, 2011.

[78] R. Bertram, A. Sherman, and L. S. Satin. Metabolic and electrical oscillations: partners in controlling pulsatile insulin secretion. *American Journal of Physiology-Endocrinology And Metabolism*, 293(4):E890–E900, 2007.

[79] M. G. Pedersen, E. Mosekilde, K. S. Polonsky, and D. S. Luciani. Complex patterns of metabolic and Ca^{2+} entrainment in pancreatic islets by oscillatory glucose. *Biophysical Journal*, 105(1):29–39, 2013.

[80] P. Lakatos, G. Karmos, A. D. Mehta, I. Ulbert, and C. E. Schroeder. Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, 320(5872):110–113, 2008.

[81] M. Ainsworth, S. Lee, M. O. Cunningham, R. D. Traub, N. J. Kopell, and M. A. Whittington. Rates and rhythms: a synergistic view of frequency and temporal coding in neuronal networks. *Neuron*, 75(4):572–583, 2012.

[82] G. Buzsáki and A. Draguhn. Neuronal oscillations in cortical networks. *Science*, 304(5679):1926–1929, 2004.

[83] J. Jacobs, M. J. Kahana, A. D. Ekstrom, and I. Fried. Brain oscillations control timing of single-neuron activity in humans. *The Journal of Neuroscience*, 27(14):3839–3844, 2007.

[84] Y. Wang and P. B. Manis. Temporal coding by cochlear nucleus bushy cells in DBA/2J mice with early onset hearing loss. *Journal of the Association for Research in Otolaryngology*, 7(4):412–424, 2006.

[85] K. S. Henry and M. G. Heinz. Diminished temporal coding with sensorineural hearing loss emerges in background noise. *Nature Neuroscience*, 15(10):1362–1364, 2012.

[86] T. Harada, H. A. Tanaka, M. J. Hankins, and I. Z. Kiss. Optimal waveform for the entrainment of a weakly forced oscillator. *Physical Review Letters*, 105(8):088301, 2010.

[87] A. Zlotnik, Y. Chen, I. Z. Kiss, H.A. Tanaka, and J. S. Li. Optimal waveform for fast entrainment of weakly forced nonlinear oscillators. *Physical Review Letters*, 111(2):024102, 2013.

[88] G. B. Ermentrout and D. H. Terman. *Mathematical Foundations of Neuroscience*, volume 35. Springer, New York, 2010.

[89] H. C. Tuckwell. *Introduction to Theoretical Neurobiology: Volume 2, Nonlinear and Stochastic Theories*. Cambridge University Press, Cambridge, 2005.

[90] M. I. Freidlin and A. D. Wentzell. *Random Perturbations of Dynamical Systems*, volume 260. Springer, Berlin, 2012.

[91] D. Wilson and J. Moehlis. An energy-optimal approach for entrainment of uncertain circadian oscillators. *Biophysical Journal*, 107(7):1744–1755, 2014.

[92] C. Ly and G. B. Ermentrout. Synchronization dynamics of two coupled neural oscillators receiving shared and unshared noisy stimuli. *Journal of Computational Neuroscience*, 26(3):425–443, 2009.

[93] G. B. Ermentrout, B. Beverlin II, T. Troyer, and T. I. Netoff. The variance of phase-resetting curves. *Journal of Computational Neuroscience*, 31(2):185–197, 2011.

[94] W. Q. Zhu. Stochastic averaging methods in random vibration. *Applied Mechanics Reviews*, 41(5):189–199, 1988.

[95] O. Miranda-Dominguez, J. Gonia, and T.I. Netoff. Firing rate control of a neuron using a linear proportional-integral controller. *Journal of Neural Engineering*, 7(6):066004, 2010.

[96] O. Miranda-Dominguez and T.I. Netoff. Parameterized phase response curves for characterizing neuronal behaviors under transient conditions. *Journal of Neurophysiology*, 109(9):2306–2316, 2013.

[97] M.S. Roulston. Estimating the errors on measured entropy and mutual information. *Physica D: Nonlinear Phenomena*, 125(3–4):285–294, 1999.

[98] K. Ota, T. Omori, S. Watanabe, H. Miyakawa, M. Okada, and T. Aonishi. Measurement of infinitesimal phase response curves from noisy real neurons. *Physical Review E*, 84(4):041902, 2011.

[99] S. Wang, M. M. Musharoff, C. C. Canavier, and S. Gasparini. Hippocampal CA1 pyramidal neurons exhibit type 1 phase-response curves and type 1 excitability. *Journal of Neurophysiology*, 109(11):2757–2766, 2013.

[100] U. S. Thounaojam, J. Cui, S. E. Norman, R. J. Butera, and C. C. Canavier. Slow noise in the period of a biological oscillator underlies gradual trends and abrupt transitions in phasic relationships in hybrid neural networks. *PLoS Computational Biology*, 10(5):e1003622, 2014.

[101] G. Ullah and S. J. Schiff. Assimilating seizure dynamics. *PLoS Computational Biology*, 6(5):e1000776, 2010.

[102] S. J. Schiff. Towards model-based control of Parkinson's disease. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1918):2269–2308, 2010.

[103] P. A. Tass. Desynchronization of brain rhythms with soft phase-resetting techniques. *Biological Cybernetics*, 87(2):102–115, 2002.

[104] S. An, R. Harang, K. Meeker, D. Granados-Fuentes, C. A. Tsai, C. Mazuski, J. Kim, F. J. Doyle, L. R. Petzold, and E. D. Herzog. A neuropeptide speeds circadian entrainment by reducing intercellular synchrony. *Proceedings of the National Academy of Sciences*, 110(46):E4355–E4361, 2013.

[105] N. Bagheri, S. R. Taylor, K. Meeker, L. R. Petzold, and F. J. Doyle. Synchrony and entrainment properties of robust circadian oscillators. *Journal of The Royal Society Interface*, 5(Suppl 1):S17–S28, 2008.

[106] W. Govaerts and B. Sautois. Computation of the phase response curve: a direct numerical approach. *Neural Computation*, 18(4):817–847, 2006.

[107] A. Dhooge, W. Govaerts, and Y. A. Kuznetsov. MATCONT: a MATLAB package for numerical bifurcation analysis of ODEs. *ACM Transactions on Mathematical Software*, 29(2):141–164, 2003.

[108] R. F. Galán, G. B. Ermentrout, and N. N. Urban. Efficient estimation of phase-resetting curves in real neurons and its significance for neural-network modeling. *Physical Review Letters*, 94(15):158101, 2005.

[109] G. T. Einevoll, C. Kayser, N. K. Logothetis, and S. Panzeri. Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nature Reviews Neuroscience*, 14(11):770–785, 2013.

[110] Y. Kawamura, H. Nakao, K. Arai, H. Kori, and Y. Kuramoto. Collective phase sensitivity. *Physical Review Letters*, 101(2):024101, 2008.

[111] Tae-Wook Ko and G Bard Ermentrout. Phase-response curves of coupled oscillators. *Physical Review E*, 79(1):016211, 2009.

[112] H. Kori, Y. Kawamura, H. Nakao, K. Arai, and Y. Kuramoto. Collective-phase description of coupled oscillators with general network structure. *Physical Review E*, 80(3):036207, 2009.

[113] Z. Levnajić and A. Pikovsky. Phase resetting of collective rhythm in ensembles of oscillators. *Physical Review E*, 82(5):056202, 2010.

[114] C. D. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, Philadelphia, 2000.

[115] G. W. Stewart. On the perturbation of pseudo-inverses, projections and linear least squares problems. *SIAM Review*, 19(4):634–662, 1977.

[116] J. L. Hindmarsh and R. M. Rose. A model of neuronal bursting using three coupled first order differential equations. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 221(1222):87–102, 1984.

[117] E. Sherwood and J. Guckenheimer. Dissecting the phase response of a model bursting neuron. *SIAM Journal on Applied Dynamical Systems*, 9:659–703, 2010.

[118] J. A. Sanders, F. Verhulst, and J. Murdock. *Averaging Methods in Nonlinear Dynamical Systems*. Springer-Verlag, New York, second edition, 2007.

[119] Z. Qu, A. Garfinkel, P. Chen, and J. N. Weiss. Mechanisms of discordant alternans and induction of reentry in simulated cardiac tissue. *Circulation*, 102(14):1664–1670, 2000.

[120] M. Hildebrand, J. Cui, E. Mihaliuk, J. Wang, and K. Showalter. Synchronization of spatiotemporal patterns in locally coupled excitable media. *Physical Review E*, 68(2):026205, 2003.

[121] M. Lauritzen, J. P. Dreier, M. Fabricius, J. A. Hartings, R. Graf, and A. J. Strong. Clinical relevance of cortical spreading depression in neurological disorders: migraine, malignant stroke, subarachnoid and intracranial hemorrhage, and traumatic brain injury. *Journal of Cerebral Blood Flow & Metabolism*, 31(1):17–35, 2011.

[122] A. C. Charles and Serapio M S. M. Baca. Cortical spreading depression and migraine. *Nature Reviews Neurology*, 9(11):637–644, 2013.

[123] E. M. Izhikevich. *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. The MIT Press, Cambridge, MA, 2006.

[124] N. Ichinose, K. Aihara, and K. Judd. Extending the concept of isochrons from oscillatory to excitable systems for modeling an excitable neuron. *International Journal of Bifurcation and Chaos*, 8(12):2375–2385, 1998.

[125] A. Rabinovitch and I. Rogachevskii. Threshold, excitability and isochrones in the Bonhoeffervan der Pol system. *Chaos*, 9(4):880–886, 1999.

[126] H. M. Osinga and J. Moehlis. Continuation-based computation of global isochrons. *SIAM Journal on Applied Dynamical Systems*, 9(4):1201–1228, 2010.

[127] J. M. Pastore, S. D. Girouard, K. R. Laurita, F. G. Akar, and D. S. Rosenbaum. Mechanism linking T-wave alternans to the genesis of cardiac fibrillation. *Circulation*, 99(10):1385–1394, 1999.

[128] M. A. Watanabe, F. H. Fenton, S. J. Evans, H. M. Hastings, and A. Karma. Mechanisms for discordant alternans. *Journal of Cardiovascular Electrophysiology*, 12(2):196–206, 2001.

[129] J. N. Weiss, A. Karma, Y. Shiferaw, P. Chen, A. Garfinkel, and Z. Qu. From pulsus to pulseless: the saga of cardiac alternans. *Circulation Research*, 98(10):1244–1253, 2006.

[130] A. Gizzi, E. M. Cherry, R. F Gilmour Jr, S. Luther, S. Filipp, and F. H. Fenton. Effects of pacing site and stimulation history on alternans dynamics and the development of complex spatiotemporal patterns in cardiac tissue. *Frontiers in Physiology*, 4:Art. No. 71, 2013.

[131] E. M. Cherry, F. H. Fenton, and R. F. Gilmour Jr. Mechanisms of ventricular arrhythmias: a dynamical systems-based perspective. *American Journal of Physiology-Heart and Circulatory Physiology*, 302(12):H2451, 2012.

[132] J. J. Fox, J. L. McHarg, and R. F. Gilmour. Ionic mechanism of electrical alternans. *American Journal of Physiology-Heart and Circulatory Physiology*, 282(2):H516–H530, 2002.

[133] J. Rogers and A. McCulloch. A collocation-Galerkin finite element model of cardiac action potential propagation. *IEEE Transactions on Biomedical Engineering*, 41:743–757, 1994.

[134] J. B. Nolasco and R. W. Dahlen. A graphic method for the study of alternation in cardiac action potentials. *Journal of Applied Physiology*, 25(2):191–196, 1968.

[135] E. G. Tolkacheva, D. G. Schaeffer, D. J. Gauthier, and W. Krassowska. Condition for alternans and stability of the 1:1 response pattern in a memory model of paced cardiac dynamics. *Physical Review E*, 67(3):031904, 2003.

[136] E. M. Cherry and F. H. Fenton. Suppression of alternans and conduction blocks despite steep APD restitution: electrotonic, memory, and conduction velocity restitution effects. *American Journal of Physiology-Heart and Circulatory Physiology*, 55(6):H2332–H2341, 2004.

[137] E. M. Cherry and S. J. Evans. Properties of two human atrial cell models in tissue: restitution, memory, propagation, and reentry. *Journal of Theoretical Biology*, 254(3):674–690, 2008.

[138] A. Garfinkel, Y. H. Kim, O. Voroshilovsky, Z. Qu, J. R. Kil, M. H. Lee, H. S. Karagueuzian, J. N. Weiss, and P. S. Chen. Preventing ventricular fibrillation by flattening cardiac restitution. *Proceedings of the National Academy of Sciences*, 97(11):6061–6066, 2000.

[139] K. E. Brack, V. H. Patel, J. H. Coote, and G. A. Ng. Nitric oxide mediates the vagal protective effect on ventricular fibrillation via effects on action potential duration restitution in the rabbit heart. *The Journal of Physiology*, 583(2):695–704, 2007.

[140] A. Mahajan, D. Sato, Y. Shiferaw, A. Baher, L. H. Xie, R. Peralta, R. Olcese, A. Garfinkel, Z. Qu, and J. N. Weiss. Modifying L-type calcium current kinetics: consequences for cardiac excitation and arrhythmia dynamics. *Biophysical Journal*, 94(2):411–423, 2008.

[141] S. Dubljevic. Optimal boundary control of cardiac alternans. *International Journal of Robust and Nonlinear Control*, 19(2):135–150, 2009.

[142] A. Garzón, R. O. Grigoriev, and F. H. Fenton. Model-based control of cardiac alternans in Purkinje fibers. *Physical Review E*, 84(4):041927, 2011.

[143] S. Wiggins. *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, volume 2. Springer, New York, 2003.

[144] N. Fenichel. Asymptotic stability with rate conditions. *Indiana Univ. Math. J*, 23(1109-1137):74, 1973.

[145] S. Wiggins. *Normally Hyperbolic Invariant Manifolds in Dynamical Systems*. Number 105. Springer, New York, 1994.

[146] A. J. Roberts. Appropriate initial conditions for asymptotic descriptions of the long term evolution of dynamical systems. *The Journal of the Australian Mathematical Society. Series B. Applied Mathematics*, 31(01):48–75, 1989.

[147] P. Constantin, C. Foias, B. Nicolaenko, and R. Temam. *Integral Manifolds and Inertial Manifolds for Dissipative Partial Differential Equations*, volume 70. Springer, New York, 1989.

[148] C. Foias, G. R. Sell, and R. Temam. Inertial manifolds for nonlinear evolutionary equations. *Journal of Differential Equations*, 73(2):309–353, 1988.

[149] C. Foias, M. S. Jolly, I. G. Kevrekidis, G. R. Sell, and E. S. Titi. On the computation of inertial manifolds. *Physics Letters A*, 131(7):433–436, 1988.

[150] M. S. Jolly, I. G. Kevrekidis, and E. S. Titi. Approximate inertial manifolds for the Kuramoto-Sivashinsky equation: analysis and computations. *Physica D: Nonlinear Phenomena*, 44(1):38–60, 1990.

[151] J. P. Keener. *Principles of Applied Mathematics*. Addison-Wesley, Redwood City, California, 1988.

[152] J. Mallet-Paret and G. R. Sell. Inertial manifolds for reaction diffusion equations in higher space dimensions. *Journal of the American Mathematical Society*, 1(4):805–866, 1988.

[153] A. L. Benabid, S. Chabardes, J. Mitrofanis, and P. Pollak. Deep brain stimulation of the subthalamic nucleus for the treatment of Parkinson's disease. *The Lancet Neurology*, 8(1):67–81, 2009.

[154] J. Rinzel. Excitation dynamics: insights from simplified membrane models. In *Federation Proceedings*, volume 44, pages 2944–2946, 1985.

[155] J. Moehlis. Canards for a reduction of the Hodgkin-Huxley equations. *J. Math. Biol.*, 52:141–53, 2006.

[156] J. Guckenheimer and P. Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, volume 42. Springer Verlag, New York, 1983.

[157] M. Li and N. F. Otani. Controlling alternans in cardiac cells. *Annals of Biomedical Engineering*, 32(6):784–792, 2004.

[158] T. Krogh-Madsen, A. Karma, M. L. Riccio, P. N. Jordan, D. J. Christini, and R. F. Gilmour Jr. Off-site control of repolarization alternans in cardiac fibers. *Physical Review E*, 81(1):011915, 2010.

[159] Y. Kobayashi, W. Peters, S. S. Khan, W. J. Mandel, and H. S. Karagueuzian. Cellular mechanisms of differential action potential duration restitution in canine ventricular muscle cells during single versus double premature stimuli. *Circulation*, 86(3):955–967, 1992.

[160] I. Adamchic, C. Hauptmann, U. B. Barnikol, N. Pawelczyk, O. Popovych, T. T. Barnikol, A. Silchenko, J. Volkmann, G. Deuschl, W. G. Meissner, M. Maarouf, V. Sturm, H. J. Freund, and P. A. Tass. Coordinated reset neuromodulation for Parkinson's disease: proof-of-concept study. *Movement Disorders*, 29(13):1679–1684, 2014.

[161] P. A. Tass, L. Qin, C. Hauptmann, S. Dovero, E. Bezard, T. Boraud, and W. G. Meissner. Coordinated reset has sustained aftereffects in Parkinsonian monkeys. *Annals of Neurology*, 72(5):816–820, 2012.

[162] R. Snari, M. R. Tinsley, D. Wilson, S. Faramarzi, T. I. Netoff, J. Moehlis, and K. Showalter. Desynchronization of stochastically synchronized chemical oscillators. *Chaos*, 25(12):123116, 2015.

[163] R. Azodi-Avval and A. Gharabaghi. Phase-dependent modulation as a novel approach for therapeutic brain stimulation. *Frontiers in Computational Neuroscience*, 9:Art. No. 26, 2015.

[164] H. Li, A. Kolpas, L. Petzold, and J. Moehlis. Parallel simulation for a fish schooling model on a general-purpose graphics processing unit. *Concurrency and Computation: Practice and Experience*, 21(6):725–737, 2009.

[165] D. Wilson and J. Moehlis. An energy-optimal methodology for synchronization of excitable media. *SIAM Journal on Applied Dynamical Systems*, 13(2):944–957, 2014.

[166] D. Wilson and J. Moehlis. Towards a more efficient implementation of low energy anti-fibrillation pacing. In review at *PLoS One*.