Firing-Rate Models in Computational Neuroscience: New Applications and Methodologies

by

Alexander G. Ginsberg

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Applied and Interdisciplinary Mathematics)
in the University of Michigan
2023

Doctoral Committee:

Professor Victoria Booth, Co-Chair
Professor Daniel B. Forger, Co-Chair
Assistant Professor Jennifer Crodelle, Middlebury College
Associate Professor Cecilia Diniz Behn, Colorado School of Mines
Professor Zaher Hani
Professor Michal Zochowski

Alexander G. Ginsberg

ginsbera@umich.edu

ORCID iD: 0009-0007-5226-7769

DEDICATION

To my Grandma Liesbeth, my Aunt Lisa, and all my family

ACKNOWLEDGMENTS

Thank you to my advisor, Victoria Booth, my co-advisor Danny Forger, my committee members, Cecilia Diniz Behn, Jen Crodelle, Michal Zochowski, and Zaher Hani, and to my family and friends, for all the support.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

TABLE

LIST OF APPENDICES

ABSTRACT

Neurons make the nervous system tick. By communicating via electrical impulses, neurons form concentrated networks in localized regions of the brain to carry out the complex tasks required of the nervous system, ranging from moving the limbs to coordinating circadian rhythms. Despite their complexity, these networks encode much information in their rates of electrical impulse production–their firing rates.

In Chapter I, we review firing-rate models. We highlight several models that spearheaded the development of the field, as well as certain applications. We augment our review with an analysis of the corresponding network of citations. In doing so, we quantify how key papers contribute to the literature. We further classify firing-rate models according to modeling methodology.

In Chapter II, we provide a new sensitivity analysis methodology which we use to study a common type of chronic pain, allodynia, where non-painful stimuli produce pain. To do so, we employ coupled firing-rate models to understand two biophysically motivated circuit structures that represent common motifs within the dorsal horn of the spinal cord. The circuit motifs, respectively, regulate the production of static and dynamic allodynia, wherein gentle pressure (static) and gentle brushing sensations (dynamic) cause pain.

To investigate variability in allodynia in each circuit motif, we identify the sets of coupling strengths that produce experimentally observed behaviors. To identify how properly behaving circuits are most vulnerable towards producing allodynia, we compute the minimal alteration in coupling strengths needed to induce the circuits to produce allodynia. We cluster the properly behaving circuits accordingly.

Results indicate that in each circuit motif, allodynia is caused by unbalancing excitation and inhibition. Results further clarify how differences in coupling strengths or circuit structure lead to different vulnerabilities towards producing allodynia.

In Chapter III, we introduce a new firing-rate model formalism capable of simultaneously addressing multiple sources of heterogeneity in a neuronal network. In particular, we apply our model to the suprachiasmatic nucleus (SCN), which likely coordinates clocks throughout the body

via 24hr oscillations in its firing rates. Further, SCN neurons intrinsically exhibit heterogeneous properties and various and non-standard forms of electrical impulses.

Our formalism consists of a system of integro-differential equations describing the time evolution of the mean and standard deviation of synaptic conductances across the network. Properties of SCN neurons are incorporated by computing responses to synaptic conductance inputs of a Hodgkin-Huxley-type SCN neuron model that exhibits these non-standard firing patterns. Such responses are then averaged over distributions of relevant quantities and included in the differential equations.

Results suggest mechanisms by which physiologically relevant changes to firing rates may appear. For instance, results show that a large spread in circadian phases across SCN neurons reduces the amplitude of the 24hr oscillations in SCN network firing activity, identifying a mechanism by which heterogeneities in neuron electrophysiology could influence circadian rhythms.

In Chapter IV, we employ the firing-rate model from Chapter III to mechanistically understand how environmental light, e.g. sunlight and smartphone light, affects SCN firing activity. In doing so, we find that we could better describe SCN output than if we use standard firing-rate models. Further, in modeling the light-to-SCN-output pathway, we identify a novel trajectory traversed throughout the day by conductances of important potassium ion channels in SCN neurons. The resulting trajectory clarifies the link between molecular clocks within SCN neurons to the electrical state of the neurons.

CHAPTER I

Introduction to Firing-Rate Models

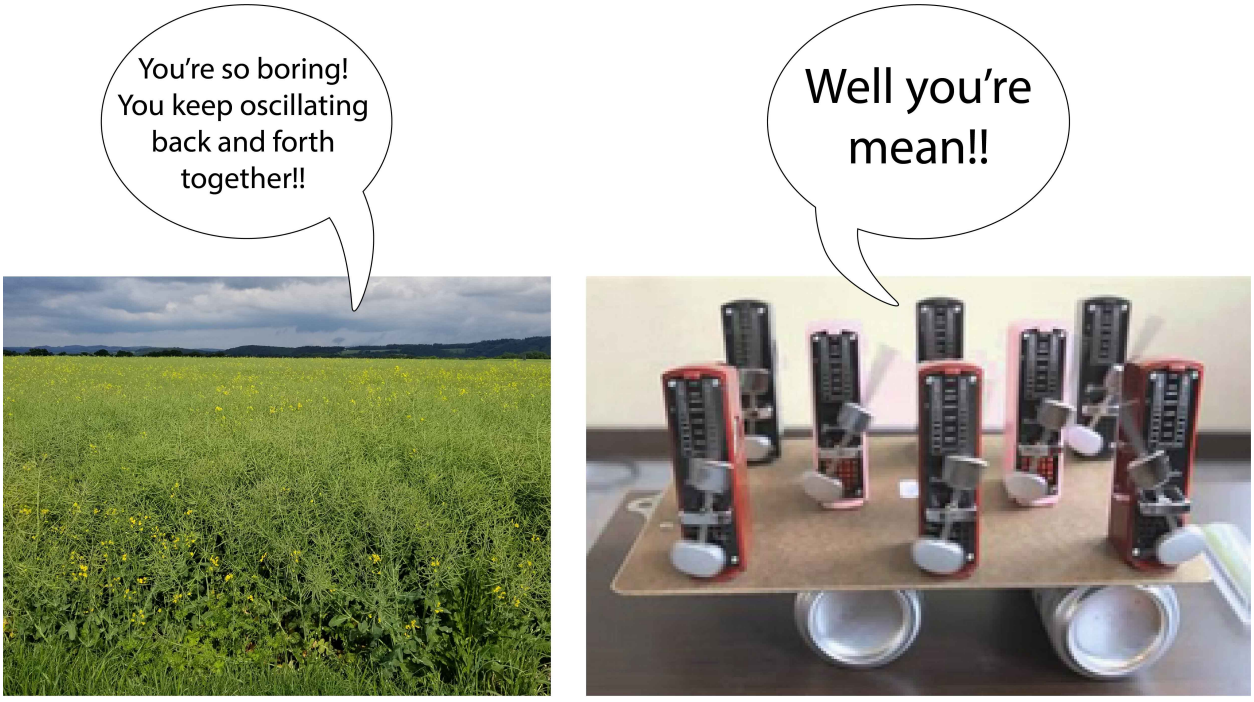


Figure I.1: Mean-field theory and coupled oscillator theory–two paradigms for reducing networks of neurons. (Coupled metronome image source: [Isa13])

1.1 The nervous system

The nervous system is fascinating. Not only does it carry out complex tasks, such as moving the limbs, coordinating circadian rhythms, and processing pain, but it is the source of consciousness. Moreover, disruptions of the nervous system underlie terrible diseases, ranging from terminal illnesses such as Alzheimer's disease, and Parkinson's disease, to psychiatric disorders such as dementia, depression, anxiety, and schizophrenia [BCP20], to chronic pain [TC10] and jet lag [FK17]. Thus, not only would understanding the nervous system clarify the human condition, but doing so would be of great importance to society.

Despite many advances, numerous important functional aspects of the nervous system remain poorly understood [BCP20]. Such a lack of understanding stems primarily from the complexity of the nervous system. Indeed, the nervous system consists of a variety of cells, including a vast network of neurons, with nearly 100 billion neurons in the human brain alone [ACG$^+$09]. To enable the nervous system to carry out tasks, such as moving the limbs, coordinating circadian rhythms, processing pain, and more, the neurons must communicate with one another.

1.1.1 Neurons and neuronal dynamics

Communication between neurons is initiated where their cell bodies, called the soma, meet a long, tube-like part of the neuron, called the axon (Figure I.2.a), which relays signals towards other neurons [BCP20]. Specifically, both the soma and the axon, and for that matter the whole neuron, are approximately enclosed by a membrane which blocks ions from entering and exiting the cell. However, embedded into this membrane are ion channels, which allow certain ions to enter and/or exit the cell. Via these ion channels, the neuron creates a charge difference, commonly referred to as the "membrane voltage" of the neuron, between its interior and the fluid surrounding the cell [BCP20]. In most neurons, ion channels maintain a charge difference so that the interior of the neuron has a negative charge relative to the extracellular fluid. This leads to a voltage difference of about -50 to -80 mV across the cell membrane (Figure I.2.b), usually referred to as the "resting membrane voltage" [BCP20].

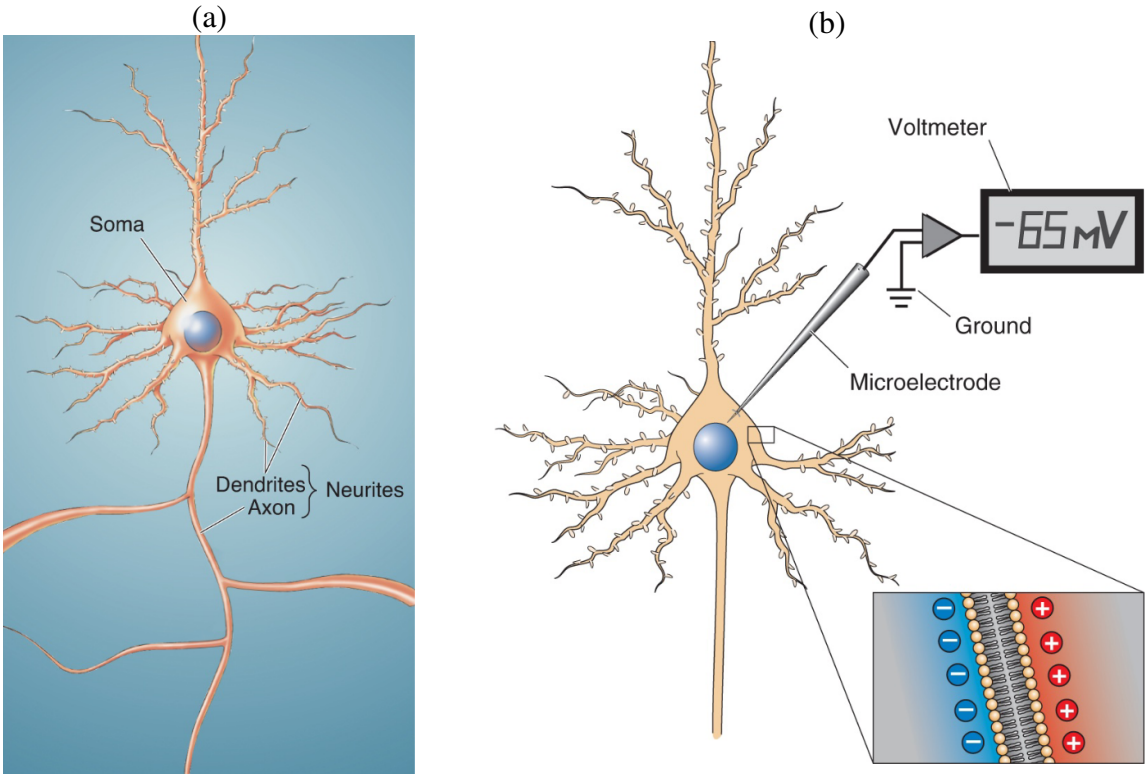


Figure I.2: Neuron structure and the voltage difference across the neuron membrane. **(a)**: The structure of a neuron. The soma is the cell body of the neuron. The soma houses cell organelles such as the nucleus and mitochondria. Dendrites and axons constitute the neurites. Dendrites receive signals from other neurons, whereas axons send signals towards other neurons. **(b)**: Charge builds up across the membrane of the cell body, so that the interior of the soma is negatively charged with respect to the exterior. Under certain conditions the neuron can rapidly depolarize, where the voltage difference across the membrane of the soma rapidly shrinks, generating an action potential. Images from [BCP20].

However, based on the value of the membrane voltage, ion channels can open or close to allow positively charged sodium and potassium ions or negatively charged chloride ions to flow into or out of the cell [BCP20]. As a result, the interior of the neuron can become more positively charged. We say then that the neuron depolarizes, as the voltage difference across the cell membrane becomes less negative. If the depolarization is large enough, it can trigger a cascade, where ion channels rapidly open, allowing positively charged ions to flow into the cell. As a result, there is a rapid increase in voltage [BCP20].

When the voltage difference across the membrane of the neuron becomes small or even positive, though, ion channel configurations change so that the flow of ions into the cell is net negative. Consequently, the voltage of the neuron rapidly falls or hyperpolarizes. The voltage can fall so rapidly that it falls below its resting membrane voltage values, after which the neuron slowly depolarizes back towards its resting membrane voltage [BCP20].

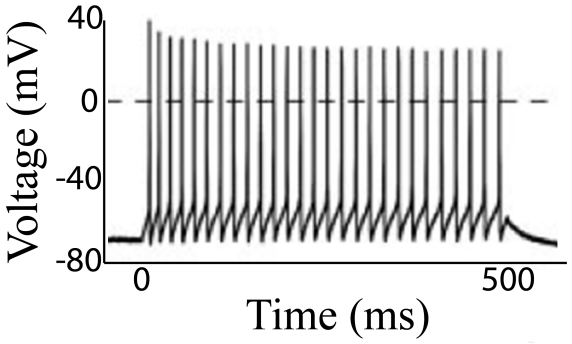


Figure I.3: Spikes in the voltage difference across the neuron membrane–action potentials. An action potential consists of a rapid rise in the voltage difference across the neuron membrane, followed by a rapid drop, therefore resembling a spike. After the rapid drop, the membrane voltage slowly rises. If the voltage rises high enough, another spike may occur, and the process repeats, leading to a series of action potentials, such as that shown in the figure. Image modified from [MSS+04].

Such a rapid increase followed immediately by a rapid decrease in the membrane voltage is often referred to as a spike, or "action potential", by neuroscientists [BCP20]. The action potential is ubiquitous across neurons and species, and as such has been the subject of much study. Most notably, the action potential was the subject of the Nobel prize winning work of Hodgkin and Huxley. Specifically they developed a mathematical model that reproduces the dynamics of ion channels and the subsequent action potentials [HH52].

To construct their model, Hodgkin and Huxley used that the neuron membrane acts as a capacitor. Indeed, charge builds up on either side of the membrane. Further, because ions can pass from the extracellular fluid through the membrane and into the cell via the ion channels, the ion channels function as resistors arranged "in parallel". Thus, the electrochemical dynamics of the cell can be modeled via a resistor-capacitor (RC) circuit.

In typical models of RC circuits, the net flow of charge through the ion channels–the ionic current I_{ionic}–is taken to be proportional to the rate of change $V'(t)$ of the membrane potential V. The corresponding proportionality constant, C, is referred to as the "capacitance":

$$C\frac{dV}{dt} = I_{ionic} \qquad \text{(Typical model of an RC circuit).}$$

This model is extremely intuitive. Indeed, the net flow of charge through the ion channels is precisely the net amount of charge that enters the neuron per unit time, which, when rescaled appropriately, is the voltage change per unit time.

Hodgkin and Huxley use the preceding model. However, instead of lumping currents from all ion channels into one quantity, they individually model the currents I_K, I_{Na}, and I_l across voltage-dependent potassium ion channels, voltage-dependent sodium ion channels, and all voltage-independent ion channels, respectively, along with the current $I_{applied}$ into the cell due to sources

external to the neuron. Thus, in the Hodgkin-Huxley model, the membrane voltage V follows:

$$C\frac{dV}{dt} + I_K + I_{Na} + I_l = I_{applied} \qquad \text{(Hodgkin-Huxley model)}.$$

To model the ionic currents across e.g. potassium channels, one uses that since all ion channels operate in parallel, one can combine them into a net current which obeys Ohm's law. Namely, current through all potassium channels is the change in voltage rescaled by the conductance of the channels:

$$I_k = g_k(V - V_k).$$

In the preceding expression, V_k is simply the voltage at which potassium current is zero. Because the direction of the net potassium current into the cell reverses as V_m increases from below V_k to above V_k, V_k is referred to as the potassium "reversal potential". Moreover, g_K is voltage and time dependent. In particular, at very low voltages, potassium ion channels are closed, or "gated". However, when voltages rise, potassium ion channels open and potassium flows out of the cell, helping the neuron to hyperpolarize again. The Hodgkin-Huxley equations incorporate the opening and closing of ion channels via gating variables, which are functions of membrane voltage and time, governed by a nonlinear ordinary differential equation. (See computational neuroscience textbooks such as [Bör17] or [ET10] for more information). An analogous description applies to I_{Na} and I_l, as well. However, because sodium-potassium pumps maintain sodium ions at higher concentrations outside of the cell, whereas they maintain potassium ions at higher concentrations inside the cell, sodium ions flow into rather than out of the cell when the voltage-gated sodium ion channels open as voltages start to rise. In all, the standard Hodgkin-Huxley model for a single neuron consists of four coupled nonlinear differential equations.

1.1.2 Synapses and signaling between neurons

It is widely believed by neuroscientists that action potentials are the basic unit of communication between neurons. Namely, the voltage spike propagates away from the soma along the axon towards other neurons (Figure I.2.a).

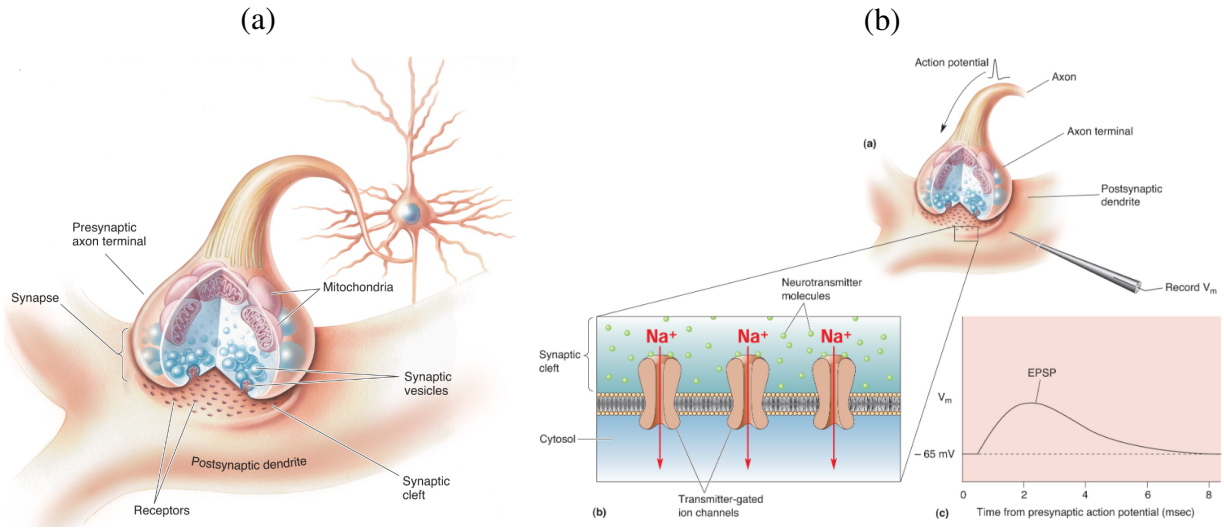


Figure I.4: Synapses, ion channels, and excitatory postsynaptic potentials. A synapse is where signals are passed from one cell (the presynaptic cell), to another cell (the postsynpaptic cell). **(a)** Typically, a synapse is where an axon from the presynaptic neuron comes nearly in contact with a dendrite from the postsynaptic neuron. Once some signal, e.g. an action potential, arrives at the end of the axon (called the axon terminal), the presynaptic neuron encapsulates chemical messengers, known as neurotransmitters, in synaptic vesicles, and sends them into the space between neurons (called the synaptic cleft). The neurotransmitters then bind to the postsynaptic neuron at specific locations known as receptors. **(b)** When a neurotransmitter binds to a receptor on a dendrite of the postsynaptic neuron, it may cause ion channels to open. For instance, if sodium channels open, positively charged sodium ions flow into the cell, causing the voltage difference across the cell wall of the dendrite to rise. Such a rise in voltage is known as an excitatory postsynaptic potential (EPSP).

Once the voltage spike reaches the end of the axon (i.e. the axon terminal) it causes the neuron to release neurotransmitters (Figure I.4.a). These neurotransmitters diffuse through the space separating the axon terminal and the dendrite (Figure I.2.a) of the neuron on the receiving end of the signal (the postsynaptic neuron). Once the neurotransmitters reach the dendrite of the postsynaptic neuron, they induce ion channels on the dendrites to open (Figure I.4.b). At some synapses referred to as excitatory, a net positive charge of ions flows through ion channels located near the dendrite, causing a small depolarization in voltage (Figure I.4.b) across the portion of the cell membrane near the synapse where the presynaptic and postsynaptic neurons meet. Given strong enough signals from presynaptic neurons, the net positive flow of ions across the dendrite of the postsynaptic neuron can cause a strong enough depolarization across the postsynaptic neuron's membrane to initiate action potential firing in the postsynaptic neuron in turn.

1.1.3 Networks of neurons in the nervous system

Neurons sending action potentials to one another in this way constitutes one of the primary methods by which neurons communicate. However, pairs of neurons communicating with one another is insufficient for completing the complex processes carried out by the nervous system. These complex processes require coordinated activity across different groups–networks–of neurons

in the brain. Such networks of neurons are often clumped together in space, allowing neuroscientists to classify brain regions according to the function they carry out. For instance, Brodmann [Bro09] famously proposed that the cortex, which roughly is the outermost layer of the brain, can be lumped according to appearance into a number of regions (Figure I.5). It turns out that such regions are also functional regions [BCP20]. For instance, region 17 is dedicated towards processing visual information. Indeed without region 17, people are blind [BCP20].

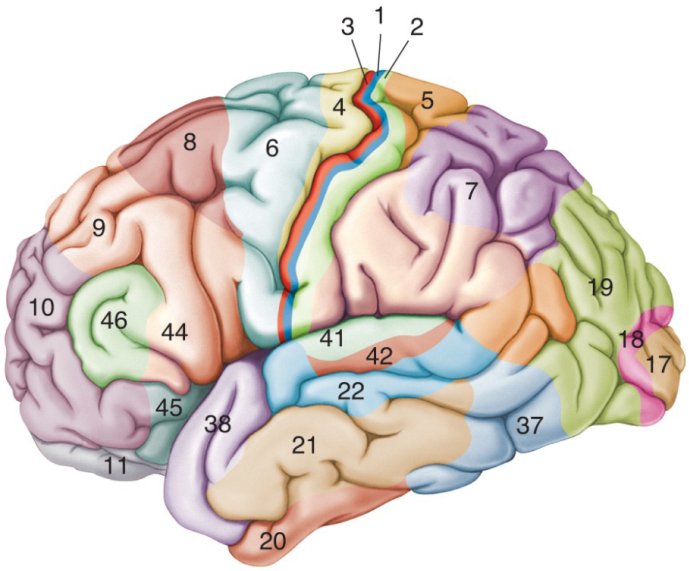


Figure I.5: Mapping the cortex via Brodmann areas. It is possible to spatially cluster the cortex into about 50 areas, based on appearance alone. It turns out that such areas constitute functional networks of neurons, as well. For instance Area 17 is the primary visual cortex, and plays a critical role in how the brain processes visual input. Image source [BCP20].

In these local neural networks, neurons wire together to generate macroscopic phenomena such as the recognition of a familiar face in the visual cortex, or to form new memories in the hippocampus. However, much as how the brain is incredibly complex, so are these smaller functional networks of neurons. Indeed in area 17, also called V1, it has been found that there are around 35 million neurons for the primate species Galagos, also known as bushbabies [CAY+10]. Moreover, neurons in these networks can be wired together in complex ways. For instance, given only 18 different neurons, as in the network shown in Figure I.6, there are about $2 \cdot 10^{76}$ possible ways to connect them via synapses to form a network.[1] Complicating matters further, each neuron integrates signals from other neurons in complex ways to produce action potentials and other interesting voltage dynamics. To make sense of such complexity, researchers have turned to mathematical modeling approaches.

[1] In fact, the number of ways d_N to connect n neurons is the number d_N of directed, weakly connected, unlabeled graphs with N nodes. Directed, because in chemical synapses, information can usually only flow in one direction (from the presynaptic to postsynaptic neuron). Weakly connected, because otherwise, the network splits cleanly into two or more networks which do not interact at all, in which case we would treat the networks it splits into separately. Unlabeled, if we are to assume there is nothing significant to distinguish one neuron from another. (If instead neurons were distinguishable, we could label them, which would increase the number of possible networks we could form). See [GW20] for an examination of networks of neurons with $N = 3$ neurons, but where two neurons are labeled. This d_N is given by A003085 in The Online Encyclopedia of Integer Sequences.

1.2 Modeling networks of neurons

To model a network of neurons, typically researchers treat the neurons as "black boxes", which receive input signals, integrate those signals, and relay the integrated signals towards other neurons in the network.

To start, researchers need to decide what "black box" to use to describe the particular means by which the neuron integrates input signals. The black box could be as simple as passing the input through an S-shaped function such as a hyperbolic tangent function, or through a "rectified linear unit". In fact, the latter function is a common approach when constructing an artificial neural network, as is done in machine learning algorithms [ESYF+20]. On the other hand, the black box could be as complicated as a full simulation of the voltage dynamics of an individual neuron, involving ion channels kinetics as in Hodgkin-Huxley models.

To describe which neurons make synaptic connections, usually researchers will specify some kind of network architecture, such as that shown in Figure I.6. In such an architecture, nodes of the network represent neurons, and edges represent synaptic connections. Such a network structure could be designed according to the network structure of neurons in the brain region in question, for example. On the other hand, if the network structure is unknown, it might simply be assumed that all neurons are connected to all other neurons (see e.g. [DJR+08]).

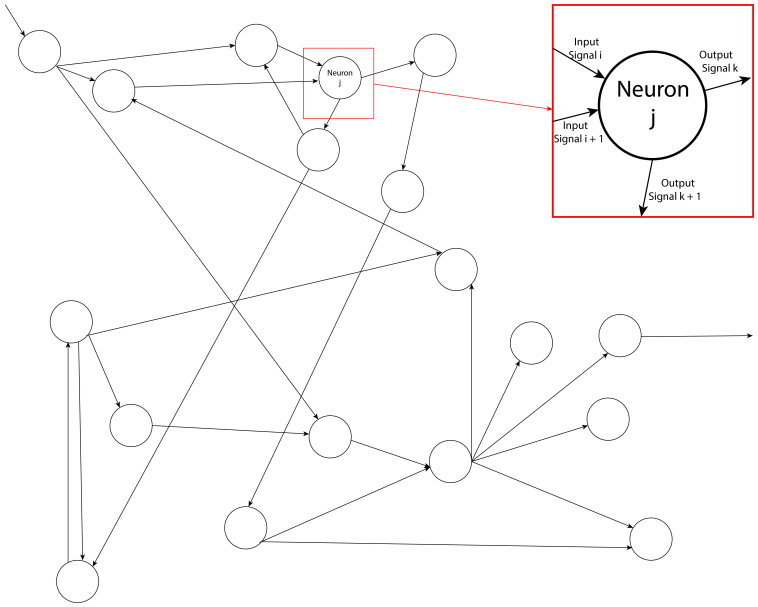


Figure I.6: A sample network of neurons. Each neuron receives signals and sends signals of its own. The network structure, formed by connecting two neurons via synapses, specifies who receives signals from who. The particular dynamics of each neuron are often treated as a black box that converts input signals into output signals.

To describe what happens when a presynaptic neuron's output signal reaches the postsynaptic neuron, researchers may use different models for the synaptic input. Such a model, could, for instance, describe the concentration of neurotransmitter output at the presynaptic axon terminal, and

describe the impact of the neurotransmitter on the voltage of the postsynaptic neuron. Or, the model could take a simpler approach, and simply rescale the output of the presynaptic neuron according to some sort of edge weight.

With these models, researchers can obtain a detailed description of how signals between neurons evolve, both across the network and over time. However, detailed network models can be computationally intensive when implemented at very large scales. To illustrate how one may go about such modeling, and the potential problems it poses, consider the following example.

Namely, consider a network of $N = 18$ neurons connected according to Figure I.6. We can describe how neuron j produces output f_j via a function g_j:

$$f_j(t) = g_j(f_1(t), f_2(t), ..., f_N(t), \eta_j(t)),$$

where $\eta_j(t)$ represents signals from external inputs. As is often done, we can assume that neurons linearly sum their inputs, so that the preceding expression can be replaced by

$$f_j(t) = g_j\left(\eta_j(t) + \sum_{i=1}^{N} w_{ij} f_i(t)\right),$$

where w_{ij} denotes the strength of the connection between neurons i and j. However, even with only 18 neurons, this is a complicated system of 18 equations, since the function g_j is nonlinear.

Moreover, 18 neurons is far fewer than we would wish to model to understand a network of neurons in the brain. Indeed, neurons in brain networks such as Brodmann Area 17 have on the order of 10^8 neurons. In which case, with this simple model we would need to deal with a system of 10^8 equations. Further, suppose we were to equip the neuron with realistic dynamics, such as Hodgkin-Huxley equations which model each neuron with four coupled nonlinear ordinary differential equations. For that model, we would have an even more unwieldy system of at least $10^8 \times 4$ differential equations that we would need to solve simultaneously.

If we were to construct such a large scale, biophysically detailed neural network model, we would then have all sorts of information available to us about what particular neurons are doing in the network. How, though, could we relate the information about the behaviors of the individual neurons to the function of the brain region relative to other brain regions? Such a question begs a reduction in the degrees of freedom of the problem.

1.3 The reduction approach: mean field theory

Researchers have used techniques borrowed from mean-field theory in statistical mechanics to reduce the dimensionality of a model network of neurons. Mean-field theory, instead of describing the dynamics of each and every neuron in the system, focuses on modeling the probability distribution of key quantities, typically relating to membrane voltage, across the network. For example, **firing-rate models** describe the evolution of the network's firing-rate statistics, and are a very popular tool for understanding the signals sent between networks of neurons. Much of the interest in firing-rate models in neuroscience has been driven by the idea of "rate-coding", wherein the information provided by signals sent out from neurons lies in the rate at which they produces action potentials.

With this key idea behind firing-rate models, namely that much of the information relayed by a neuron is not contained in the shape of an action potential but rather simply that an action potential has occurred, one can approximate the activity of a neuron as a series of instantaneous voltage spikes–a spike train (Figure I.7.a). The assumption is then that spike trains constitute the "neural code", famously the subject of the book [RWVSB99].

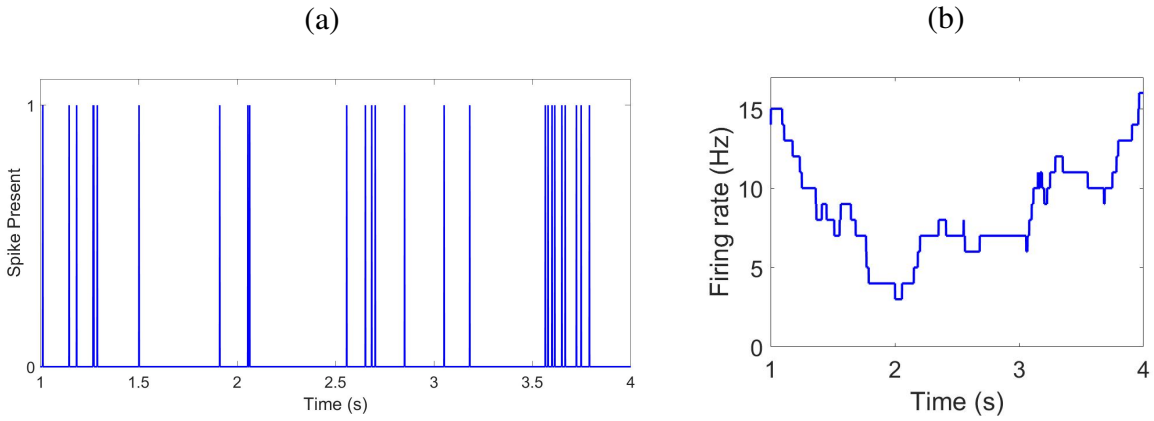


Figure I.7: Spike trains and the corresponding firing-rate. **(a)**: a sample spike train. **(b)**: The average firing rate as calculated from the spike train. The firing rate at time t is determined by counting the number of spikes occuring in the next Δt seconds and dividing by Δt. In this figure, we take $\Delta t = 1$ s.

Spike trains can then be characterized by the rate at which spikes occur. In the "rate-coding" paradigm of understanding the signals output by a neuron, one simply counts the rate at which spikes occur, i.e. the "firing-rate". That is, one views the information encoded by a neuron via

$$\text{Firing Rate} = \frac{\text{\# voltage spikes}}{\text{second}}$$

To calculate the firing-rate from a spike train, one can move a time-window across the spike train,

count the number of spikes that fall in the window at each time, and divide by the width of the window, as shown in Figure I.7.b.

We have now shown that we can summarize much of the activity of a neuron via its firing rate. We can run with this idea to represent the activity of entire networks of neurons. To do so, we look at the spike trains for all neurons in the network. We can represent the spike trains via a spike raster (Figure I.8.a), wherein for a particular time, a dot is placed in the n^{th} row of a grid to indicate that neuron n has spiked. In Figure I.8.a,the dots are most dense between times 0.2 and 0.7 s. This indicates that neurons are very active during such times, across the whole population. Such information is captured by the network average firing rate

$$\text{Average Firing Rate } = \frac{\text{\# Spikes}}{\text{second} \cdot \text{neuron}}.$$

Indeed, Figure I.8.b shows that the network average firing-rate is high between times 0.2 and 0.7, and low at other times. Thus, the network average firing-rate alone tells us that the network is most active between times 0.2 and 0.7, capturing the most important feature of the corresponding spike raster. As such, the network average firing-rate provides a tool for reducing the dynamics of a full network of neurons to a single dimension.

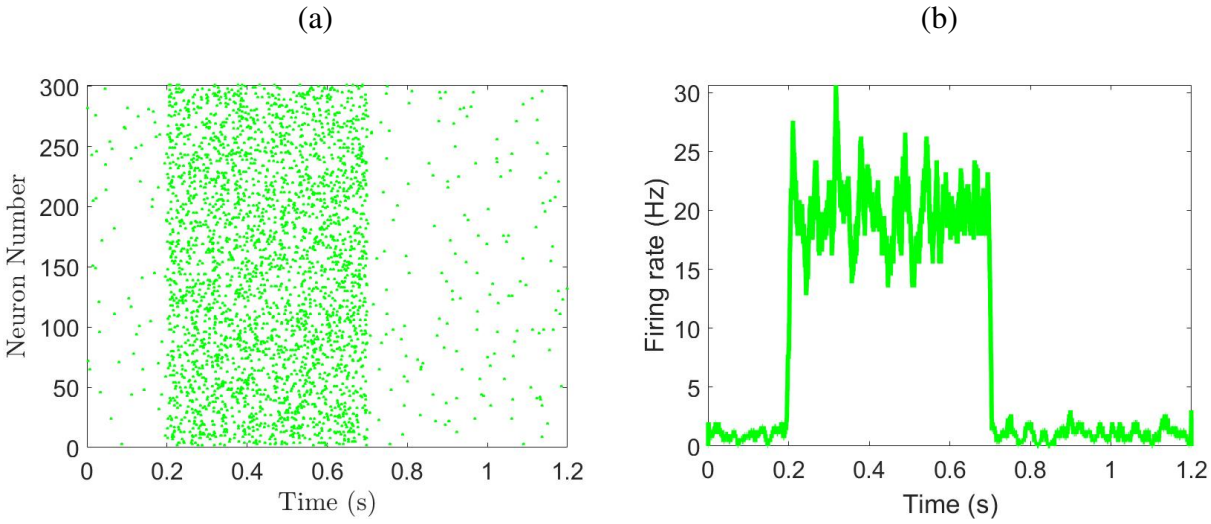


Figure I.8: Representing network activity: spike rasters and average firing-rates. **(a)**: a spike raster for a network of neurons. A dot in the n^{th} at time t indicates that neuron n fires at that time. **(b)**: The corresponding network-average firing-rate. The firing-rate at time t represents the average rate across the network at which neurons are spiking. The firing-rate captures the key information about network activity as conveyed by the spike raster–that neurons are most active between times 0.2 and 0.7.

1.4 Foundational firing-rate models

1.4.1 The Wilson-Cowan model

Firing-rate models are a relatively new development in the history of neuroscience, arising with the development of the Wilson-Cowan model [WC72] in 1972. The Wilson-Cowan model describes a cortical network consisting of two populations of neurons. One population consists of excitatory neurons, whose output excites other neurons, encouraging them to depolarize and produce action potentials. The other population consists of inhibitory neurons, whose signals inhibit other neurons, hyperpolarizing them, thereby making it more difficult for them to produce action potentials. Both of these populations are signaling each other, and at the same time are receiving external signals (P_E and P_I) from other sources, as shown in Figure I.9.

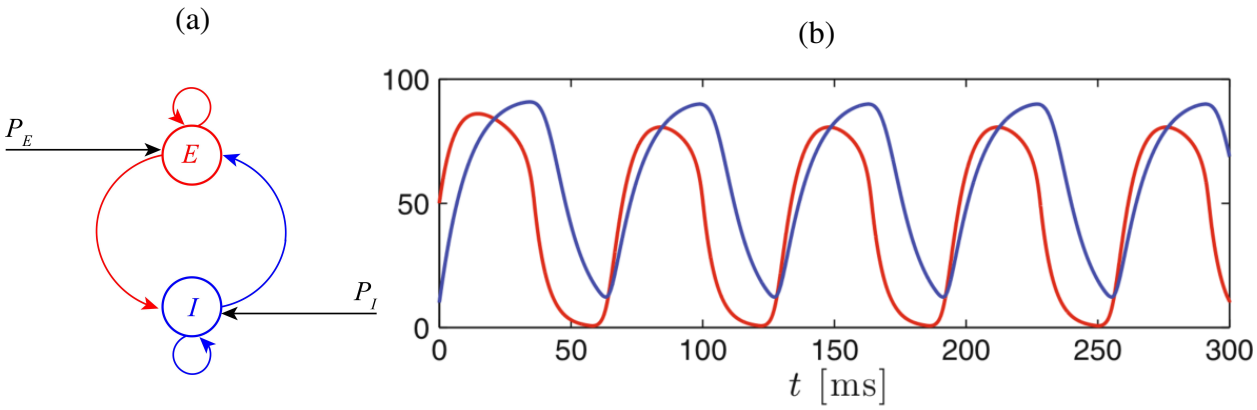


Figure I.9: The Wilson-Cowan model [WC72]. **(a)**: The Wilson-Cowan model consists of two populations of neurons. One population (I) consists of inhibitory neurons which signal other neurons to lower their activity. The other population (E) consists of excitatory neurons, which signal other neurons to increase their activity. Both E and I signal themselves and each other. The two populations receive external signals, P_E and P_I, respectively to boot. **(b)**: Firing-rates of the E and I populations show robust oscillations. Image for (a) inspired by [Bör17]. Image source for (b): [Bör17].

The Wilson-Cowan model represents neurons as "threshold neurons". In particular, if the excitation experienced by a neuron is below some threshold, the neuron does not fire. On the other hand, as soon as the excitation experienced by the neuron hits the threshold, the neuron fires, after which it cannot fire again for a time period called the refractory period. The Wilson-Cowan model then, in particular, keeps track of the average firing rates $E(t)$ and $I(t)$ for the excitatory and inhibitory populations respectively. However, because each of these neurons has a refractory period, over a small enough time window, each neuron can spike at most once. As a result, in small enough time windows, the number of neurons that are spiking is simply the number of spikes in that time

window. Thus,

$$\begin{aligned} E(t) :&= \frac{\text{\# of spikes from excitatory neurons}}{\text{\# of excitatory neurons} \cdot \text{unit time}} \\ &= \frac{\text{\# of excitatory neurons that are spiking}}{\text{\# of excitatory neurons} \cdot \text{unit time}} \\ &= \frac{\text{proportion of excitatory neurons that are spiking}}{\text{unit time}}. \end{aligned}$$

An analogous argument holds for $I(t)$. Wilson and Cowan, in their original paper, choose to frame the firing-rates in such a way. Hence, they define E and I to be

$$\begin{aligned} E(t) :=& \text{ proportion of excitatory neurons that are firing per unit time} \\ I(t) :=& \text{ proportion of inhibitory neurons that are firing per unit time} \end{aligned}$$

The Wilson-Cowan model predicts the robust oscillations in E and I activity shown in Figure I.9.b. In particular, as E increases, the excitatory neurons excite the inhibitory neurons. As a result I increases, inhibiting the excitatory neurons, and thus causing E to eventually level off and start to decrease. As a result I also decreases, releasing E from inhibition and allowing it to activate again.

The model equations that generate these oscillations rely on the "gain functions", F_E and F_I. The gain functions (Figure I.10) transform the input signal to the population into its output signal. In particular, the Wilson-Cowan model equations are:

$$\tau_e \frac{dE}{dt} = -E + (1 - r_e E) F_e(\alpha_{ee} E - \alpha_{ie} I + P_E(t)) \tag{I.1}$$

$$\tau_i \frac{dI}{dt} = -I + (1 - r_i I) F_i(\alpha_{ei} E - \alpha_{ii} I + P_I(t)), \tag{I.2}$$

where τ_E and τ_I are time constants, meaning that they determine how quickly E and I change in response to inputs; r_e and r_i are refractory periods; and α_{ei}, for example, scales the strength of the impact of signals from the excitatory population on the resulting excitation of the inhibitory population.

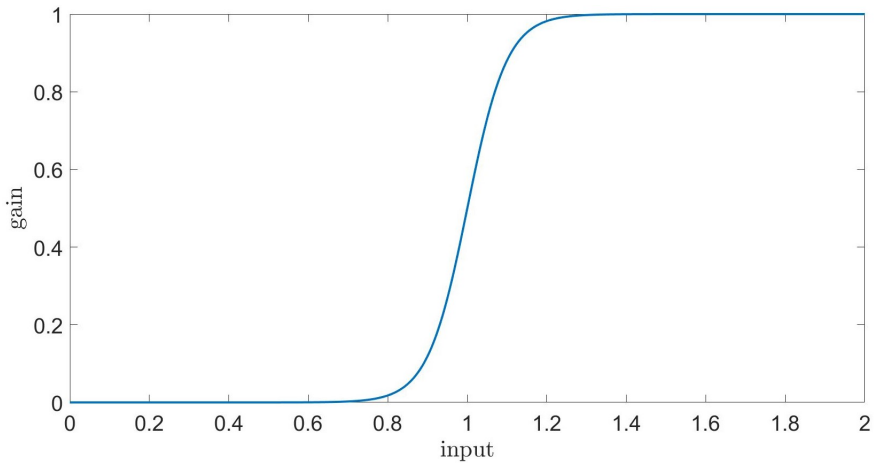


Figure I.10: Gain functions in the Wilson-Cowan model. The gain functions are traditionally taken to be sigmoidal–a bounded, differentiable, strictly increasing function with 1 inflection point. The gain function F_x, shown in this figure is sigmoidal.

1.4.2 Derivation of the Wilson-Cowan model and its sigmoidal gain function

One of the best aspects of the Wilson-Cowan model is that its derivation is fairly rigorous, and is a useful starting point for deriving other firing-rate models, such as [DSHSZ74, JR95], as we will see later in this section (Section 1.4.5). Thus, we now derive the equations for the Wilson-Cowan model, closely following [WC72]. To do so, we start by computing the response function $F(x)$, which relates the net excitation, x, coming into a population, with the proportion of neurons in that population that will be above the firing threshold per unit time.

In particular, in each subpopulation assume that there exists some distribution of neural firing thresholds $D(\theta)$, and assume that all cells receive the same large number of signals from other neurons in the population. Then, on average, all cells are subjected to the same average excitation $x(t)$. As a result, we can express the response of a subpopulation to its average excitation $x(t)$ via

$$F(x) = \int_0^{x(t)} D(\theta) d\theta$$

Alternatively, we could assume all cells within a subpopulation have the same firing threshold θ but that there exists a distribution, $C(w)$, in the number of incoming signals to each cell. Then, if $x(t)$ is the average excitation induced by one such incoming signal, all cells with at least $\theta/x(t)$ incoming synapses will be expected to receive sufficient excitation. The subpopulation response to inputs is then

$$F(x) = \int_{\theta/x(t)}^{\infty} C(w) dw$$

However, in either of the two preceding scenarios, the response functions F are strictly increasing

in x with a lower asymptote of 0 and upper asymptote of 1. In particular, if $D(\theta)$ or $C(w)$ are unimodal, the response function will be a sigmoid as shown above. Moreover, having a sigmoid is often reasonable. Indeed, single cell and population response curves are often sigmoidal functions of excitation.[2] With a sigmoidal response function, too little excitation fails to excite neurons, but very strong excitation excites all neurons.

In any case, we now have an expression for F, relating the net excitation x coming into a population to the proportion of the population that goes above threshold per unit time. However, we can do little with such an expression, unless we know the net excitation x. To find x, we make the common assumption that individual neurons sum their inputs, so that the average signal coming into a population is a weighted sum of E, I, and external signals P:

$$\text{Net signal at time } t' \text{ coming into the population} = c_1 E(t') - c_2 I(t') + P(t'),$$

where c_1 and c_2 represent the average number of excitatory and inhibitory signals, respectively, coming to neurons in the particular subpopulation under investigation.

We will also assume that the impact of signals on the population decay according to time course $\alpha(t)$. Then, at time t, the impact on the population from signaling at previous times t' is

$$\text{Average excitation due to signals from time } t' \ = \alpha(t - t')[c_1 E(t') - c_2 I(t') + P(t')].$$

As a result, the average excitation due to all signals arriving before time t is

$$x(t) = \int_{-\infty}^{t} \alpha(t - t')[c_1 E(t') - c_2 I(t') + P(t')]dt'. \tag{I.3}$$

Then, for a neuron to be firing action potentials, it must not only be above threshold, but it must not be in its refractory period. To exploit this result, we can take advantage of the law of large numbers. Namely, instead of directly computing e.g. $E(t + \Delta t)$–which we take to be the proportion of excitatory neurons which fire within Δt seconds of now–we can simply compute the probability that any neuron fires in the next unit of time. Thus, $E(t + \Delta t)$ is simply:

$$\text{Prob(the neuron fires)} \ = \text{Prob(above threshold \& not in refractory period)}.$$

We can simplify the preceding expression by assuming that the event that a neuron is above threshold

[2]For instance, firing-rates are often a roughly sigmoidal function of applied current. See [RS02] and [THR04] for experimental results relating average firing-rates of neurons, to each of which is applied a certain current. See chapter 17 of [Bör17], for example, for a description of frequency-current relations predicted by various models of single neurons.

is uncorrelated with the event the neuron is not in its refractory period. In which case, we can split the probability of the right-hand side in the preceding expression so that:

$$\text{Prob(the neuron fires)} = \text{Prob(above threshold)} \cdot \text{Prob(not in refractory period)}.$$

Indeed, if the population is richly interconnected, this absence of correlations is a reasonable assumption.

Moreover from the law of large numbers, the probability that a neuron will be above threshold due to excitatory signaling is $F(x(t))$. Thus:

$$\text{Prob(above threshold)} = F(x(t))$$

On the other hand, the proportion of neurons which *are* in the refractory period is simply the proportion of neurons that have fired during the last r units of time, where r is the duration of the refractory period. Hence, for the excitatory population:

$$\text{Prob(in refractory period)} = \int_{t-\tau}^{t} E(t')dt.$$

Thus,

$$\text{Prob(not in refractory period)} = 1 - \int_{t-\tau}^{t} E(t')dt,$$

and so,

$$E(t + \Delta t) = \left(1 - \int_{t-\tau}^{t} E(t')dt\right) F(x(t)).$$

However, the preceding expression is complicated. We can remove these integrals via temporal coarse-graining. Namely, assuming that E changes sufficiently slowly relative to the duration of the refractory period, we can express the average value $\overline{E}(t)$ of E over a very short time interval ending at time t as

$$\int_{t-r}^{t} E(t')dt' \approx r\overline{E}(t)$$

If we assume further that α is, without loss of generality, some small value[3] k_1 up to time k_2, and 0

[3] if the value is large, then the product of E and α may cease to change slowly relative to the refractory period, and our procedure would fail to work

afterwards, then for the excitatory population:

$$\int_{-\infty}^{t} \alpha(t-t')[c_1 E(t') - c_2 I(t') + P_E(t')]dt' \approx \int_{t-k_2}^{t} k_1 \left(c_1 E(t') - c_2 I(t') + P_E(t')\right) \\ \approx k_1 k_2 \left(c_1 \overline{E}(t) - c_2 \overline{I}(t) + \overline{P}(t)\right),$$

where $\bar{\cdot}$ corresponds to averaging over short time scales. Combining $k_1 k_2$ into a single variable k,

$$E(t+\Delta t) \approx (1 - r\overline{E}(t)) F_E\left(k\left[c_1 \overline{E}(t) - c_2 \overline{I}(t) + \overline{P}(t)\right]\right).$$

Dividing through by Δt and integrating both sides of the result from t to $t+\delta$ yields:

$$\frac{1}{\Delta t}\int_{t}^{t+\Delta t} E(t+\Delta t)dt \approx \frac{1}{\Delta t}\int_{t}^{t+\Delta t} (1 - r\overline{E}(t)) F_E\left(k\left[c_1 \overline{E}(t) - c_2 \overline{I}(t) + \overline{P}_E(t)\right]\right) dt.$$

However, the left-hand-side of the preceding equation is simply $\overline{E}(t+\Delta t)$. Further, because the integrand of the right-hand-side is already time coarse-grained, it is constant on short time scales. Hence,

$$\overline{E}(t+\Delta t) = (1 - r\overline{E}(t)) F_E\left(k\left[c_1 \overline{E}(t) - c_2 \overline{I}(t) + \overline{P}_E(t)\right]\right).$$

Thus,

$$\overline{E}'(t) \approx \frac{\overline{E}(t+\Delta t) - \overline{E}(t)}{\Delta t} \\ = \frac{-\overline{E}(t) + (1 - r\overline{E}(t)) F_E\left(k\left[c_1 \overline{E}(t) - c_2 \overline{I}(t) + \overline{P}_E(t)\right]\right)}{\Delta t}.$$

Taking $k = 1$ in the preceding expression yields the model equations (Equations I.1 - I.2) discussed earlier.

As is evident from the (fairly) rigorous derivation, the Wilson-Cowan model provides a good approximation to the firing activity generated by a network of excitatory and inhibitory "threshold neurons". It turns out though, that threshold neurons are not too far from reality. Namely, it is common that once a particular neuron reaches some sort of threshold voltage, it fires an action potential. Consequently, the Wilson-Cowan model has been used successfully to model cortical activity (see e.g [WC73] and [ST08]).

1.4.3 Accounting for variations over space with the Amari model

Despite the success of the Wilson-Cowan model, it is not readily adapted to describe firing-rate variations in space across a network of neurons. Firing-rate models that take into account variations in space are called "neural field models," (following the terminology of the review papers, [MPF13, PRbGF14]). These models generally consist of a system of partial differential equations, where a spatial dimension is added to describe the evolution of quantities such as average membrane voltages across a neural network as well as across time. Such models were intially introduced by Amari in 1977 [Ama77].

In particular, the Amari model pertains to neurons arranged in two-dimensional layers (as an approximation of cortical brain areas). It is assumed that each neuron receives input from neurons in their own layer as well as other layers of the network and from outside the network. The model describes the neural integration of these signals in a spatially dependent way to compute the variation of average membrane voltages and firing-rates across the network.

To describe the model, we follow the original work of [Ama77], but fill in details to make the model derivation more accessible to non-experts. To start, we introduce some notation. Namely, consider neurons located at position $\vec{x} = (x_1, x_2)$ on the i^{th} layer of the network. At time t, let

$$\text{average membrane voltage} = u_i(\vec{x}, t)$$

be the average membrane potential of the neurons in layer i at position $\vec{x}$, and let

$$\text{average firing rate} = f_i[u_i]$$

be the corresponding firing-rate, where f_i is some nonlinear function, such as a step function or sigmoidal function, as discussed earlier (see e.g. Figure I.10).

We wish to understand how the neurons at this position respond to action potentials arising throughout the network, including in other layers. To do so, we assume that the strength of stimulation from other neurons depends on their position $\vec{y}$, their layer j, and on how long in the past they fired action potentials. Thus, if the action potentials stimulating the neurons at location $\vec{x}$ in layer i occurred at time t',

$$\text{strength of stimulation} = w_{ij}(\vec{x}, \vec{y}; t - t')$$

is the resulting strength of stimulation at time t.

The impact of action potential firing activity originating at time t' in the past at location $\vec{y}$ in layer j on the neurons at location $\vec{x}$ in layer i at time t is thus the product

$$\text{impact of activity} = f_j[u_j(\vec{y}, t')]w_{ij}(\vec{x}, \vec{y}, t - t')$$

These impacts can be summed across time to yield the total impact

$$\text{impact of activity over all times} = \int f_j[u_j(\vec{y}, t')]w_{ij}(\vec{x}, \vec{y}, t - t')dt'$$

These impacts can then be summed across all neurons in layer j to yield the total impact from layer j across all times:

$$\text{impact of activity from layer } j \text{ over all times} = \int f_j[u_j(\vec{y}, t')]w_{ij}(\vec{x}, \vec{y}, t - t')d\vec{y}dt',$$

which can be summed in turn across all layers to yield the total impact:

$$\text{impact of activity from all layers over all times} = \sum_j \int f_j[u_j(\vec{y}, t)]w_{ij}(\vec{x}, \vec{y}, t - t')d\vec{y}dt'.$$

From this expression, we can begin to construct differential equations for the model. Namely, if neurons in layer i have resting membrane voltage $-r_i$ and average external stimulus intensity $\overline{s}_i$, the steady state voltage of neurons in layer i at location $\vec{x}$ at time t is

$$\text{steady state average voltage} = \overline{s}_i - r_i + \sum_j \int f_j[u_j(\vec{y}, t')]w_{ij}(\vec{x}, \vec{y}, t - t')d\vec{y}dt'.$$

Further, we assume those neurons converge to this steady state voltage at some rate, τ_i. To do so, we write the following differential equation

$$\tau_i \frac{\partial u_i(\vec{x}, t)}{\partial t} = -u_i(\vec{x}, t) + \overline{s}_i - r_i + \sum_j \int f_j[u_j(\vec{y}, t')]w_{ij}(\vec{x}, \vec{y}, t - t')d\vec{y}dt' \qquad \text{(assuming steady external stimuli).}$$

Introducing spatiotemporal deviations of the external stimuli from the mean of the form $s_i(\vec{x}, t)$, and rewriting $\overline{s}_i - r_i$ as h_i, we obtain the field equation:

$$\tau_i \frac{\partial u_i(\vec{x}, t)}{\partial t} = -u_i(\vec{x}, t) + h_i + s_i(\vec{x}, t) + \sum_{j=1}^{m} \int w_{ij}(\vec{x}, \vec{y}; t - t')f_j[u_j(\vec{y}, t)]d\vec{y}dt'. \qquad \text{[Ama77]}$$

We can then arrive at a somewhat more tractable neural field model by making several simplifying assumptions. For instance, recall that we assumed there is a lag between the time a signal from

a pre-synaptic neuron was produced and the time at which the post-synaptic neuron is affected by the signal. If we now assume that the time lag is negligible, we can drop the integration of time t and replace $w_{ij}(\vec{x}, \vec{y}; t)$ with $w_{ij}(\vec{x}, \vec{y})$. Further, we could assume that coupling between neurons is spatially homogeneous, depending in particular only on the distance between them, in which case

$$w_{ij}(\vec{x}, \vec{y}) = w_{ij}(\vec{x} - \vec{y}) \qquad \text{(simplifying assumption)}.$$

Finally, we can assume the network consists of a single layer of connected neurons, in which case we can drop i, j subscripts to obtain

$$\tau \frac{\partial u(\vec{x}, t)}{\partial t} = -u + \int w(\vec{x} - \vec{y}) f[u(\vec{y})] d\vec{y} + h + s(\vec{x}, t) \qquad \text{(simplified neural field equations)}.$$

Amari used this model to analyze traveling waves of firing activity across a cortical network layer. Traveling waves of neural firing activity are believed to play an important role in the brain and may be essential for the formation of memories, for example (see [MCRS18] for a review). As such, neural field models are a useful tool for understanding important activity patterns in the brain. Using neural field models to understand the generation of such phenomena remains a popular area of research today, such as in understanding brain imaging data via "dynamic causal modeling" [MPF13, PRbGF14].

1.4.4 Population density approaches and Fokker-Planck equations

While both the Wilson-Cowan and Amari models were developed during the 1970s, it wasn't until the 1990s that researchers developed tools for understanding the full distribution of key quantities across a network of neurons. The population density approach, introduced by Amit & Brunel [AB97], as well as by Knight et al. [KMS96], did just that.

Population density approaches typically involve starting with a simple model of a spiking neuron (see section 11.2 of [ET10]), and deriving a Fokker-Plank equation modeling the time-evolution of the probability density function of neural state variables across a network. The state variable is typically membrane voltage but may also include other variables (see [DJR$^+$08]). We explain how to arrive at such a Fokker-Planck equation by following [DJR$^+$08], but filling in details and clarifying arguments and notation to make the formalism more accessible.

Let $V(t)$ be the value of the membrane voltage at time t and consider the time evolution of the probability distribution p_{V_t} of voltages at time t across the network. One then uses the law of total probability to express the probability distribution of $V_{t+\Delta t}$ in terms of the joint distribution of $V_{t+\Delta t}$

and V_t:

$$p_{V_{t+\Delta t}}(v) = \int_{-\infty}^{\infty} p_{(V_{t+\Delta t},V_t)}(v, v')dv' \qquad \text{(Chapman-Kolmogorov equation)},$$

which is a particular instantiation of the Chapman-Kolmogorov equation. However, in many situations, the joint distribution $p_{(V_{t+\Delta t},V_t)}$ may be unknown, e.g. if V_t is a Markov process. Thus, to simplify the preceding expression, one uses Bayes' law to express the joint probability distribution of $p_{(V_{t+\Delta t},V_t)}$ in terms of the probability distribution of V_t and the probability distribution of $V_{t+\Delta t}$ conditioned on V_t:

$$p_{(V_{t+\Delta t},V_t)}(v, v') = p_{(V_{t+\Delta t}|V_t)}(v, v') \cdot p_{V_t}(v').$$

Substituting the preceding expression into the Chapman-Kolmogorov equation yields:

$$p_{V_{t+\Delta t}}(v) = \int_{-\infty}^{\infty} p_{(V_{t+\Delta t}|V_t)}(v, v') \cdot p_{V_t}(v')dv'. \tag{I.4}$$

From the preceding equation (Equation I.4), we can obtain a Fokker-Planck equation by making a Kramers-Moyal expansion of the Chapman-Kolmogorov equation (see [RTT19]). Below, we do so without assuming an advanced knowledge of probability theory, providing the necessary requisite knowledge along the way. In particular, we seek to find a series expansion of Equation I.4 by first finding a series expansion of the conditional probability distribution appearing in the integrand of Equation I.4. To do so, we introduce the small deviation ϵ, given by

$$\epsilon = v - v'.$$

so that $v' = v - \epsilon$. Then the complicated term from the integrand of Equation I.4 becomes

$$p_{(V_{t+\Delta t}|V_t)}(v, v') = p_{(V_{t+\Delta t}|V_t)}(v, v - \epsilon)$$

To simplify the preceding expression, we fix v and think of $p_{(V_{t+\Delta t}|V_t)}(v, v - \epsilon)$ as a function u_v only of ϵ:

$$u_v(\epsilon) = p_{(V_{t+\Delta t}|V_t)}(v, v - \epsilon)$$

We then seek to rewrite u_v by applying the Fourier inversion theorem to a series expansion of its

Fourier transform $\hat{u}_v$. Thus, consider the Fourier transform

$$\hat{u}_v(x) = \int_{-\infty}^{\infty} e^{-i\epsilon x} p_{(V_{t+\Delta t}|V_t)}(v, v-\epsilon) d\epsilon$$

and substitute the Taylor expansion of $e^{-ix\epsilon}$ around $\epsilon = 0$ into the preceding expression:

$$\hat{u}_v(x) = \int_{-\infty}^{\infty} \sum_{n=0}^{\infty} \frac{(-ix)^n}{n!} \epsilon^n p_{(V_{t+\Delta t}|V_t)}(v, v-\epsilon) d\epsilon.$$

Assuming that the conditional probability distribution $p_{(V_{t+\Delta t}|V_t)}$ is sufficiently nice[4], we can switch the integral with the sum:

$$\hat{u}_v(x) = \sum_{n=0}^{\infty} \frac{(-ix)^n}{n!} \int_{-\infty}^{\infty} \epsilon^n p_{(V_{t+\Delta t}|V_t)}(v, v-\epsilon) d\epsilon. \tag{I.5}$$

For notational simplicity, we then denote the integral expression with

$$\langle \epsilon^n \rangle_{t,\Delta t,v} := \int_{-\infty}^{\infty} \epsilon^n p_{(V_{t+\Delta t}|V_t)}(v, v-\epsilon) d\epsilon,$$

so that the preceding expression becomes

$$\hat{u}_v(x) = \sum_{n=0}^{\infty} \frac{(-ix)^n}{n!} \langle \epsilon^n \rangle_{t,\Delta t,v}.$$

However, we know from the Fourier inversion theorem that

$$u_v(\epsilon) = \int_{-\infty}^{\infty} e^{ix\epsilon} \hat{u}_v(x) dx.$$

[4]What constitutes "sufficiently nice" is somewhat technical, but means, more or less, that the integrals

$$\langle \epsilon^n \rangle_{t,\Delta,v} := \int_{-\infty}^{\infty} \epsilon^n p_{(V_{t+\Delta t}|V_t)}(v, v-\epsilon) d\epsilon$$

are finite for all $n = 0, 1, 2, , ...$ Specifically, $p_{(V_{t+\Delta t}|V_t)}$ is "sufficiently nice" for our purposes as long as it satisfies the conditions of the Tonelli-Fubini theorem–the usual theorem for switching the order of integration (see [Fol99]). Namely, we rewrite the integral for $\hat{u}_v$ as

$$\hat{u}_v(x) = \int_{-\infty}^{\infty} \left(\int_0^{\infty} \frac{(-ix)^n}{n!} \epsilon^n p_{(V_{t+\Delta t}|V_t)}(v, v-\epsilon) dn \right) d\epsilon,$$

to which we apply the Tonelli-Fubini theorem. The inner integral is over the counting measure, (where the measure of any non-negative integer is 1). The outer integral, on the other hand, is the usual integral over the Lebesgue measure on complex numbers. We leave the remainder of the details to the reader.

Substituting in the expression we just found for $\hat{u}_v$, we obtain:

$$u_v(\epsilon) = \int_{-\infty}^{\infty} \sum_{n=0}^{\infty} e^{ix\epsilon} \frac{(-ix)^n}{n!} \langle \epsilon^n \rangle_{t,\Delta t,v} dx \tag{I.6}$$

We have now rewritten $u_v(\epsilon)$ using the Fourier inversion theorem and the Fourier transform of $u_v(\epsilon)$.

However, the preceding expression is complex. Unfortunately, we cannot simplify it by switching the order of the sum and the integral, as we did before. Indeed, let us switch the order of integration, and see what happens:

$$u_v(\epsilon) = \sum_{n=0}^{\infty} \langle \epsilon^n \rangle_{t,\Delta t,v} \int_{-\infty}^{\infty} e^{ix\epsilon} \frac{(-ix)^n}{n!} dx.$$

However, note that when $n = 0$, the integrand from the preceding equation becomes

$$\int_{-\infty}^{\infty} e^{ix\epsilon} \frac{(-ix)^n}{n!} dx = \int_{-\infty}^{\infty} e^{ix\epsilon} dx,$$

which is the Dirac-delta function $\delta(\epsilon)$, and is thus not a function, per se[5]. Moreover, now considering any n, it turns out, as we will see shortly, that

$$\int_{-\infty}^{\infty} e^{ix\epsilon} \frac{(-ix)^n}{n!} dx = \frac{1}{n!} \delta^{(n)}(\epsilon),$$

the n^{th} derivative of the Dirac-delta function, which is also not a function. Even though we cannot switch the order of the sum and the integral, the relationship of $u_v(\epsilon)$ to the Dirac-delta function will be pertinent, going forward.

The relationship of $u_v(\epsilon)$ to the Dirac-delta function is particularly relevant upon substituting the expression for $u_v(\epsilon)$ from Equation I.6 into the integrand for the latest expression for $p_{V_{t+\Delta t}}(v)$ (Equation I.4):

$$p_{V_{t+\Delta t}}(v) = \int_{-\infty}^{\infty} p_{V_t}(v') \left(\int_{-\infty}^{\infty} \sum_{n=0}^{\infty} e^{ix\epsilon} \frac{(-ix)^n}{n!} \langle \epsilon^n \rangle_{t,\Delta t,v} dx \right) dv'.$$

To simplify the preceding expression, we take a somewhat heuristic approach to avoid getting into the details of generalized functions such as the Dirac-delta function. To start, even though we cannot

[5]Rather, the Dirac-delta function is a type of generalized function known as a *distribution*, (not to be confused with a probability distribution).

switch the order of the summation and the rightmost integral in the preceding expression, we can (somewhat heuristically) move the summation so it comes before both integrals in the preceding expression:

$$p_{V_{t+\Delta t}}(v) = \sum_{n=0}^{\infty} \int_{-\infty}^{\infty} p_{V_t}(v') \langle \epsilon^n \rangle_{t,\Delta t,v} \left(\int_{-\infty}^{\infty} e^{ix\epsilon} \frac{(-ix)^n}{n!} dx \right) dv'.$$

Substituting $\epsilon = v - v'$ back into the preceding expression, we obtain:

$$p_{V_{t+\Delta t}}(v) = \sum_{n=0}^{\infty} \int_{-\infty}^{\infty} p_{V_t}(v') \langle \epsilon^n \rangle_{t,\Delta t,v} \left(\int_{-\infty}^{\infty} e^{ix(v-v')} \frac{(-ix)^n}{n!} dx \right) dv'.$$

Conveniently, (and again somewhat heuristically), we have from integration by parts[6] in the second equality, below, that for any sufficiently differentiable function $f(v')$,

$$\begin{aligned}
\int_{-\infty}^{\infty} f(v') \left(\int_{-\infty}^{\infty} e^{ix\epsilon} \frac{(-ix)^n}{n!} dx \right) dv' &= \frac{(-1)^n}{n!} \int_{-\infty}^{\infty} f(v') \left(\int_{-\infty}^{\infty} e^{ix\epsilon} (ix)^n dx \right) dv' \\
&= -\frac{(-1)^n}{n!} \int_{-\infty}^{\infty} f'(v') \left(\int_{-\infty}^{\infty} \frac{-1}{ix} e^{ix(v-v')} (ix)^n dx \right) dv' \\
&= \frac{(-1)^n}{n!} \int_{-\infty}^{\infty} f'(v') \left(\int_{-\infty}^{\infty} e^{ix(v-v')} (ix)^{n-1} dx \right) dv'.
\end{aligned}$$

Thus, by induction,

$$\begin{aligned}
\int_{-\infty}^{\infty} f(v') \left(\int_{-\infty}^{\infty} e^{ix\epsilon} \frac{(-ix)^n}{n!} dx \right) dv' &= \frac{(-1)^n}{n!} \int_{-\infty}^{\infty} f^{(n)}(v') \left(\int_{-\infty}^{\infty} e^{ix(v-v')} dx \right) dv' \\
&= \frac{(-1)^n}{n!} \int_{-\infty}^{\infty} f^{(n)}(v') \delta(v - v') dv' \\
&= \frac{(-1)^n}{n!} f^{(n)}(v).
\end{aligned}$$

Applying the preceding result to $p_{V_t}(v') \langle \epsilon^n \rangle_{t,\Delta t,v}$, we have that

$$p_{V_{t+\Delta t}}(v) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial v^n} \left[p_{V_t}(v) \langle \epsilon^n \rangle_{t,\Delta t,v} \right].$$

Noting that the $n = 0$ term is simply $p_{V_t}(v)$, we have then from the limit definition of the

[6] Since, as mentioned before, the integral $\left(\int_{-\infty}^{\infty} e^{ix(v-v')} \frac{(-ix)^n}{n!} dx \right)$ is the n[th] derivative of the Dirac-delta function, and is thus a distribution rather than a function, integration by parts in this context is perhaps better described as taking a *distributional derivative*.

derivative that

$$\begin{aligned}\frac{\partial p_{V_{t+\Delta t}}(v)}{\partial t} &= \lim_{\Delta t \to 0} \frac{p_{V_{t+\Delta t}}(v) - p_{V_t}(v)}{\Delta t} \\ &= \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial v^n} \left[p_{V_t}(v) \lim_{\Delta t \to 0} \frac{\langle \epsilon^n \rangle_{t,\Delta t,v}}{\Delta t} \right].\end{aligned}$$

The preceding expression is the Kramers-Moyal expansion of the Chapman-Kolmogorov equation (see [DJR$^+$08]). Moreover, omitting all but the first-two terms of the preceding equation yields a Fokker-Planck equation which is commonly used in the literature of the population-density approach:

$$\frac{\partial p_{V_{t+\Delta t}}(v)}{\partial t} = -\frac{\partial}{\partial v} \left[p_{V_t}(v) \lim_{\Delta t \to 0} \frac{\langle \epsilon \rangle_{t,\Delta t,v}}{\Delta t} \right] + \frac{1}{2} \frac{\partial^2}{\partial v^2} \left[p_{V_t}(v) \lim_{\Delta t \to 0} \frac{\langle \epsilon^2 \rangle_{t,\Delta t,v}}{\Delta t} \right].$$

The task is then to calculate the $\langle \epsilon^k \rangle$ from network connectivity and the underlying voltage dynamics of neurons in the network. The dynamics simplify particularly nicely if the single neuron model follows the linear-integrate-and-fire model. In that case, and while making several other assumptions about the network, one can obtain relatively simple expressions for $\langle \epsilon^k \rangle$ which yield, in turn, a relatively simple Fokker-Planck PDE [DJR$^+$08]. Such PDEs may be difficult to solve directly [ET10][DJR$^+$08], particularly if one wishes to make the state variable $V(t)$ multidimensional, as would would be appropriate for Hodgkin-Huxley-like neuronal dynamics. However, one can simplify things by looking at the stationary solutions of the Fokker-Planck equation.

While such an approach is capable of describing the time evolution of the full probability distribution of voltages across a network, it is not an end-all solution for firing-rate models. Namely, these Fokker-Planck PDEs may be difficult to solve directly [DJR$^+$08]. Further, it is unclear how well this formalism applies to complicated neuronal dynamics. Nevertheless, Chizhov and Graham [CG07] attempted to apply the formalism to complicated neuronal dynamics by reducing a Hodgkin-Huxley-like neuron model to a resetting neuron model similar to the leaky-integrate and fire neurons. For more information on how such population density models are derived see [DJR$^+$08] for a nice review.

1.4.5 Extending the Wilson-Cowan formalism–the da Silva et al. and the Jansen-Rit models

The Wilson-Cowan model inspired the development of models critical to understanding brain activity data recorded as the electroencephalogram (EEG). Of particular importance was the use of modeling to understand alpha waves, which correspond to roughly 8 - 12 Hz oscillations of the voltage signal detected by the EEG. Namely, one key paper to investigate such waves in EEG data

was [DSHSZ74], in which da Silva et al. developed a new firing-rate model for EEG oscillatory signals.

The da Silva et al. model

The da Silva et al. model [DSHSZ74], much as the Wilson-Cowan model, describes the average activity of a population of excitatory neurons coupled with a population of inhibitory neurons. In fact, the da Silva et al. model is very nearly identical to the Wilson-Cowan model. However, to study EEG data, which tracks changes in the average voltage across networks of neurons in the brain, da Silva et al. needed to model membrane voltages, rather than model firing-rates, as did Wilson and Cowan. Thus, the key development of the da Silva et al. model was to essentially re-formulate the Wilson-Cowan model so it describes the average voltage of a network of neurons, rather than the average firing-rate.

To reformulate the Wilson-Cowan model in terms of voltage, da Silva et al. take a heuristic approach. Namely, instead of directly modeling the firing-rates $E(t)$ and $I(t)$ of the excitatory and inhibitory populations directly, they assume that the firing-rate $E(t)$ for the excitatory population, for example, depends not on the average excitation $x(t)$ the neuron receives, but instead on that population's average voltage V_E:

$$E(t) = f(V_E(t)).$$

In this case, f differs from F from the Wilson-Cowan model by directly modeling the proportion of neurons that are firing an action potential, rather than measuring the proportion of neurons that are above threshold, and does not take into account any sort of refractory period, as does the Wilson-Cowan model.

Moreover, instead of assuming that the signaling from other neurons alters the "average excitation" $x(t)$ to a neuron, da Silva et al. assume that signaling from other neurons directly modulates the membrane voltage of the population according to impulse response functions h_E and h_I, for the excitatory and inhibitory populations, respectively. These impulse response functions are the analog of the decay function $\alpha(t)$ from the Wilson-Cowan model. So, instead of having a decay α for all populations, we have decays h_E and h_I for the two populations. Then, writing the external signaling to the excitatory population as $P(t)$, and the average number of signals impinging on the excitatory population from the inhibitory population as c_2 leads to

$$\Delta V_E(t) = \int_0^t h_E(t-t')P(t') - h_I(t-t')c_2 I(t')dt', \tag{I.7}$$

where $\Delta V_E = V_E - V_{E,rest}$ is the change in voltage induced by signaling impinging on the excitatory population. This differs from the expression for the net excitation $x(t)$ from the Wilson-Cowan model (Equation I.3) only by using h_e and h_i instead of α and by not including recurrent excitation from the excitatory population to itself.

Da Silva et al. then proceed to replace the quantities in the preceding expression with the corresponding deviations from the mean of the quantities. They then take Laplace transforms of the result to arrive at closed form expressions for the Laplace transforms of V_E and V_I. From the Laplace transforms of V_E and V_I, they can then characterize $V_E(t)$ and $V_I(t)$ via the corresponding frequency powers spectrums.

While it was instrumental to their analysis to arrive at closed form expressions for V_E and V_I, it is likely that their model formalism has been more relevant to the development of computational neuroscience. Indeed, it has been established, e.g. by [Fre75], that it is reasonable to model the impact of signals on the average membrane voltage of a neuron using impulse response functions such as h_i and h_j. Moreover, the idea of modeling the average membrane voltages of neurons has been critical for understanding EEG data. This idea was revised in 1993 by Jansen et al. in [JZB93] and brought to the forefront of research via the Jansen-Rit model in [JR95].

The Jansen-Rit model

The Jansen-Rit model [JZB93] [JR95] advances the da Silva et al. model in a key way. Namely, to model the voltage-response of a population of neurons to incoming signals, they replace $h_E(t)$ and $h_I(t)$ with an alpha function, e.g

$$\alpha(t) = \begin{cases} \frac{1}{\tau} t e^{1-t/\tau} & t \geq 0 \\ 0 & t < 0 \end{cases}.$$

We can then use the preceding function, $\alpha(t)$, to write $h_E(t)$ and $h_I(t)$ as

$$h_E(t) = A_E \alpha(t)$$
$$h_I(t) = A_I \alpha(t),$$

where A_E and A_I are constants that determine the magnitude of the impulse responses.

However, Jansen and Rit interpret h_E and h_I differently from da Silva et al. in [DSHSZ74]. Namely, da Silva et al. take h_E and h_I to be the average response elicited by an excitatory and inhibitory impulse, respectively. That is, in the da Silva et al. model , h_E and h_I are intrinsic to

the presynaptic population. Jansen and Rit on the other hand, assume that a population's response to input is the same regardless of the type of input. Thus, in the Jansen-Rit model, h_E and h_I are intrinsic to the postsynaptic population. This lets them rewrite the convolution integral (Equation I.7) for, say, change in V_E, using only one response function instead of the two that appear in Equation I.7) :

$$\Delta V_E(t) = \int_0^t h_E(t - t') \left[P(t') - c_2 I(t')\right] dt'.$$

This choice of representing h_E as an α function allows the preceding convolution integral to be rewritten as a differential equation:

$$\begin{aligned}(\Delta V_E)''(t) + \frac{2}{\tau}(\Delta V_E)' + \frac{1}{\tau^2}\Delta V_E &= \frac{A_E \exp(1)}{\tau}[P(t') - c_2 I(t')] \\ &= \frac{A_E \exp(1)}{\tau}[P(t') - c_2 f(V_I(t))].\end{aligned}$$

More generally, if a population has impulse response function $h(t) = A\alpha(t)$, then, its voltage response, ΔV, is modeled as

$$(\Delta V)''(t) + \frac{2}{\tau}(\Delta V)' + \frac{1}{\tau^2}\Delta V = \frac{A \exp(1)}{\tau} \cdot \text{inputs}.$$

Jansen and Rit proceed to write down a system of differential equations that describes the time evolution of the average membrane voltages of each population of neurons in the cortical networks that they are investigating. While somewhat more heuristic, the Jansen-Rit model was used effectively to model alpha rhythms across cortical networks, much as did the da Silva et al. model .

Moreover, the Jansen-Rit model has served as a starting point for dynamic causal modeling (see [FHP03]), which seeks to model how external controls affect the dynamics of a network of neurons. For example, dynamic causal modeling reproduces data by fitting parameters from firing-rate models that describe membrane voltage, such as the Jansen-Rit model in the case of EEG data. Further, one of the first papers which applies dynamic causal modeling [DF03] , adapts the Jansen-Rit model in such a way that it produces "the whole spectrum" of EEG rhythmic activity.

1.4.6 Coupled-oscillator theory based approaches

Coupled oscillator theory is a another useful approach for reducing models for networks of neurons (see e.g. [BGLM20] for a review). In this theory, neurons are reduced to "oscillators" by assuming that the voltage of each neuron is periodic, as would occur when neurons repetitively fire action potentials. Typically, a coupled oscillator model replaces the voltage of the neuron with a

phase in $[0, 2\pi]$. This modeling approach has been used to study synchrony of coupled neurons, which occurs when the neurons have roughly the same phase (see chapter 8 of [ET10]). Whether or not a network of neurons is synchronous is believed to be one of the most physiologically relevant macroscopic descriptions of a network of neurons (see e.g. [UPL+09]). For example, synchronous neural firing is the one of the main culprits behind seizures [RBHK21]. One of the most popular coupled oscillator models is the "Kuramoto model" [Kur75], which we will discuss further in Chapter IV. While the original Kuramoto model requires a differential equation for the phase of each neuron in the network, recently Ott and Antonsen developed an analytical reduction technique to arrive at a low-dimensional model for the macroscopic or collective activity across the network [OA08].

Generally, neither the Kuramoto model nor the Ott-Antonsen reductions were initially used to explicitly describe firing rate activity in networks of neurons. More typically, coupled oscillator models have been used to instead study synchronization in networks of firing neurons. Moreover, the firing-rate model literature has remained somewhat separate from the coupled oscillator theory literature. Nevertheless, over the last decade or so, researchers have used coupled oscillator theory to describe the firing-rates across networks of neurons [BGLM20]. See [BGLM20] for a nice review of mean-field reductions of neural (and other biological) oscillators.

1.5 Present and future firing-rate models

Let us highlight the key contributions of these foundational firing-rate models from a historical perspective. In 1972, the Wilson-Cowan model [WC72] brought firing-rate models into popularity in computational neuroscience with the development of a computationally tractable, carefully derived system of equations that captures the oscillations in firing-rates for a population of excitatory neurons interacting with a population of inhibitory neurons. The da Silva et al. model [DSHSZ74] adapted the derivation of the Wilson-Cowan model to describe the dynamics of the average membrane voltage across a system of neurons. Jansen and Rit [JR95] later expanded upon the da Silva et al. model , so that it could be expressed as a system of differential equations, opening the door for studies of EEG data via firing-rate models. Meanwhile, in 1977, Amari in [Ama77] introduced spatial dynamics to firing-rate models. In 1997, Amit & Brunel [AB97] further generalized firing-rate models by borrowing the population density approach from statistical physics so that they could describe the full distribution of quantities, such as average membrane voltage[7].

Today, these foundational models are widely used in computational neuroscience. In fact, most

[7] Knight et al. [KMS96] published a population density model a year before Amit & Brunel published theirs, but cite a preprint of Amit & Brunel's work, and thus were aware of Amit & Brunel's work.

other firing-rate model formalisms that have been developed derive in one way or another from one of these foundational models.

To explore precisely how today's firing-rate model literature traces its roots back toward these foundational firing-rate models, and to identify other key firing-rate models, we have constructed a citation network of the literature.

To start our analysis, we use a citation network generator that is available online (citationgecko, link). To the generator, we provide "seed papers" consisting of review papers such as [DJR+08] as well as papers which developed the foundational firing-rate models, such as the Wilson-Cowan model. The generator then provides us with a number of possible papers that share with the seed papers a large number of citing works or a large number of works cited. Going through the resulting papers by hand, we identify which papers truly pertain to developing new firing-rate models, arriving at around 30 firing-rate model papers.

We then use those modeling papers as a new set of "seeds" to identify a representation of the full firing-rate model literature. Unfortunately, the online generator was unable to handle a large number of papers. Thus, we proceed by identifying all papers citing and cited by the 30 previously identified modeling papers. Using Web of Science[8] and custom Matlab code, we extract the citing and cited works, and compile them into a network, in the graph theoretical sense.

Of the approximately 6000 papers found, it is possible that some may not pertain directly to firing-rate models. We identify irrelevant papers by finding the in-degree (the number of citations the paper has) and the out-degree (the number of works the paper cites) for each paper in the network. If the in-degree and out-degree are either one or zero, then we know that the paper is not closely related to any of the 30 seed firing-rate model papers, except for the one that cited it or was cited by it. We thus conclude that the paper with the very low in-degree and out-degree does not focus on firing-rate models and we thus discard it from the network. The result of this process is a smaller network of about 500 papers, which we believe to be a valid representation of the firing-rate model literature, and is thus our citation network.

To begin to understand the citation network representing the firing-rate model literature, we first visualize the network. To do so, we wish to take advantage of the specific properties of citation networks. Namely, each paper was published on a specific date, and all papers which cite the paper

[8]Web of Science, while to our knowledge the only widely available tool for exporting the reference lists of papers, does not list certain articles. For instance, it does not contain information about the foundational firing-rate model of da Silva et al. [DSHSZ74], nor of Knight et al. [KMS96]. Consequently, those papers are not included in our analysis of the citation network. For future analysis of the literature of firing-rate models, we aim to include information about key papers missing from the Web of Science database.

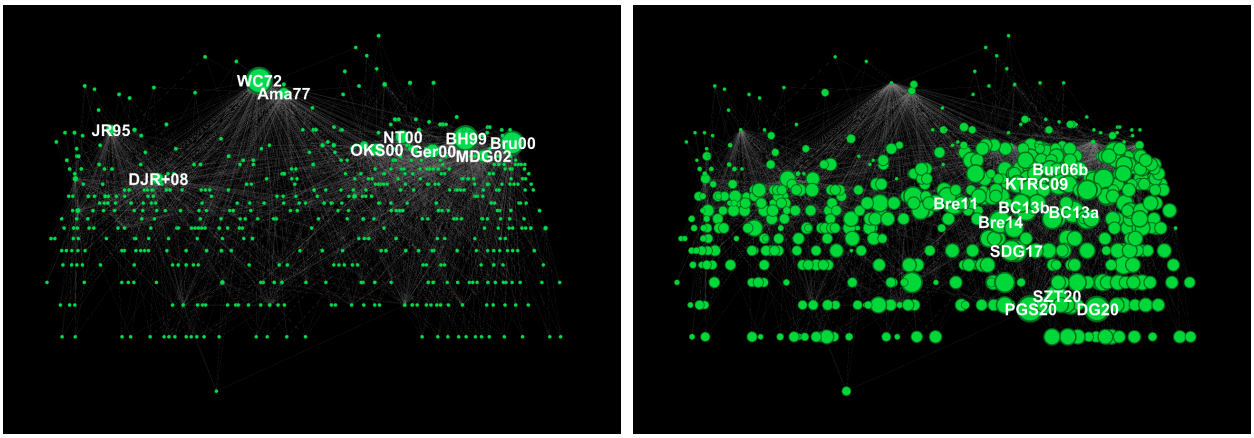
are published at later dates. Thus, if we place papers published earlier in time towards the top of the network, the flow of information from one paper to the next will be down. We can visualize the network with papers published earlier near the top of the page via the "Ordered Graph Layout" plug-in on the network visualization app, Gephi. We show the resulting visualization in Figure I.11.

From the visualization, we can rapidly identify key features of the citation network. For instance, a node in the graph with many edges exiting it represents a paper that was cited by many other papers in the network. Clearly some papers, such as the Wilson-Cowan model near the top center of each of the visualizations in Figure I.11, have many edges exiting it, and are thus well cited by other papers in the firing-rate model literature.

However, we can characterize the importance of papers a bit more cleverly. Namely, we can use two measures of importance that are geared specifically for directed graphs: hub centrality, and authority centrality [New18]. Briefly, "hubs" are papers which cite other hubs and "authorities" in the network. Thus, hubs provide information about key results in the literature. Well-known review papers would be an example of a hub. Authorities, on the other hand, are cited by key hubs and other authorities. Thus, authorities are the key results in the literature. The Wilson-Cowan model, would thus be an example of an authority in the firing-rate model literature. Using an approach similar to Google's PageRank algorithm, we can rank papers according to their hub centrality and authority centrality (see [New18], for example). Papers with higher hub centrality are thus more prominent hubs, and papers with higher authority centrality are more prominent authorities.

We compute the hub and authority centrality for each paper in the network and rank them in Table I.1. In the visualizations, we enlarge nodes with higher authority centrality in Figure I.11.a and higher hub centrality in Figure I.11.b. Unsurprisingly, we see that a number of the foundational papers discussed in the preceding section are authorities, with the Wilson-Cowan [WC72] model being the top authority, and the Amari model [Ama77] and Jansen and Rit [JR95] models appearing in the top 10. Surprisingly, the original population density model from Amit and Brunel [AB97] does not make the top 10. Instead, in the top 10, we see mainly papers from the early 2000s that advanced and/or popularized the population density approach. That the top "authorities" are dominated by firing-rate models taking the population density approach likely reflects in part the popularity of the approach, especially in the 2000s, shortly after its advent. We also speculate that many of the papers on the population density approach are focused on model analysis and developing new formalisms, and are thus perhaps more likely to cite and be cited by a larger number of papers specific to the firing-rate modeling literature. As a result, those papers would be more likely to be more central to the literature of firing-rate models.

(a) (b)



(c)

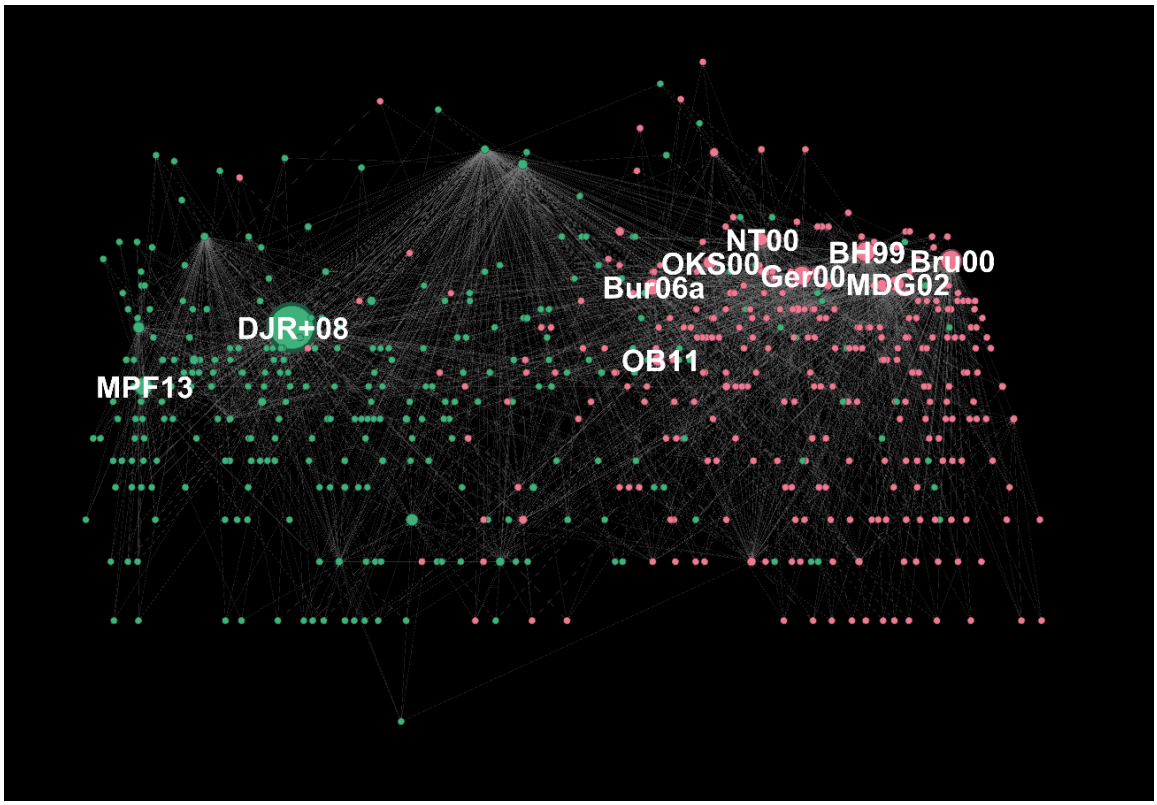


Figure I.11: Citation network for papers closely tied to firing-rate model methodologies. Each panel shows the citation network for papers closely tied to firing-rate model methodologies, where papers near the top were published earlier in time, and node size indicates a key characteristic of the paper. **(a)**: the top 10 firing-rate model authorities, where larger node size corresponds to a higher authority score, and the top 10 authorities are labeled. Authorities are papers which are well-cited by other important authorities and key review papers. **(b)**: the top 10 firing-rate model hubs, where larger node size corresponds to a higher hub score, and the top 10 hubs are labeled. Hubs are papers that cite the key authorities and other important papers. **(c)**: the top 10 firing-rate model papers in terms of betweenness centrality, where larger node size corresponds to higher betweenness centrality. Papers with high betweenness centrality scores lie on many shortest paths in the citation network, meaning that they are key in passing information from papers published earlier in time to papers published later. Papers are also colored according to membership in clusters in the network, one cluster (green), consisting mainly of Wilson-Cowan based firing-rate models, and another cluster (purple), consisting mainly of population density type models.

The hubs, on the other hand, are not the foundational papers. Instead, they are more recent papers (from 2006 and later). Being more recent papers, they can cite many of the key papers from the 2000s that arose from the foundational papers. Moreover, many of the hubs are review papers, such as the review papers from Buice and Chow in 2013, [BC13b][BC13a]. The main subjects of these hubs also span the set of topics in the firing-rate model literature, discussing neural field models, and more.

We further consider a measure of centrality used less frequently in the citation network literature: the betweenness centrality. The betweenness centrality of a node relies on the concept of shortest paths in a network. Namely, a path between two nodes in a network involves starting at one node, and following edges (in the proper direction, in the case off directed graphs), without revisiting any node or edge, until one reaches the other node. The "shortest path" is thus any path between the two nodes that uses the fewest number of edges [New18]. The betweenness centrality, in particular, is determined by the number of shortest paths between other nodes that the node lies on [New18]. Just as we did for hub and authority centrality, we rank papers according to their betweenness centrality (Table I.1), and in the network visualization we enlarge nodes that rank higher in betweenness centrality in Figure I.11.c.

To illustrate the importance of betweenness centrality, consider the review paper from Deco et al. in [DJR$^+$08]. The paper is no longer a hub, because it doesn't cite any of the many papers published after 2008. Further, while the paper is an authority, it is not one of the very top authorities–it is a review paper, after all. On the other hand, it has likely been important for facilitating the flow of information through the firing-rate model literature. Indeed, it contains sufficient details, that one might only need to read it to learn how to use the population density approach, rather than the papers from Amit and Brunel or from Knight et al. that introduced the population density approach. Likewise one might be able to read the Deco et al. review to learn about the Wilson-Cowan model, rather than reading the Wilson-Cowan model paper itself. As such, the Deco et al. review paper would lie on a large number of shortest paths from the Wilson Cowan model to other papers, as well as from the papers on population density from Amit and Brunel or from Knight et al. Thus, the Deco et al. review would have high betweenness centrality. Indeed, the Deco et al. review is the top ranked paper in the network in terms of betweenness centrality.

Firing-Rate Paper Authors	Authority rank	Betweenness rank	Hub rank	Modularity class	Review paper?
Deco et al. 2008 [DJR+08]	8	1		0	yes
Brunel 2000 [Bru00]	3	2		1	no
Brunel & Hakim [BH99]	2	3		1	no
Gerstner 2000 [Ger00]	4	4		1	no
Moran et al. 2013 [MPF13]		5		0	yes
Burkitt 2006 [Bur06a]		6		1	yes
Mattia & Del Giudice 2002 [MDG02]	7	7		1	no
Nykamp & Tranchina 2000 [NT00]	5	8		1	no
Omurtag et al. 2000 [OKS00]	6	9		1	no
Ostojic & Brunel 2011 [OB11]		10		1	no
Wilson & Cowan 1972 [WC72]	1			0	no
Amari 1977 [Ama77]	9			0	no
Jansen & Rit 1995 [JR95]	10			0	no
Pietras et al. 2020 [PGS20]			1	1	no
Burkitt 2006 [Bur06b]			2	1	yes
Dumont & Gabriel 2020 [DG20]			3	1	yes
Buice & Chow 2013 [BC13b]			4	1	yes
Bressloff 2011 [Bre11]			5	0	yes
Bressloff 2014 [Bre14]			6	1	yes
Buice & Chow 2013 [BC13a]			7	1	yes
Schwalger et al. 2017 [SDG17]			8	1	no
Kovačič et al. 2009 [KTRC09]			9	1	no
Shao et al. 2020 [SZT20]			10	1	no

Table I.1: Top 10 firing-rate models by rank in authority centrality, betweenness centrality, or hub centrality. We also show the modularity class of the paper. Roughly, class 1 consists of papers that are mainly concerned with the population density approach (as in Section 1.4.4), and class 0 is comprised of all other papers. We also indicate whether or not the papers are review papers.

Thus, papers with high betweenness centrality have been critical for facilitating the flow of information through the firing-rate model literature. We see that many of the papers with high betweenness centrality are neither the foundational models from early in the development of firing-rate models, nor are they from the last 10 years, for the most part. Indeed, the foundational models have been expanded upon enough that they are not directly used in applications as frequently today, nor have the recent papers had time to accumulate enough citations to lie on many shortest paths. Instead, the papers with high betweenness centrality consist largely of papers from the 2000s, and as a result, largely concerning the population density approach, which had just been developed and was rapidly expanding as a subfield. Nevertheless, we see that the paper with the 5th highest

betweenness centrality is the review of firing-rate models in dynamic causal modeling from Moran et al. in 2013 [MPF13], which has likely been the introduction and main source of information for researchers new to dynamic causal modeling.

These papers with high betweenness centrality are thus likely to be good for learning about the field. Hubs, on the other hand, are good places to look for references and better understand the current state of the field, whereas authorities are likely to be foundational papers that started or popularized a new style of firing-rate formalisms. In any case, all of these centrality measures can help us decide which papers in the field are important. However, these centrality measures miss some information about the field as a whole.

Indeed, to gain information about the field as a whole, we note that while the density of papers in the network visualization is approximately constant as one moves along the vertical axis. Because the vertical axis is on a logarithmic scale with respect to time, the roughly constant density indicates that the number of firing-rate papers is growing roughly exponentially. Indeed, we see from Figure I.12 that the number of papers published per year has grown very rapidly, although not quite exponentially, since the 1980s. Firing-rate modeling is a rapidly growing field.

To further gain information about the field, we can turn to a clustering analysis. To do so, we use the modularity algorithm of [BGLL08] (with resolution 1.5) as implemented in Gephi. Upon repeated iterations under a variety of choices of the resolution parameter for the algorithm, we find that there are roughly two communities in the citation network, colored green and purple, respectively in Figure I.11.b. The purple community consists mainly of models in the population density approach, whereas the green community consists mainly of models using other approaches, indicating that the population density approach has developed somewhat independently from other firing-rate models. Moreover, papers on coupled oscillator reductions are largely absent from the network, indicating a schism in terms of the firing-rate model literature, between coupled oscillator models, and those that are not coupled oscillator models.

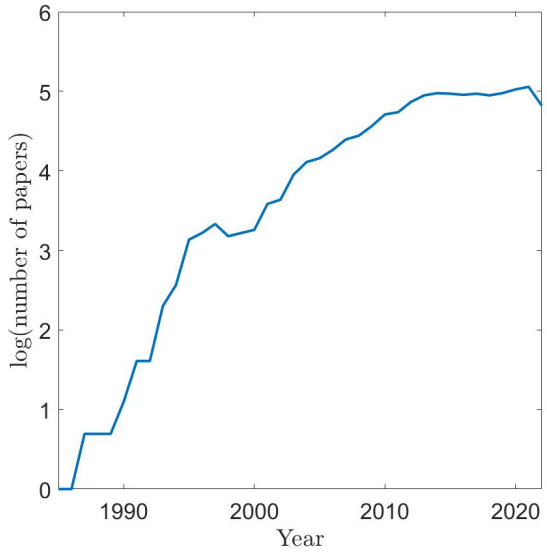


Figure I.12: Growth in the literature of firing-rate models. A log-linear plot of the number of papers identified to be closely related to firing-rate models published over the last five years, from 2022 to 1985. Our analysis was largely conducted before 2022 and thus has only one paper published in 2022 or later, resulting in the dip in the five-year moving-average at the right of the plot. If the growth in the number of papers were exponential, then the plot shown here would be a straight line. We see however, this is not quite the case.

1.5.1 Classifying firing-rate models

In addition to classifying firing-rate model studies by their citation properties, we can classify them by the properties they account for in neuronal networks. Some properties that are of particular interest to the projects in this thesis are:

1. What statistical quantities are modeled? Namely, does the model describe only mean firing rates, as in the Wilson-Cowan model? Or, does it describe means and standard deviations, or whole probability distributions? We say a model is a "neural mass model" if it treats a population of neurons as a point mass and only examines means. On the other hand, if the model considers means and at least one higher-order statistic such as variance, we call it a mean-field model. Finally, if a model describes the probability distribution of features of neurons (namely, firing-rates), and in particular if it involves Fokker-Planck equations, we call the model a population density model.

2. Does the model account for spatial evolution of neural properties in addition to evolution over time? If the model variables evolves over space, we call it a neural field model.

3. What type of underlying neuronal dynamics are assumed? For instance, does the model assume threshold neurons, as does the Wilson-Cowan model? Or does the model use the integrate-and-fire neuron, as is common for population density models. Or does the model assume more complicated neural dynamics, including Hodgkin-Huxley based dynamics, such as our model that we present in Chapter III?

4. What heterogenieties in the underlying neural network does the model consider? For instance, is there any heterogeneity in the underlying network structure, or are all neurons simply

connected to all other neurons? Or is there any heterogeneity in neuron firing properties, e.g. firing thresholds as in the Wilson-Cowan model?

5. What is the methodology for reducing network dynamics? For instance, does the model methodology express firing-rates or average membrane voltages via a convolution, as in the da Silva et al. model [DSHSZ74] or in the Amari model [Ama77]? Or does, the model use an averaging-based approach, such as in the model from Zandt [ZVvPTH14], or does it use a Fokker-Planck equation (FPE) approach?

The results of our efforts to classify firing-rate models according to the above criteria is summarized in Table I.1.

Firing-Rate Paper Name	Stat. Quant.s Modeled	Accounts for Space?	Underlying Synaptic Dynamics	Underlying Neuronal Dynamics	Where's the Heterogeneity?	Methodology
Wilson-Cowan [WC72]	Means	No	Linear summation	Threshold activation	Syn. connections/firing thresholds	Temporal coarse-graining
Amari 1977 [Ama77]		Yes	Linear Summation	Threshold activation	Weights	Sigmoidal Gain, Reduction to region boundaries
Crodelle et al. 2017		No	Linear summ.	Not treated	Noise, subpop.s	Convolutional w/ Sigmoidal Gain fun. (WC-like)
Da Silva et al. 1974 [DSHSZ74]		No	Not treated	Not treated	Multiple subpop.s	Heuristic w/ gain functions, Laplace transforms
Bhattacharya et al. 2011		No	Linear sum + α syn.	Not treated	Multiple subpop.s	Heuristic w/ Sigmoidal Gain Fun. (Da Silva et al. - like)
Heiberg et al. 2018 [HKT$^+$18]		No	Linearized activ. fun.	Izhikevich/MAT	Synapse strength, noisy inputs	Convolutional w/ Sigmoidal Gain Fun. (WC-like)
Ben-Yishai et al. 1995 [BYBOS95]		Yes	cosine coupling	stoch state switching	spatial preference	neural field
Bressloff et al. 2000 [BBC00]		Yes	variable	IF	spatial preference	neural field
Zandt 2014 [ZVvPTH14]	Multiple moments	No	α synapse	HH-like	Noise, 2 subpop.s	Averaging
Ginsberg & Booth [GB23]		No	α synapse	HH-like	Firing Regimes, HH parameters	Averaging
Amit & Brunel 1997 [AB97]	Whole PDFs	No	Exponential synapse	IF	Weights	Population density
Knight et al. 1996 [KMS96]		No	Linear Summation	LIF Neurons	Noise, multiple subpop.s	Population density
Brunel & Hakim [Bru00]		No	Linear Summation	LIF Neurons	Noise, multiple subpop.s	Population density
Brunel 2000 [Bru00]		No	Linear Summation	LIF Neurons	Noise, multiple subpop.s	Population density
Stinchcombe & Forger 2016 [SMW$^+$17]		Yes	Additive	Unspec. Oscillator	g_{CE}/α_E	Population density particle method
Chizhov & Graham 2007 [CG07]		No	α synapse	Simplified HH-like	Weights, noise	Population density refractory approach
El Boustani & Destexhe 2009 [EBD09]		No	exponential, α, etc.	LIF Neurons	Noise, multiple subpop.s, syn. connec.s	population density master equation

Table I.2: A summary of selected firing-rate models. For each of the firing-rate papers in the left column, we identify the statistical quantities modeled (2$^{\text{nd}}$ column from the left), whether it accounts for space (3$^{\text{rd}}$ column from the left), the underlying synaptic dynamics (4$^{\text{th}}$ column) and neuronal dynamics (5$^{\text{th}}$ column) assumed by the model, the sources of heterogeneity (6$^{\text{th}}$ column), and the methodology used to derive the model (7$^{\text{th}}$ column).

1.5.2 Challenges for firing-rate models

Like all models, firing-rate models make assumptions about the network of neurons that can introduce limitations in the usefulness of the model. For example, most models reviewed here have assumed very simple neuronal dynamics. This raises the question: can we construct firing-rate models that incorporate complicated neuronal dynamics? I.e. can the models incorporate behaviors more complicated than leaky-integrate and fire dynamics, such as Hodgkin-Huxley like dynamics? Another limitation of many models in the literature is that they assume all-to-all coupling of neurons or purely random coupling (as in Poisson random graphs). While all-to-all coupling and purely random coupling can be fruitful mathematically, they are fairly unrealistic of graphs in the real

world [New18] and in the nervous system (see [BS09]). Can we thus address complicated network architecture, and understand non-Poisson or non-all-to-all connectivity? Further, many models only address heterogeneous neurons by dividing networks into groups of homogeneous neurons. While such an approach may be realistic for the cortex (see e.g. [WC73]), this may not be realistic for other brain regions, as we will see in Chapter III.

The next generation of firing-rate models (see [SC19, BOF+20], has made progress addressing these challenges. For example, the Ott-Antonsen procedure produces an exact description of the dynamics of a network of coupled oscillators, and can be equipped to describe firing-rates (see [BGLM20] for review). Meanwhile, Zandt et al. [ZVvPTH14] incorporate more realistic neuronal dynamics by replacing the sigmoidal gain function with frequency-current curves computed from Hodgkin-Huxley-type model neurons by using an averaging procedure. Additionally, Byrne et al. [BOF+20] replace the sigmoidal gain functions with a Gaussian gain function using Kuramoto order parameters from coupled oscillator theory. In this thesis, we make our own efforts towards overcoming some of the challenges faced by firing-rate models (see Chapter III).

1.6 Applications of firing-rate models

Despite the challenges faced by firing-rate models, they have been used widely and successfully in computational neuroscience to understand the dynamics of networks of neurons. Indeed, much of the original firing-rate model literature was developed to model cortical neural activity [WC73] [Ama77]. Moreover, the Jansen-Rit model and the population density models have been used to understand EEG data, which reflects cortical activity as well as activity from deeper regions of the brain [SCH+19]. Further, firing-rate models are the current modeling tool of choice by The Virtual Brain Project, an international collaboration attempting to model the full human brain via an EEG-like representation [FJS16], and has allowed for clinical trials of patient-specific treatments for epilepsy [JWT+23].

Firing-rate models have also been used to understand other types of brain recording data. These types include electrocorticogram [HLVB+13] data which records brain activity via electrodes placed on the cortex rather than onto the scalp, and has produced results helpful for studying epilepsy; magnetoencephalography data [MPF13] which measures the magnetic fields produced by the brain; and functional magnetic resonance imaging [GF11] which measures changes in blood flow in the brain.

Further, firing-rate models are the basis for an important tool for understanding cycling between sleeping and waking [PR07]. Of particular interest to this thesis, firing-rate models have been used to understand how feedforward neural circuits in the spinal cord process painful stimuli [CPHB19]

(as discussed in Chapter II), and for understanding signals relayed by the master circadian clock in the brain (as discussed in Chapter III).

1.6.1 Firing-rate models for feedforward networks

As described in Chapter II, we use firing-rate models to analyze the mechanisms that induce pain-processing circuits in the dorsal horn of the spinal cord to produce chronic pain. It is widely believed that for spinal cord circuits to function properly, there must be proper balancing between excitatory signals that induce firing activity that is transmitted to the brain to indicate pain, and inhibitory signals that reduce firing activity in the circuit. Indeed, it is further believed that dysregulation of this excitatory-inhibitory balance is a primary culprit behind a common manifestation of chronic pain, known as allodynia. In allodynia, innocuous mechanical stimuli such as gentle pressure or light brushing on the skin can induce pain. Chronic allodynia is pervasive, affecting over one million[9] people in the United States alone. Therefore, there is significant motivation to be able to treat and/or prevent allodynia. To be able to do that, it is vital to understand how exactly excitatory-inhibitory balance is dysregulated in spinal cord pain-processing circuits so as to induce the circuits to erroneously relay painful signals towards the brain.

Allodynia comes in a variety of forms, and there is high interindividual variability in the success of various treatments for chronic pain. Indeed, the particular ways in which the loss of excitatory-inhibitory balance generates allodynia and the variability therein are poorly understood.

Nevertheless, because excitatory-inhibitory balance occurs at the level of populations of neurons, firing-rate models are well suited for investigating its dysregulation. Further, many recent experimental observations on the generation of allodynia in the spinal cord dorsal horn pertain to how changes in whole populations of neurons, such as ablation of certain types of neurons, can induce allodynia.

1.6.2 Firing-rate models for heterogeneous brain regions

Another brain region for which firing-rate models are particularly relevant is the suprachiasmatic nucleus (SCN)–the central circadian pacemaker in mammals–located in the hypothalamus. The SCN exhibits a 24 h variation in its average neuron firing rates that is assumed to be the primary driver for transmission of daily circadian timing information to downstream targets [JTM15, HMB18,

[9] Allodynia is estimated to affect $15 - 50\%$ of people experiencing neuropathic pain [JF14], a type of pain caused by damage or disease to the nervous system. In turn, chronic neuropathic pain is believed to affect $3 - 17\%$ of the the population worldwide [VHAK$^+$14]. That would suggest chronic allodynia would affect $0.45 - 8.5\%$ of the population. Even the lower bound of that estimate would place the number of people people in the United States that experience chronic allodynia at over a million.

LS98][10].

This transmission is critical for a variety of 24hr bodily rhythms collectively termed circadian rhythms (circa meaning about, and dies meaning a day). Indeed, SCN transmission of the circadian rhythm in combination with a homeostatic drive to enter sleep coordinate the timing and duration of sleep and wake states (see, e.g. the two-process model of sleep-wake regulation, as in [Bor22]). Other bodily circadian rhythms coordinated by SCN transmission range from variations in core body temperature to endocrine processes such as variations in appetite, to rhythms in kidney and liver function (see e.g. [FK17]).

While new and updated firing-rate model formalisms have expanded the accuracy and range of applications of firing-rate models, recent results have identified properties of SCN neurons and their network that are not easily accounted for by these models. In particular the following neural properties all vary across the SCN and over time (Figure I.13): resting voltages, the presence of spikes, and the size of spikes [BDFP09]. As a result, synaptic signaling varies as well across the SCN network and over time.

Thus, to model the SCN rate code for the circadian rhythm, there is a need to develop firing-rate model formalisms capable of accounting for the variety of heterogeneities in the SCN network. In Chapter III, we present the derivation of a new firing-rate formalism that meets this need, and apply it to model rhythms in SCN firing rates.

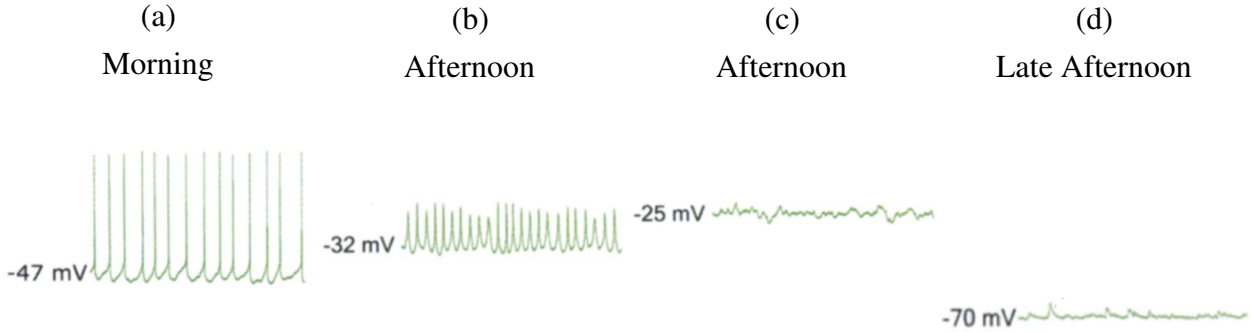


Figure I.13: SCN electrophysiological activity varies across the network and across the day. In the morning **(a)**, neurons typically display action potential firing. In the afternoon **(b)**, neurons exhibit membrane oscillations that are lower-amplitude than action potentials, and also unlike action potentials, consistently remain depolarized. At the same time, in the afternoon **(c)** neurons exhibit depolarization block, where they are depolarized but their membrane potentials roughly constant. In the late afternoon into the evening **(d)**, neurons are usually at rest, with low resting membrane voltages. Images source: [BDFP09].

[10]Although, there is also evidence of humoral signaling [LS98, YTLS21], i.e. signaling where hormones released by neurons enter the blood. In fact, according to [YTLS21], the SCN and the pituitary gland are the only brain regions involved in "brain portal pathways", where blood capillaries can convey high concentrations of hormones to specialized organs in the brain without diluting the hormones in the general circulatory system.

1.7 Overview

This thesis is organized as follows: in Chapter II, we introduce a methodology for analyzing small networks of firing-rate models in the context of pain processing in the spinal cord. We then use the methodology to uncover how disruptions of balance between inhibition and excitation in spinal cord circuit motifs can produce chronic pain involving allodynia. In Chapter III, we describe the need for new firing-rate model formalisms that can study heterogeneous networks of neurons, develop such a formalism, and apply it to the SCN. In Chapter IV, we use Chapter III's firing-rate model formalism to model the effects of changing the external light cycle on the firing rate signals sent downstream by the SCN that coordinate circadian clocks throughout the body. In particular, we show that doing so can add information not provided by existing models of the pathway by which the external light cycle affects the output from the SCN. We then establish a new link between the electrophysiological state of SCN neurons and the intracellular molecular circadian clocks keeping track of time in SCN neurons. In Chapter V, we put the work of this thesis into a broader perspective, and discuss future directions for our work. In Appendix A, we show supplementary figures for the sensitivity analysis of Chapter II. In Appendix B, we show additional figures and methods for the firing-rate model we develop in Chapter III.

CHAPTER II

New Parameter Sensitivity Analysis Methodology Applied to Neural Circuit Models in the Spinal Cord Dorsal Horn

2.1 Introduction

Chronic pain troubles over 50,000,000 Americans, according to the United States Center for Disease Control and Prevention [RSSGJ23], and estimates place the worldwide prevalence of chronic pain at around 30% [CVH21]. As such, chronic pain poses a significant economic burden, costing the United States alone around $600 billion in 2010 (chapter 2 of [Sim12]) and is often disabling (chapter 2 of [Sim12]). Thus, there is a pressing need to understand, treat, and prevent the conditions that lead to chronic pain.

2.1.1 Pain-processing in the spinal cord

Understanding pain-processing circuitry in the spinal cord is vital for understanding the physiological mechanisms responsible for chronic pain conditions [PWZ+15, PDT20]. Indeed, the spinal cord is responsible for the initial processing of both tactile and pain-inducing stimuli at the periphery, and then for relaying them towards the brain [BCP20, Tod10, Lec17]. In particular, tactile and pain-inducing signals travel from the periphery to the spinal cord along different classes of afferent nerve fibers [BCP20, Tod10, Lec17]. Normally, $A\beta$-fibers transmit non-painful signals due to innocuous stimuli such as gentle pressure or the brush of clothing on skin [BCP20, Tod10]. On the other hand, C-fibers, and to some extent $A\delta$ fibers [BCP20, Tod10], transmit pain-inducing signals in response to heat, noxious chemicals, or intense mechanical stimuli (Figure II.1). These signals, upon arriving at the spinal cord, are filtered by intermediate circuitry in superficial laminae (primarily laminae I and II) of the spinal cord's dorsal horn [DCM18]. From there, they are relayed towards the brain by projection neurons often in lamina I (but also in deeper laminae, namely laminae III and IV), of the dorsal horn [DCM18].

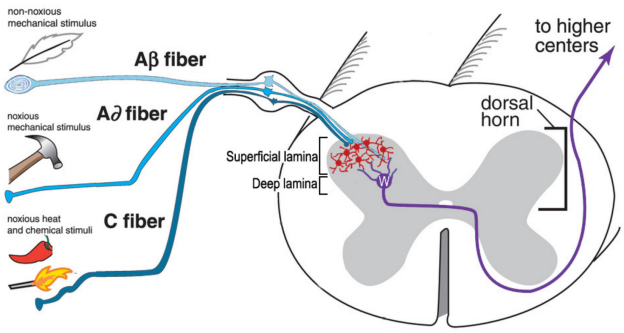


Figure II.1: Overview of processing of painful signals in the spinal cord. $A\beta$ fibers signal non-painful tactile stimuli. $A\delta$ fibers transmit pain from noxious mechanical stimuli. C-fibers transmit pain inducing stimuli. Intermediate circuitry in the superficial laminae of the spinal cord dorsal horn processes such signals, and relays them towards the brain via projection neurons in lamina I of the dorsal horn. Image source: [BSY+19]

It is widely believed that normal processing of afferent signals in the pain-processing circuit of the dorsal horn relies on a balance between excitation and inhibition [Woo22, PWZ+15, DCM18] as manifested by the conceptual model of "gate control", originally due to Melzack and Wall in 1965 [MW65]. While the original formulation of "gate control" has failed to keep up with the knowledge gained since 1965 about pain processing circuits in the spinal cord, "gate control" remains influential, and certain aspects of the theory remain relevant [BSWB14, Woo22, DCM18]. Namely, inhibitory neurons "gate" activity of excitatory neurons that relay pain signals towards the brain [PWZ+15]. In particular, the paradigm states that $A\beta$ stimuli excite both excitatory and inhibitory interneuron populations, but that the inhibitory populations can be sufficiently activated so as to "close" the gate, sufficiently inhibiting the excitatory populations so as to block responses to pain-inducing signals from C-fibers.

Thus, it should not come as a surprise that some pathological change in spinal cord circuitry is frequently the culprit behind chronic pain. In particular, two types of chronic pain, neuropathic pain–pain arising from damage or disease of the nervous system [CVH21]–and pain arising from chronic inflammation [TC10], are associated with changes in spinal cord circuitry. Such changes in spinal cord circuitry are termed central sensitization, and involve long-term alterations in the properties of individual neurons, synapses, glial cells, and even the overall network structure of the dorsal horn circuitry [TC10]. Such diverse alterations combine to disrupt the excitatory-inhibitory balance in spinal-cord circuitry, disrupting inhibitory control of excitatory cells.

In allodynia, a type of chronic pain frequently associated with alterations to the spinal cord pain-processing circuitry [TC10], individuals feel pain in response to normally innocuous stimuli [HK21]. Allodynia is classified according to the type of innocuous stimuli that is painful [HK21]. In this work, we focus on two well-studied types of allodynia: static and dynamic [San09, DCB+14]. In static allodynia, individuals feel pain in response to gentle pressure that normally wouldn't be

painful [DCB$^+$14]. In dynamic allodynia, on the other hand, individuals feel pain in response to brushing sensations, e.g. the brushing of clothing on skin, that again normally wouldn't be painful [DCB$^+$14].

Recent rodent experiments [DCB$^+$14, PPF$^+$15, CDH$^+$17, PBT$^+$19] have shown that either type of allodynia can be induced by activating or inactivating (e.g. ablating) specific populations of excitatory or inhibitory interneurons in the dorsal horn. In clinical conditions, allodynia likely occurs through more subtle circuit disruptions. Since E-I balance in a circuit can be achieved by diverse contributions of different excitatory and inhibitory circuit components, its disruption leading to allodynia can potentially occur through multiple pathways. In this work, using biophysically-based mathematical modeling, we develop a methodology to identify likely mechanisms by which spinal cord dorsal horn circuits may be dysregulated to produce allodynia. We construct models of neural circuits associated with static and dynamic allodynia and constrain model parameters to obtain E-I balance in normal conditions. These constraints result in distributions of parameter sets representing circuits that achieve E-I balance in different ways. We then identify the sensitivity by which E-I balance can be disrupted to result in allodynia in the circuit population. We find that the particular means of disruption varies across the circuit population, thus suggesting multiple likely mechanisms responsible for allodynia.

2.1.2 Proposed pain-processing circuitry in the dorsal horn mediating allodynia

We first identify circuit motifs that reflect recent experimental evidence for the structure of dorsal horn layer I-II networks mediating static allodynia[1] and dynamic allodynia (Figure II.2). In particular, the aforementioned experiments probing these circuits in rodents have shown that static allodynia is reliant on activity in somatostatin-positive (SOM+) [DCB$^+$14], calretinin-positive (CR+) [PBT$^+$19], and protein kinase c γ-positive (PKCγ+) [PPF$^+$15] excitatory interneurons. Inactivation of either dynorphin-positive (DYN+) [DCB$^+$14] or parvalbumin-positive (PV+) [PPF$^+$15] inhibitory interneurons is sufficient to produce static allodynia, suggesting that these inhibitory cells usually gate the activation of SOM+, CR+ and PKCγ+ excitatory cells. That these inhibitory cells gate those excitatory cells is further evidenced by direct synaptic connections between PV+ and PKCγ+ neurons [PPF$^+$15].

We consider a simplified circuit motif (Figure II.2b) consisting of three populations. One

[1] In the literature, there is some confusion as to what "static allodynia" means. The confusion lies in the types of innocuous stimuli that evoke pain. Some sources such as [DCB$^+$14] and [PBT$^+$19] allow static allodynia to be evoked from a "von Frey device", which is essentially a narrow-tipped probe that puts pressure on a very small area of skin. Other sources, however, such as [CDH$^+$17], term that type of allodynia instead as "punctate" allodynia, and only term the allodynia as "static" if it is evoked from a broad-tipped rather than narrow-tipped probe. The sources we use to study "static" allodynia–[DCB$^+$14, PPF$^+$15, CDH$^+$17, PBT$^+$19]–all use the von Frey device.

population, E, consists of excitatory interneurons, representing the collective activity of SOM+, CR+ and PKCγ+ excitatory cells in layer I dorsal horn that is inhibited by two distinct inhibitory interneuron populations, I_1 representing DYN+ cells and I_2 representing PV+ cells. All three populations receive inputs from $A\beta$ fibers, (as suggested in [DCB$^+$14, HSK$^+$12, PBT$^+$19]), while the E population is presumed to be additionally targeted by C fiber input (see e.g. [DCB$^+$14, PBT$^+$19]).

In contrast, dynamic allodynia relies on the activity of SOM+/CR- excitatory interneurons. Indeed, ablating neurons that express SOM "abolishes or greatly reduces" dynamic allodynia in mice [CDH$^+$17]. Thus SOM+ neurons are necessary for dynamic allodynia. Further, ablating neurons expressing the calbindin 2/calretinin gene had no effect on dynamic allodynia in mice [CDH$^+$17]. Thus, the SOM+ neurons that are necessary for dynamic allodynia must also be negative for calretinin (CR), hence they are SOM+/CR- neurons.

Dynamic allodynia also is related to DYN+ inhibitory neurons, just as is static allodynia. Namely, ablation of DYN+ inhibitory neurons is sufficient to induce dynamic allodynia [DCB$^+$14], suggesting that these inhibitory cells usually gate the SOM+/CR- excitatory interneurons. Inhibitory neurons, likely including DYN+, also gate VT3+ neurons [CDH$^+$17]. Indeed, when inhibitory signalling between neurons is blocked, VT3+ neurons begin to fire action potentials in response to $A\beta$ input.

Further, even when DYN+ neurons are still intact, either ablating or silencing neurons expressing vesicular glutamate transporter 3, (VT3+ neurons), eliminates or attenuates dynamic allodynia induced by nerve injury or ablation in mice [CDH$^+$17]. Thus, at least with DYN+ neurons intact, VT3+ neurons are necessary for dynamic allodynia. It is not clear, however, whether VT3+ neurons are necessary for dynamic allodynia when DYN+ neurons have been ablated. Moreover, it has been proposed that the VT3+ neurons themsleves synapse onto the excitatory SOM+/CR- neurons [CDH$^+$17].

Thus, much as for the static circuit, we have constructed a simplified circuit motif (Figure II.2) consisting of four populations. One is an excitatory population $E2$ representing the collective activity of SOM+/CR- excitatory neurons that are exctited by a population ($E1$) represeting representing VT3+ neurons and inhibited by a population representing DYN+ cells. Meanwhile, $E1$ receives inhibitory signals from a second population $I1$ representing the collective activity of DYN+ neurons that inhibit it. Further, as for the static circuit, all populations receive $A\beta$ input (as suggested in [DCB$^+$14, CDH$^+$17]), and again the E population is presumed to be additionally targeted by C fiber input (see [CDH$^+$17]).

We acknowledge that these simple representations (Figure II.2) of the neural circuitry associated with static and dynamic allodynia do not encompass the full range of neuron types and processes involved in their respective types of allodynias. Thus, our model is more similar to [ZJG14] rather than the detailed model of [MSH+22]. Nevertheless, the simplicity of these representations provides ample opportunities to explore precisely how allodynia may be generated from pain-processing circuitry in the spinal cord.

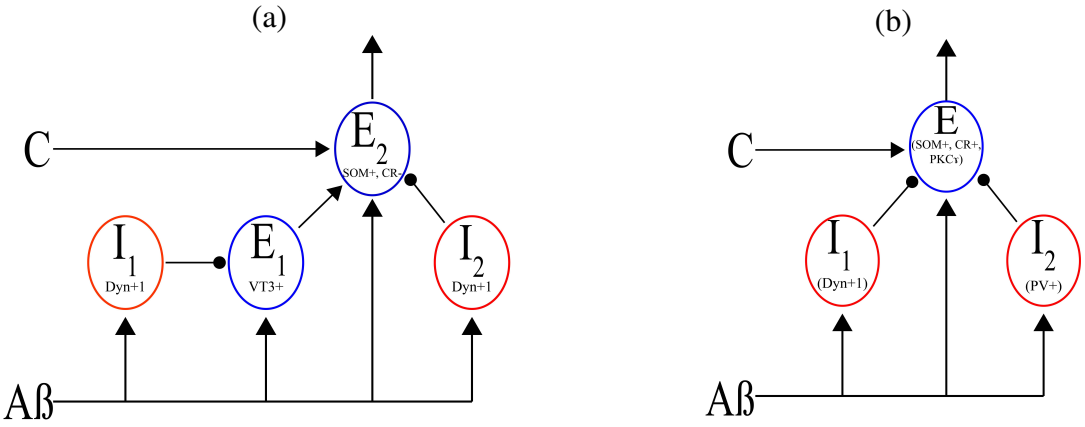


Figure II.2: (a) A schematic of the proposed circuit mediating dynamic allodynia. I_1 and I_2 represent the populations of dynorphin-positive inhibitory neurons, and E_1 and E_2 represent the populations of Vglut3-negative and somatostatin-negative/calretonin-negative excitatory neurons. $A\beta$ and C represent inputs to the circuit relayed from the periphery along $A\beta$ and C fibers, respectively. **(b)** A schematic of the proposed circuit mediating static allodynia. I_1 and I_2 represent the populations of dynorphin-positive and parvalbumin-positive inhibitory neurons, respectively, and E represents the population of excitatory neurons. $A\beta$ and C represent inputs to the circuit relayed from the periphery along $A\beta$ and C fibers, respectively.

2.1.3 Chapter overview

For these circuit motifs, we develop population firing-rate models based on the biophysics of the dorsal horn of the spinal cord. We use these firing-rate models to analyze mechanisms that maintain E-I balance under normal conditions and to analyze mechanisms that lead to the disruption of E-I balance which can occur with allodynia.

In these circuit motifs, as well as in a simple circuit representing a canonical gate control model, we determine parameter sets that replicate behaviors observed under normal conditions and observed in response to experimental manipulations. Specifically, we focus on the parameters representing the synaptic coupling between neural populations in the circuit. We refer to the resulting sets of coupling strengths as the "allowable parameter space". To identify the most likely mechanisms by which E-I balance may be disrupted, we compute the minimal parameter changes needed to produce responses associated with allodynia. Analysis of the minimal parameter changes reveals clustering of the allowable parameter space whereby subpopulations of circuits show different

parameter sensitivities, thus suggesting that the most likely mechanisms for allodynia differ form subcircuit to subcircuit.

Results show that, in the allowable parameter space, a considerable range of parameter combinations can reproduce the experimentally-observed behavior required of the circuit motif under normal conditions. This highlights the diversity of ways that E-I balance can be obtained within a circuit. Results further indicate that across and within circuit motifs, there is considerable variation in the means by which the circuits are most vulnerable to E-I dysregulation that results in allodynia. For instance, in the proposed circuit mediating static allodynia, there are four different means of such dysregulation, represented as a clustering of the allowable parameter space into four regions. Generally, in the static circuit, those means of dysregulation occur primarily via over-excitation of E in response to $A\beta$, along with simultaneous disinhibition of the excitatory populations. In particular, the disinhibition primarily occurs by lowering the activity of the inhibitory population. However, the degree of over-excitation, and the cause of disinhition vary across the corresponding allowable parameter space.

As for the proposed circuit mediating dynamic allodynia, such dysregulation can also occur by four separate mechanisms, illustrated by clustering the allowable parameter space into four regions. However, unlike in the static circuit, these mechanisms can involve the $E1$ population rather than the population ($E2$) which relays pain-inducing towards the brain. Further, these mechanisms occur mainly by over-excitation of $E2$, and disinhibition, without altering the activity of the inhibitory population. Much as for the static circuit, the degree of over-excitation, and the cause of disinhition vary across the corresponding allowable parameter space.

These results identify the diverse ways by which excitatory and inhibitory components within the circuits combine to maintain E-I balance. These results also identify the sensitivity of these combinations to different mechanisms of disruption that can lead to pathological responses.

The chapter is organized as follows: in Section 2.2, we describe and apply our sensitivity analysis methodology to the simple circuit representing a canonical gate control model. In Sections 2.3, and 2.4, we do likewise for the proposed circuits mediating dynamic and static allodynia, respectively. In Section 2.5, we summarize the results and put them into a greater biological context. Finally, in Sections 2.6 and 2.7, we discuss the methods we use to conduct the sensitivity analysis in more detail, and provide supporting proofs, respectively.

2.2 Analysis of a simple "gate control" circuit

Before analyzing the proposed circuits mediating dynamic and static allodynia, we illustrate our methodology by applying it to a simple circuit representing the "gate control" conceptual model. The simple circuit consists of two neural populations: one for inhibitory interneurons (I) and one for excitatory (E) interneurons, where it is assumed that the E population relays pain-inducing signals to projection neurons that transmit towards the brain. Both populations receive input from $A\beta$ fibers. $A\beta$ fiber input is scaled by the coupling strength parameters $g_{A\beta E}$ and $g_{A\beta I}$ between the $A\beta$ fibers and the excitatory and inhibitory population, respectively. Under normal "gate control" conditions, $A\beta$ signals stimulate the inhibitory population sufficiently so that the net signaling to the E population is inhibitory. Should pain inducing signals traveling along C fibers arrive at the E population, the E population would then only relay the pain-inducing signals if the resulting excitation of the E neurons were sufficiently large to overcome the inhibition arising from $A\beta$ signaling. Thus, the $A\beta$ signals "gate" pain-inducing signals relayed by the E population. A schematic of the circuit is shown in Figure II.3.a.

2.2.1 Simple circuit model

In the circuit, we model each population with a firing-rate model. The modeling formalism is based on the work of Crodelle et al. in [CPHB19]. However, unlike [CPHB19], our model uses voltage as the state variable, as in the well-used firing-rate models of Jansen and Rit [JR95] and Phillips and Robinson [PR07]. Modeling the state-variable as a voltage provides certain advantages. Namely, it allows us to replace the heuristic input-output relation from [CPHB19] with a data-driven relationship between the population voltage and population firing-rates.

Specifically, from the average population voltage V_E of some population, e.g. E, we model the average population firing rate, f_E (in Hz), as a sigmoidal function of V_E:

$$f_E(V_E) = 0.5max_E \left(1 + \tanh\left(\frac{V_x - \beta_E}{\alpha_E}\right)\right).$$

In the preceding expression, max_E represents the maximum possible firing-rate of the excitatory population; β_E represents the "half-activation" voltage, at which the population fires at half of its maximum possible firing-rate; and α_E represents the "flatness" of the firing-rate's response to changes in V_E. We determine the parameters of this sigmoidal activation function to match experimentally determined frequency-voltage relationships (see Section 2.6.2 for details).

We expect that in the absence of input, V_E remains at some rest value $V_{E,rest}$, which is also based on experimental data (Section 2.6.2). On the other hand, in the presence of steady inputs, we

expect that V_E eventually deviates from rest by an amount equal to the net input signal:

$$V_E = \text{inputs} + V_{E,rest} \text{ (steady state)} .$$

We take the inputs to be the output firing-rates of populations pre-synaptic to E, for example I, along with the input $A\beta$ firing-rate $f_{A\beta}$, weighted by the corresponding coupling strengths:

$$\text{inputs} = g_{IE} f_I + g_{A\beta E} f_{A\beta}$$

We add time dynamics to the system via the following ordinary differential equation:

$$\frac{dV_E}{dt} = \frac{\text{inputs} - V_E + V_{E,rest}}{\tau_E}. \tag{II.1}$$

where τ_E governs how quickly V_E moves towards its steady-state value.

We model the inhibitory population analogously, except we choose different parameters for inhibitory populations. Namely, in this section and throughout this chapter, we take all parameters for inhibitory populations to be the same and as shown in [RS02], and do likewise for all excitatory populations. Further, in Section 2.6.1, we provide a general description of how we model the inputs, voltage, outputs, and dynamics for an arbitrary population x in an arbitrary circuit. Although, rather than any arbitrary circuit, we expect our model formalism and the methodology, which we will describe shortly, to be most useful for feedforward circuits, such as those presented in this chapter.

2.2.2 Conditions on simple circuit parameters

To constrain the circuit to display realistic behaviors, we impose conditions on the steady-state voltages of the E and I populations. To start, we require that steady-state voltages of the E and I populations remain within reasonable bounds, which we describe further in the methods, Section 2.6.2. Further, under control conditions, we require that the I population fires in response to typical activity on $A\beta$ fibers, which we choose to be $f_{A\beta} \in [10, 20]$ Hz (see Section 2.6.2). Notably, we require that such a response by the I population must happen for all reasonable $f_{A\beta}$. We then require that the signal from the I population relayed to the E population is strong enough that it not only prevents the E population from firing, but that the net signals to the E population in response to any typical $A\beta$ activity are inhibitory, causing the voltage of the E population to drop. Such $A\beta$ signaling would require pain-inducing inputs from other sources to excite the E population more than normal to elicit pain, thereby "gating" the E population response. On the other hand, if the I population is experimentally ablated, then we expect $A\beta$ signaling to cause the E population to fire, relaying pain-inducing signals from normally innocuous input, thereby creating allodynia.

Each such condition leads to an inequality that average voltages must satisfy. To illustrate this, consider the pain-inhibition condition discussed above. Innocuous stimuli make E less likely to fire and transmit painful signals. Thus, typical $A\beta$ firing activity ($f_{A\beta} \in [10, 20]$) should make V_E decrease:

$$\begin{aligned} V_{E,rest} &\geq V_{E,steady} \\ &= \text{ Change in } V_E \text{ due to weighted inputs } + V_{E,rest} \\ &= g_{A\beta E} f_{A\beta} - g_{IE} f_I(f_{A\beta}) + V_{E,rest} \end{aligned}$$

We summarize these conditions and the corresponding inequalities in Table (II.1). These conditions on steady state population voltages can be re-written as a system of 6 inequalities, some nonlinear, on the coupling strengths $g_{A\beta I}$, g_{IE} and $g_{A\beta E}$) (see Section 2.6.3.1).

3-tuples of coupling strengths, $(g_{A\beta I}, g_{IE}, g_{A\beta E})$, which satisfy these inequalities constitute an allowable parameter space (APS), in which we expect parameters to lie for the circuit to function properly. However, because these inequalities must be satisfied for a range of input $f_{A\beta}$ signals, solving them explicitly requires finding the solution to nonlinear optimization problems. For instance, we can rewrite the pain inhibition condition so it is expressed as an inequality on $g_{A\beta E}$:

$$g_{A\beta E} \leq \frac{g_{IE} f_I(f_{A\beta})}{f_{A\beta}}.$$

However, because the pain inhibition condition should hold for all $f_{A\beta} \in [10, 20]$ Hz, the preceding inequality must hold for all $f_{A\beta} \in [10, 20]$ Hz. If the preceding inequality is satisfied for the value of $f_{A\beta}$ that minimizes the right-hand side, the inequality is satisfied for all $f_{A\beta} \in [10, 20]$. Thus, the preceding inequality holds if and only if

$$g_{A\beta E} \leq \min_{f_{A\beta} \in [10,20]} \left[\frac{g_{IE} f_I(f_{A\beta})}{f_{A\beta}} \right].$$

While the solutions of these optimization problems may often be expressed using Lambert functions (see Section 2.6.3.1), the resulting system of inequalities including Lambert function solutions, while considerably simpler, remains difficult to solve explicitly.

2.2.3 Computing the allowable parameter space (APS)

Since the above constraints that define the APS are difficult to solve explicitly, we instead seek a representative sample of the APS, namely one for which sampled points are distributed uniformly in space. By sampling uniformly, we can ensure that statistics of the sample, such as distributions of

Condition Type	Condition	Steady state voltage inequality
Control conditions	V_I upper bound	$V_{I,max} \geq V_I = g_{A\beta I} f_{A\beta} + V_{I,rest}$
	I fires	$V_{I,thr} \leq V_I = g_{A\beta I} f_{A\beta} + V_{I,rest}$
	Pain inhibition	$V_{E,rest} \geq V_E = g_{A\beta E} f_{A\beta} - g_{IE} f_I + V_{E,rest}$
	V_E lower bound	$V_{E,min} \leq V_E = g_{A\beta E} f_{A\beta} - g_{IE} f_I + V_{E,rest}$
I ablation conditions	V_E upper bound	$V_{E,max} \geq V_E = g_{A\beta E} f_{A\beta} + V_{E,rest}$
	E fires	$V_{E,thr} \leq V_E = g_{A\beta E} f_{A\beta} + V_{E,rest}$

Table II.1: Conditions that the simple gate control circuit must satisfy and the resulting inequalities on steady-state voltages. We ensure that the circuit exhibits appropriate behaviors by imposing conditions (middle column) on the circuit. Each condition is exhibited in either control or inhibitory-ablation scenarios (left-most column), and is realized as an inequality (right-most column) on the steady-state voltage of a population. The inequalities in the right-most column must hold for any reasonable $f_{A\beta}$, which we take to be $f_{A\beta} \in [10, 20]$ Hz.

coupling strengths across the parameter space or the proportion of sampled points falling in certain clusters, are representative of the APS as a whole. However, because the APS may be non-convex (see Figure II.3.d), and since to our knowledge there are no widely-used algorithms for uniformly sampling from non-convex solids, we develop a sampling algorithm and use it to arrive at the sample described in Figure II.3.b - e. See the Methods, Section 2.6.4 for a description of the sampling algorithm.

We illustrate that sampled parameter sets correspond to circuits which yield the desired behaviors by simulating the simple circuit model with 20 sampled points, in response to a noisy $A\beta$ signal with mean firing rate in $[10, 20]$ Hz. In doing so, we see pain inhibition in action (see Figure II.3.b). Indeed, in each simulation, the firing-rate of the I populations rises to at least 80% of the maximum firing-rate of 80 Hz for the I population (left panel). The inhibition from the I population then causes voltages of the E population to drop (middle panel), as we require for pain inhibition. Then, as expected, the E population doesn't fire (right panel). On the other hand, if we ablate the I population, we lose pain inhibition (see Figure A.1 in the supplementary material). In this case, the voltage of the E population sufficiently increases so that the E population fires in response to innocuous $A\beta$ stimuli, thus simulating allodynia.

To analyze the distributions of the coupling strengths across the sample, we construct violin plots (Figure II.3.c). In these plots, we see that $g_{A\beta I}$ has the largest range of allowable values (from about 2.6 - 7.1 mV/Hz), with $g_{A\beta E}$ having the second largest range (about 3.5 - 6.9), and g_{IE} having

the smallest range (about 0.9 - 2.1). Because $g_{A\beta I}$ can thus be changed the most without exiting the APS, whereas g_{IE} can be changed the least, they are in some sense the least and most sensitive coupling strengths, respectively, underlying this circuit.

To begin to understand mechanisms that can induce allodynia in terms of relative changes in parameter values, we normalize each coupling strength to the range $[0, 1]$. To do so, we find the distance between the coupling strength g_{xy} and its smallest allowable value, and set the normalized coupling strength $\hat{g}_{xy}$ to be that distance expressed as a proportion of the coupling strength's allowable range:

$$\hat{g}_{xy} = \frac{g_{xy} - \min\{g_{xy}\}}{\max\{g_{xy}\} - \min\{g_{xy}\}}.$$

A change in $\hat{g}_{xy}$ by 0.1, for example, thus corresponds to a change in g_{xy} by 10% of its natural range. By normalizing parameters as such, not only do we control for the effects that having a larger range of values has on parameter sensitivity, but we are better positioned to compare the effects of changing one parameter to the effects of changing another. In Figure II.3.d, we show the full set of normalized sampled points. Notably, we have constructed our sampling algorithm so sets of normalized coupling strengths are uniformly distributed in the unit cube $[0, 1] \times [0, 1] \times [0, 1]$.

From the illustration of the sampled points in Figure II.3.d and from the violin plots in Figure II.3.c, we see that distributions of individual coupling strengths across the APS are skewed towards zero. To understand the degree to which the distrbutions are skewed towards zero, we can examine normalized violin plots (Figure A.2). For instance, if the normalized violin plots are symmetric about 0.5, they are not skewed at all. However, we see that the means of each coupling strength, after normalization, are roughly the same, and are all around 0.4. Thus, the extent to which distributions of coupling strengths are skewed towards zero is roughly the same for all three coupling strengths.

Further, from Figure II.3.d, we see relationships between coupling strengths. For instance, we see that $\hat{g}_{A\beta E}$ has the largest allowable range when $g_{A\beta I}$ is at its smallest. Looking at correlations between coupling strengths (see Figure II.3.e), we see further relationships, namely that when g_{IE} is near the top of its range, $g_{A\beta E}$ is also near the top of its range. This highlights that under normal conditions, the inhibition of E from I is balanced by the excitation of E from $A\beta$ inputs.

2.2.4 Mechanisms for generating allodynia in the simple circuit

Thus far, we have investigated the APS showing which sets of coupling strengths reproduce the desired behaviors necessary for normal E-I balance in the circuit and for the disrupted E-I balance arising upon ablation of the inhibitory population. To analyze vulnerability of these model circuits

to allodynia, we next determine the coupling strength changes that most easily result in E cell firing in response to innocuous input ($f_{A\beta} \in [10, 20]$ Hz).

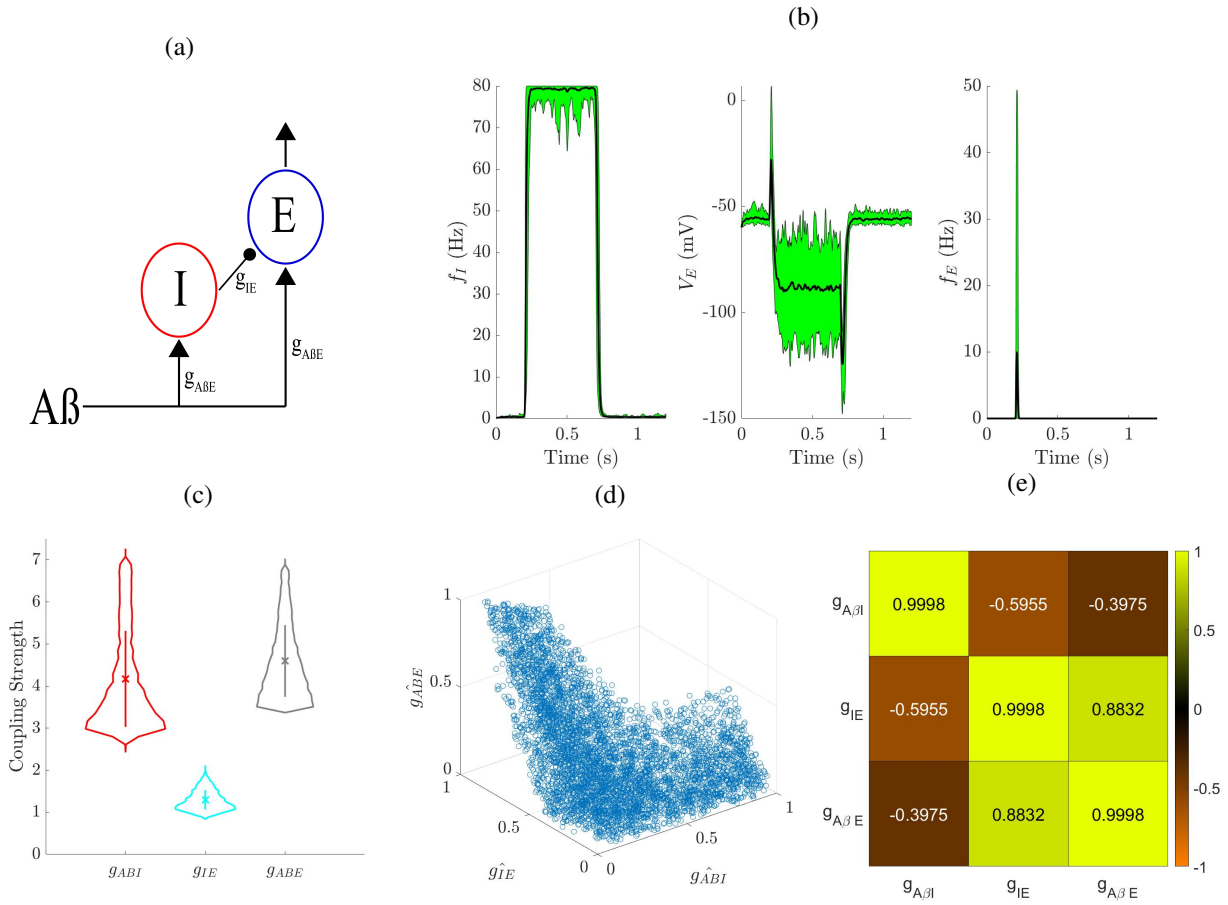


Figure II.3: Allowable parameter space for the simple "gate control" circuit. **(a)** A schematic of the simple "gate control" circuit, where I and E represent the populations of excitatory and inhibitory neurons, respectively. **(b)** Simulation results showing the mean (black lines) and range (shaded green areas) for the firing rate (left) of the I population, as well as for the voltage (middle) and firing rate (right) of the E population, calculated for 20 sampled sets of coupling strengths each with a different random input $A\beta$ stimulus in range $[10, 20]$ Hz. **(c)** Violin plots showing the distribution of each coupling strength, the mean (x-marker), and the range of values that lie within one standard deviation (vertical bar). **(d)** A scatter plot of the 5000 sampled sets of normalized $(g_{A\beta I}, g_{IE}, g_{A\beta E})$ values that satisfy the conditions listed in Table II.1. Samples were computed such that normalized samples are uniformly distributed in the three-dimensional $(g_{A\beta I}, g_{IE}, g_{A\beta E})$-space. **(e)** Normalized Pearson correlation coefficients between coupling strengths among normalized sampled sets of parameters.

Defining allodynia in the simple circuit

In our formalism, allodynia is realized when the steady-state average voltage of the E population, V_E, exceeds the firing threshold for some typical $f_{A\beta}$:

$$V_{E,thr} \leq V_{E,rest} + g_{A\beta E} f_{A\beta} - g_{A\beta I} \cdot f_I(f_{A\beta}).$$

We use this inequality to define a surface in $(g_{A\beta I}, g_{IE}, g_{A\beta E})$ space, which we term the "allodynia surface", above which the corresponding circuit produces allodynia for at least some typical $f_{A\beta} \in [10, 20]$ Hz. We use "above" in the sense that the $g_{A\beta E}$ component of a point in parameter space defines a height for that point. This surface, S, is then the following sets of points:

$$S := \left\{ (g_{A\beta I}, g_{IE}, g_{A\beta E}) : g_{A\beta E} = \min_{f_{A\beta} \in [10,20]} \frac{V_{E,thr} - V_{E,rest} + g_{A\beta I} \cdot f_I(f_{A\beta})}{f_{A\beta}} \right\}.$$

The allodynia surface S always lies above the APS, as illustrated in Figure II.4.a. We discuss the preceding formula for S and the properties of allodynia surfaces further in Section 2.6.5.

For each circuit associated with a point in the APS, we identify its vulnerability to allodynia by computing the shortest path from the point to the allodynia surface S. The shortest path is a straight line joining the point in the APS to the corresponding nearest point on S, which we compute using a customized global optimization scheme described further in Section 2.6.6. Because the shortest path indicates precisely how to reach the allodynia surface by altering coupling strengths as little as possible, it represents the direction in parameter space in which the circuit is most vulnerable to allodynia. In this way, the components of the shortest path vector suggest which circuit components, along with the corresponding magnitudes of their relative change, will disrupt E-I balance in the circuit resulting in allodynia.

Summary statistics for the shortest paths to the allodynia surface

We see in Figure II.4.a that in some portions of the APS, the direction needed to travel to reach the allodynia surface fastest is mainly the $-\hat{g}_{A\beta I}$ direction. However, in other portions of the APS, the direction is mainly the $-\hat{g}_{IE}$ direction. Using density-based scanning [EKSX96], we assign points to clusters based on the directions of their shortest path to S. This cluster analysis identifies two clusters. For points in cluster 1 (red) the fastest direction to the allodynia surface primarily involves decreasing $\hat{g}_{A\beta I}$, while for points in cluster 2 (cyan), the shortest path involves decreasing $\hat{g}_{IE}$ and increasing $\hat{g}_{A\beta E}$.

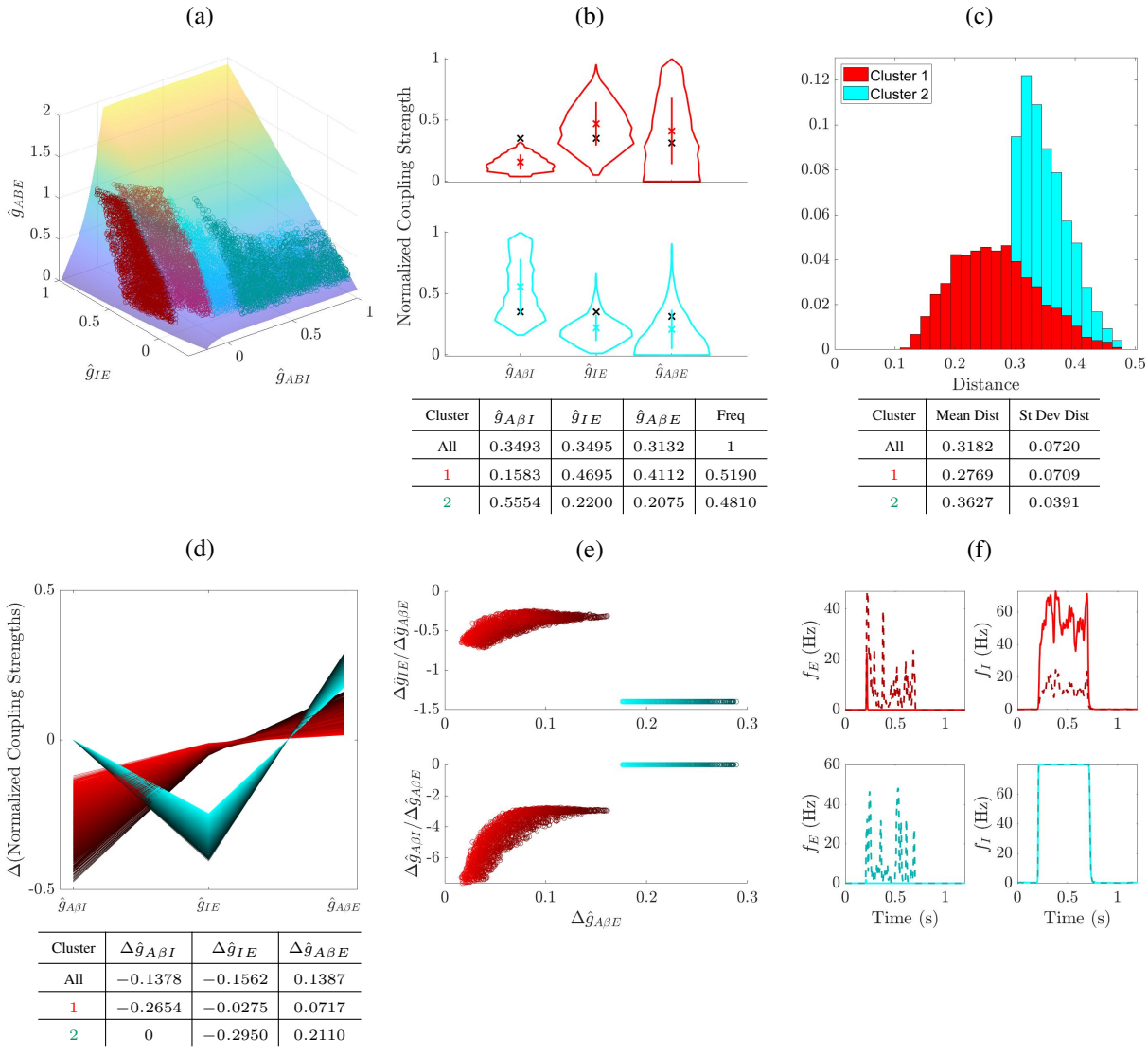
Within the APS, cluster 1 is characterized by relatively larger $\hat{g}_{IE}$ and $\hat{g}_{A\beta E}$ values but considerably smaller $\hat{g}_{A\beta I}$ values (Figure II.4.b). In fact, $\hat{g}_{A\beta I}$ values for cluster 1 are so small that they all lie below the mean $\hat{g}_{A\beta I}$ for the whole APS. We also see that cluster 1 is generally closer to S than cluster 2 (Figure II.4.c), indicating that cluster 1 is the more sensitive of the two clusters to dysregulation. A closer look at the shortest paths from points in cluster 1 to the allodynia surface reveals that shortest paths involve traveling in a mix of all component directions. The largest magnitude components of the mix are the decreasing $\hat{g}_{A\beta I}$- and increasing $\hat{g}_{A\beta E}$-directions (Figure

II.4.d). Otherwise, the mix varies from point to point within the cluster, exhibiting no consistent correlation in the magnitude of changes among the directions (Figure II.4.e). Thus, while the most efficient means of producing allodynia for circuits in cluster 1 involve disinhibition by lowering the response of the I population to $A\beta$ input (Figure II.4.f), the most efficient means also involves reducing $\hat{g}_{IE}$, thereby reducing the effect of I's output on E. The most efficient means further involves increasing $\hat{g}_{A\beta E}$, over-exciting E.

Cluster 2, on the other hand, is characterized by larger $g_{A\beta I}$ values and smaller $\hat{g}_{IE}$ and $\hat{g}_{A\beta E}$ values, (Figure II.4.b). Cluster 2 is further characterized by longer distances to the allodynia surface, indicating that these circuits may be more protected against E-I disruption (Figure II.4.c). Notably, shortest paths from points in cluster 2 to S *only* involve changing $\hat{g}_{IE}$ and $\hat{g}_{A\beta E}$ (Figure II.4.d), and the changes are in a fixed ratio (Figure II.4.e). Thus, the shortest paths to the allodynia surface from points in cluster 2 are always in exactly the same direction, namely a $\sim 30\%$ decrease in g_{IE} and a $\sim 20\%$ increase in $g_{A\beta E}$ (the direction of the unit vector $(0, -0.2950, 0.2110)/||(0, -0.2950, 0.2110)||$).

The difference in the allodynia-sensitivity of circuits in cluster 1 vs circuits in cluster 2 likely lies in the differences in $\hat{g}_{A\beta I}$ between clusters. In particular, we see in Figure II.4.f that a typical point in cluster 1 (top panels) exhibits inhibitory firing-rates around 50 Hz, just over half of the maximum firing-rate. When a population is at its half-maximum firing-rate, it is engaging its activation function where the activation function's slope is steepest, and thus firing rates will be most sensitive to changes in input. Also considering that $\hat{g}_{A\beta I}$ modulates the strength of the inputs to the I activation function, it follows that the I population firing rates for cluster 1 circuits are particularly sensitive to changes in $\hat{g}_{A\beta I}$. On the other hand, for a typical circuit in cluster 2 (bottom panels), the firing-rate of the I population is saturated at its maximum value (80 Hz). In this case, the population is engaging the activation function where the activation function's slope is very flat, and thus firing rates will be insensitive to changes in input. Hence, I population firing rates for cluster 2 circuits are very insensitive to changes in $\hat{g}_{A\beta I}$.

In summary, all circuits in the APS are most vulnerable to allodynia by the combined effect of disinhibition and over-excitation. However, the specific circuit components involved are different. The circuits in cluster 1, for instance, are most easily dysregulated by changes in the responses of the I and E populations to $A\beta$ input. Specifically, decreased I cell responses and potentiated E cell responses to $A\beta$ input open the inhibitory gate on E cell activity (see Figure II.4.f, top row). For circuits in cluster 2, E-I balance is disrupted solely at the E cell population via a decrease in the efficacy of the inhibitory input from the I population coupled with potentiated responses to $A\beta$ input (Figure II.4.f, bottom row).



Cluster	$\hat{g}_{A\beta I}$	$\hat{g}_{IE}$	$\hat{g}_{A\beta E}$	Freq
All	0.3493	0.3495	0.3132	1
1	0.1583	0.4695	0.4112	0.5190
2	0.5554	0.2200	0.2075	0.4810

Cluster	Mean Dist	St Dev Dist
All	0.3182	0.0720
1	0.2769	0.0709
2	0.3627	0.0391

Cluster	$\Delta\hat{g}_{A\beta I}$	$\Delta\hat{g}_{IE}$	$\Delta\hat{g}_{A\beta E}$
All	−0.1378	−0.1562	0.1387
1	−0.2654	−0.0275	0.0717
2	0	−0.2950	0.2110

Figure II.4: Examining the shortest paths to the allodynia surface yields two clusters in the simple circuit. **(a)** A scatter plot of the sampled sets of coupling strengths with the allodynia surface overlaid (5000 points colored red and blue). Overlaid on top is the allodynia surface, and for each sampled set of coupling strengths, the nearest point on the allodynia surface is also shown (darker red and blue points). Clearly, the data divides into two clusters (red points versus blue points) based on the direction needed to travel to reach the corresponding closest points on the allodynia surface. **(b)** Violin plots (top) of the mean coupling strengths for sampled sets in each cluster. A black "x" indicates the mean of the corresponding data from points across both clusters. Bottom: mean normalized coupling strengths for each cluster and for all data.**(c)** The probability distribution (top) of the shortest distances to the allodynia surface (overall profile). Contributions to the area under the curve from each cluster are shaded according to the cluster. Bottom: the mean and standard deviations of the distance to the allodynia surface for each cluster. **(d)** A parallel plot representation (top) of the shortest paths from sampled sets of coupling strengths to the corresponding nearest points on the allodynia surface. Bottom: mean-values of the components of the displacements for sampled sets of coupling strengths in each cluster. **(e)** Ratios of components of the displacement vector in the $\hat{g}_{IE}$-direction (top) and $\hat{g}_{A\beta I}$-direction (bottom) to the component of the displacement vector in the $\hat{g}_{A\beta E}$-direction versus the component of displacement in the $\hat{g}_{A\beta E}$-direction. Such ratios are roughly constant for cluster 2 but not cluster 1. **(f)** Firing-rate responses to a noisy $A\beta$ input signal for each population (columns) and for each cluster (rows). Solid lines correspond to a circuit with coupling strengths given by the mean clustering strengths for the particular cluster, and dashed lines correspond to the circuit with coupling strengths given by the corresponding closest point on the allodynia surface. $A\beta$ input frequencies are chosen to be the smallest possible $A\beta$ that induces allodynia for each cluster.

2.3 Analysis of the dynamic allodynia circuit

As discussed in Section 2.1.2, we propose a circuit motif which reflects experimentally identified layer I-II dorsal horn circuitry involved in mediating dynamic allodynia. The proposed circuit, which we refer to as the "dynamic circuit", consists of four populations of neurons: two inhibitory ($I1$ and $I2$), and two excitatory ($E1$ and $E2$). All four populations receive an input signal from $A\beta$ fibers. We assume the excitatory population, $E2$, is responsible for relaying signals to layer I projection neurons that signal to the brain. Essentially, the dynamic circuit is two simple circuits coupled together, with the $I1$ and $E1$ subcircuit providing additional excitation to $E2$, which relays signals to projection neurons. A schematic of the circuit is shown in Figure II.5.a.

We approach modeling and analyzing the dynamic circuit analogously to how we modeled and analyzed the simple circuit. As before, we model each population with a firing-rate model, equipped with activation functions and voltage dynamics as in Section 2.6.1. This time, though, there are 4 populations–$I1$, $E1$, $I2$, and $E2$–rather than two to worry about, and 7 coupling strengths–$g_{A\beta I1}$, g_{I1E1}, $g_{A\beta E1}$, g_{E1E2}, $g_{A\beta I2}$, g_{I2E2}, and $g_{A\beta E2}$–to worry about, rather than the three from the simple circuit. We illustrate these 7 coupling strengths, along with the four populations, in the schematic of the circuit.

2.3.1 Conditions on dynamic circuit parameters

As for the simple circuit, we impose conditions on the steady-state voltages of the E1, E2, I1, and I2 populations so the circuit displays experimentally observed behaviors under normal conditions. Specifically, we again require that all steady-state voltages remain within reasonable bounds. Since we are considering the I1-E1 portion of the circuit to behave as a simple circuit, we impose the same conditions as on the simple circuit to make it display pain inhibition. Likewise, under normal conditions, where $E1$ population activity is gated by inhibition from the $I1$ population, the I2-E2 portion of the circuit is essentially disconnected from the I1-E1 portion of the circuit, and also behaves like a simple circuit. Thus, we impose conditions on the steady states of the I2 and E2 populations accordingly. However, if the $I1$ population is ablated, then the $E1$ population is expected to fire in response to typical $A\beta$ stimuli, and in turn excite the $E2$ population. In this case, we expect the $E2$ population to also fire, thereby displaying allodynia.

Each of these conditions leads to an inequality on steady state population voltages, as summarized in Table (II.2). This system can be re-written as a set of inequalities for the coupling strengths that must be satisfied for all $f_{A\beta}$ input levels in the range $[10, 20]$ Hz. The resulting system of inequalities and optimization problems is derived in Section 2.6.3.2, and summarized in Table II.7.

Proceeding as we do for the simple circuit, we define the APS for the dynamic circuit as the sets of 7-tuples of coupling strengths $(g_{A\beta I1}, g_{I1E1}, g_{A\beta E1}, g_{E1E2}, g_{A\beta I2}, g_{I2E2}, g_{A\beta E2})$ which satisfy the system of inequalities and optimization problems. As solving the system explicitly is prohibitively difficult, we simplify the optimization problems using Lambert functions (see Section 2.6.3.2) and then randomly and uniformly-in-space sample from the defined APS using our customized sampling algorithm (Section 2.6.4). After analyzing the APS, we then define an allodynia surface (Section 2.6.5) and compute the shortest paths (Section 2.6.6) from points in the APS to the allodynia surface. Clustering analysis based on these shortest paths identifies different sensitivities across the circuit population to disruption of E-I balance and hence allodynia.

Condition Type	**Condition**	**Steady state voltage inequality–for all $f_{A\beta}$**
Control conditions	V_{I1} upper bound	$V_{I1,max} \geq V_{I1} = g_{A\beta I1} f_{A\beta} + V_{I1,rest}$
	$I1$ fires	$V_{I1,thr} \leq V_{I1} = g_{A\beta I1} f_{A\beta} + V_{I1,rest}$
	V_{I2} upper bound	$V_{I2,max} \geq V_{I2} = g_{A\beta I2} f_{A\beta} + V_{I2,rest}$
	$I2$ fires	$V_{I2,thr} \leq V_{I2} = g_{A\beta I2} f_{A\beta} + V_{I2,rest}$
	Pain inhibition (E2)	$V_{E2,rest} \geq V_{E2} = g_{A\beta E2} f_{A\beta} - g_{I2E2} f_{I2} + V_{E2,rest}$
	Pain inhibition (E1)	$V_{E1,rest} \geq V_{E1} = g_{A\beta E1} f_{A\beta} - g_{I1E1} f_{I1} + V_{E1,rest}$
E1 ablation conditions	V_{E2} lower bound	$V_{E2,min} \leq V_{E2} = g_{A\beta E2} f_{A\beta} - g_{I2E2} f_{I2} + V_{E2,rest}$
I1 ablation conditions	E1 fires	$V_{E1,thr} \leq V_{E1} = g_{A\beta E1} f_{A\beta} + V_{E1,rest}$
	V_{E2} upper bound	$V_{E2,max} \geq V_{E2} = g_{A\beta E1} f_{A\beta} + g_{E1E2} f_{E1,I_1 abl} - g_{I2E2} f_{I2} + V_{E2,rest}$
	E2 fires	$V_{E2,thr} \leq V_{E2} = g_{A\beta E1} f_{A\beta} + g_{E1E2} f_{E1,I_1 abl} - g_{I2E2} f_{I2} + V_{E2,rest}$
I2 ablation conditions	V_{E2} upper bound	$V_{E2,max} \geq V_{E2} = g_{A\beta E2} f_{A\beta} + g_{E1E2} f_{E1} + V_{E2,rest}$
	E2 fires	$V_{E2,thr} \leq V_{E2} = g_{A\beta E2} f_{A\beta} + g_{E1E2} f_{E1} + V_{E2,rest}$

Table II.2: Conditions that the proposed circuit mediating dynamic allodynia must satisfy and the resulting inequalities on steady-state voltages in response to $A\beta$ input. We ensure that the circuit exhibits these behaviors by imposing conditions (middle column) on the circuit. Each condition is exhibited in either control, E1-ablation, I1-ablation, or I2-ablation conditions (left-most column), and is realized as an inequality (right-most column) on the steady-state voltage of a population.

2.3.2 The allowable parameter space (APS) for the dynamic circuit

To illustrate the response of dynamic circuits in the APS to $A\beta$ input, we choose 20 points sampled from the APS. We then simulate the corresponding circuit models, taking the $A\beta$ input to be noisy and to have time-averaged firing-rates $f_{A\beta}$ in $[10, 20]$ Hz across the network on average. In

each simulation (see Figure II.5.b), the firing-rates of the $I1$ and $I2$ populations typically rise to nearly their maximum firing-rates of 80 Hz and always rise well above 0. The inhibition from the $I1$ population prevents the $E1$ population from firing, and the inhibition from I2 causes voltages of the $E2$ population to drop, as we require for pain inhibition. Then, as expected, the $E2$ population doesn't fire. Hence, the sampled sets of coupling strengths lead to the expected behaviors under control conditions.

Violin plots (Figure II.5.c)) of the sampled coupling strength values in the APS show that $g_{A\beta E1}$ has the largest range of allowable values (from about 2 - 6.9). Although $g_{A\beta E1}$ has a long tail of low values, along which the distribution of $g_{A\beta E1}$ is very narrow. If we ignore the tail, the range of $g_{A\beta E1}$ would be comparable to that of $g_{A\beta E2}$ (3.5 - 6.9 mV/Hz). The second largest ranges of coupling strengths belong to $g_{A\beta I1}$ and $g_{A\beta I2}$s (both about 2.8 - 7.1 mV/Hz), whereas g_{I1E1} and g_{I2E2} have the smallest ranges of allowable values (1.0 - 2.0 and 0.9 - 2.1 mV/Hz, respectively). Because $g_{A\beta E1}$ can thus be changed the most without exiting the APS, it is in some sense the least sensitive coupling strength underlying this circuit. In the same sense, g_{I1E} is the most sensitive coupling strength underlying this circuit.

To glean additional information about parameter sensitivity, we control for relative parameter variations by normalizing all parameters so their ranges are $[0, 1]$, as discussed in (Section 2.2.3). In the parallel plot of normalized parameter sets (Figure II.5.d) showing the full set of sampled points in the APS, distinct clustering is not apparent. However, we can begin to see correlations in the relative values of coupling strengths. Specifically, segments of the lines between two coupling strength values are flat, such as between $\hat{g}_{I1E1}$ and $\hat{g}_{A\beta E1}$, indicates a strong correlation between the two coupling strengths. Indeed, correlations (Figure II.5. e)) indicate for instance that $\hat{g}_{I1E1}$ and $\hat{g}_{A\beta E1}$ are strongly positively correlated, as are $\hat{g}_{I2E2}$ and $\hat{g}_{A\beta E2}$, indicating an E-I balance within the $I1$-$E1$ subcircuit and the $I2$-$E2$ subcircuit, similar to what we find for the simple circuit. Further, coupling strengths pertaining specifically to the $I2$-$E2$ subcircuit are only weakly correlated with coupling strengths pertaining to the $I1$-$E1$ subcircuit, indicating further that excitation and inhibition are being balanced within each subcircuit, more so than across the two subcircuits.

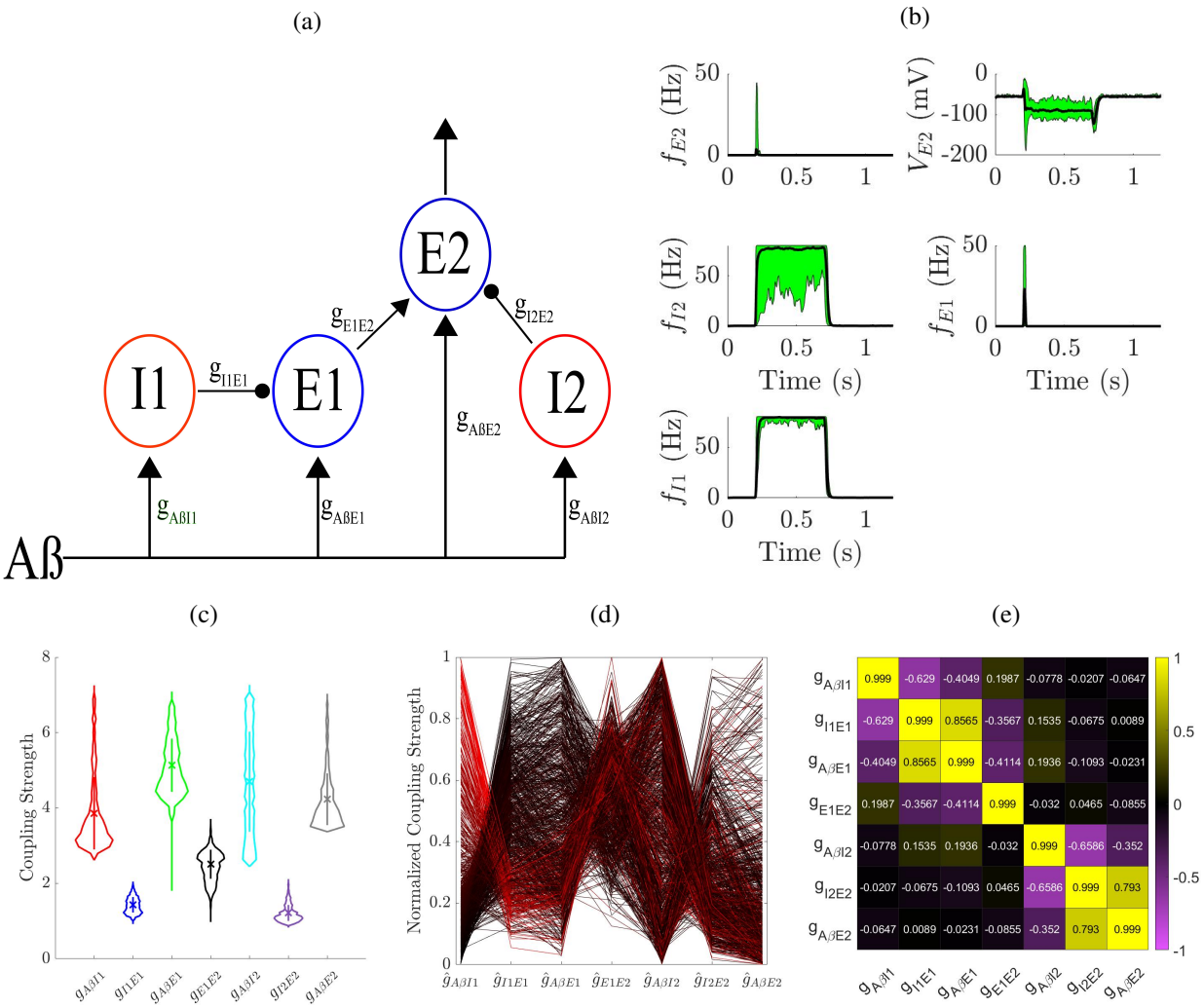


Figure II.5: Allowable parameter space for the proposed circuit mediating dynamic allodynia. **(a)** A schematic of the dynamic circuit. $I1$ and $I2$ represent the populations of inhibitory neurons, and $E1$ and $E2$ represent the populations of excitatory neurons. We assume the $E2$ population relays signals to projection neurons. $A\beta$ represents inputs to the circuit relayed from the periphery along $A\beta$ fibers. **(b)** The mean (black lines) and range (shaded green areas) for the firing rate (top left) and voltage (top right) of the $E2$ population, as well as for the firing rates of the I2 (middle left), E1 (middle right), and I1 (bottom left) populations calculated across 20 sampled sets of coupling strengths each with a different random input $A\beta$ stimulus. **(c)** Violin plots for each coupling strength across all sampled sets of coupling strengths. An "x" and a vertical bar represent the mean and the values within one standard deviation of the mean for the corresponding coupling strength. **(d)** A "parallel plot" representation of the sampled sets of normalized coupling strengths. A line gives the values of each coupling strength in a sampled set. **(e)** Normalized Pearson's correlation coefficients between coupling strengths across sampled sets of normalized coupling strengths.

2.3.3 Mechanisms of allodynia in the dynamic circuit

To conduct the analysis of sensitivity to allodynia, we define an allodynia surface S_{dyn} based on the condition that the average voltage of the $E2$ population increases past the firing threshold in response to $A\beta$ input in the normal range ($f_{A\beta}$ in $[10, 20]$ Hz). We then re-write this as a condition on $g_{A\beta E2}$ as follows:

$$S_{dyn} := \left\{ (g_{A\beta I1}, g_{I1E1}, ..., g_{A\beta E2}) : g_{A\beta E2} = \min_{f_{A\beta} \in [10,20]} \left[\frac{V_{E2,thr} - V_{E2,rest} - g_{E1E2} \cdot f_{E1}(f_{A\beta}) + g_{I2E2} \cdot f_{I2}(f_{A\beta})}{f_{A\beta}} \right] \right\}.$$

Recall that the allodynia surface separates circuits which cannot produce allodynia for any typical $f_{A\beta}$ from those that can for at least some $f_{A\beta} \in [10, 20]$ Hz. We discuss the preceding formula for the allodynia surface S_{dyn} and the properties of the allodynia surface further in Section 2.6.5.

Clustering based on the shortest path to the allodynia surface

Much as for the simple circuit, we can divide the APS into clusters according to the shortest paths from points in the APS to the allodynia surface S_{dyn}. Doing so via density-based scanning yields four clusters: cluster 1 (Figure II.6, green), cluster 2 (blue), cluster 3 (red), and cluster 4 (cyan).

Clusters 1 and 2 are the $I2$-$E2$ subcircuit's analog of clusters 1 and 2 appearing in the basic circuit. In particular, parameter sets belonging to cluster 1 (green) are characterized by having $\hat{g}_{A\beta I2}$ values and $\hat{g}_{I2E2}$ values considerably smaller and larger than their respective means across the whole APS. Just as for cluster 1 in the simple circuit, values of $\hat{g}_{A\beta I1}$ are so small, that they lie below the mean value of $\hat{g}_{A\beta I1}$ across the whole APS. Also just as for cluster 1 in the simple circuit, cluster 1 in the dynamic circuit is generally closer to S_{dyn} than are the other clusters.

A closer look at the shortest paths to S_{dyn}, again much as for cluster 1 in the simple circuit, reveals that shortest paths for cluster 1 in the dynamic circuit involve traveling in a mix of all component directions pertaining to the $I2$-$E2$ subcluster (Figure II.6.c). That is, to reach the allodynia surface, $g_{A\beta I2}$ and g_{I2E2} decrease, $g_{A\beta E2}$ increases, and none of the other coupling strengths change. However, $g_{A\beta I2}$ decreases more than $g_{A\beta E2}$ increases, and far more than g_{I2E2} decreases. As such, we expect that upon traveling to the end of the shortest path, the activity of $I2$ is dramatically reduced in response to normal $A\beta$ input, thereby disinhibiting $E2$, and producing allodynia. This is reflected in simulations of cluster 1 circuits (Figure II.6.d, top row), where upon traveling to the end of the shortest path, in response to normal $A\beta$ input, activity of the $I2$ population indeed goes down dramatically and $E2$ fires, thereby producing allodynia.

On the other hand, the distribution of individual coupling strengths for points in cluster 2 do not greatly differ from the mean or from the overall distribution in the APS. This is mainly because the majority of points in the APS belong to cluster 2 (Figure II.6.a,b). Nevertheless, there are some small differences between the means of coupling strengths in cluster 2 versus the whole APS. Namely, these differences are in $\hat{g}_{A\beta I2}$, which is somewhat larger than normal, and in $\hat{g}_{I2E2}$, which is somewhat smaller than normal.

Thus, it should not come as a surprise that the most efficient allodynia-producing mechanism for cluster 2 involves disrupting E-I balance in the $I2$ - $E2$ portion of the circuit, as seen in Figure II.6.c, without changing coupling strengths in the $I1$ - $E1$ portion of the circuit at all. In particular, the most allodynia-vulnerable direction involves primarily decreasing $\hat{g}_{I2E2}$ and increasing $\hat{g}_{A\beta E2}$, thereby simultaneously disinhibiting and over-exciting the $E2$ population solely. Similarly to the cluster 1, this is reflected in Figure II.6.d, where on the allodynia surface (dashed curves), the $E1$ population doesn't fire while the $E2$ population does fire.

Whereas clusters 1 and 2 are the $I1$-$E1$ subcircuit's analog of clusters 1 and 2 from the simple circuit, clusters 3 and 4 from the dynamic circuit are the $I2$-$E2$ subcircuit's analog of those same clusters from the simple circuit. Namely, similarly to cluster 1, parameter sets from cluster 3 are characterized by very small $\hat{g}_{A\beta I1}$ and $\hat{g}_{I1E1}$ (Figure II.6.a). Cluster 3 parameter sets also have a smaller $\hat{g}_{A\beta E1}$ than normal, much more so than does $\hat{g}_{A\beta E2}$ in points belonging to cluster 1. Points from cluster 3 are generally quite close to the allodynia surface, almost as close as points from cluster 1.

Because coupling strengths in cluster 3 mainly differ from the APS as a whole in terms of $g_{A\beta I1}$, g_{I1E1} and $g_{A\beta E1}$, it makes sense that that the most allodynia-sensitive direction for cluster 3 involves disrupting E-I balance specifically in the $I1$-$E1$ subcircuit. Namely, similarly to cluster 1 from the dynamic circuit, that direction mainly involves decreasing $\hat{g}_{A\beta I1}$, to a lesser extent increasing $\hat{g}_{A\beta E1}$. However, because changes in the $I1$-$E1$ subcircuit affect the activity of $I2$ and $E2$, the most allodynia-sensitive direction involves at least small changes to all coupling strengths. Thus, as seen in Figure II.6, bottom row, upon moving coupling strengths to S_{dyn} along the most allodynia-sensitive direction (dashed curves), there is no significant change in the firing activity of $I2$. However, $I1$ activity decreases towards 0, disinhibiting $E1$, and causing $E1$ to fire strongly in response to typical $A\beta$ input. As a result $E2$ becomes over-excited and fires as well, thus producing allodynia.

Parameter sets belonging to cluster 4 (cyan) are characterized by having considerably larger values of $\hat{g}_{I1E1}$ and $\hat{g}_{A\beta E1}$ values compared to mean APS values (Figure II.6.a). Cluster 4 is also significantly further from the allodynia surface than the remainder of the APS (Figure II.6).

The most efficient allodynia mechanism for cluster 4 mainly involves disrupting E-I balance in the $I1$ - $E1$ portion of the circuit, with only small changes in the coupling strengths in the $I2$ - $E2$ portion (Figure II.6.c)). Indeed, the most allodynia-vulnerable direction for points in cluster 4 primarily involves decreasing $\hat{g}_{I1E1}$ and increasing $\hat{g}_{A\beta E1}$, thereby simultaneously disinhibiting and over-exciting the $E1$ population. This is reflected in Figure II.6.d, bottom row, showing that with parameter values on the allodynia surface (dashed curves), the $E1$ population fires in response to

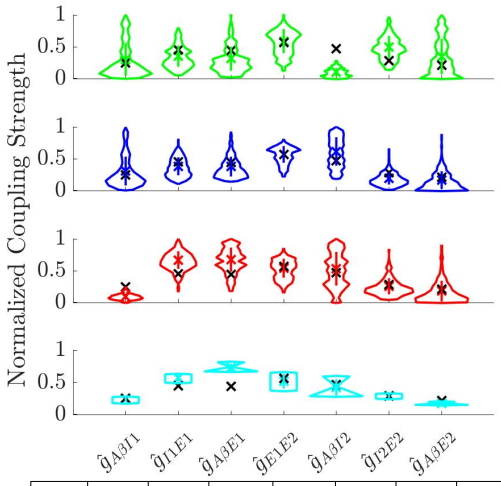
normal $A\beta$ input, leading to $E2$ population firing as well. However, the $E1$ population itself does not decrease due to a reduction in inhibitory activity of $I1$, instead, it decreases due to a reduction in the impact of inhibitory signaling from $I1$ to $E1$.

What seems to distinguish circuits corresponding to cluster 4, and circuits corresponding to all other clusters, for that matter, is the firing-rates of inhibitory populations in response to normal $A\beta$ signals under control conditions, much as it does for the simple circuit. Indeed, for cluster 1, firing-rates of $I2$ are about half-their maximum firing-rate of 80 Hz. Because the slope of the activation function is steepest when firing-rates are at that point in their trajectories, the activity of $I2$ is particularly sensitive to the inputs that are rescaled by $g_{A\beta I2}$. Hence, the system for cluster 1 circuits is particularly sensitive to changing $g_{A\beta I2}$. Indeed, our clustering indicates that upon minimally altering our parameters to reach S_{dyn}, it is $\hat{g}_{A\beta I2}$ primarily that is altered. A similar argument explains why cluster 3 is so sensitive to $\hat{g}_{A\beta I1}$.

On the other hand, if the activity for $I1$ and $I2$ are saturated near their 80 Hz maximums, then the system will be particularly sensitive to changing either $\hat{g}_{I1E1}$ or $\hat{g}_{I2E2}$. Indeed, because these two coupling strengths re-scale the firing-rates output by $I1$ and $I2$, and because the firing-rates are so high, small changes in firing activity can have a large effect on the system. Thus, cluster 2 is particularly sensitive to changing $\hat{g}_{I2E2}$. However, cluster 2 is not so allodynia-sensitive to changing $\hat{g}_{I1E1}$ because the resulting changes on the system could only affect the firing-activity of $E2$ by acting through $E1$ and overcoming the inhibition on $I1$, which isn't feasible because $E2$ itself isn't coupled very strongly with $A\beta$ signaling for cluster 2. On the other hand, $g_{A\beta E1}$ is large for cluster 4, hence why cluster 4 is so allodynia-sensitive to changes in $\hat{g}_{I1E1}$ (while simultaneously changing $\hat{g}_{A\beta E1}$, to give the circuit an extra boost).

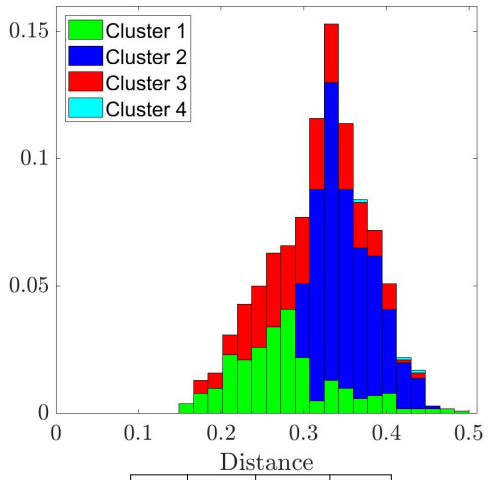
Nevertheless, the majority of circuits in the APS are most vulnerable to allodynia by altering the E-I balance on the $I2 - E2$ portion of the circuit. We believe this is because any changes regarding the activity of the $I1$-$E1$ portion of the circuit can only affect $E2$ by acting through the effect of $E1$. Also considering that the distribution of $\hat{g}_{A\beta E1}$ is skewed towards 0, it seems to be the case that in the majority of the APS, $E1$ is not sufficiently excited by $A\beta$ firing to overcome the direct inhibition of $E2$ from $I2$, even when $E1$ is disinhibited.

(a)



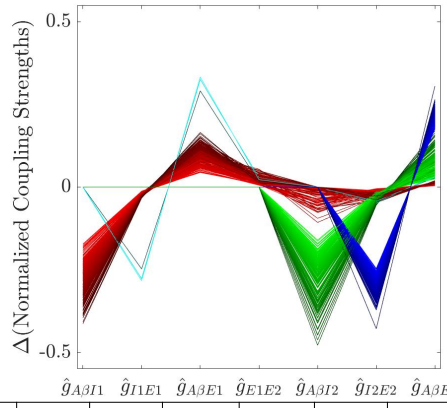
Cluster	$\hat{g}_{A\beta I1}$	$\hat{g}_{I1E1}$	$\hat{g}_{A\beta E1}$	$\hat{g}_{E1E2}$	$\hat{g}_{A\beta I2}$	$\hat{g}_{I2E2}$	$\hat{g}_{A\beta E2}$
All	0.2491	0.4495	0.4399	0.5655	0.4645	0.2838	0.2131
1	0.3000	0.3508	0.3122	0.5892	0.1005	0.4970	0.3448
2	0.3063	0.3812	0.3761	0.5678	0.6144	0.1905	0.1641
3	0.0954	0.6674	0.6747	0.5394	0.5311	0.2548	0.1802
4	0.2394	0.5666	0.7349	0.5195	0.4231	0.2925	0.1708

(b)



Cluster	Rel. Freq.	Dist. Mean	Dist. SD
All	1	0.3290	0.0601
1	0.2480	0.2872	0.0663
2	0.4860	0.3613	0.0346
3	0.2630	0.3076	0.0572
4	0.0030	0.4196	0.0299

(c)



Cluster	$\Delta\hat{g}_{A\beta I1}$	$\Delta\hat{g}_{I1E1}$	$\Delta\hat{g}_{A\beta E1}$	$\Delta\hat{g}_{E1E2}$	$\Delta\hat{g}_{A\beta I2}$	$\Delta\hat{g}_{I2E2}$	$\Delta\hat{g}_{A\beta E2}$
All	−0.0755	−0.0064	0.0270	0.0055	−0.0705	−0.1562	0.1245
1	0	0	0	0	−0.2737	−0.0282	0.0803
2	0	0	0	0	0	−0.2942	0.2097
3	−0.2872	−0.0211	0.0990	0.0205	−0.0100	−0.0232	0.0099
4	−0.0004	−0.2694	0.3165	0.0221	0	−0.0436	0.0311

(d)

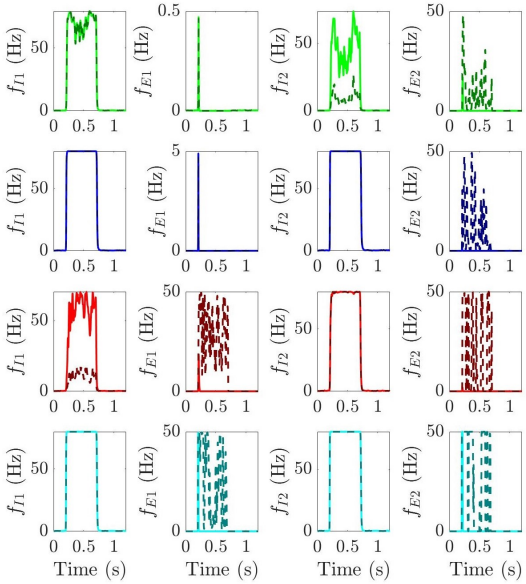


Figure II.6: Examining the shortest paths to the allodynia surface yields two clusters in the proposed circuit mediating dynamic allodynia. **(a)** Violin plots (top) for each coupling strength in the sampled 7-tuples of coupling strengths. Each set of rows corresponds to a cluster, with the top row corresponding to points in cluster 1, etc. Each black "x" represents the mean coupling strength across all sampled points in the corresponding cluster. Mean coupling strengths (bottom) are shown for each cluster. (**b**) The probability distribution (top) of the shortest distances to the allodynia surface (overall profile). Contributions to the area under the curve from each cluster are shaded according to cluster. Bottom: Relative frequencies of each cluster and values of the mean and standard deviation of the distances from sampled points in the APS to the allodynia surface. **(c)** A "parallel plot" representation of the shortest paths to the allodynia surface from each sampled point in the APS. A kine gives the components of the displacement vector corresponding to one such shortest path. Displacements cleanly divide into four clusters: cluster 1 (green), cluster 2 (purple). **(d)** Firing-rate responses to a noisy $A\beta$ input signal for each population (columns) and for each cluster (rows). Solid lines correspond to a circuit with coupling strengths given by the mean clustering strengths for the particular cluster, and dashed lines correspond to the circuit with coupling strengths given by the corresponding closest point on the allodynia surface. $A\beta$ input frequencies are chosen to be the smallest possible $A\beta$ that induces allodynia for each cluster.

2.4 Analysis of the proposed circuit mediating static allodynia

As described in Section 2.1.2, we propose a circuit motif that reflects experimentally identified layer I-II dorsal horn circuitry involved in mediating static allodynia. Going forward, we refer to the proposed circuit as the "static circuit". The static circuit consists of three populations of neurons: two inhibitory ($I1$ and $I2$), and one excitatory (E) population with all three populations receiving input from $A\beta$ fibers. We assume the excitatory population is responsible for relaying signals to projection neurons. We allow for the two inhibitory populations to have different underlying parameters, but for the analysis presented in this chapter, we continue to take all parameters for inhibitory populations to be the same and as shown in [RS02]. In this way, the circuit is symmetric in the sense that there is nothing to distinguish the two inhibitory populations from one another. This symmetry is reflected in the structure of the APS for this circuit, as well as in the ways in which allowable circuits may be induced to produce allodynia. A schematic of the population is shown in Figure II.7.a.

We approach modeling and analyzing the circuit analogously as in the previous cases. As before, we model each population with a firing-rate model, equipped with activation functions and voltage dynamics as in Section 2.6.1. This time, though, there are 3 populations–$I1$, $I2$, and E–to worry about, and there are 5 coupling strengths–$g_{A\beta I1}$, g_{I1E}, $g_{A\beta I2}$, g_{I2E}, and $g_{A\beta E}$–to worry about. As we do for both the simple and dynamic circuits, we illustrate these 5 coupling strengths along with the 3 populations in the circuit schematic.

2.4.1 Conditions on static circuit parameters

Also much as for the simple and dynamic circuits, we impose conditions on the steady-state voltages of the E, $I1$, and $I2$ populations so the circuit displays desired experimentally identified behaviors. In particular, we again require that all steady-state voltages remain within reasonable bounds. Unlike in the simple or dynamics circuit, in this circuit we require both of the inhibitory populations to be active in order to maintain pain inhibition. Thus, under control conditions, typical $A\beta$ stimuli induce both $I1$ and $I2$ to fire, preventing E from firing, and reducing V_E below its resting voltage. However, if the $I1$ population is ablated, we then presume that the excitatory signaling from $A\beta$ fibers is strong enough to overcome the remaining inhibition of E from $I2$ to induce E firing. Likewise, if $I2$ is ablated, we expect E to fire. We do not incorporate ablation of both $I1$ and $I2$ in developing our conditions for the static circuit, as that would lead to extreme over-excitation of E.

Each of these conditions results in an inequality on steady state population voltages, leading to a system of 10 inequalities. We summarize these inequalities and the corresponding condition on

steady-state voltages in Table (II.3).

We proceed with the analysis just as we did for the simple and dynamic circuits. See Section 2.6.3.3 for the resulting inequalities and optimization problems for the circuit coupling strengths derived from the conditions in Table II.3. From this system, we define the APS for the static circuit. See 2.6.5 for the definition of the allodynia surface S_{stat} for the static circuit.

Condition Type	**Condition**	**Steady state voltage inequality–for all $f_{A\beta}$**
Control conditions	V_{I1} upper bound	$V_{I1,max} \geq V_{I1} = g_{A\beta I1} f_{A\beta} + V_{I1,rest}$
	$I1$ fires	$V_{I1,thr} \leq V_{I1} = g_{A\beta I1} f_{A\beta} + V_{I1,rest}$
	V_{I2} upper bound	$V_{I2,max} \geq V_{I2} = g_{A\beta I2} f_{A\beta} + V_{I2,rest}$
	$I2$ fires	$V_{I2,thr} \leq V_{I2} = g_{A\beta I2} f_{A\beta} + V_{I2,rest}$
	Pain inhibition	$V_{E,rest} \geq V_E = g_{A\beta E} f_{A\beta} - g_{I1E} f_{I1} - g_{I2E} f_{I2} + V_{E,rest}$
	V_E lower bound	$V_{E,min} \leq V_E = g_{A\beta E} f_{A\beta} - g_{I1E} f_{I1} - g_{I2E} f_{I2} + V_{E,rest}$
I1 ablation conditions	E fires	$V_{E,thr} \leq V_E = g_{A\beta E} f_{A\beta} - g_{I2E} f_{I2} + V_{E,rest}$
	V_E upper bound	$V_{E,max} \geq V_E = g_{A\beta E} f_{A\beta} - g_{I2E} f_{I2} + V_{E,rest}$
I2 ablation conditions	E fires	$V_{E,thr} \leq V_E = g_{A\beta E} f_{A\beta} - g_{I1E} f_{I1} + V_{E,rest}$
	V_E upper bound	$V_{E,max} \geq V_E = g_{A\beta E} f_{A\beta} - g_{I1E} f_{I1} + V_{E,rest}$

Table II.3: Conditions that the proposed circuit mediating static allodynia must satisfy and the resulting inequalities on steady-state voltages. Much as for the simple circuit and the proposed circuit mediating dynamic allodynia, we ensure that the circuit exhibits these behaviors by imposing conditions (middle column) on the circuit. Each condition is exhibited in either control, I1-ablation, or I2-ablation conditions (left-most column), and is realized as an inequality (right-most column) on the steady-state voltage of a population.

2.4.2 The allowable parameter space (APS) for the static circuit

As previously, to illustrate the response of circuits in the APS to $A\beta$ input, we choose 20 points sampled from the APS. We then simulate the corresponding circuit models, taking the $A\beta$ input to be noisy and to have time-averaged firing-rates $f_{A\beta}$ in $[10, 20]$ Hz, across the network on average. In each simulation (see Figure II.7.b), the firing-rates of the $I1$ and $I2$ populations always rise to at least 25% of the maximum firing-rate, and typically approach the 80 Hz maximum. The resulting inhibition from $I1$ and $I2$ prevents the E population from firing, and cause E's voltage to drop, as

we require for pain inhibition. Hence, the sampled sets of coupling strengths lead to the expected behaviors.

Violin plots (Figure II.7.c) of the distributions of coupling strengths in the allowable parameter space highlight the symmetry of the static circuit. In particular, the distribution of $g_{A\beta I1}$ is very similar to that of $g_{A\beta I2}$. Likewise, the distribution of g_{I1E} is very similar to the distribution of g_{I2E}.

Also notable is that the range of $g_{A\beta E}$ is largest, about 4 mV/Hz; and excluding outliers, the ranges of $g_{A\beta I1}$, $g_{A\beta I2}$, and next largest, both about 2 mV/Hz. The smallest ranges belong to g_{I1E} and g_{I2E}. This indicates that in a sense, the circuit is least sensitive to changes in $g_{A\beta I1}$, $g_{A\beta I2}$, and $g_{A\beta E}$, but most sensitive to changes in g_{I1E} and g_{I2E}.

However, from the violin plots, we see that $g_{A\beta E}$ is far larger than in the simple or dynamic circuit: it varies between roughly 9 and 13 mV/Hz, whereas for the simple circuit and for the dynamic circuit, coupling strengths between $A\beta$ fibers and excitatory populations were closer to 5 mV/Hz. In the static circuit, however, $g_{A\beta E}$ needs to balance out two sources of inhibition under control conditions, and needs to be large enough to overcome the inhibition from one inhibitory population as would happen in either of the $I1$ or $I2$ ablation scenarios.

Further, as seen in Figure II.7.c and d, there are a few points with values of $g_{A\beta I1}$ and $g_{A\beta I2}$ much larger than is typical. This indicates that there are long and narrow regions of the APS with very low-volume. These low volume regions appear for high values of $g_{A\beta I1}$ and $g_{A\beta I2}$, and would likely look like tails extending in the positive $g_{A\beta I1}$ and positive $g_{A\beta I2}$ directions. In the corresponding circuits for these two tails in the APS, the activity of $I1$ and $I2$, respectively, would be very strong, but because g_{I1E} and g_{I2E} are respectively very small, the impact of signals from the I population on the E population is likely relatively weak.

We also see from the violin plots that values of $g_{A\beta I1}$, $g_{A\beta I2}$, g_{I1E}, and g_{I2E} are strongly bimodal. This bimodality is reflected in the two separate trends seen in the parallel plot (Figure II.5.d) representing the sampled sets of coupling strengths in normalized parameter space. In one trend, $\hat{g}_{I1E}$ is large (blacker lines), whereas the other trend (redder lines) involves larger $\hat{g}_{I2E}$ values, further reflecting symmetry in the circuit. From the parallel plot, we can also begin to see a number of correlations between coupling strengths. For instance, when $\hat{g}_{A\beta I1}$ is large, $\hat{g}_{I1E}$ is small. Further, as symmetry demands, when $\hat{g}_{A\beta I2}$ is large, $\hat{g}_{I2E}$ is small. Additionally, when $\hat{g}_{I1E}$ is large, $\hat{g}_{I2E}$ is small. These correlations are confirmed by calculating the Pearson's correlation coefficients between each coupling strength (Figure II.5.e). Generally these correlations reflect an E-I balance manifested in the previous circuits as a balance of $A\beta$ and inhibitory input to excitatory populations. However, for the static circuit, we see additional balancing between $I1$ and $I2$, where

if the inhibition from $I1$ is large, then the inhibition from $I2$ is small.

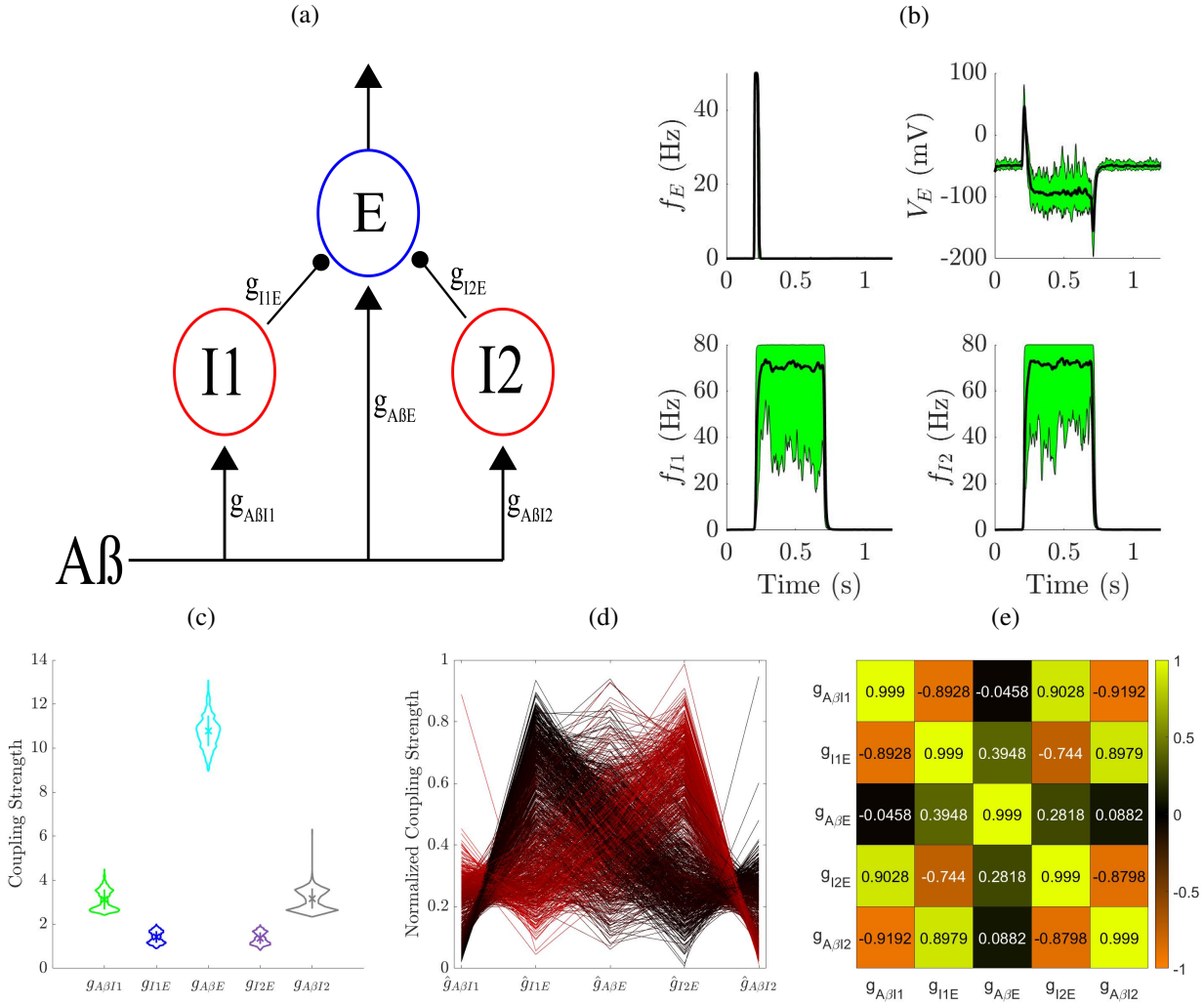


Figure II.7: The allowable parameter space for the proposed circuit mediating static allodynia. **(a)** A schematic of the static circuit. $I1$ and $I2$ represent the populations of inhibitory neurons and E represents the population of excitatory neurons. $A\beta$ represents inputs to the circuit relayed from the periphery along $A\beta$ fibers. **(b)** The mean (black lines) and range (shaded green areas) for the firing rate (top left) and voltage (top right) of the E population, as well as for the firing rates of the $I1$ (bottom left) and $I2$ (bottom right) populations calculated from 20 sampled sets of coupling strengths each with a different random $A\beta$ input stimulus. **(c)** Violin plots for each coupling strength across all sampled sets of coupling strengths. The thickness of the violin near values indicates how likely the corresponding coupling strength is to be near the values. An "x" and a vertical bar respectively represent the mean and the values within one standard deviation of the mean for the corresponding coupling strength. **(d)** A "parallel plot" representation of the sampled sets of normalized coupling strengths. A line gives the values of each coupling strength in a sampled set. **(e)** Normalized Pearson correlation coefficients between sampled sets of normalized coupling strengths.

2.4.3 Mechanisms of allodynia in the static circuit

To identify sensitivities of the circuits to allodynia, we define the allodynia surface S_{stat} as the set of points in the $(g_{A\beta I1}, g_{A\beta I2}, g_{I1E}, g_{I2E}, g_{A\beta E})$-space that satisfy the following (see Section 2.6.5):

$$S_{stat} := \left\{ (g_{A\beta I1}, ..., g_{A\beta E}) : g_{A\beta E} = \min_{f_{A\beta} \in [10,20]} \frac{V_{E,thr} - V_{E,rest} + g_{I1E} f_{I1}(f_{A\beta}) + g_{I2E} f_{I2}(f_{A\beta})}{f_{A\beta}} \right\}.$$

We then compute the shortest distance from each sampled point in the APS to S_{stat}.

Clustering based on the shortest path to the allodynia surface yields 4 clusters

Density-based scanning indicates that the APS for the static circuit divides into four clusters based on the shortest paths from APS points to the allodynia surface: cluster 1 (green), cluster 2 (blue), cluster 3 (cyan), and cluster 4 (gray) (Figure II.8).

Cluster 1 (green) is characterized by $\hat{g}_{A\beta I1}$ values larger than the APS mean but $\hat{g}_{I1E}$ smaller than the APS mean (Figure II.8.a, top panel). Conversely, $\hat{g}_{A\beta I2}$ values are smaller than the mean and $\hat{g}_{I2E}$ values are larger. This reflects that circuits in cluster 1 have a highly active $I1$ population whose effect on E is weak, and a weakly active $I2$ population whose effect on E is strong.

Generally portions of the APS corresponding to cluster 1 are further from the allodynia surface than typical (Figure II.8.b). To actually reach the allodynia surface most efficiently, cluster 1 circuits primarily need to decrease their $\hat{g}_{A\beta I2}$ values (Figure II.8.c). Simultaneously, but to a lesser extent, they also need to decrease their $\hat{g}_{I2E}$ values, increase their $\hat{g}_{A\beta E}$ values, and decrease their $\hat{g}_{I1E}$ values. This makes sense, because the firing activity of $I2$ is high but not saturated (Figure II.8.c), and the impact g_{I2E} of signaling from $I2$ on E is very strong. Thus, mainly decreasing $g_{A\beta I2}$ decreases $I2$ activity. The reduction in inhibitory signaling coming out of $I2$ is then magnified by the high g_{I2E} values, leading to even less inhibitory control of E.

Inducing allodynia by primarily reducing $g_{A\beta I2}$, and moreover, by not reducing $g_{A\beta I1}$ at all, makes sense because it is inefficient to decrease the $\hat{g}_{A\beta I1}$ values (Figure II.8.c). Indeed, as seen in Figure II.8.d, in a typical circuit in cluster 1, the firing-rate of I1 is saturated near it's maximum. Hence, f_{I1} is insensitive to changes in the inputs to $I1$, and thus is also insensitive to changes in $\hat{g}_{A\beta I1}$. Indeed, cluster 1 is most easily dysregulated by disinhibition of the $I2$ population.

Due to the symmetry between the $I1$ and $I2$ halves of the static circuit, cluster 2 (blue) is similar to cluster 1 except with the characteristics of $I1$ and $I2$ reversed. Namely cluster 2 is characterized by large $\hat{g}_{A\beta I2}$ values but small $\hat{g}_{I2E}$ values (Figure II.8.a, 2nd panel from the top) as well as small $\hat{g}_{A\beta I1}$ values and large $\hat{g}_{I1E}$ values. Cluster 2 is similarly as distant from the allodynia surface as is cluster 1 (Figure II.8.b), and allodynia occurs primarily by $I1$ disinhibition rather than $I2$ disinhibition (Figure II.8.c and d).

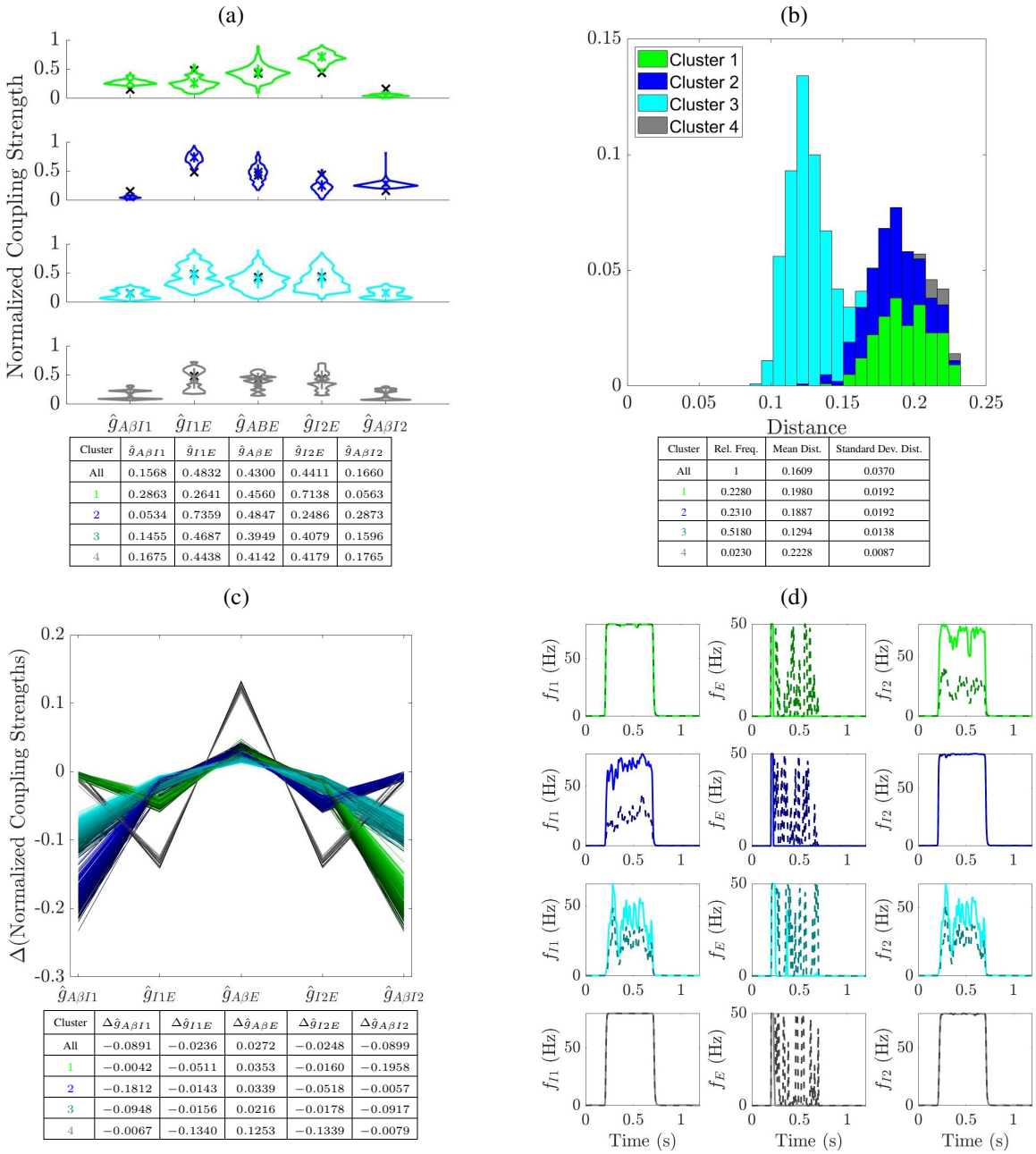
Whereas cluster 1 and cluster 2 both exhibit considerable asymmetry across the two halves of

the static circuit, cluster 3 (cyan) is symmetric across the two halves of the circuit. Indeed, in cluster 3, the coupling strength distributions for $\hat{g}_{A\beta I1}$ and its $I2$-population analog $\hat{g}_{A\beta I2}$ are very similar (Figure II.8.a, third panel). Likewise, the distributions for values of $\hat{g}_{I1E}$ and $\hat{g}_{I2E}$ are very similar. Further, the means of coupling strengths for cluster 3 are similar to the APS mean values, with slightly less bimodal distributions than for the APS as a whole.

Unlike clusters 1 and 2, portions of the APS in cluster 3 are close to the allodynia surface (Figure II.8.b). In fact, so much of cluster 3 is so much closer to the allodynia surface than are clusters 1 and 2, that cluster 3 causes the full distribution of distances from the allowable parameter space to the allodynia surface to be bimodal.

To actually induce allodynia in cluster 3 circuits, it is most efficient to decrease $\hat{g}_{A\beta I1}$ and $\hat{g}_{A\beta I2}$ by roughly equal amounts and to increase $\hat{g}_{A\beta E}$ (Figure II.8.c). As shown in Figure II.8.d, (third row from the top), for typical circuits in cluster 3, $I1$ and $I2$ firing-rates are both around halfway between their minimum and maximum values and are thus very sensitive to changing the inputs scaled by $\hat{g}_{A\beta I1}$ and $\hat{g}_{A\beta I2}$. Thus E firing occurs due to decreased firing in both the $I1$ and $I2$ populations.

Cluster 4 (gray) is likewise symmetric in its distributions of coupling strengths across the two halves of the static circuit, and has mean coupling strengths typical of the overall parameter space. However, unlike cluster 3, the distributions of coupling strengths are more bimodal than for the APS as a whole. This is reflected in the violin plots in Figure II.8.a (bottom panel) showing that the distributions of $\hat{g}_{A\beta I1}$ and $\hat{g}_{A\beta I2}$ vanish near their mean values. This suggests that cluster 4 has at least two spatially disconnected components, which can be confirmed by the parallel plot displayed in Figure A.3 shown in the supplementary material (Appendix A). In fact, within any one of these two disconnected components, particular cluster 4 circuits are asymmetric across the two halves of the static circuit. Nevertheless, allodynia is induced for all circuits in cluster 4 by altering coupling strengths in almost exactly the same way, thereby underscoring the differences between clustering in space versus clustering according to the "most sensitive direction".



Cluster	$\hat{g}_{A\beta I1}$	$\hat{g}_{I1E}$	$\hat{g}_{A\beta E}$	$\hat{g}_{I2E}$	$\hat{g}_{A\beta I2}$
All	0.1568	0.4832	0.4300	0.4411	0.1660
1	0.2863	0.2641	0.4560	0.7138	0.0563
2	0.0534	0.7359	0.4847	0.2486	0.2873
3	0.1455	0.4687	0.3949	0.4079	0.1596
4	0.1675	0.4438	0.4142	0.4179	0.1765

Cluster	Rel. Freq.	Mean Dist.	Standard Dev. Dist.
All	1	0.1609	0.0370
1	0.2280	0.1980	0.0192
2	0.2310	0.1887	0.0192
3	0.5180	0.1294	0.0138
4	0.0230	0.2228	0.0087

Cluster	$\Delta\hat{g}_{A\beta I1}$	$\Delta\hat{g}_{I1E}$	$\Delta\hat{g}_{A\beta E}$	$\Delta\hat{g}_{I2E}$	$\Delta\hat{g}_{A\beta I2}$
All	−0.0891	−0.0236	0.0272	−0.0248	−0.0899
1	−0.0042	−0.0511	0.0353	−0.0160	−0.1958
2	−0.1812	−0.0143	0.0339	−0.0518	−0.0057
3	−0.0948	−0.0156	0.0216	−0.0178	−0.0917
4	−0.0067	−0.1340	0.1253	−0.1339	−0.0079

Figure II.8: Examining the shortest paths to the allodynia surface yields four clusters in the proposed circuit mediating static allodynia. **(a)** Violin plots for each coupling strength in the sampled 5-tuples of coupling strengths. Each set of rows corresponds to a cluster, with the top row corresponding to points in cluster 1, etc. Each black "x" represents the mean coupling strength for all sampled points in the corresponding cluster. **(b)** The probability distribution (top) of the shortest distances to the allodynia surface (overall profile). Contributions to the area under the curve from each cluster are shaded according to cluster. Relative frequencies of the clusters (bottom), and the mean and standard deviation of the distances from sampled points in the APS to the allodynia surface. **(c)** A "parallel plot" (top) representation of the shortest paths to the allodynia surface from each sampled point from the APS. A line gives the components of the displacement vector corresponding to one such shortest path. Displacements cleanly divide into four clusters: cluster 1 (green), cluster 2 (blue), cluster 3 (cyan), and cluster 4 (gray). Mean displacements (bottom) are shown for each cluster. **(d)** Firing-rate responses to a noisy $A\beta$ input signal for each population (columns) and for each cluster (rows). Solid lines correspond to a circuit with coupling strengths given by the mean clustering strengths for the particular cluster (except for cluster 4, which is spatially disjoint, where we use a representative sampled set of clustering strengths), and dashed lines correspond to the circuit with coupling strengths given by the corresponding closest point on the allodynia surface. $A\beta$ input frequencies are chosen to be the smallest possible $A\beta$ that induces allodynia for each cluster.

Indeed, all circuits in cluster 4 are considerably further from the allodynia surface–0.22–than is typical of the allowable parameter space (Figure II.8.b). To actually induce allodynia in cluster 4 circuits, it is most efficient to decrease $\hat{g}_{I1E}$ and $\hat{g}_{I2E}$, and simultaneously increase $\hat{g}_{A\beta E}$, all in roughly equal amounts (Figure II.8.c). As shown in Figure II.8.d, (bottom panel), a typical circuit in cluster 4 has inhibitory populations saturated near the maximum firing-rates. Thus, decreasing $\hat{g}_{I1E}$ and $\hat{g}_{I2E}$ by just a little greatly decreases the firing-rate signal relayed by the two inhibitory populations, whereas decreasing decreasing $\hat{g}_{A\beta I1}$ and $\hat{g}_{A\beta I2}$ has little effect on those firing-rates. Thus, with a fairly large decrease in the effect of inhibition from $I1$ and $I2$ on E, and with a fairly large increase in $\hat{g}_{A\beta E}$, (but with no change in $\hat{g}_{A\beta I1}$ and $\hat{g}_{A\beta I2}$), the excitatory population relays pain-inducing signals towards the brain in response to typically innocuous $A\beta$ stimuli, thereby producing allodynia.

Further, again because the majority of APS points belong to cluster 3, the static circuits as a whole are most sensitive to breakage by disinhibition, where the two inhibitory populations are less activated by innocuous $A\beta$ stimuli. As the remainder of the APS largely lies in equal proportions in cluster 1 and cluster 2, another likely mechanism for allodynia is through disinhibition caused by only one of the inhibitory populations becoming less activated by innocuous stimuli. Thus, the static circuit is most sensitive to breakage by disabling the inhibitory control on E cells, by making either one or both of the two inhibitory populations less active under typical innocuous $A\beta$ stimuli.

2.5 Discussion, summary and takeaways

In this work, we construct a methodology for conducting parameter sensitivity analyses on small neural population ircuits. The methodology involves the following steps:

1. Translating normal and pathological experimental behaviors that the circuits should replicate into analytical conditions that model variables must satisfy;

2. Re-framing these conditions into systems of inequalities and optimization problems that the parameters of interest must satisfy. In our work, we focused on the parameters governing the coupling strengths between populations. These analytically determined conditions describe distinct regions of parameter space in which the corresponding circuits display normal behaviors (the APS) and above which display pathological behavior (above the allodynia surface).

3. Determine the most likely mechanisms that induce the pathological condition, in our case allodynia, by finding the shortest path from each parameter set in the APS to the surface at the boundary of the pathological region

We apply this analysis to pain-processing circuits in layer I-II of the dorsal horn to identify different likely mechanisms for allodynia.

Biological interpretation of our results

Our results indicate that under normal conditions, all three of the circuits we investigate – the simple "gate control" circuit, and the circuits mediating dynamic and static allodynia – exhibit E-I balance. In such a balance, direct excitatory signaling from $A\beta$ fibers towards pain-relaying excitatory populations is balanced by inhibition from inhibitory populations.

Each circuit motif shows E-I balance in different ways

In the simple circuit, when the impact of excitatory $A\beta$ signals on the pain-relaying population, E, is large, generally the impact of inhibitory signaling from I on E is large, as well. This is shown by a strong positive correlation between the coupling strengths $g_{A\beta E}$ and $g_{A\beta I}$.

For the dynamic circuit, we see E-I balance in a similar form. However, the dynamic circuit consists of two subcircuits–one subcircuit which consists of the pain-relaying $E2$ population and the $I2$ population which directly inhibits $E2$, and a second subcircuit which consists of $I1$ and $E1$ coupled together to form a simple circuit. As a result, E-I balance occurs in both the $E2$-$I2$ and the $E1$-$I1$ subcircuits. Namely, when the impact of excitatory $A\beta$ signals on the $E1$ and $E2$ populations are large, generally so are the impacts of direct inhibitory signaling from $I2$ on $E2$ and from $I1$ on $E1$.

For the static circuit, E-I balance is obtained similarly. Indeed, under normal conditions, strong inhibitory signaling from either of the two inhibitory populations ($I1$ and $I2$) to the excitatory population (E) is balanced by strong excitation from $A\beta$ fibers to E. However, because there are two inhibitory populations simultaneously signaling E, the $A\beta$ to E excitatory signaling needs to be stronger than in either the simple or the dynamic circuit, so as to balance the extra inhibition.

Disruption of E-I balance produces allodynia

Our results indicate that allodynia is induced by disrupting these E-I balance states. Further, the particular ways in which the E-I balance is most efficiently disrupted in these circuits to create allodynia varies within and across circuit motifs. Such variations in pathological disruptions of E-I balance are indicated by clustering of the shortest paths from the allowable parameter space in which a circuit functions properly to the boundary of the pathological region, namely the allodynia surface.

Indeed, we see two different trends in the means of E-I disruption by which circuits are most vulnerable to producing allodynia. One trend involves reducing the impact of inhibitory signaling on excitatory populations while simultaneously over-exciting them. In this case, E-I balance is disrupted by excitatory populations "escaping" from inhibitory control. The other trend involves reducing the activity of inhibitory populations as well as over-exciting excitatory populations. In this case the disruption of E-I balance occurs due to excitatory populations being "released" from inhibitory control.

In particular, the over-excitation can be caused from an increased impact of signaling from $A\beta$ fibers on the excitatory population. Such over-excitation could arise from sprouting of synapses [WSC92], or upregulating the neurotransmitters corresponding to the synaptic inputs, for example (see [San09]). Reducing the impact of inhibitory signaling on the excitatory population, on the other hand, could be caused by loss of synapses from the inhibitory to the excitatory population [PPF$^+$15], reduced neurotransmitter output from inhibitory cell populations [MKK$^+$02], dysregulation of the post-synaptic effects of inhibitory neurotransmission (due to chloride ion concentration dysregulation) [Wes22], or even loss of inhibitory cells [MKK$^+$02], although there is some debate about the relevance of losing inhibitory cells to allodynia [San09].

E-I balance disruptions through escape from inhibitory control occur for all three circuits. For simple circuits, the most likely disruption is manifested by simultaneously reducing $\hat{g}_{IE}$ and increasing $\hat{g}_{A\beta E}$. About half of all simple circuits are most vulnerable towards producing allodynia via the escape disruption. Such circuits generally have larger $\hat{g}_{A\beta I}$ values, causing the firing activity of inhibitory populations to saturate in response to typical $A\beta$ signaling, making them particularly sensitive to loss of inhibitory gating of the E population by reducing the impact of inhibitory signaling from I to E.

For dynamic circuits, such disruptions are manifested either in the $I2$-$E2$ subcircuit by simultaneously reducing the impact of inhibitory signaling on the pain-relaying $E2$ population (decreasing g_{I2E2}) and increasing the impact of $A\beta$ signaling on $E2$ (increasing $g_{A\beta E2}$).The $I1$-$E1$ subcircuit can be similarly disrupted by increasing the impact of $A\beta$ signaling on $E1$ (increasing $g_{A\beta E1}$) and decreasing the efficacy of inhibitory input to $E1$ (decreasing g_{I1E1}). For the plurality of circuits (about 49%), the most likely disruption is in the $I2$-$E2$ subcircuit, whereas or only 0.3% of circuits would the disruption be due to escape from inhibitory control in the $I1$-$E1$ subcircuit.

For static circuits, E-I balance disruption due to escape from inhibition can be manifested by reducing $\hat{g}_{A\beta E}$ and in approximately equal proportions decreasing the impact of $I1$ on E (decreasing g_{I1E}) and decreasing the impact of $I2$ on E (decreasing g_{I2E}). However, only about 5% of static circuits are most vulnerable to such a disruption.

In the second trend, where E-I balance occurs due to "release from inhibitory control", circuits are most vulnerable to allodynia by reducing the activity of inhibitory populations themselves and simultaneously over-exciting an excitatory population. Such reduction of activity of inhibitory populations would likely arise by pruning of $A\beta$ - I synapses, downregulation of synaptic signaling on the I cells, or even reduced neurotransmitter output in synaptic clefts in $A\beta$ - I synapses.

Such a disruption in E-I balance also occurs mainly in all circuit types, and generally requires weaker excitation of inhibitory populations from $A\beta$ signals so that inhibitory firing activity doesn't saturate. For the simple circuit, such a disruption is manifested primarily by reducing $\hat{g}_{A\beta I}$ and simultaneously but to a lesser extent increasing $\hat{g}_{A\beta E}$, (and to an even lesser extent decreasing $\hat{g}_{IE}$).

For static circuits, such a disruption is manifested in several different ways. For instance, in cluster 2 of static circuits, the disruption happens by primarily decreasing the impact of $A\beta$ signaling on the $I1$ inhibitory population (by decreasing $g_{A\beta I1}$), whereas for cluster 1, the disruption happens primarily by decreasing the strength of the impact of $A\beta$ fibers on $I2$ (by decreasing $g_{A\beta I2}$). For yet another cluster (cluster 3), such disruption happens by simultaneously and in roughly equal proportions decreasing the strength of impact of $A\beta$ on both $I1$ and $I2$. Unsurprisingly, for the cluster where we produce such disruption by decreasing $g_{A\beta I1}$, $g_{A\beta I1}$ is only moderately large, so that the corresponding activation function is near its steepest value (where the firing-rate is about half of its maximum), making the circuit quite sensitive to changing $g_{A\beta I1}$. Similarly, for the cluster where such disruption happens by decreasing $g_{A\beta I2}$, $g_{A\beta I2}$ is only moderately large, whereas in the third cluster, both $g_{A\beta I1}$ and $g_{A\beta I2}$ are moderately large.

Hence, we see indeed that the means by which circuits are most sensitive to producing allodynia varies both according to circuit motif and within the allowable parameter space for a specific motif. For instance, all three circuit motifs–simple, dynamic, and static–have properly behaving circuits which are most sensitive to allodynia via "escape from inhibitory control" and others most sensitive to allodynia via "release from inhibitory control". Hence, there is indeed variation with each motif. On the other hand, disrupted E-I balance due to "release from inhibitory control", for example, can occur in static circuits by simultaneous reducing the activity of two inhibitory populations ($I1$ and $I2$), whereas in the dynamic circuit, "release from inhibitory control" involves reducing the activity from predominantly one inhibitory population at a time. Thus, there is indeed variation across circuit motifs.

While our analysis has centered around the minimial coupling strength changes that result in allodynia, it is important to remember that allodynia can be induced by other, much more disruptive mechanisms. For instance, in all three circuits, it is possible to induce allodynia by sufficiently increasing the impact of $A\beta$ signaling, alone, on the excitatory population that relays pain-inducing

signals towards the brain so that it overcomes inhibitory control and fires in response to innocuous stimuli. Additionally, allodynia can be induced by sufficiently decreasing the impact of inhibitory input on the pain-relaying excitatory population. However, as larger magnitude changes would be required in these scenarios, it may be presumed that they are less likely to occur.

Limitations

While our work identifies how E-I balance may be disrupted so as to produce allodynia, and suggests underlying biological mechanisms by which such disruptions occur, our work does not specifically identify such biological mechanisms. For instance, our work suggests that to most efficiently produce allodynia, the firing activity of inhibitory neurons in the static allodynia circuit typically decreases, and thus that the impact of $A\beta$ signaling on inhibitory neurons decreases. This in turn suggests that the effect of the neurotransmitters released at the corresponding synapses could be downregulated, that the number of $A\beta$-inhibitory interneuron synapses has decreased, or that the amount of neurotransmitter released at $A\beta$ has gone down, for instance. However, this does not specify which of such specific biological mechanisms has occurred.

Further, while the work we have conducted focuses on $A\beta$-fibers, it is possible to incorporate the pain-inducing signals from C fibers into our model. Upon doing so, and upon incorporating slowly-varying magnesium currents driven by NMDA released by the C-fibers as in [CPHB19], the model reproduces wind-up and shows reduced C-fiber-induced firing activity on the pain-relaying excitatory population upon receiving $A\beta$ signals, further confirming that the model is realistic.

Possible applications of the study outside of pain-processing

The sensitivity analysis methodology that we have presented in this paper doesn't only apply to pain-processing circuit in the spinal cord dorsal horn, although we do have follow through studies in mind related to understanding how spinal cord stimulation can relieve chronic pain (see Section 5.1). Indeed, the methodology can be used to study any small system of firing-rate models. Although, because implementing the methodology can be somewhat computationally intensive, the methodology may be primarily useful for studying small rather than larger systems of firing-rate models.

Namely, the methodology can be seen to generalize beyond systems of firing-rate model by thinking of circuits of firing-rate models instead as a network. Then, populations of neurons become "nodes" in the network, and coupling strengths become "edge weights". Each node would then linearly sum weighted inputs from other nodes, and would in turn provide signals to other nodes. In particular, each node would be equipped with an activation function, which applies a nonlinear

transform, such as the sigmoidal frequency-voltage relations used in this paper or such as the variety of activation functions, e.g. sigmoids, hyperbolic tangent, and ReLU [GBB11] commonly used in today's artificial neural networks [ESYF+20, AZH+21, CCKC+20].

Our methodology would then be useful for investigating "state changes", such as when the inputs for a particular node exceed a certain threshold (which e.g. would cause allodynia). In particular, our methodology would provide a means to quantify the sensitivity of the network towards changing states. Our methodology would further break the space of edge weights into regions, according to the specific changes in edge weights which most efficiently induce the network to reach the target state. As stated, our methodology is particularly applicable to feedforward networks, as this would allow our sampling algorithm to be applied.

2.6 Methods

2.6.1 Population firing rate model

For our models of layer I-II dorsal horn neuronal circuits, we implement a well-established firing rate model formalism that models the average membrane voltage and average firing rates of neuronal populations[2] (see e.g. [Fre75] or the Phillips-Robinson model [PR07]). The efficacy of synaptic coupling between populations is modeled as proportional to pre-synaptic population firing rates. Let the populations in the circuit be labeled as $x_1, x_2, ..., x_n$. Strengths of synaptic coupling between these populations are given by the parameters $g_{x_i x_j}$, where $g_{x_i x_j} < 0$ is equivalent to population x_i inhibiting population x_j and likewise $g_{x_i x_j} > 0$ is equivalent to population x_i exciting population x_j. In our circuits, populations also receive external input signals f_{input} within some normal range, $f_{input} \in [f_{input,l}, f_{input,u}]$.

Average firing rates f_{x_i} are computed from average voltage, V_{x_i} with a sigmoidal activation function of the form:

$$f_{x_i}(V_{x_i}) = 0.5 max_i \left(1 + \tanh\left(\frac{V_{x_i} - \alpha_i}{\beta_i}\right)\right), \tag{II.2}$$

where max_i is the maximum firing rate of the population, β_i is the half-activation voltage and α_i governs the slope of the population's firing-rate response to voltage changes. We match these

[2] This, in fact, is effectively the Jansen-Rit model [JR95], where instead of using an α function (see Section 1.4.5) to model the response of a population of neurons to presynaptic signaling, we use an exponential function of the form

$$h(t) = \begin{cases} \frac{1}{\tau} e^{-t/\tau} & t \geq 0 \\ 0 & t < 0 \end{cases} \qquad \text{(exponential response function)},$$

where just as for the α function, τ is a time constant. τ governs the rate at which the response decays. h can be used in place of [Fre75] or as an approximation to [ZVvPTH14] an α function.

parameters to experimental measurements of frequency-voltage relationships for dorsal horn neurons as described in Section 2.6.2. We expect that in the absence of inputs, V_{x_i} remains at a rest value $V_{i,rest}$. In response to synaptic inputs from other populations, V_{x_i} deviates from its rest value according to the following differential equation[3]:

$$\frac{dV_{x_i}}{dt} = \frac{\text{inputs} - V_{x_i} + V_{i,rest}}{\tau_i}.$$

where the inputs are the output firing-rates of populations pre-synaptic to x_i weighted by the corresponding coupling strengths as well as external input:

$$\text{inputs} = g_{y_1 x_i} f_{y_1} + g_{y_2 x_i} f_{y_2} + ... + f_{input}.$$

We thus expect that in the presence of steady inputs, V_{x_i} approaches a value equal to the net input signal above its resting voltage. The time constant τ_i governs how quickly V_x moves towards its steady state value given by

$$V_{x_i} = \text{inputs} + V_{i,rest} \text{ (steady state)}$$

Table II.4 summarizes the model equations for each circuit we analyze.

[3]This equation is analogous to the equation for change in voltage from the Jansen-Rit model [JR95], except where we have replaced the α-function response with the exponential function h. Indeed, using the exponential function, the convolution integral describing the average voltage $V(t)$ due to signaling from inputs to the neuron,

$$\Delta V(t) = \int_0^t h(t - t') \cdot \text{inputs}(t') dt'$$

is the solution, (as seen via Laplace transforms), to the following ordinary differential equation:

$$(\Delta V)'(t) = [\text{inputs} - \Delta V(t)] / \tau$$

Re-writing the differential equation back in terms of $V(t)$, and we obtain

$$V'(t) = [\text{inputs} - V + V_{rest}] / \tau,$$

as desired.

Circuit Motif	Population	Equation
Simple circuit	I	$\frac{dV_I}{dt} = \frac{(g_{A\beta I} f_{A\beta} + V_{I,rest}) - V_I}{\tau_I}$
	E	$\frac{dV_E}{dt} = \frac{(g_{A\beta E} f_{A\beta} - g_{IE} f_I(V_I) + V_{I,rest}) - V_E}{\tau_E}$
Dynamic Circuit	$I1$	$\frac{dV_{I1}}{dt} = \frac{(g_{A\beta I1} f_{A\beta} + V_{I1,rest}) - V_{I1}}{\tau_I}$
	$E1$	$\frac{dV_{E1}}{dt} = \frac{(g_{A\beta E1} f_{A\beta} - g_{I1E1} f_{I1}(V_{I1}) + V_{E1,rest}) - V_{E1}}{\tau_E}$
	$I2$	$\frac{dV_{I2}}{dt} = \frac{(g_{A\beta I2} f_{A\beta} + V_{I2,rest}) - V_{I2}}{\tau_I}$
	$E2$	$\frac{dV_{E2}}{dt} = \frac{(g_{A\beta E2} f_{A\beta} + g_{E1E2} f_{E2} - g_{I2E2} f_{I2} + V_{E2,rest}) - V_{E2}}{\tau_E}$
Static Circuit	$I1$	$\frac{dV_{I1}}{dt} = \frac{(g_{A\beta I1} f_{A\beta} + V_{I1,rest}) - V_{I1}}{\tau_I}$
	$I2$	$\frac{dV_{I2}}{dt} = \frac{(g_{A\beta I2} f_{A\beta} + V_{I2,rest}) - V_{I2}}{\tau_I}$
	E	$\frac{dV_E}{dt} = \frac{(g_{A\beta E} f_{A\beta} - g_{I1E} f_{I1} - g_{I2E} f_{I2} + V_{E,rest}) - V_E}{\tau_E}$

Table II.4: Model equations for each circuit. The model equations consist of a system of ordinary differential equations specifying how the voltages of populations of neurons evolve over time. Thus, for the simple circuit, the system contains two equations–one for the I population, and the other for the E population. Likewise, for the dynamic circuit, the system contains four differential equations, and for the static circuit, the system contains 3 differential equations.

2.6.2 Parameters of neuronal population models

We choose parameters for the activation functions of excitatory and inhibitory populations based on the experimental measurements of membrane properties and firing behavior in rat dorsal horn neurons reported in Ruscheweyh et al. in [RS02]. In our circuits, we assume all excitatory populations have the same parameters, and assume the same for inhibitory populations. We assume that average resting voltages $V_{I,rest}$ and $V_{E,rest}$ are approximately the values reported in [RS02]

$$V_{I,rest} = V_{E,rest} = -60 \text{ mV}.$$

Maximum firing rates of excitatory and inhibitory populations are set to

$$max_e = 50 \text{ Hz}$$
$$max_i = 80 \text{ Hz}$$

based on [MSH+22].

We use the frequency-current relations and current-voltage relations reported in [RS02] to extract frequency-voltage relations so as to fit firing rate activation functions given in Equation

II.2. Using Matlab's curve-fitting functionality, we obtain the values of β_i and α_i ($i = E, I$) for excitatory and inhibitory populations as shown in the Table (II.5).

Maximum and minimum voltages are set to $12\alpha_i$ mV above and below, respectively, the half-activation voltage value β_i. The voltage thresholds for firing are defined as $\beta_i - \alpha_i$. This ensures that these properties are defined similarly in both populations.

Population type	**Parameter category**	**Parameter name**	**Parameter symbol**	**Parameter value**
I populations	Activation function parameter	flatness	α_I	9.3 mV^{-1}
		half-activation	β_I	-30 mV
		Max firing rate	max_I	80 Hz
	Voltage cutoff	Min voltage	$V_{I,min}$	$\beta_I - 12\alpha_I$ mV
		Resting voltage	$V_{I,rest}$	-60 mV
		Firing threshold	$V_{I,thr}$	$\beta_I - \alpha_I$ mV
		Max voltage	$V_{I,max}$	$\beta_I + 12\alpha_I$ mV
	ODE specific parameters	Time constant	τ_I	0.02 s
E populations	Activation function parameter	flatness	α_E	7.9 mV^{-1}
		half-activation	β_E	-17 mV
		Max firing rate	max_E	50 Hz
	Voltage cutoff	Min voltage	$V_{E,min}$	$\beta_E - 12\alpha_E$ mV
		Resting voltage	$V_{E,rest}$	-60 mV
		Firing threshold	$V_{E,thr}$	$\beta_E - \alpha_E$ mV
		Max voltage	$V_{E,max}$	$\beta_E + 12\alpha_E$ mV
	ODE specific parameters	Time constant	τ_E	0.01 s

Table II.5: Model parameter values for each circuit. We divide parameters for the type of population (left column) they describe into 3 categories (2nd leftmost column): those that pertain to activation functions, those that give voltage cutoffs, and those that specifically concern the model differential equations. We further list the name of the parameter (middle column) and the symbol (2nd rightmost column) that we use to represent the parameter in the text. Finally we give the value (rightmost column) of the parameter that we use in simulations. All these parameters are fixed at the same value throughout the analysis and simulations.

Finally, we choose membrane time constants so they are roughly on the same time-scale as those

of [ZJG14]. Notably, as long as the time constants τ_E and τ_I for E and I populations, respectively, are both roughly on that time scale, the particular choice of τ_E and τ_I has little bearing on the qualitative behavior of the results. We choose $\tau_E = 0.01$ seconds and $\tau_I = 0.02$ seconds.

2.6.3 Defining the allowable parameter space (APS) for the circuits

Our parameter sensitivity analysis method consists of translating normal experimental behaviors that the circuits should replicate into analytical conditions on model variables, specifically on average voltages. These conditions are then re-written into systems of inequalities and optimization problems that coupling strength parameters must satisfy. The parameter sets that satisfy these systems constitute the allowable parameter space (APS). In this section, we derive these systems for our model circuits. Full details of the derivation are shown for the simple circuit. For the dynamic and static circuits, the systems are derived similarly and only some details are given here.

2.6.3.1 Simple circuit

For the simple circuit, it is possible to make considerable progress towards deriving an explicit definition of the allowable parameter space. As described in Table II.1 in Section 2.2.2, the conditions on average voltages that ensure that the circuit replicates experimentally appropriate behaviors are given as follows:

$$\begin{cases} V_{I,max} & \geq g_{A\beta I} f_{A\beta} + V_{I,rest} \\ V_{I,thr} & \leq g_{A\beta I} f_{A\beta} + V_{I,rest} \\ V_{E,rest} & \geq g_{A\beta E} f_{A\beta} - g_{IE} f_I + V_{E,rest} \\ V_{E,min} & \leq g_{A\beta E} f_{A\beta} - g_{IE} f_I + V_{E,rest} \\ V_{E,max} & \geq g_{A\beta E} f_{A\beta} + V_{E,rest} \\ V_{E,thr} & \leq g_{A\beta E} f_{A\beta} + V_{E,rest} \end{cases} \tag{II.3}$$

The first two inequalities can be rewritten to yield bounds on $g_{A\beta I}$:

$$\frac{V_{I,thr} - V_{I,rest}}{f_{A\beta}} \leq g_{A\beta I} \leq \frac{V_{I,max} - V_{I,rest}}{f_{A\beta}}.$$

However, as these conditions must hold for all $f_{A\beta} \in [f_{A\beta,min}, f_{A\beta,max}]$ Hz (considered to be $[10, 20]$ Hz), we need that

$$\max_{f_{A\beta}} \frac{V_{I,thr} - V_{I,rest}}{f_{A\beta}} \leq g_{A\beta I} \leq \min_{f_{A\beta}} \frac{V_{I,max} - V_{I,rest}}{f_{A\beta}},$$

which can be re-written as

$$\frac{V_{I,thr} - V_{I,rest}}{f_{A\beta,min}} \leq g_{A\beta I} \leq \frac{V_{I,max} - V_{I,rest}}{f_{A\beta,max}}.$$

Likewise, the last two inequalities from the set in Equation II.3 yield analogous bounds on $g_{A\beta E}$:

$$\frac{V_{E,thr} - V_{E,rest}}{f_{A\beta,min}} \leq g_{A\beta E} \leq \frac{V_{E,max} - V_{E,rest}}{f_{A\beta,max}}.$$

The middle two inequalities can similarly be rewritten so they give bounds on $g_{A\beta E}$ in terms of f_I and g_{IE}:

$$\max_{f_{A\beta}} \frac{g_{IE} f_I - (V_{E,rest} - V_{E,min})}{f_{A\beta}} \leq g_{A\beta E} \leq \min_{f_{A\beta}} \frac{g_{IE} f_I}{f_{A\beta}}. \tag{II.4}$$

We summarize the resulting system of inequalities in Table II.6. Notably, writing the inequalities as we do in the table induces a hierarchy on the inequalities, where the E inequalities depend on the solution to the I inequalities, and are thus higher in the hierarchy. This will prove helpful when sampling from the allowable parameter space.

Population	**Condition**	**Inequalities**
I	Upper bound (UB) + fires	$\frac{V_{I,thr}-V_{I,rest}}{f_{A\beta,min}} \leq g_{A\beta I} \leq \frac{V_{I,max}-V_{I,rest}}{f_{A\beta,max}}$
E	UB + fires (I1 abl.)	$\frac{V_{E,thr}-V_{E,rest}}{f_{A\beta,min}} \leq g_{A\beta E} \leq \frac{V_{E,max}-V_{E,rest}}{f_{A\beta,max}}$
	LB + pain inhibition	$\max_{f_{A\beta}} \frac{g_{IE}f_I-(V_{E,rest}-V_{E,min})}{f_{A\beta}} \leq g_{A\beta E} \leq \min_{f_{A\beta}} \frac{g_{IE}f_I}{f_{A\beta}}$

Table II.6: Inequalities expressed as upper and lower bounds on coupling strengths define the allowable parameter space for the simple circuit. These inequalities are obtained by algebraically manipulating the inequalities on the voltages of various populations from (Table II.1) that define the APS for the simple circuit so that the inequalities are written explicitly in terms of coupling strengths.

However, because f_I is a rescaled $\tanh$ function whose inputs depend on both $f_{A\beta}$ and $g_{A\beta I}$, the maximization and minimization problems in Equation II.4 are generally difficult to solve explicitly. Nevertheless, we are able to find the explicit solution to the minimization problem. To do so, note that both expressions to be optimized in Equation II.4 can be rewritten as:

$$
\begin{aligned}
\frac{g_{IE} max_I f_I + C}{f_{A\beta}} &= \frac{g_{IE} 0.5 \left(1 + \tanh\left((V_I - \beta_I)/\alpha_I\right)\right) + C}{f_{A\beta}} \\
&= \frac{g_{IE} max_I 0.5 \left(1 + \tanh\left((g_{A\beta I} f_{A\beta} - \beta_I)/\alpha_I\right)\right) + C}{f_{A\beta}} \\
&= \frac{g_{IE} max_I 0.5 \left(1 + \tanh\left(g_{A\beta I} f_{A\beta}/\alpha_I + (V_{I,rest} - \beta_I)/\alpha_I\right)\right) + C}{f_{A\beta}} \\
&= \frac{g_{IE} max_I \left(g_{A\beta I}/\alpha_I\right) 0.5 \left(1 + \tanh\left(g_{A\beta I} f_{A\beta}/\alpha_I + (V_{I,rest} - \beta_I)/\alpha_I\right)\right) + C\left(g_{A\beta I}/\alpha_I\right)}{g_{A\beta I} f_{A\beta}/\alpha_I} \\
&= \frac{g_{IE} max_I \left(g_{A\beta I}/\alpha_I\right) 0.5 \tanh\left(g_{A\beta I} f_{A\beta}/\alpha_I + (V_{I,rest} - \beta_I)/\alpha_I\right) + \left(C\left(g_{A\beta I}/\alpha_I\right) + g_{IE} max_I \left(g_{A\beta I}/\alpha_I\right) 0.5\right)}{g_{A\beta I} f_{A\beta}/\alpha_I} \\
&= \frac{a \tanh\left(x - b\right) + c}{x},
\end{aligned}
$$

where

$$
\begin{aligned}
a &= 0.5 g_{IE} max_I \left(g_{A\beta I}/\alpha_I\right) \\
b &= -(V_{I,rest} - \beta_I)/\alpha_I \\
c &= C\left(g_{A\beta I}/\alpha_I\right) + 0.5 g_{IE}\left(\alpha_I/g_{A\beta I}\right).
\end{aligned}
$$

Consequentially, the solutions to the preceding optimization problems occur either at $f_{A\beta,min}$, $f_{A\beta,max}$, or at one of the critical points given by

$$
0 = \frac{d}{dx}\left[\frac{a \tanh\left(x - b\right) + c}{x}\right]. \tag{II.5}
$$

We show in Section 2.7.1 that when $C = 0$, the solutions of Equation II.5 are given in terms of the 0^{th}, W_0, and -1^{st}, W_{-1} branches of the Lambert-W function:

$$
x_{0,-1} = \frac{1}{2}\left(c/a - W_{0,-1}\left(-e^{1-2b}\right)\right) \tag{II.6}
$$

so long as x_0 or $x_{-1} \in [f_{A\beta,min}, f_{A\beta,max}]$ or at $x = f_{A\beta,min}$ or $x = f_{A\beta,max}$. We address the case that $C \neq 0$, also in Section 2.7.1, by showing how to take advantage of the structure of the optimization problem to solve it numerically. We also describe an alternative approach to solving the system that defines the APS or the basic circuit. We do not use such an alternative approach when conducting our analysis, but it motivates a formula that we do use elsewhere

The alternative approach involves rewriting the middle two inequalities so they give bounds on g_{IE} in terms of f_I and $g_{A\beta I}$:

$$
\max_{f_{A\beta}} \frac{g_{A\beta E} f_{A\beta} + V_{E,rest} - V_{E,min}}{f_I} \leq g_{IE} \leq \min_{f_{A\beta}} \frac{g_{A\beta E} f_{A\beta}}{f_I} \tag{II.7}
$$

Both of the expressions to be optimized in Equation II.7 can be rewritten as:

$$\frac{g_{A\beta E}f_{A\beta} + V_{E,rest} - V_{E,min}}{0.5\text{max}_I\left(1+\tanh((g_{A\beta I}f_{A\beta} + V_{I,rest} - \beta_I)/\alpha_I)\right)} = \frac{g_{A\beta E}\alpha_I}{0.5\text{max}_I g_{A\beta I}} \frac{g_{A\beta I}f_{A\beta}/\alpha_I + (V_{E,rest} - V_{E,min})/(g_{A\beta E}\alpha_I)}{1+\tanh((g_{A\beta I}f_{A\beta} + V_{I,rest} - \beta_I)/\alpha_I)},$$

whose local extrema are shared with the following:

$$\frac{g_{A\beta I}f_{A\beta}/\alpha_I + (V_{E,rest} - V_{E,min})/(g_{A\beta E}\alpha_I)}{1+\tanh((g_{A\beta I}f_{A\beta} + V_{I,rest} - \beta_I)/\alpha_I)} = \frac{x+A}{1+\tanh(x+B)}$$

Consequentially, the solutions to the preceding optimization problems (Equation II.7) occur either at $f_{A\beta,min}$, $f_{A\beta,max}$, or at one of the critical points which solves

$$0 = \frac{d}{dx}\left[\frac{x+A}{1+\tanh(x+B)}\right]. \tag{II.8}$$

where

$$A = (V_{E,rest} - V_{E,min})/(g_{A\beta E}\alpha_I)$$
$$B = (V_{I,rest} - \beta_I)/\alpha_I.$$

We show in section 2.7.1 that the solution to Equation II.8 is expressed in terms of the principal branch W_0 of the Lambert-W function:

$$x_0 = -\frac{1}{2}\left(1 + 2A + W_0\left(e^{2B-2A-1}\right)\right). \tag{II.9}$$

We make use of the solution (Equation II.9) to Equation II.8 to make sampling from the static circuit more computationally efficient.

2.6.3.2 Dynamic circuit

For the dynamic circuit, the inequalities on population voltages needed for the circuit to replicate experimentally observed behaviors are given in Table II.2 in Section 2.3.1. Following a similar derivation as for the simple circuit, we can rewrite the conditions as the system of inequalities and optimization problems for the coupling strengths given in Table II.7.

Population	Condition	Inequalities
I1	Fires + upper bound (UB)	$\frac{V_{I1,rest}-V_{I1,min}}{f_{A\beta,min}} \leq g_{A\beta I1} \leq \frac{V_{I1,max}-V_{I1,rest}}{f_{A\beta,max}}$
E1	LB + pain inhib (no abl.)	$\max_{f_{A\beta}}\left[\frac{g_{I1E1}f_{I1}-(V_{E1,rest}-V_{E1,min})}{f_{A\beta}}\right] \leq g_{A\beta E1} \leq \min_{f_{A\beta}}\left[\frac{g_{I1E}f_{I1}}{f_{A\beta}}\right]$
	Fires + UB (I1 abl.)	$\frac{V_{E1,thr}-V_{E1,rest}}{f_{A\beta,min}} \leq g_{A\beta E1} \leq \frac{V_{E1,max}-V_{E1,rest}}{f_{A\beta,max}}$
I2	Fires + UB	$\frac{V_{I2,rest}-V_{I2,min}}{f_{A\beta,min}} \leq g_{A\beta I2} \leq \frac{V_{I2,max}-V_{I2,rest}}{f_{A\beta,max}}$
E2	LB (E1 abl.) + pain inhib (no abl.)	$\max_{f_{A\beta}}\left[\frac{g_{I2E2}f_{I2}-(V_{E2,rest}-V_{E2,min})}{f_{A\beta}}\right] \leq g_{A\beta E2} \leq \min_{f_{A\beta}}\left[\frac{g_{I2E2}f_{I2}-g_{E1E2}f_{E1}}{f_{A\beta}}\right]$
	Fires + UB (I1 abl.)	$\max_{f_{A\beta}}\left[\frac{g_{I2E2}f_{I2}-g_{E1E2}f_{E1,abl}+(V_{E2,thr}-V_{E2,rest})}{f_{A\beta}}\right] \leq g_{A\beta E2} \leq \min_{f_{A\beta}}\left[\frac{g_{I2E2}f_{I2}-g_{E1E2}f_{E1,abl}+(V_{E2,max}-V_{E,rest})}{f_{A\beta}}\right]$
	Fires + UB (I2 abl.)	$\max_{f_{A\beta}}\left[\frac{-g_{E1E2}f_{E1}+(V_{E2,thr}-V_{E2,rest})}{f_{A\beta}}\right] \leq g_{A\beta E2} \leq \min_{f_{A\beta}}\left[\frac{-g_{E1E2}f_{E1}+(V_{E2,max}-V_{E2,rest})}{f_{A\beta}}\right]$

Table II.7: Inequalities expressed as upper and lower bounds on coupling strengths define the allowable parameter space for the dynamic circuit. These inequalities are obtained by algebraically manipulating the inequalities on the voltages of various populations from (Table II.2) that define the APS for the dynamic circuit so that the inequalities are written explicitly in terms of coupling strengths. "abl." means ablation, and "inhib" means inhibition.

These conditions were found in the same manner as described for the simple circuit. For instance, the lower bound for $g_{A\beta E1}$ on the second line of Table II.7 is arrived at from the condition that the voltage of the $E1$ population must not be too hyperpolarized under control conditions. This is analogous to how we arrive at the lower bound (bottom-most inequality in Table II.6) on $g_{A\beta E}$ from the simple circuit from the analogous lower bound on the voltage of E.

Notably, much as for the simple circuit, rewriting conditions in such a way induces a hierarchy on the sets of conditions. For the dynamic circuit in particular, $I1$ and $I2$ conditions are lowest in the hierarchy, followed by $E1$ conditions, whereas $E2$ conditions are highest in the hierarchy. This hierarchical structure, much as for the simple circuit, proves helpful for sampling from the allowable parameter space (APS).

Further, lower and upper bounds explicitly on g_{I1E1} and g_{I2E2} are absent from the Table. However, to be able to sample from the APS, we will need to be able to find upper and lower bounds on such parameters. Nevertheless, upper and lower bounds on g_{I1E1} are straightforward to find, because the $I1$-$E1$ portion of the dynamic circuit has identical constraints to those of the simple circuit as described in the preceding subsection. To find upper and lower bounds on g_{I2E2} specifically when sampling from the APS, on the other hand, we use a more computationally intensive approach to find the true upper and lower bounds. Namely, given the parameters for the $I1$-$E1$ portion of the circuit, and given g_{E1E2}, we find the set of g_{I2E2} values such that the inequalities in the bottom-most three lines of the the table (Table II.7) have a solution. That is, we choose g_{I2E2} so that the upper bounds on $g_{A\beta E2}$ appearing in the last three lines of the table (Table II.7) are indeed larger than each of the lower bounds on $g_{A\beta E2}$ also appearing in the last three lines of the table.

2.6.3.3 Static circuit

As we do for the simple and dynamic circuits, we can re-write the inequalities that population voltages in the static circuit must satisfy (Table II.3 in Section 2.4.1) as a system of inequalities and optimization problems for the coupling strengths to define the APS for the static circuit (Table II.8). To arrive at this system of inequalities and optimization problems for the coupling strengths, we proceed exactly as we did to arrive at the analogous systems for the dynamic and simple circuits.

As for the other circuits, rewriting conditions in such a way induces a hierarchy on the sets of conditions. For the static circuit, $I1$ and $I2$ conditions are lowest in the hierarchy, whereas E conditions are highest in the hierarchy. However, writing the inequalities as we do in the table (Table II.8) creates computationally difficulties.

One issue with the expressions in the table is that expressions such as

$$g_{A\beta E} \geq \max_{f_{A\beta}} \frac{(V_{E,thr} - V_{E,rest}) + g_{I2E} f_{I2}}{f_{A\beta}} \tag{II.10}$$

may be reduced to optimization problems of the form presented in Equation II.5 from Section 2.6.3.1:

$$0 = \frac{d}{dx}\left[\frac{a\tanh(x-b)+c}{x}\right],$$

where $c \neq 1$. However, we are unable to find a closed form solution for the preceding expression (Section 2.7.1). Such difficulties may be overcome either numerically, as outlined in Section 2.7.1, or by rewriting the original expression in terms of g_{I1E}, as in Equation II.7:

$$g_{I2E} \leq \min_{f_{A\beta}} \frac{g_{A\beta E} f_{A\beta} - (V_{E,thr} - V_{E,rest})}{f_{I2}}.$$

The preceding equation can be reduced to an optimization problem of the form described in Equation II.8:

$$0 = \frac{d}{dx}\left[\frac{x+A}{1+\tanh(x+B)}\right],$$

whose solution is given explicitly by Equation II.9. Because we can explicitly find the solution to the rewritten version of Equation II.10, we can gain a computational speed-up by using it when sampling the APS. To take advantage of that to some extent, we rewrite the third and fourth top-most lines of Table II.8 so they are expressed as inequalities on g_{I2E}. As a result, those equations cease

Population	Condition	Inequalities
I1	Fires + upper bound	$\frac{V_{I1,rest}-V_{I1,min}}{f_{A\beta,min}} \leq g_{A\beta I1} \leq \frac{V_{I1,max}-V_{I1,rest}}{f_{A\beta,max}}$
I2	Fires + upper bound	$\frac{V_{I2,rest}-V_{I2,min}}{f_{A\beta,min}} \leq g_{A\beta I2} \leq \frac{V_{I2,max}-V_{I2,rest}}{f_{A\beta,max}}$
E	Lower bound + pain inhib (no ablations)	$\max_{f_{A\beta}}\left[\frac{g_{I1E}f_{I1}+g_{I2E}f_{I2}-(V_{E,rest}-V_{E,min})}{f_{A\beta}}\right] \leq g_{A\beta E} \leq \min_{f_{A\beta}}\left[\frac{g_{I1E}f_{I1}+g_{I2E}f_{I2}}{f_{A\beta}}\right]$
	Fires + upper bound (I1 ablation)	$\max_{f_{A\beta}}\left[\frac{g_{I2E}f_{I2}+(V_{E,thr}-V_{E,rest})}{f_{A\beta}}\right] \leq g_{A\beta E} \leq \min_{f_{A\beta}}\left[\frac{g_{I2E}f_{I2}+(V_{E,max}-V_{E,rest})}{f_{A\beta}}\right]$
	Fires + upper bound (I2 ablation)	$\max_{f_{A\beta}}\left[\frac{g_{I1E}f_{I1}+(V_{E,thr}-V_{E,rest})}{f_{A\beta}}\right] \leq g_{A\beta E} \leq \min_{f_{A\beta}}\left[\frac{g_{I1E}f_{I1}+(V_{E,max}-V_{E,rest})}{f_{A\beta}}\right]$

Table II.8: Inequalities expressed as upper and lower bounds on coupling strengths define the allowable parameter space for the static circuit. These inequalities are obtained by algebraically manipulating the inequalities on the voltages of various populations from (Table II.3) that define the APS for the static circuit so that the inequalities are written explicitly in terms of coupling strengths.

being at the same level in the hierarchy as the inequality from the fifth line, and become highest in the hierarchy.

However, there is still another computational difficulty to overcome when sampling from the APS. Indeed, no conditions appearing in Table II.8 nor in the re-written version of Table II.8 as proposed in the preceding paragraph directly yield explicit bounds on g_{I1E}. Nevertheless, they do yield implicit bounds. Namely, it is possible to use combinations of conditions to extract upper and lower bounds on these two coupling strengths.

To find an upper bound on g_{I1E}, we can use the upper bound on E during $I1$-ablation along with the E lower bound given no ablations to obtain that

$$\max_{f_{A\beta}}\left[\frac{g_{I1E}f_{I1}+g_{I2E}f_{I2}-(V_{E,rest}-V_{E,min})}{f_{A\beta}}\right] \leq g_{A\beta E} \leq \min_{f_{A\beta}}\left[\frac{g_{I2E}f_{I2}+(V_{E,max}-(V_{E,rest})}{f_{A\beta}}\right] \Rightarrow$$
$$\max_{f_{A\beta}}\left[\frac{g_{I1E}f_{I1}+g_{I2E}f_{I2}-(V_{E,rest}-V_{E,min})}{f_{A\beta}}\right] \leq \min_{f_{A\beta}}\left[\frac{g_{I2E}f_{I2}+(V_{E,max}-V_{E,rest})}{f_{A\beta}}\right],$$

which requires in particular that for all $f_{A\beta} \in [f_{A\beta,min}, f_{A\beta,max}]$

$$\begin{aligned}
\frac{g_{I1E}f_{I1}+g_{I2E}f_{I2}-(V_{E,rest}-V_{E,min})}{f_{A\beta}} &\leq \frac{g_{I2E}f_{I2}+(V_{E,max}-V_{E,rest})}{f_{A\beta}} \Leftrightarrow \\
g_{I1E}f_{I1}+g_{I2E}f_{I2}-(V_{E,rest}-V_{E,min}) &\leq g_{I2E}f_{I2}+(V_{E,max}-V_{E,rest}) \Leftrightarrow \\
g_{I1E}f_{I1}-(V_{E,rest}-V_{E,min}) &\leq V_{E,thr}-V_{E,rest} \Leftrightarrow \\
g_{I1E}f_{I1} &\leq V_{E,max}-V_{E,min} \Leftrightarrow \\
g_{I1E}f_{I1,max} &\leq V_{E,max}-V_{E,min} \Leftrightarrow \\
g_{I1E} &\leq \frac{V_{E,max}-V_{E,min}}{f_{I1,max}}.
\end{aligned}$$

To find a lower bound on g_{I1E}, on the other hand, we can use the E lower bound under $I1$ ablation along with the pain inhibition condition, i.e. that:

$$\max_{f_{A\beta}}\left[\frac{g_{I2E}f_{I2}+(V_{E,thr}-V_{E,rest})}{f_{A\beta}}\right]\leq g_{A\beta E}\leq\min_{f_{A\beta}}\left[\frac{g_{I1E}f_{I1}+g_{I2E}f_{I2}}{f_{A\beta}}\right]\Rightarrow$$
$$\max_{f_{A\beta}}\left[\frac{g_{I2E}f_{I2}+(V_{E,thr}-V_{E,rest})}{f_{A\beta}}\right]\leq\min_{f_{A\beta}}\left[\frac{g_{I1E}f_{I1}+g_{I2E}f_{I2}}{f_{A\beta}}\right].$$

The preceding condition requires that for all $f_{A\beta}\in[f_{A\beta,min},f_{A\beta,max}]$,

$$\begin{aligned}\frac{g_{I2E}f_{I2}+(V_{E,thr}-V_{E,rest})}{f_{A\beta}}&\leq\frac{g_{I1E}f_{I1}+g_{I2E}f_{I2}}{f_{A\beta}}\Leftrightarrow\\ g_{I2E}f_{I2}+(V_{E,thr}-V_{E,rest})&\leq g_{I1E}f_{I1}+g_{I2E}f_{I2}\Leftrightarrow\\ V_{E,thr}-V_{E,rest}&\leq g_{I1E}f_{I1}\Leftrightarrow\\ V_{E,thr}-V_{E,rest}&\leq g_{I1E}f_{I1,min}\Leftrightarrow\\ \frac{V_{E,thr}-V_{E,rest}}{f_{I1,min}}&\leq g_{I1E}.\end{aligned}$$

An analogous procedure produces bounds on g_{I2E}. However, the bounds on g_{I1E} and g_{I2E} from the preceding procedure are loose. For example, g_{I1E} has to be larger than such a lower bound, but there might never be a circumstance where g_{I1E} actually nears that lower bound. We can attempt to find maximally tight bounds on g_{I1E} and g_{I2E}, but doing so proves prohibitively difficult, and involves finding the zeros of expressions wherein Lambert-W functions appear multiple times.

2.6.4 Sampling from the allowable parameter space (APS)

To obtain statistics about the allowable parameter space (APS), we want to sample from it uniformly in space. However, because the APS is complicated, high-dimensional, can be non-convex, and is generally of an unknown shape, it is difficult to sample uniformly. In this section we describe the algorithm we developed to sample the APS's for the model circuits. We describe the algorithm in terms of sampling of a general space in $\mathbb{R}^n$ constrained by a set of inequalities that follow a hierarchical structure.

In particular, we first rewrite the sets of inequalities from Table II.10 as a "simple" set of inequalities (Table II.11) on the coupling strengths. We describe simple sets of inequalities in the next subsubsection. From those simple sets of inequalities, we determine a rectangular subspace of $\mathbb{R}^n$ that contains the solution to the inequalities, i.e. the APS. We describe how to find those subspaces in the 2nd subsubsection. We use the resulting subspace to constrain the search space for sampling from the APS. We then describe in the third subsubsection the procedure to uniformly

sample from the APS by using the subspaces.

Simple Sets of Inequalities

Roughly, a simple set of inequalities is one that can be rewritten as a hierarchy in the following sense. Consider a set of inequalities on points in $\mathbb{R}^n$. We say the set of inequalities is "simple" if the inequalities induce bounds on coordinates for $\mathbb{R}^n$ in the following way:

- $a_1 \leq x_1 \leq b_1$
- $L_2(x_1) \leq x_2 \leq U_2(x_1)$ for $x_2 \in [a_2, b_2]$
- $L_3(x_1, x_2) \leq x_3 \leq U_3(x_1, x_2)$ for $x_3 \in [a_3, b_3]$

$\vdots$

- $L_n(x_1, x_2, \ldots, x_{n-1}) \leq x_n \leq U_n(x_1, x_2, \ldots, x_{n-1})$ for $x_n \in [a_n, b_n]$

for some labeling $x_1, x_2, \ldots x_n$ of the populations in the circuit, some real numbers $a_1 \leq b_1, \ldots a_n \leq b_n$, and some functionals $L_1 \leq U_1, \ldots, L_n \leq U_n$. The x_n inequalities would then be the highest in the hierarchy, because it depends on $x_1, \ldots, x_{n-1}$; the x_{n-1} would be 2nd highest, because it depends on all coupling strengths lower in the hierarchy; and so on.

We say that if the original inequalities are rewritten in the preceding form, than they are in "normal form". We see examples of such hierarchically written systems of inequalities for the simple, static, and dynamic circuits in Section 2.6.3. For the simple circuit for example, (Table II.6), $g_{A\beta I}$ inequalities are lowest in the hierarchy, followed by g_{IE} inequalities, and then by $g_{A\beta E}$ inequalities. For each circuit, we show the full hierarchies of inequalities in Table II.9.

Note, however, that for any particular set of simple inequalities, there might exist multiple normal forms. Further, note that we do not require that given any $x_1, \ldots x_{j-1}$ satisfying the first $j-1$ inequalities given above, that $U_{i_j}(x_1, x_2, \ldots, x_{j-1}) > L_n(x_1, x_2, \ldots, x_{j-1})$. Hence, the set of inequalities given above might not describe the exact solution to the original set of inequalities.

Circuit	Hierarchy of inequalities labeled by coupling strength
Simple	$g_{A\beta I}, g_{IE}, g_{A\beta E}$
Dynamic	$g_{A\beta I1}, g_{A\beta I2}, g_{I1E1}, g_{E1E2}, g_{I2E}, g_{A\beta E2}.$
Static	$g_{A\beta I1}, g_{A\beta I2}, g_{I1E}, g_{I2E}, g_{A\beta E}$

Table II.9: The hierarchy of inequalities on each coupling strength for each circuit. We see in Table II.6, Table II.7, and Table II.8, that the inequalities that define the APS for the simple, dynamic, and static circuits, respectively, may each be rewritten as a simple set of inequalities on the corresponding coupling strengths. This table shows the resulting hierarchy of inequalities, where inequalities higher in the hierarchy implicitly depend on the coupling strengths corresponding to inequalities lower in the hierarchy. Namely, we list the inequalities from lowest to highest in the hierarchy.

Finding a rectangular subspace that contains the APS

If we can write our set of inequalities in the normal form described above, then finding a rectangular subspace that contains the APS is fairly straightforward. Namely, for a simple set of inequalities, we can fairly easily identify for each coordinate x_i, (which in the context of our pain-processing circuits would be a coupling strength), the interval of values $[x_{i,min}, x_{i,max}]$ in which x_i lies when $\vec{x}$ is in the APS. The rectangular subspace may then be expressed as a Cartesian product:

$$[x_{1,min}, x_{1,max}] \times [x_{2,min}, x_{2,max}] \times \cdots \times [x_{n,min}, x_{n,max}]$$

In particular, to identify the values in which x_i lies when $\vec{x}$ is in the APS, we use the simple set of inequalities to generate random (not necessarily uniform-in-space) samples from the APS as follows:

- Uniformly at random choose a value of x_1 in $[a_1, b_1]$
- Given the value of x_1, uniformly at random choose a value of $x_2 \in [L_2(x_1), U_2(x_1)]$ if such an interval exists. If such an interval doesn't exist, start over.
- Repeat for $x_3, \ldots, x_n$

For each circuit, we generate at least 1000 samples. For each coupling strength, we take it's minimum value as the minimum value across all samples, and its maximum value as the maximum value across all such samples. This defines a rectangular subspace of $\mathbb{R}^n$ that contains the APS.

Naive sampling is inefficient

We now have the upper and lower bounds on each coupling strengths which specify a rectangular subspace of $\mathbb{R}^n$ which contains the APS. From here, we could now implement the naive strategy, where we uniformly at random sample a point from the rectangular subspace, check to see if the point is in the APS, and keep it if it is. However, recall that we wish to normalize the APS so that we can more easily compare the effects of changing coupling strengths. Thus, before applying the "naive strategy", we normalize coupling strengths so the upper bound is 1 and the lower bound is 0, so that all points in the APS lie in the unit hypercube after being normalized.

After normalizing[4], we could now implement the naive strategy. However, the naive sampling strategy has the potential to be computationally inefficient. Indeed, checking if a point is in the APS requires computing $L_1, U_1, L_2, U_2, \ldots L_n, U_n$, which, as discussed in Section 2.7.2, takes $O(n^2)$ to do, where n is the number of dimensions. Further, the point will be in the sample space with probability given by $volume(APS)$, and thus the number of iterations needed to find a point in the APS will be $1/volume(APS)$. Thus, the order of the naive algorithm is $O(n^2/volume(APS))$.

However, $O(n^2/volume(APS))$ is likely very bad. Heuristically, if the average width of the APS in the each component direction is α for some $\alpha \in (0, 1)$, then it is reasonable to expect that $volume(APS) \approx \alpha^n$, and the algorithm becomes $O(n^2/\alpha^n)$. Thus, the naive sampling algorithm is exponential in n!

More efficient sampling algorithm

Thus, we seek to construct a spatially uniform sampling algorithm whose complexity is independent of the volume of the APS. To do so, instead of blindly sampling from the unit hypercube, we sample from a cover of the normalized[5] APS by hyperrectangles that approximates the normalized APS. In particular, our algorithm takes the following strategy:

1. Define a set of hyperrectangles that contains and approximates the allowable parameter space.

2. Uniformly at random select a hyperrectangle

3. Uniformly at random select a point from the hyperrectangle.

[4]normalizing is not required to implement the naive strategy. However, the step in the naive algorithm where we uniformly at random sample points from rectangular space containing the APS is very easy if the rectangular space is a hypercube. Namely, one can simply choose the x_1 coordinate of the point by uniformly sampling from $[0, 1]$, and choose each other coordinate likewise.

[5]normlizing is also not required to implement this more efficient sampling algorithm. However, for the pain-processing circuits presented in this paper, we wish to normalize so that we can more effectively compare the effects of changing various coupling strengths. Thus, we present this more efficient sampling algorithm specifically after normalizing the APS

- If the point is not in the parameter space, discard it

4. Otherwise, keep the point with probability proportional to the volume of the hyperrectangle.

Step (4) ensures that the sample is indeed uniform-in-space. We illustrate the sampling algorithm in Figure II.9.

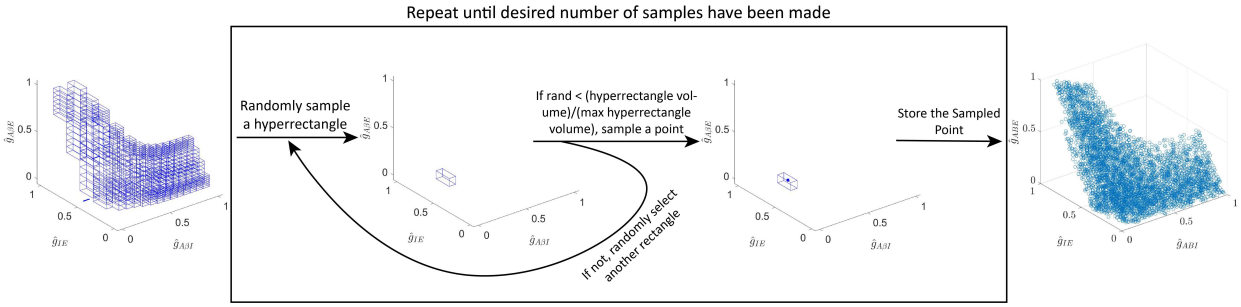


Figure II.9: Schematic of the sampling algorithm. Cover the allowable parameter space with hyperrectangles (left). Then, while the number of sampled points is smaller than the desired number of sampled points, randomly select a hyperrectangle (2nd from left). If a random number, "rand", is less than the ratio of the hyperrectangle's volume to the maximum volume of such hyperrectangles, keep the hyperrectangle (2nd from right); otherwise, randomly select another hyperrectangle. If the hyperrectangle is kept, randomly and uniformly-in-space select a point from the hyperrectangle, and store the point as part of the sample (right). In practice, rather than computing the full set of hyperrectangles that cover the allowable parameter space in advance, it is more efficient compute individual hyperrectangles as they are being randomly selected, as described in the implementation described in the main body of the text.

To implement the algorithm, we uniformly select a hyperrectangle from a set of hyperrectangles that covers the APS via steps (1) - (3) below. Specifically, to randomly select the hyperrectangle, in steps (1) - (3) we randomly select the subintervals whose Cartesian product defines the hyperrectangle. As a result, we avoid the need to precompute the hyperrectangle-cover of the APS. In step (4), we randomly select a point from the hypercube. In steps (5) - (6), we randomly keep the point with probability proportional to the hyperrectangles volume, and then discard the point if it fails to fall in the APS. We describe the details below:

1. Partition $[a_1, b_1]$ into m_1 equal-sized subintervals

   - Uniformly at random select a subinterval
   - Deterministically select a value of x_1 in that subinterval, e.g. by taking x_1 to be the left endpoint.

2. Given the value of x_1, compute $L_2(x_1)$ and $U_2(x_1)$.

   - If $L_2(x_1) \geq U_2(x_1)$, start over.
   - Otherwise, partition $[L_2(x_1), U_2(x_1)]$ into m_2 equal-sized subintervals.

- Append an interval of size $(U_2(x_1) - L_2(x_1))/m_2$ to either side of the partition
- Uniformly at random select a subinterval from the partition
- Deterministically select a value of x_2 in that sub-interval.

3. Repeat the preceding step for $x_3, \ldots, x_n$. We will have now obtained n subintervals whose cartesian product forms a hypercube
4. Uniformly at random select a point in the hypercube
5. Randomly keep the point with probability proportional to the volume of the hypercube. To do so,
   - Compute the maximal possible volume of the hypercube.
     - Because the whole normalized parameter space lies in the unit hypercube $[0, 1]\times...\times[0, 1]$ in $\mathbb{R}^n$, and because the selected hypercube has maximal length in dimension j given by $[U_j(x_1, x_2, \ldots, x_{j-1}) - L_j(x_1, x_2, \ldots, x_{j-1})]/m_j \leq 1/m_j$, the maximal possible volume of the hypercube is $\left(\Pi_{j=1}^n m_j\right)^{-1}$
   - Randomly select a number in [0, 1]. If that number is smaller than the volume of the selected hypercube divided by the maximum possible volume of selected hyperrectangles, keep the point.
6. If the selected point is not in the allowable parameter space, discard it.

We show in Section 2.7.2 that the time-complexity of the implementation is much better than that of the naive algorithm, so long as for all $j = 1, .., n$, the probability that the upper bound U_j is larger than the lower bound L_j is close to 1. Indeed, we show that:

$$\text{Time complexity of selecting a point} = \begin{cases} O(n^3) & U_j \geq L_j \text{ always for } j = 1, 2, \ldots, n \\ O\left(n\left(\frac{1}{\beta}\right)^{n-2}\right) & U_j < L_j \text{ with probability } \beta \text{ for } j = 1, 2, \ldots, n. \end{cases}$$

We prove that the algorithm and implementation indeed result in a uniform sample, in Section 2.7.3.

2.6.5 Defining the allodynia surface

In this section, we derive the conditions describing the allodynia surface for each circuit. We present the derivation in terms of a generalized circuit and finding conditions for a generalized "target state" to occur. In particular, we restrict our attention to circuits where the output signal is relayed by a single population of neurons, which we denote by y. The "target state" is represented

by the voltage of this output population V_y increasing above a specified threshold. For our firing rate model, the steady-state average voltage V_y is given by

$$V_{y,steady} = g_{in,y} f_{in} + \sum_{k=1}^{n} g_{x_k y} f_{x_k}(\vec{g}, f_{in}) + V_{y,rest}.$$

where f_{in} is an external input and f_{x_k} are firing rates of the populations pre-synaptic to population y. The circuit is in the "target state" when

$$V_{y,thr} \leq V_{y,steady}.$$

For the simple, dynamic and static circuits, the output populations are E, $E2$, and E, respectively, namely the excitatory populations which relay signals to projection neurons that send input to the brain. The target state for allodynia is that the average voltage of these excitatory populations increases above the firing threshold. These allodynia conditions are summarized in Table II.10.

Circuit	**Conditions under which allodynia is produced**
Simple	$V_{E,thr} \leq V_{E,steady} = g_{A\beta E} f_{A\beta} - g_{IE} f_I + V_{E,rest}$
Dynamic	$V_{E2,thr} \leq V_{E2,steady} = g_{A\beta E2} f_{A\beta} + g_{E1E2} f_{E1} - g_{I2E2} f_{I2} + V_{E2,rest}$
Static	$V_{E,thr} \leq V_{E,steady} = g_{A\beta E} f_{A\beta} - g_{I1E} f_{I1} - g_{I2E} f_{I2} + V_{E,rest}$

Table II.10: Allodynia conditions on each circuit. In particular, the allodynia conditions specify precisely when the circuit is producing allodynia–when it is relaying pain-inducing stimuli in response to innocuous $f_{A\beta}$ signals. These conditions are given by inequalities on the voltage of the population E, $E2$, and E, for the simple, dynamic, and static circuits, respectively, which relay signals towards the brain.

Using the allodynia condition, we can identify the circuit coupling strengths $\vec{g}$ for which the target state is attainable. To do so, we rewrite the condition defining the target state as an inequality on the coupling strength $g_{in,y}$ between the input signal and the population y:

$$g_{in,y} f_{in} \geq (V_{y,thr} - V_{y,rest}) - \sum_{k=1}^{n} g_{x_k y} f_{x_k}(\vec{g}, f_{in}) \Leftrightarrow \tag{II.11}$$

$$g_{in,y} \geq \frac{(V_{y,thr} - V_{y,rest}) - \sum_{k=1}^{n} g_{x_k y} f_{x_k}(\vec{g}, f_{in})}{f_{in}}. \tag{II.12}$$

For the target state to be attainable, we don't need to reach the target state for all values of f_{in}. Instead, we can reach the target state for the value of f_{in} that minimizes the right-hand side of the

preceding equation. Thus, the circuit corresponding to a set of coupling strengths $\vec{g}$ can attain the target state if and only if

$$g_{in,y} \geq \min_{f_{in} \in [f_{in,min}, f_{in,max}]} \frac{(V_{y,thr} - V_{y,rest}) - \sum_{k=1}^{n} g_{x_k y} f_{x_k}(\vec{g}, f_{in})}{f_{in}}. \tag{II.13}$$

To illustrate this more concretely, the conditions on the coupling strengths for which allodynia is attainable are summarized in Table II.11 for the simple, dynamic, and static circuits.

Circuit	**Condition for possible allodynia production**
Simple	$g_{A\beta E} \geq \min_{f_{A\beta} \in [f_{A\beta,min}, f_{A\beta,max}]} \frac{g_{IE} f_I + V_{E,thr} - V_{E,rest}}{f_{A\beta}}$
Dynamic	$g_{A\beta E2} \geq \min_{f_{A\beta} \in [f_{A\beta,min}, f_{A\beta,max}]} \frac{g_{I2E2} f_{I2} - g_{E1E2} f_{E1} + (V_{E,thr} - V_{E2,rest})}{f_{A\beta}}$
Static	$g_{A\beta E} \geq \min_{f_{A\beta} \in [f_{A\beta,min}, f_{A\beta,max}]} \frac{g_{I1E} f_{I1} + g_{I2E} f_{I2} + (V_{E,thr} - V_{E,rest})}{f_{A\beta}}$

Table II.11: The conditions under which a circuit can produce allodynia. These conditions no longer specify precisely when the circuit produces allodynia, but instead specify the sets of coupling strengths when the corresponding circuit will produce allodynia for some typical $f_{A\beta}$. As such, these conditions are defined by inequalities on the coupling strengths $g_{A\beta E}$, $g_{A\beta E2}$, and $g_{A\beta E}$ for the simple, dynamic and static circuits, respectively, between the pain-relaying population and the $A\beta$ input signal. Notably, the inequalities involve an optimization problem.

Equation II.13 thus defines a target state boundary surface S which divides the set of $\vec{g}$'s for which the circuit is in the target state from the set of $\vec{g}$'s for which it is not. We can express this surface as

$$S = \left\{ \vec{g} : g_{in,y} = \min_{f_{in} \in [f_{in,min}, f_{in,max}]} \left[\frac{(V_{y,thr} - V_{y,rest}) - \sum_{k=1}^{n} g_{x_k y} f_{x_1}(\vec{g}, f_{in})}{f_{in}} \right] \right\} \tag{II.14}$$

For the simple, dynamic, and static circuits, this boundary (Table II.12 separates circuits which can produce allodynia from those than cannot.

Circuit	Allodynia surface
Simple	$\left\{\vec{g} : g_{A\beta E} = \min_{f_{A\beta}\in[f_{A\beta,min},f_{A\beta,max}]} \frac{g_{IE}f_I+V_{E,thr}-V_{E,rest}}{f_{A\beta}}\right\}$
Dynamic	$\left\{\vec{g} : g_{A\beta E2} = \min_{f_{A\beta}\in[f_{A\beta,min},f_{A\beta,max}]} \frac{g_{I2E2}f_{I2}-g_{E1E2}f_{E1}+(V_{E,thr}-V_{E2,rest})}{f_{A\beta}}\right\}$
Static	$\left\{\vec{g} : g_{A\beta E} = \min_{f_{A\beta}\in[f_{A\beta,min},f_{A\beta,max}]} \frac{g_{I1E}f_{I1}+g_{I2E}f_{I2}+(V_{E,thr}-V_{E,rest})}{f_{A\beta}}\right\}$

Table II.12: The allodynia surface for the simple, dynamic and static circuits, respectively. The allodynia surface separates circuits which can produce allodynia in response to typical $f_{A\beta}$ signaling from those that cannot.

Properties of the target state boundary surface S

To understand the target state boundary surface, S, it is convenient to think of $g_{in,y}$ as a height. That is, the height h of a point $\vec{g}$ is its coupling strength $g_{in,y}$:

$$h(\vec{g}) := g_{in,y}.$$

We can then express S as the graph of the function

$$h_S(\vec{g}) := \min_{f_{in}} \left[\frac{(V_{y,thr} - V_{y,rest}) - \sum_{k=1}^{n} g_{x_k y} f_{x_1}(\vec{g}, f_{in})}{f_{in}} \right],$$

noting that h_S may depend on any coupling strength except for $g_{in,y}$. Then, circuits whose corresponding coupling strengths $\vec{g}$ with $h(\vec{g}) \geq h_S(\vec{g})$ are capable of producing the target state, and other circuits are not.

This notion of height makes it intuitively clear that S divides the space of coupling strengths into a part that can produce the target state, and a part that cannot. However, rigorously justifying that S does so is tricky. One approach towards such a justification would be to use an appropriate extension of the Jordan Curve Theorem (see [Hat02]) from algebraic topology. However, such "heavy machinery" is not necessary.

For a simple proof of the result, note that we are working in $\mathbb{R}^{n+1}$-dimensional space, and that

$$h : \mathbb{R}^{n+1} \to \mathbb{R}$$
$$h_S : \mathbb{R}^n \to \mathbb{R},$$

and that both h and h_S are continuous. In Section 2.7.4, we show that S, as well as any other graph of a continuous function from $\mathbb{R}^n \to \mathbb{R}$, divides space into two path connected components:

one that lies above the graph and another that lies below the graph. Moreover, S is the boundary between these two components.

2.6.6 Computing the distance between sampled points and the allodynia surface

In this section, we discuss the computational algorithm implemented to compute the shortest path from points in the APS to the allodynia surface. Continuing as above in Section 2.6.5, we describe the algorithm for a generalized circuit and a generalized target state boundary surface S.

The length of the shortest path between a point $\vec{g}$ in the space of coupling strengths to the target state boundary surface S indicates how easy it is to move the circuit into the target state. It also identifies which coupling strengths need to change to reach the target state. We illustrate the problem of finding the shortest path to S by considering an arbitrary set of coupling strengths or point in the parameter space

$$\vec{gp} = (gp_{x_1y}, \cdots, gp_{x_ny}, gp_{in,y}).$$

To compute the shortest path from $\vec{gp}$ to S, we need to find the point on S closest to $\vec{gp}$:

$$\vec{gs}_{nearest} = (gs_{x_1y}, \cdots, gs_{x_ny}, gs_{in,y}).$$

To do so, we need to solve the optimization problem:

$$\vec{gs}_{nearest} = \text{argmin}_{\vec{gs}\in S}||\vec{gp} - \vec{gs}||, \tag{II.15}$$

where $||\cdot||$ represents the 2-norm.

2.6.6.1 A more computationally efficient strategy

Solving the optimization problem in Equation II.15 directly would require knowing the target state surface itself or knowing important properties of it such as its gradient. However, the target state boundary surface in our work is defined via solving a minimization problem over the space of coupling strengths (excluding the coupling strength $g_{in,y}$) and is thus difficult to include in existing optimization algorithms. Thus, we seek to solve this problem without computing an explicit representation of the target state surface S.

Instead, we solve the higher dimensional problem

$$(\vec{gs^*}_{nearest}, f_{in,nearest}) := \text{argmin}_{(\vec{gv},f_{in})\in S^*}||\vec{gp} - \vec{gv}||, \tag{II.16}$$

where S^* is the following set of coupling strength-input signal pairs defined by removing the minimization from Equation II.14:

$$S^* := \left\{ (\vec{g}, f_{in}) : g_{in,y} = \frac{(V_{y,rest} - V_{y,thr}) + \sum_{j=1}^{n} g_{x_i y} f_{x_i}(\vec{g}, f_{in})}{f_{in}} \text{ and } f_{in} \in [f_{in,min}, f_{in,max}] \right\}.$$

Equation II.16 presents a constrained optimization problem that can be solved with high likelihood using global optimization algorithms based on stochastic gradient descent. As a result, Equation II.16 is far more tractable than the original problem posed in Equation II.15. Moreover, if $\vec{gs}^*_{nearest} = \vec{gs}_{nearest}$, the two minimization problems are equivalent, and by solving Equation II.16, we will have solved the original minimization problem Equation II.15. We show in Section 2.7.5 that if we solve Equation II.16, we do indeed solve Equation II.15.

2.6.6.2 Implementing the computationally efficient strategy

To implement our strategy, we solve the original optimization problem (Equation II.15) by finding a solution to the secondary optimization problem (Equation II.16). In particular, recall that the optimization problem from Equation II.16 is to find

$$\text{argmin}||\vec{gp} - \vec{g}||$$

over pairs $(\vec{g}, f_{in})$ subject to the conditions that

1. $g_{in,y} f_{in} = (V_{y,rest} - V_{y,thr}) + \sum_{j=1}^{n} g_{x_i y} f_{x_i}(\vec{g}, f_{in})$ for some $f_{in} \in [f_{in,min}, f_{in,max}]$, i.e. that a circuit with coupling strengths $\vec{g}$ and input signal f_{in} reaches the target state
2. $g_{x_1 y}, \ldots, g_{x_n y}, g_{in,y} \geq 0$
3. $f_{in} \in [f_{in,min}, f_{in,max}]$

To solve the problem numerically, we use a multi-start stochastic global optimization scheme:

1. For efficiency, we restrict the search space by requiring that
   (a) The coupling strengths $\hat{g}_{x_1 y}, \ldots, \hat{g}_{x_n y}, \hat{g}_{in,y} \in [-1, 2]$
   (b) Since we know that the desired $\hat{g}_{in,y} \geq gp_{in,y}$, we further restrict $\hat{g}_{in,y} \geq gp_{in,y}$
2. We apply Matlab's stochastic gradient descent-based algorithm *fmincon* to the preceding optimization problem with the preceding extra restrictions

3. We repeat the algorithm 15 times using start points selected uniformly at random from the preceding constraints, and take as the solution to the optimization problem the value of $\vec{g}$ yielding the smallest $||\vec{gp} - \vec{g}||$.

2.6.7 Clustering the data based on shortest paths to the allodynia surface

We use a density-based scanning clustering algorithm coupled with data visualization to identify clusters in the APS. To do so, we work with the points uniformly sampled from the APS. However, because we are interested in how circuits may be induced to produce allodynia, we do not simply cluster points in the APS by location. Rather, we cluster the APS by a different property of points: the corresponding shortest paths to the allodynia surface. Thus, when clustering the APS, we need not worry explicitly about the coordinates of the points, and instead we need only consider the shortest paths from the points to the allodynia surface. Thus, we apply the density-based scanning algorithm and the data visualization specifically to those shortest paths.

For the simple, dynamic, and static circuits, we see that the shortest paths to the allodynia surface, represented as displacements, divide into clusters (Figures II.4.d, II.6.c, and II.8.c), and thus so do points in the APS. Fortunately, for the simple circuits in particular, those displacements are three dimensional, allowing us to easily visualize the shortest paths as points in $\mathbb{R}^3$. We can use the visualizations of the shortest paths, then, to check which clustering algorithms will be successful. However we see for the simple circuit, as per Figure II.4.e, that the clusters of shortest paths are not necessarily spherical. In fact, for the simple circuits, the clusters are flat, with the blue cluster being a line. As a result, the popular k-means algorithms may fail to detect the clusters. Nevertheless, that shortest paths are tightly packed within spatially separated clusters bodes well for detecting clusters via density-based scanning.

Indeed, the clusters for this simple circuit may be detected using density-based scanning with perfect accuracy. For more complicated circuits, we see that density-based scanning similarly detects clusters with perfect accuracy. Namely, for the dynamic and static circuits, respectively, the colors of lines in the parallel plots of shortest paths shown in Figures II.6.c and II.8.c indicate the clusters assigned by density-based scanning. In particular, dividing the shortest paths into clusters as density-based scanning does exactly matches how one would spatially divide shortest paths in the parallel plot. Therefore, all data we present regarding clusters specifically uses the cluster membership specified via density-based scanning.

To apply density-based clustering to the data, we use Matlab's *dbscan* function. Briefly, *dbscan* divides the data into equivalence classes, where two datapoints are equivalent if they are sufficiently close, and identifies equivalence classes as a cluster if they contain a point which exceeds a minimum

number of sufficiently close neighbors. The function takes three arguments:

1. The data to be clustered: We use the set of shortest path vectors from sampled points in the APS to the allodynia surface.

2. A sufficiently close distance ϵ: we use the smallest ϵ under the euclidean metric that leads to no outliers in the data.

3. The minimum number of sufficiently close neighbors to identify an equivalence class as a cluster: we take this to be 5.

2.7 Proofs

2.7.1 Using Lambert-W functions to simplify conditions on coupling strengths

To solve the optimization problems contained in the condition that specify reasonable behaviors for circuits such as the simple, dynamic and static circuits, it is necessary to find the critical points x_0, x_{-1} of

$$\frac{a\tanh(x-b)+c}{x}.$$

Such critical points are given by the 0 and -1 branch W_0 and W_{-1} of the Lambert-W function (Equation II.6):

$$x_{0,-1} = \frac{1}{2}\left(c/a - W_{0,-1}\left(-e^{1-2b}\right)\right)$$

We need to find critical points of $\frac{a\tanh(x-b)+c}{x}$. That is, we need

$$\begin{aligned} 0 &= \frac{d}{dx}\left[\frac{a\tanh\left(x-b\right)+c}{x}\right] \\ &= \frac{a\,\text{sech}^2(x-b)x - \left(a\tanh\left(x-b\right)+c\right)}{x^2}. \end{aligned}$$

Multiplying through by x^2, we thus need that

$$0 = a\,\text{sech}^2(x-b)x - a\tanh\left(x-b\right) - c.$$

and thus that

$$0 = \text{sech}^2(x-b)x - \tanh\left(x-b\right) - c/a,$$

which denoting c/a by d becomes

$$\begin{aligned} 0 = &\ \text{sech}^2(x-b)x - \tanh(x-b) - d \\ = &\ \text{sech}^2(x-b)(x-b) + b\,\text{sech}^2(x-b) - \tanh(x-b) - d \end{aligned}$$

which denoting $u = x - b$ becomes

$$0 = u\,\text{sech}^2(u) + b\,\text{sech}^2(u) - \tanh(u) - d$$

Multiplying through by $\cosh^2(x-b)$, we then have that

$$\begin{aligned} 0 &= u + b - \sin(u)\cosh(u) - d\cosh^2(u) \\ &= u + b - \sinh(2u)/2 - d(1+\cosh(2u))/2 \Leftrightarrow \\ 0 &= 2u + 2b - d(1+\cosh(2u)) - \sinh(2u) \\ &= 2u + (2b-d) - d\cosh(2u) - \sinh(2u) \\ &= 2u + (2b-d) - d\left(e^{2u}+e^{-2u}\right)/2 - \left(e^{2u}-e^{-2u}\right)/2 \Leftrightarrow \\ 0 &= 4u + (4b-2d) - d\left(e^{2u}+e^{-2u}\right) - \left(e^{2u}-e^{-2u}\right). \end{aligned}$$

Multiplying through by e^{-2u}, we obtain

$$\begin{aligned} 0 &= 4ue^{2u} + (4b-2d)e^{2u} - d\left(e^{4u}+1\right) - \left(e^{4u}-1\right) \\ &= 4ue^{2u} + (4b-2d)e^{2u} + e^{4u}(-d-1) + (1-d) \\ &= (4u+4b-2d)e^{2u} - e^{4u}(1+d) + (1-d) \Leftrightarrow \\ 0 &= \left(e^{d-2b}(2u+2b-d)\right)e^{2u+2b-d} - \left(\frac{1+d}{2}e^{2d-4b}\right)e^{4u+4b-2d} + (1-d)/2 \\ &= \frac{1-d}{2} + \left(e^{d-2b}\right)ve^{v} - \left(\frac{1+d}{2}e^{2d-4b}\right)e^{2v} \Leftrightarrow \\ 0 &= \left(\frac{1-d}{2}e^{2b-d}\right) + ve^{v} - \left(\frac{1+d}{2}e^{d-2b}\right)e^{2v} \end{aligned}$$

If $d = 1$, the preceding equation simplifies to the following:

$$\begin{aligned} 0 &= ve^v - e^{1-2b}e^{2v} \Leftrightarrow \\ 0 &= ve^{2b-1} - e^v \Leftrightarrow \\ 0 &= e^{2b-1} - \frac{1}{v}e^v \Leftrightarrow \\ e^{2b-1} &= \frac{1}{v}e^v \Leftrightarrow \\ e^{1-2b} &= ve^{-v} \Leftrightarrow \\ -e^{1-2b} &= -ve^{-v}. \end{aligned}$$

Then, as long as

$$\begin{aligned} -e^{1-2b} &\geq -e^{-1} \Leftrightarrow \\ e^{1-2b} &\leq e^{-1} \Leftrightarrow \\ 1 - 2b &\leq -1 \Leftrightarrow \\ 2 &\leq 2b \Leftrightarrow \\ 1 &\leq b, \end{aligned}$$

the preceding equation has a solution given by the principal W_0 and -1 W_{-1} branches of the Lambert-w function:

$$\begin{aligned} W_{0,-1}\left(-e^{1-2b}\right) &= -v \Leftrightarrow \\ v &= -W_{0,-1}\left(-e^{1-2b}\right) \Leftrightarrow \\ 2u + 2b - d &= -W_{0,-1}\left(-e^{1-2b}\right) \Leftrightarrow \\ 2u &= d - 2b - W_{0,-1}\left(-e^{1-2b}\right) \Leftrightarrow \\ x - b &= \frac{1}{2}\left(d - 2b - W_{0,-1}\left(-e^{1-2b}\right)\right) \Leftrightarrow \\ x &= b + \frac{1}{2}\left(d - 2b - W_{0,-1}\left(-e^{1-2b}\right)\right) \\ &= b + \frac{1}{2}\left(c/a - 2b - W_{0,-1}\left(-e^{1-2b}\right)\right) \end{aligned}$$

Otherwise, the preceding equation has no solution, and there are no local extrema of the function being minimized. However if $d \neq 1$, i.e. if $c \neq a$, then there does not appear to be a general solution to the equation in terms of known functions. The solution to Equation II.5

thus occurs at

$$x_{0,-1} = \frac{1}{2}\left(c/a - W_{0,-1}\left(-e^{1-2b}\right)\right)$$

so long as x_0 or $x_{-1} \in [f_{A\beta,min}, f_{A\beta,max}]$ or at $x = f_{A\beta,min}$ or $x = f_{A\beta,max}$

If instead $c \neq a$, I claim that $\frac{a\tanh(x-b)+c}{x}$ may have at most three critical points.

Indeed, we know from the preceding proof that the extrema occur when

$$0 = q(u) := (1-d) + (4u + 4b - 2d)e^{2u} - (1+d)e^{4u}.$$

However, it turns out that $q(u)$ can be zero at most three times. To see why, we will show that $q'(u)$ has at most two extremum on $(-\infty, \infty)$. Indeed,

$$\begin{aligned} q'(u) &= (8u + 8b - 4d + 4)e^{2u} - 4(1+d)e^{4u} \\ &= 4e^{2u}(2u + 2b - d + 1 - (1+d)e^{4u}), \end{aligned}$$

which is zero only when

$$\begin{aligned} (1+d)e^{4u} &= 2u + 2b - d + 1 \Leftrightarrow \\ e^{4u} &= \frac{2u + 2b - d + 1}{1+d}. \end{aligned}$$

Now, make the substitution

$$r := 2u + 2b - d + 1.$$

Then,

$$\begin{aligned} e^{2r-4b+2d-2} &= \frac{r}{1+d} \Leftrightarrow \\ (1+d)e^{-4b+2d-2} &= re^{-2r} \Leftrightarrow \\ -2(1+d)e^{-4b+2d-2} &= -2re^{-2r}. \end{aligned}$$

However, we can write the solution to the preceding expression using the 0 (W_0) and -1

(W_{-1}) branches of the Lambert-W function

$$-2r = W_{0,-1}\left(-2(1+d)e^{-4b+2d-2}\right) \Leftrightarrow$$
$$r = -\frac{1}{2}W_{0,-1}\left(-2(1+d)e^{-4b+2d-2}\right).$$

Rewriting r back in terms of u, we have

$$2u + 2b - d + 1 = -\frac{1}{2}W_{0,-1}\left(-2(1+d)e^{-4b+2d-2}\right) \Leftrightarrow$$
$$u_{0,-1} = \frac{1}{2}\left(-2b + d - 1\right) - \frac{1}{4}W_{0,-1}\left(-2(1+d)e^{-4b+2d-2}\right).$$

In each of $(-\infty, \min\{u_0, u_{-1}\}]$, $(\max\{u_0, u_{-1}\}, \max\{u_0, u_{-1}\}]$, $(\max\{u_0, u_{-1}\}, \infty)$, q is monotonic, and thus can have at most one zero. Thus, q has at most three zeros. Moreover to search for critical points, we only need to find zeros in those three intervals, which is a fairly easy task because q is monotonic in the intervals. Once we find the zeros of $q(u)$ in those intervals, we change $u = x - b$ back to x, and compute $\frac{a\tanh(x-b)+c}{x}$ at each of the zeros of q and at the smallest and largest possible values of x. The largest and smallest of those values of $\frac{a\tanh(x-b)+c}{x}$ are thus the maximum and minimum values of $\frac{a\tanh(x-b)+c}{x}$.

To solve the optimization problems contained in conditions that circuit parameters need to satisfy to produce the desired behaviors, it is also helpful to find the critical point x_0 of

$$\frac{x + A}{1 + \tanh(x + B)}.$$

Such a critical point is given by the principal branch W_0 of the Lambert-W function (Equation II.9):

$$x_0 = -\frac{1}{2}\left(1 + 2A + W_0\left(e^{2B-2A-1}\right)\right).$$

Indeed, to solve Equation II.8, we need to solve

$$0 = \frac{d}{dx}\left[\frac{x + A}{1 + \tanh(x + B)}\right]$$
$$= \frac{1 + \tanh(x + B) + \text{sech}^2(x + B)(x + A)}{(1 + \tanh(x + B))},$$

which, multiplying through by the denominator and taking x and B to be positive so $\tanh(x +$

$B) \neq -1$, occurs if and only if ($\Leftrightarrow$)

$$0 = 1 + \tanh(x + B) + (x + A)\text{sech}^2(x + B).$$

Multiplying through by $\cosh^2(x + B)$, we can simplify the preceding expression so it can be solved using the Lambert-W function:

$$\begin{aligned}
0 &= \cosh^2(x + B) + \sinh(x + B)\cosh(x + B) + (x + A) \\
&= \frac{1}{2}\left(\cosh(2x + 2B) + 1\right) + \frac{1}{2}\sinh(2x + 2B) + (x + A) \\
&= x + A + \frac{1}{2} + \frac{1}{2}\left(\cosh(2x + 2B) + \sinh(2x + 2B)\right) \\
&= x + A + \frac{1}{2} + \frac{1}{4}\left(2e^{2x+2B}\right) \\
&= x + A + \frac{1}{2} + \frac{1}{2}e^{2x+2B} \Leftrightarrow \\
0 &= 2x + 2A + 1 + e^{2x+2B} \Leftrightarrow \\
-1 - 2x - 2A &= e^{2x+2B} \Leftrightarrow \\
1 &= \frac{1}{-1 - 2x - 2A}e^{2x+2B} \Leftrightarrow \\
1 &= (-1 - 2x - 2A)e^{-2x-2B} \\
&= (-1 - 2x - 2A)e^{-2x-2A-1}e^{-2B+2A+1} \Leftrightarrow \\
e^{2B-2A-1} &= (-2x - 2A - 1)e^{-2x-2A-1}.
\end{aligned}$$

Noting that the left-hand side of the preceding equation is always positive, it follows that the preceding equation has its only real-valued solution given by the principal branch W_0 of the Lambert-W function

$$\begin{aligned}
-2x - 2A - 1 &= W_0\left(e^{2B-2A-1}\right) \Leftrightarrow \\
-2x &= 1 + 2A + W_0\left(e^{2B-2A-1}\right).
\end{aligned}$$

The solution to Equation II.8 is thus

$$x_0 = -\frac{1}{2}\left(1 + 2A + W_0\left(e^{2B-2A-1}\right)\right).$$

2.7.2 Proof that the sampling algorithm has time-complexity O(n^3) under ideal conditions

In this subsection, we seek to show that the time complexity of running one iteration of the sampling algorithm we have constructed is

$$\text{Time complexity of selecting a point} = \begin{cases} O(n^3) & U_j \geq L_j \text{ always for } j = 1, 2, \ldots, n \\ O\left(n\left(\frac{1}{\beta}\right)^{n-2}\right) & U_j < L_j \text{ with probability } \beta \text{ for } j = 1, 2, \ldots, n, \end{cases}$$

where n is the number of coupling strengths underlying the circuit.

In particular, below we show that selecting a hypercube in a cover of the allowable parameter space (APS) is $O(n^2)$, so long as the simple set of inequalities which we are using to define the APS is a "proper" description of the APS in the sense that the upper bounds on each coupling strength are indeed larger than the corresponding lower bounds. We then show that the time-complexity of sampling from the rectangle is $O(n)$ and then argue that the amount of times the algorithm requires us to toss out the point to ensure we are sampling uniformly from the APS is independent of n:

> Indeed, step 1 can be accomplished randomly selecting a number between 0 and 1, multiplying the number by m_1, taking the ceiling of the result, and using that as the index for a preallocated, linearly spaced partition of $[a_1, b_1]$. Hence step 1 is O(1).
>
> Likewise, partitioning $[L_j(x_1, \ldots, x_{j-1}), U_j(x_1, \ldots, x_{j-1})]$ and selecting a value of x_j from the partition is O(1). However, the number of operations needed to compute L_j and U_j likely scales with the number of coordinates that L_j and U_j depend upon, which is $j - 1$. Thus, the whole process of selecting x_j–computing L_j and U_j, partitioning $[L_j, U_j]$, randomly selecting a subinterval from the partition, and deterministically selecting an x_j from the subinterval–is $O(j)$.
>
> Thus, if $U_j \geq L_j$ always, steps (1) - (3) collectively take:
>
> $$\begin{aligned} \text{Order of steps (1) - (3)} &= O(1) + O(2) + \cdots + O(n) \\ &= O(n^2) \end{aligned}$$
>
> However if $U_n < L_n$ with some fixed probability β, then, we need to recompute $x_1, \ldots x_{n-1}$, L_n, and U_n with probability $\beta \in (0, 1)$. Thus, we would need to recompute U_n and L_n $O(\beta^{-1})$ times. However, we would need to recompute x_{n-1} an additional $O(\beta^{-1})$ times, as well. If we seem that $U_{n-1} < L_{n-1}$ with probability β as well, for each time we would need to recompute U_n and L_n, we would need to recompute U_{n-1} and L_{n-1} an additional $O(\beta^{-1})$ times, leading to a total of $O(\beta^{-2})$ times that we would need to recompute U_{n-1} and L_{n-1}. Thus, by induction, we would need to recompute U_{n-j} and L_{n-j} a total of $O(\beta^{-j})$ times. Since computing U_{n-j} and L_{n-j} once is $O(n-j)$, we would spend a total time $O((n-j)\beta^{-j})$ computing U_{n-j} and L_{n-j}. Thus, the total amount of time spent computing

the bounds is

$$\text{Order of steps (1) - (3)} = O\left(\sum_{j=0}^{n-2}(n-j)\beta^{-j}\right).$$

However, we can simplify the preceding sum by replacing β^{-1} with some constant $\kappa \in (1,\infty)$ and splitting the resulting sum into two sums:

$$\sum_{j=0}^{n-2}(n-j)\kappa^j = n\sum_{j=0}^{n-2}\kappa^j - \sum_{j=0}^{n-2}j\kappa^j.$$

The former sum in the split is a geometric series and thus sums to:

$$n\sum_{j=0}^{n-2}\kappa^j = n\frac{\kappa^{n-1}-1}{\kappa-1}$$

The latter sum in the split, on the other hand, is the derivative of a geometric series. Indeed:

$$\begin{aligned}\sum_{j=0}^{n-2}j\kappa^j &= \kappa\sum_{j=0}^{n-2}j\kappa^{j-1}\\ &= \kappa\sum_{j=0}^{n-2}\frac{d}{d\kappa}\left[\kappa^j\right]\\ &= \kappa\frac{d}{d\kappa}\left[\sum_{j=0}^{n-2}\kappa^j\right].\end{aligned}$$

Using the formula for the sum of a geometric series and differentiating we obtain:

$$\begin{aligned}\sum_{j=0}^{n-2}j\kappa^j &= \kappa\frac{d}{d\kappa}\left[\frac{\kappa^{n-1}-1}{\kappa-1}\right]\\ &= \frac{(n-2)\kappa^n-(n-1)\kappa^{n-1}+\kappa}{(\kappa-1)^2}.\end{aligned}$$

Adding the two sums together and simplifying, we obtain:

$$\begin{aligned}\sum_{j=0}^{n-2}(n-j)\kappa^j &= n\frac{\kappa^{n-1}-1}{\kappa-1} - \frac{(n-2)\kappa^n-(n-1)\kappa^{n-1}+\kappa}{(\kappa-1)^2}\\ &= n\frac{\kappa^n-\kappa^{n-1}-\kappa+1}{(\kappa-1)^2} - \frac{(n-2)\kappa^n-(n-1)\kappa^{n-1}+\kappa}{(\kappa-1)^2}\\ &= \frac{2\kappa^n+\text{ lower order terms}}{(\kappa-1)^2},\end{aligned}$$

which has complexity given by

$$\sum_{j=0}^{n-2}(n-j)\kappa^j \sim O\left(\kappa^{n-2}\right) \sim O\left(\frac{1}{\beta^{n-2}}\right).$$

However, in step (4), we need to uniformly at random selecting a point from the hyperrectangle, which requires $O(n)$ operations, thus, the time complexity of randomly sampling one point from the allowable parameter space is:

$$\text{Time complexity of selecting a point} = O(n) \cdot \text{Order of steps (1) - (3)} = \begin{cases} O(n^3) & U_j \geq L_j \text{ always for } j = 1, 2, \ldots, n \\ O\left(n\left(\frac{1}{\beta}\right)^{n-2}\right) & U_j < L_j \text{ with probability } \beta \text{ for } j = 1, 2, \ldots, n. \end{cases}$$

In step (5), discarding point from the hypercube with probability proportional to the volume of the hyperrectangle will require the algorithm to take additional iterations. In particular, it will add a number of iterations proportional to one divided by the ratio of the average rectangle volume and the max possible rectangle volume. However, we see no reason for such a ratio to be dependent on either $m_1, m_2, \ldots, m_n$ or on the number of dimensions as long as the max possible rectangle volume is chosen well. Indeed, if the APS wound up being a hyperrectangle, then each selected hyperrectangle in the cover of the APS would have the same volume, and the ratio of the seelected hyperrectangle volume to the max hyperrectangle volume would be 1. Thus, each rectangle would be kept with probability 1, so the number of times we would need to repeat steps (1) - (4) added by discarding the point with probability proportional to the volume of the hyperrectangle would be O(1).

In step (6) Discarding the point from the hyperrectangle if the point fails to fall in the APS may increase the number of iterations the algorithm takes. However, as long as the hyperrectangle cover of the APS is a good approximation of the APS, the probability that the point fails to fall in the APS is low. For instance, if the APS is itself a hyperrectangle, such a probability is 0. Thus, we expect step(6) to increase the number of iterations in a dimension-independent manner.

2.7.3 Proof that the sampling algorithm indeed produces a uniform sample

Recall that we have constructed an algorithm for randomly and uniformly-in-space sampling that allowable parameter space (APS), in Section 2.6.4.

To prove that the algorithm and implementation work we first show that in steps (1) - (3) of the implementation, there are $m_1 \Pi_{j=1}^{n}(2 + m_{i_j})$ hyperrectangles that can be selected, and that these hyperrectangles are selected uniformly at random in each iteration of the sampling algorithm.

The hyperrectangles that can be selected are the same across all iterations of the sample. These hyperrectangles in particular are specified as the Cartesian product of the n-subintervals selected in one iteration of our algorithm.

Indeed, recall that in the first step of an iteration of our algorithm, we partition $[a_1, b_1]$ into m_1 equal-sized subintervals and uniformly at random select a subinterval. However, uniformly at random selecting a subinterval amounts to uniformly at random selecting one endpoint c_1 of the form

$$c_1 = a_1 + d\frac{b_1 - a_1}{m_1} \text{ for any } d = 0 \ldots, m_1 - 1$$

from the partition. The corresponding subinterval is

$$\left[c_1, c_1 + \frac{b_1 - a_1}{m_1}\right],$$

which forms the x_1 portion of the Cartesian product that specifies the hyperrectangle. In the next iteration of the algorithm, in selecting a subinterval from $[U_2(c_1) - L_2(c_1)]$, we uniformly at random select an endpoints of the form

$$c_2 = L_2(c_1) + d\frac{U_2(c_1) - L_2(c_1)}{m_2} \text{ for any } d = -1, 0 \ldots, m_2$$

for which the corresponding subinterval is

$$\left[c_2, c_2 + \frac{U_2(c_1) - L_2(c_1)}{m_2}\right]$$

which in turn forms the x_2 portion of the Cartesian product that specifies the hyperrectangle. Repeating this process, we select a hyperrectangle of the form:

$$\left[c_1, c_1 + \frac{b_1 - a_1}{m_1}\right] \times \left[c_2, c_2 + \frac{U_2(c_1) - L_2(c_1)}{m_2}\right] \times \cdots \times \left[c_n, c_n + \frac{U_n(c_1, \ldots c_{n-1}) - L_n(c_1, \ldots, c_{n-1})}{m_n}\right]$$

for any

$$c_1 = a_1 + d\frac{b_1 - a_1}{m_1} \text{ for any } d = 0 \ldots, m_1 - 1$$
$$c_2 = L_2(c_1) + d\frac{U_2(c_1) - L_2(c_1)}{m_2} \text{ for any } d = -1, 0 \ldots, m_2$$
$$\vdots$$
$$c_n = L_n(c_1, \ldots, c_{n-1}) + d\frac{U_n(c_1, \ldots c_{n-1}) - L_n(c_1, \ldots, c_{n-1})}{m_n} \text{ for any } d = -1, 0 \ldots, m_n.$$

Note that such hyperrectangles are disjoint. Moreover, from Bayes' law, any set $\{c_1, c_2, \ldots, c_n\}$ is selected with probability

$$\begin{aligned} Pr\left(\{c_1, c_2, \ldots, c_n\}\right) &= Pr(c_1) \cdot Pr(c_2|c_1) \cdot \ldots \cdot Pr(c_n|c_1, \ldots, c_{n-1}) \\ &= \frac{1}{m_1} \cdot \frac{1}{m_2 + 2} \cdot \ldots \cdot \frac{1}{m_n + 2}, \end{aligned}$$

where the last step follows because each c_{i_j} is the left endpoint of a uniformly at random selected subinterval given values of $c_1, \ldots, c_{j-1}$. Thus, each set $\{c_1, c_2, \ldots, c_n\}$ is selected uniformly at random, and so each hyperrectangle is selected uniformly at random.

Having shown that each hyperrectangle is selected uniformly at random, we can now show that the algorithm selects points uniformly at random from the allowable parameter space. To see why, we can compute the probability distribution P of selected points by considering some point x in some hyperrectangle R_j. In particular, from the law of total probability:

$$\begin{aligned} P(x) &= Pr(R_j \text{ is selected}) \cdot Pr(x \text{ is selected } | R_j \text{ is selected}) \cdot Pr(x \text{ is kept } | x \text{ is selected and } R_j \text{ is selected} \\ &= \left(\frac{1}{m_1} \cdot \frac{1}{m_2 + 2} \cdot \ldots \cdot \frac{1}{m_n + 2}\right) \cdot \frac{1}{\text{Volume}(R_j)} \cdot \frac{\text{Volume}(R_j)}{\text{Maximum Possible Rectangle Volume}} \\ &= \left(\frac{1}{m_1} \cdot \frac{1}{m_2 + 2} \cdot \ldots \cdot \frac{1}{m_n + 2}\right) \frac{1}{\text{Maximum Possible Rectangle Volume}}, \end{aligned}$$

which is constant. Therefore, steps (1) - (4) indeed select points randomly and uniformly in space within the set of hyperrectangles approximating the allowable parameter space.

Thus, if the union of hyperrectangles from which we sample points contains the APS, we can obtain a uniform-in-space sample of the APS simply by discarding any of the points contained in steps (1) - (4) that are not actually in the allowable parameter space. We show next that under the right conditions, it is indeed the case that such a union of hyperrectangles does indeed contain the APS. Namely, if $U_2, U_3, \ldots, U_n$ and $L_2, L_3, \ldots, L_n$ change sufficiently slowly, and if $m_1, \ldots, m_n$

are sufficiently large, the set of hyperrectangles approximating the allowable parameter space contains the allowable parameter space.

Indeed, consider some point x in the APS, and denote its coordinates by $(x_1, \ldots, x_n)$. We now show that each x_1, x_2,... lie within the intervals we partitioned in steps (1) - (3) of the algorithm. Namely, because $x_1 \in [0, 1]$, there exists c_1 such that

$$x_1 \in \left[c_1, c_1 + \frac{b_1 - a_1}{m_1}\right]$$

We know that

$$L_2(x_1) \leq x_2 \leq U_2(x_1).$$

However, for x to fall in one of the desired rectangles, we want to show that x_2 falls between the upper and lower boundaries of the partition constructed in step (2) of the algorithm:

$$\left[L_2(c_1) - \frac{U_2(c_1) - L_2(c_1)}{m_2}, U_2(c_1) + \frac{U_2(c_1) - L_2(c_1)}{m_2}\right].$$

This is clearly the case if $x_1 = c_1$, since then we would have that

$$L_2(c_1) \leq x_2 \leq U_2(c_1).$$

Then assuming U_2 and L_2 change sufficiently slowly on $\left[c_1, c_1 + \frac{b_1 - a_1}{m_1}\right]$,

$$\begin{aligned} L_2(c_1) - \frac{U_2(c_1) - L_2(c_1)}{m_2} &\leq L_2(x_1) \\ U_2(x_1) &\leq U_2(c_1) + \frac{U_2(c_1) - L_2(c_1)}{m_2} \end{aligned}$$

so

$$L_2(c_1) - \frac{U_2(c_1) - L_2(c_1)}{m_2} \leq x_2 \leq U_2(c_1) + \frac{U_2(c_1) - L_2(c_1)}{m_2}$$

Likewise, if for each $j = 1, \ldots, n$ it's true that $U_{i_j}(c_1, \ldots, c_{j-1})$ and $L_{i_j}(c_1, \ldots, c_{j-1})$

vary sufficiently slowly, then

$$L_{i_j}(c_1, \ldots, c_{j-1}) - \frac{U_{i_j}(c_1, \ldots, c_{j-1}) - L_{i_j}(c_1, \ldots, c_{j-1})}{m_{i_j}} \leq x_2 \text{ and}$$
$$x_2 \leq U_j(c_{1,\ldots,c_{j-1}}) + \frac{U_{i_j}(c_1, \ldots, c_{j-1}) - L_{i_j}(c_1, \ldots, c_{j-1})}{m_{i_j}}.$$

Therefore, x lies in one of the desired hyperrectangles. Thus, if $U_1, \ldots, U_n$ and $L_1, \ldots, L_n$ do not change sufficiently slowly, then choosing $m_1, \ldots, m_n$ sufficiently large then the hypercube approximation of the allowable parameter space will be a good approximation of the actual allowable parameter space

Thus, simply discarding points not in the allowable parameter space leads to a sample of points in the allowable parameter space that is uniformly distributed in space. □

2.7.4 Proof that surface S divides parameter space into two disconnected components

In this section, we attempt to show that the graph S of any continuous function $k : \mathbb{R}^n \to \mathbb{R}$ divides $\mathbb{R}^{n+1}$ into two portions: one connected portion that lies above S,and another connected portion that lies below S. Moreover, the union of the two portions is disconnected, and all paths between the two cross S. To start, we attempt to validate my claim that the portion of $\mathbb{R}^{n+1}$ lying strictly below the graph of any continuous function is connected:

Indeed, consider arbitrary points

$$\vec{x} = (x_1, x_2, \ldots, x_n, x_h)$$
$$\vec{y} = (y_1, y_2, \ldots, y_n, y_h)$$

in the portion of R^{n+1} lying strictly below the graph S of k and think of the coordinate x_h and y_h as the heights of the respective vectors. We attempt to find a path lying strictly below the parameter space and joining $\vec{x}$ to $\vec{y}$.

To do so, we find a path contained in S that joins the points $\vec{X}$ and $\vec{Y}$ in S directly above $\vec{x}$ and $\vec{y}$, respectively, where

$$\vec{X} = (x_1, x_2, \ldots, x_n, X_h)$$
$$\vec{Y} = (y_1, y_2, \ldots, y_n, Y_h)$$

In particular, let P be a path in $\mathbb{R}^n$ joining $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ with a parametrization

p of P where

$$p(0) = (x_1, x_2, \ldots, x_n),$$
$$p(1) = (y_1, y_2, \ldots, y_n).$$

By the continuity of k, $(P, k(P))$ is a path in S joining $\vec{X}$ and $\vec{Y}$. For instance,

$$(p(0), k(p(0))) = (x_1, x_2, \ldots, x_n, X_h),$$
$$(p(1), k(p(1))) = (y_1, y_2, \ldots, y_n, Y_h)$$

Then, if we can shift the path $(P, k(P))$ vertically by a continuous function $r : P \to \mathbb{R}^{n+1}$ such that the height of $r(P)$ is strictly negative and all other components of $r(P)$ are identically zero, then $(P, k(P))$ shifted by r lies strictly below S. Moreover, if we choose r such that

$$r(x_1, \ldots, x_n, x_h) = (0, 0, \ldots, 0, x_h - X_h)$$
$$r(y_1, \ldots, y_n, y_h) = (0, 0, \ldots, 0, y_h - Y_h),$$

then $k(P)$ shifted by r joins $\vec{x}$ and $\vec{y}$. Indeed, if p is parametrization of P where $p(0) = \vec{x}$ and $p(1) = \vec{y}$,

$$(p(0) + r(p(0)), k(p(0)) + r(p(0)) = (x_1, x_2, \ldots, x_n, X_h + (x_h - Y_h)) = \vec{x}$$
$$(p(0) + r(p(0)), k(p(1)) + r(p(1) = (y_1, y_2, \ldots, y_n, Y_h + (y_h - Y_h)) = \vec{y},$$

as desired. Such an r is easy to construct. Indeed, a parametrization of $r(P)$ is given by

$$r(p(t)) = (0, \ldots, 0, (x_h - X_h)(1 - t) + (y_h - Y_h)t).$$

for any $t \in [0, 1]$.

Hence, there is a path contained in the "below space" that joins $\vec{x}$ and $\vec{y}$, hence the "below-space" of parameter space is path-connected. Likewise, the portion of parameter space lying above the graph of h, the "above-space", is also path-connected. Since both the "above-space" and "below-space" lie in $\mathbb{R}^n$, they each are connected spaces. $\square$

I further claim that the "above-space" and the "below-space" are not path connected.

To do so, we show that any path joining arbitrary points $\vec{x}$ and $\vec{y}$ be above and below S,

respectively, must intersect S. Let $\tilde{P} \in \mathbb{R}^{n+1}$ be a path parameterized as

$$(p_1(t), p_2(t), \ldots, p_n(t), p_h(t)) \qquad \text{(parameterization of } \tilde{P})$$

for $t \in [0, 1]$ that connects $\vec{x}$ and $\vec{y}$ and let $p(t) = (p_1(t), p_2(t), \ldots, p_n(t))$ be the projection of that parametrization into $\mathbb{R}^n$. Then the path $(P, k(P))$ parameterized by

$$(p_1(t), p_2(t), \ldots, p_n(t), k(p_1(t), p_2(t), \ldots, p_n(t)) \qquad \text{(parameterization of } (P, k(P)))$$

for $t \in [0, 1]$ lies in S.

Define the vertical displacement between $\tilde{P}$ and

$$(P, k(P))$$

to be $d(t)$. Note that

$$d(t) = k(p_1(t), p_2(t), \ldots, p_n(t)) - p_h(t)$$

and note that d is continuous and its magnitude is the distance between the two paths. Note further that

$$d(0) = k(p_1(0), p_2(0), \ldots, p_n(0)) - p_h(0) = X_h - x_h > 0$$
$$d(1) = k(p_1(1), p_2(1), \ldots, p_n(1)) - p_h(1) = Y_h - y_h < 0$$

because $\vec{x}$ and $\vec{y}$ lie below and above S, respectively. Since d is continuous and changes sign, it must at some point be 0. In other words, $\tilde{P}$ and $(P, k(P))$ must intersect. At such an intersection point, $\tilde{P}$ thus intersects S.

Hence, all paths joining $\vec{x}$ and $\vec{y}$ intersect S and thus the "above-space" and "below-space" must not be path-connected. □

Therefore, the graph S of any continuous function k partitions $\mathbb{R}^{n+1}$ into two connected spaces: the space above S, and the space below S. Therefore, the surface S which is the graph of h_S truly does divide parameter space into space lying above (and including) S in which the target state is attainable and a space lying strictly below S where the target space is unattainable. □

2.7.5 Proof that the primary and secondary optimization problems are equivalent

Recall that the primary optimization problem (Equation II.15) is to find the closest point $\vec{gs}$ on the attainable-state boundary S to the point $\vec{gp}$ in the allowable parameter space:

$$\vec{gs}_{nearest} = \text{argmin}_{\vec{gs} \in S} ||\vec{gp} - \vec{gs}||,$$

where we denote

$$\vec{gs}_{nearest} = (gs_{x_1 y}, \cdots, gs_{x_n y}, gs_{in,y}).$$

However, because of the difficulty in finding S, recall that we instead try to solve the secondary optimization problem (Equation II.16) is

$$(\vec{gs^*}_{nearest}, f_{in,nearest}) = \text{argmin}_{(\vec{gv}, f_{in}) \in S^*} ||\vec{gp} - \vec{gv}||,$$

where S^* is the following set of coupling strength-input signal pairs defined by removing the minimization from Equation II.14:

$$S^* := \left\{ (\vec{g}, f_{in}) : g_{in,y} = \frac{(V_{y,rest} - V_{y,thr}) + \sum_{j=1}^{n} g_{x_i y} f_{x_i}(\vec{g}, f_{in})}{f_{in}} \text{ and } f_{in} \in [f_{in,min}, f_{in,max}] \right\},$$

and where going forward we will denote the solution $\vec{gs^*}_{nearest}$ as

$$\vec{gs^*}_{nearest} = (gs^*_{x_1 y}, \cdots, gs^*_{x_n y}, gs^*_{in,y}).$$

Thus, for the preceding strategy to work, we need

$$\vec{gs^*}_{nearest} = \vec{gs}_{nearest}.$$

Indeed, it can be proven that $\vec{gs^*}_{nearest} = \vec{gs}_{nearest}$ via the steps outlined below:

1. We show that the two optimization problems are "well-behaved" in that solutions to the primary and secondary optimization problems exist and lie in S and S^* (the projection of S^* onto the space of coupling strengths), respectively.

2. We then provide a simpler way of showing that the solutions to the secondary optimization problem solve the original optimization problem. Namely, we show that $\vec{gs^*}_{nearest}$ solves the original optimization problem if and only if $\vec{gs^*}_{nearest} \in S$.

3. We establish that $\vec{gs^*}_{nearest} = \vec{gs}_{nearest}$ by showing that $\vec{gs^*}_{nearest} \in S$.

The details are as follows:

1. Lemma: $\vec{gs}_{nearest}$ exists and $\vec{gs}_{nearest} \in S$. Likewise, $\vec{gs^*}_{nearest}$ exists and $\vec{gs^*}_{nearest} \in S^*$, the projection of S^* onto the space of coupling strengths:

$$S^* := \left\{ \vec{g} : g_{in,y} = \frac{(V_{y,rest} - V_{y,thr}) + \sum_{j=1}^{n} g_{x_j y} f_{x_j}(\vec{g}, f_{in})}{f_{in}} \text{ for some } f_{in} \in [f_{in,min}, f_{in,max}] \right\}.$$

   Proof:

   (a) Since S is the graph of a continuous function on $\mathbb{R}$, it follows that S is a closed and connected n-dimensional manifold embedded in $\mathbb{R}^{n+1}$. Further, the minimization function $||gp - gs||$ is continuous and bounded away from infinity. Therfore, the map from $S \to \mathbb{R}$ given by $\vec{gs} \to ||\vec{gp} - \vec{gs}||$ is continuous on a bounded and connected set, and therefore maps to a bounded and connected set in $\mathbb{R}$ which has a maximum and attains the maximum for some $\vec{gs} \in S$. Thus $\vec{gs}_{nearest}$ exists, and $\vec{gs}_{nearest} \in S$.

   (b) Likewise, S^* is the graph of a continuous function on $\mathbb{R}^n \times [f_{in,min}, f_{in,max}]$ (where $g_{in,y}$ defines the height of the graph, much as discussed previously). Hence S^* is closed and connected, and so its projection onto the space of coupling strengths S^* is a closed $(n+1)$-dimensional manifold embedded in $\mathbb{R}^{n+1}$. Thus, following the argument given for $\vec{gs}_{nearest}$, $\vec{gs^*}_{nearest}$ exists, and $\vec{gs^*}_{nearest} \in S^*$

2. Lemma: $\vec{gs^*}_{nearest}$ exists and solves the original optimization problem Equation II.15 if and only if ($\Leftrightarrow$) $gv_{nearest} \in S$.

   (a) "$\Rightarrow$": If $\vec{gs^*}_{nearest}$ solves Equation II.15, then, $gv_{nearest} \in S$ by the preceding lemma.

   (b) "$\Leftarrow$": If $gv_{nearest} \in S$, then by construction,

$$\min_{\vec{gs} \in S} ||\vec{g} - \vec{gs}|| \leq ||\vec{g} - \vec{gs^*}_{nearest}||.$$

   However, since $S \subset S^*$ (as per the proof in (c), below),

$$\min_{(\vec{gv}, f_{in}) \in S^*} ||\vec{g} - \vec{gv}|| \leq \min_{\vec{gs} \in S} ||\vec{g} - \vec{gs}||.$$

   Thus,

$$\min_{\vec{gs} \in S} ||\vec{g} - \vec{gs}|| = ||\vec{g} - \vec{gs^*}_{nearest}||,$$

and so $\vec{gs^*}_{nearest}$ solves the original minimization problem Equation II.15. $\square$

(c) Lemma for "$\Leftarrow$": $S \subset S^*$. Proof:

$\vec{gs}$ satisfies

$$gs_{in,y} = \frac{(V_{y,rest} - V_{y,thr}) + \sum_{j=1}^{n} gs_{x_j y} f_{x_j}(\vec{g}, f_{in})}{f_{in}}$$

for some $f_{in} \in [f_{in,min}, f_{in,max}]$. Indeed, as long as $0 \notin [f_{in,min}, f_{in,max}]$, the function being maximized in Equation II.14 is continuous and thus must attain its maximum at some $f_{in} \in [f_{in,min}, f_{in,max}]$. Hence, any point $gs \in S$ must also be in S^*, and so $S \subset S^*$. $\square$

3. I claim that $\vec{gs^*}_{nearest}$ is the point $\vec{gc} \in S$ given by

$$\vec{gc} := \left(gs^*_{x_1 y}, \ldots, gs^*_{x_n y}, \min_{f_{in}} \left[\frac{(V_{y,rest} - V_{y,thr}) + \sum_{j=1}^{n} gs^*_{x_j y} f_{x_j}(\vec{gs^*}_{nearest}, f_{in})}{f_{in}} \right] \right).$$

If we can establish such a result, we will have shown that $\vec{gs^*}_{nearest} \in S$, and so $\vec{gs^*}_{nearest}$ solves the original optimization problem (Equation II.15). To establish such a result, we will use that because $\vec{gc}$ and $\vec{gs^*}_{nearest}$ are vertically displaced from one another, $\vec{gc} = \vec{gs^*}_{nearest}$ if and only if

$$||\vec{gs^*}_{nearest} - \vec{gp}|| = ||\vec{gc} - \vec{gp}||.$$

To show this, we show (a) that $||\vec{gs^*}_{nearest} - \vec{gp}|| \leq ||\vec{gc} - \vec{gp}||$. We then show (b) that $||\vec{gs^*}_{nearest} - \vec{gp}|| \geq ||\vec{gc} - \vec{gp}||$ using the result shown in (c):

(a) Indeed, by the construction of $\vec{gs^*}_{nearest}$,

$$||\vec{gs^*}_{nearest} - \vec{gp}|| \leq ||\vec{gc} - \vec{gp}||.$$

(b) To show that $||\vec{gc} - \vec{gp}|| \geq ||\vec{gc} - \vec{gp}||$ we want to determine whether the following argument holds:

$$\begin{aligned}
\|\vec{gc} - \vec{gp}\| &= \| \left(gs^*_{x_1 y}, \ldots, gs^*_{x_n y}, \min_{f_{in}} \left[\frac{(V_{y,rest} - V_{y,thr}) + \sum_{j=1}^{n} gs^*_{x_j y} f_{x_j}(\vec{gs^*}_{nearest}, f_{in})}{f_{in}} \right] \right) - \vec{gp} \| \\
&= || \left(gs^*_{x_1 y} - gp_{x_1 y}, \ldots, gs^*_{x_n y} - gp_{x_n y}, \min_{f_{in}} \left[\frac{(V_{y,rest} - V_{y,thr}) + \sum_{j=1}^{n} gs^*_{x_j y} f_{x_j}(\vec{gv}, f_{in})}{f_{in}} \right] - gp_{in,y} \right) || \\
&\leq || \left(gs^*_{x_1 y} - gs^*_{x_1 y}, \ldots, gs^*_{x_n y} - gp_{x_n y}, gs^*_{in,y} - gp_{in,y} \right) || \\
&= ||\vec{gs^*}_{nearest} - \vec{gp}||.
\end{aligned}$$

The issue is the third of the four equations listed above. It turns out that to show the third equation holds, it suffices to show that

$$gs^*_{in,y} - gp_{in,y} > \min_{f_{in}} \left[\frac{(V_{y,rest} - V_{y,thr}) + \sum_{j=1}^{n} gv_{x_j y} f_{x_j}(\vec{gs^*}_{nearest}, f_{in})}{f_{in}} \right] - gp_{in,y} > 0,$$

or equivalently, that

$$gs^*_{in,y} > \min_{f_{in}} \left[\frac{(V_{y,rest} - V_{y,thr}) + \sum_{j=1}^{n} gs^*_{x_j y} f_{x_j}(\vec{gs^*}_{nearest}, f_{in})}{f_{in}} \right] > gp_{in,y}.$$

To show that the former of such inequalities holds, note that because $\vec{gs^*}_{nearest} \in S^*$, it follows that for some $f_{in} \in [f_{in,min}, f_{in,max}]$,

$$gs^*_{in,y} = \frac{(V_{y,rest} - V_{y,thr}) + \sum_{j=1}^{n} gs^*_{x_j y} f_{x_j}(\vec{gs^*}_{nearest}, f_{in})}{f_{in}}.$$

Hence,

$$gs^*_{in,y} \geq \min_{f_{in}} \left[\frac{(V_{y,rest} - V_{y,thr}) + \sum_{j=1}^{n} gv_{x_j y} f_{x_j}(\vec{gs^*}_{nearest}, f_{in})}{f_{in}} \right].$$

We show that the latter of such inequalities holds in step (c), below, thereby completing the proof.

(c) Claim: it is always true that $\min_{f_{in}} \left[\frac{(V_{y,rest} - V_{y,thr}) + \sum_{j=1}^{n} gs^*_{x_j y} f_{x_j}(\vec{gs^*}_{nearest}, f_{in})}{f_{in}} \right] \geq gp_{in,y}$. To prove this, we need only show that the point $(gs^*_{x_1 y}, \ldots, gs^*_{x_n y}, gp_{in,y})$, which is vertically displaced from $\vec{gc}$, lies on or beneath the allodynia surface. Proof:

> Suppose not. Then, $(gs^*_{x_1 y}, \ldots, gs^*_{x_n y}, gp_{in,y})$ lies directly above the allodynia surface.
>
> However, by construction, $\vec{gp}$ lies below the allodynia surface. Thus, having shown that the space above the allodynia surface and the space below the allodynia surface are not path connected, it follows that the line segment joining $\vec{gp}$ and $(gs^*_{x_1 y}, \ldots, gs^*_{x_n y}, gp_{in,y})$ passes through the allodynia surface.
>
> In particular, there is a point on the line segment that is on the allodynia surface,

with displacement from $\vec{gp}$ given by

$$\begin{aligned}\vec{e} &:= a\left[(gs^*_{x_1y}, \ldots, gs^*_{x_ny}, gp_{in,y}) - \vec{gp}\right] \\ &= (gs^*_{x_1y} - gp_{x_1y}, \ldots, gs^*_{x_ny} - gp_{x_ny}, 0),\end{aligned}$$

for some $a \in (0, 1)$.

Consequently,

$$\begin{aligned}||\vec{e}|| &= a||(gs^*_{x_1y} - gp_{x_1y}, \ldots, gs^*_{x_ny} - gp_{x_ny}, 0)|| \\ &< ||(gs^*_{x_1y} - gp_{x_1y}, \ldots, gs^*_{x_ny} - gp_{x_ny}, 0)|| \\ &\leq ||(gs^*_{x_1y} - gp_{x_1y}, \ldots, gs^*_{x_ny} - gp_{x_ny}, gs^*_{in,y} - gp_{in,y})|| \\ &= ||\vec{gs^*} - \vec{gp}||.\end{aligned}$$

This contradicts that $\vec{gs^*}$ is the closest point in S^* to $\vec{gp}$! Therefore, it must be the case that $(gs^*_{x_1y}, \ldots, gs^*_{x_ny}, gp_{in,y})$ lies on or directly beneath the allodynia surface. $\square$

Since $\vec{gc}$ lies on the surface and is vertically displaced from $(gs^*_{x_1y}, \ldots, gs^*_{x_ny}, gp_{in,y})$, it follows that

$$gp_{in,y} \leq \hat{gs^*}_{in,y} \text{ i.e. that}$$
$$gp_{in,y} \leq \min_{f_{in}} \left[\frac{(V_{y,rest} - V_{y,thr}) + \sum_{j=1}^{n} gs^*_{x_jy} f_{x_j}(\vec{gs^*}, f_{in})}{f_{in}} \right]. \square$$

Thus,

$$||\vec{gc} - \vec{gp}|| = ||\vec{gs^*} - \vec{gp}||.$$

But $\vec{gc}$ and $\vec{gs^*}_{nearest}$ are vertically displaced from one other. Hence, $\vec{gc} = \vec{gs^*}_{nearest}$. Consequently, $\vec{gs^*}_{nearest} \in S$. So, it is indeed the case that $\vec{gs^*}_{nearest} = \vec{gs}_{nearest}$, and the two optimization problems are equivalent. $\square$

CHAPTER III

A Mean-Field Firing-Rate Model for the Suprachiasmatic Nucleus and Other Heterogeneous Networks of Neurons

3.1 Introduction

As we discuss in Chapter I, firing-rate models are successful mathematical reductions for the high-dimensional nonlinear systems representing neuronal networks and their activity. These models describe average neural firing activity levels across a network. Thus, they are particularly relevant for networks of neurons where the assumed pertinent information is carried in the neural firing rate, rather than in the timing of individual neuron spikes relative to each other. The suprachiasmatic nucleus (SCN) in the hypothalamus, the central circadian pacemaker in mammals, is an example of a rate-coding neural population. Indeed, its 24-hour variation in average neuron firing rates is assumed to be the primary driver for transmission of daily circadian timing information to downstream targets [HMB18]. There is evidence for this transmission to occur via synaptic signaling and humoral signaling [LS98].

Perhaps the most well-known class of firing-rate models are "convolutional models", which assume that a network of neurons may be reduced to several interacting populations. The firing activity of each population is modeled as a deterministic, often sigmoidal, gain function of the population's inputs. The time evolution of such firing activity is then typically modeled using differential equations whose solutions are convolution integrals over time. The most well known among such models is the Wilson-Cowan neural mass model [WC72], as discussed in Section 1.4.2. Many next-generation firing-rate models (Section 1.5) seek to improve upon these models by replacing the logistic gain functions with more accurate functions of neural responses (such as frequency-current curves) [ZVvPTH14] or of network population activity [BOF+20].

In classic convolutional firing rate models, such as the foundational firing-rate models [WC73, Ama77, DSHSZ74, JR95] discussed in Section 1.4, state variables represent averages of important quantities across the network, such as average membrane voltages and average firing rates. Thus, they do not describe how variability in those important quantities across the network are affected by

variability of network properties, as may be caused by heterogeneity in neural responses, synaptic signals or connectivity. Population density approaches account for variability by starting with a simple model of a spiking neuron, and deriving a Fokker-Plank equation for the time-evolution of a probability density function of membrane voltages (see for example [DJR+08, ET10], and the discussion in Chapter 1.4.4). However, obtaining expressions for such probability densities is often computationally intensive. Furthermore, obtaining resulting average firing rates can be complex, especially for more biophysically accurate neural responses. An alternate neural mass model formalism was proposed by Zandt et al. [ZVvPTH14] that accounts for the time-evolution of distributions of synaptic currents by assuming a normal distribution of currents and firing rates across the network. This results in integro-differential equations that model the time evolution of approximations of the mean and standard deviation of synaptic currents and firing rates across the network.

While such formalisms have expanded the accuracy and range of applications of firing-rate models, recent results have identified properties of SCN neurons and their network that are not easily accounted for by these models. For example, while SCN population firing rates display a $\sim$24 h cycle with higher rates during the day (light period) and lower rates during the night (dark period), recordings of individual SCN neurons have identified atypical firing patterns across the circadian cycle [BDFP09]. Specifically, one class of SCN neurons displays the familiar action potential firing at varying rates throughout most of the 24h day, but during mid-day hours they enter electrically excited states. These states include depolarization block (DB) or a state characterized by low-amplitude oscillatory fluctuations in membrane potential, referred to as Depolarized Low-Amplitude Membrane Oscillations (DLAMOs) [BOMB+21, BDFP09, JYLA97, PdJB+02]. In the DLAMO state, membrane voltages oscillate nearly sinusoidally with a minimum well above resting membrane potential and peak well below the maximum voltage in typical action potential firing.

Another property of SCN population activity that challenges firing-rate reductions is heterogeneity of neural responses across the network. A primary facet of such heterogeneity lies in the "molecular clocks" contained in SCN neurons. Namely, the daily variations in SCN neuron firing patterns are generated by the intracellular transcription-translation feedback loops (TTFLs) of circadian clock genes and proteins. These molecular clocks modulate the responses of SCN neurons to synaptic signalling, so that for some clock states the neuron never fires, and in others the neuron fires spontaneously or exhibits DLAMOs given strong synaptic signaling. Under normal conditions, when circadian rhythms are entrained to the 24 h environmental light cycle, cellular TTFL circadian clocks are approximately synchronized. As a result, SCN neurons may be assumed to be exhibiting oscillating firing patterns synchronized to the circadian clocks. However, this synchronization can be degraded by disruptions in light schedules as occurs in shift work and jet lag [NAN+03].

Additionally, degradation of clock cell synchrony may occur during aging [NNT+15]. This desynchrony of intracellular clocks presumably results in wide variation in neural firing patterns and responses, i.e. action potential spiking, DLAMOs or DB, across the SCN network. Physiologically, the resulting variation of the firing activity in the SCN has important consequences as it has been correlated with reduced amplitude of circadian signaling [FMD+12].

A further facet of heterogeneity in the SCN network lies in the function of GABA, the primary synaptic neurotransmitter in the SCN [MS93]. GABA plays a somewhat unusual role in the SCN [DMB+15], wherein GABA elicited post-synaptic currents (PSCs) can be either excitatory or inhibitory. This variability in GABA-mediated PSCs is due to variations in intracellular chloride concentrations which alter the synaptic current reversal potential in post-synaptic cells [DMB+15]. As a result, neuronal responses to synaptic currents can vary across the network leading to more variation in neural firing activity.

Thus, the rate code of circadian signaling by the SCN network is influenced by multiple factors not typically accounted for in firing rate models: atypical neuron firing states as well as variations in neuron firing states and synaptic currents across the network.

In this chapter, we derive a firing-rate model framework for the average network synaptic and firing activity, and their standard deviations, of a network of SCN neurons that exhibit diverse firing patterns across the circadian cycle. A strength of our approach is that it accounts for atypical neuron firing patterns, and incorporates distributions of neuron firing properties and post-synaptic currents. While we derive the model for an SCN network, the formalism can be applied generally to other neural populations with unique and variable neural firing properties.

The chapter is organized as follows. We derive and describe our firing-rate model framework in the most general context in Section 3.2. We apply our model to the SCN network in Section 3.3, and show the results of the application to the SCN in Section 3.4. Finally, we provide concluding remarks that contextualize our formalism within the firing-rate model literature and highlight predictions made by our model about the SCN in Section 3.5.

3.2 Firing-rate model derivation and description

Our firing-rate model formalism consists of a system of integro-differential equations describing the time evolution of the means and standard deviations of the firing rates and synaptic conductances across a single population of neurons, similarly to [ZVvPTH14]. In the model, the synaptic conductance statistics are the "model variables", in the sense that integrating the differential equations explicitly yields the synaptic conductance statistics as a function of time. The firing-

rate means and standard deviations, on the other hand, are computed implicitly during numerical integration of the model.

While the statistics of the firing rates and synaptic conductances are the main variables in the model, a variety of other parameters play an important role. Foremost are the parameters whose values vary across the network. One such varying parameter is the number of incoming synaptic connections to each neuron, N_{syn}. Other such parameters, which we collectively refer to with the parameter vector $\vec{P}(t)$ of dimension m, may include synaptic reversal potentials, such as the varying GABA reversal potential in the SCN, or parameters that influence the firing states of neurons, such as the time-varying phases of molecular clocks in the SCN. Notably, our model allows for time dependence in the vector of parameters $\vec{P}(t)$, so long as $\vec{P}(t)$ varies on a much slower time-scale than the time scale of synaptic integration.

To explain the derivation of the model, we first describe the post-synaptic response of a single neuron to synaptic input from multiple presynaptic cells (Figure III.1 and Section 3.2.1). We then describe the calculation of means and standard deviations of the neural responses across the network (Figure III.2 and Section 3.2.2). But first, we briefly summarize the derivation.

Consider the response of a post-synaptic cell i, at time t, that receives synaptic input from cell j. As described in Section 3.2.1, we compute the synaptic conductance $g_{ji}(t)$ which depends on two factors. One is the firing rate of presynaptic cell j just prior to time t, $f_j(s)$, and the other factor is the synaptic response function H_α which describes the time-varying post-synaptic response to a pre-synaptic spike. To account for modulation of synaptic current when a pre-synaptic cell is in an altered firing state, such as the DLAMO or DB states, we scale H_α by a synaptic gating variable $y_j(s)$ that depends on membrane voltage of pre-synaptic cell j and represents the fraction of activated post-synaptic current receptors, as is frequently used in spiking neuron networks [ET10].

From the synaptic conductances $g_{ji}(t)$ from all presynaptic neurons j to neuron i, we calculate the total synaptic conductance $G_i(s)$ received by the postsynaptic neuron i. The resulting firing frequency of postsynaptic cell i may depend on some or all of the parameters in $\vec{P}(t)$, in addition to $G_i(t)$. The firing rate $f_i(G_i(t), \vec{P}_i(t))$ is then computed from a biophysical, conductance-based single-neuron model. We additionally compute the synaptic gating variable $y_i(G_i(t), \vec{P}_i(t))$ based on the membrane voltage predicted from the biophysical SCN neuron model. These variables then feed back into the model to compute model variables at the next time-step.

To arrive at a time-evolving, mean-field model as described in Section 3.2.2, we compute the means and standard deviations of the quantities described above at each time step. In particular, we compute mean synaptic conductance $\overline{g}(t)$ and its standard deviation $\sigma_g(t)$ from mean population

firing rate $\overline{f}(t)$ and its standard deviation $\sigma_f(t)$, as well as from the synaptic response function H_α scaled by the mean synaptic gating variable $\overline{y}(t)$. From $\overline{g}(t)$ and $\sigma_g(t)$, we use network connectivity statistics to compute the mean and standard deviation of total synaptic conductance, $\overline{G}(t)$ and $\sigma_G(t)$, respectively. Assuming the form of distributions of parameters in $\vec{P}$ across the network as well as a Gaussian distribution for total synaptic conductance $G_i(t)$, we compute the means and standard deviations of firing frequency ($\overline{f}(t)$, $\sigma_f(t)$) and synaptic gating ($\overline{y}(t)$, $\sigma_y(t)$) from the frequency and voltage responses of the biophysical model for individual neurons in the network.

3.2.1 Computing neural firing-rates

To model the post-synaptic response of a neuron to a presynaptic spike, we let g_0 be the maximum synaptic conductance and assume synaptic currents are governed by an α function (see [ET10]). We define the synaptic response function $H_\alpha(t)$ as:

$$H_\alpha(t) = \begin{cases} g_0 \frac{t}{\tau} e^{1-t/\tau} & t \geq 0 \\ 0 & t < 0 \end{cases}, \tag{III.1}$$

where time constant τ determines the duration of the synaptic response. However, to allow for variations in synaptic conductance based on potentially altered presynaptic cell firing states, we scale g_0 by a synaptic gating variable. This synaptic gating variable for pre-synaptic cell j, $y_j(t)$, depends on cell j's firing rate $f_j(t)$ which, in turn, depends on the total synaptic conductance $G_j(t)$ that cell j receives as well as the parameter values for cell j, $\vec{P}_j(t)$. Thus, postsynaptic current conductance due to firing activity in presynaptic cell j is described by:

$$y_j(G_j(t), \vec{P}_j(t)) \cdot H_\alpha(t).$$

Synaptic conductance (g)

Continuing from above, for postsynaptic cell i, the time-dependent conductance of synaptic current induced by presynaptic cell j firing at time s is modeled as:

$$g_{ji}(t) = y_j(G_j(s), \vec{P}_j(s)) \cdot H_\alpha(t - s).$$

Assuming linear temporal summation of conductances due to multiple presynaptic spikes, the synaptic conductance due to all spikes from presynaptic neuron j in the small time interval $(s, s+\Delta s)$

is

$$g_{ji}(t) \approx y_j(G_j(s), \vec{P}_j(s)) \cdot H_\alpha(t-s) \cdot \text{\# spikes fired}$$
$$= y_j(G_j(s), \vec{P}_j(s)) \cdot H_\alpha(t-s) \cdot f_j(G_j(s), \vec{P}_j(s))\Delta s.$$

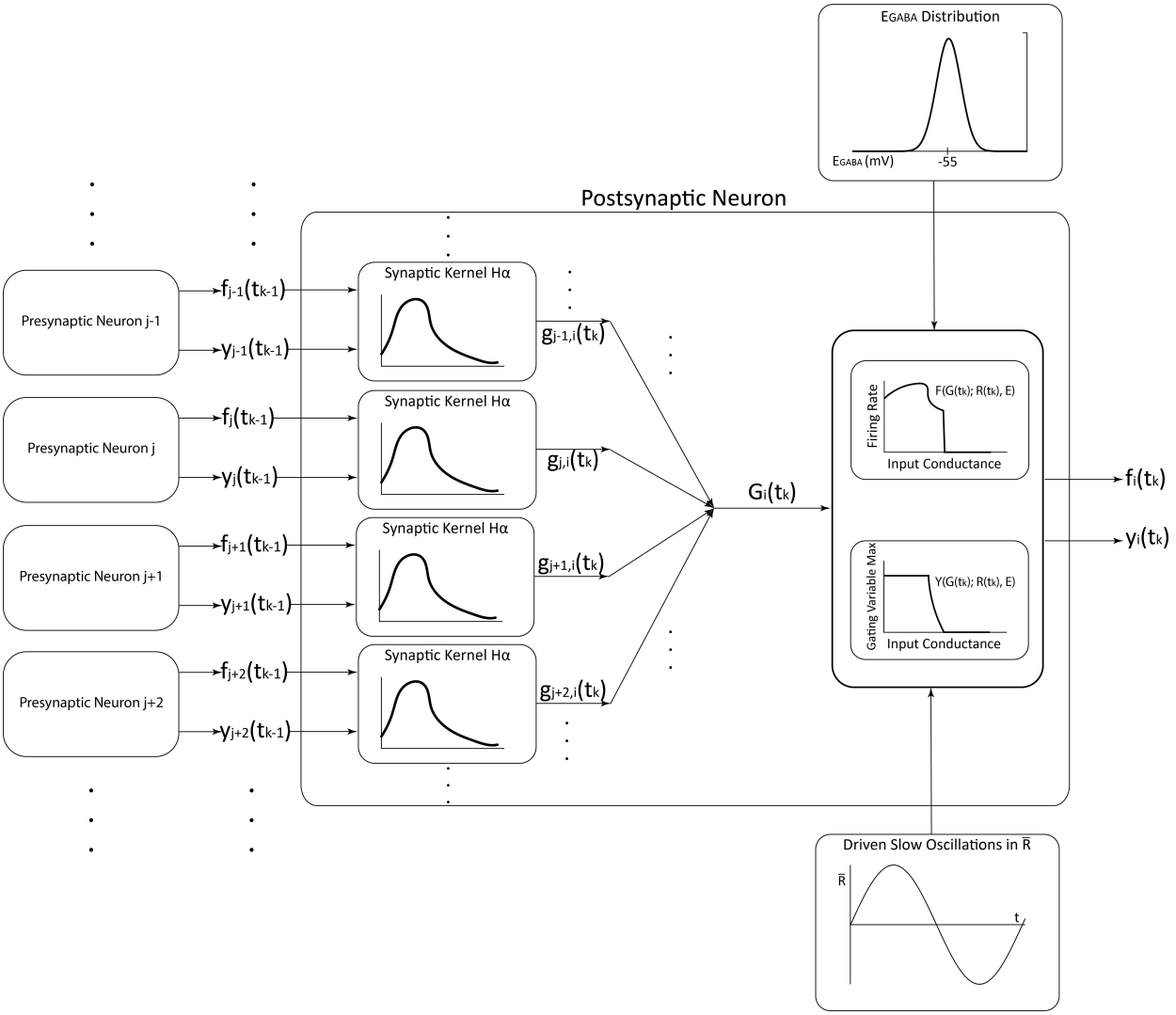


Figure III.1: Schematic of model algorithm for computing the postsynaptic neuron response. Postsynaptic neuron i (large rectangle) receives synaptic inputs from multiple presynaptic neurons j as determined by the network structure. The conductance of the postsynaptic current $g_{j,i}(t)$ induced by presynaptic cell j firing $f_j(t)$ is modeled by its convolution with the synaptic response function H_α scaled by the synaptic current gating variable $y_j(t)$. We assume linear temporal summation of synaptic conductances. Postsynaptic firing rate and maximum synaptic gating variable of neuron i at time t_k, $f_i(t_k)$ and $y_i(t_k)$, respectively, are computed from a biophysical SCN neuron model and synaptic gating functions which depend on parameters $\vec{P}(t_k)$. Illustrated here in the case of the SCN, $\vec{P}$ consists of a circadian phase proxy parameter called R (see Section 3.3) and GABA reversal potential E_{GABA}.

We now arrive at a convolution integral[1] describing the synaptic conductance response of postsynaptic cell i to all spikes generated by presynaptic neuron j before time t. Namely, dividing $[0, t]$ into n equal intervals with left endpoints $s_0, s_1, \ldots, s_{n-1}$, we obtain that

$$g_{ji}(t) \approx \sum_{k=0}^{n-1} y_j(G_j(s_k), \vec{P}_j(s_k)) \cdot H_\alpha(t - s_k) \cdot f_j(G_j(s_k), \vec{P}_j(s_k))\Delta s.$$

Letting $\Delta s \to 0$ gives the synaptic conductance $g_{ji}(t)$ in postsynaptic neuron i induced from firing of presynaptic neuron j as

$$g_{ji}(t) \approx \int_0^t H_\alpha(t - s) \cdot y_j(G_j(s), \vec{P}_j(s)) \cdot f_j(G_j(s), \vec{P}_j(s))ds. \tag{III.2}$$

$$= (H_\alpha \star y_j f_j)(t). \tag{III.3}$$

Total synaptic conductance (G) in a postsynaptic neuron

To obtain the total synaptic conductance, we treat the system of synaptic inputs to a postsynaptic neuron i as a parallel circuit. Hence, we have linear summation of conductances, and the total synaptic conductance G_i in postsynaptic neuron i is

$$G_i(t) = \sum_{j=1}^{n} g_{ji}(t), \tag{III.4}$$

where n is the total number of presynaptic cells.

Neuron firing-rate (f) and synaptic gating (y)

From the total synaptic conductance $G_i(t)$, along with the neuron-dependent parameters $\vec{P}_i(t)$, we model the neuronal firing rate of postsynaptic neuron i as a deterministic function F of ($G_i(t)$, $\vec{P}_i(t)$):

$$f_i(t) = F(G_i(t), \vec{P}_i(t)). \tag{III.5}$$

F is computed from a biophysical conductance-based model for the individual neurons in the network of the form

$$C_i \frac{dV_i}{dt} = I_{syn,i} - I_{ion,i}, \tag{III.6}$$

where $C_i, V_i, I_{syn,i}$ and $I_{ion,i}$ describe the capacitance of the cell membrane, the voltage across the cell membrane, the incoming current via synapses with other neurons, and the membrane ionic

[1]This convolution integral is similar to the convolution integrals from Section 1.4 describing population level responses to signaling.

currents, respectively, for the i^{th} neuron (see Section 3.3 for the SCN neuron model). Synaptic input is modeled by the synaptic current $I_{syn,i}$ which for postsynaptic cell i is given by

$$I_{syn,i}(t) = G_i(t) \cdot (E_{syn,i} - V_i(t)). \tag{III.7}$$

To compute the function F, the neuron model is numerically integrated across appropriate ranges of G and $\vec{P}$ values, and the value of F is set to the firing frequency of the stationary solution (obtained when numerical integration reaches a periodic orbit or steady state). We note that since firing rate is assigned based on the stationary response to $G_i(t)$, there is the underlying assumption that $G_i(t)$ varies sufficiently slowly so that any transient behavior is negligible.

The synaptic gating variable $y_i(t)$ for neuron i is defined by the deterministic function Y that depends on the same variables and parameters that affect firing rate

$$y_i(t) = Y(G_i(t), \vec{P}_i(t)). \tag{III.8}$$

To compute Y, the model equations for the synaptic gating variable $\hat{y}(t)$ are numerically integrated in conjunction with the neuron model across the ranges of G and $\vec{P}$ values (see Section 3.3 for the model used in the SCN network). The synaptic gating function Y is then set to the maximum value of $\hat{y}(t)$ obtained in the stable solution:

$$Y(G, \vec{P}) = \max_{V} \left\{ \hat{y}(V(t); G, \vec{P}) \right\}. \tag{III.9}$$

3.2.2 Computing statistics of network firing rate

To construct the model for the population firing-rate of the network, we calculate averages and standard deviations for synaptic conductance and firing rate variables across all neurons in the network, accounting for distributions in the parameters $\vec{P}(t)$. To do so, we treat these variables, such as g, G, $\vec{P}$, N_{syn} f, and yf, as random quantities drawn from (potentially time-varying) distributions. The distributions for each variable are fixed across all synapses and neurons in the network.

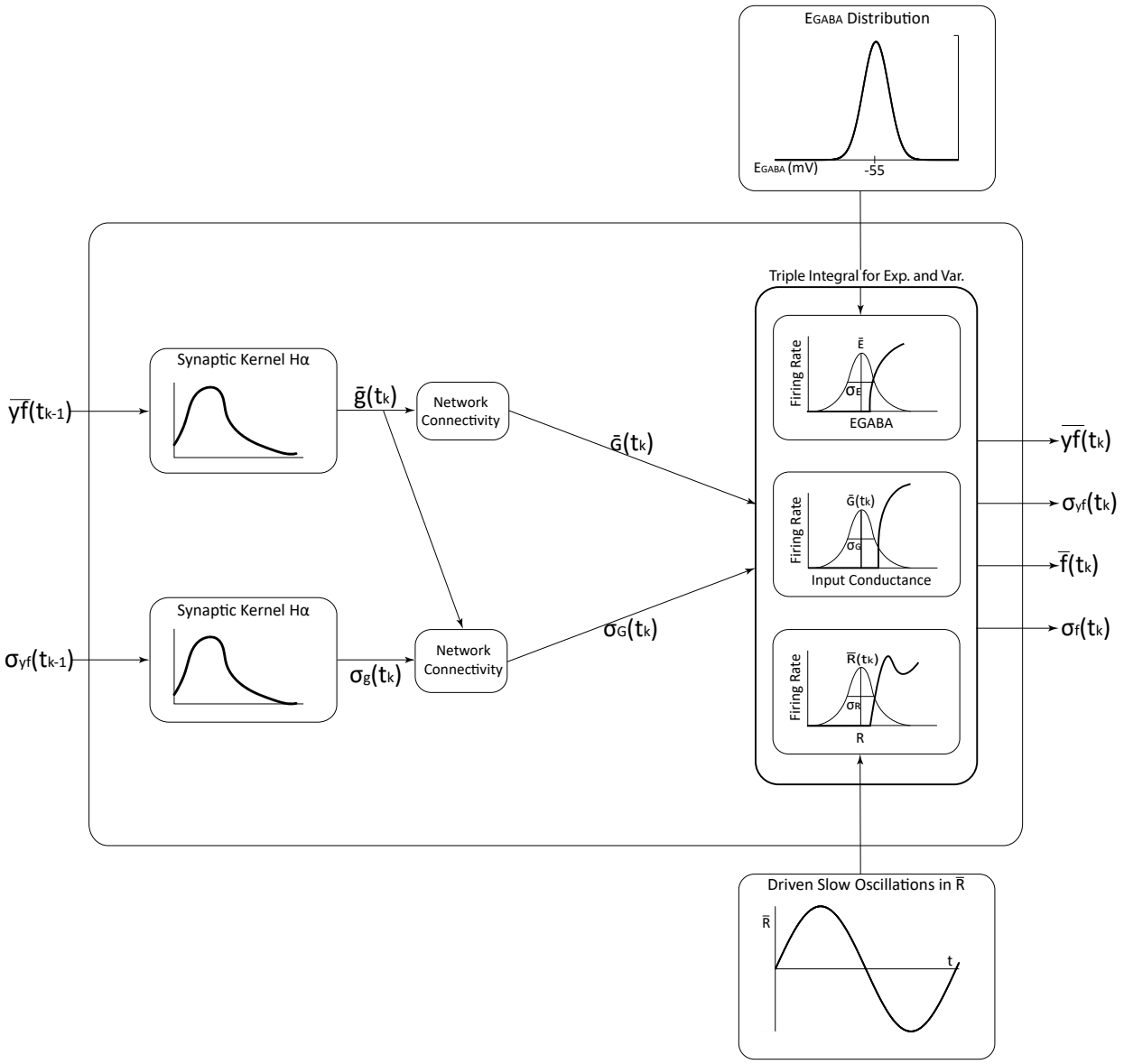


Figure III.2: Schematic of model algorithm for computing network averages and standard deviations of synaptic conductances and firing rate variables. Mean firing rate scaled by synaptic gating $\overline{yf}(t_{k-1})$ and its standard deviation $\sigma_{yf}(t_{k-1})$ at the $(k-1)^{\text{th}}$ step are convolved with the synaptic kernel H_α to yield the mean synaptic conductance $\overline{g}(t_k)$ and standard deviation $\sigma_g(t_k)$ at the k^{th} time step. Applying network connectivity statistics yields the mean total synaptic conductance $\overline{G}(t_k)$ and its standard deviation $\sigma_G(t_k)$. A uniform distribution for G along with the defined distributions for $\vec{P}$ form the joint distribution for $(G, \vec{P})$. This joint distribution is then used to calculate the expectations $\overline{f}(t_k)$, $\overline{y}(t_k)$ of the firing rate and the synaptic gating, respectively, as well as their standard deviations $\sigma_f(t_k)$ and $\sigma_y(t_k)$. To provide a concrete example, here we take $\vec{P}$ to represent the parameters constituting $\vec{P}$ in the SCN model (see Section 3.3): the circadian proxy R and the GABA reversal potential E_{GABA}.

Statistics for firing rate (f) and synaptic gating (y)

Since firing rate depends on total synaptic conductance G as well as $\vec{P}$, its mean across the SCN network and its standard deviation depend on the distributions of those variables and parameters

in the network. Viewing $G, \vec{P}$ together as a $(1+m)$-dimensional stochastic process $(G(t), \vec{P})$, denote the joint distribution at time t of $G(t)$ and $\vec{P}$ by $\gamma_t(\vec{x})$, where $\vec{x}$ represents a particular value of $(G(t), \vec{P}(t))$. Letting $U = \mathbb{R}^{1+m}$, the mean firing rate $\overline{f}(t)$ and its standard deviation $\sigma_f(t)$ are computed as:

$$\overline{f}(t) = \int_U F(\vec{x})\gamma_t(\vec{x})d\vec{x} \tag{III.10}$$

$$\sigma_f^2(t) = \int_U [F(x,y,z) - \overline{f}(t)]^2\gamma_t(\vec{x})d\vec{x}. \tag{III.11}$$

Similarly, the mean and standard deviation of the product, $yf(t)$, of firing rate $f(t)$ and synaptic gating $y(t)$ is computed as

$$\overline{yf}(t) = \int_U YF(\vec{x})\gamma_t(\vec{x})d\vec{x} \tag{III.12}$$

$$\sigma_{yf}^2(t) = \int_U [YF(\vec{x}) - \overline{yf}(t)]^2\gamma_t(\vec{x})d\vec{x}. \tag{III.13}$$

Statistics for synaptic conductance g

The mean synaptic conductance $\overline{g}$ and its standard deviation σ_g across all neurons are computed using the following equations with the derivations given below:

$$\overline{g}(t) = (H \star \overline{yf})(t) := \int_0^t H(t-s)\overline{yf}(s)ds \tag{III.14}$$

$$\sigma_g(t) = (H \star \sigma_{yf})(t) := \int_0^t H(t-s)\sigma_{yf}(s)ds, \tag{III.15}$$

where we assume each $g_{i,j}$ is drawn from the same distribution across all synapses. The $\overline{g}(t)$ equation is derived by using the Tonelli-Fubini theorem [Fol99] to pass the expectation through the integral in Equation III.2:

$$\begin{aligned}\overline{g}(t) &= \mathbb{E}\left[\int_0^t H(t-s)\cdot yf(s)ds\right] \\ &= \int_0^t H(t-s)\mathbb{E}\left[yf(s)\right]ds \\ &= \int_0^t H(t-s)\cdot\overline{yf}(s)ds.\end{aligned}$$

The derivation of the $\sigma_g(t)$ equation, on the other hand, is more complicated. Namely, it is true if the autocorrelation coefficient $\rho_{yf}(s_1, s_2) \approx 1$, as shown by [ZVvPTH14]. Indeed, following the derivation from [ZVvPTH14], rewriting the variance using the normalized autocorrelation

coefficient of $yf(t)$ at times s_1 and s_2, $\rho_{yf}(s_1, s_2)$, we have that

$$\begin{aligned}
Var\left(H \star yf\right)(t) &= \mathbb{E}\left[\int_0^t H(t-s)\cdot yf(s)ds - \int_0^t H(t-s)\cdot \overline{yf}(s)ds\right]^2 \\
&= \mathbb{E}\left[\int_0^t H(t-s)\cdot (yf(s) - \overline{yf}(s))ds\right]^2 \\
&= \mathbb{E}\left[\int_0^t H(t-s_1)(yf(s_1) - \overline{yf}(s_1))ds_1 \int_0^t H(t-s_2)(yf(s_2) - \overline{yf}(s_2))ds_2\right] \\
&= \mathbb{E}\left[\int_0^t \int_0^t H(t-s_1)H(t-s_2)\cdot (yf(s_1) - \overline{yf}(s_1))(yf(s_2) - \overline{yf}(s_2))ds_1 ds_2\right] \\
&= \int_0^t \int_0^t H(t-s_1)H(t-s_2)\cdot \mathbb{E}\left[(yf(s_1) - \overline{yf}(s_1))(yf(s_2) - \overline{yf}(s_2))\right] ds_1 ds_2 \\
&= \int_0^t \int_0^t H(t-s_1)H(t-s_2)\cdot \rho_{yf}(s_1, s_2)\sigma_{yf}(s_1)\sigma_{yf}(s_2)ds_1 ds_2.
\end{aligned}$$

The convolution integral, however, only depends on the last few seconds because for $t - s \geq 2$ seconds, $H(t-s)$ is on the order of machine precision. But, assuming circadian phase varies slowly (on the scale of hours), yf should vary slowly enough that it is approximately constant on the scale of the convolution integral. Therefore, the autocorrelation coefficient $\rho_{yf}(s_1, s_2) \approx 1$ for times s_1, s_2 within about 2 seconds of one another, and thus

$$\begin{aligned}
Var\left(H \star yf\right) &\approx \int_0^t \int_0^t H(t-s_1)H(t-s_2)\sigma_{yf}(s_1)\sigma_{yf}(s_2)ds_1 ds_2 \\
&= \int_0^t H(t-s_1)\sigma_{yf}(s_1)ds_1 \int_0^t H(t-s_2)\sigma_{yf}(s_2)ds_2 \\
&= \left(\int_0^t H(t-s)\sigma_{yf}(s)ds\right)^2,
\end{aligned}$$

which yields Equation III.15.

For computational simplicity, we rewrite the convolution integrals Equations III.14 and III.15 as differential equations:

$$\overline{g}'' = -2\tau^{-1}\overline{g}' - \tau^{-2}\overline{g} + \tau^{-1}eg_0\overline{yf}(t) \tag{III.16}$$

$$\sigma_g'' = -2\tau^{-1}\sigma_g' - \tau^{-2}\sigma_g + \tau^{-1}eg_0\sigma_{yf}(t). \tag{III.17}$$

The veracity of these differential equations is easily checked (see Appendix 2.5).

Statistics for total synaptic conductance (G)

From the statistics for synaptic conductance g, we calculate total synaptic conductance G_i of neuron i by treating it as the compound random variable $G_i = \sum_{j=1}^{N_{syn,i}} g_{j,i}$, where $N_{syn,i}$ is the number of incoming synaptic connections and $g_{j,i}$ is the conductance induced at the synapse in neuron i by the j$^{\text{th}}$ synaptic connection. Then, if we assume $N_{syn,i}$ and $g_{j,i}$ to be uncorrelated for all synapses $j = 1, ..., N_{syn,i}$ incoming into neuron i, we have

$$\overline{G}(t) = \overline{N}_{syn}\overline{g}(t), \tag{III.18}$$

where $\overline{N_{syn}}$ is the average number of incoming synapses to a neuron across the network.

To compute the standard deviation, $\sigma_G(t)$, of the total synaptic conductance, we assume that $g_{j,i}(t)$ for all $j = 1, ..., N_{syn,i}$ are pairwise conditionally uncorrelated with one another given $N_{syn,i}$. This means that for all i, j, l,

$$\mathbb{E}[g_{j,i}g_{l,i}|N_{syn,i}] = \mathbb{E}[g_{l,i}|N_{syn,i}]\mathbb{E}[g_{j,i}|N_{syn,i}]. \tag{III.19}$$

We further need to assume that $g_{j,i}(t)$ for all $j = 1, ..., N_{syn,i}$ are independent of $N_{syn,i}$. Or, more simply we could replace these two assumptions with the stronger assumption that $g_{j,i}(t)$ for all $j = 1, ..., N_{syn,i}$ and $N_{syn,i}$ form a mutually independent collection of random variables. In either case, it suffices to calculate the mean and variance of G for a single neuron, and we have from the law of total variance that

$$\begin{aligned}
Var[G] &= \mathbb{E}\left[Var(G_i|N_{syn,i})\right] + Var\left(\mathbb{E}\left[G_i|N_{syn,i}\right]\right) \\
&= \mathbb{E}\left[Var\left(\sum_{j=1}^{N_{syn_i}} g_{j,i}|N_{syn,i}\right)\right] + Var\left(\mathbb{E}\left[\sum_{j=1}^{N_{syn,i}} g_{j,i}|N_{syn,i}\right]\right) \\
&= \mathbb{E}\left[\sum_{j=1}^{N_{syn,i}} Var\left(g_{j,i}|N_{syn,i}\right) + \sum_{j=1;\, j\neq l}^{N_{syn,i}} \left(\mathbb{E}[g_{j,i}g_{l,i}|N_{syn,i}] - \mathbb{E}[g_{l,i}|N_{syn,i}]\mathbb{E}[g_{j,i}|N_{syn,i}]\right)\right] \\
&\quad + Var\left(\sum_{j=1}^{N_{syn,i}} \mathbb{E}\left[g_{j,i}|N_{syn,i}\right]\right),
\end{aligned}$$

where we use the definition of the variance to expand and simply the first term on the right-hand side. Using Equation III.19 to further simplify the first term on the right hand side above, and using the independence of $g_{j,i}(t)$ and $N_{syn,i}$ for all $j = 1, ..., N_{syn,i}$ to remove the conditionality, we have

that

$$Var[G] = \mathbb{E}\left[\sum_{j=1}^{N_{syn,i}} Var\left(g_{j,i}|N_{syn,i}\right)\right] + Var\left(\sum_{j=1}^{N_{syn,i}} \mathbb{E}\left[g_{j,i}|N_{syn,i}\right]\right)$$
$$= \mathbb{E}\left[\sum_{j=1}^{N_{syn,i}} Var\left(g_{j,i}\right)\right] + Var\left(\sum_{j=1}^{N_{syn,i}} \mathbb{E}\left[g_{j,i}\right]\right).$$

Assuming that each $N_{syn,i}$ is drawn from the same distribution, just as we have assumed for each $g_{j,i}$, we can drop the subscripts from the preceding calculation, obtaining that

$$Var[G] = \mathbb{E}\left[N_{syn}\sigma_g^2\right] + Var\left(N_{syn}\overline{g}\right)$$
$$= \sigma_g^2\mathbb{E}\left[N_{syn}\right] + \overline{g}^2 Var\left(N_{syn}\right),$$

and thus obtaining the expression for the variance of G as appears in [ZVvPTH14]:

$$\sigma_G^2(t) = Var(N_{syn})(\overline{g}(t))^2 + \overline{N_{syn}}\sigma_g^2(t). \tag{III.20}$$

While Equations III.18 and III.20 give the statistics needed for our model algorithm, the preceding discussion does not specify the shape of the distribution beyond its first and second moments. However, if $\overline{N_{syn}}$ is large and $Var(N_{syn})$ is small, then $N_{syn,i} \approx N >> 0$, some large constant for all neurons i, and we can invoke the central limit theorem to conclude that G is roughly normally distributed. If, on the other hand, such conditions fail to hold, yet the distributions of g and N_{syn} are known, one may instead obtain the shape of the distribution of G in terms of its characteristic function ϕ_G (see Appendix 2.3).

Joint distribution of $(G, \vec{P})$

We have now discussed how one may find the marginal distribution of $(G, \vec{P})$ with respect to G, but to define Equations III.10 - III.13, one must determine the full joint distribution $\gamma_t(\vec{x})$ of the stochastic process $(G(t), \vec{P}(t))$ at time t. To do so, we assume that G, and the parameters P_1, P_2, ..., P_M constituting $\vec{P}$, are independent (see further comments in the Discussion). Hence, the joint probability distribution function γ_t at time t is

$$\gamma_t(\vec{x}) = PDF_{G(t)}(x_1) \cdot PDF_{P_1(t)}(x_2) \cdot \cdots \cdot PDF_{P_m}(x_{m+1}), \tag{III.21}$$

where $PDF_{X(t)}(\cdot)$ represents the probability density function at time t of a stochastic process $X(t)$ for which the random variable X_t is defined by $X_t = X(t)$ with mean $\overline{X}(t)$ and standard deviation $\sigma_X(t)$ at time t.

A summary of our mean-field firing-rate model is given below.

Model Summary

Evolve synaptic conductance statistics $\overline{g}(t)$ and $\sigma_g(t)$ as per Equations III.16 - III.17:

$$\overline{g}'' = -2\tau^{-1}\overline{g}' - \tau^{-2}\overline{g} + e\tau^{-1}eg_0\overline{yf}(t)$$
$$\sigma_g'' = -2\tau^{-1}\sigma_g' - \tau^{-2}\sigma_g + \tau^{-1}eg_0\sigma_{yf}(t),$$

where $\overline{yf}(t)$ and $\sigma_{yf}(t)$ are given by Equations III.12 - III.13:

$$\overline{yf}(t) = \int_U YF(\vec{x})\gamma_t(\vec{x})d\vec{x}$$
$$\sigma_{yf}^2(t) = \int_U [YF(\vec{x}) - \overline{yf}(t)]^2\gamma_t(\vec{x})d\vec{x},$$

and $\vec{x}$ represents values of $(G(t), \vec{P}(t))$. Firing-rate statistics are obtained at each time by Equations III.10 - III.11

$$\overline{f}(t) = \int_U F(\vec{x})\gamma_t(\vec{x})d\vec{x}$$
$$\sigma_f(t) = \int_U [F(\vec{x}) - \overline{f}(t)]\gamma_t(\vec{x})d\vec{x}.$$

Calculation of $\overline{yf}(t)$, $\sigma_{yf}(t)$, $\overline{f}(t)$, and $\sigma_f(t)$ requires previous identification of:

1. The neuronal firing rate F and the product of F and the synaptic gating Y, YF, as functions of total synaptic conductance G and parameters $\vec{P}$.

    - Computed by numerically integrating the single neuron model and synaptic gating equation over relevant ranges of G and $\vec{P}$, and linearly interpolating or fitting a polynomial to the data as appropriate.

2. The joint probability distribution γ_t of $(G(t), \vec{P}(t))$.

    - Assume that G and the components $P_1(t), \ldots, P_m(t)$ of $\vec{P}(t)$ are independent to obtain Equation III.21:

    $$\gamma_t(\vec{x}) = PDF_{G(t)}(x_1) \cdot PDF_{P_1(t)}(x_2) \cdot \cdots \cdot PDF_{P_m}(x_{m+1}),$$

    - To calculate the statistics for G, make the simplifying assumptions (see Section 3.2.2) to obtain Equations III.18 and III.20:

    $$\overline{G}(t) = \overline{N}_{syn}\overline{g}(t)$$
    $$\sigma_G^2(t) = Var(N_{syn})(\overline{g}(t))^2 + \overline{N_{syn}}\sigma_g^2(t).$$

3.3 Applying the firing-Rate model to the SCN

We implement the mean-field firing-rate model formalism described above to simulate time-varying circadian modulation of SCN population activity. To do so, we must describe the following, as discussed below:

- The parameters $\vec{P}$ representing neural and synaptic characteristics that vary across the network (excluding parameters that specify network connectivity).
- The neuronal firing rate F and the synaptic gating Y, as functions of total synaptic conductance G and parameters $\vec{P}$.
- The joint distribution γ_t of $(G(t), \vec{P}(t))$.

Values for parameters and quantities used to apply the model to the SCN are given in Table B.1 in Appendix 2.1, and Appendix 2.2 gives details of the numerical methods used.

3.3.1 Parameters $\vec{P}$ which vary across the SCN network

As discussed in the introduction, we consider two sources of heterogeneity in the SCN network. One is heterogeneity in the GABA reversal potential of SCN neurons (E_{GABA}) which affects the response to GABAergic synaptic input. For simplicity, we assume that all synaptic inputs in the SCN are GABAergic and that E_{GABA} is constant for each neuron across the circadian cycle. The other source of heterogeneity in the SCN network is the circadian phase experienced by different neurons across the SCN. To describe this circadian phase, we use the quantity $R(t)$, which is a measure of the EBOX activity in the molecular clock of individual SCN neurons, and which varies with a 24hr period corresponding to the circadian cycle.

Hence, in applying the formalism to the SCN, we define $\vec{P}(t)$ as:

$$\vec{P}(t) = (R(t), E_{GABA}). \tag{III.22}$$

3.3.2 Neuron firing Rate (F) and synaptic gating (Y) functions for the SCN network

To compute the neuron firing rate function F, we use the conductance-based SCN neuron model of Diekman et al. [DBI$^+$13] governed by:

$$C\frac{dV}{dt} = I_{syn} - I_{Na} - I_K - I_{CaL} - I_{CaNonL} - I_{KCa}(R) - I_{K-leak}(R) - I_{Na-leak} \text{ [DBI}^+13] \tag{III.23}$$

This Hodgkin-Huxley-type model contains spike-generating Na^+ and K^+ currents, two inward Ca^{2+} currents, a Ca^{2+}-dependent K^+ current and Na^+- and K^+-mediated leak currents. This model replicates the firing responses of SCN neurons across the circadian cycle by accounting for the effects of circadian clock genes and proteins on membrane K^+ currents. Specifically, the circadian

proxy parameter $R(t)$ modulates the conductance of the Ca^{2+}-dependent K^+ current I_{KCa} and the K^+-mediated leak current I_{K-leak}. Low values of R correspond to circadian phases in the night and result in no firing or low firing rates in response to low values of input, while higher R values correspond to circadian phases in the middle of the day and can result in DLAMO firing and DB states and generally lead to higher signalling strength (Figure III.3).

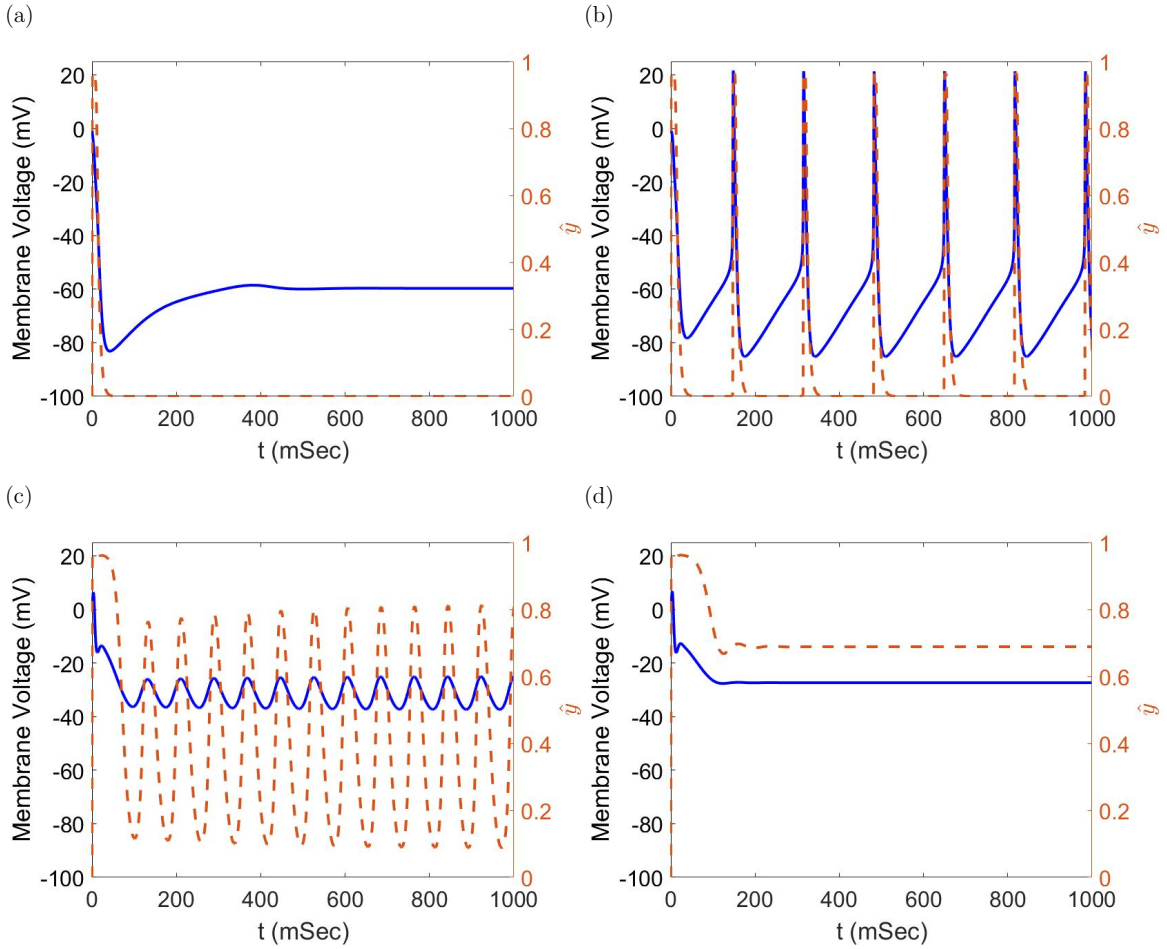


Figure III.3: Diverse firing and synaptic gating behavior at different circadian phases in the biophysical SCN neuron model [DBI+13]. Voltage $V(t)$ traces (left axes, solid blue lines) and synaptic gating $\hat{y}(t)$ as a function of $V(t)$ (right axes, dashed red lines) for values of the circadian proxy R and external applied current I_{app} exhibiting four firing regimes. **(a)**: A neuron at rest, with circadian proxy $R = -5$ and applied current $I_{app} = 2\ \mu A/cm^2$. Such behavior fails to induce postsynaptic current, and is expected of SCN neurons at night [BDFP09]. **(b)**: A neuron with $R = -5$ and $I_{app} = 8\ \mu A/cm^2$, exhibiting typical action potential (AP) firing. Such behavior results in large synaptic gating during the AP, and is expected in circadian morning and circadian evening [BDFP09]. **(c)** A neuron with $R = 4$ and $I_{app} = 8\ \mu A/cm^2$, exhibiting depolarized low-amplitude membrane oscillations (DLAMOs), where the membrane voltage oscillates roughly sinusoidally at depolarized voltages. Such behavior decreases synaptic gating and is observed in some neurons around circadian mid-day. **(d)** A neuron exhibiting depolarization block, with $R = 5$ and $I_{app} = 8\ \mu A/cm^2$, observed in some SCN neurons in circadian afternoon [BDFP09], where membrane voltage is constant and elevated due to too much excitation, often resulting in non-zero synaptic gating.

GABAergic synaptic current is modeled as:

$$I_{syn}(t) = G(t) \cdot (E_{GABA} - V(t)). \tag{III.24}$$

Synaptic gating $\hat{y}$ is modeled as in [DMB$^+$15, ET10]:

$$\frac{d\hat{y}}{dt} = a_r \cdot \frac{1}{(1 + exp(-(V(t) - (-20))/3))} \cdot (1 - \hat{y}) - a_d \cdot \hat{y}, \tag{III.25}$$

where a_r and a_d characterize the rise and decay rates, respectively, of synaptic gating in response to presynaptic neuron voltage $V(t)$. See Appendix 2.4 for full equations and parameter values.

To explicitly obtain the functions F and Y (defined by Equation III.9), values of $F(G, R, E_{GABA})$ and $Y(G, R, E_{GABA})$ were computed for each combination of G, R, and E_{GABA} on the 201-by-201-by-201 rectangular grid for $G \in [0, 1]$, $R \in [-8.5, 8.5]$, and $E_{GABA} \in [-110, 0]$. We did so by numerically integrating Eqs III.23 and III.25, holding G, R and E_{GABA} constant until the numerical solution settled down to a stationary state. Level surfaces of the resulting data $F(G, R, E_{GABA})$, $Y(G, R, E_{GABA})$ and the product $YF(G, R, E_{GABA})$ for different values of E_{GABA} are shown in Figure III.4. In this figure, boundaries between the different firing regimes–action potential firing, DLAMO, DB, and resting–are indicated.

Generally, higher values of E_{GABA} have a depolarizing effect on neural activity. For instance, when $E_{GABA} = -110$ mV for all neurons in the simulation (leftmost column), synaptic currents are exclusively inhibitory, causing neurons to exit excited states such as DB, DLAMOs, and AP firing as G increases, and significantly reducing the size of (G, R) phase space in which such excited states occur. When $E_{GABA} = -55$ mV (middle column), synaptic currents are primarily inhibitory, so that neural activity is in a resting state for high G values, but the region of (G, R) phase space corresponding to each firing regime is significantly larger. On the other hand, when $E_{GABA} = 0$ mV (rightmost column), synaptic currents are primarily excitatory, leading to AP firing for low values of R, and to DB for sufficiently large total synaptic conductances G at all values of R. Hence, for lower values of E_{GABA}, increasing G generally increases inhibition due to synaptic currents, pushing neurons from the more excited states, DB, DLAMOs or AP firing, to the least excited state, resting. Increasing G at higher values of E_{GABA}, on the other hand, increases excitation due to synaptic currents, having the opposite effect on neural activity.

3.3.3 Joint distribution for $G(t)$, $R(t)$, and E_{GABA} for the SCN network

To apply the model formalism to the SCN network, we write the joint probability distribution function γ_t at time t given by Equation III.21 using the parameters composing $\vec{P}(t)$:

$$\gamma_t(x, y, z) = PDF_{G(t)}(x) \cdot PDF_{R(t)}(y) \cdot PDF_{E_{GABA}}(z).$$

For the E_{GABA} distribution, we follow [DMB$^+$15] and assume a normal distribution across the SCN, with a mean of -55 mV and standard deviation of 7 mV.

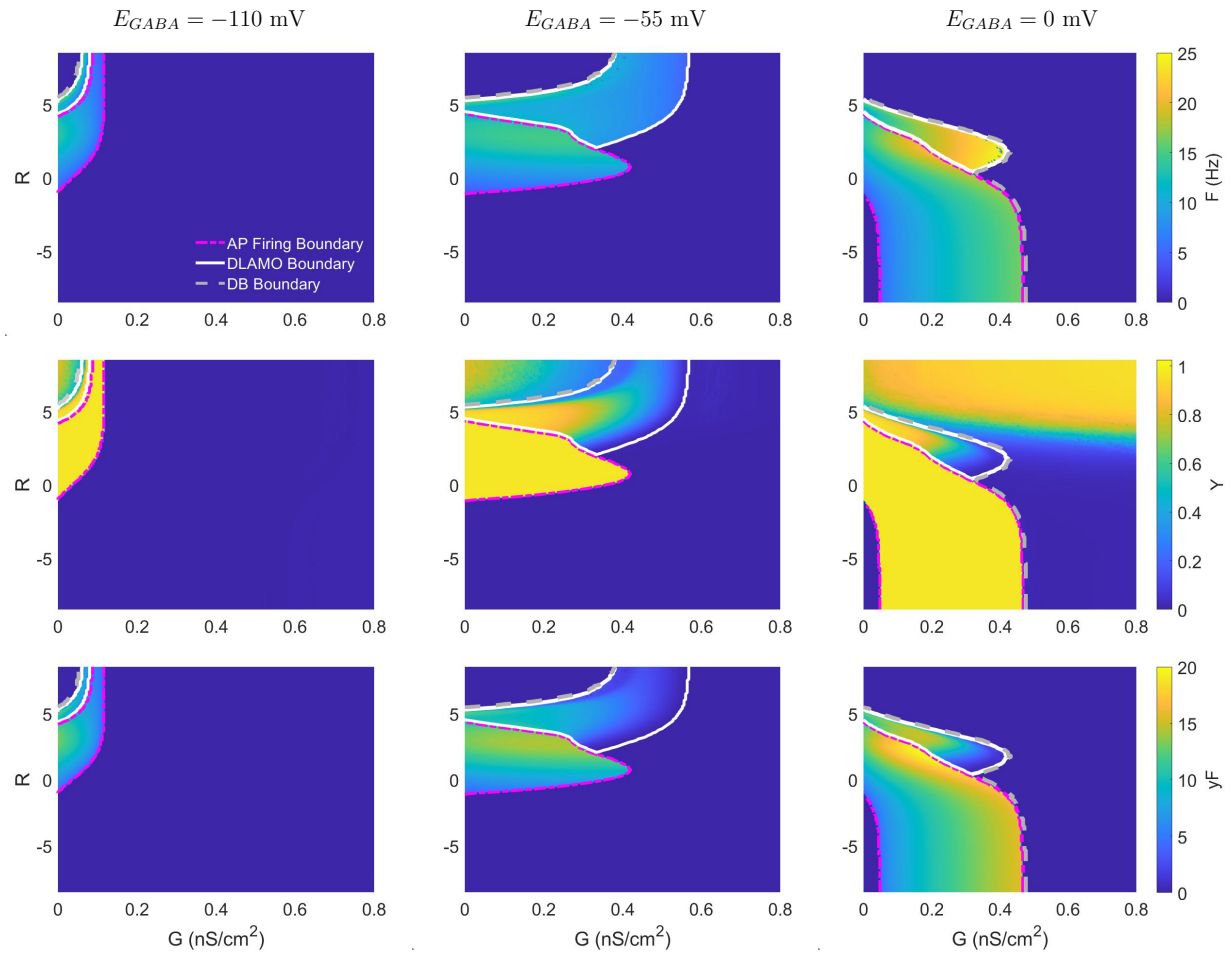


Figure III.4: Level surfaces (for fixed E_{GABA}) of firing rate function $F(G, R, E_{GABA})$ (top row), synaptic gating function $Y(G, R, E_{GABA})$ (middle row), and their product $YF(G, R, E_{GABA})$ (bottom row) for $E_{GABA} = -110$ (1st column), -55 (2nd column), and 0 mV (3rd column). Regions of each panel between magenta lines correspond to action potential firing activity, whereas regions between white lines correspond to DLAMOs, regions between gray lines and the edges of the panel correspond to DB, and the remaining regions correspond to resting. Synaptic gating Y varies in these regions (row 2): in action potential firing, $Y \sim 0.965$; in DLAMOs, $Y \sim 0.965$ near the boundary between action potential firing and DLAMOs, but rapidly decays away from the boundary; in DB, the firing rate is zero but synaptic gating ranges from about 0 to about 0.9. When the neuron is at rest, both F and Y are zero.

For the time-varying circadian proxy parameter $R(t)$, we let R vary periodically in the range $[-5, 5]$ over a circadian period, as suggested in [DMB$^+$15]. To model this simply, we assume that for any neuron, j, its R value, $R_j(t)$, varies sinusoidally, driven by circadian phase $\theta_j(t)$:

$$R_j(t) = R_{ampl} \sin(\theta_j(t)). \tag{III.26}$$

We take each $\theta_j(t)$ to follow a Gaussian distribution with mean $\overline{\theta(t)}$ varying linearly over a circadian period from 0 to 2π. In this way, $\theta = 3\pi/2$ corresponds to the trough in circadian activity, for example. We further assume that the standard deviation σ_θ of the Gaussian distribution is constant with respect to time. The distribution of $R_j(t)$ is then inherited from the distribution of $\theta_j(t)$ according to Equation III.26. For computational expediency, we consider a "circadian" period

of about 43 s, as this is the minimal period needed to eliminate hysteresis in model variables across the cycle.

Finally, for the G distribution, we make the assumptions under which Equation III.20 holds and G is normally distributed with mean and standard deviation given by Equations III.18 and III.20, respectively. To use such equations, though, we must calculate the mean $\overline{N_{syn}}$ and variance $Var(N_{syn})$ of the number of presynaptic neurons impinging on each neuron. To carry out the calculation, we use that the SCN has roughly 20,000 neurons [AMGF$^+$16], $N = 10,000$ per hemisphere. We also use that there are about $3 - 12 \cdot 10^6$ [Gül76] synaptic connections per SCN half, with [MB89] putting the number of synapses closer to the higher end of that range at $11 \cdot 10^6$. Further, since each synaptic connection represents 1 outgoing connection (from the presynaptic neuron) and 1 incoming connection (to the postsynaptic neuron), there are 11,000,000 incoming synaptic connections, and the average number of incoming synapses for SCN neurons should be

$$\overline{N_{syn}} = \frac{\text{\# Incoming Synaptic Connections}}{\text{\# Neurons}} = 1100.$$

We recognize that some recent studies such as [SMW$^+$17] suggest that $\overline{N_{syn}}$ should be closer to 10. Nevertheless, we have opted to use $\overline{N_{syn}} = 1100$ until a consensus in the SCN community has been reached, noting that a reduction in N_{syn} leads to similar results after compensating by making a proportional increase in coupling strength g_0.

In any case, to use Equations III.20 for the variance of G, observe that if the SCN were to be modeled as a Poisson random graph with probability that any two neurons share a synapse being $p = \overline{N_{syn}}/N = 0.11$, then via basic properties of expectations and variances:

$$Var(N_{syn}) = \overline{N_{syn}} \cdot (p - p^2) = 979.$$

On the other hand, if for instance the number of incoming synapses to SCN neurons follows a pure exponential distribution $p_k = (1 - r)^k r$ where $1 > r > 0$ is an unspecified constant, then

$$\begin{aligned} \overline{N_{syn}} &= 1/r \\ Var(N_{syn}) &= (1 - r)/r^2 = \overline{N_{syn}}^2(1 - 1/\overline{N_{syn}}) = \overline{N_{syn}}^2 - \overline{N_{syn}}, \end{aligned}$$

and hence if $\overline{N_{syn}} = 1100$,

$$Var(N_{syn}) \approx 1100^2.$$

$Var(N_{syn})$ is also $\approx 1100^2$ if the underlying distribution is log-normal, a common assumption for cortical brain areas [BM14]. We explore the effects of changing $Var(N_{syn})$ in Section 3.4.2.

3.4 Results

Here we present numerical results for our mean-field firing-rate model applied to the SCN network, considering variation of different neural and network parameters, including maximum synaptic conductance, mean and variance of synaptic reversal potentials, variance of circadian phases across the population, and variance of the number of incoming synapses per cell. We consider different amplitudes of the circadian proxy parameter $R(t)$ to illustrate different ranges of neural activity across the circadian cycle caused by the diverse firing states of the SCN neuron model. We summarize the effects of changing these parameters on the SCN firing rate statistics in Table III.1, and validate the model results against a spiking neuronal network in Appendix 2.6.

3.4.1 Effect of changing g_0

To investigate effects of changing maximum synaptic conductance g_0, we consider four different values of g_0 (in $\frac{mS}{cm^2}$): 10^{-6} (Very weak), 10^{-5} (weak), 10^{-4} (middling strength), and $5 \cdot 10^{-4}$ (strong). To isolate the effects of g_0, we take $E_{GABA} = -55$mV for all neurons in the network ($\sigma_{EGABA} = 0$) and let all neurons have the same circadian phase ($\sigma_\theta = 0$) with R values that vary periodically with different amplitudes ($R_{ampl} = 5, 3$ and 1). We illustrate our findings in Figure III.5.

As expected, increasing g_0 from weak coupling to stronger coupling generally leads to higher mean firing rates as well as higher standard deviations in firing rates. However, for large R_{ampl} (left column) and particularly for stronger coupling, firing rates dip at peak $\overline{R}(t)$, the assumed circadian mid-day, due to neuronal activity entering the DLAMO region, thus leading to a decrease in synaptic signaling. For smaller R_{ampl} (middle column), firing rates plateau at "mid-day" but still dip for sufficiently large g_0, whereas for still smaller R_{ampl} values (right column), firing rates vary nearly sinusoidally, with firing rates being significantly lower for strong coupling than weaker coupling.

To clarify the process by which g_0 influences firing rates, the trajectories of mean network conductance $\overline{G}(t)$ across the $R(t)$ cycle are plotted on the neuronal firing rate surface $F(G, R, E_{GABA})$ (curves in panels (g) - (i)). From such plots, it is clear that maximum synaptic conductance g_0 greatly affects which neuronal firing regions are sampled as $R(t)$ varies. These plots also clarify the observed behavior of standard deviations in the firing rates. Notably, comparing panels (g) - (i) to the standard deviations in panels (d) - (f), there is non-zero standard deviation when the total synaptic conductances G are well above zero (see black and green curves in panels (g) - (i)). Moreover, peak standard deviations in panels (d) and (e) for stronger g_0 occur when $G(t)$ nears the boundary of firing regimes. For instance, in panels (g) and (h), firing rate standard deviations are large near transitions from zero to non-zero firing rates, and remain elevated until $G(t)$ is firmly in the DLAMO regime. Further, for lowest R_{ampl} (panel (i)), the $(\overline{G}(t), R(t))$ trajectory remains near

the threshold for action potential firing throughout the circadian cycle for strong coupling, resulting in non-zero standard deviations. These higher standard deviations arise because when we integrate over the G distribution to calculate the standard deviations of the firing rates, we integrate over the transition from action potential firing to rest.

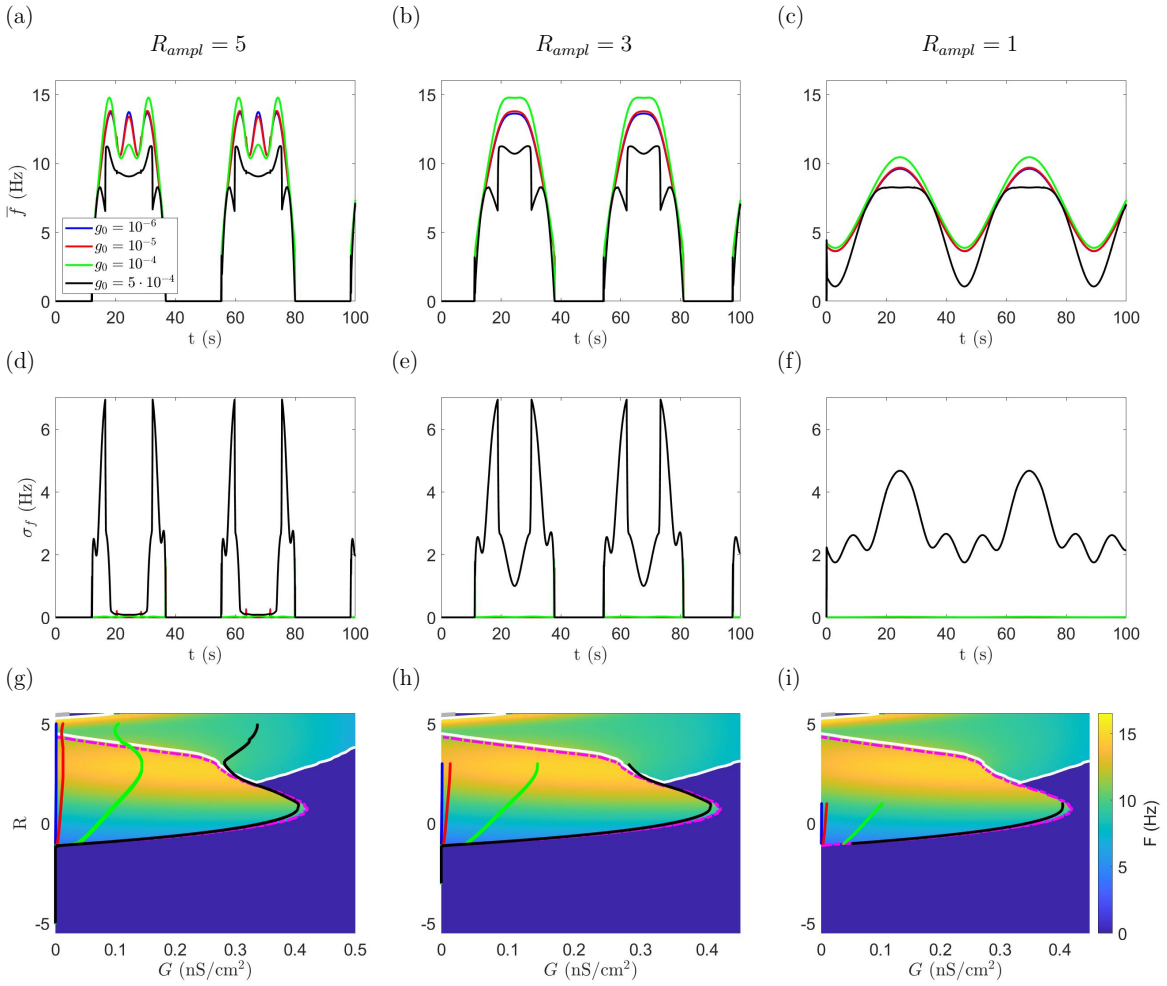


Figure III.5: Effect of maximum synaptic conductance g_0 on network firing rate and standard deviation. **(a) - (f)**: Mean firing rates $f(t)$ (first row) and standard deviations of firing rates $\sigma_f(t)$ (2nd row) over multiple $R(t)$ cycles (circadian period shortened to $\approx$ 43s) for 4 different values of maximum synaptic conductance g_0 (in $\frac{mS}{cm^2}$) ranging from strong coupling ($g_0 = 5 \cdot 10^{-4}$, black curves) to weak coupling ($g_0 = 10^{-6}$, blue curves). Three amplitudes of circadian variation are simulated, $R_{ampl} = 5$ (left column), 3 (middle column) and 1 (right column). **(g) - (i)**: $(\overline{G(t)}, R(t))$ trajectories are plotted (solid lines) on the neuronal firing rate surface $F(G, R, -55)$ for the same values of g_0 as in (a) - (f). Dashed lines represent boundaries between firing regimes, as in Figure III.4. In all panels, $E_{GABA} = -55$ mV and there is zero variance in circadian phase across the network.

3.4.2 Effect of N_{syn} distribution

Assuming a different network structure in the SCN without changing the number of neurons or synapses changes $Var(N_{syn})$, the variance in the number of incoming synapses per SCN neuron. In Figure III.6, we consider 4 different values for $Var(N_{syn})$ representing different network con-

nectivity structures but assuming $N = 10,000$ neurons in the network and $\overline{N_{syn}} = 1100$ incoming synapses per neuron on average. Specifically we consider $Var(N_{syn}) = 0$ representing a network where all neurons receive the same number of incoming synapses, $Var(N_{syn}) = 979$ corresponding to a Poisson distribution in N_{syn} characteristic of Poisson random graphs, $Var(N_{syn}) = 1.21 \cdot 10^6$ corresponding roughly to exponential or log-normal distributions for N_{syn}, and $Var(N_{syn}) = 9.79 \cdot 10^6$ corresponding roughly to the maximum possible variance according to the inequality $Var(N_{syn}) \leq (N_{syn,max} - \overline{N_{syn}})(\overline{N_{syn}} - N_{syn,min})$ where $N_{syn,min}$ and $N_{syn,max}$, the smallest and largest possible values of N_{syn}, are 0 and $10,000$, respectively. As seen in the figure, increasing $Var(N_{syn})$ slightly reduces firing rate amplitudes but can increase firing rate standard deviations considerably. For example, when $R_{ampl} = 5$, increasing $Var(N_{syn})$ from $1.21 \cdot 10^6$ to $9.76 \cdot 10^6$ more than doubles firing rate standard deviations.

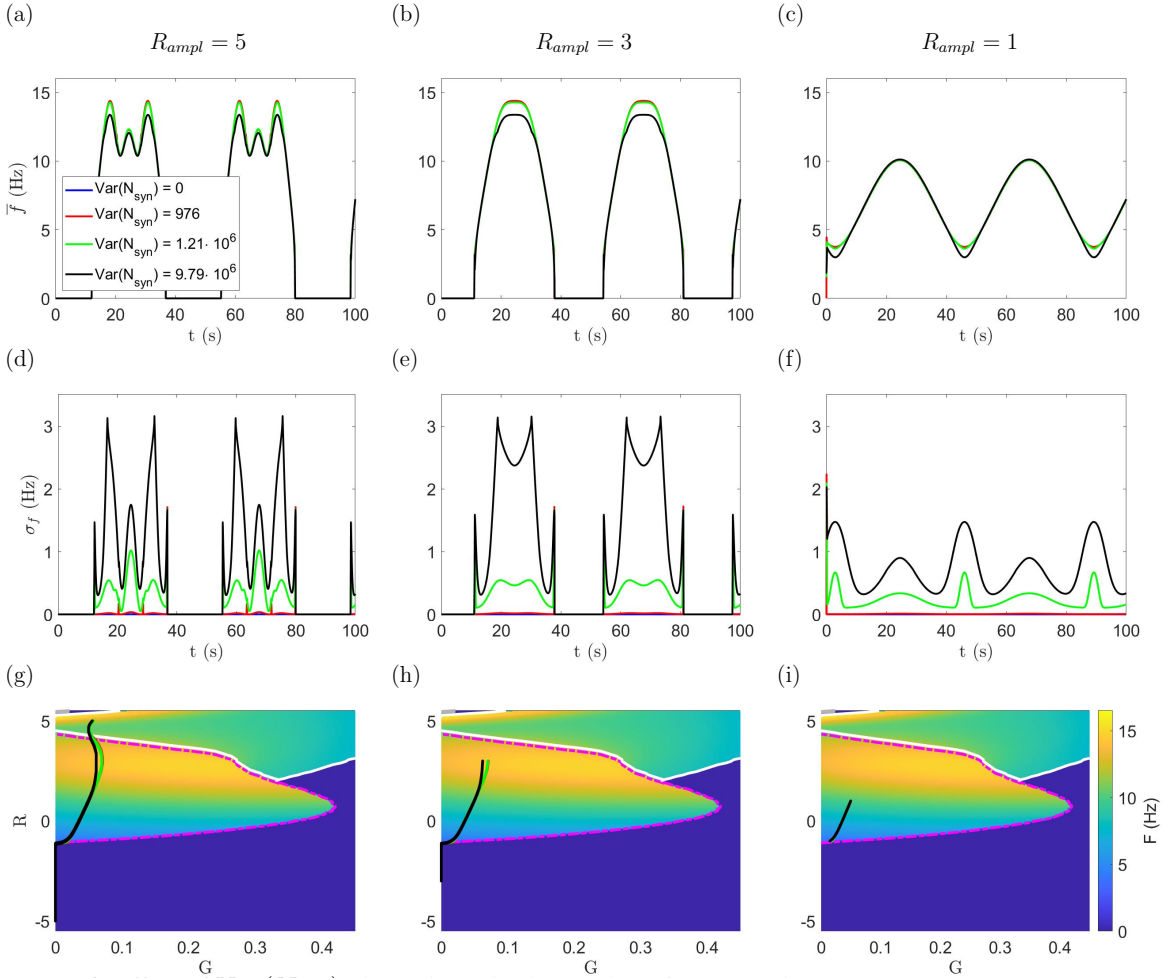


Figure III.6: Effect of $Var(N_{syn})$, the variance in the number of presynaptic neurons per SCN neuron, on network firing rate and standard deviation given weak coupling ($g_0 = 5 \cdot 10^{-5}$). **(a) - (f)**: In the top row and middle row, firing rates and firing rate standard deviations, respectively, are graphed for each of four values of $Var(N_{syn})$. **(g) - (i)**: $(\overline{G(t)}, R(t))$ trajectories are plotted (solid lines) on the neuronal firing rate surface $F(G, R, -55)$ for the same values of $Var(N_{syn})$ as in (a) - (f). Dashed lines represent boundaries between firing regimes, as in Figure III.4. The left, middle, and right columns correspond to values of 5, 3, and 1, respectively, for R_{ampl}, simulated over two shortened circadian cycles. In all panels, $E_{GABA} = -55$ mV and there is zero variance in both circadian phase and GABA reversal potential across the network.

3.4.3 Effect of circadian phase distribution

Introducing variance in the circadian phases of neurons in the SCN network results in a distribution of R values. In Figure III.7, circadian phase standard deviation σ_θ values of 0 (no variation), 0.3 (middling), 0.6 (somewhat large), and 0.9 (very large) radians were considered. Increasing σ_θ strongly reduces the amplitude of firing rate oscillations. Even for the largest R_{ampl}, increased σ_θ removes the "mid-day" dip because the spread in $R(t)$ values attenuates the influence of DLAMO and DB behaviors on overall firing rate. Additionally, increasing circadian phase standard deviation keeps firing rates above zero throughout the $R(t)$ cycle. Standard deviations in firing rate are higher with increased σ_θ and remain elevated throughout the circadian cycle for largest σ_θ.

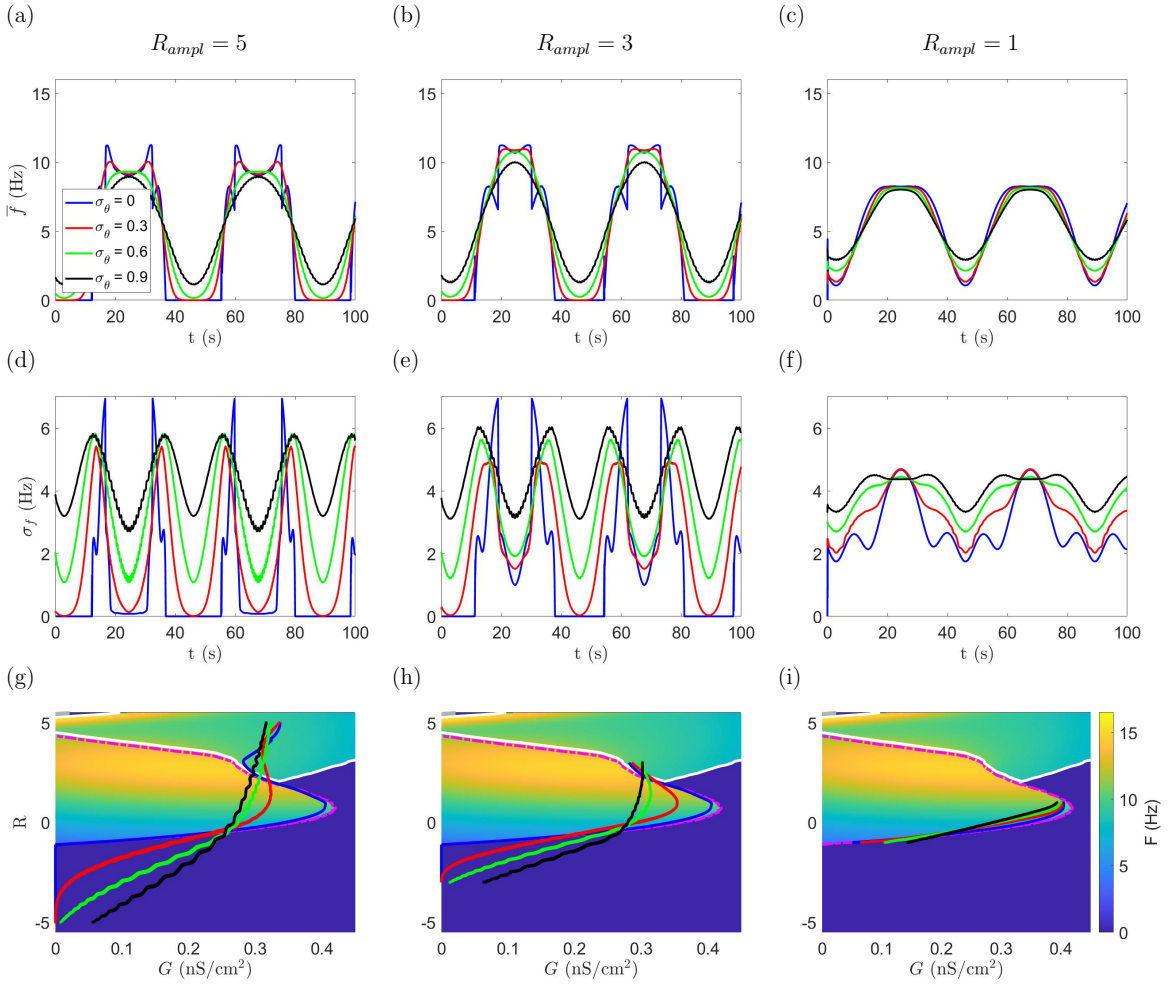


Figure III.7: *Effect of introducing variance in circadian phase (given $g_0 = 5 \cdot 10^{-5} \frac{m}{cm^2}$, $E_{GABA} = -55mV$, and $\sigma_{EGABA} = 0mV$). **(a) - (f)**: In the top row and middle row, firing rates and firing rate standard deviations, respectively, are graphed for each of four values of σ_θ. **(g) - (i)**: $(\overline{G(t)}, R(t))$ trajectories are plotted (solid lines) on the neuronal firing rate surface $F(G, R, -55)$ for the same values of σ_θ as in (a) - (f). Dashed lines represent boundaries between firing regimes, as in Figure III.4. The left, middle, and right columns correspond to R_{ampl} values of 5, 3, and 1, respectively, simulated over two shortened circadian cycles.*

3.4.4 Effect of E_{GABA} distribution

From our simulations, the standard deviation in E_{GABA}, σ_{EGABA}, has non-negligible effects on firing rate statistics mainly for large maximal synaptic conductances ($g_0 \geq 10^{-4}$ nS/cm^2). However, even with strong synaptic coupling, introducing a distribution in E_{GABA} only slightly dampens firing rate oscillation amplitude (Figure III.8). Indeed, as displayed in panels (a) - (c), firing rates corresponding to higher values of σ_{EGABA} have slightly lower peaks mid-day and higher troughs in the evening (for lowest R_{ampl}, right column). More visible are the changes in firing rate standard deviation, where increasing σ_{EGABA} significantly increases firing rate standard deviations especially during mid-day of the circadian cycle (panels (d) - (f) of Figure III.8). Further simulations (not shown) attempting to isolate the effect of σ_{EGABA} on firing rate statistics suggest that, particularly for weaker coupling strengths, increasing σ_{EGABA} primarily plays a role in increasing firing rate standard deviations around circadian mid-day.

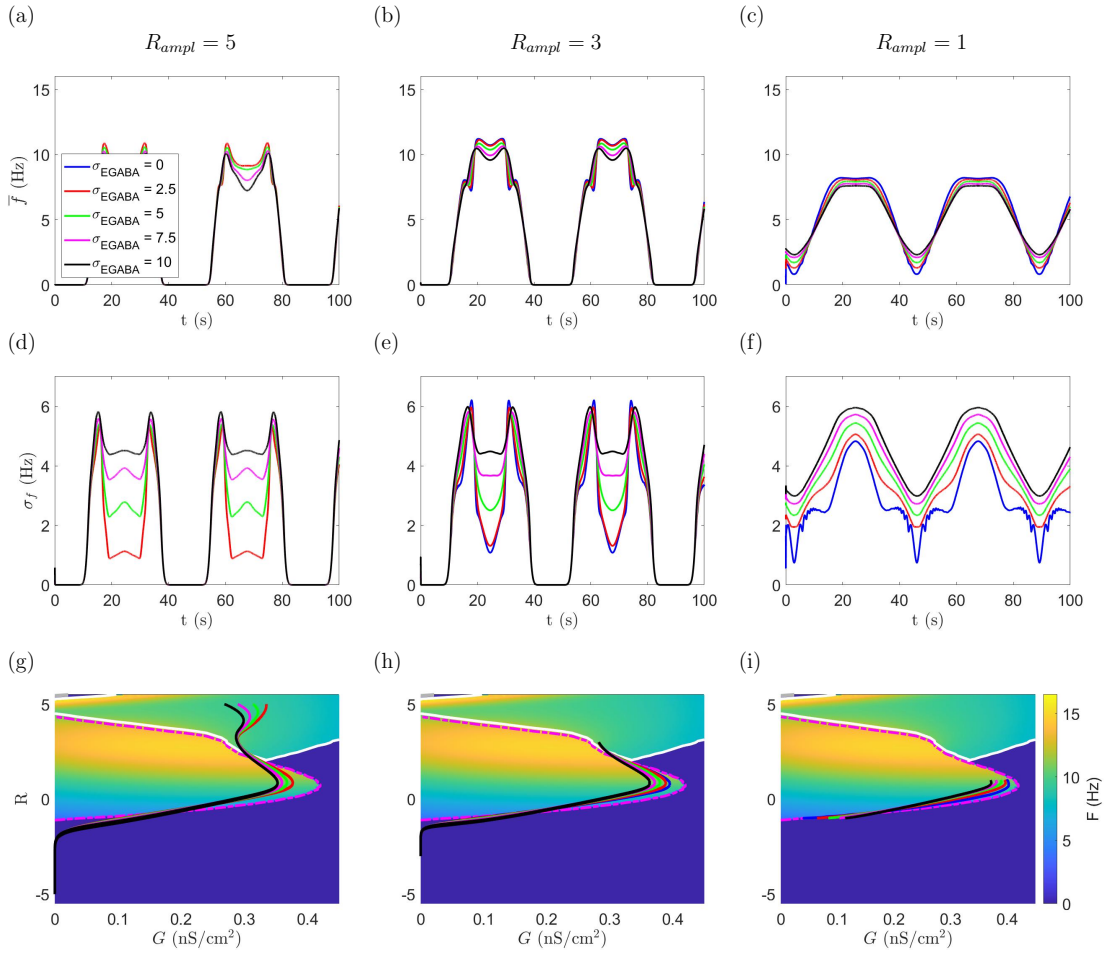


Figure III.8: Effect of introducing standard deviation in E_{GABA} given strong coupling ($g_0 = 5 \cdot 10^{-4} \frac{mS}{cm^2}$, $E_{GABA} = -55mV$, and $\sigma_\theta = 0.1$). **(a) - (f)**: In the top row and middle row, firing rates and firing rate standard deviations, respectively, are graphed for each of four values of σ_{EGABA}. **(g) - (i)**: $(\overline{G(t)}, R(t))$ trajectories are plotted (solid lines) on the neuronal firing rate surface $F(G, R, -55)$ for the same values of σ_{EGABA} as in (a) - (f). Dashed lines represent boundaries between firing regimes, as in Figure III.4. The left, middle, and right columns correspond to R_{ampl} values of 5, 3, and 1, respectively.

3.4.5 Combined effects of distributions in circadian phase and E_{GABA}

Here we consider the presumably more physiologically accurate condition where there are distributions of circadian phase and GABA reversal potential across the SCN network. Given a more realistic $\sigma_{EGABA} = 7$ along with strong coupling $g_0 = 5 \cdot 10^{-4}$ and some variance in σ_θ, (Figure III.9), we see firing-rate oscillations qualitatively similar to those without any variance in E_{GABA} ($\sigma_{EGABA} = 0$) but still with variance in circadian phase, as in Figure III.7. With variance in both circadian phase and E_{GABA}, firing rates are generally several Hz lower than with only variance in σ_θ, have means with somewhat lower peaks and higher troughs mid-day, and no secondary peaks at mid-day. Moreover, the standard deviations of firing rates are considerably larger than in Figure III.7, particularly for small σ_θ values and at mid-day in the circadian cycle. This suggests that while σ_θ has larger effects on mean firing rates than does σ_{EGABA}, both variances increase firing-rate standard deviations, with σ_{EGABA} having a larger effect than σ_θ at mid-day.

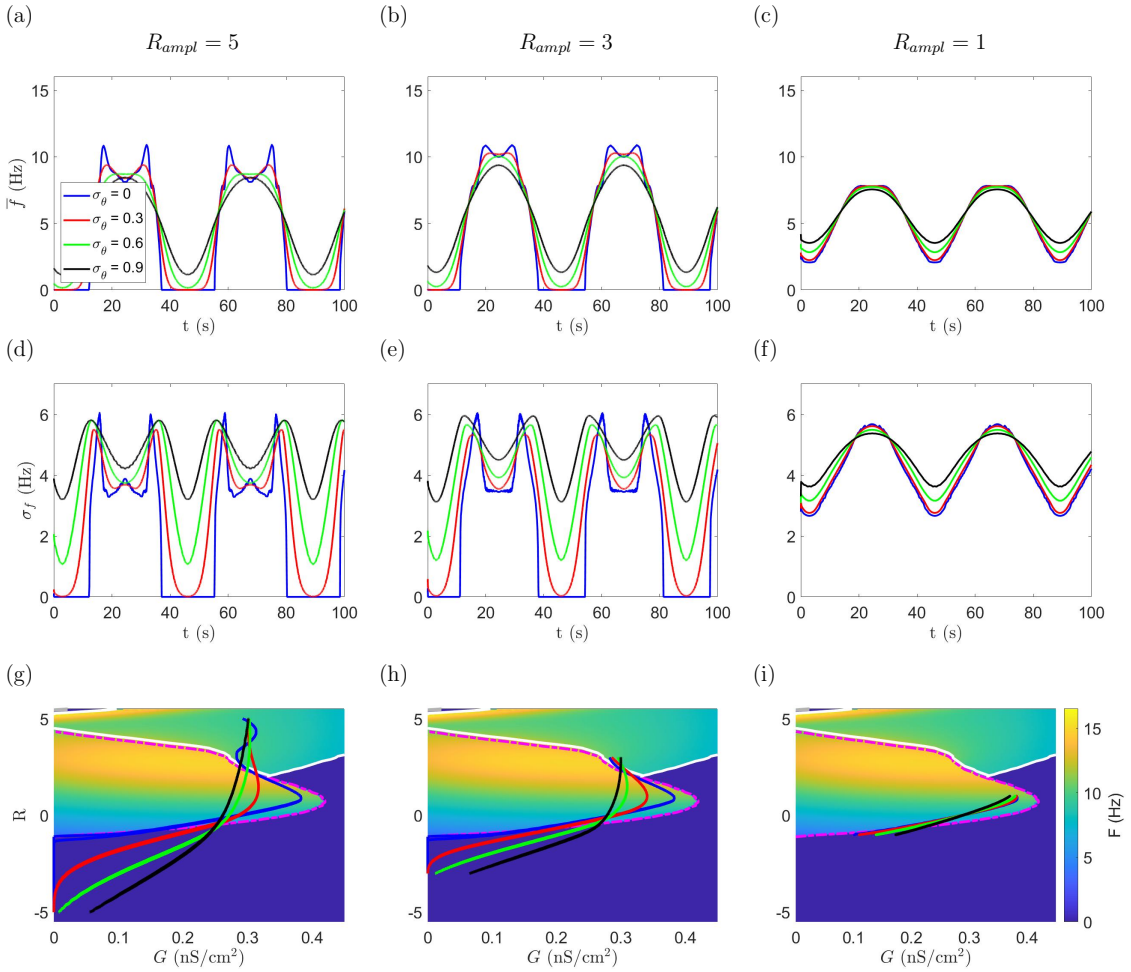


Figure III.9: Role of changing variance in circadian phase given significant variance in E_{GABA} and strong coupling between neurons ($g_0 = 5 \cdot 10^{-4} \frac{mS}{cm^2}$, $\overline{E_{GABA}} = -55mV$, $\sigma_{EGABA} = 7mV$). **(a) - (f)**: In the top row and middle row, firing rates and firing rate standard deviations, respectively, are graphed for each of four values of σ_{EGABA}. **(g) - (i)**: $(\overline{G(t)}, R(t))$ trajectories are plotted (solid lines) on the neuronal firing rate surface $F(G, R, -55)$ for the same values of σ_{EGABA} as in (a) - (f). Dashed lines represent boundaries between firing regimes, as in Figure III.4. The left, middle, and right columns correspond to R_{ampl} values of 5, 3, and 1, respectively.

3.5 Discussion

We have introduced a mean-field firing-rate formalism for a population of neurons whose electrophysiological characteristics, such as firing responses and synaptic current reversal potentials, are modeled as probability distributions across the network. Such distributions arise in our formalism from variability in underlying network parameters, such as network connectivity, or in the case of the mammalian SCN, molecular clock phase and GABA reversal potential. In applying the formalism to the SCN, not only have we incorporated variation in circadian phases and GABA reversal potentials across SCN neurons, but we have incorporated into the model the unusual firing activity of SCN neurons, including DLAMOs and DB. In doing so, we have provided and illustrated an unusually flexible firing-rate model formalism which accounts for more varied and complex neuronal properties than is typical of standard firing-rate models.

3.5.1 Advantages of our mean-field firing-rate formalism

Indeed, the primary advantage of our model is that it allows any electrophysiological properties which can be incorporated into an underlying single-neuron model to vary roughly on a continuum across the network. In other words, our model can incorporate heterogeneity in far more diverse ways than can other firing-rate models. To start, our model allows for arbitrary degree distributions in the underlying neuronal network, that is, in the number of synapses impinging on postsynaptic neurons, thereby including the primary source of heterogeneity in firing-rate models appearing in the treatment of population density approaches from [DJR$^+$08] (see Section 1.4.4), although our model does not include any noise terms. Further, our model expands upon the treatment of the two sources of heterogeneity appearing in the famous Wilson-Cowan model [WC72] (see Section 1.4.2). One such source of heterogeneity is that one population of neurons is excitatory while the other is inhibitory. Our model, on the other hand, by assuming varying GABAergic reversal potentials, allows postsynaptic currents to excite/inhibit postsynaptic neurons to different degrees within a single population! The other source of heterogeneity in the Wilson-Cowan model is that the firing thresholds of neurons differ. Our model also accounts for variable firing thresholds and variable types of firing activity by allowing parameters such as R and E_{GABA} underlying individual neuronal dynamics to vary. In doing so, our model also captures the heterogeneities in the "intrinsic frequencies" of neurons in Ott-Antonsen derived oscillator models appearing in [MPR15] or [BOF$^+$20] (as introduced in Section 1.4.6), and expands upon this by allowing for different types of activity such as depolarized low-amplitude membrane oscillations.

To capture all those heterogeneities, our model directly incorporates the dynamics of individual neurons through model variables and the deterministic "transfer" functions $F(G, \vec{P})$ and $Y(G, \vec{P})$

describing the firing-rate and gating of synaptic efficacy, respectively, for a neuron with total synaptic conductance G and parameters $\vec{P}$. Consequently, our model allows for the neuronal dynamics to be complex, so long as F and Y may be computed from appropriate single-neuron models. Only a few other firing-rate models have successfully incorporated complex neuronal dynamics into firing-rate models, including the convolutional model with averaging of Zandt et al. [ZVvPTH14], the population density model of [CG07], and the "master equation" population density model of [CCDP$^+$20].

A further advantage is that our model describes the mean and variance of firing-rates of neurons across the network without attempting to calculate the full probability distribution. In this sense, our model can capture the effects of heterogeneity among neurons in the network better than models that only describe the means, such as classic firing-rate models, including "convolutional models". Describing the variance of firing-rates also informs us about the coherence of the signals output by the modeled network, which has potentially important physiological outcomes, particularly for the SCN (see Section 3.5.2). Such descriptions are therefore lacking from models that describe only mean firing rates, including the foundational models of Wilson-Cowan [WC72], Amari [Ama77], da Silva et al. [DSHSZ74], and Jansen and Rit [JR95] as discussed in Section 1.4. Nevertheless, such convolutional models remain popular, being used to model neurological processes in a variety of contexts, including in dynamic causal modeling as used for understanding brain recording data (as discussed in Section 1.6), and more (see Section 1.6).

On the other hand, there are firing-rate models which directly output both the means and standard deviations of the firing rates across the network just as our model does, of which the most relevant to this discussion is that of Zandt et al. [ZVvPTH14]. Other models, such as the population density approach (as discussed in Section 1.4.4) describe even more information about the variability in firing rates across the network–they describe the full probability distribution of firing rates. While these models provide more information about the effects of heterogeneities, they have largely not yet been used to describe heterogeneities as varied as the SCN network requires.

Despite these advantages, our model formalism does not account for varying synaptic strengths across the network. Instead, we considered a homogeneous maximum synaptic conductance g_0 for all synaptic connections. Alternatively, heterogeneous synaptic strengths could be included in our formalism by considering a weighted sum for the total synaptic conductance $G_i(t)$ in Equation III.4. For two cases, including weighted synaptic strengths can be directly carried through our model development. For the case where all outgoing synapses from a presynaptic neuron have the same weight, a scaling factor w_j can be included within the summation of Equation III.4. For the model derivation, the product of the synaptic gating and firing rate functions $YF(G(t), \vec{P})$ in Equations

III.12 and III.13 would be scaled by this synaptic weight and its distribution would be included in the joint distribution γ_t (defined in Equation III.21). For the second case where all incoming synapses to a postsynaptic neuron have the same weight, a scaling factor w_i can be included within the summation of Equation III.4. For the model derivation, the statistics of the distribution of w_i would be incorporated into the distribution of total synaptic conductance $G(t)$ in Equations III.18 and III.20, thus modifying the distribution of $G(t)$ in Equation III.21. For the case of completely heterogeneous synaptic strengths, it is possible to extend the model formalism and derive appropriate statistics for total synaptic conductance [ZVvPTH14]. However, implementation of the model for this case would depend on the specific weighted connectivity matrix for the network.

Our model formalism also does not include a spatial component, and thus cannot account for differences in firing behavior between different portions of a network of neurons. Firing-rate models which include spatial dependencies are typically called neural field models, as discussed in Section 1.5. Examples of neural field models include that of Amari [Ama77], and more recently the model of wandering "bumps" by Bressloff [Bre19]. Accounting for spatial distributions of firing rates could improve analysis of SCN network firing activity, as the SCN is divided into dorsal (shell) and ventral (core) regions, which some evidence suggests could form functionally distinct compartments within the SCN [YKL$^+$07]. Extending our formalism to account for spatial statistics of firing rates may be considered in future work.

3.5.2 Summary of effects of SCN cellular variation

To apply our mean-field firing-rate formalism to the SCN, we have considered the scenario where SCN firing rates exhibit daily rhythmic oscillations, as observed experimentally in recordings of SCN neural activity [BP09, MSWA97, MGR86, MRG92, VHM$^+$07]. To generate rhythmic oscillations, we have forced periodic variation in the circadian proxy parameter $R(t)$. However, the best time profile for $R(t)$ to accurately simulate the effects of the molecular clock on the electrophysiological properties of SCN neurons has not been identified. Hence, we have simply modeled $R(t)$ as sinusoidally varying. Furthermore, experimental recordings of SCN electrical activity indicate that its rhythmic oscillations can vary under different conditions, such as light intensity [BOMH$^+$21] and seasonality [VHM$^+$07]. Thus, we have considered multiple sinusoidal time-profiles for $R(t)$ which induce the various ranges of activity levels that can be generated by the different firing states observed in the single SCN neuron model.

We have investigated the effects of variation in SCN network properties by simulating the model under a variety of values for four physiologically relevant parameters: synaptic coupling strength (g_0), variance in the number of synapses incoming to SCN neurons $(Var(N_{syn}))$, standard deviation

	g_0	$\overline{E_{GABA}}$	$Var(N_{syn})$	σ_θ	σ_{EGABA}
Minimum of $\overline{f}$	− or None	− or None	− or None	+	+ or none
Maximum of $\overline{f}$	− or None	+	−	−	−
Amplitude of $\overline{f}$	Variable	Variable	Variable	−	−
Minimum of σ_f	+ or None	None	None	+	− or none
Maximum of σ_f	+ or None	+	+	+	−
Amplitude of σ_f	+	+	−	−	Variable
Amplitude of Mid-Day Dip	−	Variable	Variable	−	+

Table III.1: Summary of Effects of Various Parameters on SCN Network Firing Rate Statistics. Parameters correspond to columns, the types of effects on firing rate statistics correspond to rows, and the entry in each box designates the type of correlation between the corresponding parameter and corresponding aspect of firing-rate statistics. A "+" indicates positive correlation, a "-" indicates negative correlation.

in GABA reversal potential (σ_{EGABA}), and standard deviation in circadian phase (σ_θ). We have also investigated the effects of varying mean GABA reversal potential ($\overline{E_{GABA}}$), shown in Appendix 2.7. The primary effects of changing these parameters are reflected in the maximum, minimum, and amplitude of oscillations of mean firing rates $\overline{f}$ and standard deviations in firing rates σ_f across the simulated circadian cycle. Interestingly, we also find that parameter variation affected the occurrence of a "mid-day" dip in mean firing rates. The observed effects of these parameters on firing-rate statistics are summarized in Table III.1, below.

Of all the observed effects on firing rates, perhaps the most physiologically significant are the effects on the amplitude of firing rate oscillations across the circadian cycle. Indeed, our results highlight mechanisms which shrink the amplitudes of oscillations in SCN firing rates. Decreased amplitude of SCN firing rate cycling has been observed in aged rodents [FMD+12], and would likely lead to reduced amplitude of behavioral circadian rhythms. Such blunting of behavioral circadian rhythms has likewise been associated with aging [FDR+14, HA+17] as well as with neurodegenerative diseases such as Alzheimer's disease [CSH+13, LMH+19, PMvGG18] and Parkinson's Disease [CSH+13, LMH+19, vOLH+12, WSW95].

Our results indicate that increasing σ_θ, the standard deviation in circadian phases across SCN neurons, strongly decreases the amplitude of firing rates. In terms of our model, this happens because upon increasing standard deviations in circadian phase, the range of neuron firing states grows wider. Indeed, with a large standard deviation in circadian phase, in the middle of the circadian night, a typical neuron might be at rest, but other neurons whose phases correspond to

earlier or later circadian times could be in the action potential firing regime, thus increasing the minimum average firing rate. Similarly, lower average firing rates may occur in the middle of the circadian day due to some neurons firing at lower frequencies. Physiologically, wider standard deviations in circadian phase would correspond to desynchronization in the molecular clocks of SCN neurons. Desynchronization of gene expression rhythms in the SCN has been observed in aged rodents [NNT$^+$15] and also in rodents in response to abrupt shifts in the light:dark cycle, as would occur with jet lag [NAN$^+$03].

Another parameter whose growth reduced the amplitude of firing rates is σ_{EGABA}, the standard deviation of GABA reversal potentials across the SCN network. GABAergic neurotransmission is fundamental in the SCN [MS93] and consists of both synaptic GABA currents as well as extrasynaptic tonic GABA-receptor mediated currents [MCA21]. Differences in the locations of synaptic and extrasynaptic GABA receptors and in the local chloride concentrations may lead to differences in effective GABA reversal potentials. In our formalism, accounting for variance in E_{GABA} can reflect such diversity of GABA-mediated signaling in the network. Moreover, wider variations in E_{GABA} values lead to wider variations in synaptic signaling and thus firing responses across the network. However, large effects of σ_{EGABA} on mean firing rates were observed only when coupling strength between neurons was high, i.e. when total synaptic conductance was high.

While coupling strength g_0 had some effect on the amplitude of oscillations in mean SCN firing rates, its effects were more nonlinear than those of other parameters. Most notable is that g_0 mainly affects firing rates near circadian "mid-day". This is because the coupling g_0 between two neurons only matters when the presynaptic neuron is active. Since most SCN neurons are intrinsically active at times near "mid-day", whereas few neurons are active near circadian "midnight", neuronal signaling has the largest capacity to be influenced by g_0 near "mid-day". In particular, for low values of g_0, mean firing rates show three peaks around "mid-day", one shortly before "mid-day", a second at circadian maximum (i.e. for maximal R), and the third shortly after mid-day. These peaks reflect variations in activity and synaptic signaling as $R(t)$ and $G(t)$ values traverse the DLAMO region of the $F(G, R, E_{GABA})$ surfaces. Increasing g_0 from weak coupling reduces and eventually eliminates the second peak at "mid-day", resulting in a "mid-day" dip in firing rates.

Variability in the level of high SCN neural activity, including dips, has been observed in recordings of SCN neural firing activity particularly under long photoperiods [MZJS00, VHM$^+$07]. While the significance of such mid-day dips in SCN firing rates is unclear, we speculate that they could contribute to the well-documented increased tendency to sleep in afternoons [BD89] and the afternoon dip in alertness [MH94] in humans.

Other phenomena predicted by our model are that the standard deviations of firing rates are

highly variable, varying in patterns similar to the variations in firing rates, but often surging at circadian "sunrise" and "sunset". Tracking trajectories of total synaptic conductances show that these surges occur when neural firing states approach and cross boundaries between firing regimes–where sunrise corresponds to the transition from rest to action potential firing, and sunset corresponds to the transition from AP firing to rest. Recordings of individual SCN neural activity show high variance, particularly in the dorsal region, with some units reaching peak firing at the transition when lights turn on or off, in contrast to the majority of units firing during lights on [BP09]. The high firing-rate variance at these transitions suggests that SCN signaling may become more incoherent and perhaps contribute to "sundowning" [KTSK11], in which certain neuropsychiatric conditions become exacerbated in dementia patients, for example, at or around sunset. There is evidence that sundowning is mediated by degeneration of the SCN [KTSK11] and can be improved with bright light therapy which increases amplitude and reduces variance in SCN neural activity [BOMH$^+$21].

Outside of this surging tendency at boundaries between firing regimes, firing rate standard deviations were strongly positively correlated with the total synaptic conductance G. Hence, increasing parameters which increased total synaptic conductance, such as maximum synaptic conductance g_0, tended to increase standard deviations of firing rates. Firing rate standard deviations also increased as standard deviations in E_{GABA}, in circadian phases, and in N_{syn} increased. Notably, the increase in firing-rate standard deviation as $Var(N_{syn})$ increased appeared to be the only strong effect on network firing rate statistics of changing the network connectivity structure, when average connection density $\overline{N}_{syn}$ was held constant.

3.5.3 SCN model limitations

Despite the potential predictive power of our model, our results have several limitations. In particular, we have assumed that the mean circadian proxy $R(t)$ varies sinusoidally throughout the day. As mentioned above, this assumption captures the observed oscillations in firing characteristics of SCN neurons [BOMB$^+$21, BDFP09] and generates the observed oscillations in firing rates [BP09, MSWA97, MGR86, MRG92, VHM$^+$07]. However, a more accurate time-profile for $R(t)$ may be determined by matching experimental data on the time course of daily changes in SCN firing properties with the effects of R on firing state in the SCN neuron model. Existing data provides information on SCN neural firing at different time points across the day, and current work is focused on estimating the appropriate continuous time-profile to capture this data. Further work is needed, however, to understand how external light schedules may influence the course of SCN neuron firing patterns across the day.

Model results would also be improved by incorporating additional distributions of properties

of SCN neurons. For example, variation in the period of the molecular circadian clocks across the SCN could be included. It may also be appropriate to incorporate spatial heterogeneity into our model. Indeed, as mentioned earlier in the Discussion, it may be appropriate to model the SCN as two populations–one population each for the dorsal (core) and ventral (shell) portions of the SCN. Not only may these regions play functionally different roles in the SCN [YKL$^+$07], but important neuronal properties may differ between these regions. For instance, E_{GABA} is typically lower in the ventral than the dorsal SCN [DMB$^+$15], and differences in circadian phases between the dorsal and ventral SCN have been reported [ELCCD13].

Further, studies such as [AMGF$^+$16] indicate that ventral SCN neurons have more functional connections than do dorsal SCN neurons. This in turn suggests that neurons in the ventral SCN may share a larger number of synaptic connections with other SCN neurons than do dorsal SCN neurons. Due to the larger number of synapses, ventral SCN neurons would likely experience more synaptic signaling and thus higher total synaptic conductances G than dorsal SCN neurons. Because ventral SCN neurons, as mentioned above, also have lower values of E_{GABA}, there would be an anticorrelation between G and E_{GABA} in the SCN. Such anticorrelation would violate the assumption of independence we have made to calculate the probability distribution of (G, R, E_{GABA}) across SCN neurons, potentially altering predicted mean firing-rates.

Finally, even if such assumptions of independence are not violated, it is not clear what the distribution of the total synaptic conductance G, alone, should be at any particular time. Indeed, even though we have provided conditions that would make G normally distributed, such conditions require that $Var(N_{syn})$ be much smaller than $\overline{N_{syn}}$, which is unknown. Nevertheless, the aforementioned functional connectivity study finds that functional connections in the SCN follow a power-law degree distribution [AMGF$^+$16], which, if also true of the synaptic connections, would help us better understand the distribution of G.

Also potentially affecting the distribution of G and SCN firing activity in general is the presence of gap junctions in the SCN. Indeed, it has been shown that gap junctions play a role, in particular, in synchronizing SCN neurons [WCW14]. However, our model only takes into account chemical synapses in its current formulation, and as it only summarizes firing-rate statistics, is not able to directly account for synchrony between SCN neurons.

3.5.4 Conclusions

Despite the limitations, our mean-field firing-rate model for the SCN network makes predictions about the effects on population firing statistics of the unique properties of the SCN neural network, such as circadian phase effects on firing state, distributions of $GABA$ reversal potentials, and

modulation of synaptic signaling due to atypical neuron firing states. In doing so, our model clarifies how these properties can lead to circadian disruptions, such as the blunted circadian rhythm amplitude associated with aging and neurophysiological diseases. More generally, our model formalism provides a new method to capture the effects of heterogeneity in the electrophysiological properties across neurons on the network firing rate statistics.

CHAPTER IV

Modeling the Effect of Varying Light Schedules on SCN Firing-Rates

4.1 Introduction

4.1.1 Light and circadian rhythms

The behavior of mammals is adapted to the environmental light cycle of the daily sunrise and sunset [MB22]. For example, humans are typically active during the day. Many rodents, on the other hand, are typically active during the night. Such synchronization of behavioral rhythms with the environmental light cycle likely serves an evolutionary purpose [BTDB17]. Indeed, many mammals that are awake during the day, such as humans, lack the visual acuity needed to find food at night.

Along with activity patterns a number of other mammalian physiological processes are also synchronized to the light-dark cycle [FK17]. For instance, in humans, body temperature, the levels of the stress-inducing hormone cortisol, and metabolic activity all peak during the light phase [FK17]. Such 24 hour rhythmic processes, as discussed in the previous chapter, are called circadian rhythms.

The proper timing and robustness of circadian rhythms with daily rest and waking activity patterns is critical for health [FK17]. In particular, it is critical that the circadian rhythms in different physiological processes are correctly aligned with one another. Indeed, circadian misalignment has been shown to be a risk factor for diabetes [PR22], heart disease [CVWS19], and cancer [SK19].

The importance of circadian alignment is especially clear when individuals fly across a number of time zones [FK17]. With the change in the timing of rest and waking activity, individuals may feel groggy while awake, feel hungry at unusual times of day, and have difficulty sleeping when it is dark outside. This experience of jet-lag arises because circadian rhythms may be more aligned with the usual light-dark cycle at their origin location, rather than with the new environmental light cycle at their destination location.

However, circadian misalignment isn't limited to long-distance air travel. For instance, many high-school and college students go to bed earlier and wake up earlier on school nights. On the

other hand, on the weekend, they stay up late, sleep longer in the mornings and may sleep for longer durations to make up for any sleep deprivation they incurred during the week. Unfortunately, the rapid change in sleep schedule going from weekdays to weekends and vice versa disrupts circadian rhythms and can lead to jet-lag like symptoms. Hence, this phenomenon has been termed social jet-lag [WDMR06].

Another source of circadian misalignment arises from shift work [BBK22]. Of particular concern is shift work that involves switching between day shifts and night shifts (see [BBK22]). Such rapid changes in activity rhythms lead to jet-lag like issues, where circadian rhythms fail to change fast enough to keep up with the rapid shift in the times of day where the workers need to be awake and alert. Moreover, because switching between day and night shifts may occur frequently for months or years, the frequent jet-lag like symptoms can negatively impact quality of life (see [NKS$^+$18, KKL$^+$16])[1].

Thus, it is vital to understand how circadian misalignment arises and how it is corrected. In particular, it is crucial to understand what aspects of our environment affect circadian rhythms, and to understand how such interactions occur. Much research has been dedicated towards doing so over the past fifty years or so (see e.g. [GR10] and [GRPK19] for reviews). Such research has identified that the most important environmental stimulus for altering our circadian rhythms is likely environmental light [MM97].

In particular, researchers have identified the pathway by which external light alters circadian rhythms [MM97], as discussed briefly in the preceding chapter. This pathway starts when light, particularly blue light, hits the retina, and interacts with cells known as intrinsically photosensitive retinal ganglion cells (ipRGCs). These cells are similar to cones and rods in the retina, which react to photons by sending signals towards the visual cortex, and help you develop a visual representation of the world around you. However, unlike cones and rods, ipRGCs generally do not contribute to vision, nor are their signals initially processed by the thalamus, as most other sensory input would be. Instead, signals from the ipRGCs are transmitted directly to the mammalian master clock–the suprachiasmatic nucleus (SCN)–via the retinohypothalamic tract (RHT) [MM97].

The signals from the ipRGCs, after passing through the RHT, arrive at particular neurons throughout the SCN [FCHC16]. The signals then coordinate intracellular transcription and translation processes of genes and proteins, such as PER genes in the Period family, within these neurons [HMB18]. Specifically, these molecular interactions result in periodic gene and protein expression that has an approximately 24 hour period. Namely, experiments suggest that inputs from the ipRGCs

[1] A range of types of shift work, beyond rotating shift work, negatively affects quality of life. [BBK22, FK17, NKS$^+$18, KKL$^+$16]

synchronize the rhythms of these intracellular transriptional-translational feedback loops (TTFLs) in SCN cells to environmental light (see chapter 5 of [FK17]).

The proteins and genes expressed through the intracellular TTFLs affect the electrophysiological membrane properties of SCN neurons resulting in a daily rhythm of firing activity [JM16]. As a result of the entrainment of TTFL rhythms by light input, both PER1-expressing [BDFP09, JM16] and non-PER1-expressing SCN neurons are most active during the light cycle, and least active during the dark cycle [BDFP09], as seen in Figure IV.1. PER1-expressing neurons may, in particular, be critical for phasing the rhythms of the intracellular TTFLs with firing-rate rhythms in the SCN [JM16]. This is important because coordination of "downstream" processes, such as rhythms in the sleep-promoting hormone melatonin (but not so much in behavioral activity rhythms, see [SLTL96]), by the SCN is in large part due to synaptic signaling [SLCG05]. Indeed, SCN neurons are known to project to brain regions related to sleep regulation (see [AM01]), and transplanting healthy SCN tissue fails to restore circadian rhythms, particularly in endocrine rhythms [MBJM+99], in mice with damaged SCNs. Moreover, much of the information about such synaptic signaling is carried along the firing-rates of the SCN [JTM15], (see Figure IV.1).

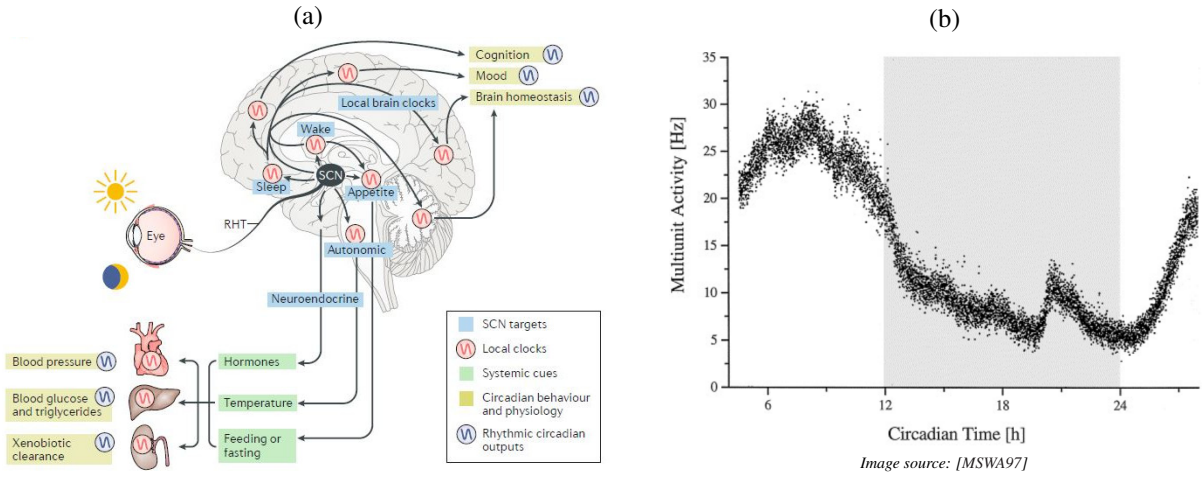


Figure IV.1: How light from the eye coordinates circadian processes throughout the body. **(a)**: Light reaches the eye, and induces signals to be sent from the retina to the suprachiasmatic nucleus (SCN) via the retinohypothalamic tract (RHT). The light entrains SCN neuron firing activity to the external light-dark cycle, which then coordinates a variety of bodily processes with 24-hr rhythms. Such processes include the sleep-wake cycle, appetite rhythms, and blood pressure rhythms. **(b)**: The SCN likely coordinates such processes by sending out robust 24-hr oscillatory signals described by the firing-rate of its neurons. Panel (a) is from [HMB18], and (b) is from [MSWA97].

4.1.2 Overview of modeling goals

Indeed, the pathway by which light influences firing rates of SCN neurons is complicated, and has not been completely experimentally identified. To contribute to our understanding of how varying the external light schedule affects SCN firing rates, we turn to modeling. In particular, we

seek to create a mechanistic model of how external light specifically affects SCN firing rate (see [DMB+15] for a related model). To do so, we extend the the previous chapter's firing rate model developed for the SCN.

In particular, we model light effects on SCN firing rate as a 3 step process:

1. Input light schedules to a network model for SCN molecular clock phases (θ) [HBF19a]. This step models how altering the external light cycle affects statistics of the phases of the molecular clocks contained within individual PER-expressing SCN neurons.

2. Convert molecular clock phase to the circadian phase proxy (R) discussed in the preceding chapter. This step involves mapping the phases of the molecular clocks to the firing activity of the SCN neurons. In particular, from the statistics for molecular clock phase, this step should output the statistics for R.

3. Input R statistics into our SCN firing rate model. The statistics for R will influence the mean and standard deviation of the firing rates across SCN neurons.

We illustrate the model in Figure IV.2.

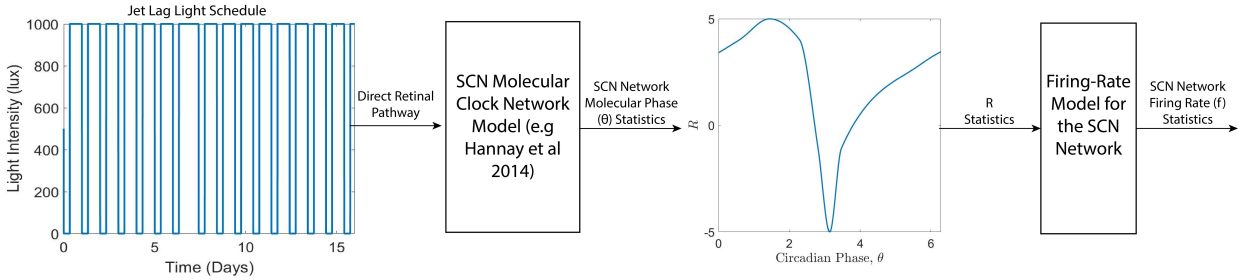


Figure IV.2: Model of the pathway by which light entrains SCN firing rate statistics. We propose (1) using a model to describe how external light affects the intracellular molecular clock that keep track of time within individual SCN neurons. Then, we propose (2) arriving at statistics for the quantity R which alters the electrophysiological state of SCN neurons in the single neuron model of [DBI+13] from the molecular clock phase statistics in SCN neurons. Finally, (3), we propose using the statistics of R in the SCN firing rate model we described in the preceding chapter.

This project is still a work in progress. Hence, in this chapter, we discuss our preliminary results. We begin by discussing step (1), where we provide background for the Hannay et al. model [HBF19a] in Section 4.2.1, and discuss how it links environmental light to phases of oscillators underlying the generation of circadian rhythms. We then provide background for step (2), discussing preliminary efforts to link phases of molecular clocks to the electrophysiological state of neurons via the SCN single neuron model of [DBI+13], all in Section 4.2.2. We briefly explain how to combine steps (1) and (2) to find more realistic inputs to our SCN firing-rate model in step (3) in Section 4.2.3, although, largely, we explain how to combine steps (1) and (2) in Sections 4.2.2 and

4.2.3. In Section 4.2.4, we show simulations of the preliminary model of the full pathway by which environmental light affects SCN outputs. In doing so, we illustrate that this model has the potential to capture important information not described by existing models of the pathway.

For the remainder of this chapter, we summarize the work we have done towards completing step (2). Namely, we present results on expressing the electrophysiological activity of PER1-expressing SCN neurons in terms of rhythms in two potassium conductances according to data from [BDFP09]. These potassium conductances have been found to be the primary physiological changes to SCN neurons throughout the day [BDFP09], and thus may be the means by which the molecular clock affects the firing-activity of the neurons [DBI$^+$13]. In particular, we simulate the model from [DBI$^+$13] for a wide range of combinations of those potassium conductances, and identify firing-rates and average membrane voltages for each of those combinations. We incorporate heterogeneity in those potassium conductances by integrating the firing-rate and membrane voltages against a multivariate normal distribution of those potassium conductances. We then match the resulting population-average firing-rates and population-average membrane voltages to those computed from experimental recordings of firing-rates and membrane voltages from actual SCN neurons across the 24 hr day.

We conclude the chapter by discussing the work we have done (Section 4.4). We further discuss limitations of the work in Section 4.4.1, and begin to discuss how we can go about improving our model.

4.2 Firing-rate models have the potential to provide information not captured by standard biological clock oscillator models

4.2.1 Model for statistics of molecular circadian phase

We use the model of Hannay et al. [HBF19a] to complete step (1). In [HBF19a], the authors start by treating the SCN as a system of coupled oscillators according to [HBF19b]. In particular, the authors assume that the intracellular molecular clock in each SCN neuron produces regular oscillations of protein expression so that each neuron i has a phase ϕ_i. The neurons are assumed to be coupled sufficiently weakly that the oscillator network dynamics can be reduced to the system of coupled Kuramoto phase oscillators [HBF19b]. As such, ϕ_i can be expressed as an ordinary differential equation, wherein in the absence of external input, ϕ_i increases linearly at a rate ω_i, which corresponds to the oscillator's intrinsic frequency:

$$\frac{d\phi_i}{dt} = \omega_i \qquad \text{(in the absence of noise and external inputs)}$$

However, as neuron dynamics are intrinsically noisy due to stochastic opening of ion channels, the authors assume ϕ_i dynamics are subject to the influence of additive white noise of the form

$$\text{additive white noise } = \sqrt{D}\eta_i(t),$$

where $\sqrt{D}$ is the strength of the noise and η is the white noise itself. Coupling between SCN neurons is modeled as in Kuramoto networks wherein a larger difference in phases between neurons i and j causes a faster increase in the rate at which the phase of neuron i tends toward the phase of neuron j, according to

$$\text{coupling between neurons } = K\sin(\phi_j - \phi_i + \beta),$$

where K is the strength of the coupling, and β represents an offset in the phases at which the coupling is strongest. Combining these terms describes a Kuramoto network of N weakly coupled oscillators, where the phase of oscillator i is modeled as

$$\frac{d\phi_i}{dt} = \omega_i + \sqrt{D}\eta_i(t) + \frac{K}{N}\sum_{i=1}^{N}\sin(\phi_j - \phi_i + \beta) \quad .$$

Hannay et al. [HBF19a] use this model to describe the dynamics of molecular clock phases in SCN neurons. However, the preceding description is lacking the response of SCN molecular clocks to external light.

To incorporate light into the preceding Kuramoto-type system of equations, the authors assume that the clock phase of neuron i responds to light according to its phase response curve. The phase response curve $Q(\phi_i)$, describes how much the clock phase of the neuron advances, (i.e. the phase value becomes what it would be later in the day), or delays, (i.e. the phase value becomes what it was earlier in the day), in response to external light stimuli. Thus, if light were to advance the phase of neuron i, the rate of change of ϕ_i should increase proportionally to the size of the advance. The authors assume the corresponding proportionality is given by the strength $B(t)$ of the light signal arriving from the RHT:

$$\text{change in } \frac{d\phi_i}{dt} \propto B(t)Q(\phi_i),$$

where $B(t)$ is derived from the Process L formalism described in [KFJ99]. The resulting phase

oscillator network model is then

$$\frac{d\phi_i}{dt} = \omega_i + \sqrt{D}\eta_i(t) + \frac{K}{N}\sum_{i=1}^{N}\sin(\phi_j - \phi_i + \beta) + B(t)Q(\phi_i).$$

The authors then reduce this N-dimensional system to a low-dimensional system of differential equations using an Ott-Antonsen [OA08] approach, adapted to fit a more realistic distribution of the intrinsic frequencies ω_i according to the reduction from [HBF19a]. The resulting macroscopic description of the network dynamics describes the time evolution of the (first order) Kuramoto order parameters ψ and R_k. R_k describes the coherence of the oscillators, or amplitude of the collective oscillations in phase, and is computed by mapping the phases onto the complex unit circle, finding the average of the resulting complex numbers, and computing the corresponding magnitude:

$$R_k := \left| \frac{1}{N}\sum_{i=1}^{N} e^{i\phi_i} \right|.$$

Notably, when $R_k = 1$, all neurons have exactly the same phase, and thus the network is completely synchronous. In contrast, when $R_k = 0$, phases are uniformly distributed in $[0, 2\pi]$, and so the network is completely asynchronous. The order parameter ψ, on the other hand, describes the mean phase of the oscillators across the network, and is given by mapping the phases on the complex unit circle, taking the average, and computing the corresponding argument:

$$\psi = Arg\left(\frac{1}{N}\sum_{i=1}^{N} e^{i\phi_i} \right).$$

Because under robust light schedules and behavioral rhythms, the molecular clock in SCN cells essentially keeps track of the time of day (see e.g. [HMB18]), we expect ψ for the SCN to progress more or less linearly from 0 to 2π throughout the 24 hr day. The Hannay et al. model [HBF19a] tracks the impact of light on the Kuramoto order parameters ψ and R_k (see [HBF19a] for a full description of model equations and parameters).

To incorporate light effects on clock phases in the SCN from the Hannay model into the model of the pathway for light effects on the firing rates of SCN neurons, we treat the Kuramoto order parameters as describing the statistics of the phases of the molecular clocks contained in SCN neurons (below in Section 4.4.1 we discuss limitations of this interpretation).

Besides the mean and amplitude of the molecular clock phases, we can describe the full

distribution of phases from the complete set of Kuramoto order parameters Z_n defined by:

$$Z_n := \frac{1}{N}\sum_{k=1}^{N} e^{in\phi_k}.$$

That is, Z_n is simply the average of $e^{in\phi}$ across the network. We can re-express Z_n in terms of the circular probability distribution $\rho(\phi, t)$, i.e. the probability that a molecular clock has phase ϕ at time t. To do so, we simply use the standard computation of the average across the network applied to $e^{in\phi}$:

$$Z_n \approx \int_0^{2\pi} \rho(\phi, t) e^{in\phi} d\phi,$$

which happens to be (2π times) the n^{th} Fourier coefficient of the distribution. Hannay et al. [HBF19b] show how the system for Z_n can be closed so that phase dynamics appropriately fit experimental measurements of SCN cell molecular phases.

To describe this closure of Z_n, it helps to introduce the higher order parameters $R_{k,n}$ and ψ_n:

$$\begin{aligned} R_{k,n} &:= |Z_n| \\ \psi_n &= Arg(Z_n) \end{aligned}$$

so that

$$Z_n = R_{k,n} e^{i\psi_n}.$$

Hannay et al. [HBF19b] show that for molecular phases of SCN neurons the following closure is appropriate:

$$\begin{aligned} R_{k,n} &= R_k^{n^2} \\ \psi_n &= n\psi \end{aligned}$$

Using this, we rewrite ρ in terms of its Fourier series,

$$\rho(\phi, t) = \sum_{n=-\infty}^{\infty} \frac{Z_n}{2\pi} e^{-in\phi}$$

which can now be expressed in terms of R_k and ψ:

$$\begin{aligned}\rho(\phi,t) &= \frac{Z_0}{2\pi} + \sum_{n=-\infty,n\neq 0}^{\infty} \frac{Z_n}{2\pi} e^{-in\phi} \\ &\approx \frac{R_{k,0}}{2\pi} + \sum_{n=-\infty,n\neq 0}^{\infty} \frac{R_k^{n^2} e^{in\psi}}{2\pi} e^{-in\phi}.\end{aligned}$$

Using that $R_{k,0} = 1$ by definition, applying Euler's formula, and then rearranging and canceling terms yields the phase-amplitude form of the Fourier series for $\rho(\phi, t)$:

$$\begin{aligned}\rho(\phi,t) &= \frac{1}{2\pi} + \frac{1}{2\pi} \sum_{n=-\infty,n\neq 0}^{\infty} R_k^{n^2} e^{in(\psi-\phi)} \\ &= \frac{1}{2\pi} + \frac{1}{2\pi} \left(\sum_{n=-\infty,n\neq 0}^{\infty} R_k^{n^2} \left[\cos(n(\psi-\phi)) + i\sin(n(\psi-\phi))\right] \right) \\ &= \frac{1}{2\pi} + \frac{1}{2\pi} \left(\sum_{n=-\infty}^{-1} R_k^{n^2} \left[\cos(n(\psi-\phi)) + i\sin(n(\psi-\phi))\right] + \sum_{n=1}^{\infty} R_k^{n^2} \left[\cos(n(\psi-\phi)) + i\sin(n(\psi-\phi))\right] \right) \\ &= \frac{1}{2\pi} + \frac{1}{2\pi} \left(\sum_{n=1}^{\infty} R_k^{n^2} \left[\cos(n(\psi-\phi)) + \cos(-n(\psi-\phi)) + i[\sin(n(\psi-\phi)) + \sin(-n(\psi-\phi))]\right] \right) \\ &= \frac{1}{2\pi} + \frac{1}{2\pi} \left(\sum_{n=1}^{\infty} R_k^{n^2} 2\cos(n(\psi-\phi)) \right) \\ &= \frac{1}{2\pi} + \frac{1}{\pi} \sum_{n=1}^{\infty} R_k^{n^2} \cos(n(\psi-\phi)).\end{aligned}$$

This series converges as long as $|Z_1| := R_k < 1$, which is almost always the case. Note that the distribution can be expressed using the Jacobi-theta function:

$$\rho(\phi,t) = \frac{1}{2\pi} \Theta\left(\frac{-(\psi-\phi)}{2\pi}, R_k \right),$$

where

$$\Theta(z,q) := \sum_{n=-\infty}^{\infty} q^{n^2} e^{inz}.$$

Hence, we can obtain the distribution of circadian phases ϕ_i from R_k and ψ computed by the Hannay model.

It turns out that the Fourier series we have just identified is the Fourier series for a wrapped normal distribution. Indeed, if σ is defined so that

$$R_k = e^{-\sigma^2/2},$$

then,

$$R_k^{n^2} e^{i\psi n} = e^{-n^2\sigma^2/2} e^{i\psi n},$$

which is (2π times) the n^{th} Fourier coefficient for the wrapped normal distribution of ϕ_i values with mean μ and variance σ^2 given by

$$\begin{aligned} \mu &= \psi \\ \sigma^2 &= -2\ln(R_k). \end{aligned}$$

Because $\rho(\phi, t)$ has the same Fourier series as does the wrapped normal distribution with the preceding parameters, $\rho(\phi, t)$ is the preceding wrapped normal distribution.

The fact that $\rho(\phi, t)$ follows a wrapped normal distribution allows us to seamlessly integrate the mean phase ψ and amplitude R_k variables of the Hannay model into our firing-rate model for the SCN. It is not so clear at first, though, that in our firing-rate model, that "seamless integration" is possible. Indeed, in our firing-rate model, we never explicitly state that we are assuming a wrapped-Gaussian distribution for circadian phase. Specifically, in our firing-rate model, we have assumed that circadian phases θ are normally distributed with mean $\overline{\theta}$ and standard deviation σ_θ, and that θ can (theoretically) take any value in $(-\infty, \infty)$. However, θ is a phase, and thus effectively lives on the unit circle. The aforementioned normal distribution, thus is an "unwrapped" version of the corresponding wrapped normal distribution which describes the distribution of phases on the unit circle. Thus, for all intents and purposes, we treat θ as being distributed according to a wrapped Gaussian. Namely, in our firing-rate model, we take θ into account only when we calculate the firing-rate statistics by integrating across the distribution of R. In particular, in Equations III.10, III.11, III.12, III.13 from the firing-rate model (see IV), we integrate, for example, the firing-rate function $F(G, R(\theta), E_{Gaba})$ against the distribution of θ, integrating over an interval 10 standard deviations of θ long, centered at the mean θ. However, because $R(\theta)$ is 2π-periodic, integrating against θ over that range amounts to integrating over the wrapped Gaussian.

4.2.2 Mapping molecular clock phases to firing activity of SCN neurons

To incorporate the wrapped-normal distribution of circadian phases output by the Hannay model into our firing-rate model for the SCN, we need to find the relationship between circadian phase θ and the quantity R which modulates the electrophysiological state of SCN neurons, as per Diekman et al. in [DBI$^+$13]. Recall that in the simulations of our firing-rate model in the preceding chapter,

we assumed that this relationship was given by

$$R = 5\sin(\theta).$$

To improve upon the choice of the function that maps θ to R, we use a model developed by Alicia Colclasure (formerly at Colorado School of Mines) and Cecilia Diniz Behn (Colorado School of Mines) in unpublished work.

To find an improved relationship between θ and R, they examined the daily time course of SCN neuron firing changes predicted by the model of [DBI+13]. Diekman et al. model individual SCN neurons in the Hodgkin-Huxley formalism with an applied current, I_{app}, two sodium currents, two calcium currents, as well as three potassium currents:

$$C\frac{dV}{dt} = I_{app} - I_{Na} - I_K - I_{CaL} - I_{CaNonL} - I_{KCa}(R) - I_{K-leak}(R) - I_{Na-leak}.$$

We discuss the details of this model in the appendix and Table B.3. Notably, two of these currents–the potassium leak current I_{K-leak} and the current through calcium mediated potassium channels I_{KCa}– depend on the parameter R. R controls these potassium currents by modulating their maximum conductances, g_{K-leak} and g_{KCa}. Thus,

$$I_{KCa} = g_{KCa}(R)(V - E_K)$$
$$I_{K-leak} = g_{K-leak}(R)(V - E_K),$$

where E_K is the reversal potential of potassium, (i.e. the membrane potential at which the diffusion force of K^+ ions due to concentration differences across the cell membrane is balanced with the electrical forces generated by the potential difference across the cell membrane). Specifically, the two conductances are modeled to be inversely proportional to $\exp(R)$, according to:

$$g_{KCa}(R) = 2 + \frac{198}{1 + \exp(R)}$$
$$g_{K-leak}(R) = \frac{0.2}{1 + \exp(R)}.$$

Moreover, as shown by [BDFP09], out of a number of the known currents affecting the electrophysiological behavior of SCN neurons, it is variations in the potassium currents driven by variations in the corresponding conductances that vary across the circadian day (also see [POM06, MWM+06, KM04] for evidence indicating the circadian variability of these currents, and [Col11] for a review of the link between molecular clocks and electrophysiological state of SCN

neurons). Accordingly, Diekman et al. drive the potassium conductances g_{KCa} and g_{K-leak} by varying R across the circadian day (hence we call R the "circadian proxy", in Chapter III). For the time-dependence of R, they adapt a well-known model of biochemical oscillations–the Goodwin oscillator (see [GR21] for a review)–to describe the molecular clocks within SCN cells. Thus, as time progresses across the 24 hr day, so does R. As a result, the values of (g_{KCa}, g_{K-leak}) trace out a closed loop in (g_{KCa}, g_{K-leak}) parameter space.

Further, as (g_{KCa}, g_{K-leak}) vary with R, the electrophysiological state of the neuron changes, so that when R is low, the neuron is in a hyperpolarized state, such as rest. However, as R increases, the neuron enters more depolarized states, transitioning from rest, to action potential spiking, to depolarized low amplitude membrane oscillations (DLAMOs, as discussed in chapter IV), to depolarization block.

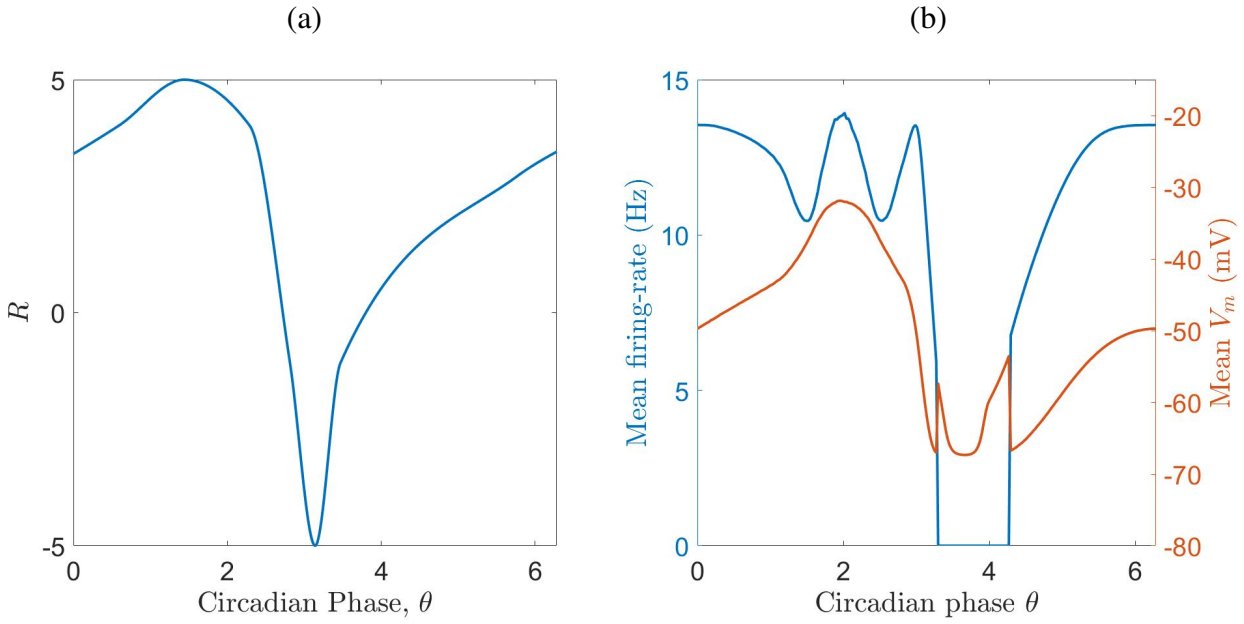


Figure IV.3: The relationship between R and θ proposed by Colclasure and Diniz Behn. **(a)**: R as a function of θ. **(b)**: The corresponding firing-rate and time-averaged membrane voltages for an SCN neuron as described by the Diekman et al. model of [DBI$^+$13].

However, Alicia Colclasure and Cecilia Diniz Behn note that if R varies as predicted by [DBI$^+$13], then the timing of the predicted electrophysiological states of the neuron fail to match the timing of the electrophysiological states observed by Belle et al. in [BDFP09]. To correct this, Colclasure and Diniz Behn revise the dependence of R on circadian phase θ, as pictured in Figure IV.3. The resulting neuron firing rates and membrane voltage traces are shown in Figure IV.3.b. As can be seen from the figure, the model of SCN neurons using the new R function produces consistent firing activity throughout the day. This new function is somewhat in agreement with data from [BDFP09] concerning non-PER1-expressing neurons, but is not in agreement with experimental

data from Belle et al. in [BDFP09] concerning PER1 neurons. Thus, revisions of this $R(\theta)$ function need to be made to be able to describe PER1-expressing neurons. We discuss such revisions in Section 4.3. In the next section, we implement the Colclasure-Diniz Behn $R(\theta)$ function to predict how light-induced changes in circadian phases affect SCN firing rates.

4.2.3 Modeling light's effects on statistics of SCN firing rates

With a defined $R(\theta)$ function, we can directly input the time-varying statistics of circadian phases output by the Hannay model into our SCN firing-rate model to predict the effects of changing light schedules. To obtain preliminary results, we use the Colclasure-Diniz Behn $R(\theta)$ function.

4.2.4 Preliminary results

To illustrate the potential usefulness of our model, we apply the model to a jet lag event. In this jet lag event, an individual travels 8 time zones west via plane, so that they are exposed to light for an extra 8 hours late in the day. For simplicity, we set the intensity of light exposure to around 1000 lux whenever it is light outside (such a light intensity corresponds to sitting near a window during the day), and to 0 lux whenever it is dark outside. During the jet lag event, we simply extend the 1000 lux period so it lasts an additional 8 hrs at the end of the day. The resulting light schedule would then be the one shown in Figure IV.4, where we take the jet lag event to occur on day 6.

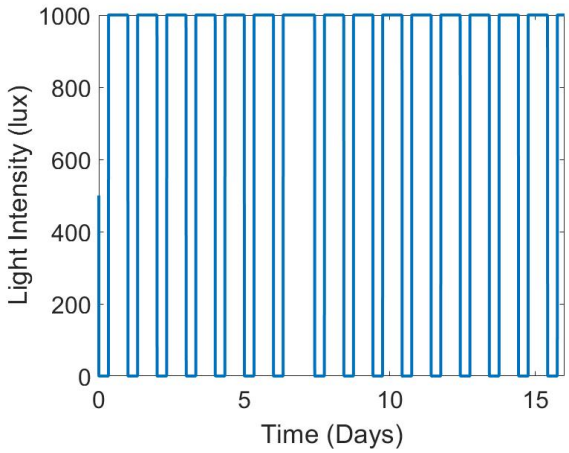


Figure IV.4: A jet lag light schedule for someone traveling 8 time zones west. From days 0 - 5, the individual is in complete darkness (0 light intensity) from midnight to 8 a.m, and in light with intensity resembling filtered sunlight from 8 a.m to midnight. However, on day 6, the individual travels 8 time zones west. As a result, daylight lasts 8 hours later, hence the time in which they are in 1000 lux light on day six lasts an additional 8 hours. From then on, they resume their schedule of being in the dark from midnight - 8 a.m, and being in the dark from 8 a.m to midnight, but in terms of time in the new time zone. Such a situation amounts to an 8 hour phase-delaying event.

Such a situation amounts to an 8 hour phase delay. The resulting effects on SCN mean firing rate and its standard deviation of this 8 hour phase delay are shown in Figure IV.5. The results illustrate that the phase delay in the light schedule shifts mean firing-rates ($\overline{f}$), so their peaks occur later and later relative to the control case, where the individual would have stayed at home rather than hopped on a plane. Eventually, though, the firing rates cease shifting, indicating that the circadian clock

model has entrained to the new environmental light cycle. A similar shift occurs in the standard deviations σ_f of SCN firing-rates (Panel (b) of Figure IV.5).

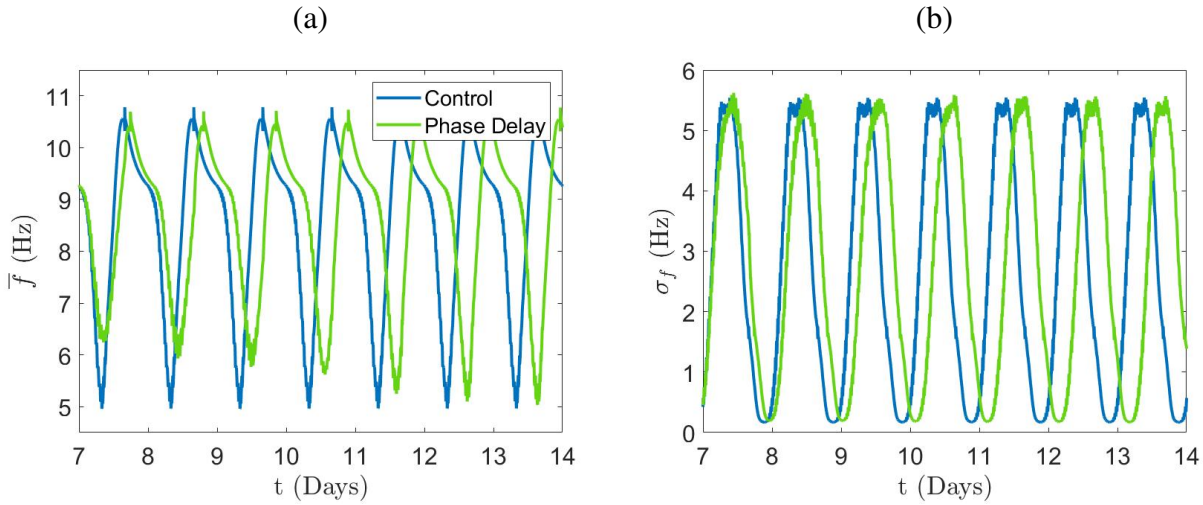


Figure IV.5: Jet lag by traveling 8 time zones west: an 8 hr phase delay. **(a)**: the mean firing-rate output by the SCN of the phase-delayed individual (green), shown alongside the mean firing-rate of the same individual if they had not been phase-delayed (blue). **(b)**: analogous plots illustrating the standard deviation of firing rates across the SCN.

Interestingly, during days 7 - 11, when the entrainment to the new environmental light cycle is occuring, the amplitude of firing-rate rhythms is smaller than the control. This can be expected from variations in circadian clock phase amplitude R_k of the Hannay model during re-entrainment to the new light schedule, but the magnitude of the reduction in amplitude of the mean firing-rate of the SCN may differ, as we discuss below. On the other hand, if the individual flies 8 hrs east, they experience light 8 hrs earlier in the day than in their origin location, and thus experience an 8 hr phase advance. In this case we obtain much more interesting results (Figure IV.6).

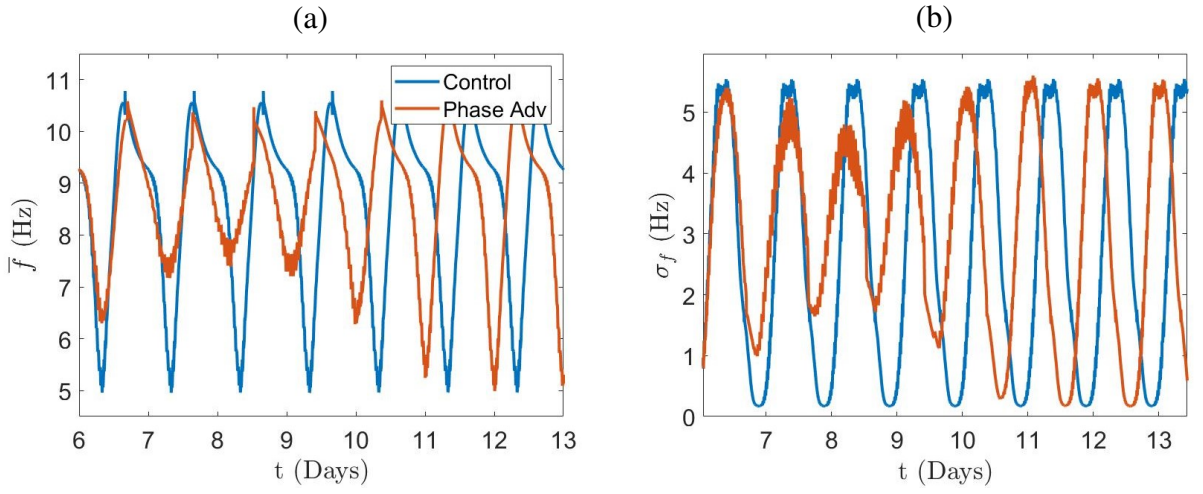


Figure IV.6: Jet lag by traveling 8 time zones east: an 8 hr phase advance. **(a)**: the mean firing-rate output by the SCN of the phase-advanced individual (green), shown alongside the mean firing-rate of the same individual if they had not been phase-advanced (blue). **(b)**: analogous plots illustrating the standard deviation of firing rates across the SCN.

Namely, not only is there a large reduction in the amplitude of the oscillations in $\overline{f}$, there is a large decrease in the amplitude of σ_f. In particular, during the dark phase, instead of nearly all neurons being comparatively inactive (having a firing-rate around 4-6 Hz), firing rates are higher with a much greater standard deviation.

Relationship between the amplitudes of oscillations in mean SCN firing-rates and in circadian clock phases

To further analyze how circadian clock phases affect SCN firing rates in our modeled pathway, we compare the amplitude of oscillations in $\overline{f}$ and R_k. Assuming a fixed amplitude R_k for the oscillations in circadian clock phase and using that as input to our firing-rate model, we can compute the resulting oscillations in $\overline{f}$ output by our firing-rate model. When assuming the sinusoidal $R(\theta)$ function, this amplitude relationship is linear, indicating that firing rate oscillation amplitudes are directly proportional to circadian clock oscillation amplitudes (Figure IV.7)b. However, when using the Colclasure – Diniz Behn $R(\theta)$ function there is a nonlinear relationship in the amplitudes (Figure IV.7).a.

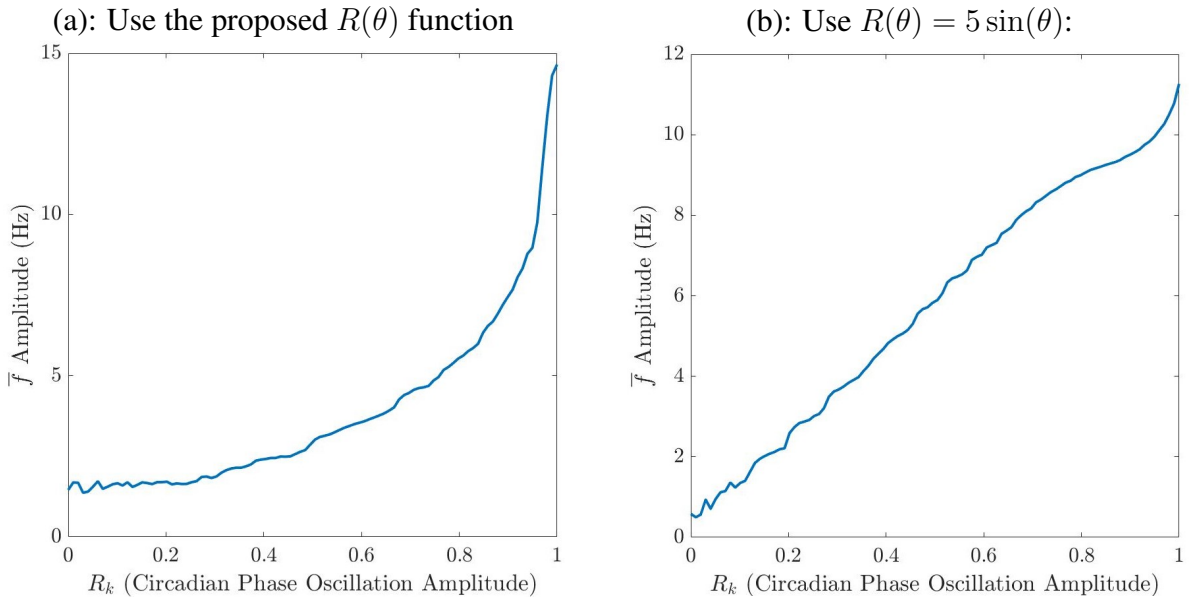


Figure IV.7: Relationships between amplitudes in mean firing-rate $\overline{f}$ oscillations and the amplitude R_k of the corresponding oscillations in circadian molecular clock phases (θ) across the network. **(a)**: the relationship between $\overline{f}$ amplitude and R_k is nonlinear, when using the $R(\theta)$ function proposed by Colclasure and Diniz Behn. **(b)**: the relationship between R_k and the amplitude of oscillations in $\overline{f}$ is roughly linear, when using $R(\theta) = 5\sin(\theta)$.

Accounting for σ_f increases the information in the signal relayed by the SCN

As seen in Figures IV.5 and IV.6, there may be some correlation between $\overline{f}$ and σ_f, where e.g. the peaks of $\overline{f}$ align with the peaks of σ_f. This could lead one to suspect that one can gain

information about the standard deviation of the firing-rate output by our model from the mean firing-rate alone. However, we show here that that is not the case.

To do so, we choose fixed values of circadian clock amplitude R_k and mean phase θ, and compute the corresponding statistics for R using the Colclasure – Diniz Behn function. We then input these statistics into our SCN firing-rate model and simulate it to steady state. We show the results of these simulations in Figure IV.8.

From Figure IV.8, we see that $\overline{f}$ and σ_f vary differently with R_k and θ. For instance, when R_k is high and θ is at low values, σ_f varies with changes in θ, whereas R_k varies little. Thus, one would expect we could gain considerable information about the state of the SCN by considering σ_f in addition to $\overline{f}$. We verify this using information theory.

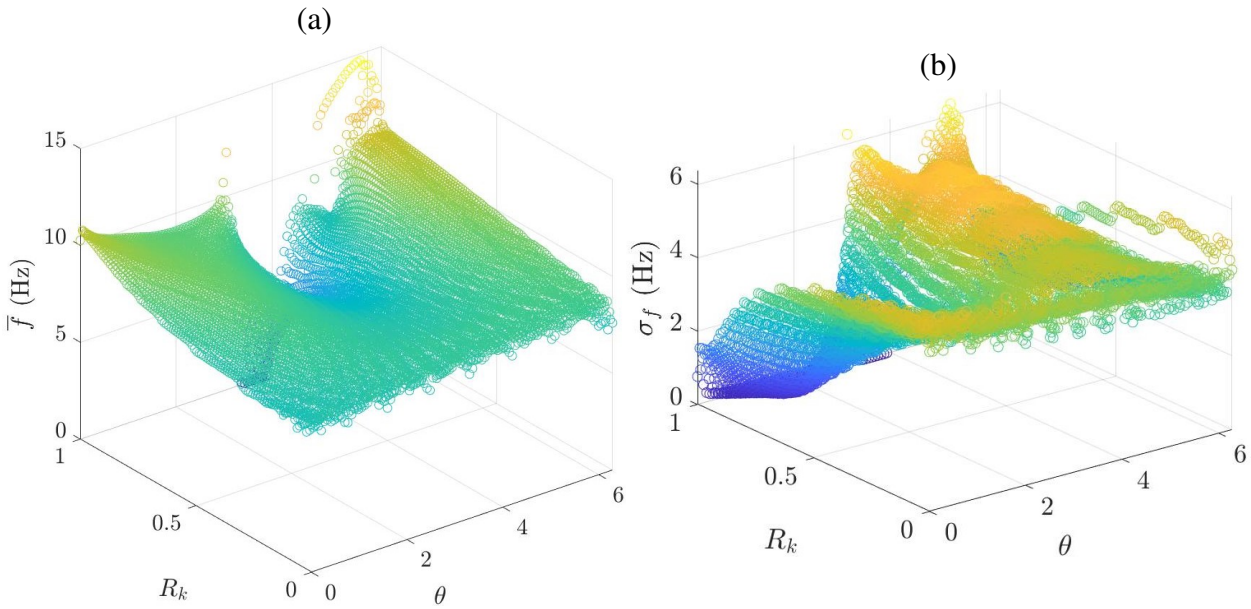


Figure IV.8: The relationship between circadian clock phase properties and the associated SCN firing-rate statistics. **(a)**: Mean firing-rate versus the circadian clock phase amplitude R_k and mean phase ψ (which we take to be circadian phase θ), assuming the $R(\theta)$ function proposed by Colclasure and Diniz Behn. **(b)**: The analogous plot, except involving the standard deviation σ_f in firing-rates.

To start, we quantify the total amount of uncertainty in $(\overline{f}, \sigma_f)$ via the joint entropy:

$$\text{Joint Entropy of } (\overline{f}, \sigma_f) = 5.3180.$$

This joint entropy is, in other words, the expected amount of information gained upon measuring both $(\overline{f}, \sigma_f)$. We can compare this joint entropy to the total amount of information gained about $(\overline{f},$

σ_f) by measuring $\overline{f}$, alone. Such information is described by the mutual information:

$$\text{Mutual Information of } (\overline{f}, \sigma_f) = 0.8758.$$

Thus, the amount of information gained by measuring $\overline{f}$ is only a small proportion of the total expected amount of information gained by measuring both $\overline{f}$ and σ_f. Indeed, the proportion of expected information conveyed by $(\overline{f}, \sigma_f)$ that would be gained by measuring σ_f in addition to $\overline{f}$ is given by

$$\text{Rajski Distance between } \overline{f} \text{ and } \sigma_f \ = 0.8353.$$

In particular, the Rajski distance (see [Raj61]), calculated as

$$\text{Rajski distance} := \frac{\text{joint entropy} - \text{mutual information}}{\text{joint entropy}},$$

is essentially a normalized mutual information, and lies in $[0, 1]$, (since joint entropy $\geq$ mutual information ≥ 0, which in turn follows fairly straightforwardly from the definitions of joint entropy and mutual entropy). The closer the Rajski distance is to 1, the greater the proportion of expected information conveyed by $(\overline{f}, \sigma_f)$ that would be gained by measuring σ_f in addition to $\overline{f}$. Thus, since the Rajski distance between $\overline{f}$ and σ_f is close to 1, most of the information about the $(\overline{f}, \sigma_f)$ signal output by the SCN stands to be gained by measuring σ_f instead of only measuring $\overline{f}$. Moreover, even if we were to use $R(\theta) = 5\sin(\theta)$ instead of the Colclasure-Diniz Behn $R(\theta)$ function, we would see that the Rajski distance would still be high (0.6501), but not as high.

By directly comparing the values of $\overline{f}$ and σ_f (Figure IV.9), we see that σ_f and $\overline{f}$ are indeed related. In fact, the set of possible values of $(\overline{f}, \sigma_f)$ lie roughly on a parabola. Thus, there is some relationship between $\overline{f}$ and σ_f, hence why the Rajski distance is not 1[2]. However, we can also see that for given values of one of these quantities, there is considerable uncertainty in the other. For instance if $\overline{f} = 9$, σ_f could be almost anywhere in its range of 0 to ≈ 6. We see similar uncertainty when using $R(\theta) = 5\sin(\theta)$ relationship (data not shown).

[2]While the information theoretic measures such as mutual information and the Rajski distance, can effectively describe any kind of relation between two random variables, correlation coefficients can only describe linear relations, and hence would miss the parabolic relationship between $\overline{f}$ and σ_f.

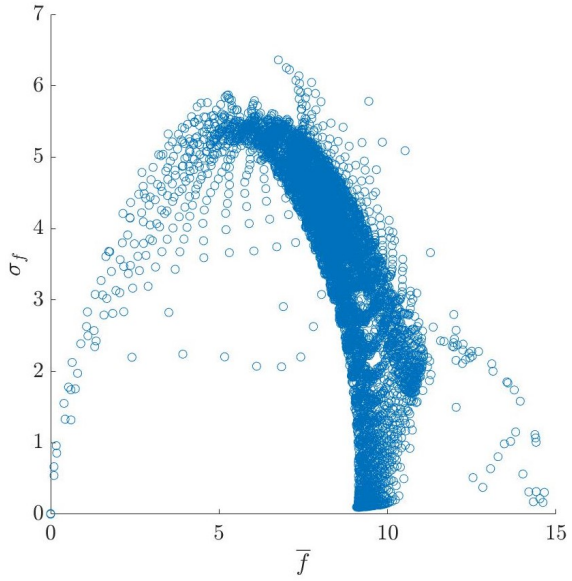


Figure IV.9: The relationship between mean firing-rates ($\overline{f}$) and the standard deviations σ_f of firing-rates. While the set of $\overline{f}$-σ_f pairs is roughly parabolic, there remains considerable uncertainty in the value of σ_f given $\overline{f}$, and vice-versa.

4.3 Improving the mapping of molecular clock phases to firing activity of SCN neurons

As discussed above, the Colclasure-Diniz Behn function $R(\theta)$ for the effect of mean circadian phase on electrophysiological properties of SCN neurons erroneously produces spiking activity for large stretches of time when per1-expressing SCN neurons should be more-or-less quiescent (compare Figure IV.3 to Figure IV.10).

Thus, in this section, we construct a more physiologically sound relationship between R and θ. To do so, we computationally extracted experimental data of firing rates and average membrane voltage measured across the circadian cycle reported in Figures 1.A, 1.A1, 2.C, and 2.F from [BDFP09], and report the extracted data in Figure IV.10, (where to illustrate the periodicity of firing rate and average membrane voltages, we duplicate the data from [BDFP09] so it runs over two circadian days instead of one). Constraining firing rates and average voltages of the SCN neuron model [DBI$^+$13] with this data, we determine the changes throughout the day occurring in the potassium conductances g_{KCa} and g_{K-leak} which are driven by changes in R. Namely, we identify a probable closed loop in the space of values of (g_{K-leak}, g_{KCa}), which represents the daily time variation in (g_{K-leak}, g_{KCa}) in SCN neurons, thereby relating the electrophysiological states of the neurons as per [BDFP09] to the molecular clock phase across SCN neurons.

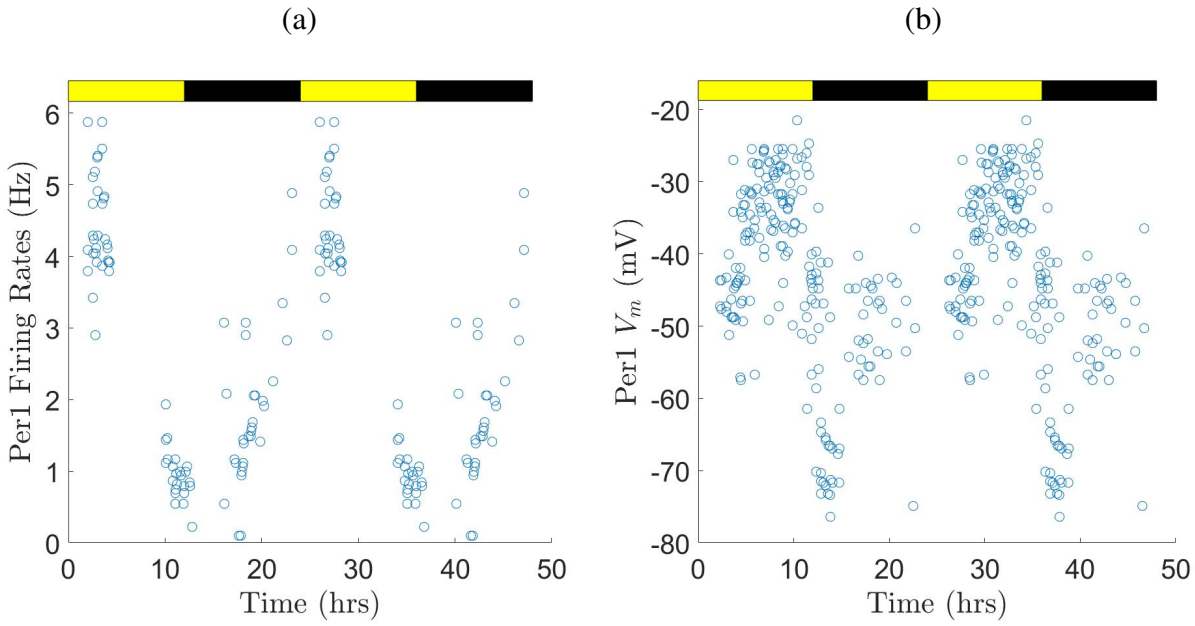


Figure IV.10: Data on electrophysiological activity of per1-expressing SCN neurons extracted from [BDFP09]. **(a)**: Firing rates vs time across various per1-expressing SCN neurons. Notably, data from about 4.25 to 10 hrs after lights were turned on are missing. **(b)**: Average membrane voltages vs time across various per1-expressing SCN neurons. For both (a) and (b), data for day 1 (times 0 - 24) are copied and also shown as data for day 2 (times 24 - 48), for illustrative purposes. Yellow and black bars above each panel respectively indicate the light and dark phases of the day.

To start, we seek to relate (g_{K-leak}, g_{KCa}) values to the electrophysiological activity of SCN neurons using the single SCN neuron model [DBI$^+$13]. For ranges of g_{KCa} and $g_{K,leak}$ values, we simulate the single neuron model until it reaches a stable solution. For each run, we calculate the corresponding firing-rate F by counting the number of peaks (if any), and calculate the corresponding average membrane voltage V_m by averaging the membrane voltage across one of its oscillations (assuming it oscillates). We show the results of these simulations in Figure IV.11, panels (a.i) - (a.iii).

Using the experimental measurements of average membrane voltage and firing rate of SCN neurons across the circadian cycle reported in [BDFP09], we can gain considerable information about the neuron's g_{KCa} and $g_{K,leak}$ by matching to the simulation results. To do so, we attempt to approximate the actual voltage dynamics of the neuron with the voltage dynamics from the model. Namely, for a given voltage value $V_m = v$, we know that the g_{KCa}-$g_{K,leak}$ combination responsible for that V_m value must lie on the $V_m = v$ contour in $(g_{KCa}, g_{K,leak})$ space. Such contours are shown in Figure IV.11. Likewise, for a given firing rate value $F = f$, we know that the g_{KCa}-$g_{K,leak}$ combination generating that firing rate must lie on the $F = f$ contour in $(g_{KCa}, g_{K,leak})$ space. The (g_{KCa}, g_{KLeak}) combination that generates this average voltage and firing rate then lies at the intersection of those contours.

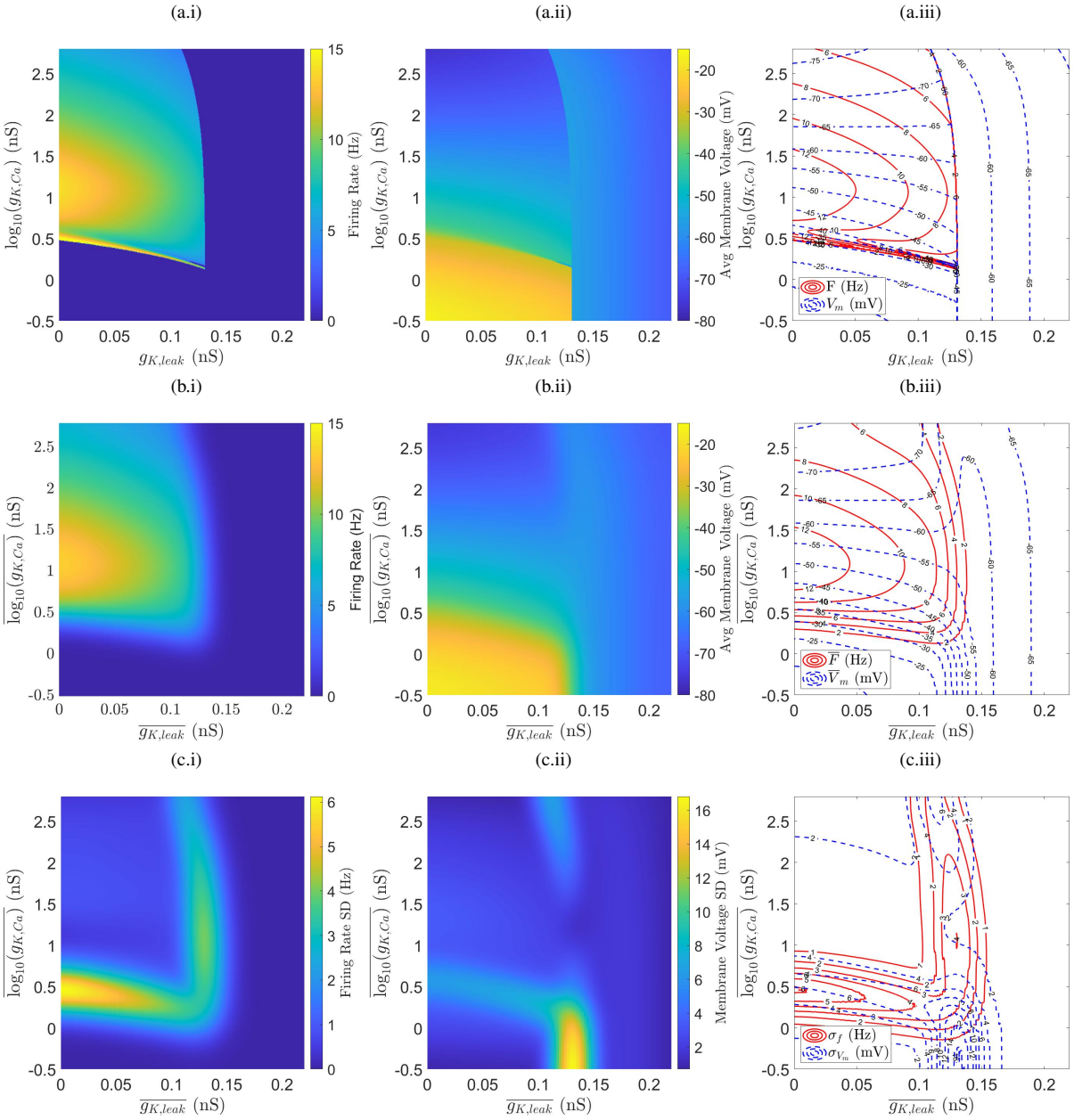


Figure IV.11: Modeled firing-rates and time-averaged membrane voltages of model SCN neurons as a function of the potassium conductances g_{KCa} and g_{K-leak} which are believed to vary across the 24 hr day [DBI$^+$13]. **(a)**: simulations of the Diekman model at each (g_{K-leak}, g_{KCa}) combination on an evenly spaced 501×501 grid. We start with resting voltage of -67 mV in each simulation. Panel (i) shows a heatmap of the firing-rates for each simulation, and panel (ii) shows a heatmap of the average membrane voltage across one oscillation for each conductance combination. Panel (iii) shows the corresponding contour curves for firing-rates (red) and time-averaged membrane voltages (blue). **(b)**: Assuming that g_{K-leak} and g_{KCa} each follow a normal distribution with means corresponding to points on the grid and with fixed standard deviations, we compute the expected means of firing-rates (i) and of time-averaged membrane voltages (ii), and show the corresponding contours in (iii). **(c)**: we show the same plots as in (b), except instead of means, we show standard deviations.

For instance if $F = 12$ Hz and $V_m = -50$ mV, then we find the intersection of the $F = 12$

and $V_m = -50$ contours which occurs, for instance, at $\log(g_{KCa}), g_{K,leak} = (1, 0.05)$. On the other hand, if $F = 0$ Hz and $V_m = -25$ mV, then we don't have enough information to identify a specific (g_{KCa}, g_{KLeak}) combination. Indeed, the $F = 0$ "contour" is a wide swath of (g_{KCa}, g_{KLeak}) space which contains the $V_m = -25$ contour. Thus, the intersection of the two "contours" is the $V_m = -25$ contour, which corresponds to a 1-dimensional set of (g_{KCa}, g_{KLeak}) combinations.

This poses a problem. Namely, as would be the case for $F = 12$ Hz and $V_m = -25$ mV, data only about the firing-rate F and membrane voltage V_m of a neuron may be insufficient to specify the values of g_{K-leak} and g_{KCa} for the neuron. Moreover, we don't even have data describing the V_m and F for any one particular neuron. Instead, we have F for a variety of neurons across the day, and we have V_m for a variety of neurons across the data, but we don't have any (F, V_m) combinations.

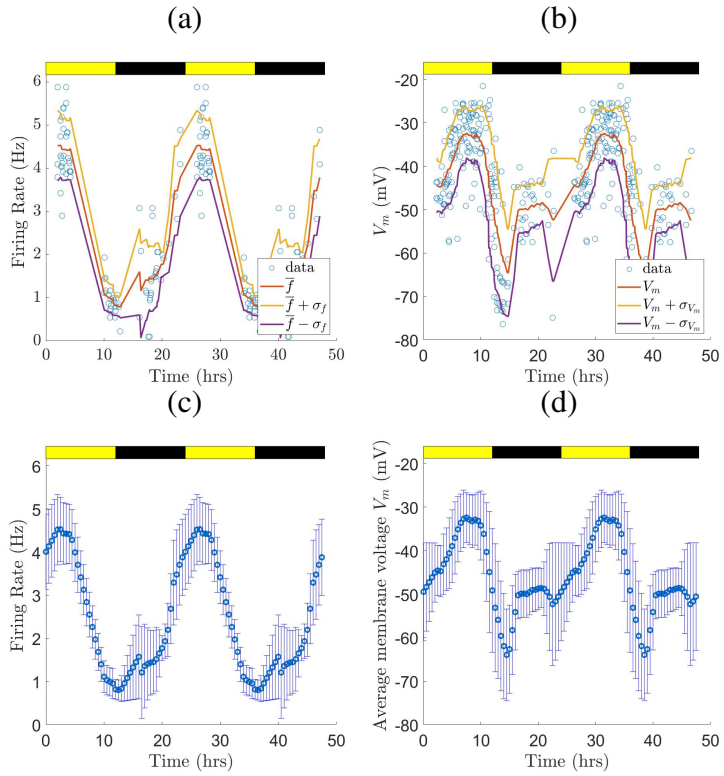


Figure IV.12: Firing-rate and average membrane voltage means and standard deviations for per1-expressing SCN neurons. **(a)**: Using running time-windows, we compute the mean firing rate ($\overline{f}$ red) and the standard deviation σ_f of the firing rates, which we visualize by showing $\overline{f}+\sigma_f$ (yellow) and $\overline{f}-\sigma_f$ (purple). **(b)**: we show the analogous results, but applied to time-averaged membrane voltages, rather than firing-rates. Data points (blue circles) in both panels are those from Figure IV.10, the particular firing-rates and membrane voltages recorded from SCN neurons as extracted from figures from [BDFP09]. **(c)**: Mean firing-rate data (circles) and the correponding standard deviations (error bars) from (a) interpolated at 0.5 hour intervals throughout the day. **(d)**: The same as (c), except for time-averaged membrane voltages instead of firing rates. Interpolating ensures that every 0.5 hours, we simultaneously know the mean and standard deviation across the network of both firing-rates and time-averaged membrane voltages, which we can then use to match to simulations of the Diekman et al. model [DBI$^+$13] as in Figure IV.11 panels (b.i)-(c.iii). Yellow and black bars above each panel respectively indicate the light and dark phases of the day.

We can attempt to instead determine (F, V_m) combinations by computing the average V_m and the average F across several time bins. Binning the data not only allows us to extract the V_m-F pairs we are looking for, but we can compute the standard deviations of V_m and F across data in those bins. We can use the extra information provided by the standard deviations of V_m and F in those bins to further constrain our estimates of (F, V_m) when most neurons lie in a state such as rest

or depolarization block where their firing-rates are zero. In Figure IV.12, we show an example of how we can bin the data across running time windows of 3 hrs to extract combinations of F and V_m which we can subsequently use to estimate g_{KCa} and g_{K-leak}.

However, to make use of the extra information provided to us by the mean and standard deviations of F and V_m across a time bin in the experimental data, we need to be able to compute means and standard deviations of F and V_m across model SCN neurons. To do so, we assume that some sort of heterogeneity in g_{K-leak} and g_{KCa} will produce a spread in firing-rates and time-averaged membrane voltages. In particular, to expedite computations, we assume that at any particular time, g_{K-leak} and g_{KCa} are both normally distributed and are uncorrelated, so that their joint distribution is

$$\gamma_t(x, y) = P_{g_{K-leak}}(x) \cdot P_{g_{KCa}}(y).$$

In this expression, $P_{g_{K-leak}}(x)$ and $P_{g_{KCa}}(y)$ are the normal distributions for the respective potassium conductances. We further assume that the standard deviations of each of the two conductances are 5% of their respective ranges, (although in the future we hope to be able to fit those standard deviations to the Belle et al. data [BDFP09], in addition to fitting the corresponding means). Then, for each possible (g_{K-leak}, g_{KCa}) combination across a 501×501 uniform grid of (g_{K-leak}, g_{KCa}) values, we compute the resulting joint distribution γ_t centered at those combinations. Moreover, for each of those combinations, we integrate the firing-rate and membrane voltages from Figure IV.11 panels (a.i) and (a.ii) against the joint distribution γ_t o firing-rates and membrane voltages to compute mean values of firing-rates and time-averaged membrane voltages. Wesimilarly integrate the squared deviation from the mean against the joint distribution γ_t to compute the standard deviation of firing-rates and of time-averaged membrane voltages. We show the results of such simulations in Figure IV.11, panels (b.i), (b.ii), (c.i), and (c.ii).

Then, given $\overline{f}$, $\overline{V_m}$, σ_f, and σ_{V_m} for the bin as computed from experimental data from [BDFP09], we can find a point in (g_{K-leak}, g_{KCa}) parameter space for which the simulated $\overline{f}$, $\overline{V_m}$, σ_f, and σ_{V_m} using the SCN neuron model integrated against the joint distribution best match the original data. To find the best match, we want to make the simulated values of $\overline{V}_m$ and $\overline{f}$ as close to the experimentally measured values of $\overline{V}_m$ and $\overline{f}$, respectively, as possible.

We also want the simulated values of σ_{V_m} and σ_f to be as close as possible to the experimentally measured values. However, we have chosen a fixed value of the standard deviations of g_{K-leak} and g_{KCa} which might not be truly representative of the actual spread of g_{K-leak} and g_{KCa} values. Because the particular values of the predicted g_{K-leak} and g_{KCa} depend on the magnitudes of the standard deviations of g_{K-leak} and g_{KCa}, it is thus possible that the simulated σ_{V_m} and σ_f are not

representative of the σ_{V_m} and σ_f calculated from the experimental data. Thus, until we can get a better estimate of the magnitudes of the standard deviations in g_{K-leak} and g_{KCa}, we focus on matching simulated values of $\overline{f}$ and $\overline{V}_m$ to the $\overline{f}$ and $\overline{V}_m$ calculated from the experimental data.

To do so, we compute an error measurement for each value of (g_{K-leak}, g_{KCa}), which describes the quality of the match between $\overline{f}$ and $\overline{V}_m$ calculated from the experiments in [BDFP09] with the simulated $\overline{f}$ and $\overline{V}_m$. To compute the error measurement, we first normalize all data by dividing by its range of values. We then find the squared error between simulated $\overline{f}$ and actual $\overline{f}$, add that to the squared error between simulated $\overline{V}_m$ and actual $\overline{V}_m$, then take the square root of the result. The resulting error measurement is

$$\text{error } = \sqrt{(\text{simulated } \overline{f} - \overline{f})^2 + (\text{simulated } \overline{V}_m - \overline{V}_m)^2} \tag{IV.1}$$

where data and predicted data have been normalized appropriately.

We then minimize the error across possible (g_{K-leak}, g_{KCa}) combinations, by computing the error for all points in the simulated data from Figure IV.11, and selecting the smallest of those errors. In doing so, we arrive at the trajectory of (g_{K-leak}, g_{KCa}) values that is traversed during one circadian cycle illustrated in panels (a.i), (a.ii) and (a.iii) of Figure IV.13. Notably, the corresponding daily variation in mean firing rates and mean membrane voltages obtained with this trajectory loop are in good agreement with the data from [BDFP09] (panels (b.i) - (b.ii) in the figure), and the simulated standard deviations in voltage are close to that of the data (panel (c.ii) in the figure). On the other hand, the simulated standard deviations in firing-rates are quite a bit off (panel (c.i) in the figure).

We can understand this trajectory in (g_{K-leak}, g_{KCa}) space by considering its predictions for firing rates and average voltages in different daily time windows. To start, when the lights are turned on (time 0 h, as reported from the experimental data [BDFP09]), per1-expressing SCN neurons are all active, which requires that trajectories are in the non-purple region of panel (a.i). Further, at time 0 h, the corresponding time-averaged membrane voltages are around -50 to -40 mV, which corresponds to the greenish-blue region in panel (a.ii) which separates depolarization block (yellow) from other firing regimes. However, because near time 0 h there is action potential firing [BDFP09], the trajectory must be on the action-potential firing side of the greenish-blue region in the membrane voltage panel (a.ii) of figure IV.13.

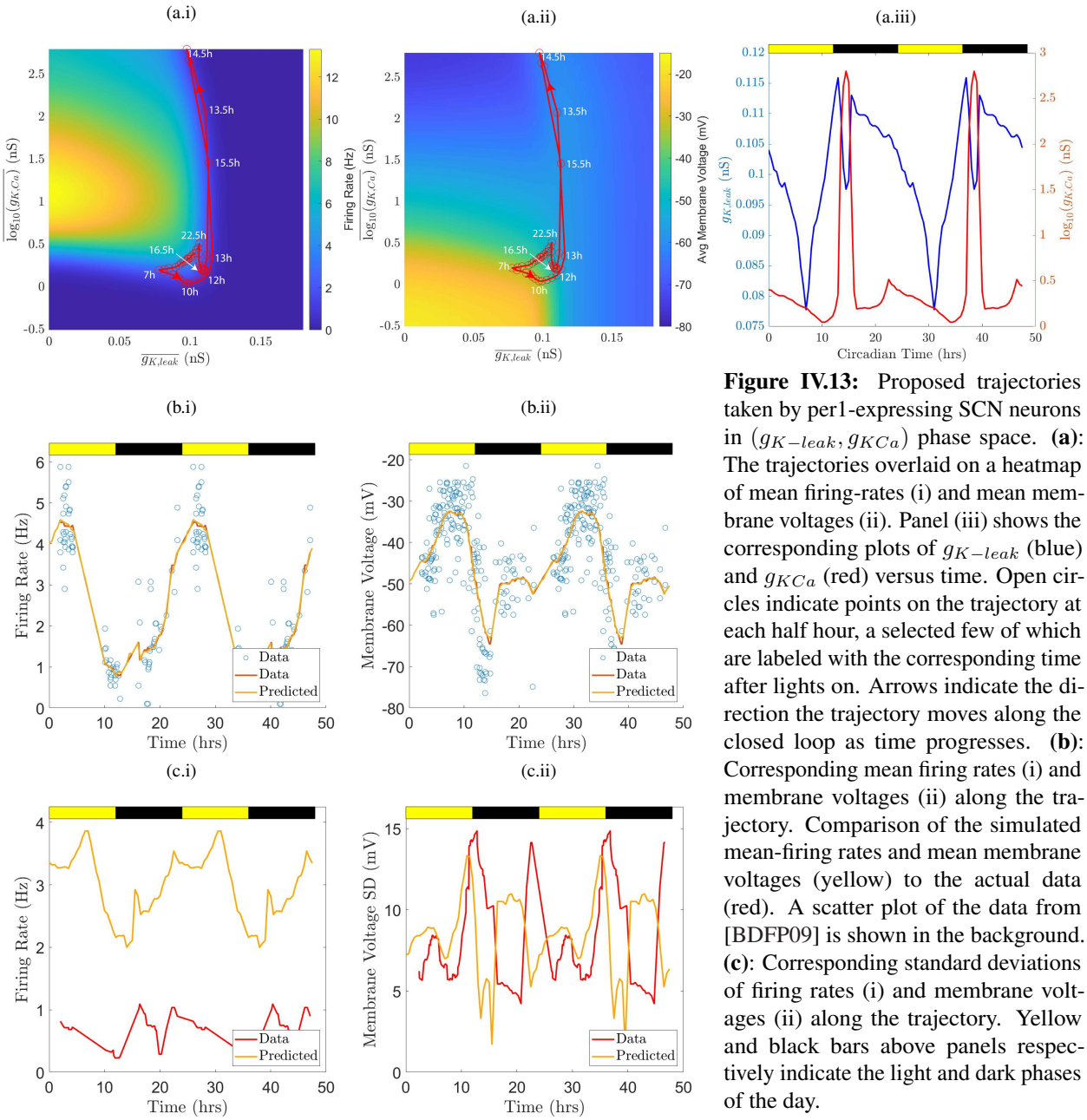


Figure IV.13: Proposed trajectories taken by per1-expressing SCN neurons in (g_{K-leak}, g_{KCa}) phase space. **(a)**: The trajectories overlaid on a heatmap of mean firing-rates (i) and mean membrane voltages (ii). Panel (iii) shows the corresponding plots of g_{K-leak} (blue) and g_{KCa} (red) versus time. Open circles indicate points on the trajectory at each half hour, a selected few of which are labeled with the corresponding time after lights on. Arrows indicate the direction the trajectory moves along the closed loop as time progresses. **(b)**: Corresponding mean firing rates (i) and membrane voltages (ii) along the trajectory. Comparison of the simulated mean-firing rates and mean membrane voltages (yellow) to the actual data (red). A scatter plot of the data from [BDFP09] is shown in the background. **(c)**: Corresponding standard deviations of firing rates (i) and membrane voltages (ii) along the trajectory. Yellow and black bars above panels respectively indicate the light and dark phases of the day.

However, between times 5.25 h and 10.75 h, no per1-expressing neurons produce action potentials [BDFP09]. Instead, all neurons are either in depolarization block, or are producing depolarized low-amplitude membrane oscillations–the DLAMOs discussed in the preceding chapter. As such, it must be the case that between times 5.25 h and 10.75 h, the trajectory must be on the yellow-side of the greenish-yellow region in panel (a.ii) separating depolarization block from other firing regimes. Reaching that point in (g_{K-leak}, g_{KCa}) parameter space requires moving towards the bottom left corner of the parameter space, which amounts to decreasing g_{K-leak} and g_{KCa}, as shown in panel

(a.iii).

Further, as we near the lights off time (12 h), average voltages drop (panel (b.ii) of the figure). We see from panel (a.ii) that there are two ways in which voltages can drop: either only increasing g_{K-leak}, or moving towards the top left corner of the parameter space by simultaneously increasing g_{KCa} and decreasing g_{K-leak}. In the latter way, the trajectory would have to cross the region where there is elevated action potential firing. However, this is not the case, because around time 10 h, the firing activity across the network transitions from mainly DLAMOs, to very low amplitude firing, around 1 Hz. This requires the trajectory to move away from the DB-DLAMOs boundary, and towards the AP-rest boundary. This is achieved by slightly decreasing g_{KCa} and significantly increasing g_{K-leak}, as shown in panel (a.iii) of the figure. Moreover, because as the time of lights off (12 h) approaches, neurons across the SCN exhibit a very wide range of membrane voltages, ranging from around -35 to -65 mV, the trajectory must be in a region of (g_{K-leak}, g_{KCa})-space which produces high standard deviations in membrane voltages, yet produces some AP firing. Such a region lies only near the point at which the AP firing regime, DB regime, and rest regime meet. Thus, during this time g_{K-leak} must increase and g_{KCa} must slightly decrease, so as to move towards such a point.

In the times immediately preceding time 12 h, though, we see that firing-rates reach their minimum, which requires decreasing g_{KCa}, increasing g_{K-leak}, or both, so as to move away from the AP boundary in panel (a.i). In particular, because membrane voltages are starting to drop, it must be the case that g_{K-leak} increases so as to move to the right in panel (a.ii). However, V_m measurements extracted from [BDFP09] continue to display a large spread. This spread in experimentally measured V_m occurs because while many neurons are at rest, others continue to fire action potentials with elevated membrane potentials, but none are in depolarization block. Thus, while g_{K-leak} increases, (g_{KCa}, g_{K-leak}) remains near the AP-rest boundary. Voltages are relatively high still, hence to reproduce the data, while trajectories should move away from the AP-rest boundary, they should still stay relatively close to the portion of the AP-rest boundary corresponding to lower values of $\log(g_{KCa})$ in panels (a.i)-(a.ii).

From times 13 - 15.25 h, however, neurons are presumably largely silent [BDFP09]. Moreover, membrane voltages are quite low, and have a significant spread. The only portion of (g_{K-leak}, g_{KCa})-space at which this happens, though, is near the top of the pictured parameter space, attained by rapidly increasing g_{KCa} and slightly decreasing g_{K-leak}. This portion of the trajectory is represented by the spike in the red curve in panel (a.iii) of the figure.

From times 15.25 h to the start of the next day, on the other hand, we see that the neurons are firing action potentials at a low-rate, indicating that the trajectory must be near the boundary

between action potential firing and rest. However, membrane voltages approach -40 mV, hence the trajectory must move back towards the greenish-blue portion of figure (a.ii). Hence, g_{KCa} must rapidly decrease, and g_{K-leak} must slowly decrease (panel (a.iii)) so as to move towards the boundary between firing and action potential spiking (panel (a.i)). This would result in slowly increasing firing-rates, and a gradual approach to the (g_{K-leak}, g_{KCa}) values associated with time 0h.

This daily trajectory loop in (g_{K-leak}, g_{KCa}) space replicates well the daily variations in firing rates and average voltages reported in [BDFP09] (panels (b.i) and (b.ii) in Figure IV.13). However, it results in less accurate daily fluctuations in standard deviations (panels (c.i) and (c.ii)). In particular, consider σ_f between roughly times 5 and 10 h. For this interval, firing-rate data was not included in [BDFP09]. Hence, the extrapolated red line corresponding to these times in panel (c.i) may not be representative of the true data. Indeed, Belle et al. in [BDFP09] state that in this time interval there actually is a wide spread of cellular firing rates ranging between 0 and 7 Hz. This, in fact, agrees with our prediction that firing-rate standard deviations should be largest during such times.

Additionally, our simulated firing rate standard deviations are in general far higher than is reported in the data. The cause of this discrepancy lies at least in part with the firing behavior of the SCN neuron model. Namely, the model produces low firing rates around 1 or 2 Hz only in a very small range of (g_{K-leak}, g_{KCa}) space along the boundary of the firing regime. Consequently, it is very difficult, and likely biologically unrealistic, to tune g_{K-leak} and $g_{KCa})$ so that the model neuron reproduces firing rates of 1 - 2 Hz observed between times 10 and 20 by [BDFP09]. Instead, to simulate mean firing rates around 1 - 2 Hz, the predicted trajectory passes through (g_{K-leak}, g_{KCa}) space near the boundary of firing activity which would represent a situation where most neurons are either at rest or firing at rates greater than 2 Hz. Since firing rate standard deviations are necessarily high along this boundary, our simulated results produces higher standard deviations of firing-rates than does the data.

Also complicating our simulated results is the presence of bistability in the SCN neuron model. Namely, for a range of (g_{K-leak}, g_{KCa}) space extending from the depolarization block-rest boundary in panels (a.i) and (a.ii) of Figure IV.13, towards the bottom right of the panels, the corresponding SCN neuron can be in one of either two states. In particular, the neuron is either in depolarization block, with membrane voltage around -30 mV, or it can be at rest, with membrane voltage around -60 mV. This region of bistability has the potential to create large standard deviations in membrane voltage values. However, because data from experiments in [BDFP09] indicate that transitions out of depolarization block occur by producing DLAMOs, and then producing action potentials, actual neuron transitions in firing states seem to avoid the boundary between depolarization block and rest.

In summary, despite some limitations in the computed trajectory that we predict SCN neurons to take in (g_{K-leak}, g_{KCa}) space across the day, we believe we have made significant progress towards finding a good representation of it. Implementing this trajectory into our SCN firing rate model requires a parameterization of the trajectory loop with respect to circadian phase θ. As a result, implementing this trajectory into our firing-rate model would also require re-computing steady state firing rates of the SCN neuron model for ranges of values of synaptic conductance G, circadian phase θ and GABA reversal potential E_{GABA}, namely the function $F(G, R, E_{GABA})$ from Figure III.4 top row, (as well as the maximum synaptic signaling function $Y(G, R, E_{GABA})$ from Figure III.4, bottom row). With these modifications to our SCN firing-rate model, we can obtain a more realistic understanding of the daily variation in SCN firing-rates and thus be one step closer to understanding how the SCN modulates its output as the environmental light-dark cycle changes.

4.4 Discussion

We have made significant progress towards modeling the pathway by which the SCN modulates its firing output in response to changes in the light cycle. In particular, we have proposed using the Hannay model [HBF19a] to predict how changes in the environmental light-dark cycle affect the statistics of the intracellular molecular circadian clock which keeps track of time within SCN neurons. We have then elucidated the connection between the phase of the molecular clock and the electrophysiology of SCN neurons. Finally, we demonstrated how the statistics of circadian phases can be accounted for in computing statistics of SCN firing rates using our SCN firing-rate model.

Our preliminary results using the Colclasure-Diniz Behn function $R(\theta)$ demonstrate that while phase delays can disrupt the signals output by the SCN, phase advances disrupt them much more strongly. In particular, not only do phase advances reduce the amplitude of the firing-rate output of the SCN, as would accompany the reduced amplitude in behavioral rhythms, but phase advances also increase the firing activity of SCN neurons when they would normally be quiescent. As such, our model identifies a mechanism by which large phase advances, as would occur with air travel across many time zones to the east, could contribute to the circadian disruptions associated with jet lag.

We have further clarified that it is important to describe the signal output by the SCN in terms of both the mean and standard deviation of the firing rate. In particular, we show using information theoretic techniques that most of the information described by the mean and standard deviation of the firing-rates of neurons across the SCN network is not captured by considering the mean alone. As such, our model of the dependence of SCN firing rate on circadian phase adds considerable information about the ways in which the environmental light cycle affects the signal output by the

SCN.

Moreover, in putting our model of the pathway together we, have made significant progress towards elucidating the effect of the phases of the molecular clocks contained in PER1 expressing SCN neurons on the firing state of SCN neurons. We have leveraged the SCN single neuron model of Diekman et al. in [DBI+13] to fit daily rhythms in the conductances of the potassium leak current and calcium-dependent potassium current that successfully reproduce the daily rhythms in both the mean firing rates and membrane voltages reported experimentally for SCN neurons [BDFP09].

4.4.1 Limitations

Our model for light's effects on SCN firing rates has several limitations that ought to be addressed. To start, it is likely not reasonable to use the Hannay model [HBF19a] "as is" to describe the statistics of the phases of the intracellular molecular TTFL rhythm across PER1-expressing SCN neurons. Indeed, the Hannay model, rather than being fit to data regarding molecular clock phases, is fit to phase response curves of downstream circadian markers including core body temperature and dim light melatonin onset (DLMO). These phase response curves are likely a reflection of multiple physiological processes that would be influenced by SCN firing rates. As a result, it is likely that the parameters of the Hannay model are tuned to describe phenomena other than the intracellular TTFL phases within SCN neurons.

To overcome such a limitation, we could tune the parameters of the Hannay model differently. For example, we could tune the parameters of the Hannay model so that when it is used in the model of the full pathway, the phase shifts obtained in SCN firing rates reproduce the circadian marker phase response curves. Alternatively, we could replace the Hannay model with a model of the clock TTFLs themselves, such as [LA21].

Another issue with using the Hannay model "as is" is that it is tuned to human data. Our SCN firing rate model, on the other hand, relies on the Diekman SCN neuron model that is tuned to rodent data. This issue can be addressed as implementations of the Hannay model fit to the appropriate rodent phase response curves exist, as per a personal communication with Kevin Hannay.

Future work, as described in the next chapter, will address these limitations as well as complete the implementation of our improved mapping of molecular clock phases to SCN neuron firing activity into our SCN firing rate model.

CHAPTER V

Summary and Discussion of Future Work

In this chapter, we summarize the work completed on all projects, reiterate the significance of the projects, and detail future directions regarding pain processing in the spinal cord.

The work shown in this dissertation advances the literature of firing-rate models. Firing-rate models are significant because they reduce the complicated behavior of individual neurons and their interactions across a whole network of neurons. In particular, they summarize the network's behavior in terms of firing-rates: the number of voltage spikes per second.

Chapter I sets the stage. To start, we provide the necessary neuroscience prerequisites for understanding the paper. We then review the firing-rate model literature, and augment our review with an analysis of the citation network of the literature and an identification of key traits of a selection of models across the literature. The review clarifies where our firing-rate model fits in the literature. In particular, the firing-rate model used in both Chapters II and III hearken back to the Jansen-Rit model, which in turn derives from the da Silva et al. model, which itself derives from the Wilson-Cowan model. The review also clarifies that the models we use belong to the class of models which derive from the Wilson-Cowan approach rather than the population density approach or the coupled oscillator approach.

Chapter II uses a system of Jansen and Rit type firing-rate models to understand pain processing in the spinal cord. Whereas the firing-rate models used are not novel, the analysis of them is. Our analysis begins by mapping experimentally constrained behavior of feed-forward networks of firing-rate models to the corresponding set of coupling strengths. We then identify the two regions in the space of coupling strengths that reproduce "healthy" behaviors and "allodynia" behaviors, respectively. We characterize the sensitivity of healthy circuits towards producing allodynia, by solving a control problem: how can we induce a healthy circuit to produce allodynia while changing coupling strengths as little as possible. The answer is to take the shortest path in coupling strength space between the healthy circuit space and the allodynia space. The result then characterizes in which ways a circuit is most vulnerable to producing allodynia. By clustering circuits according

to those shortest paths, we can identify trends in how the circuit can be dysregulated. Because we are using firing-rate models that represent whole populations of e.g. excitatory neurons, those trends represent alterations in the balance between excitation and inhibition in the network. As a result, our analysis identifies the specific disruptions in excitatory-inhibitory balance that most effectively produce allodynia. Thus, the use of firing-rate models enables us to uncover mesoscopic mechanisms of E-I balance.

Whereas Chapter II introduces a novel analysis of firing-rate models, Chapter III introduces a novel firing-rate model formalism. The firing-rate formalism adopts the approach of [ZVvPTH14], a firing-rate model descending from the Jansen - Rit model, wherein firing rates are averaged against a gaussian distribution of underlying currents to obtain both the mean and standard deviation of firing rates across the network. However, our model formalism captures the effects of heterogeneity in the electrophysiological properties across neurons on the network firing rate, well beyond only the heterogeneity of synaptic currents. In particular, recall that our mean-field firing-rate model for the SCN network makes predictions about the effects on population firing statistics of the unique properties of the SCN neural network, such as circadian phase effects on firing state, distributions of $GABA$ reversal potentials, and modulation of synaptic signaling due to atypical neuron firing states. In doing so, our model clarifies how these properties can lead to circadian disruptions, such as the blunted circadian rhythm amplitude associated with aging and neurophysiological diseases. Moreover, because our model describes the statistics specifically of the firing rate across the SCN, it is well-suited for modeling how the SCN coordinates "downstream" circadian processes, such as the sleep-wake cycle.

Chapter IV extends the results found in Chapter III so as to model the pathway by which environmental light affects the firing rate outputs of the SCN. Namely, we couple our firing-rate model for the SCN from Chapter III to a coupled-oscillator based mean-field reduction describing the phases of intracellular molecular clocks across the SCN. Thus, the model of our pathway is essentially two mean-field models coupled together. The way in which the two mean-field models is coupled, though, is a difficult question. We answer it by fitting experimental data to find a relationship between the phases of intracellular clocks to key potassium conductances that control the electrophysiological state of the neuron.

5.1 Future Directions

The analysis of Chapter IV, however, is not yet complete. In particular, while we have found a link between the phases of intracellular molecular clocks and the electrophysiological state of the SCN, we have not yet used that link to model the full pathway. To implement the relationship,

we need to recompute SCN neuron firing rates as a function of model parameters, as described in section 4.3.

Moreover, the relationship between intracellular molecular clock phase and the electrophysiological state of a neuron was determined only for per1-expressing neurons in the SCN. However, many neurons in the SCN that also contain molecular clocks do not express per1. Thus, it may be fruitful to add a third mean-field reduction to the model of the pathway. In particular, whereas the second mean-field model in the pathway would describe the firing-rate statistics output the by per1-expressing neurons in the SCN as described via Section 4.3, the third model would describe the resulting statistics of the firing-rates output by non-per1-expressing neurons in the SCN. Another approach would be to simply use the data from [BDFP09] on non-per1-expressing neurons to describe the signal output by the SCN, or to lump the two regions together.

Future improvements to our model of the pathway by which environmental light affects SCN output include resolving the issues with using the Hannay model "as is" to describe the responses of the phases of intracellular molecular clocks to environmental light, as discussed in Section 4.4.1.

After resolving the issues with using the Hannay model "as is" and the other issues discussed above, we will have completed the model of the pathway. We can then use the resulting firing rates output by the SCN in models of sleep-wake regulation. Namely, leading models of sleep-wake regulation, such as the Phillips-Robinson model [PR07], or the model of [APDBB22], describe mean SCN activity heuristically. By using the outputs of our model pathway, not only would the sleep-wake models be more realistic, but they would incorporate important information about the SCN relayed only by the standard deviation in SCN firing-rates. As a result, those models would be well-equipped to describe how environmental light affects the sleep-wake cycle. We could then use our model to better describe how purposefully altering environmental light could improve sleep in shift workers and those experiencing jet lag, and better understand the effects of e.g. blue light from electronic devices on sleep.

The future directions of our research are not limited to applications to circadian rhythms only. In particular, we also hope to be able to use our analysis of firing-rate models representing pain-processing circuits in the spinal cord to treat chronic pain. While allodynia, as we investigate in this work, is a widespread and for many people, chronic, condition, work is underway to alleviate allodynia. Namely, spinal cord stimulation (SCS) has been shown to alleviate allodynia, although the mechanisms by which such alleviation happens are poorly understood. Further, spinal cord stimulation only works for some patients. We speculate that the success of SCS depends on how the parameters underlying spinal cord circuit motifs have changed from a normal range (e.g. in the allowable parameter space, as discussed earlier) to a pathological range (e.g. in moving above

the allodynia surface). Indeed, in this work, we have established that dynamic allodynia and static allodynia may be produced by altering E-I balance in a variety of ways, and have identified the most likely allodynia-producing alterations. If we can understand how spinal cord stimulation interacts with such alterations, e.g. by stimulating axon sprouting between $A\beta$-fibers and inhibitory interneurons, or by down-regulating the impact of $A\beta$ signaling on excitatory interneurons, we will have made significant progress towards understanding how SCS can alleviate allodynia, and moreover how its success can vary from patient to patient. Our experimental collaborators at the University of Michigan are engaged in such an effort to understand how SCS affects $A\beta$ signaling and other aspects of pain-processing circuitry in the spinal cord. We hope that we will be able to pair together our sensitivity analysis of circuit motifs underlying the spinal cord pain processing circuitry with our collaborator's analysis of how SCS affects signaling related to pain-processing circuitry, to much better understand the mechanisms underlying the alleviation of chronic pain via SCS.

APPENDIX A

Supplementary Figures for "A New Parameter Sensitivity Analysis Methodology Applied to Neural Circuits in the Spinal Cord Dorsal Horn" (Chapter II)

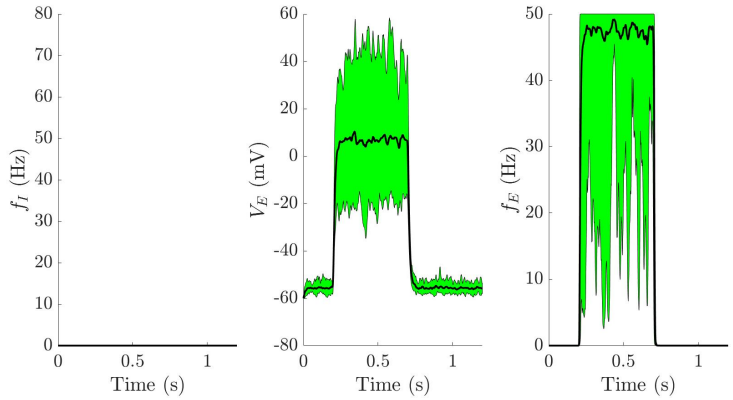


Figure A.1: Range of voltages and firing-rates for populations in the simple circuit under inhibitory ablation conditions. (Leftmost panel): Because the inhibitory ablation population is ablated, we represent its firing-rate as 0, in our model. (Middle panel): Range of voltages and (rightmost panel) firing-rates of the excitatory population. We see that under inhibitory ablation, the voltage of the excitatory population increases sufficiently that it induces the firing-rate of the excitatory population to increase. As a result, the circuit relays pain despite the input signals being innocuous ($f_{A\beta} \in [10, 20]$), thereby producing allodynia.

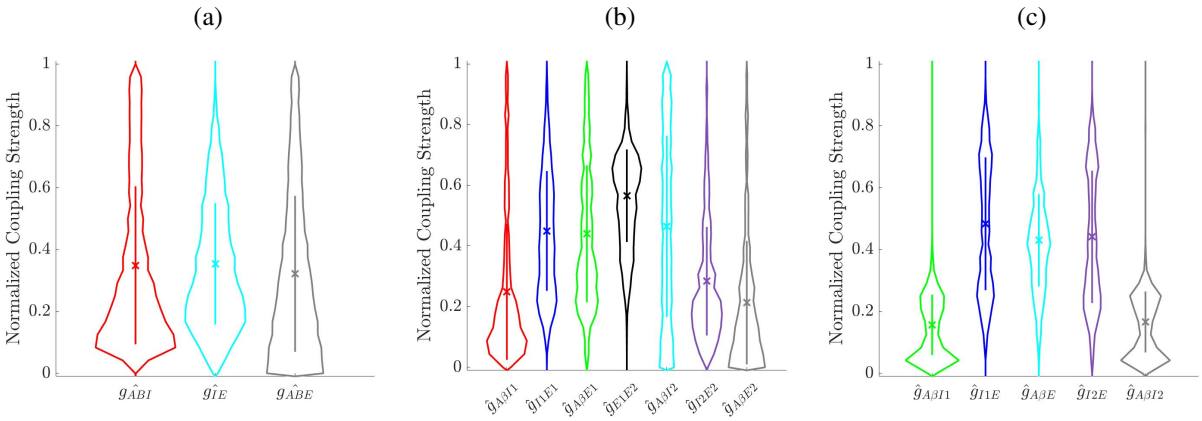


Figure A.2: Violin plots of the sampled sets of normalized coupling strengths for **(a)** the simple gate control circuit, **(b)** the proposed circuit mediating dynamic allodynia, and **(c)** the proposed circuit mediating static allodynia.

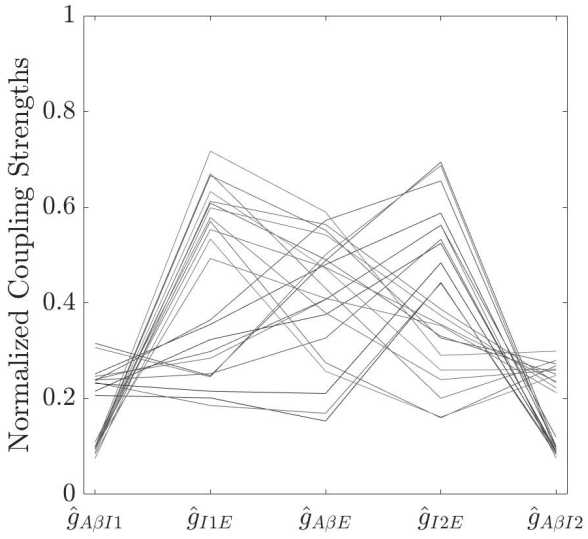


Figure A.3: Parallel plot representation of sampled sets of coupling strength for the cluster 4 for the proposed circuit mediating static allodynia. We see that cluster 4 consists of at least two spatially disconnected regions: one in which $\hat{g}_{A\beta I1}$ is low (less than about 0.2) and $g_{A\beta I2}$ is high (greater than about 0.2), and the other where conversely $\hat{g}_{A\beta I1}$ is high and $\hat{g}_{A\beta I2}$ is low. Nevertheless, the corresponding circuits all are most vulnerable towards producing allodynia in almost exactly the same ways.

APPENDIX B

Appendix to "A Mean-Field Firing-Rate Model for the Suprachiasmatic Nucleus and Other Heterogeneous Networks of Neurons" (Chapter III)

2.1 Parameters for the application to the SCN

In Table B.1 below, we summarize the abbreviations, default values, valid range, and default units for parameters used to apply the model specifically to the SCN. Note that some parameters have parameter ranges outside of which our model does not function. Namely, R_{ampl}, $\overline{E_{GABA}}$, and σ_{EGABA} are restricted by the values of R and E_{GABA} over which we have calculated $Y(G, R, E_{GABA})$ and $F(G, R, E_{GABA})$. On the other hand, σ_θ, and g_0 make the model impractically slow or subject to noise if they exceed the parameter range.

2.2 Numerical simulations

All simulations were conducted using Matlab, and code can be made available upon request.

The convolutions used to find the statistics for synaptic conductance, g, rewritten as differential equations, were solved using an Euler-step algorithm. Notably, switching to more stable algorithms such as Runge-Kutta in no cases significantly improved the smoothness of solutions. Further, at times the step-size for the Euler-step algorithm had to be reduced to increase the smoothness of solutions.

The data underlying the F-G-R-E_{GABA}, Y-G-R-E_{GABA}, and YF-G-R-E_{GABA} surfaces was generated by solving the Hodgkin-Huxley type differential equations of Diekman [DBI$^+$13] over about 2 seconds for roughly 6,000,000 combinations of $G, R,$ and E_{GABA} values. To do so we took advantage Matlab's parallel computing capabilities by running the code on the Great Lakes Computing Cluster using 12 cores over the course of about 20 hours.

To extrapolate from this data to the continuous functions $F(G, R, E_{GABA})$ and $Y(G, R, E_{GABA})$, we linearly interpolated piece-wise between data points. However, to reduce numerical noise introduced by integrating over piece-wise linear interpolations of the original data

in simulations for model results, polynomial fits of $F(G, R, E_{GABA})$ and $YF(G, R, E_{GABA})$ data, as well as of the corresponding boundaries between firing regimes, were computed for each firing regime present when $E_{GABA} = -55$ mV. In simulations of our firing-rate model where $E_{GABA} = -55$ mV for all neurons in the network, these fits were used instead of the linear interpolations of the data . In particular, the polynomial fits can be seen in the figures displaying model simulations where $\sigma_{EGABA} = 0$ and $\overline{E_{GABA}} = -55$ mV (Figures III.5, III.6, III.7, and the second column of Figure B.2). Fits for each firing regime seen in $F(G, R, -55)$ and $YF(G, R, -55)$ as well as for the corresponding boundaries were made using Matlab's "fit" command, using a mixture of built-in and custom fit-types.

Parameter	Abbreviation	Default	Range	Units
# of Neurons in Simulation	N	10,000[a]	-	# Neurons
Average # Synapses into a Neuron	$\overline{N_{syn}}$	1100[b]	-	# Synapses
Variance in # of Synapses into Neurons	$Var(N_{syn})$	979	0 - $(N - \overline{N_{syn}})\overline{N_{syn}}$	# Synapses
Circadian Proxy	R	$R_{ampl}\sin(\theta(t))$	-8.5 - 8.5	N/A
Circadian Proxy Amplitude	R_{ampl}	5[c]	0 - 8.5	N/A
Mean Circadian Phase	$\overline{\theta}$	Time-dependent	0 - 2π	Radians
Standard Deviation in Circadian Phase	σ_θ	0.1	0 - ≈ 3	Radians
Mean GABA Reversal Potential	$\overline{E_{GABA}}$	-55[b]	-110 - 0	mV
Standard Deviation in E_{GABA}	σ_{EGABA}	7[b]	0 - 11[e]	mV
Maximum Synaptic Conductance	g_0	$5 \cdot 10^{-4}$	0 - $\approx 10^{-2}$	nS/cm^2
Synaptic Time Constant	τ	34[d]	-	ms

Table B.1: Summary of Parameters for the Application to the SCN. In the table, [a], [b], [c], [d] indicate that respective values are due to [AMGF+16], [MB89], [DMB+15], and [ZVvPTH14]. [e] indicates that in particular, we need that $\sigma_{EGABA} \leq \frac{1}{5}\max\left(|\overline{E_{GABA}} + 110)|, |\overline{E_{GABA}}|\right)$, or else when integrating over the E_{GABA} distribution when calculating expectations of firing rates, we exceed the valid parameter range. The maximum value of σ_{EGABA} therefore is 11 mV, occurring when $\overline{E_{GABA}} = -55$ mV.

Finally, to calculate the integrals for standard deviations and means of firing rates f and yf, we used trapezoidal integration modified to take into account the sparsity of and to interpolate over the F-G-R-E_{GABA} and YF-G-R-E_{GABA} surfaces. To further increase the speed of our algorithm, we only integrated within five standard deviations of the center of the distributions for G, R, and E_{GABA}, as the contributions of the integrand from other values of G, R, and E_{GABA} were negligibly small in our simulations.

2.3 Characteristic function for total synaptic conductance G

The characteristic function for total synaptic conductance G may be used to specify the shape of the G distribution, and is a function of the statistics for N_{syn} and g. In particular, if we strengthen our assumptions that $g_{j,i}(t)$ for $j = 1, \dots N_{syn,i}$ are pairwise uncorrelated with one another and are independent of $N_{syn,i}$ by instead assuming that $g_{j,i}(t)$for $j = 1, \dots N_{syn,i}$ and $N_{syn,i}$ form a mutually independent collection of random variables, we have that the characteristic function is

$$\begin{aligned}
\phi_G(x) &= \mathbb{E}\left[e^{ixG(t)}\right] \\
&= \mathbb{E}\left[e^{ix\sum_{j=1}^{N_{syn_i}} g_{j,i}}\right] \\
&= \mathbb{E}\left[\Pi_{j=1}^{N_{syn_i}} e^{ixg_{j,i}}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\Pi_{j=1}^{N_{syn_i}} e^{ixg_{j,i}} | N_{syn,i}\right]\right] \\
&= \mathbb{E}\left[\Pi_{j=1}^{N_{syn_i}} \mathbb{E}\left[e^{ixg_{j,i}} | N_{syn,i}\right]\right] \\
&= \mathbb{E}\left[\Pi_{j=1}^{N_{syn_i}} \mathbb{E}\left[e^{ixg}\right]\right] \\
&= \mathbb{E}\left[(\phi_g(x))^{N_{syn}}\right]
\end{aligned}$$

and hence

$$\phi_G = \mathcal{G}_{N_{syn}}(\phi_g)$$

where $\mathcal{G}_{N_{syn}}$, the probability generating function of N_{syn}, is determined exclusively by the degree distribution of the network.

2.4 Single-cell SCN neuron model

Diekman's single-cell SCN neuron model [DBI+13] states that the membrane voltage V of an SCN neuron is modeled by

$$C\frac{dV}{dt} = I_{app} - I_{Na} - I_K - I_{CaL} - I_{CaNonL} - I_{KCa}(R) - I_{K-leak}(R) - I_{Na-leak},$$

where $C = 5.7$ pF, voltage is in mV, and membrane currents I_x in pA are given in Table B.2.

Current (pA)	Formula	Max Conductance (nS)	Reversal potential (mV)
I_{Na}	$g_{Na}m^3(t)h(t)\cdot(V(t)-E_{Na})$	$g_{Na}=229$	$E_{Na}=45$
I_K	$g_K n^4(t)\cdot(V(t)-E_K)$	$g_K=3$	$E_K=-97$
I_{CaL}	$g_{CaL}r_L(t)f_L(Ca_s(t))\cdot(V(t)-E_{Ca})$	$g_{CaL}=6$	$E_{Ca}=54$
I_{CaNonL}	$g_{CaNonL}r_{NonL}(t)f_{NonL}(t)\cdot(V(t)-E_{Ca})$	$g_{CaNonL}=20$	$E_{Ca}=54$
$I_{KCa}(R)$	$g_{KCa}(R)s^2(t)\cdot(V(t)-E_K)$	$g_{KCa}=\frac{198}{1+\exp(R)}+2$	$E_K=-97$
$I_{K-leak}(R)$	$g_{K-leak}(R)\cdot(V(t)-E_K)$	$g_{K-leak}=\frac{0.2}{1+\exp(R)}$	$E_K=-97$
$I_{Na-leak}$	$g_{Na-leak}\cdot(V(t)-E_{Na})$	$g_{Na-leak}=0.0576$	$E_{Na}=45$

Table B.2: Here we present the formulas (2^{nd} column) for ionic currents (1^{st}) column and corresponding constants (3^{rd} and 4^{th} columns) as appearing in the single neuron model of [DBI+13].

The gating variables $q=m,h,n,r_L,r_{NonL},f_{NonL}$ follow

$$\frac{dq}{dt}=\frac{q_\infty(V(t))-q(t)}{\tau_q(V(t))},$$

the gating variable s follows

$$\frac{ds}{dt}=\frac{s_\infty(Ca_s(t))-s(t)}{\tau_q(Ca_s(t))},$$

f_L follows

$$f_L=\frac{K_1}{K_2+Ca_s(t)},$$

where $K_1=3.93E-5$mM, $K_2=6.55E-4$mM, and Ca_s follows

$$\frac{dCa_s}{dt}=-k_s(I_{CaL}(t)+I_{CaNonL}(t))-Ca_s/\tau_s+b_s,$$

where $k_s=1.65E-4$ mM/fC and $b_s=5.425E-4$ mM/ms. Steady state activation functions q_∞ and time constants τ_q for the gating variables are given in Table B.3:

Current Type (pA)	Gating Variable q	q_∞	τ_q
I_{Na}	m	$\frac{1}{1+\exp(-(V(t)+35.2)/8.1)}$	$\exp(-(V(t)+286)/160)$
I_{Na}	h	$\frac{1}{1+\exp((V(t)+62)/2)}$	$0.51+\exp(-(V(t)+26.6)/7.1)$
I_K	n	$\frac{1}{[1+\exp(-(V(t)-14)/17)]^{0.25}}$	$\exp(-(V(t)-67)/68)$
I_{CaL}	r_L	$\frac{1}{1+\exp(-(V(t)+36)/5.1)}$	3.1
I_{CaNonL}	r_{NonL}	$\frac{1}{1+\exp(-(V(t)+21.6)/6.7)}$	3.1
I_{CaNonL}	f_{NonL}	$\frac{1}{1+\exp((V(t)+260)/65)}$	$\exp(-(V(t)-444)/220)$
$I_{KCa}(R)$	s	$\frac{10^7(Ca_s)^2}{10^7(Ca_s)^2+5.6}$	$\frac{500}{10^7(Ca_s)^2+5.6}$

Table B.3: Here we present the gating variables (2nd column) involved in the dynamics of each type of ionic current (1st column). The 3rd and 4th columns roughly describe the gain functions and the speeds at which gating variables change, for each gating variable.

2.5 Verification of differential equations

We show that the convolutions from Equations III.14 and III.15,

$$\overline{g}(t) = \overline{H \star yf} = (H \star \overline{yf})(t) = \int_0^t H(t-s)\overline{yf}(s)ds$$
$$\sigma_g(t) = (H \star \sigma_{yf})(t) = \int_0^t H(t-s)\sigma_{yf}(s)ds,$$

may be written as the differential equations from Equations III.16 and III.17,

$$\overline{g}'' = -2\tau^{-1}\overline{g}' - \tau^{-2}\overline{g} + e\tau^{-1}g_0\overline{yf}(t)$$
$$\sigma_g'' = -2\tau^{-1}\sigma_g' - \tau^{-2}\sigma_g + e\tau^{-1}g_0\sigma_{yf}(t).$$

To verify these equations, observe that

$$\begin{aligned}
\frac{d}{dt}\left[\int_0^t H(t-s)\overline{yf}(s)ds\right] &= \frac{d}{dt}\left[\int_0^t e\cdot g_0\frac{t-s}{\tau}e^{-(t-s)/\tau}\overline{yf}(s)ds\right] \\
&= \frac{g_0 e}{\tau}\frac{d}{dt}\left[e^{-t/\tau}t\int_0^t e^{s/\tau}\overline{yf}(s)ds - e^{-t/\tau}\int_0^t se^{s/\tau}\overline{yf}(s)ds\right] \\
&= \frac{g_0 e}{\tau}\left[e^{-t/\tau}\left(1-\frac{t}{\tau}\right)\int_0^t e^{s/\tau}\overline{yf}(s)ds + t\overline{yf}(t) - \left(-e^{-t/\tau}\int_0^t \frac{s}{\tau}e^{s/\tau}\overline{yf}(s)ds + t\overline{yf}(t)\right)\right] \\
&= \frac{g_0 e}{\tau}\left[e^{-t/\tau}\left(1-\frac{t}{\tau}\right)\int_0^t e^{s/\tau}\overline{yf}(s)ds + \frac{1}{\tau}e^{-t/\tau}\int_0^t se^{s/\tau}\overline{yf}(s)ds\right] \\
&= e^{-t/\tau}\frac{g_0 e}{\tau}\left[\left(1-\frac{t}{\tau}\right)\int_0^t e^{s/\tau}\overline{yf}(s)ds + \frac{1}{\tau}\int_0^t se^{s/\tau}\overline{yf}(s)ds\right] \\
&= \frac{g_0 e}{\tau}\int_0^t\left(1-\frac{t}{\tau}+\frac{s}{\tau}\right)e^{-(t-s)/\tau}\overline{yf}(s)ds
\end{aligned}$$

and

$$\begin{aligned}
&\frac{d^2}{dt^2}\left[\int_0^t H(t-s)\overline{yf}(s)ds\right] \\
&= \frac{g_0 e}{\tau}\left(\frac{e^{-\frac{t}{\tau}}}{-\tau}\left[\left(1-\frac{t}{\tau}\right)\int_0^t e^{\frac{s}{\tau}}\overline{yf}(s)ds + \frac{1}{\tau}\int_0^t se^{\frac{s}{\tau}}\overline{yf}(s)ds\right] + e^{-\frac{t}{\tau}}\left[\frac{d}{dt}\left(\left(1-\frac{t}{\tau}\right)\int_0^t e^{\frac{s}{\tau}}\overline{yf}(s)ds\right) + \frac{t}{\tau}e^{\frac{t}{\tau}}\overline{yf}(t)\right]\right) \\
&= e^{-\frac{t}{\tau}}\frac{g_0 e}{\tau}\left[\left(\frac{t}{\tau^2}-\frac{1}{\tau}\right)\int_0^t e^{\frac{s}{\tau}}\overline{yf}(s)ds - \frac{1}{\tau^2}\int_0^t se^{\frac{s}{\tau}}\overline{yf}(s)ds - \frac{1}{\tau}\int_0^t e^{\frac{s}{\tau}}\overline{yf}(s)ds + \left(1-\frac{t}{\tau}\right)e^{\frac{t}{\tau}}\overline{yf}(t) + \frac{t}{\tau}e^{\frac{t}{\tau}}\overline{yf}(t)\right] \\
&= e^{-t/\tau}\frac{g_0 e}{\tau}\left[\left(\frac{t}{\tau^2}-\frac{2}{\tau}\right)\int_0^t e^{s/\tau}\overline{yf}(s)ds - \frac{1}{\tau^2}\int_0^t se^{s/\tau}\overline{yf}(s)ds + e^{t/\tau}\overline{yf}(t)\right] \\
&= e^{-t/\tau}\frac{g_0 e}{\tau}\left[\left(\frac{t}{\tau^2}-\frac{2}{\tau}\right)\int_0^t e^{s/\tau}\overline{yf}(s)ds - \frac{1}{\tau^2}\int_0^t se^{s/\tau}\overline{yf}(s)ds\right] + \frac{g_0 e}{\tau}\overline{yf}(t) \\
&= e^{-t/\tau}\frac{g_0 e}{\tau}\left[\left(\frac{t}{\tau^2}-\frac{2}{\tau}\right)\int_0^t e^{s/\tau}\overline{yf}(s)ds - \frac{1}{\tau^2}\int_0^t se^{s/\tau}\overline{yf}(s)ds\right] + \frac{g_0 e}{\tau}\overline{yf}(t) \\
&= -e^{-t/\tau}\frac{g_0 e}{\tau^2}\left[\left(2-\frac{t}{\tau}\right)\int_0^t e^{s/\tau}\overline{yf}(s)ds + \frac{1}{\tau}\int_0^t se^{s/\tau}\overline{yf}(s)ds\right] + \frac{g_0 e}{\tau}\overline{yf}(t) \\
&= -\frac{g_0 e}{\tau^2}\int_0^t\left(2-\frac{t}{\tau}+\frac{s}{\tau}\right)e^{-(t-s)/\tau}\overline{yf}(s)ds + \frac{g_0 e}{\tau}\overline{yf}(t).
\end{aligned}$$

and thus

$$\begin{aligned}
&\frac{d^2}{dt^2}\left[\int_0^t H(t-s)\overline{yf}(s)ds\right] + \frac{2}{\tau}\frac{d}{dt}\left[\int_0^t H(t-s)\overline{yf}(s)ds\right] + \frac{1}{\tau^2}\int_0^t H(t-s)\overline{yf}(s)ds - \frac{g_0 e}{\tau}\overline{yf}(t) \\
&\quad = \frac{g_0 e}{\tau^2}\int_0^t\left[-\left(2-\frac{t}{\tau}+\frac{s}{\tau}\right)+2\left(1-\frac{t}{\tau}+\frac{s}{\tau}\right)+\left(\frac{t-s}{\tau}\right)\right]e^{-(t-s)/\tau}\overline{yf}(s)ds \\
&\quad = 0,
\end{aligned}$$

as desired. An analagous argument shows that Equation III.15 may be rewritten as Equation III.17.

2.6 Validating model results against a spiking neuronal network

To validate the predictions of our firing-rate model, we simulated a network of biophysical model SCN neurons. The 100-cell network had connectivity similar to the networks underlying results shown in Figures III.5 - B.1. In particular, synaptic structure was based on a Poisson random graph with 11% connectivity. The dynamics of each neuron was modeled using the Hodgkin-Huxley-type SCN neuron model from Equation III.23, and GABA synaptic gating was modeled with Equation III.25. Total synaptic conductance $G(t)$ for each neuron was computed by linearly summing synaptic gating variables over all presynaptic neurons and multiplying by a maximum synaptic conductance g_0. The value of g_0 was chosen to be $0.5/\overline{N_{syn}}$ nS/cm^2, similar to the value used in the firing-rate model. This resulted in values of $G(t)$ between 0 and roughly 0.5 nS/cm^2. Synaptic current in each postsynaptic cell was modeled by Equation III.24 where $V(t)$ is the membrane potential of the postsynaptic cell.

To validate the ability of the firing-rate model to capture the effects of heterogeneities in the spiking network model, we simulated the spiking network model for two combinations of standard deviations in circadian phase values, σ_θ, and in GABA reversal potential values, σ_{EGABA}. Specifically, we simulated a network with no variance in circadian phases or E_{GABA} ($\sigma_\theta = 0$ and $\sigma_{EGABA} = 0$ mV) and a network where the variances were set to their largest values considered in Figure III.9, ($\sigma_\theta = 0.9$ and $\sigma_{EGABA} = 7$ mV). In both these networks, mean E_{GABA} was set to -55 mV, the amplitude R_{ampl} of the circadian proxy was 5, and in the latter network both E_{GABA} and circadian phase followed Gaussian distributions.

To compare the firing-rate statistics of the firing-rate model to the average activity of the spiking network, we calculated the mean $\overline{f}$ and standard deviations σ_f of the firing rates across the spiking network. To do so, we binned the spiking activity into 0.125 s intervals, counted the number of spikes for each neuron in each bin, and divided the spike counts by the bin width to arrive at time-varying, mean firing rates for each neuron. We then computed the mean and standard deviations of cellular firing rates across all neurons during each bin.

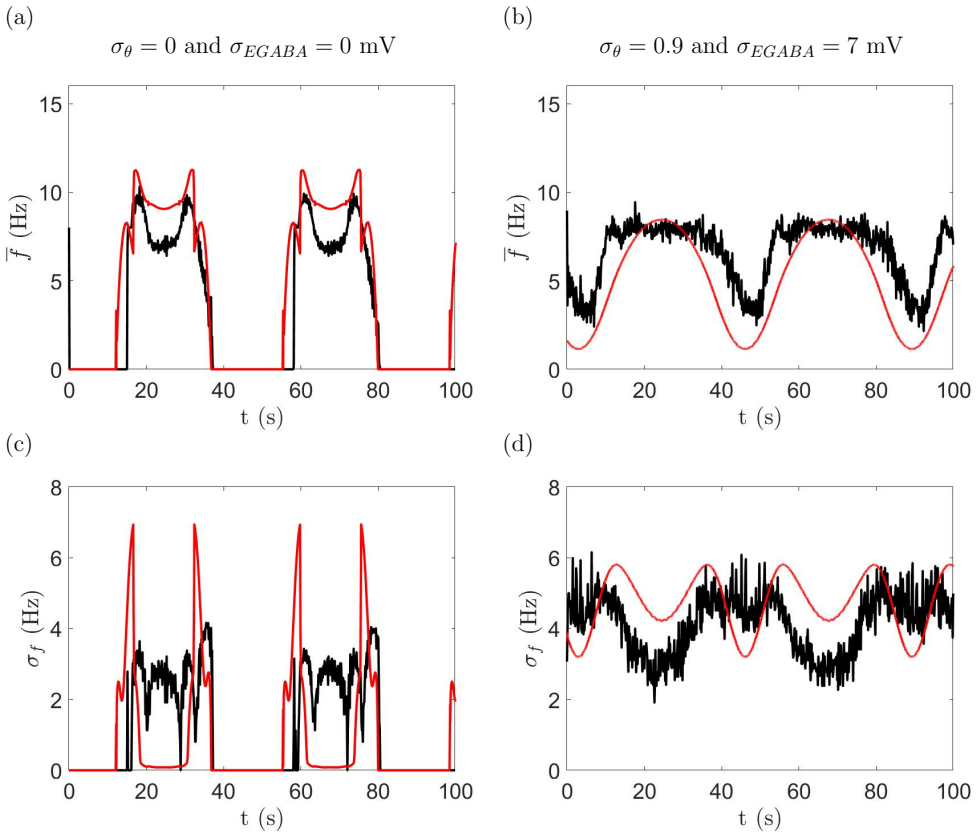


Figure B.1: Comparison of Firing-Rate Model Results to Activity in a Spiking Neuronal Network. Mean firing-rates (top row) and standard deviations (bottom row) for the firing-rate model (red curves) and for a network of 100 SCN model neurons (black curves). **(a), (c)**: no variance in circadian phases or E_{GABA} ($\sigma_\theta = 0$ and $\sigma_{EGABA} = 0$ mV). **(b), (d)**: $\sigma_\theta = 0.9$ and $\sigma_{EGABA} = 7$ mV as in Figure 9. In the firing-rate model, $g_0 = 5 \cdot 10^{-4}$ nS/cm^2, $R_{ampl} = 5$, $\overline{N_{syn}} = 1100$ and $Var(N_{syn}) = 979$.

As shown in Figure B.1, the firing-rate statistics of the spiking network qualitatively reproduce the predictions of our firing-rate model. In particular, the peaks and troughs of $\overline{f}$ and σ_f predicted by the spiking network occur roughly at the same time as those predicted by our firing-rate model. Moreover, each model's predictions of $\overline{f}$ typically differ by only up to a few Hz for both the simulation with $\sigma_\theta = 0$ and $\sigma_{EGABA} = 0$ mV and the simulation with $\sigma_\theta = 0.9$ and $\sigma_{EGABA} = 7$ mV. However, while both model's predictions of σ_f have similar magnitudes, it should be noted that since the spiking network is small and the simulation time is relatively short (lasting 100 seconds), the magnitudes of σ_f predicted by the spiking network increase with decreasing bin width. Consequently, comparisons between σ_f output by the two types of models are best made qualitatively.

2.7 Effect of changing $\overline{E_{GABA}}$

The effects of E_{GABA} on network synaptic conductance and firing rates are shown in Figure B.2 where three different values of E_{GABA} are considered: -80 mV, -55 mV, and -32 mV. To isolate the effects of changing E_{GABA}, we assume there to be no variance in circadian phase $\sigma_\theta = 0$ or in E_{GABA} itself ($\sigma_{EGABA} = 0$). We find that higher values of E_{GABA} lead to higher firing rates, with a "mid-day" dip occurring for the largest R_{ampl} (left column) when maximum synaptic conductance is at a moderate level ($g_0 = 5 \cdot 10^{-4}$ mS/cm^2). For lower R_{ampl}, the mid-day dip is flattened out since network activity doesn't reach the DLAMO region (middle column), and for smallest R_{ampl}, firing rate varies approximately sinusoidally. Firing rate standard deviations (middle row) remain small, except for brief spikes as network activity transitions between AP firing and subthreshold regimes.

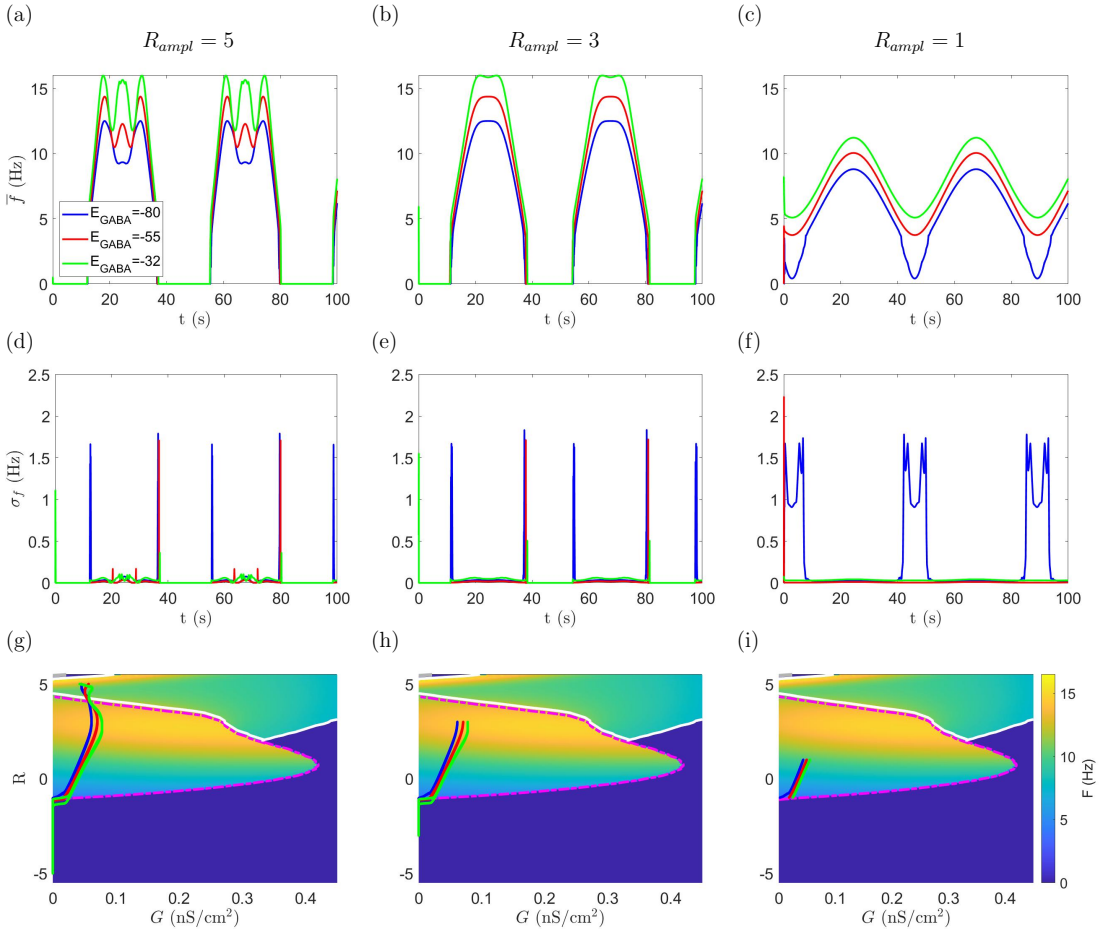


Figure B.2: Effect of E_{GABA} on firing rates when variance in E_{GABA} and circadian phase are not present and synaptic coupling is weak, fixed at $g_0 = 5 \cdot 10^{-5}$ for all neurons. **(a) - (f)**: Mean firing rates $f(t)$ (first row) and standard deviations of firing rates $\sigma_f(t)$ (2nd row) over multiple $R(t)$ cycles (shortened circadian cycles) for 3 different values of GABA reversal potential E_{GABA} (in mV), -55 mV likely being the mean E_{GABA} in the SCN. Three amplitudes of circadian variation are simulated, $R_{ampl} = 5$ (left column), 3 (middle column) and 1 (right column). **(g) - (i)**: Trajectories $(G(t), R(t))$ are plotted on the neuronal firing rate surface $F(G, R, -55)$ for the same values of E_{GABA} as in (a) - (f). Firing rate standard deviations tend to be largest near transitions between firing regimes.

BIBLIOGRAPHY

[AB97] Daniel J Amit and Nicolas Brunel. Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral cortex (New York, NY: 1991)*, 7(3):237–252, 1997.

[ACG+09] Frederico AC Azevedo, Ludmila RB Carvalho, Lea T Grinberg, José Marcelo Farfel, Renata EL Ferretti, Renata EP Leite, Wilson Jacob Filho, Roberto Lent, and Suzana Herculano-Houzel. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, 513(5):532–541, 2009.

[AM01] Eric E Abrahamson and Robert Y Moore. Suprachiasmatic nucleus in the mouse: retinal innervation, intrinsic organization and efferent projections. *Brain research*, 916(1-2):172–191, 2001.

[Ama77] Shun-ichi Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern.*, 27(2):77–87, 1977.

[AMGF+16] John H Abel, Kirsten Meeker, Daniel Granados-Fuentes, Peter C St John, Thomas J Wang, Benjamin B Bales, Francis J Doyle, Erik D Herzog, and Linda R Petzold. Functional network inference of the suprachiasmatic nucleus. *Proc. Natl. Acad. Sci. USA*, 113(16):4512–4517, 2016.

[APDBB22] Christina Athanasouli, Sofia H Piltz, Cecilia G Diniz Behn, and Victoria Booth. Bifurcations of sleep patterns due to homeostatic and circadian variation in a sleep-wake flip-flop model. *SIAM Journal on Applied Dynamical Systems*, 21(3):1893–1929, 2022.

[AZH+21] Laith Alzubaidi, Jinglan Zhang, Amjad J Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, José Santamaría, Mohammed A Fadhel, Muthana Al-Amidie, and Laith Farhan. Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. *Journal of big Data*, 8:1–74, 2021.

[BBC00] Paul C. Bressloff, Neil W Bressloff, and Jack D. Cowan. Dynamical mechanism for sharp orientation tuning in an integrate-and-fire model of a cortical hypercolumn. *Neural computation*, 12(11):2473–2511, 2000.

[BBK22] Diane B Boivin, Philippe Boudreau, and Anastasi Kosmadopoulos. Disturbance of the circadian system in shift work and its health impact. *Journal of biological rhythms*, 37(1):3–28, 2022.

[BC13a] Michael A Buice and Carson C Chow. Beyond mean field theory: statistical field theory for neural networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(03):P03003, 2013.

[BC13b] Michael A Buice and Carson C Chow. Dynamic finite size effects in spiking neural networks. *PLoS computational biology*, 9(1):e1002872, 2013.

[BCP20] Mark Bear, Barry Connors, and Michael A Paradiso. *Neuroscience: exploring the brain, enhanced edition: exploring the brain*. Jones & Bartlett Learning, 2020.

[BD89] Roger J Broughton and David F Dinges. *Sleep and alertness: chronobiological, behavioral, and medical aspects of napping*. New York: Raven Press, 1989.

[BDFP09] Mino DC Belle, Casey O Diekman, Daniel B Forger, and Hugh D Piggins. Daily electrical silencing in the mammalian circadian clock. *Science*, 326(5950):281–284, 2009.

[BGLL08] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[BGLM20] Christian Bick, Marc Goodfellow, Carlo R Laing, and Erik A Martens. Understanding the dynamics of biological and neural oscillator networks through exact mean-field reductions: a review. *The Journal of Mathematical Neuroscience*, 10(1):9, 2020.

[BH99] Nicolas Brunel and Vincent Hakim. Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural computation*, 11(7):1621–1671, 1999.

[BM14] György Buzsáki and Kenji Mizuseki. The log-dynamic brain: how skewed distributions affect network operations. *Nat. Rev. Neurosci.*, 15(4):264–278, 2014.

[BOF+20] Áine Byrne, Reuben D O'Dea, Michael Forrester, James Ross, and Stephen Coombes. Next-generation neural mass and field modeling. *J. Neurophysiol.*, 123(2):726–742, 2020.

[BOMB+21] Beatriz Bano-Otalora, Matthew J Moye, Timothy Brown, Robert J Lucas, Casey O Diekman, and Mino DC Belle. Daily electrical activity in the master circadian clock of a diurnal mammal. *Elife*, 10:e68179, 2021.

[BOMH+21] Beatriz Bano-Otalora, Franck Martial, Court Harding, David A Bechtold, Annette E Allen, Timothy M Brown, Mino DC Belle, and Robert J Lucas. Bright daytime light enhances circadian amplitude in a diurnal mammal. *Proc. Natl. Acad. Sci. USA*, 118(22):e2100094118, 2021.

[Bör17] Christoph Börgers. *An introduction to modeling neuronal dynamics*, volume 66. Springer, 2017.

[Bor22] Alexander Borbély. The two-process model of sleep regulation: Beginnings and outlook. *Journal of Sleep Research*, 31(4):e13598, 2022.

[BP09] TM Brown and HD Piggins. Spatiotemporal heterogeneity in the electrical activity of suprachiasmatic nuclei neurons and their response to photoperiod. *J. Biol. Rhythms*, 24(1):44–54, 2009.

[Bre11] Paul C Bressloff. Spatiotemporal dynamics of continuum neural fields. *Journal of Physics A: Mathematical and Theoretical*, 45(3):033001, 2011.

[Bre14] Paul C Bressloff. Waves in neural media. *Lecture notes on mathematical modelling in the life sciences*, pages 18–19, 2014.

[Bre19] Paul C Bressloff. Stochastic neural field model of stimulus-dependent variability in cortical neurons. *PLoS Comput. Biol.*, 15(3):e1006755, 2019.

[Bro09] Korbinian Brodmann. *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Barth, 1909.

[Bru00] Nicolas Brunel. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of computational neuroscience*, 8:183–208, 2000.

[BS09] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews neuroscience*, 10(3):186–198, 2009.

[BSWB14] João Braz, Carlos Solorzano, Xidao Wang, and Allan I Basbaum. Transmitting pain and itch messages: a contemporary view of the spinal cord circuits that generate gate control. *Neuron*, 82(3):522–536, 2014.

[BSY+19] Christine Beauchene, Pierre Sacré, Fei Yang, Yun Guan, and Sridevi V Sarma. Modeling responses to peripheral nerve stimulation in the dorsal horn. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2324–2327. IEEE, 2019.

[BTDB17] Utpal Bhadra, Nirav Thakkar, Paromita Das, and Manika Pal Bhadra. Evolution of circadian rhythms: from bacteria to human. *Sleep medicine*, 35:49–61, 2017.

[Bur06a] Anthony N Burkitt. A review of the integrate-and-fire neuron model: I. homogeneous synaptic input. *Biological cybernetics*, 95:1–19, 2006.

[Bur06b] Anthony N Burkitt. A review of the integrate-and-fire neuron model: Ii. inhomogeneous synaptic input and network properties. *Biological cybernetics*, 95:97–112, 2006.

[BYBOS95] Rani Ben-Yishai, R Lev Bar-Or, and Haim Sompolinsky. Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences*, 92(9):3844–3848, 1995.

[CAY+10] Christine E Collins, David C Airey, Nicole A Young, Duncan B Leitch, and Jon H Kaas. Neuron densities vary across and within cortical areas in primates. *Proceedings of the National Academy of Sciences*, 107(36):15927–15932, 2010.

[CCDP+20] Mallory Carlu, Omar Chehab, Leonardo Dalla Porta, Damien Depannemaecker, Charlotte Héricé, Maciej Jedynak, E Köksal Ersöz, Paolo Muratore, Selma Souihel, Cristiano Capone, et al. A mean-field approach to the dynamics of networks of complex neurons, from nonlinear integrate-and-fire to hodgkin–huxley models. *J. Neurophysiol.*, 123(3):1042–1051, 2020.

[CCKC+20] Rene Y Choi, Aaron S Coyner, Jayashree Kalpathy-Cramer, Michael F Chiang, and J Peter Campbell. Introduction to machine learning, neural networks, and deep learning. *Translational Vision Science & Technology*, 9(2):14–14, 2020.

[CDH+17] Longzhen Cheng, Bo Duan, Tianwen Huang, Yan Zhang, Yangyang Chen, Olivier Britz, Lidia Garcia-Campmany, Xiangyu Ren, Linh Vong, Bradford B Lowell, et al. Identification of spinal circuits involved in touch-evoked dynamic mechanical pain. *Nature neuroscience*, 20(6):804–814, 2017.

[CG07] Anton V Chizhov and Lyle J Graham. Population model of hippocampal pyramidal neurons, linking a refractory density approach to conductance-based neurons. *Phys. Rev. E*, 75(1):011924, 2007.

[Col11] Christopher S Colwell. Linking neural activity and molecular oscillations in the scn. *Nature Reviews Neuroscience*, 12(10):553–569, 2011.

[CPHB19] Jennifer Crodelle, Sofia H Piltz, Megan Hastings Hagenauer, and Victoria Booth. Modeling the daily rhythm of human pain processing in the dorsal horn. *PLoS Comput. Biol.*, 15(7):e1007106, 2019.

[CSH+13] Andrew N Coogan, Barbora Schutová, Susanne Husung, Karolina Furczyk, Bernhard T Baune, Peter Kropp, Frank Häßler, and Johannes Thome. The circadian system in alzheimer's disease: disturbances, mechanisms, and opportunities. *Biol. Psychiatry*, 74(5):333–339, 2013.

[CVH21] Steven P Cohen, Lene Vase, and William M Hooten. Chronic pain: an update on burden, best practices, and new advances. *The Lancet*, 397(10289):2082–2097, 2021.

[CVWS19] Sarah L Chellappa, Nina Vujovic, Jonathan S Williams, and Frank AJL Scheer. Impact of circadian disruption on cardiovascular function and disease. *Trends in Endocrinology & Metabolism*, 30(10):767–779, 2019.

[DBI+13] Casey O. Diekman, Mino D. C. Belle, Robert P. Irwin, Charles N. Allen, Hugh D. Piggins, and Daniel B. Forger. Causes and consequences of hyperexcitation in central clock neurons. *PLoS Comput. Biol.*, 9(8):1–11, 08 2013.

[DCB+14] Bo Duan, Longzhen Cheng, Steeve Bourane, Olivier Britz, Christopher Padilla, Lidia Garcia-Campmany, Michael Krashes, Wendy Knowlton, Tomoko Velasquez, Xiangyu Ren, et al. Identification of spinal circuits transmitting and gating mechanical pain. *Cell*, 159(6):1417–1432, 2014.

[DCM18] Bo Duan, Longzhen Cheng, and Qiufu Ma. Spinal circuits transmitting mechanical pain and itch. *Neuroscience bulletin*, 34:186–193, 2018.

[DF03] Olivier David and Karl J Friston. A neural mass model for meg/eeg:: coupling and neuronal dynamics. *NeuroImage*, 20(3):1743–1755, 2003.

[DG20] Grégory Dumont and Pierre Gabriel. The mean-field equation of a leaky integrate-and-fire neural network: measure solutions and steady states. *Nonlinearity*, 33(12):6381, 2020.

[DJR+08] Gustavo Deco, Viktor K Jirsa, Peter A Robinson, Michael Breakspear, and Karl Friston. The dynamic brain: from spiking neurons to neural masses and cortical fields. *PLoS Comput. Biol.*, 4(8):e1000092, 2008.

[DMB+15] Daniel DeWoskin, Jihwan Myung, Mino D. C. Belle, Hugh D. Piggins, Toru Takumi, and Daniel B. Forger. Distinct roles for gaba across multiple timescales in mammalian circadian timekeeping. *Proc. Natl. Acad. Sci. USA*, 112(29):E3911–E3919, 2015.

[DSHSZ74] FH Lopes Da Silva, A Hoeks, H Smits, and LH Zetterberg. Model of brain rhythmic activity. *Kybernetik*, 15(1):27–37, 1974.

[EBD09] Sami El Boustani and Alain Destexhe. A master equation formalism for macroscopic modeling of asynchronous irregular activity states. *Neural Comput.*, 21(1):46–100, 2009.

[EKSX96] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In Evangelos Simoudis, Jiawei Han, and Usama Fayyad, editors, *KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pages 226–231. AAAI press, 1996.

[ELCCD13] Jennifer A Evans, Tanya L Leise, Oscar Castanon-Cervantes, and Alec J Davidson. Dynamic interactions mediated by nonredundant signaling mechanisms couple circadian clock neurons. *Neuron*, 80(4):973–983, 2013.

[ESYF+20] Frank Emmert-Streib, Zhen Yang, Han Feng, Shailesh Tripathi, and Matthias Dehmer. An introductory review of deep learning for prediction models with big data. *Frontiers in Artificial Intelligence*, 3:4, 2020.

[ET10] G Bard Ermentrout and David H Terman. *Mathematical foundations of neuroscience*, volume 35. Springer Science & Business Media, 2010.

[FCHC16] Diego Carlos Fernandez, Yi-Ting Chang, Samer Hattar, and Shih-Kuo Chen. Architecture of retinal projections to the central circadian pacemaker. *Proceedings of the National Academy of Sciences*, 113(21):6047–6052, 2016.

[FDR+14] Sahar Farajnia, Tom Deboer, Jos HT Rohling, Johanna H Meijer, and Stephan Michel. Aging of the suprachiasmatic clock. *Neuroscientist*, 20(1):44–55, 2014.

[FHP03] Karl J Friston, Lee Harrison, and Will Penny. Dynamic causal modelling. *Neuroimage*, 19(4):1273–1302, 2003.

[FJS16] Maria I Falcon, Viktor Jirsa, and Ana Solodkin. A new neuroinformatics approach to personalized medicine in neurology: The virtual brain. *Current opinion in neurology*, 29(4):429, 2016.

[FK17] Russell Foster and Leon Kreitzman. *Circadian rhythms: a very short introduction*. Oxford University Press, 2017.

[FMD+12] Sahar Farajnia, Stephan Michel, Tom Deboer, Henk Tjebbe vanderLeest, Thijs Houben, Jos HT Rohling, Ashna Ramkisoensing, Roman Yasenkov, and Johanna H Meijer. Evidence for neuronal desynchrony in the aged suprachiasmatic nucleus clock. *J. Neurosci.*, 32(17):5891–5899, 2012.

[Fol99] Gerald B Folland. *Real analysis: modern techniques and their applications*, volume 40. John Wiley & Sons, 1999.

[Fre75] Walter J Freeman. *Mass action in the nervous system*, volume 2004. Citeseer, 1975.

[GB23] Alexander G Ginsberg and Victoria Booth. A mean-field firing-rate model for the suprachiasmatic nucleus. *SIAM Journal on Applied Dynamical Systems*, 22(1):90–128, 2023.

[GBB11] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323. JMLR Workshop and Conference Proceedings, 2011.

[Ger00] Wulfram Gerstner. Population dynamics of spiking neurons: fast transients, asynchronous states, and locking. *Neural computation*, 12(1):43–89, 2000.

[GF11] Christian Grefkes and Gereon R Fink. Reorganization of cerebral networks after stroke: new insights from neuroimaging with connectivity approaches. *Brain*, 134(5):1264–1276, 2011.

[GR10] Diego A Golombek and Ruth E Rosenstein. Physiology of circadian entrainment. *Physiological reviews*, 90(3):1063–1102, 2010.

[GR21] Didier Gonze and Peter Ruoff. The goodwin oscillator and its legacy. *Acta Biotheoretica*, 69:857–874, 2021.

[GRPK19] Andrea G Gillman, George V Rebec, Norman C Pecoraro, and Ann EK Kosobud. Circadian entrainment by food and drugs of abuse. *Behavioural processes*, 165:23–28, 2019.

[Gül76] Fritz-H Güldner. Synaptology of the rat suprachiasmatic nucleus. *Cell Tissue Res.*, 165(4):509–544, 1976.

[GW20] Martin Golubitsky and Yangyang Wang. Infinitesimal homeostasis in three-node input–output networks. *Journal of mathematical biology*, 80(4):1163–1185, 2020.

[HA+17] Suzanne Hood, Shimon Amir, et al. The aging clock: circadian rhythms and later life. *J. Clin. Invest.*, 127(2):437–446, 2017.

[Hat02] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.

[HBF19a] Kevin M Hannay, Victoria Booth, and Daniel B Forger. Macroscopic models for human circadian rhythms. *Journal of Biological Rhythms*, 34(6):658–671, 2019.

[HBF19b] Kevin M Hannay, Victoria Booth, and Daniel B Forger. Macroscopic models for human circadian rhythms. *J. Biol. Rhythms*, 34(6):658–671, 2019.

[HH52] Alan L Hodgkin and Andrew F Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4):500, 1952.

[HK21] Yusi He and Peggy Y Kim. Allodynia. In *StatPearls [Internet]*. StatPearls Publishing, 2021.

[HKT+18] Thomas Heiberg, Birgit Kriener, Tom Tetzlaff, Gaute T Einevoll, and Hans E Plesser. Firing-rate models for neurons with a broad repertoire of spiking behaviors. *J. Comput. Neurosci.*, 45(2):103–132, 2018.

[HLVB+13] Gatien Hocepied, Benjamin Legros, Patrick Van Bogaert, Francis Grenez, and Antoine Nonclercq. Early detection of epileptic seizures based on parameter identification of neural mass model. *Computers in biology and medicine*, 43(11):1773–1782, 2013.

[HMB18] Michael H Hastings, Elizabeth S Maywood, and Marco Brancaccio. Generation of circadian rhythms in the suprachiasmatic nucleus. *Nat. Rev. Neurosci.*, 19(8):453–469, 2018.

[HSK+12] DI Hughes, S Sikander, CM Kinnon, KA Boyle, M Watanabe, RJ Callister, and BA Graham. Morphological, neurochemical and electrophysiological features of parvalbumin-expressing cells: a likely source of axo-axonic inputs in the mouse spinal dorsal horn. *The Journal of physiology*, 590(16):3927–3951, 2012.

[Isa13] Katie Isaacson. Coupled metronomes, 2013. Accessed on: 2023-07-18. URL: https://www.youtube.com/watch?v=2EAZ3VH_hNU.

[JF14] Troels S Jensen and Nanna B Finnerup. Allodynia and hyperalgesia in neuropathic pain: clinical manifestations and mechanisms. *The Lancet Neurology*, 13(9):924–935, 2014.

[JM16] Jeff R Jones and Douglas G McMahon. The core clock gene per1 phases molecular and electrical circadian rhythms in scn neurons. *PeerJ*, 4:e2297, 2016.

[JR95] Ben H Jansen and Vincent G Rit. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biological cybernetics*, 73(4):357–366, 1995.

[JTM15] Jeff R Jones, Michael C Tackenberg, and Douglas G McMahon. Manipulating circadian clock neuron firing rate resets molecular circadian rhythms and behavior. *Nat. Neurosci.*, 18(3):373–375, 2015.

[JWT+23] Viktor Jirsa, Huifang Wang, Paul Triebkorn, Meysam Hashemi, Jayant Jha, Jorge Gonzalez-Martinez, Maxime Guye, Julia Makhalova, and Fabrice Bartolomei. Personalised virtual brain models in epilepsy. *The Lancet Neurology*, 22(5):443–454, 2023.

[JYLA97] Zhi Gen Jiang, Y Yang, Zhao Ping Liu, and Charles N Allen. Membrane properties and synaptic inputs of suprachiasmatic nucleus neurons in rat brain slices. *J. Physiol.*, 499(1):141–159, 1997.

[JZB93] Ben H Jansen, George Zouridakis, and Michael E Brandt. A neurophysiologically-based mathematical model of flash visual evoked potentials. *Biological cybernetics*, 68:275–283, 1993.

[KFJ99] Richard E Kronauer, Daniel B Forger, and Megan E Jewett. Quantifying human circadian pacemaker response to brief, extended, and repeated light stimuli over the phototopic range. *Journal of biological rhythms*, 14(6):501–516, 1999.

[KKL+16] Woorim Kim, Tae Hyun Kim, Tae-Hoon Lee, Jae Woo Choi, and Eun-Cheol Park. The impact of shift and night work on health related quality of life of working women: findings from the korea health panel. *Health and quality of life outcomes*, 14:1–6, 2016.

[KM04] Sandra J Kuhlman and Douglas G McMahon. Rhythmic regulation of membrane potential and potassium current persists in scn neurons in the absence of environmental input. *European Journal of Neuroscience*, 20(4):1113–1117, 2004.

[KMS96] Bruce W Knight, Dimitri Manin, and Lawrence Sirovich. Dynamical models of interacting neuron populations in visual cortex. *Robot Cybern*, 54:4–8, 1996.

[KTRC09] Gregor Kovačič, Louis Tao, Aaditya V Rangan, and David Cai. Fokker-planck description of conductance-based integrate-and-fire neuronal networks. *Physical Review E*, 80(2):021904, 2009.

[KTSK11] Nina Khachiyants, David Trinkle, Sang Joon Son, and Kye Y Kim. Sundown syndrome in persons with dementia: an update. *Psychiatry Investig.*, 8(4):275, 2011.

[Kur75] Yoshiki Kuramoto. Self-entrainment of a population of coupled non-linear oscillators. In *International Symposium on Mathematical Problems in Theoretical Physics: January 23–29, 1975, Kyoto University, Kyoto/Japan*, pages 420–422. Springer, 1975.

[LA21] Yannuo Li and Ioannis P Androulakis. Light entrainment of the scn circadian clock and implications for personalized alterations of corticosterone rhythms in shift work and jet lag. *Scientific reports*, 11(1):17929, 2021.

[Lec17] Stefan G Lechner. An update on the spinal and peripheral pathways of pain signalling. *e-Neuroforum*, 23(3):131–136, 2017.

[LMH+19] Yue Leng, Erik S Musiek, Kun Hu, Francesco P Cappuccio, and Kristine Yaffe. Association between circadian rhythms and neurodegenerative diseases. *Lancet Neurol.*, 18(3):307–318, 2019.

[LS98] Joseph LeSauter and Rae Silver. Output signals of the scn. *Chronobiol Int.*, 15(5):535–550, 1998.

[MB89] RY Moore and Mary E Bernstein. Synaptogenesis in the rat suprachiasmatic nucleus demonstrated by electron microscopy and synapsin i immunoreactivity. *J. Neurosci.*, 9(6):2151–2162, 1989.

[MB22] Mirjam Münch and Vivien Bromundt. Light and chronobiology: implications for health and disease. *Dialogues in clinical neuroscience*, 2022.

[MBJM+99] Elizabeth L Meyer-Bernstein, Amy E Jetton, Shin-ichiro Matsumoto, Jeffrey F Markuns, Michael N Lehman, and Eric L Bittman. Effects of suprachiasmatic transplants on circadian rhythms of neuroendocrine function in golden hamsters. *Endocrinology*, 140(1):207–218, 1999.

[MCA21] Michael Moldavan, Olga Cravetchi, and Charles N Allen. Diurnal properties of tonic and synaptic gabaa receptor-mediated currents in suprachiasmatic nucleus neurons. *J. Neurophysiol.*, 126(2):637–652, 2021.

[MCRS18] Lyle Muller, Frédéric Chavane, John Reynolds, and Terrence J Sejnowski. Cortical travelling waves: mechanisms and computational principles. *Nature Reviews Neuroscience*, 19(5):255–268, 2018.

[MDG02] Maurizio Mattia and Paolo Del Giudice. Population dynamics of interacting spiking neurons. *Physical Review E*, 66(5):051917, 2002.

[MGR86] Johanna H Meijer, Gerard A Groos, and Benjamin Rusak. Luminance coding in a circadian pacemaker: the suprachiasmatic nucleus of the rat and the hamster. *Brain Res.*, 382(1):109–118, 1986.

[MH94] V Mavjee and JA Home. Boredom effects on sleepiness/alertness in the early afternoon vs. early evening and interactions with warm ambient temperature. *Br. J. Psychol.*, 85(3):317–333, 1994.

[MKK+02] Kimberly A Moore, Tatsuro Kohno, Laurie A Karchewski, Joachim Scholz, Hiroshi Baba, and Clifford J Woolf. Partial peripheral nerve injury promotes a selective loss of gabaergic inhibition in the superficial dorsal horn of the spinal cord. *Journal of Neuroscience*, 22(15):6724–6731, 2002.

[MM97] Robert Y Moore MD. Circadian rhythms: basic neurobiology and clinical applications. *Annual review of medicine*, 48(1):253–266, 1997.

[MPF13] Rosalyn Moran, Dimitris A Pinotsis, and Karl Friston. Neural masses and fields in dynamic causal modeling. *Frontiers in computational neuroscience*, 7:57, 2013.

[MPR15] Ernest Montbrió, Diego Pazó, and Alex Roxin. Macroscopic description for networks of spiking neurons. *Phys. Rev. X*, 5(2):021028, 2015.

[MRG92] Johanna H Meijer, Benjamin Rusak, and Gabriëlla Gänshirt. The relation between light-induced discharge in the suprachiasmatic nucleus and phase shifts of hamster circadian rhythms. *Brain Res.*, 598(1-2):257–263, 1992.

[MS93] Robert Y Moore and Joan C Speh. Gaba is the principal neurotransmitter of the circadian system. *Neurosci. Lett.*, 150(1):112–116, 1993.

[MSH+22] Laura Medlock, Kazutaka Sekiguchi, Sungho Hong, Salvador Dura-Bernal, William W Lytton, and Steven A Prescott. Multiscale computer model of the spinal dorsal horn reveals changes in network processing associated with chronic pain. *Journal of Neuroscience*, 42(15):3133–3149, 2022.

[MSS+04] Igor V Melnick, Sónia FA Santos, Karolina Szokol, Péter Szûcs, and Boris V Safronov. Ionic basis of tonic firing in spinal substantia gelatinosa neurons of rat. *Journal of neurophysiology*, 91(2):646–655, 2004.

[MSWA97] JH Meijer, J Schaap, K Watanabe, and H Albus. Multiunit activity recordings in the suprachiasmatic nuclei: in vivo versus in vitro models. *Brain Res.*, 753(2):322–327, 1997.

[MW65] Ronald Melzack and Patrick D Wall. Pain mechanisms: A new theory: A gate control system modulates sensory input from the skin before it evokes pain perception and response. *Science*, 150(3699):971–979, 1965.

[MWM+06] Andrea L Meredith, Steven W Wiler, Brooke H Miller, Joseph S Takahashi, Anthony A Fodor, Norman F Ruby, and Richard W Aldrich. Bk calcium-activated potassium channels regulate circadian behavioral rhythms and pacemaker output. *Nature neuroscience*, 9(8):1041–1049, 2006.

[MZJS00] Maciej Mrugala, Piotr Zlomanczuk, Anita Jagota, and William J Schwartz. Rhythmic multiunit neural activity in slices of hamster suprachiasmatic nucleus reflect prior photoperiod. *Am. J. Physiol. Regul. Integr. Comp. Physiol.*, 278(4):R987–R994, 2000.

[NAN+03] Mamoru Nagano, Akihito Adachi, Ken-ichi Nakahama, Toru Nakamura, Masako Tamada, Elizabeth Meyer-Bernstein, Amita Sehgal, and Yasufumi Shigeyoshi. An abrupt shift in the day/night cycle causes desynchrony in the mammalian circadian center. *J. Neurosci.*, 23(14):6141–6151, 2003.

[New18] Mark Newman. *Networks*. Oxford university press, 2018.

[NKS+18] Evangelia Nena, Maria Katsaouni, Paschalis Steiropoulos, Evangelos Theodorou, Theodoros C Constantinidis, and Grigorios Tripsianis. Effect of shift work on sleep, health, and quality of life of health-care workers. *Indian journal of occupational and environmental medicine*, 22(1):29, 2018.

[NNT+15] Takahiro J Nakamura, Wataru Nakamura, Isao T Tokuda, Takahiro Ishikawa, Takashi Kudo, Christopher S Colwell, and Gene D Block. Age-related changes in the circadian system unmasked by constant conditions. *eNeuro*, 2(4), 2015.

[NT00] Duane Q Nykamp and Daniel Tranchina. A population density approach that facilitates large-scale modeling of neural networks: Analysis and an application to orientation tuning. *J. Comput. Neurosci.*, 8(1):19–50, 2000.

[OA08] Edward Ott and Thomas M Antonsen. Low dimensional behavior of large systems of globally coupled oscillators. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 18(3):037113, 2008.

[OB11] Srdjan Ostojic and Nicolas Brunel. From spiking neuron models to linear-nonlinear models. *PLoS computational biology*, 7(1):e1001056, 2011.

[OKS00] Ahmet Omurtag, Bruce W. Knight, and Lawrence Sirovich. On the simulation of large populations of neurons. *Journal of computational neuroscience*, 8(1):51–63, 2000.

[PBT+19] Hugues Petitjean, Farin B Bourojeni, Deborah Tsao, Albena Davidova, Susana G Sotocinal, Jeffrey S Mogil, Artur Kania, and Reza Sharif-Naeini. Recruitment of spinoparabrachial neurons by dorsal horn calretinin neurons. *Cell reports*, 28(6):1429–1438, 2019.

[PdJB+02] Cyriel MA Pennartz, Marcel TG de Jeu, Nico PA Bos, Jeroen Schaap, and Alwin MS Geurtsen. Diurnal modulation of pacemaker potentials and calcium current in the mammalian circadian clock. *Nature*, 416(6878):286–290, 2002.

[PDT20] Cedric Peirs, Radhouane Dallel, and Andrew J Todd. Recent advances in our understanding of the organization of dorsal horn neuron populations and their contribution to cutaneous mechanical allodynia. *Journal of neural transmission*, 127(4):505–525, 2020.

[PGS20] Bastian Pietras, Noé Gallice, and Tilo Schwalger. Low-dimensional firing-rate dynamics for populations of renewal-type spiking neurons. *Physical Review E*, 102(2):022407, 2020.

[PMvGG18] Jodi R Paul, Hira A Munir, Thomas van Groen, and Karen L Gamble. Behavioral and scn neurophysiological disruption in the tg-swdi mouse model of alzheimer's disease. *Neurobiol. Dis.*, 114:194–200, 2018.

[POM06] Gilbert R Pitts, Hidenobu Ohta, and Douglas G McMahon. Daily rhythmicity of large-conductance ca2+-activated k+ currents in suprachiasmatic nucleus neurons. *Brain research*, 1071(1):54–62, 2006.

[PPF+15] Hugues Petitjean, Sophie Anne Pawlowski, Steven Li Fraine, Behrang Sharif, Doulia Hamad, Tarheen Fatima, Jim Berg, Claire M Brown, Lily-Yeh Jan, Alfredo Ribeiro-da Silva, et al. Dorsal horn parvalbumin neurons are gate-keepers of touch-evoked pain after nerve injury. *Cell reports*, 13(6):1246–1257, 2015.

[PR07] AJK Phillips and Peter A Robinson. A quantitative model of sleep-wake dynamics based on the physiology of the brainstem ascending arousal system. *Journal of Biological Rhythms*, 22(2):167–179, 2007.

[PR22] Gokul Parameswaran and David W Ray. Sleep, circadian rhythms, and type 2 diabetes mellitus. *Clinical endocrinology*, 96(1):12–20, 2022.

[PRbGF14] Dimitris Pinotsis, Peter Robinson, Peter beim Graben, and Karl Friston. Neural masses and fields: modeling the dynamics of brain activity, 2014.

[PWZ+15] Cedric Peirs, Sean-Paul G Williams, Xinyi Zhao, Claire E Walsh, Jeremy Y Gedeon, Natalie E Cagle, Adam C Goldring, Hiroyuki Hioki, Zheng Liu, Paulina S Marell, et al. Dorsal horn circuits for persistent mechanical pain. *Neuron*, 87(4):797–812, 2015.

[Raj61] C Rajski. A metric space of discrete probability distributions. *Information and Control*, 4(4):371–377, 1961.

[RBHK21] Xin Ren, Anastasia Brodovskaya, John L Hudson, and Jaideep Kapur. Connectivity and neuronal synchrony during seizures. *Journal of Neuroscience*, 41(36):7623–7635, 2021.

[RS02] Ruth Ruscheweyh and Jürgen Sandkühler. Lamina-specific membrane and discharge properties of rat spinal dorsal horn neurones in vitro. *The Journal of physiology*, 541(1):231–244, 2002.

[RSSGJ23] S Michaela Rikard, Andrea E Strahan, Kristine M Schmit, and Gery P Guy Jr. Chronic pain among adults—united states, 2019–2021. *Morbidity and Mortality Weekly Report*, 72(15):379, 2023.

[RTT19] M Reza Rahimi Tabar and M Reza Rahimi Tabar. Kramers–moyal expansion and fokker–planck equation. *Analysis and Data-Based Reconstruction of Complex Nonlinear Dynamical Systems: Using the Methods of Stochastic Processes*, pages 19–29, 2019.

[RWVSB99] Fred Rieke, David Warland, Rob de Ruyter Van Steveninck, and William Bialek. *Spikes: exploring the neural code*. MIT press, 1999.

[San09] Jurgen Sandkuhler. Models and mechanisms of hyperalgesia and allodynia. *Physiological reviews*, 89(2):707–758, 2009.

[SC19] Tilo Schwalger and Anton V Chizhov. Mind the last spike—firing rate models for mesoscopic populations of spiking neurons. *Current opinion in neurobiology*, 58:155–166, 2019.

[SCH+19] Martin Seeber, Lucia-Manuela Cantonas, Mauritius Hoevels, Thibaut Sesia, Veerle Visser-Vandewalle, and Christoph M Michel. Subcortical electrophysiological activity is detectable with high-density eeg source imaging. *Nature communications*, 10(1):753, 2019.

[SDG17] Tilo Schwalger, Moritz Deger, and Wulfram Gerstner. Towards a theory of cortical columns: From spiking neurons to interacting neural populations of finite size. *PLoS computational biology*, 13(4):e1005507, 2017.

[Sim12] Lee S Simon. Relieving pain in america: A blueprint for transforming prevention, care, education, and research. *Journal of pain & palliative care pharmacotherapy*, 26(2):197–198, 2012.

[SK19] Ayesha A Shafi and Karen E Knudsen. Cancer and the circadian clock. *Cancer research*, 79(15):3806–3814, 2019.

[SLCG05] Clifford B Saper, Jun Lu, Thomas C Chou, and Joshua Gooley. The hypothalamic integrator for circadian rhythms. *Trends in neurosciences*, 28(3):152–157, 2005.

[SLTL96] Rae Silver, Joseph LeSauter, Patrick A Tresco, and Michael N Lehman. A diffusible coupling signal from the transplanted suprachiasmatic nucleus controlling circadian locomotor rhythms. *Nature*, 382(6594):810–813, 1996.

[SMW+17] Adam R Stinchcombe, Joshua W Mouland, Kwoon Y Wong, Robert J Lucas, and Daniel B Forger. Multiplexing visual signals in the suprachiasmatic nuclei. *Cell Rep.*, 21(6):1418–1425, 2017.

[ST08] Vladimir Shusterman and William C Troy. From baseline to epileptiform activity: a path to synchronized rhythmicity in large-scale neural networks. *Physical Review E*, 77(6):061911, 2008.

[SZT20] Yuxiu Shao, Jiwei Zhang, and Louis Tao. Dimensional reduction of emergent spatiotemporal cortical dynamics via a maximum entropy moment closure. *PLoS computational biology*, 16(6):e1007265, 2020.

[TC10] Hsinlin Thomas Cheng. Spinal cord mechanisms of chronic pain and clinical implications. *Current pain and headache reports*, 14:213–220, 2010.

[THR04] T Tateno, A Harsch, and HPC Robinson. Threshold firing frequency–current relationships of neurons in rat somatosensory cortex: type 1 and type 2 dynamics. *Journal of neurophysiology*, 92(4):2283–2294, 2004.

[Tod10] Andrew J Todd. Neuronal circuitry for pain processing in the dorsal horn. *Nature Reviews Neuroscience*, 11(12):823–836, 2010.

[UPL+09] Peter Uhlhaas, Gordon Pipa, Bruss Lima, Lucia Melloni, Sergio Neuenschwander, Danko Nikolić, and Wolf Singer. Neural synchrony in cortical networks: history, concept and current status. *Frontiers in integrative neuroscience*, page 17, 2009.

[VHAK+14] OASKR Van Hecke, Sophie K Austin, Rafi A Khan, BH Smith, and N Torrance. Neuropathic pain in the general population: a systematic review of epidemiological studies. *PAIN®*, 155(4):654–662, 2014.

[VHM+07] Henk Tjebbe VanderLeest, Thijs Houben, Stephan Michel, Tom Deboer, Henk Albus, Mariska J Vansteensel, Gene D Block, and Johanna H Meijer. Seasonal encoding by the circadian pacemaker of the scn. *Curr. Biol.*, 17(5):468–473, 2007.

[vOLH+12] Floor van Oosterhout, Eliane A Lucassen, Thijs Houben, Henk Tjebbe vanderLeest, Michael C Antle, and Johanna H Meijer. Amplitude of the scn clock enhanced by the behavioral activity rhythm. *PLoS One*, 7(6):e39693, 2012.

[WC72] Hugh R Wilson and Jack D Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.*, 12(1):1–24, 1972.

[WC73] Hugh R Wilson and Jack D Cowan. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, 13(2):55–80, 1973.

[WCW14] Ming-Hui Wang, Na Chen, and Jin-Hui Wang. The coupling features of electrical synapses modulate neuronal synchrony in hypothalamic superachiasmatic nucleus. *Brain Res.*, 1550:9–17, 2014.

[WDMR06] Marc Wittmann, Jenny Dinich, Martha Merrow, and Till Roenneberg. Social jetlag: misalignment of biological and social time. *Chronobiology international*, 23(1-2):497–509, 2006.

[Wes22] Karin N Westlund. Neurophysiology of pain: Peripheral, spinal, ascending, and descending pathways. In *Practical management of pain*, pages 95–109. Elsevier, 2022.

[Woo22] Clifford J Woolf. Pain modulation in the spinal cord. *Frontiers in Pain Research*, 3:984042, 2022.

[WSC92] Clifford J Woolf, Peter Shortland, and Richard E Coggeshall. Peripheral nerve injury triggers central sprouting of myelinated afferents. *Nature*, 355(6355):75–78, 1992.

[WSW95] Akihito Watanabe, Shigenobu Shibata, and Shigenori Watanabe. Circadian rhythm of spontaneous neuronal activity in the suprachiasmatic nucleus of old hamster in vitro. *Brain Res.*, 695(2):237–239, 1995.

[YKL+07] Lily Yan, Ilia Karatsoreos, Joseph LeSauter, DK Welsh, S Kay, D Foley, and R Silver. Exploring spatiotemporal organization of scn circuits. In *Cold Spring Harbor symposia on quantitative biology*, volume 72, pages 527–541. Cold Spring Harbor Laboratory Press, 2007.

[YTLS21] Yifan Yao, Alana B'nai Taub, Joseph LeSauter, and Rae Silver. Identification of the suprachiasmatic nucleus venous portal system in the mammalian brain. *Nature Communications*, 12(1):5643, 2021.

[ZJG14] Tianhe C Zhang, John J Janik, and Warren M Grill. Modeling effects of spinal cord stimulation on wide-dynamic range dorsal horn neurons: influence of stimulation frequency and gabaergic inhibition. *Journal of neurophysiology*, 112(3):552–567, 2014.

[ZVvPTH14] Bas-Jan Zandt, Sid Visser, Michel van Putten, and Bennie Ten Haken. A neural mass model based on single cell dynamics to model pathophysiology. *J. Comput. Neurosci.*, 37, 08 2014.